SIAM J. CONTROL OPTIM.
Vol. 42, No. 2, pp. 596–636

BREGMAN MONOTONE OPTIMIZATION ALGORITHMS*

HEINZ H. BAUSCHKE†, JONATHAN M. BORWEIN‡, AND PATRICK L. COMBETTES§

Abstract. A broad class of optimization algorithms based on Bregman distances in Banach spaces is unified around the notion of Bregman monotonicity. A systematic investigation of this notion leads to a simplified analysis of numerous algorithms and to the development of a new class of parallel block-iterative surrogate Bregman projection schemes. Another key contribution is the introduction of a class of operators that is shown to be intrinsically tied to the notion of Bregman monotonicity and to include the operators commonly found in Bregman optimization methods. Special emphasis is placed on the viability of the algorithms and the importance of Legendre functions in this regard. Various applications are discussed.

Key words. Banach space, block-iterative method, Bregman distance, Bregman monotone, Bregman projection, $\mathfrak{B}$-class operator, convex feasibility problem, essentially smooth function, essentially strict convex function, Fejér monotone, Legendre function, monotone operator, proximal mapping, proximal point algorithm, resolvent, subgradient projection

AMS subject classifications. 90C25, 90C48, 47H05

PII. S0363012902407120

1. Introduction.

A sequence $(x_n)_{n\in\mathbb{N}}$ in a Banach space $\mathcal{X}$ is *Fejér monotone* with respect to a set $S \subset \mathcal{X}$ if

$$(\forall x \in S)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\| \le \|x_n - x\|. \tag{1.1}$$

In Hilbert spaces, this notion has proven to be remarkably useful and successful in attempts to unify and harmonize the convergence proofs of a large number of optimization algorithms; see, e.g., [5, 6, 9, 40, 41, 49, 60]. A classical example is the method of cyclic projections for finding a point in the intersection $S \neq \emptyset$ of a finite family of closed convex sets $(S_i)_{1\le i\le m}$. In 1965, Bregman [14, Thm. 1] showed that for every initial point $x_0 \in \mathcal{X}$ the sequence $(x_n)_{n\in\mathbb{N}}$ generated by the cyclic projections algorithm

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = P_{n\,(\mathrm{mod}\,m)+1}x_n, \tag{1.2}$$

where P_i denotes the metric projector onto S_i and where the mod m function takes values in $\{0, \dots, m-1\}$, is Fejér monotone with respect to S and converges weakly to a point in that set. Two years later [15], the same author investigated the convergence of this method in a general topological vector space $\mathcal{X}$. To this end, he introduced a distance-like function $D\colon E \times E \to \mathbb{R}$, where E is a convex subset of $\mathcal{X}$ such that $S = E \cap \bigcap_{i=1}^m S_i \neq \emptyset$. The conditions defining D require, in particular, that for

*Received by the editors May 6, 2002; accepted for publication (in revised form) January 8, 2003; published electronically May 29, 2003.
http://www.siam.org/journals/sicon/42-2/40712.html
†Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario N1G 2W1, Canada (hbauschk@uoguelph.ca). This author's research was supported by the Natural Sciences and Engineering Research Council of Canada.
‡Centre for Experimental & Constructive Mathematics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada (jborwein@cecm.sfu.ca). This author's research was supported by the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chair Programme.
§Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie – Paris 6, 75005 Paris, France (plc@math.jussieu.fr).

every $i \in \{1, \ldots, m\}$ and every $y \in E$, there exists a point $P_i y \in E \cap S_i$ such that $D(P_i y, y) = \min D(E \cap S_i, y)$. In this broader context, Bregman showed that for every initial point $x_0 \in E$ the cyclic projections algorithm (1.2) produces a sequence that satisfies the monotonicity property

$$(\forall x \in S)(\forall n \in \mathbb{N}) \quad D(x, x_{n+1}) \leq D(x, x_n) \tag{1.3}$$

and whose cluster points are in S [15, eq. (1.2) and Thm. 1]. If $\mathcal{X}$ is a Hilbert space, an example of a D-function satisfying the required conditions relative to the weak topology is $D\colon \mathcal{X}^2 \to \mathbb{R}\colon (x, y) \mapsto \|x - y\|^2/2$. In this case, we recover the previous convergence result [15, Example 1] and observe that (1.3) reduces to (1.1). If $\mathcal{X}$ is the Euclidean space $\mathbb{R}^N$, another example of a suitable D-function is

$$D\colon E \times E \to \mathbb{R}\colon (x, y) \mapsto f(x) - f(y) - \langle x - y, \nabla f(y) \rangle, \tag{1.4}$$

where $f\colon E \subset \mathbb{R}^N \to \mathbb{R}$ is a convex function which is differentiable on E and satisfies a set of auxiliary properties [15, Example 2]. Due to its importance in applications, this particular type of D-function was further studied in [30] and has since been known as a *Bregman distance* (see [33] for an historical account). In $\mathbb{R}^N$, various investigations have focused on the use of Bregman distances in projection, proximal point, and fixed point algorithms; see [7, 31, 32, 33, 46, 47, 83]. (See also [58, 59], where extensions of (1.4) to nondifferentiable functions were studied.) Extensions to Hilbert [18, 20, 61] and Banach [1, 8, 21, 23, 24, 25, 26, 27, 55, 56, 75] spaces have also been considered more recently. In the present paper, we adopt the following definition for Bregman distances.

DEFINITION 1.1. *Let $\mathcal{X}$ be a real Banach space and let $f\colon \mathcal{X} \to]-\infty, +\infty]$ be a lower semicontinuous convex function which is Gâteaux-differentiable on* $\operatorname{int} \operatorname{dom} f \neq \varnothing$. *The* Bregman distance *(for brevity* D-distance*) associated with f is the function*

$$D\colon \mathcal{X} \times \mathcal{X} \to [0, +\infty], \qquad (x, y) \mapsto \begin{cases} f(x) - f(y) - \langle x - y, \nabla f(y) \rangle & \text{if } y \in \operatorname{int} \operatorname{dom} f, \\ +\infty & \text{otherwise.} \end{cases} \tag{1.5}$$

In addition, the Bregman distance to a set $C \subset \mathcal{X}$ is the function

$$D_C\colon \mathcal{X} \to [0, +\infty], \quad y \mapsto \inf D(C, y). \tag{1.6}$$

In Hilbert spaces, one recovers $D\colon (x, y) \mapsto \|x - y\|^2/2$ by setting $f = \|\cdot\|^2/2$. This observation suggests that the following natural variant of the notion of Fejér monotonicity suits the environment described in Definition 1.1.

DEFINITION 1.2. *A sequence $(x_n)_{n \in \mathbb{N}}$ in $\mathcal{X}$ is* Bregman monotone *(for brevity* D-monotone*) with respect to a set $S \subset \mathcal{X}$ if the following conditions hold:*

(i) $S \cap \operatorname{dom} f \neq \varnothing$.
(ii) $(x_n)_{n \in \mathbb{N}}$ *lies in* $\operatorname{int} \operatorname{dom} f$.
(iii) $(\forall x \in S \cap \operatorname{dom} f)(\forall n \in \mathbb{N}) \;\; D(x, x_{n+1}) \leq D(x, x_n)$.

Let us note that item (ii) is stated only for the sake of clarity and that it could be replaced by $x_0 \in \operatorname{int} \operatorname{dom} f$ since, in view of (1.5), (iii) then forces the whole sequence $(x_n)_{n \in \mathbb{N}}$ to lie in $\operatorname{int} \operatorname{dom} f$.

The importance of the notion of Bregman monotonicity is implicit in [15]. In the Euclidean space setting of [32] (see also [33, page 55]), Bregman monotone sequences were called "D_f Fejér monotone" by analogy with (1.1).

The goal of this paper is to provide a broad framework for the design and the analysis of algorithms based on Bregman distances around the notion of D-monotonicity. This framework not only will lead to a unified convergence analysis for existing algorithms, but also will serve as a basis for the development of a new class of parallel, block-iterative, surrogate Bregman projection methods for solving convex feasibility problems involving variational inequalities, convex inequalities, equilibrium constraints, and fixed point constraints. The tools developed in this paper also provide the main building blocks for the algorithms proposed in [10] to find best Bregman approximations from intersections of closed convex sets in reflexive Banach spaces.

Guide to the paper. We proceed towards our goal of constructing a broad framework for Bregman distance-based algorithms in several steps.

We collect assumptions, notation, and basic results in section 2. The standing assumptions on the underlying space $\mathcal{X}$ and the function f that generates the Bregman distance are stated in section 2.1. In sections 2.2–2.6, we introduce basic notation and terminology, including D-viable operators and Legendre functions. Useful identities for the Bregman distance are provided in section 2.7.

A general and powerful class of operators based on Bregman distances is introduced and analyzed in section 3. This so-called $\mathfrak{B}$-class includes types of operators fundamental in Bregman optimization such as D-firm operators, D-resolvents, D-prox operators, and (subgradient) D-projections, which correspond to their classical counterparts when $\mathcal{X}$ is a Hilbert space and $f = \|\cdot\|^2/2$. For example, it is shown that if $\mathcal{X}$ is reflexive and f is Legendre, then D-prox operators belong to $\mathfrak{B}$ (Corollary 3.25). This result underscores the importance of Legendreness. Moreover, $\mathfrak{B}$-class operators are stable under a certain type of parallel combination, which will be crucial in the formulation of a new block-iterative algorithmic framework in section 5.

Section 4 is devoted to D-monotonicity. This is a central notion in the analysis of Bregman optimization methods because it describes the behavior of a wide class of algorithms based on Bregman distances. Assumptions are given under which simple characterizations can be established for the weak and strong convergence of D-monotone sequences. In conjunction with the results of section 3, D-monotonicity provides a global framework for the development and analysis of algorithms. Indeed, we show that D-monotone sequences can be generated systematically via the iterative scheme

$$x_0 \in \operatorname{int}\operatorname{dom} f \;\text{ and }\; (\forall n \in \mathbb{N})\;\; x_{n+1} \in T_n x_n, \;\text{ where }\; T_n \in \mathfrak{B}. \tag{1.7}$$

A detailed convergence analysis of this unifying model is carried out which, in turn, covers and extends known convergence results.

Finally, in section 5, we are in a position to construct a new block-iterative algorithmic framework. Results obtained in sections 3 and 4 are combined to construct and investigate new classes of parallel, block-iterative methods for solving convex feasibility problems. The main result, Theorem 5.7, provides conditions sufficient for the weak and strong convergence of sequences generated by the new algorithm. Section 5.4 presents several scenarios in which these sufficient conditions are satisfied, including the frequently encountered situation when f is a separable Legendre function on $\mathbb{R}^N$ such that $\operatorname{dom} f^*$ is open (Example 5.14). The concluding sections, sections 5.5 and 5.6, discuss how the main result can be applied to specific optimization problems such as solving convex inequalities, finding common zeros of maximal monotone operators, finding common minimizers of convex function, and finding common fixed points of D-firm operators.

2. Notation, assumptions, and basic facts.

2.1. Standing assumptions. We assume throughout the paper that $\mathcal{X}$ is a real Banach space and that $f\colon \mathcal{X} \to \left]-\infty, +\infty\right]$ is a lower semicontinuous convex function which is Gâteaux-differentiable on $\operatorname{int}\operatorname{dom} f \neq \varnothing$.

2.2. Basic notation. Throughout, $\mathbb{N}$ is the set of nonnegative integers. The norm of $\mathcal{X}$ and that of its topological dual $\mathcal{X}^*$ is denoted by $\|\cdot\|$, the associated metric distance by d, and the canonical bilinear form on $\mathcal{X} \times \mathcal{X}^*$ by $\langle\cdot,\cdot\rangle$. (If $\mathcal{X}$ is a Hilbert space, $\langle\cdot,\cdot\rangle$ denotes also its scalar (or inner) product.) The metric distance function to a set $C \subset \mathcal{X}$ is $d_C\colon \mathcal{X} \to [0, +\infty] : y \mapsto \inf_{x \in C} \|x - y\|$ where, by convention, $\inf \varnothing = +\infty$. For every $y \in \operatorname{int}\operatorname{dom} f$, we set $f_y = f - \nabla f(y)$. The symbols $\rightharpoonup$, $\overset{*}{\rightharpoonup}$, and $\to$ denote, respectively, weak, weak$*$, and strong convergence. $\mathfrak{S}(x_n)_{n\in\mathbb{N}}$ and $\mathfrak{W}(x_n)_{n\in\mathbb{N}}$ are, respectively, the sets of strong and weak cluster points of a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$. $\operatorname{bdry} C$ denotes the boundary of a set $C \subset \mathcal{X}$, $\operatorname{int} C$ its interior, and $\overline{C}$ its closure. The closed ball of center x and radius ρ is denoted by $B(x;\rho)$. The normalized duality mapping J of $\mathcal{X}$ is defined by

$$(\forall x \in \mathcal{X}) \quad J(x) = \big\{x^* \in \mathcal{X}^* \mid \|x\|^2 = \langle x, x^*\rangle = \|x^*\|^2\big\}. \tag{2.1}$$

$\mathbb{R}^N$ is the standard N-dimensional Euclidean space.

2.3. Set-valued operators. Let $\mathcal{Y}$ be a Banach space and $2^{\mathcal{Y}}$ the family of all subsets of $\mathcal{Y}$. A set-valued operator from $\mathcal{X}$ to $\mathcal{Y}$ is an operator $A\colon \mathcal{X} \to 2^{\mathcal{Y}}$. It is characterized by its graph $\operatorname{gr} A = \{(x,u) \in \mathcal{X} \times \mathcal{Y} \mid u \in Ax\}$; its domain is $\operatorname{dom} A = \{x \in \mathcal{X} \mid Ax \neq \varnothing\}$ (with closure $\overline{\operatorname{dom}}\,A$); its range is $\operatorname{ran} A = \bigcup_{x\in\mathcal{X}} Ax$ (with closure $\overline{\operatorname{ran}}\,A$); and, if $\mathcal{Y} = \mathcal{X}$, its fixed point set is $\operatorname{Fix} A = \{x \in \mathcal{X} \mid x \in Ax\}$ (with closure $\overline{\operatorname{Fix}}\,A$). The graph of the inverse A^{-1} of A is $\{(u,x) \in \mathcal{Y} \times \mathcal{X} \mid (x,u) \in \operatorname{gr} A\}$. If $B\colon \mathcal{X} \to 2^{\mathcal{Y}}$ and $\alpha \in \mathbb{R}$, then $\operatorname{gr}(\alpha A + B) = \{(x, \alpha u + v) \in \mathcal{X} \times \mathcal{Y} \mid (x,u) \in \operatorname{gr} A, (x,v) \in \operatorname{gr} B\}$. As is customary, if $x \in \operatorname{dom} A$ and A is single-valued on $\operatorname{dom} A$, we shall denote the unique element in Ax by Ax. Finally, A is locally bounded at $x \in \mathcal{X}$ if there exists $\rho \in \left]0, +\infty\right[$ such that $A\big(B(x;\rho)\big)$ is bounded. (We adopt the same definition as in [79, section 17]; it differs slightly from Phelps' definition [71, Chap. 2] which requires $x \in \operatorname{dom} A$.)

2.4. Orbits and suborbits of algorithms. In section 4 and subsequent sections, we shall discuss various algorithms. Sequences generated by algorithms are called orbits, and their subsequences are referred to as suborbits.

2.5. Functions. The domain of a function $g\colon \mathcal{X} \to \left]-\infty, +\infty\right]$ is $\operatorname{dom} g = \{x \in \mathcal{X} \mid g(x) < +\infty\}$ (with closure $\overline{\operatorname{dom}}\,g$), and g is proper if $\operatorname{dom} g \neq \varnothing$. Moreover, g is subdifferentiable at $x \in \operatorname{dom} g$ if its subdifferential at this point,

$$\partial g(x) = \big\{x^* \in \mathcal{X}^* \mid (\forall y \in \mathcal{X}) \;\; \langle y - x, x^*\rangle + g(x) \leq g(y)\big\}, \tag{2.2}$$

is not empty; a subgradient of g at x is an element of $\partial g(x)$. The domain of continuity of g is

$$\operatorname{cont} g = \big\{x \in \mathcal{X} \mid \; |g(x)| < +\infty \text{ and } g \text{ is continuous at } x\big\}, \tag{2.3}$$

and its lower level set at height $\eta \in \mathbb{R}$ is $\operatorname{lev}_{\leq\eta} g = \{x \in \mathcal{X} \mid g(x) \leq \eta\}$. Recall that the value of g^*, the conjugate of g, at point $x^* \in \mathcal{X}^*$ is defined by

$$g^*(x^*) = \sup_{x\in\mathcal{X}} \langle x, x^*\rangle - g(x); \tag{2.4}$$

g is cofinite if $\operatorname{dom} g^* = \mathcal{X}^*$. Furthermore, g is coercive if $\lim_{\|x\|\to+\infty} g(x) = +\infty$, supercoercive if $\lim_{\|x\|\to+\infty} g(x)/\|x\| = +\infty$, (weak) lower semicontinuous if its lower level sets $\left(\operatorname{lev}_{\leq\eta} g\right)_{\eta\in\mathbb{R}}$ are (weakly) closed, and (weak) inf-compact if they are (weakly) compact. If $\mathcal{X}$ is reflexive, the notions of weak inf-compactness and coercivity coincide for weak lower semicontinuous functions. The set of minimizing sequences of g is denoted by

$$\mathcal{M}(g) = \left\{(x_n)_{n\in\mathbb{N}} \text{ in } \operatorname{dom} g \mid g(x_n) \to \inf g(\mathcal{X})\right\} \tag{2.5}$$

and the set of global minimizers of g by $\operatorname{Argmin} g$. (If it is a singleton, its unique element is denoted by $\operatorname{argmin} g$.) The inf-convolution of two functions $g_1, g_2 \colon \mathcal{X} \to \left]-\infty, +\infty\right]$ is $g_1 \,\square\, g_2 \colon \mathcal{X} \to [-\infty, +\infty] : x \mapsto \inf_{y\in\mathcal{X}} g_1(y) + g_2(x-y)$.

The indicator function of a set $C \subset \mathcal{X}$ is the function $\iota_C \colon \mathcal{X} \to \{0, +\infty\}$ that takes value 0 on C and $+\infty$ on its complement, and its normal cone is

$$N_C = \partial\iota_C \colon \mathcal{X} \to 2^{\mathcal{X}^*} : x \mapsto \begin{cases} \left\{x^* \in \mathcal{X}^* \mid (\forall y \in C) \quad \langle y - x, x^*\rangle \leq 0\right\} & \text{if } x \in C, \\ \varnothing & \text{otherwise.} \end{cases} \tag{2.6}$$

2.6. *D*-viability and Legendre functions. Operators based on Bregman distances are not defined outside of $\operatorname{int}\operatorname{dom} f$. Thus, using the terminology of [3], for an algorithm such as (1.7) to be viable in the sense that its iterates remain in $\operatorname{int}\operatorname{dom} f$, the operators involved must satisfy the following viability condition.

DEFINITION 2.1. *An operator $T \colon \mathcal{X} \to 2^{\mathcal{X}}$ is D-*viable *if* $\operatorname{ran} T \subset \operatorname{dom} T = \operatorname{int}\operatorname{dom} f$.

It was shown in [7] that a sufficient condition for Bregman projection operators onto closed convex sets in Euclidean spaces to be D-viable is that f be a Legendre function. (In this context, "D-viability" was called "zone consistency" after [30].) The classical finite-dimensional definition of a Legendre function, as introduced by Rockafellar in [77, section 26], is of limited use in general Banach spaces since the resulting class of functions loses some of its remarkable finite-dimensional properties. In the context of Banach spaces, we introduced in [8] the following notion a Legendre function. It not only generalizes Rockafellar's classical definition but also preserves its salient properties in reflexive spaces. (For results on Legendre functions in nonreflexive spaces, see [13].)

DEFINITION 2.2 ([8, Def. 5.2]). *The function f is*

(i) essentially smooth *if ∂f is both locally bounded and single-valued on its domain;*
(ii) essentially strictly convex *if $(\partial f)^{-1}$ is locally bounded on its domain and f is strictly convex on every convex subset of* $\operatorname{dom}\partial f$*;*
(iii) Legendre *if it is both essentially smooth and essentially strictly convex.*

Such functions will be of prime importance in our analysis as they will be shown to provide a simple and convenient sufficient condition for the D-viability of the operators commonly encountered in Bregman optimization methods in Banach spaces.

2.7. Basic properties of Bregman distances. The following properties follow directly from (1.5).

PROPOSITION 2.3. *Let $\{x, y\} \subset \mathcal{X}$ and $\{u, v\} \subset \operatorname{int}\operatorname{dom} f$. Then*

(i) $D(u,v) + D(v,u) = \langle u - v, \nabla f(u) - \nabla f(v)\rangle$;
(ii) $D(x,u) = D(x,v) + D(v,u) + \langle x - v, \nabla f(v) - \nabla f(u)\rangle$;
(iii) $D(x,v) + D(y,u) = D(x,u) + D(y,v) + \langle x - y, \nabla f(u) - \nabla f(v)\rangle$.

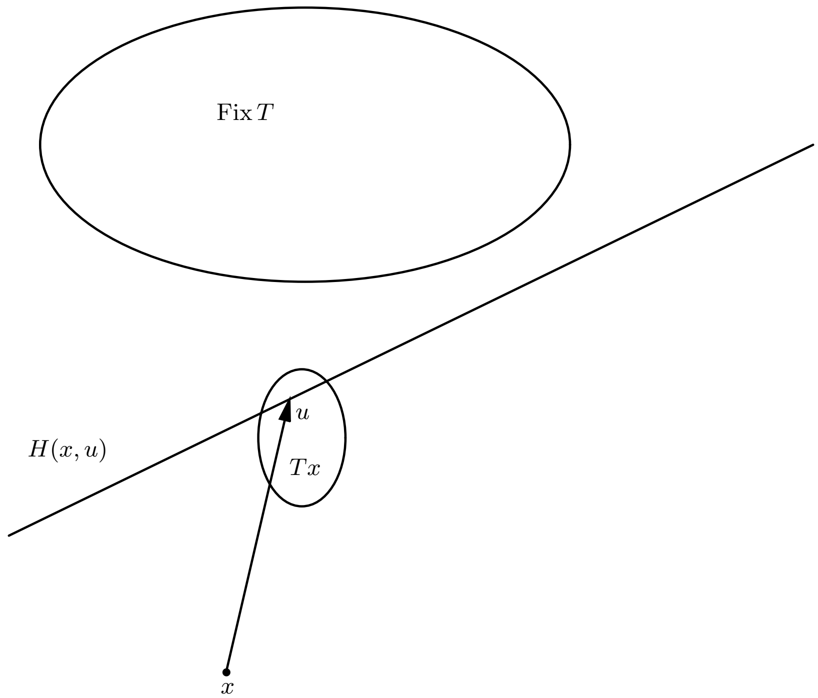

FIG. 1. *If $T \in \mathfrak{B}$, $x \in \operatorname{int}\operatorname{dom} f$, and $u \in Tx$, the half-space $H(x,u)$ contains* $\operatorname{Fix} T$.

3. Operators associated with Bregman distances. In Hilbert spaces, various nonlinear operators are involved in the design of algorithms, including projection operators, proximal operators, resolvents, subgradient projection operators, firmly nonexpansive operators, and combinations of these. Such operators arise in convex feasibility problems, in equilibrium theory, in systems of convex inequalities, in variational inequalities, as well as in numerous fixed point problems [5, 6, 9, 17, 35, 40, 41, 60, 72, 78]. Intrinsically tied to the very definition of these operators is the use of the standard notion of metric distance to measure the proximity between two points. In the context of Bregman distances, it is therefore natural to attempt to define variants of these operators. This effort has been undertaken by several authors at various levels of generality. In this section, we systematically study nonlinear operators associated with Bregman distances in order to bring together and extend a collection of results disseminated in the literature. Specifically, we investigate when D-firm operators, D-resolvents, D-prox operators, D-projectors, and subgradient D-projectors belong to class $\mathfrak{B}$. (For relationships among these operators in the classical case, i.e., when $\mathcal{X}$ is a Hilbert space and $f = \|\cdot\|^2/2$, see [9, Prop. 2.3].) Moreover, the class $\mathfrak{B}$ is shown to be closed under a certain type of relaxed parallel combination. The discussion is not limited to convex problems as nonconvex extensions of standard algorithms have been found to be quite useful in a number of applications; see [12, 28, 43, 52, 62].

3.1. The class $\mathfrak{B}$. Ultimately, our goal is to define a class of operators for which (1.7) systematically generates D-monotone sequences. In this perspective, the operators employed in (1.7) must be D-viable (see Definition 2.1) and induce a certain monotonicity property (see Definition 1.2). These requirements lead to the following class of operators (see Figure 1).

DEFINITION 3.1. *For every x and u in* $\operatorname{int}\operatorname{dom} f$*, set*

$$H(x,u) = \big\{y \in \mathcal{X} \mid \langle y-u, \nabla f(x) - \nabla f(u)\rangle \leq 0\big\}. \tag{3.1}$$

Then

$$\mathfrak{B} = \big\{T\colon \mathcal{X} \to 2^{\mathcal{X}} \mid \operatorname{ran} T \subset \operatorname{dom} T = \operatorname{int}\operatorname{dom} f,\ (\forall (x,u) \in \operatorname{gr} T)\ \operatorname{Fix} T \subset H(x,u)\big\}.$$

If $\mathcal{X}$ is Hilbertian, $f = \|\cdot\|^2/2$, and only single-valued operators are considered, then $\mathfrak{B}$ reverts to the class $\mathfrak{T}$ of operators introduced in [9] and further investigated in this context in [41, 42]. In these studies, $\mathfrak{T}$ was shown to play a central role in the analysis of Fejér monotone algorithms. Because of Proposition 3.3(i) below, there is some overlap between the "paracontractions" introduced in [31, 75] (see also [24, 26]) and operators in $\mathfrak{B}$. Furthermore, if f satisfies certain conditions and $T \in \mathfrak{B}$ is single-valued with $\operatorname{Fix} T \neq \varnothing$, then T is "totally nonexpansive" in the sense of [24].

LEMMA 3.2. *Let C_1 and C_2 be two convex subsets of $\mathcal{X}$ such that C_1 is closed and $C_1 \cap \operatorname{int} C_2 \neq \varnothing$. Then $\overline{C_1 \cap \operatorname{int} C_2} = C_1 \cap \overline{C_2}$.*

Proof. Since C_2 is convex with nonempty interior, $\overline{C_1 \cap \operatorname{int} C_2} \subset \overline{C_1} \cap \overline{\operatorname{int} C_2} = C_1 \cap \overline{C_2}$. To show the reverse inclusion, fix $x_0 \in C_1 \cap \operatorname{int} C_2$ and $x_1 \in C_1 \cap \overline{C_2}$. By convexity, $[x_0, x_1] \subset C_1$ and $[x_0, x_1[\subset \operatorname{int} C_2$. Therefore, $(\forall \alpha \in [0,1[)\ x_\alpha = (1-\alpha)x_0 + \alpha x_1 \in C_1 \cap \operatorname{int} C_2$. Consequently $x_1 = \lim_{\alpha \uparrow 1^-} x_\alpha \in \overline{C_1 \cap \operatorname{int} C_2}$, and we conclude $C_1 \cap \overline{C_2} \subset \overline{C_1 \cap \operatorname{int} C_2}$. ☐

PROPOSITION 3.3. *Let T be an operator in $\mathfrak{B}$ and let $F = \bigcap_{(x,u) \in \operatorname{gr} T} H(x,u)$. Then*

(i) $(\forall (x,u) \in \operatorname{gr} T)(\forall y \in \operatorname{Fix} T)\ \ D(y,u) \leq D(y,x) - D(u,x)$;
(ii) $(\forall (x,u) \in \operatorname{gr} T)\ \ D(u,x) \leq D_{\operatorname{Fix} T}(x)$;
(iii) $(\forall (x,u) \in \operatorname{gr} T)(\forall y \in \overline{\operatorname{Fix}} T)\ \ D(x,u) + D(u,x) \leq \langle y - x, \nabla f(u) - \nabla f(x)\rangle$.

Now suppose that $f|_{\operatorname{int}\operatorname{dom} f}$ is strictly convex; then

(iv) $\operatorname{Fix} T = F \cap \operatorname{int}\operatorname{dom} f$;
(v) $\operatorname{Fix} T$ *is convex;*
(vi) T *is single-valued on* $\operatorname{Fix} T$.

If, in addition, $\operatorname{Fix} T \neq \varnothing$*, then*

(vii) $\overline{\operatorname{Fix}} T = F \cap \overline{\operatorname{dom}} f$;
(viii) $(\forall (x,u) \in \operatorname{gr} T)(\forall y \in \overline{\operatorname{Fix}} T)\ \ D(y,u) \leq D(y,x) - D(u,x)$.

Proof. (i) Take $(x,u) \in \operatorname{gr} T$ and $y \in \operatorname{Fix} T$. Then Proposition 2.3(ii) and the inclusion $y \in H(x,u)$ yield $D(y,u) = D(y,x) - D(u,x) + \langle y-u, \nabla f(x) - \nabla f(u)\rangle \leq D(y,x) - D(u,x)$. (ii) By (i), $(\forall (x,u) \in \operatorname{gr} T)(\forall y \in \operatorname{Fix} T)\ D(u,x) \leq D(y,x)$. (iii) Take $(x,u) \in \operatorname{gr} T$ and $y \in \overline{\operatorname{Fix}} T$, and suppose $y_n \to y$ for some sequence $(y_n)_{n \in \mathbb{N}}$ in $\operatorname{Fix} T$. Then it follows from Proposition 2.3(i) that

$$\begin{aligned}(\forall n \in \mathbb{N})\ \ D(x,u) + D(u,x) &= \langle x-u, \nabla f(x) - \nabla f(u)\rangle \\ &= \langle x-y_n, \nabla f(x) - \nabla f(u)\rangle + \langle y_n - u, \nabla f(x) - \nabla f(u)\rangle \\ &\leq \langle x - y_n, \nabla f(x) - \nabla f(u)\rangle. \end{aligned} \tag{3.2}$$

Since $\langle x-y_n, \nabla f(x) - \nabla f(u)\rangle \to \langle x-y, \nabla f(x) - \nabla f(u)\rangle$, the proof is complete. (iv) Take $y \in F \cap \operatorname{int}\operatorname{dom} f$. Then $y \in \bigcap_{u \in Ty} H(y,u)$ and, in turn,

$$(\forall u \in Ty)\ \ \langle y-u, \nabla f(y) - \nabla f(u)\rangle \leq 0. \tag{3.3}$$

However, $\{y\} \cup Ty \subset \operatorname{int}\operatorname{dom} f$ and, since $f|_{\operatorname{int}\operatorname{dom} f}$ is strictly convex, ∇f is strictly monotone on $\operatorname{int}\operatorname{dom} f$. Therefore $Ty = \{y\}$ and $y \in \operatorname{Fix} T$. Thus, $F \cap \operatorname{int}\operatorname{dom} f \subset$

$\operatorname{Fix} T$. Since $T \in \mathfrak{B}$, the reverse inclusion is clear. (iv) $\Rightarrow$ (v) Since the sets $\big(H(x,u)\big)_{(x,u)\in \operatorname{gr} T}$ and $\operatorname{int}\operatorname{dom} f$ are convex, so is their intersection $\operatorname{Fix} T$. (vi) was proved in the proof of (iv). (iv) $\Rightarrow$ (vii) Observe that F is closed and apply Lemma 3.2. (viii) Take $(x,u) \in \operatorname{gr} T$, $y_0 \in \operatorname{Fix} T$, and $y \in \overline{\operatorname{Fix}}\, T$. By (iv) and (vii), $\operatorname{Fix} T = F \cap \operatorname{int}\operatorname{dom} f$ and $\overline{\operatorname{Fix}}\, T = F \cap \overline{\operatorname{dom}}\, f$. Since F and $\operatorname{dom} f$ are convex, $[y_0, y] \subset F$ and $[y_0, y[\subset \operatorname{int}\operatorname{dom} f$. Therefore,

$$(\forall \alpha \in [0,1[)\ \ y_\alpha = (1-\alpha)y_0 + \alpha y \in \operatorname{Fix} T. \tag{3.4}$$

Invoking the lower semicontinuity and convexity of f, we get

$$f(y) \le \underline{\lim}_{\alpha\uparrow 1^-} f(y_\alpha) \le \overline{\lim}_{\alpha\uparrow 1^-} f(y_\alpha) \le \overline{\lim}_{\alpha\uparrow 1^-} (1-\alpha)f(y_0) + \alpha f(y) = f(y). \tag{3.5}$$

Hence $\lim_{\alpha\uparrow 1^-} f(y_\alpha) = f(y)$ and, in turn,

$$(\forall z \in \operatorname{int}\operatorname{dom} f)\ \ \lim_{\alpha\uparrow 1^-} D(y_\alpha, z) = D(y,z). \tag{3.6}$$

On the other hand, since $u \in Tx$ and $T \in \mathfrak{B}$, (3.4) and (i) yield

$$(\forall \alpha \in [0,1[)\ \ D(y_\alpha, u) \le D(y_\alpha, x) - D(u,x). \tag{3.7}$$

Consequently, $D(y,u) \le D(y,x) - D(u,x)$. ☐

3.2. *D*-firm operators. An operator $T\colon \mathcal{X} \to \mathcal{X}$ is said to be firmly nonexpansive if for all x and y in $\operatorname{dom} T$ one has [51]

$$(\forall \alpha \in \,]0,+\infty[)\ \ \|Tx - Ty\| \le \|\alpha(x-y) + (1-\alpha)(Tx-Ty)\|. \tag{3.8}$$

For the sake of notational simplicity, let us now suppose that $\mathcal{X}$ is smooth. Then its normalized duality map J is single-valued and, upon invoking the equivalence $(\forall \alpha \in \,]0,+\infty[)\ \|u\| \le \|u + \alpha v\| \Leftrightarrow 0 \le \langle v, Ju\rangle$ [51], we observe that (3.8) is equivalent to

$$\langle Tx - Ty, J(Tx-Ty)\rangle \le \langle x - y, J(Tx-Ty)\rangle\,. \tag{3.9}$$

If $\mathcal{X}$ is not a Hilbert space, then J is not linear and this type of inequality may be difficult to manipulate. In Hilbert spaces, $J = \operatorname{Id} = \nabla f$ for $f = \|\cdot\|^2/2$, and (3.9) can therefore be written

$$\langle Tx - Ty, \nabla f(Tx) - \nabla f(Ty)\rangle \le \langle Tx - Ty, \nabla f(x) - \nabla f(y)\rangle\,. \tag{3.10}$$

In the framework of Bregman distances, this inequality suggests the following definition.

DEFINITION 3.4. *An operator* $T\colon \mathcal{X} \to 2^{\mathcal{X}}$ *with* $\operatorname{dom} T \cup \operatorname{ran} T \subset \operatorname{int}\operatorname{dom} f$ *is* D-firm *if*

$$(\forall (x,u) \in \operatorname{gr} T)(\forall (y,v) \in \operatorname{gr} T)\ \ \langle u - v, \nabla f(u) - \nabla f(v)\rangle \le \langle u - v, \nabla f(x) - \nabla f(y)\rangle\,. \tag{3.11}$$

PROPOSITION 3.5. *Let* $T\colon \mathcal{X} \to 2^{\mathcal{X}}$ *be a D-firm operator. Then*

(i) $(\forall (x,u) \in \operatorname{gr} T)\ \ \operatorname{Fix} T \subset H(x,u)$;

(ii) $T \in \mathfrak{B}$ *if* $\operatorname{int}\operatorname{dom} f = \operatorname{dom} T$;

(iii) T *is single-valued on its domain if* $f|_{\operatorname{int}\operatorname{dom} f}$ *is strictly convex;*

(iv) $(\forall (x,u) \in \operatorname{gr} T)(\forall (y,v) \in \operatorname{gr} T) \quad D(u,v) + D(v,u) \leq D(u,y) + D(v,x) - D(u,x) - D(v,y)$.

Proof. (i) Suppose $y \in Ty$. Then (3.11) implies that

$$(\forall (x,u) \in \operatorname{gr} T) \quad \langle y-u, \nabla f(x) - \nabla f(u) \rangle \leq 0. \tag{3.12}$$

(i) ⇒ (ii) is clear. (iii) Fix $x \in \operatorname{dom} T$ and $\{u,v\} \subset Tx$. Then (3.11) implies that

$$\langle u-v, \nabla f(u) - \nabla f(v) \rangle \leq 0. \tag{3.13}$$

Since ∇f is strictly monotone on $\operatorname{int} \operatorname{dom} f \supset \{u,v\}$, we obtain $u = v$. (iv) follows from Proposition 2.3(i), (3.11), and Proposition 2.3(iii). □

Remark 3.6. For single-valued operators in Hilbert spaces and f strongly convex (i.e., $f - \beta\|\cdot\|^2/2$ is convex for some $\beta \in \,]0,+\infty[$), item (iv) above was used to define D-firmness in [18].

3.3. *D*-resolvents. The resolvent of an operator $A\colon \mathcal{X} \to 2^{\mathcal{X}}$ is $(\operatorname{Id} + A)^{-1}$. It is known that an operator $T\colon \mathcal{X} \to \mathcal{X}$ is firmly nonexpansive if and only if it is the resolvent of an accretive operator $A\colon \mathcal{X} \to 2^{\mathcal{X}}$ [19].

Now let $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be a nontrivial operator, i.e., $\operatorname{gr} A \neq \varnothing$. Then, in the context of Bregman distances, it is reasonable to introduce the following variant of the notion of a resolvent to obtain an operator from $\mathcal{X}$ to $\mathcal{X}$ (this definition appears to have first been proposed in $\mathbb{R}^N$ in [46]).

DEFINITION 3.7. *The* D-resolvent *associated with* $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ *is the operator*

$$R_A = (\nabla f + A)^{-1} \circ \nabla f\colon \mathcal{X} \to 2^{\mathcal{X}}. \tag{3.14}$$

An a posteriori motivation for (3.14) is that it preserves the usual fixed point characterization of the zeros of A, namely,

$$(\forall x \in \mathcal{X})(\forall \gamma \in \,]0,+\infty[) \quad 0 \in Ax \quad \Leftrightarrow \quad x \in \operatorname{Fix} R_{\gamma A}, \tag{3.15}$$

as $0 \in Ax \Leftrightarrow \nabla f(x) \in \nabla f(x) + \gamma A(x) = (\nabla f + \gamma A)(x) \Leftrightarrow x \in (\nabla f + \gamma A)^{-1}\big(\nabla f(x)\big)$. It is also consistent with previous attempts to define resolvents for monotone operators:

- Let $\mathcal{X}$ be smooth and set $f = \|\cdot\|^2/2$. Then $\nabla f = J$ and $R_A = (J + A)^{-1} \circ J$. This type of resolvent was used in [57].
- If $\mathcal{X}$ is Hilbertian and $f\colon x \mapsto \|\Pi x\|^2/2$, where Π is the metric projector onto a closed vector subspace of $\mathcal{X}$, then $\nabla f = \Pi$ and $R_A = (\Pi + A)^{-1} \circ \Pi$. This generalized resolvent was used in [54].

PROPOSITION 3.8. *R_A satisfies the following properties:*

(i) $\operatorname{dom} R_A \subset \operatorname{int} \operatorname{dom} f$.

(ii) $\operatorname{ran} R_A \subset \operatorname{int} \operatorname{dom} f$.

(iii) $\operatorname{Fix} R_A = (\operatorname{int} \operatorname{dom} f) \cap A^{-1}0$.

(iv) *Suppose A is monotone. Then the following conditions hold:*

  (a) *R_A is D-firm.*

  (b) *R_A is single-valued on its domain if $f|_{\operatorname{int} \operatorname{dom} f}$ is strictly convex.*

  (c) *Suppose* $\operatorname{ran} \nabla f \subset \operatorname{ran}(\nabla f + A)$. *Then* $R_A \in \mathfrak{B}$. *If, in addition, $f|_{\operatorname{int} \operatorname{dom} f}$ is strictly convex, then* $\operatorname{Fix} R_A$ *is convex.*

Proof. (i) is clear. (ii) We have

$$\begin{aligned} \operatorname{ran} R_A \subset \operatorname{ran}(\nabla f + A)^{-1} &= \operatorname{dom}(\nabla f + A) = \operatorname{dom} \nabla f \cap \operatorname{dom} A \subset \operatorname{dom} \nabla f \\ &= \operatorname{int} \operatorname{dom} f. \end{aligned} \tag{3.16}$$

(iii) $\operatorname{Fix} R_A \subset \operatorname{int}\operatorname{dom} f$ by (i) and $(\forall x \in \operatorname{int}\operatorname{dom} f)\ 0 \in Ax \Leftrightarrow x \in R_A x$ by (3.15). Hence, $A^{-1}0 \cap \operatorname{int}\operatorname{dom} f = \operatorname{Fix} R_A \cap \operatorname{int}\operatorname{dom} f = \operatorname{Fix} R_A$. (iv) Suppose that A is monotone. (a) In view of (i) and (ii), let us show that (3.11) is satisfied. Fix (x, u) and (y, v) in $\operatorname{gr} R_A$. Then $\nabla f(x) - \nabla f(u) \in Au$ and $\nabla f(y) - \nabla f(v) \in Av$. Consequently, since A is monotone, we get $\langle u-v, \nabla f(x)-\nabla f(u)-(\nabla f(y)-\nabla f(v))\rangle \geq 0$. (b) follows from (a) and Proposition 3.5(iii). (c) $\operatorname{ran}\nabla f \subset \operatorname{ran}(\nabla f + A) \Leftrightarrow \operatorname{ran}\nabla f \subset \operatorname{dom}(\nabla f + A)^{-1} \Leftrightarrow \operatorname{dom} R_A = \operatorname{dom}\nabla f = \operatorname{int}\operatorname{dom} f$. In view of (a) and Proposition 3.5(ii), $R_A \in \mathfrak{B}$. Proposition 3.3(v) implies the convexity of $\operatorname{Fix} R_A$. ☐

Definition 3.9 (see [86, sections 32.14 and 32.21]). *A is*

(i) weakly coercive *if* $\lim_{\|x\|\to+\infty} \inf \|Ax\| = +\infty$;

(ii) strongly coercive *if*

$$(\forall x \in \operatorname{dom} A) \qquad \lim_{\|y\|\to+\infty} \inf \frac{\langle y - x, Ax\rangle}{\|y\|} = +\infty;$$

(iii) 3-monotone *if*

$$\big(\forall\big((x, x^*), (y, y^*), (z, z^*)\big) \in (\operatorname{gr} A)^3\big) \ \ \langle x - y, x^*\rangle + \langle y - z, y^*\rangle + \langle z - x, z^*\rangle \geq 0;$$

(iv) 3^*-monotone *if it is monotone and*

$$(\forall (x, x^*) \in \operatorname{dom} A \times \operatorname{ran} A) \ \ \sup\big\{\langle x - y, y^* - x^*\rangle \mid (y, y^*) \in \operatorname{gr} A\big\} < +\infty.$$

Lemma 3.10 (see [86, section 32.21], [16]). *Suppose that $\mathcal{X}$ is reflexive and that A is monotone and satisfies one of the following properties:*

(i) *A is 3-monotone.*

(ii) *A is strongly coercive.*

(iii) *$\operatorname{ran} A$ is bounded.*

(iv) *$A = \partial\varphi$, where $\varphi\colon \mathcal{X} \to \left]-\infty, +\infty\right]$ is a proper function.*

Then A is 3^*-monotone.

The following lemma is Reich's extension to a reflexive Banach space setting of the Brézis–Haraux theorem [16] on the range of the sum of two monotone operators.

Lemma 3.11 (see [74, Thm. 2.2]). *Suppose that $\mathcal{X}$ is reflexive and let $A_1, A_2\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be two monotone operators such that $A_1 + A_2$ is maximal monotone and A_1 is 3^*-monotone. In addition, suppose that* $\operatorname{dom} A_2 \subset \operatorname{dom} A_1$ *or A_2 is 3^*-monotone. Then* $\operatorname{int}\operatorname{ran}(A_1 + A_2) = \operatorname{int}(\operatorname{ran} A_1 + \operatorname{ran} A_2)$ *and* $\overline{\operatorname{ran}}\,(A_1 + A_2) = \overline{\operatorname{ran} A_1 + \operatorname{ran} A_2}$.

Proposition 3.12. *Let $\gamma \in \left]0, +\infty\right[$. Suppose that $\mathcal{X}$ is reflexive and that A is maximal monotone with* $(\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom} A = \operatorname{dom}\partial f \cap \operatorname{dom} A \neq \varnothing$. *Then $\nabla f + \gamma A$ is maximal monotone. Moreover, the inclusions*

$$\begin{cases} \operatorname{int}(\operatorname{ran}\nabla f + \gamma \operatorname{ran} A) \subset \operatorname{ran}(\nabla f + \gamma A) \\ \operatorname{ran}\nabla f + \gamma \operatorname{ran} A \subset \overline{\operatorname{ran}}\,(\nabla f + \gamma A) \end{cases} \tag{3.17}$$

are satisfied if one of the following conditions holds:

(i) $\operatorname{dom} A \subset \operatorname{int}\operatorname{dom} f$.

(ii) *A is 3^*-monotone.*

Proof. Since f is proper, lower semicontinuous, and convex, ∂f is maximal monotone [79, Thm. 30.3] and $\operatorname{int}\operatorname{dom} f = \operatorname{cont} f \subset \operatorname{dom}\partial f \subset \operatorname{dom} f$ [48, Chap. I]. Since $(\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom} A = \operatorname{dom}\partial f \cap \operatorname{dom} A \neq \varnothing$, we have $(\operatorname{int}\operatorname{dom}\partial f) \cap \operatorname{dom}\gamma A = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom} A \neq \varnothing$, and it follows from Rockafellar's sum theorem [79, section 23] that $\partial f + \gamma A$ is maximal monotone. However, the above assumption implies that $\operatorname{dom}(\nabla f + \gamma A) = \operatorname{dom}(\partial f + \gamma A)$ and, in turn, that $\nabla f + \gamma A = \partial f + \gamma A$ since $\{\nabla f\} = \partial f|_{\operatorname{int}\operatorname{dom} f}$. Thus, $\nabla f + \gamma A$ is maximal monotone. The second assertion is an application of Lemma 3.11 with $A_1 = \nabla f$ and $A_2 = \gamma A$. Indeed, $\operatorname{dom}\nabla f = \operatorname{int}\operatorname{dom} f$ and, by Lemma 3.10(iv), ∂f is 3^*-monotone and so is, therefore, ∇f since $\operatorname{gr}\nabla f \subset \operatorname{gr}\partial f$. ☐

THEOREM 3.13. *Let $\gamma \in]0, +\infty[$. Suppose that $\mathcal{X}$ is reflexive, that A is maximal monotone with* $(\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom} A = \operatorname{dom}\partial f \cap \operatorname{dom} A \neq \varnothing$*, and that one of the following conditions holds:*

(i) *$\mathcal{X}$ is smooth and $f = \|\cdot\|^2/2$.*
(ii) *$(\nabla f + \gamma A)^{-1}$ is locally bounded at every point in $\mathcal{X}^*$.*
(iii) *$\nabla f + \gamma A$ is weakly coercive.*
(iv) $\operatorname{dom} A \subset \operatorname{int}\operatorname{dom} f$ *or A is 3^*-monotone, and one of the following conditions holds:*
  (a) $\operatorname{ran}\nabla f + \gamma \operatorname{ran} A = \mathcal{X}^*$.
  (b) *f is Legendre and cofinite.*
  (c) $\operatorname{ran}(\nabla f + \gamma A)$ *is closed and* $0 \in \operatorname{ran} A$.
  (d) $\operatorname{ran}\nabla f$ *is open and* $0 \in \operatorname{ran} A$.

Then $R_{\gamma A} \in \mathfrak{B}$.

Proof. In view of Proposition 3.8(iv)(c), it suffices to show that $\operatorname{ran}\nabla f \subset \operatorname{ran}(\nabla f + \gamma A)$. (i) Since $\mathcal{X}$ is smooth, $\nabla f = J$ [34, Corollary I.4.5] and Rockafellar's surjectivity theorem [79, Thm. 10.7] yields $\operatorname{ran}(\nabla f + \gamma A) = \mathcal{X}^*$. (ii) Proposition 3.12 asserts that $\nabla f + \gamma A$ is maximal monotone. It therefore follows from the Brézis–Browder surjectivity theorem (see [34, Thm. V.3.8] or [86, Thm. 32.G]) that $\operatorname{ran}(\nabla f + \gamma A) = \mathcal{X}^*$. (iii) $\Rightarrow$ (ii) follows from [86, Cor. 32.35] since $\nabla f + \gamma A$ is maximal monotone. (iv) By Proposition 3.12, (3.17) holds. (a) By (3.17), $\mathcal{X}^* = \operatorname{int}(\operatorname{ran}\nabla f + \gamma \operatorname{ran} A) \subset \operatorname{ran}(\nabla f + \gamma A)$. (b) $\Rightarrow$ (a) By [8, Thm. 5.10], Legendreness guarantees $\operatorname{ran}\nabla f = \operatorname{int}\operatorname{dom} f^*$ while cofiniteness gives $\operatorname{int}\operatorname{dom} f^* = \mathcal{X}^*$. Consequently, $\operatorname{ran}\nabla f + \gamma \operatorname{ran} A = \mathcal{X}^*$. (c) By (3.17), $\operatorname{ran}\nabla f = \operatorname{ran}\nabla f + \{0\} \subset \operatorname{ran}\nabla f + \gamma \operatorname{ran} A \subset \overline{\operatorname{ran}}\,(\nabla f + \gamma A) = \operatorname{ran}(\nabla f + \gamma A)$. (d) By (3.17), $\operatorname{ran}\nabla f = \operatorname{int}(\operatorname{ran}\nabla f + \{0\}) \subset \operatorname{int}(\operatorname{ran}\nabla f + \gamma \operatorname{ran} A) \subset \operatorname{ran}(\nabla f + \gamma A)$. ☐

In connection with the problem of finding zeros of maximal monotone operators, the following corollary is particularly useful.

COROLLARY 3.14. *Let $\gamma \in]0, +\infty[$. Suppose that $\mathcal{X}$ is reflexive, that A is maximal monotone with $0 \in \operatorname{ran} A$, and that one of the following conditions holds:*

(i) $\operatorname{ran}\nabla f$ *is open and* $\operatorname{dom} A \subset \operatorname{int}\operatorname{dom} f$.
(ii) *f is Legendre and* $\operatorname{dom} A \subset \operatorname{int}\operatorname{dom} f$.
(iii) *f is Legendre, A is 3^*-monotone, and* $\operatorname{dom} A \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$.

Then $R_{\gamma A} \in \mathfrak{B}$.

Proof. The assertions follow from Theorem 3.13(iv)(d). Indeed, in (i), $\operatorname{dom} A \subset \operatorname{int}\operatorname{dom} f = \operatorname{cont} f \subset \operatorname{dom}\partial f \Rightarrow (\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom} A = \operatorname{dom}\partial f \cap \operatorname{dom} A = \operatorname{dom} A \neq \varnothing$. On the other hand, in (ii) and (iii), $\operatorname{ran}\nabla f$ is open since Legendreness yields $\operatorname{ran}\nabla f = \operatorname{int}\operatorname{dom} f^*$ [8, Thm. 5.10]. Consequently, if $\operatorname{dom} A \subset \operatorname{int}\operatorname{dom} f$, then (ii) is a consequence of (i). Otherwise, if A is 3^*-monotone and $(\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom} A \neq \varnothing$, then it suffices to note that essential smoothness yields $\operatorname{dom}\partial f = \operatorname{int}\operatorname{dom} f$ [8, Thm. 5.6], whence $(\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom} A = \operatorname{dom}\partial f \cap \operatorname{dom} A \neq \varnothing$. ☐

Remark 3.15. In $\mathbb{R}^N$, Corollary 3.14(i) corresponds to [46, Thm. 4].

3.4. *D*-prox operators. The classical notion of a proximal operator was introduced by Moreau [64, 65, 67] in Hilbert spaces. The proximal operator associated with a function $\varphi\colon \mathcal{X} \to]-\infty, +\infty]$ is $\operatorname{prox}^{\varphi}\colon y \mapsto \operatorname{argmin}\varphi + \|\cdot - y\|^2/2$. Outside of

Hilbert spaces, this notion is of less interest since Fermat's rule for the minimization of $\varphi + \|\cdot - y\|^2/2$ becomes a nonseparable inclusion, namely, $0 \in \partial\varphi(x) + J(x-y)$.

In $\mathbb{R}^N$, the idea of defining proximal operators based on D-distance—rather than quadratic—penalizations was introduced in [32]. In our setting, they will be defined as follows.

DEFINITION 3.16. *Let* $\varphi\colon \mathcal{X} \to \,]-\infty,+\infty]$. *The* D-prox operator *of index* $\gamma \in \,]0,+\infty[$ *associated with* φ *is the operator*

$$\operatorname{prox}_\gamma^\varphi\colon \mathcal{X} \to 2^{\mathcal{X}},$$
$$y \mapsto \left\{ x \in \operatorname{dom} f \cap \operatorname{dom}\varphi \mid \varphi(x) + \frac{1}{\gamma}D(x,y) = \min\left(\varphi + \frac{1}{\gamma}D(\cdot,y)\right)(\mathcal{X}) < +\infty \right\}.$$

It follows from this definition that

$$\operatorname{dom}\operatorname{prox}_\gamma^\varphi \subset \operatorname{int}\operatorname{dom} f \quad \text{and} \quad \operatorname{ran}\operatorname{prox}_\gamma^\varphi \subset \operatorname{dom} f \cap \operatorname{dom}\varphi. \tag{3.18}$$

Recall (see section 2.5) that a function is weak inf-compact if all its lower level sets are weakly compact.

LEMMA 3.17. *Suppose that* $g_1\colon \mathcal{X} \to \,]-\infty,+\infty]$ *is weak lower semicontinuous and bounded from below and that* $g_2\colon \mathcal{X} \to \,]-\infty,+\infty]$ *is weak inf-compact. Then* $g_1 + g_2$ *is weak inf-compact.*

Proof. Set $\beta = \inf g_1(\mathcal{X})$ and let $\eta \in \mathbb{R}$. Since g_1 and g_2 are weak lower semicontinuous, so is their sum, and therefore $\operatorname{lev}_{\leq\eta}(g_1+g_2)$ is weakly closed. On the other hand, $\operatorname{lev}_{\leq\eta}(g_1+g_2)$ is contained in the weakly compact set $\operatorname{lev}_{\leq\eta-\beta} g_2$. We conclude that $\operatorname{lev}_{\leq\eta}(g_1+g_2)$ is weakly compact. ◻

The following result concerns the domain requirement for the D-viability of D-prox operators. Recall (see sections 2.5 and 2.2) that $\mathcal{M}$ denotes the set of minimizing sequences of a function and that $\mathfrak{W}$ is the set of weak cluster points of a sequence.

THEOREM 3.18. *Let* $\gamma \in \,]0,+\infty[$*, let* $\varphi\colon \mathcal{X} \to \,]-\infty,+\infty]$ *be such that* $\operatorname{dom} f \cap \operatorname{dom}\varphi \neq \varnothing$*, and assume that one of the following conditions holds:*

(i) $(\forall y \in \operatorname{int}\operatorname{dom} f)(\exists\,(x_n)_{n\in\mathbb{N}} \in \mathcal{M}(f_y + \gamma\varphi))(\exists\, x \in \mathfrak{W}(x_n)_{n\in\mathbb{N}})$ $f + \gamma\varphi$ *is weak lower semicontinuous at* x.

(ii) $(\forall y \in \operatorname{int}\operatorname{dom} f)$ $f_y + \gamma\varphi$ *is weak inf-compact.*

(iii) φ *is weak lower semicontinuous and bounded from below, and, for every* $y \in \operatorname{int}\operatorname{dom} f$, f_y *is weak inf-compact.*

(iv) φ *is weak inf-compact.*

Then $\operatorname{dom}\operatorname{prox}_\gamma^\varphi = \operatorname{int}\operatorname{dom} f$.

Proof. Fix $y \in \operatorname{int}\operatorname{dom} f$ and set $g = f_y + \gamma\varphi$. (i) Pick $(x_n)_{n\in\mathbb{N}} \in \mathcal{M}(g)$ such that $x_{k_n} \rightharpoonup x$ and g is weak lower semicontinuous at x. It follows that $g(x) \leq \underline{\lim}\, g(x_{k_n}) = \inf g(\mathcal{X})$ and hence $g(x) = \inf g(\mathcal{X})$. Therefore, g achieves its infimum and the result holds since $\operatorname{prox}_\gamma^\varphi y = \operatorname{Argmin}(f_y + \gamma\varphi) = \operatorname{Argmin}(g)$. (ii) $\Rightarrow$ (i) Take $(x_n)_{n\in\mathbb{N}} \in \mathcal{M}(g)$. Then it follows from weak inf-compactness of g that $(x_n)_{n\in\mathbb{N}}$ lies in a weakly compact set and therefore that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \neq \varnothing$. On the other hand, as g is weak inf-compact, it is weak lower semicontinuous and so is $f + \gamma\varphi = f_y + \gamma\varphi + \nabla f(y) = g + \nabla f(y)$. (iii) $\Rightarrow$ (ii) follows from Lemma 3.17. (iv) $\Rightarrow$ (ii) It is clear that f_y is weak lower semicontinuous. On the other hand, it follows from the convexity of f that, for every $x \in \mathcal{X}$, $\langle x - y, \nabla f(y)\rangle + f(y) \leq f(x)$ and, therefore, $f_y(x) \geq f_y(y)$. Hence $\inf f_y(\mathcal{X}) \geq f_y(y) > -\infty$ and, by Lemma 3.17, g is weak inf-compact. ◻

The following fundamental result is due to Moreau [66] and Rockafellar [76].

LEMMA 3.19. *Let* $y^* \in \mathcal{X}^*$. *Then* $f - y^*$ *is coercive if and only if* $y^* \in \operatorname{int}\operatorname{dom} f^*$.

LEMMA 3.20. *Let* $g_1, g_2\colon \mathcal{X} \to \,]-\infty,+\infty]$ *be two convex functions. Then*

(i) *if* g_1 *and* g_2 *are lower semicontinuous and* $0 \in \operatorname{int}\big(\operatorname{dom} g_1 - \operatorname{dom} g_2\big)$, *then* $(g_1+g_2)^* = g_1^* \,\square\, g_2^*$ [2];

(ii) *if* $\operatorname{cont} g_1 \cap \operatorname{dom} g_2 \neq \varnothing$, *then* $\partial(g_1+g_2) = \partial g_1 + \partial g_2$ [79, Thm. 28.2].

PROPOSITION 3.21. *Let* $\varphi\colon \mathcal{X} \to \,]-\infty,+\infty]$ *be a lower semicontinuous convex function such that* $\operatorname{dom} f \cap \operatorname{dom}\varphi \neq \varnothing$ *and let* $\gamma \in \,]0,+\infty[$. *Suppose that* $\mathcal{X}$ *is reflexive and that one of the following conditions holds:*

(i) $(\forall y \in \operatorname{int}\operatorname{dom} f)(\exists\,(x_n)_{n\in\mathbb{N}} \in \mathcal{M}(f_y+\gamma\varphi))\quad \sup_{n\in\mathbb{N}}\|x_n\| < +\infty$.

(ii) $(\forall y \in \operatorname{int}\operatorname{dom} f)\;\; f_y + \gamma\varphi$ *is coercive.*

(iii) $\operatorname{ran}\nabla f \subset \operatorname{int}\operatorname{dom}(f+\gamma\varphi)^*$.

(iv) $f+\gamma\varphi$ *is cofinite.*

(v) $0 \in \operatorname{int}(\operatorname{dom} f - \operatorname{dom}\varphi)$ *and* $\operatorname{dom} f^* + \gamma\operatorname{dom}\varphi^* = \mathcal{X}^*$.

(vi) φ *is bounded from below and* f *is essentially strictly convex.*

(vii) $f+\gamma\varphi$ *is supercoercive.*

(viii) φ *is bounded from below and* f *is supercoercive.*

(ix) φ *is coercive.*

Then $\operatorname{dom}\operatorname{prox}_\gamma^\varphi = \operatorname{int}\operatorname{dom} f$.

Proof. Let y be an arbitrary point in $\operatorname{int}\operatorname{dom} f$. Note that, since φ is weak lower semicontinuous, so are $f+\gamma\varphi$ and $f_y+\gamma\varphi$ and that, since $\mathcal{X}$ is reflexive, coercive weak lower semicontinuous functions are weak inf-compact. (i) is a consequence of Theorem 3.18(i). Indeed, take a bounded sequence $(x_n)_{n\in\mathbb{N}} \in \mathcal{M}(f_y+\gamma\varphi)$. Then it follows from the reflexivity of $\mathcal{X}$ that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \neq \varnothing$. (ii) follows at once from Theorem 3.18(ii). (iii) $\Leftrightarrow$ (ii) $\nabla f(y) \in \operatorname{int}\operatorname{dom}(f+\gamma\varphi)^* \Leftrightarrow f+\gamma\varphi-\nabla f(y)$ is coercive by Lemma 3.19. (iv) $\Rightarrow$ (iii) is clear. (v) $\Rightarrow$ (iv) Lemma 3.20(i) yields

$$\begin{aligned}\operatorname{dom} f^* + \gamma\operatorname{dom}\varphi^* &= \operatorname{dom} f^* + \operatorname{dom}\gamma\varphi^*(\cdot/\gamma) = \operatorname{dom} f^* + \operatorname{dom}(\gamma\varphi)^* \\ &= \operatorname{dom}\big(f^* \,\square\, (\gamma\varphi)^*\big)\end{aligned} \tag{3.19}$$

and

$$0 \in \operatorname{int}(\operatorname{dom} f - \operatorname{dom}\varphi) \quad\Rightarrow\quad f^* \,\square\, (\gamma\varphi)^* = (f+\gamma\varphi)^*. \tag{3.20}$$

Hence $\operatorname{dom} f^* + \gamma\operatorname{dom}\varphi^* = \mathcal{X}^* \Rightarrow \operatorname{dom}(f+\gamma\varphi)^* = \mathcal{X}^*$. (vi) is a consequence of Theorem 3.18(iii): indeed, by [8, Thm. 5.9(ii)], $\nabla f(y) \in \operatorname{int}\operatorname{dom} f^*$ and f_y is therefore coercive by Lemma 3.19. (vii) $\Rightarrow$ (iv) See [8, Thm. 3.4]. (viii) $\Rightarrow$ (vii) is clear. (ix) is a consequence of Theorem 3.18(iv). ∎

The next result gathers some facts concerning D-prox operators for convex functions.

PROPOSITION 3.22. *Let* $\varphi\colon \mathcal{X} \to \,]-\infty,+\infty]$ *be convex and let* $\gamma \in \,]0,+\infty[$. *Then the following hold:*

(i) $\operatorname{prox}_\gamma^\varphi = \big(\partial(f+\gamma\varphi)\big)^{-1} \circ \nabla f$.

(ii) *If, in addition,* $\operatorname{ran}\operatorname{prox}_\gamma^\varphi \subset \operatorname{int}\operatorname{dom} f$, *then*

  (a) $\operatorname{prox}_\gamma^\varphi = R_{\gamma\partial\varphi}$;

  (b) $\operatorname{Fix}\operatorname{prox}_\gamma^\varphi = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{Argmin}\varphi$;

  (c) $\operatorname{prox}_\gamma^\varphi$ *is D-firm;*

  (d) $\operatorname{prox}_\gamma^\varphi$ *is single-valued on its domain if* $f|_{\operatorname{int}\operatorname{dom} f}$ *is strictly convex.*

Proof. Fix $y \in \operatorname{int}\operatorname{dom} f$. (i) By (3.18), $\operatorname{ran}\operatorname{prox}_\gamma^\varphi \subset \operatorname{dom} f \cap \operatorname{dom}\varphi$. If $\operatorname{dom} f \cap \operatorname{dom}\varphi = \varnothing$, both sides of the desired identity reduce to the trivial operator $z \mapsto \varnothing$. If not, take $x \in \operatorname{dom} f \cap \operatorname{dom}\varphi$. Since $\operatorname{cont}\nabla f(y) = \mathcal{X}$, Lemma 3.20(ii) yields

$\partial(f_y+\gamma\varphi)(x)=\partial(f+\gamma\varphi)(x)-\nabla f(y)$. Consequently,

$$\begin{aligned} x\in \operatorname{prox}_\gamma^\varphi y &\Leftrightarrow 0\in\partial(f_y+\gamma\varphi)(x)\\ &\Leftrightarrow \nabla f(y)\in\partial(f+\gamma\varphi)(x)\\ &\Leftrightarrow x\in\big(\partial(f+\gamma\varphi)\big)^{-1}\big(\nabla f(y)\big). \end{aligned}\tag{3.21}$$

(ii) Suppose $\operatorname{ran}\operatorname{prox}_\gamma^\varphi\subset\operatorname{int}\operatorname{dom} f$. (a) On the one hand, it follows from (3.18) that $\operatorname{ran}\operatorname{prox}_\gamma^\varphi\subset(\operatorname{int}\operatorname{dom} f)\cap\operatorname{dom}\varphi$. On the other hand, $\operatorname{ran} R_{\gamma\partial\varphi}\subset\operatorname{dom}(\nabla f+\gamma\partial\varphi)\subset(\operatorname{int}\operatorname{dom} f)\cap\operatorname{dom}\varphi$. Therefore, if $(\operatorname{int}\operatorname{dom} f)\cap\operatorname{dom}\varphi=\varnothing$, both sides of the desired identity reduce to the trivial operator $z\mapsto\varnothing$. If not, take $x\in(\operatorname{int}\operatorname{dom} f)\cap\operatorname{dom}\varphi=\operatorname{cont} f\cap\operatorname{dom}\varphi$. Lemma 3.20(ii) now yields $\partial(f+\gamma\varphi)(x)=\nabla f(x)+\gamma\partial\varphi(x)$ and (3.21) becomes

$$x\in\operatorname{prox}_\gamma^\varphi y\Leftrightarrow\nabla f(y)\in\nabla f(x)+\gamma\partial\varphi(x)\Leftrightarrow x\in R_{\gamma\partial\varphi}y.\tag{3.22}$$

(a) $\Rightarrow$ (b) follows from Proposition 3.8(iii). (a) $\Rightarrow$ (c) Since $\partial\varphi$ is monotone, $R_{\gamma\partial\varphi}$ is D-firm by Proposition 3.8(iv)(a). (a) $\Rightarrow$ (d) follows from Proposition 3.8(iv)(b). ◻

We now turn our attention to the range requirement for the D-viability of D-prox operators.

PROPOSITION 3.23. *Let* $\varphi\colon\mathcal{X}\to\,]-\infty,+\infty]$ *be convex such that* $\operatorname{dom} f\cap\operatorname{dom}\varphi\neq\varnothing$*, and let* $\gamma\in\,]0,+\infty[$*. Assume that one of the following conditions holds:*
- (i) $\operatorname{dom}\partial(f+\gamma\varphi)\subset\operatorname{int}\operatorname{dom} f$.
- (ii) $\operatorname{dom} f\cap\operatorname{dom}\varphi\subset\operatorname{int}\operatorname{dom} f$.
- (iii) $\operatorname{dom} f$ *is open.*
- (iv) $\operatorname{dom}\varphi\subset\operatorname{int}\operatorname{dom} f$.
- (v) $(\operatorname{int}\operatorname{dom} f)\cap\operatorname{dom}\varphi\neq\varnothing$ *and one of the following conditions holds:*
  - (a) $\operatorname{dom}\partial f\cap\operatorname{dom}\partial\varphi\subset\operatorname{int}\operatorname{dom} f$.
  - (b) f *is essentially smooth.*
  - (c) $\operatorname{dom}\partial\varphi\subset\operatorname{int}\operatorname{dom} f$.

Then $\operatorname{ran}\operatorname{prox}_\gamma^\varphi\subset\operatorname{int}\operatorname{dom} f$.

Proof. (i) By Proposition 3.22(i),

$$\operatorname{ran}\operatorname{prox}_\gamma^\varphi\subset\operatorname{ran}\big(\partial(f+\gamma\varphi)\big)^{-1}=\operatorname{dom}\partial(f+\gamma\varphi)\subset\operatorname{int}\operatorname{dom} f.\tag{3.23}$$

(ii) $\Rightarrow$ (i) $\operatorname{dom}\partial(f+\gamma\varphi)\subset\operatorname{dom}(f+\gamma\varphi)=\operatorname{dom} f\cap\operatorname{dom}\varphi\subset\operatorname{int}\operatorname{dom} f$. (iii) $\Rightarrow$ (ii) and (iv) $\Rightarrow$ (ii) are clear. (v) $\Rightarrow$ (i) It results from Lemma 3.20(ii) that $\partial(f+\gamma\varphi)=\partial f+\gamma\partial\varphi$. Whence, (a) $\Rightarrow$ (i). (b) $\Rightarrow$ (a) Essential smoothness $\Rightarrow\operatorname{dom}\partial f=\operatorname{int}\operatorname{dom} f$ [8, Thm. 5.6(iii)]. (c) $\Rightarrow$ (a) is clear. ◻

Upon combining Propositions 3.23, 3.22(ii)(c), 3.21, and 3.5(ii), we obtain the following theorem.

THEOREM 3.24. *Let* $\varphi\colon\mathcal{X}\to\,]-\infty,+\infty]$ *be a lower semicontinuous convex function such that* $\operatorname{dom} f\cap\operatorname{dom}\varphi\neq\varnothing$*, and let* $\gamma\in\,]0,+\infty[$*. Suppose that* $\mathcal{X}$ *is reflexive and that one of conditions* (i)–(ix) *in Proposition* 3.21 *holds together with one of conditions* (i)–(v) *in Proposition* 3.23*. Then* $\operatorname{prox}_\gamma^\varphi\in\mathfrak{B}$.

The following special case underscores the importance of the notion of Legendreness.

COROLLARY 3.25. *Let* $\varphi\colon\mathcal{X}\to\,]-\infty,+\infty]$ *be a lower semicontinuous convex function such that* $(\operatorname{int}\operatorname{dom} f)\cap\operatorname{dom}\varphi\neq\varnothing$*, and let* $\gamma\in\,]0,+\infty[$*. Suppose that* $\mathcal{X}$ *is reflexive, that* f *is Legendre, and that* φ *is bounded below. Then*
- (i) $\operatorname{prox}_\gamma^\varphi$ *is single-valued on its domain and* $\operatorname{prox}_\gamma^\varphi\in\mathfrak{B}$;

(ii) *for every x and y in* $\operatorname{int}\operatorname{dom} f$,

$$x = \operatorname{prox}_\gamma^\varphi y \;\Leftrightarrow\; (\forall z \in \operatorname{dom}\varphi) \quad \langle z - x, \nabla f(y) - \nabla f(x)\rangle /\gamma + \varphi(x) \le \varphi(z).$$

Proof. (i) Combine Propositions 3.23(v)(b), 3.22(ii)(c) and (d), 3.21(vi), and 3.5(ii). (ii) By (3.22), $x = \operatorname{prox}_\gamma^\varphi y \Leftrightarrow \nabla f(y) - \nabla f(x) \in \gamma\partial\varphi(x)$. ☐

Remark 3.26. A special case of Theorem 3.18(iii) in $\mathbb{R}^N$ can be found in [32, Prop. 3.1]. In $\mathbb{R}^N$, assertions (iv) and (v)(b) of Proposition 3.23 appear in [58, Lemma 3.3]. In the case when $\mathcal{X}$ is Hilbertian and $f = \|\cdot\|^2/2$, the characterization supplied by Corollary 3.25(ii) is well known; see, e.g., [48, section II.2].

3.5. *D*-projections. The following concept goes back to Bregman's original paper [15].

Definition 3.27. *The* D-projector *onto a set $C \subset \mathcal{X}$ is the operator*

$$(3.24)\qquad \begin{aligned} P_C\colon \mathcal{X} &\to 2^{\mathcal{X}},\\ y &\mapsto \big\{x \in C \cap \operatorname{dom} f \mid D(x,y) = D_C(y) < +\infty\big\}. \end{aligned}$$

It is clear that, for any $\gamma \in \left]0,+\infty\right[$, $P_C = \operatorname{prox}_\gamma^{\iota_C}$. Hence, the results of section 3.4 will automatically yield results on D-projections when specialized to $\varphi = \iota_C$. Before we proceed in this direction, let us introduce a couple of definitions, which are natural adaptations of standard ones in metric approximation theory [81].

Definition 3.28. *A set $C \subset \mathcal{X}$ is* D-proximinal *if* $\operatorname{dom} P_C = \operatorname{int}\operatorname{dom} f$ *and* D-semi-Chebyshev *if P_C is single-valued on its domain. C is* D-Chebyshev *if it is D-proximinal and D-semi-Chebyshev.*

Definition 3.29. *A set $C \subset \mathcal{X}$ is* D-approximately weakly compact *if*

$$(\forall y \in \operatorname{int}\operatorname{dom} f)(\forall (x_n)_{n\in\mathbb{N}} \text{ in } C \cap \operatorname{dom} f) \quad D(x_n, y) \to D_C(y) \;\Rightarrow\; \mathfrak{W}(x_n)_{n\in\mathbb{N}} \cap C \ne \varnothing.$$

Theorem 3.30. *Let C be a subset of $\mathcal{X}$ such that $C \cap \operatorname{dom} f \ne \varnothing$ and assume that one of the following conditions holds:*

(i) *C is D-approximately weakly compact.*
(ii) $(\forall y \in \operatorname{int}\operatorname{dom} f)(\exists\, \eta \in \mathbb{R})$ *$C \cap \operatorname{lev}_{\le\eta} f_y$ is nonempty and weakly compact.*
(iii) *C is weakly closed and, for every $y \in \operatorname{int}\operatorname{dom} f$, f_y is weak inf-compact.*
(iv) *C is weakly compact.*

Then C is D-proximinal.

Proof. (i) Since f is weak lower semicontinuous, $f + \iota_C$ is weak lower semicontinuous at every point in C. Now fix $y \in \operatorname{int}\operatorname{dom} f$ and $(x_n)_{n\in\mathbb{N}} \in \mathcal{M}(f_y + \iota_C)$. Then $D(x_n, y) \to D_C(y)$ and Definition 3.29 yields $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \cap C \ne \varnothing$. Now take $x \in \mathfrak{W}(x_n)_{n\in\mathbb{N}} \cap C$. Since $f + \iota_C$ is weak lower semicontinuous at x, the claims follow from Theorem 3.18(i) with $\varphi = \iota_C$. (ii) Fix $y \in \operatorname{int}\operatorname{dom} f$. As minimizing $D(\cdot, y)$ over C is equivalent to minimizing the weak lower semicontinuous function f_y over the weakly compact set $C \cap \operatorname{lev}_{\le\eta} f_y$, the result follows. Assertions (iii) and (iv) follow, respectively, from assertions (iii) and (iv) in Theorem 3.18 with $\varphi = \iota_C$. ☐

Upon setting $\varphi = \iota_C$, Proposition 3.21 becomes the following.

Proposition 3.31. *Let C be a closed and convex subset of $\mathcal{X}$ such that $C \cap \operatorname{dom} f \ne \varnothing$. Suppose that $\mathcal{X}$ is reflexive and that one of the following conditions holds:*

(i) $(\forall y \in \operatorname{int}\operatorname{dom} f)(\forall (x_n)_{n\in\mathbb{N}} \in \mathcal{M}(f_y + \iota_C))$ $\sup_{n\in\mathbb{N}} \|x_n\| < +\infty$.
(ii) $(\forall y \in \operatorname{int}\operatorname{dom} f)$ *$f_y + \iota_C$ is coercive.*
(iii) $\operatorname{ran}\nabla f \subset \operatorname{int}\operatorname{dom}(f + \iota_C)^*$.

(iv) $f + \iota_C$ *is cofinite.*
(v) $0 \in \operatorname{int}(\operatorname{dom} f - C)$ *and* $\operatorname{dom} f^* + \operatorname{dom} \iota_C^* = \mathcal{X}^*$.
(vi) f *is essentially strictly convex.*
(vii) $f + \iota_C$ *is supercoercive.*
(viii) f *is supercoercive.*
(ix) C *is bounded.*

Then C *is* D*-proximinal.*

Likewise, Proposition 3.22 with $\varphi = \iota_C$ yields the following.

PROPOSITION 3.32. *Let* C *be a convex subset of* $\mathcal{X}$. *Then the following hold:*
(i) $P_C = \big(\partial(f + \iota_C)\big)^{-1} \circ \nabla f$.
(ii) *If, in addition,* $\operatorname{ran} P_C \subset \operatorname{int} \operatorname{dom} f$, *then*
  (a) $P_C = R_{N_C}$.
  (b) $\operatorname{Fix} P_C = C \cap \operatorname{int} \operatorname{dom} f$.
  (c) P_C *is* D*-firm.*
  (d) C *is* D*-semi-Chebyshev if* $f|_{\operatorname{int} \operatorname{dom} f}$ *is strictly convex.*

The D-viability requirements for the range of P_C are obtained by setting $\varphi = \iota_C$ in Proposition 3.23.

PROPOSITION 3.33. *Let* $C \subset \mathcal{X}$ *be convex such that* $C \cap \operatorname{dom} f \neq \varnothing$. *Assume that one of the following conditions holds:*
(i) $\operatorname{dom} \partial(f + \iota_C) \subset \operatorname{int} \operatorname{dom} f$.
(ii) $C \cap \operatorname{dom} f \subset \operatorname{int} \operatorname{dom} f$.
(iii) $\operatorname{dom} f$ *is open.*
(iv) $C \subset \operatorname{int} \operatorname{dom} f$.
(v) $C \cap \operatorname{int} \operatorname{dom} f \neq \varnothing$ *and one of the following conditions holds:*
  (a) $C \cap \operatorname{dom} \partial f \subset \operatorname{int} \operatorname{dom} f$;
  (b) f *is essentially smooth.*

Then $\operatorname{ran} P_C \subset \operatorname{int} \operatorname{dom} f$.

THEOREM 3.34. *Let* $C \subset \mathcal{X}$ *be a closed convex set such that* $C \cap \operatorname{dom} f \neq \varnothing$. *Suppose that* $\mathcal{X}$ *is reflexive and that one of conditions* (i)–(ix) *in Proposition* 3.31 *holds together with one of conditions* (i)–(v) *in Proposition* 3.33. *Then* $P_C \in \mathfrak{B}$.

Proof. Since Proposition 3.31 parallels Proposition 3.21 and Proposition 3.33 parallels Proposition 3.23, it suffices to set $\varphi = \iota_C$ in Theorem 3.24. ☐

We conclude this section with the following result.

COROLLARY 3.35. *Suppose that* $\mathcal{X}$ *is reflexive, that* f *is Legendre, and that* C *is a closed convex subset of* $\mathcal{X}$ *such that* $C \cap \operatorname{int} \operatorname{dom} f \neq \varnothing$. *Then*
(i) C *is* D*-Chebyshev and* $P_C \in \mathfrak{B}$;
(ii) *for every* x *and* y *in* $\operatorname{int} \operatorname{dom} f$,

$$x = P_C y \quad \Leftrightarrow \quad \begin{cases} x \in C, \\ C \subset H(y, x). \end{cases} \tag{3.25}$$

Proof. Take $\varphi = \iota_C$ in Corollary 3.25. ☐

Remark 3.36. Proposition 3.31(vii)–(ix) can be found in [1, Prop. 2.1]. Corollary 3.35(i) covers [8, Cor. 7.9] (see also [7, section 3] in the special case of Euclidean spaces), which was obtained via different arguments. If $\mathcal{X}$ is Hilbertian and $f = \|\cdot\|^2/2$, Corollary 3.35(ii) reduces to the classical characterization of metric projections onto closed convex sets.

3.6. Subgradient *D*-projections. The D-projection onto a closed convex set may be hard to compute. If the set is specified as a lower level set, it can be approximated by the D-projection onto a separating hyperplane, which is much easier to compute. In the traditional case when $\mathcal{X}$ is Hilbertian and $f = \|\cdot\|^2/2$, this is a standard approach which goes back to [73] (see also [6, 37, 60]). In the context of Bregman distances, we shall define subgradient D-projections as follows (see also [27, 59] for special instances).

DEFINITION 3.37. *Suppose that*

$$\begin{cases} \mathcal{X} \text{ is reflexive and } f \text{ is Legendre,} \\ g\colon \mathcal{X} \to \left]-\infty,+\infty\right] \text{ is lower semicontinuous and convex,} \\ \operatorname{lev}_{\leq 0} g \cap \operatorname{int\,dom} f \neq \varnothing \text{ and } \operatorname{dom} f \subset \operatorname{dom} g. \end{cases} \tag{3.26}$$

For every $x \in \operatorname{int\,dom} f$ *and* $x^* \in \partial g(x)$, *set*

$$G(x,x^*) = \big\{ y \in \mathcal{X} \mid \langle x - y, x^* \rangle \geq g(x) \big\}. \tag{3.27}$$

The operator

$$Q_g\colon \operatorname{int\,dom} f \to \mathcal{X}\colon x \mapsto \big\{ P_{G(x,x^*)} x \mid x^* \in \partial g(x) \big\} \tag{3.28}$$

is the subgradient D-projector *onto* $\operatorname{lev}_{\leq 0} g$.

Note that $G(x,x^*)$ is a proper closed half-space if $x^* \neq 0$ and the whole space $\mathcal{X}$ otherwise; the latter may occur only when $x \in \operatorname{Argmin} g$.

PROPOSITION 3.38. *Suppose that* (3.26) *is in force and let* Q_g *be the subgradient* D*-projector onto* $\operatorname{lev}_{\leq 0} g$. *Then*

(i) $\operatorname{Fix} Q_g = \operatorname{lev}_{\leq 0} g \cap \operatorname{int\,dom} f$;

(ii) $Q_g \in \mathfrak{B}$.

Proof. Fix $x \in \operatorname{int\,dom} f$ and $x^* \in \partial g(x)$. Since $\operatorname{int\,dom} f \subset \operatorname{int\,dom} g \subset \operatorname{dom} \partial g$, $\partial g(x) \neq \varnothing$ and the closed convex set $G(x,x^*)$ is well defined. Moreover, (2.2) yields

$$(\forall y \in \operatorname{lev}_{\leq 0} g) \quad \langle y - x, x^* \rangle \leq g(y) - g(x) \leq -g(x). \tag{3.29}$$

Therefore, $\operatorname{lev}_{\leq 0} g \subset G(x,x^*)$ and, in turn, $G(x,x^*) \cap \operatorname{int\,dom} f \neq \varnothing$. Hence, Corollary 3.35(i) asserts that $P_{G(x,x^*)}$ is single-valued with $\operatorname{ran} P_{G(x,x^*)} \subset \operatorname{int\,dom} f = \operatorname{dom} P_{G(x,x^*)}$, whence $\operatorname{ran} Q_g \subset \operatorname{int\,dom} f = \operatorname{dom} Q_g$. (i) Take $y \in \mathcal{X}$. Then it follows from Proposition 3.32(ii)(b) that

$$\begin{aligned} y \in \operatorname{Fix} Q_g &\Leftrightarrow (\exists\, y^* \in \partial g(y)) \;\; y = P_{G(y,y^*)} y \\ &\Leftrightarrow (\exists\, y^* \in \partial g(y)) \;\; y \in G(y,y^*) \cap \operatorname{int\,dom} f \\ &\Leftrightarrow (\exists\, y^* \in \partial g(y)) \;\; 0 = \langle y - y, y^* \rangle \geq g(y) \text{ and } y \in \operatorname{int\,dom} f \\ &\Leftrightarrow y \in \operatorname{lev}_{\leq 0} g \cap \operatorname{int\,dom} f. \end{aligned}$$

Thus, $\operatorname{Fix} Q_g = \operatorname{lev}_{\leq 0} g \cap \operatorname{int\,dom} f$. (ii) To show that $Q_g \in \mathfrak{B}$ observe that Corollary 3.35(ii) implies that $G(x,x^*) \subset H(x, P_{G(x,x^*)} x)$. Consequently, $\operatorname{Fix} Q_g \subset \operatorname{lev}_{\leq 0} g \subset G(x,x^*) \subset H(x, P_{G(x,x^*)} x)$, where $(x, P_{G(x,x^*)} x)$ is an arbitrary point in $\operatorname{gr} Q_g$. Altogether, $Q_g \in \mathfrak{B}$. ◻

3.7. Relaxed parallel combination of $\mathfrak{B}$-class operators. The following proposition describes a scheme to aggregate $\mathfrak{B}$-class operators in order to create a new $\mathfrak{B}$-class operator.

PROPOSITION 3.39. *Suppose that* $\mathcal{X}$ *is reflexive and that* f *is Legendre. Let* $(T_i)_{i \in I}$ *be a finite family of operators in* $\mathfrak{B}$ *such that* $\bigcap_{i \in I} \operatorname{Fix} T_i \neq \varnothing$, *let* $(\omega_i)_{i \in I}$ *be* *weights in* $]0,1]$ *such that* $\sum_{i\in I}\omega_i = 1$*, and let* λ *be a relaxation parameter in* $]0,1]$*. For every* $x \in \operatorname{int}\operatorname{dom} f$*, select* $(u_i)_{i\in I} \in \times_{i\in I} T_i x$*, put*

$$H(x) = \big\{y \in \mathcal{X} \mid \langle y, x^*\rangle \le \eta(x)\big\}, \tag{3.30}$$

where

$$\begin{cases} x^* = \nabla f(x) - \sum_{i\in I}\omega_i \nabla f(u_i),\\ \eta(x) = \sum_{i\in I}\omega_i \left\langle x + \lambda(u_i - x), \nabla f(x) - \nabla f(u_i)\right\rangle, \end{cases} \tag{3.31}$$

and define $T\colon \operatorname{int}\operatorname{dom} f \to \mathcal{X}\colon x \mapsto P_{H(x)}x$*. Then the following hold:*

(i) T *is single-valued on* $\operatorname{dom} T = \operatorname{int}\operatorname{dom} f \supset \operatorname{ran} T$*.*

(ii) *For every* $x \in \operatorname{int}\operatorname{dom} f$*, the following statements are equivalent:*

- (a) $x \in \bigcap_{i\in I}\operatorname{Fix} T_i$*.*
- (b) $x^* = 0$*.*
- (c) $H(x) = \mathcal{X}$*.*
- (d) $x \in H(x)$*.*
- (e) $x \in \operatorname{Fix} T$*.*

(iii) $\operatorname{Fix} T = \bigcap_{i\in I}\operatorname{Fix} T_i$*.*

(iv) $\overline{\operatorname{Fix}}\, T = \bigcap_{i\in I}\overline{\operatorname{Fix}}\, T_i$*.*

(v) $(\forall x \in \operatorname{int}\operatorname{dom} f)\ H(x) = H(x, Tx)$*.*

(vi) $T \in \mathfrak{B}$*.*

Proof. Fix $x \in \operatorname{int}\operatorname{dom} f$. (i) We first observe that the operator T is well defined. Indeed, since $(T_i)_{i\in I}$ lies in $\mathfrak{B}$, x^* and $\eta(x)$ are well defined and we have

$$\begin{aligned}
\varnothing &\neq \bigcap_{i\in I}\operatorname{Fix} T_i\\
&\subset (\operatorname{int}\operatorname{dom} f)\cap \bigcap_{i\in I} H(x,u_i)\\
&\subset (\operatorname{int}\operatorname{dom} f)\cap \bigcap_{i\in I}\big\{y\in\mathcal{X} \mid \langle y-u_i, \nabla f(x)-\nabla f(u_i)\rangle\\
&\qquad\qquad \le (1-\lambda)\,\langle x-u_i, \nabla f(x)-\nabla f(u_i)\rangle\big\}\\
&\subset (\operatorname{int}\operatorname{dom} f)\cap \Big\{y\in\mathcal{X} \mid \sum_{i\in I}\omega_i\,\langle y-u_i, \nabla f(x)-\nabla f(u_i)\rangle\\
&\qquad\qquad \le (1-\lambda)\sum_{i\in I}\omega_i\,\langle x-u_i, \nabla f(x)-\nabla f(u_i)\rangle\Big\}\\
&= (\operatorname{int}\operatorname{dom} f)\cap H(x),
\end{aligned} \tag{3.32}$$

where the second inclusion follows from the inequality $\lambda \le 1$ and the monotonicity of ∇f. Whence, $(\operatorname{int}\operatorname{dom} f)\cap H(x) \neq \varnothing$, and it follows from Corollary 3.35(i) that $P_{H(x)}x$ is a well-defined point in $\operatorname{int}\operatorname{dom} f$. (ii) Since f is essentially strictly convex, it is strictly convex on $\operatorname{int}\operatorname{dom} f$ and it follows from Proposition 3.3(vi) that (a) $\Rightarrow$ $(\forall i \in I)\ u_i = x \Rightarrow$ (b). (b) $\Rightarrow$ (c) Suppose $x^* = 0$ and fix $y \in \bigcap_{i\in I}\operatorname{Fix} T_i$. Then, since $(T_i)_{i\in I}$ lies in $\mathfrak{B}$,

$$\begin{aligned} 0 &\leq \sum_{i\in I} \omega_i \langle u_i - y, \nabla f(x) - \nabla f(u_i)\rangle \\ &= \eta(x) - \langle y, x^*\rangle - (1-\lambda)\sum_{i\in I} \omega_i \langle x - u_i, \nabla f(x) - \nabla f(u_i)\rangle \\ &\leq \eta(x). \end{aligned} \tag{3.33}$$

Accordingly, $H(x) = \mathcal{X}$. The implications (c) $\Rightarrow$ (d) $\Rightarrow$ $x = P_{H(x)}x$ $\Rightarrow$ (e) are clear in view of Proposition 3.32(ii)(b). (e) $\Rightarrow$ (a) We have

$$\begin{aligned} x \in \operatorname{Fix} T &\Leftrightarrow x = P_{H(x)}x \\ &\Leftrightarrow x \in H(x) \\ &\Leftrightarrow \langle x, x^*\rangle \leq \eta(x) \\ &\Leftrightarrow \lambda \sum_{i\in I} \omega_i \langle x - u_i, \nabla f(x) - \nabla f(u_i)\rangle \leq 0 \\ &\Leftrightarrow (\forall i \in I)\ \ x = u_i \in T_i x \\ &\Leftrightarrow x \in \bigcap_{i\in I} \operatorname{Fix} T_i, \end{aligned}$$

where the next to last equivalence follows from the strict monotonicity of ∇f on $\operatorname{int}\operatorname{dom} f$ (f is strictly convex on $\operatorname{int}\operatorname{dom} f$) and the inequalities $\lambda > 0$ and $\min_{i\in I} \omega_i > 0$. (iii) (i) and (ii) yield $\operatorname{Fix} T = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{Fix} T = \bigcap_{i\in I}(\operatorname{Fix} T_i \cap \operatorname{int}\operatorname{dom} f) = \bigcap_{i\in I} \operatorname{Fix} T_i$. (iv) Set $(\forall i \in I)\ F_i = \bigcap_{(x,u)\in \operatorname{gr} T_i} H(x,u)$. Then (iii) and Proposition 3.3(iv) yield $\operatorname{Fix} T = (\operatorname{int}\operatorname{dom} f) \cap \bigcap_{i\in I} F_i$. Therefore, by Lemma 3.2 and Proposition 3.3(vii),

$$\overline{\operatorname{Fix}}\, T = \overline{\operatorname{dom}}\, f \cap \bigcap_{i\in I} F_i = \bigcap_{i\in I}(F_i \cap \overline{\operatorname{dom}}\, f) = \bigcap_{i\in I} \overline{\operatorname{Fix}}\, T_i. \tag{3.34}$$

(v) By Corollary 3.35(ii), we always have $H(x) \subset H(x, P_{H(x)}x) = H(x, Tx)$. Now suppose $x \in H(x)$. Then (ii) yields $H(x) = \mathcal{X} = H(x,x) = H(x, P_{H(x)}x) = H(x, Tx)$. Next, suppose $x \notin H(x)$. Then (ii) yields $x^* \neq 0$ and $H(x)$ is therefore a proper closed half-space in $\mathcal{X}$. On the other hand, $x \neq P_{H(x)}x = Tx$ and, since ∇f is injective [8, Thm. 5.10], $\nabla f(x) \neq \nabla f(Tx)$. Consequently, $H(x, Tx)$ is also a proper closed half-space in $\mathcal{X}$. Since $Tx \in H(x) \cap \operatorname{bdry} H(x, Tx)$ and $H(x) \subset H(x, Tx)$, we conclude $H(x) = H(x, Tx)$. (vi) It follows successively from (iii), (3.32), and (v) that $\operatorname{Fix} T = \bigcap_{i\in I} \operatorname{Fix} T_i \subset H(x) = H(x, Tx)$. In view of (i), the proof is complete. ∎

4. Bregman monotonicity.

4.1. Properties. D-monotonicity was introduced in Definition 1.2. We first collect some elementary properties.

PROPOSITION 4.1. *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{X}$ which is D-monotone with respect to a set $S \subset \mathcal{X}$. Then the following hold:*

(i) $(\forall x \in S \cap \operatorname{dom} f)\ \big(D(x, x_n)\big)_{n\in\mathbb{N}}$ *converges.*
(ii) $(\forall n \in \mathbb{N})\ \ D_S(x_{n+1}) \leq D_S(x_n)$.
(iii) $\big(D_S(x_n)\big)_{n\in\mathbb{N}}$ *converges.*
(iv) $(\forall (x, x') \in (S \cap \operatorname{dom} f)^2)\ \big(\langle x - x', \nabla f(x_n)\rangle\big)_{n\in\mathbb{N}}$ *converges.*

(v) $(x_n)_{n\in\mathbb{N}}$ *is bounded if, for some* $z \in S \cap \operatorname{dom} f$, *the set* $\operatorname{lev}_{\leq D(z,x_0)} D(z,\cdot)$ *is bounded. This is true in particular if* $S \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$, $\mathcal{X}$ *is reflexive, and one of the following properties is satisfied:*
(a) f *is supercoercive;*
(b) $\dim \mathcal{X} < +\infty$ *and* $\operatorname{dom} f^*$ *is open.*

Proof. (i) and (ii) are immediate consequences of Definition 1.2, and (iii) follows from (ii). (iv) Take x and x' in $S \cap \operatorname{dom} f$. By (i), the sequences $\big(f(x_n) + \langle x - x_n, \nabla f(x_n)\rangle\big)_{n\in\mathbb{N}}$ and $\big(f(x_n) + \langle x' - x_n, \nabla f(x_n)\rangle\big)_{n\in\mathbb{N}}$ converge and so does their difference $\big(\langle x - x', \nabla f(x_n)\rangle\big)_{n\in\mathbb{N}}$. (v) By definition, for every $x \in S\cap \operatorname{dom} f$, $(x_n)_{n\in\mathbb{N}}$ lies in $\operatorname{lev}_{\leq D(z,x_0)} D(z,\cdot)$. The second assertion follows from [8, Lemma 7.3(viii) and (ix)], which asserts that $D(z,\cdot)$ is coercive under the stated assumptions if $z \in \operatorname{int}\operatorname{dom} f$. ∎

The following example shows that the conclusion of Proposition 4.1(v) may hold even though the properties (a) and (b) are not satisfied.

EXAMPLE 4.2. *Let* $\mathcal{X} = \ell_2(\mathbb{N})$ *and define*

(4.1)
$$f\colon \mathcal{X} \to \,]-\infty,+\infty] : x = (\xi_k)_{k\in\mathbb{N}} \mapsto \begin{cases} \sum_{k\in\mathbb{N}} \xi_k - \ln(1+\xi_k) & \text{if } (\forall k\in\mathbb{N})\ \ \xi_k > -1,\\ +\infty & \text{otherwise.}\end{cases}$$

Then f *is Legendre and* $\operatorname{dom} f$ *is open. Moreover,* $\operatorname{lev}_{\leq\eta} D(0,\cdot)$ *is bounded for* $\eta > 0$ *sufficiently small.*

Proof. We only sketch the arguments, as the example is not utilized elsewhere. Observe that f is separable: $(\forall x \in \mathcal{X})\ f(x) = \sum_{k\in\mathbb{N}} h(\xi_k)$, where

$$(\forall \xi \in \mathbb{R})\ \ h(\xi) = \begin{cases} \xi - \ln(1+\xi) & \text{if } \ \xi > -1,\\ +\infty & \text{otherwise.}\end{cases} \tag{4.2}$$

Using calculus, one verifies that $\operatorname{dom} f = \big\{x \in \mathcal{X} \mid (\forall k \in \mathbb{N})\ \ \xi_k > -1\big\}$, which is open. Also, f is Gâteaux-differentiable on its domain with $\nabla f(x) = \big(\xi_k/(1+\xi_k)\big)_{k\in\mathbb{N}}$. Hence f is essentially smooth. Now $(\forall x \in \mathcal{X})\ f^*(x) = f(-x)$. Thus f^* is essentially smooth as well. By [8, Thm. 5.4], f is essentially strictly convex. Altogether, f is Legendre. Let $\alpha = \ln(2) - 1/2$. A careful analysis of the Bregman distance D_h associated with h reveals that $D_h(0,\xi) < \alpha \Rightarrow |\xi| < 1 \Rightarrow D_h(0,\xi) \geq \alpha|\xi|^2$. (In passing, we point out that $D_h(0,\cdot)$ is convex precisely on $]-1,+1[$.) Fix $\eta \in [0,\alpha[$ and $x \in \mathcal{X}$ such that $D(0,x) \leq \eta$. Then $(\forall k \in \mathbb{N})\ D_h(0,\xi_k) \geq \alpha|\xi_k|^2$. Summing yields $\eta \geq D(0,x) \geq \alpha\|x\|^2$, whence $x \in B(0;\sqrt{\eta/\alpha})$. ∎

The next two assumptions will be quite helpful in the analysis of the convergence of D-monotone sequences.

CONDITION 4.3. *Given* $S \subset \mathcal{X}$, *for every bounded sequence* $(x_n)_{n\in\mathbb{N}}$ *in* $\operatorname{int}\operatorname{dom} f$, *one has*

$$\begin{cases} x \in \mathfrak{W}(x_n)_{n\in\mathbb{N}} \cap S,\\ x' \in \mathfrak{W}(x_n)_{n\in\mathbb{N}} \cap S,\\ (x_n)_{n\in\mathbb{N}} \text{ is } D\text{-monotone with respect to } S\end{cases} \quad\Rightarrow\quad x = x'. \tag{4.3}$$

CONDITION 4.4. *For all bounded sequences* $(x_n)_{n\in\mathbb{N}}$ *and* $(y_n)_{n\in\mathbb{N}}$ *in* $\operatorname{int}\operatorname{dom} f$, *one has*

$$D(x_n,y_n) \to 0 \quad\Rightarrow\quad x_n - y_n \to 0. \tag{4.4}$$

These two assumptions cover familiar situations, as the following examples show.

EXAMPLE 4.5. *Suppose that S is a subset of $\mathcal{X}$ such that $S \cap \overline{\operatorname{dom}} f$ is a singleton. Then Condition* 4.3 *is satisfied.*

Proof. Take $(x_n)_{n\in\mathbb{N}}$ in $\operatorname{int}\operatorname{dom} f$. Then $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \overline{\operatorname{dom}} f$ and, therefore, $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \cap S$ is at most a singleton. ∎

EXAMPLE 4.6. *Suppose that $S \subset \operatorname{int}\operatorname{dom} f$ is convex, $f|_S$ is strictly convex, and ∇f is sequentially weak-to-weak$*$ continuous at every point in S. Then Condition* 4.3 *is satisfied.*

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence which is D-monotone with respect to S. Then $x_{k_n} \rightharpoonup x \in S$ and $x_{l_n} \rightharpoonup x' \in S$ imply $\nabla f(x_{k_n}) \overset{*}{\rightharpoonup} \nabla f(x)$ and $\nabla f(x_{l_n}) \overset{*}{\rightharpoonup} \nabla f(x')$. Proposition 4.1(iv) therefore forces $\langle x - x', \nabla f(x)\rangle = \langle x - x', \nabla f(x')\rangle$; hence $\langle x - x', \nabla f(x) - \nabla f(x')\rangle = 0$. Since ∇f is strictly monotone on S, we get $x = x'$. ∎

Our next example requires the following lemma.

LEMMA 4.7. *Suppose that $\varepsilon \in \,]0, +\infty[$, $x \in \operatorname{dom} f$, and $y \in \operatorname{int}\operatorname{dom} f$. Then there exists $z \in \operatorname{int}\operatorname{dom} f$ such that $\|x - z\| \leq \varepsilon$ and $|D(x,y) - D(z,y)| \leq \varepsilon$.*

Proof. Put $(\forall \alpha \in [0,1[)\ x_\alpha = (1-\alpha)y + \alpha x$. Then $(x_\alpha)_{\alpha\in[0,1[}$ lies in $\operatorname{int}\operatorname{dom} f$, $\lim_{\alpha\uparrow 1^-} x_\alpha = x$ and, by (3.6), $\lim_{\alpha\uparrow 1^-} D(x_\alpha, y) = D(x,y)$. Thus, for α sufficiently close to 1, we can take $z = x_\alpha$. ∎

We now recall the notion of a Bregman/Legendre function in $\mathbb{R}^N$, which covers numerous functions of importance in convex optimization [7]. This notion will allow us to describe a finite-dimensional setting in which Condition 4.3 holds.

DEFINITION 4.8. *Suppose that $\mathcal{X} = \mathbb{R}^N$ and f is Legendre. Then f is* Bregman/Legendre*, if each of the following conditions is satisfied:*

(i) $\operatorname{dom} f^*$ *is open.*

(ii) $(\forall x \in \operatorname{dom} f \smallsetminus \operatorname{int}\operatorname{dom} f)$ $D(x,\cdot)$ *is coercive.*

(iii) $$\begin{cases} x \in \operatorname{dom} f \smallsetminus \operatorname{int}\operatorname{dom} f, \\ (y_n)_{n\in\mathbb{N}} \text{ in } \operatorname{int}\operatorname{dom} f, \\ y_n \to y \in \operatorname{bdry}\operatorname{dom} f, \\ \big(D(x,y_n)\big)_{n\in\mathbb{N}} \text{ bounded} \end{cases} \Rightarrow \quad D(y,y_n) \to 0.$$

(iv) $$\begin{cases} (x_n)_{n\in\mathbb{N}} \text{ in } \operatorname{int}\operatorname{dom} f, \\ (y_n)_{n\in\mathbb{N}} \text{ in } \operatorname{int}\operatorname{dom} f, \\ x_n \to x \in \operatorname{dom} f \smallsetminus \operatorname{int}\operatorname{dom} f, \\ y_n \to y \in \operatorname{dom} f \smallsetminus \operatorname{int}\operatorname{dom} f, \\ D(x_n,y_n) \to 0 \end{cases} \Rightarrow \quad x = y.$$

EXAMPLE 4.9. *Suppose that $\mathcal{X} = \mathbb{R}^N$, f is Bregman/Legendre, and S is a subset of $\mathcal{X}$ such that $S \cap \operatorname{dom} f \neq \varnothing$. Then Condition* 4.3 *is satisfied.*

Proof. Let us start with two useful facts, namely

$$\begin{cases} x \in \operatorname{dom} f, \\ (y_n)_{n\in\mathbb{N}} \text{ in } \operatorname{int}\operatorname{dom} f, \\ y_n \to y, \\ \big(D(x,y_n)\big)_{n\in\mathbb{N}} \text{ bounded} \end{cases} \Rightarrow \quad \begin{cases} D(y,y_n) \to 0, \\ y \in \operatorname{dom} f \end{cases} \tag{4.5}$$

and

$$
\begin{cases} x \in \operatorname{dom} f, \\ (y_n)_{n\in\mathbb{N}} \text{ in } \operatorname{int}\operatorname{dom} f, \\ y_n \to y \in \operatorname{dom} f, \\ D(x, y_n) \to 0 \end{cases} \quad \Rightarrow \quad x = y. \tag{4.6}
$$

If $x \in \operatorname{int}\operatorname{dom} f$, (4.5) follows from [7, Thm. 3.8(ii)]. On the other hand, if $x \in \operatorname{dom} f \smallsetminus \operatorname{int}\operatorname{dom} f$, (4.5) follows from [7, Prop. 3.3] if $y \in \operatorname{int}\operatorname{dom} f$ and from [7, Def. 5.2.BL2] if $y \in \operatorname{bdry}\operatorname{dom} f$. We now turn to (4.6). If x or y belongs to $\operatorname{int}\operatorname{dom} f$, it suffices to apply [7, Thm. 3.9(iii)]. Otherwise, $\{x, y\} \subset \operatorname{dom} f \smallsetminus \operatorname{int}\operatorname{dom} f$ and Lemma 4.7 ensures that, for every $n \geq 1$, we can find a point $x_n \in \operatorname{int}\operatorname{dom} f$ such that $\|x - x_n\| \leq 1/n$ and $|D(x, y_n) - D(x_n, y_n)| \leq 1/n$. Therefore, $x_n \to x$ and, since $D(x, y_n) \to 0$ by assumption, $D(x_n, y_n) \to 0$. It then follows from [7, Def. 5.2.BL3] that $x = y$. Now let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence which is D-monotone with respect to S and let $z \in S \cap \operatorname{dom} f$. Suppose $x_{k_n} \to x \in S$ and $x_{l_n} \to x' \in S$. Since by D-monotonicity the sequences $\big(D(z, x_{k_n})\big)_{n\in\mathbb{N}}$ and $\big(D(z, x_{l_n})\big)_{n\in\mathbb{N}}$ are bounded, (4.5) yields $D(x, x_{k_n}) \to 0$, $D(x', x_{l_n}) \to 0$, and $\{x, x'\} \subset S \cap \operatorname{dom} f$. However, it follows from Proposition 4.1(i) that $D(x, x_{k_n}) \to 0 \Rightarrow D(x, x_n) \to 0 \Rightarrow D(x, x_{l_n}) \to 0$. In view of (4.6), we conclude $x = x'$, as required. ◻

Following [25], we say that f is uniformly convex on bounded sets if, for every bounded set $B \subset \mathcal{X}$, one has

$$
(\forall t \in \,]0, +\infty[) \quad \inf \mu\big(B \cap \operatorname{dom} f, t\big) > 0, \tag{4.7}
$$

where

$$
\mu\colon \operatorname{dom} f \times [0, +\infty[\,\to [0, +\infty] : (x, t) \mapsto \inf_{\substack{\|x-y\|=t \\ y\in\operatorname{dom} f}} \frac{f(x) + f(y)}{2} - f\left(\frac{x+y}{2}\right). \tag{4.8}
$$

Examples of such functions are given in [84].

The next result gives sufficient conditions for Condition 4.4 to hold. (See also [22] and [82] for item (ii).)

EXAMPLE 4.10. *Condition* 4.4 *is satisfied whenever one of the following is true:*

(i) f *is uniformly convex on bounded sets.*

(ii) $\mathcal{X} = \mathbb{R}^N$, $\operatorname{dom} f$ *is closed, and* $f|_{\operatorname{dom} f}$ *is strictly convex and continuous.*

(iii) $\mathcal{X} = \mathbb{R}$ *and* $f|_{\operatorname{dom} f}$ *is strictly convex.*

Proof. (i) is a direct consequence of [25, Prop. 4.2]. (ii) and (iii) are special cases of (i) by [85, Prop. 3.6.6(i)]. ◻

In passing, we note that it follows from [85, Thm. 3.5.13] that item (i) of Example 4.10 forces the underlying space $\mathcal{X}$ to be reflexive.

The above assumptions lead to remarkably simple weak and strong convergence criteria for D-monotone sequences. In the case when $\mathcal{X}$ is Hilbertian and $f = \|\cdot\|^2/2$, Conditions 4.3 and 4.4 are satisfied and these criteria can essentially be found in [53] (see also [6] and [40]). Recall (see section 2) that $\mathfrak{S}$ denotes the set of strong cluster points of a sequence.

THEOREM 4.11. *Let* $(x_n)_{n\in\mathbb{N}}$ *be a bounded sequence in* $\mathcal{X}$ *which is* D*-monotone with respect to a set* $S \subset \mathcal{X}$. *Suppose that* $\mathcal{X}$ *is reflexive and Condition* 4.3 *is satisfied. Then*

(i) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $S \cap \overline{\operatorname{dom}} f$ *if and only if* $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset S$;

(ii) *supposing that* $x_n \rightharpoonup x \in S \cap \operatorname{int}\operatorname{dom} f$ *and Condition* 4.4 *is satisfied, then* $x_n \to x$ *if and only if* $\mathfrak{S}(x_n)_{n\in\mathbb{N}} \neq \varnothing$.

Proof. (i) Necessity is clear. To prove sufficiency, suppose that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset S$ and take x and x' in $\mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x$ and $x_{l_n} \rightharpoonup x'$. Then x and x' lie in S and (4.3) forces $x = x'$. Since $\mathcal{X}$ is reflexive and $(x_n)_{n\in\mathbb{N}}$ is bounded, we conclude $x_n \rightharpoonup x$. Furthermore, since $\overline{\operatorname{dom}}\, f \ni x_n \rightharpoonup x$ and $\overline{\operatorname{dom}}\, f$ is weakly closed, $x \in \overline{\operatorname{dom}}\, f$. (ii) Necessity is clear. To prove sufficiency, suppose that Condition 4.4 is satisfied, $x \in S \cap \operatorname{int}\operatorname{dom} f$, and $\mathfrak{S}(x_n)_{n\in\mathbb{N}} \neq \varnothing$, i.e., some subsequence $(x_{k_n})_{n\in\mathbb{N}}$ converges strongly. Since $x_n \rightharpoonup x$, we must have $x_{k_n} \to x$. In turn, [8, Lemma 7.3(x)] yields $D(x, x_{k_n}) \to 0$ and it follows from Proposition 4.1(i) that $D(x, x_n) \to 0$. In view of (4.4), we conclude $x_n \to x$. ☐

4.2. Construction.

ALGORITHM 4.12. *Starting with* $x_0 \in \operatorname{int}\operatorname{dom} f$, *at every iteration* $n \in \mathbb{N}$, *select first* $T_n \in \mathfrak{B}$ *and then* $x_{n+1} \in T_n x_n$.

PROPOSITION 4.13. *Let* $(x_n)_{n\in\mathbb{N}}$ *be an arbitrary orbit of Algorithm* 4.12. *Suppose that*

$$\bigcap_{n\in\mathbb{N}} \operatorname{Fix} T_n \neq \varnothing, \quad S \subset \bigcap_{n\in\mathbb{N}} \overline{\operatorname{Fix}}\, T_n, \quad \textit{and} \quad S \cap \operatorname{dom} f \neq \varnothing. \tag{4.9}$$

Then

(i) *if* $f|_{\operatorname{int}\operatorname{dom} f}$ *is strictly convex,* $(x_n)_{n\in\mathbb{N}}$ *is* D*-monotone with respect to* S;
(ii) $\sum_{n\in\mathbb{N}} D(x_{n+1}, x_n) < +\infty$.

Proof. (i) Proposition 3.3(viii) yields $(\forall n \in \mathbb{N})(\forall y \in \overline{\operatorname{Fix}}\, T_n)\ D(y, x_{n+1}) \leq D(y, x_n)$. (ii) Fix $y \in \bigcap_{n\in\mathbb{N}} \operatorname{Fix} T_n$. Then Proposition 3.3(i) yields the stronger statement

$$(\forall n \in \mathbb{N})\ \ D(y, x_{n+1}) \leq D(y, x_n) - D(x_{n+1}, x_n). \tag{4.10}$$

Therefore $\sum_{n\in\mathbb{N}} D(x_{n+1}, x_n) \leq D(y, x_0)$. ☐

THEOREM 4.14. *Let* $(x_n)_{n\in\mathbb{N}}$ *be an arbitrary bounded orbit of Algorithm* 4.12. *Suppose that* $\mathcal{X}$ *is reflexive, that* $f|_{\operatorname{int}\operatorname{dom} f}$ *is strictly convex, and that* (4.9) *is satisfied. Suppose in addition that Condition* 4.3 *is satisfied and that*

$$\sum_{n\in\mathbb{N}} D(x_{n+1}, x_n) < +\infty \quad \Rightarrow \quad \mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset S. \tag{4.11}$$

Then

(i) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point* $x \in S$;
(ii) *the convergence is strong in* (i) *if* $x \in \operatorname{int}\operatorname{dom} f$, *Condition* 4.4 *is satisfied, and*

$$\sum_{n\in\mathbb{N}} D(x_{n+1}, x_n) < +\infty \quad \Rightarrow \quad \mathfrak{S}(x_n)_{n\in\mathbb{N}} \neq \varnothing. \tag{4.12}$$

Proof. Combine Theorem 4.11 and Proposition 4.13. ☐

5. Parallel block-iterative *D*-monotone algorithm.

5.1. Objective. For the remainder of this paper, we assume that

$$\begin{cases} \mathcal{X} \text{ is reflexive and } f \text{ is Legendre,} \\ (S_i)_{i\in I} \text{ is a countable family of closed convex subsets of } \mathcal{X}, \\ (\operatorname{int}\operatorname{dom} f) \cap \bigcap_{i\in I} S_i \neq \varnothing, \\ S = \overline{\operatorname{dom}}\, f \cap \bigcap_{i\in I} S_i. \end{cases} \tag{5.1}$$

The purpose of this section is to develop a relaxed, parallel, block-iterative algorithm to solve the convex feasibility problem

$$\text{Find } x \in S. \tag{5.2}$$

5.2. Algorithm.

ALGORITHM 5.1. *Starting with* $x_0 \in \operatorname{int} \operatorname{dom} f$*, take at every iteration* n

① *a nonempty finite index set* $I_n \subset I$,
② *operators* $(T_{i,n})_{i \in I_n}$ *in* $\mathfrak{B}$ *such that* $(\forall i \in I_n)\ S_i \cap \operatorname{int}\operatorname{dom} f \subset \operatorname{Fix} T_{i,n}$,
③ *points* $(u_{i,n})_{i\in I_n} \in \times_{i\in I_n} T_{i,n}x_n$,
④ *weights* $(\omega_{i,n})_{i\in I_n}$ *in* $[0,1]$ *such that* $\sum_{i\in I_n}\omega_{i,n} = 1$,
⑤ *a relaxation parameter* $\lambda_n \in \,]0,1]$

and put

⑥ $x_n^* = \nabla f(x_n) - \sum_{i\in I_n}\omega_{i,n}\nabla f(u_{i,n})$,
⑦

$$\eta_n = \left\langle x_n, \nabla f(x_n) - \sum_{i\in I_n}\omega_{i,n}\nabla f(u_{i,n})\right\rangle - \lambda_n \sum_{i\in I_n}\omega_{i,n}\left\langle u_{i,n}-x_n, \nabla f(u_{i,n}) - \nabla f(x_n)\right\rangle,$$

⑧ $H_n = \big\{ y \in \mathcal{X} \mid \langle y, x_n^*\rangle \le \eta_n \big\}$.

Then set $x_{n+1} = P_{H_n}x_n$.

We now motivate this algorithm geometrically. At iteration n, x_n is given and a finite block of indices I_n is retained. Set $I_n^+ = \big\{ i \in I_n \mid \omega_{i,n} > 0 \big\}$. Then, using Lemma 3.2 for the first and last equality, step ② for the third inclusion, and (3.32) for the fourth inclusion,

$$\begin{aligned} S &= \overline{(\operatorname{int}\operatorname{dom} f) \cap \bigcap_{i\in I} S_i} \subset \overline{(\operatorname{int}\operatorname{dom} f) \cap \bigcap_{i\in I_n} S_i} \subset \overline{(\operatorname{int}\operatorname{dom} f) \cap \bigcap_{i\in I_n^+} S_i} \\ &\subset \overline{\bigcap_{i\in I_n^+} \operatorname{Fix} T_{i,n}} \subset \overline{(\operatorname{int}\operatorname{dom} f) \cap H_n} = \overline{\operatorname{dom}} f \cap H_n \subset H_n. \end{aligned} \tag{5.3}$$

Thus, H_n acts as an outer approximation to the intersection of the block of constraint sets $(\overline{\operatorname{dom}} f \cap S_i)_{i\in I_n}$ and, therefore, to S. More precisely, the block constraint $y \in \overline{\operatorname{dom}} f \cap \bigcap_{i\in I_n} S_i$ is replaced by the surrogate affine constraint $\langle y, x_n^*\rangle \le \eta_n$. The update x_{n+1} is then the D-projection of x_n onto H_n, i.e., the D-closest point to x_n which satisfies the surrogate constraint. (x_{n+1} is well defined by virtue of (5.1) and Corollary 3.35(i).) Naturally, such a point is considerably simpler to find than a point in $\overline{\operatorname{dom}} f \cap \bigcap_{i\in I_n} S_i$. In spirit, this type of surrogate constraint construction can be found—explicitly or implicitly—in several places in the literature, although not in the context of Bregman distances. (See, for instance, [39, 60] and the references therein.)

The parallel nature of the algorithm stems from the fact that the points $(u_{i,n})_{i\in I_n}$ at step ③ can be computed independently on concurrent processors. In addition, the algorithm has the ability to process variable blocks of constraints, which makes it possible to match closely the computational load of each iteration to the parallel processing architecture at hand. A discussion on the importance of block-processing for task scheduling on parallel architectures can be found in [33].

To shed more light on Algorithm 5.1, we first consider the case when $\mathcal{X}$ is Hilbertian and $f = \|\cdot\|^2/2$. Then, steps ⑥ and ⑦ become

$$\begin{cases} x_n^* = x_n - \sum_{i\in I_n}\omega_{i,n}u_{i,n}, \\ \eta_n = \big\langle x_n, x_n - \sum_{i\in I_n}\omega_{i,n}u_{i,n}\big\rangle - \lambda_n\sum_{i\in I_n}\omega_{i,n}\|u_{i,n}-x_n\|^2. \end{cases} \tag{5.4}$$

Furthermore, the updating step is explicitly given as

$$x_{n+1} = P_{H_n} x_n = x_n + \frac{\eta_n - \langle x_n, x_n^* \rangle}{\|x_n^*\|^2} x_n^* = x_n + \lambda_n L_n \left(\sum_{i \in I_n} \omega_{i,n} (u_{i,n} - x_n) \right), \tag{5.5}$$

where

$$L_n = \begin{cases} \dfrac{\sum_{i \in I_n} \omega_{i,n} \|u_{i,n} - x_n\|^2}{\| \sum_{i \in I_n} \omega_{i,n} (u_{i,n} - x_n)\|^2} & \text{if } x_n \notin \bigcap_{i \in I_n} S_i, \\ 1 & \text{otherwise.} \end{cases} \tag{5.6}$$

This is essentially the algorithm proposed in [41, section 6] (in this setting, the range of λ_n can be extended to $]0, 2[$), which itself contains those of [5, 6, 35, 37, 38, 60, 69] as special cases. In particular, if I is finite, $I_n \equiv I$, $\omega_{i,n} = \omega_i$, and $u_{i,n} = P_i x_n$, where P_i is the metric projector onto S_i, then (5.5)–(5.6) reduces to Pierra's classic extrapolated parallel projection method [72], which in turn can be traced back to Merzlyakov's method [63] for solving systems of linear inequalities in $\mathbb{R}^N$. Since $L_n \geq 1$ in (5.6), large extrapolations are possible in this algorithm by selecting $\lambda_n \approx 1$. It is known that these extrapolations yield significantly accelerated convergence in numerical experiments [36, 37, 50, 72] in comparison with purely averaged iterations, i.e.,

$$x_{n+1} = \sum_{i \in I_n} \omega_{i,n} u_{i,n}, \tag{5.7}$$

which can be derived from (5.5) by setting $\lambda_n = 1/L_n$.

Returning to the standing assumptions, let us now consider the parallel block-iterative update rule

$$\nabla f(x_{n+1}) = \sum_{i \in I_n} \omega_{i,n} \nabla f(u_{i,n}). \tag{5.8}$$

This alternative method for solving (5.2) was recently proposed by Censor and Herman in [29] (see also [31]) for the special case when $\mathcal{X} = \mathbb{R}^N$, I is finite, and $u_{i,n}$ is the D-projection of x_n onto S_i. If we assume that $\mathcal{X}$ is a Hilbert space and $f = \| \cdot \|^2/2$, then (5.8) reduces to (5.7) which, as noted above, is itself a special case of (5.5)–(5.6), hence of Algorithm 5.1. In general, however, we do not know whether (5.8) is always a particularization of Algorithm 5.1.

We now turn to Butnariu and Iusem's algorithmic framework [24] for solving (5.2). (In fact, they study the so-called stochastic convex feasibility problem, which is similar to (5.2) but allows for an uncountable index set I. Their framework requires measure theory for a precise formulation and their assumptions on the underlying function f are different from the ones made here. The reader is referred to [24] for further details.) Let $(R_i)_{i \in I}$ be a family of totally nonexpansive operators in the sense of [24]. (See also the paragraph following Definition 3.1.) Specialized to the case when I is finite, the update step in this algorithm is

$$x_{n+1} = \sum_{i \in I} \omega_i R_i x_n. \tag{5.9}$$

This resembles (5.8), except for notably absent gradients on both sides of the equation and for weights that do not depend on n. (If the R_i's are D-projectors, then

(5.9) can also be interpreted as a sequential algorithm in the product space X^I; see [11].) Note that if $\mathcal{X}$ is a Hilbert space and $f = \|\cdot\|^2/2$, then (5.9) once again corresponds to a parallel Cimmino-type algorithm, which is genuinely more restrictive than Algorithm 5.1 for this set-up.

While a detailed numerical and theoretical comparison of these algorithms lies beyond the scope of this paper, we remark that preliminary experiments suggest that Algorithm 5.1 is more flexible and faster than the one given by (5.8) and that Algorithm 5.1 is genuinely different from the method given by (5.9).

5.3. Convergence. The following notions were introduced in [6, Def. 3.7] and [41, Def. 6.5], respectively, to study the asymptotic behavior of Fejér monotone algorithms in Hilbert spaces. The former can be interpreted as an extension of the notion of demiclosedness at 0 [68] and the latter as an extension of the notion of demicompactness at 0 [70].

DEFINITION 5.2. *Algorithm* 5.1 *is*

- focusing *if for every bounded suborbit* $(x_{k_n})_{n\in\mathbb{N}}$ *it generates and every index* $i \in I$*,*

$$\begin{cases} i \in \bigcap_{n\in\mathbb{N}} I_{k_n}, \\ x_{k_n} \rightharpoonup x, \\ u_{i,k_n} - x_{k_n} \to 0 \end{cases} \quad \Rightarrow \quad x \in S_i; \tag{5.10}$$

- demicompactly regular *if there exists* $i \in I$*, called an* index of demicompact regularity*, such that for every bounded suborbit* $(x_{k_n})_{n\in\mathbb{N}}$ *it generates,*

$$\begin{cases} i \in \bigcap_{n\in\mathbb{N}} I_{k_n}, \\ u_{i,k_n} - x_{k_n} \to 0 \end{cases} \quad \Rightarrow \quad \mathfrak{S}(x_{k_n})_{n\in\mathbb{N}} \neq \varnothing. \tag{5.11}$$

We now describe the context in which the convergence of Algorithm 5.1 will be investigated.

CONDITION 5.3.

(i) *For some* $z \in \operatorname{dom} f \cap \bigcap_{i\in I} S_i$*,* $C = \operatorname{lev}_{\leq D(z,x_0)} D(z,\cdot)$ *is bounded.*

(ii) *For all sequences* $(u_n)_{n\in\mathbb{N}}$ *and* $(v_n)_{n\in\mathbb{N}}$ *in* C *such that* $(\forall n \in \mathbb{N})\ u_n \neq v_n$*, one has*

$$\frac{\langle u_n - v_n, \nabla f(u_n) - \nabla f(v_n)\rangle}{\|\nabla f(u_n) - \nabla f(v_n)\|} \to 0 \quad \Rightarrow \quad \nabla f(u_n) - \nabla f(v_n) \to 0. \tag{5.12}$$

CONDITION 5.4.

(i) $(\exists\, \delta_1 \in \,]0,1[)(\forall n \in \mathbb{N})(\exists\, j \in I_n)$

$$\|\nabla f(u_{j,n}) - \nabla f(x_n)\| = \max_{i\in I_n} \|\nabla f(u_{i,n}) - \nabla f(x_n)\| \ \text{ and } \ \omega_{j,n} \geq \delta_1.$$

(ii) $(\exists\, \delta_2 \in \,]0,1[)(\forall n \in \mathbb{N})\ \ \lambda_n \geq \delta_2$.

(iii) $(\forall i \in I)(\exists\, M_i \in \mathbb{N} \smallsetminus \{0\})(\forall n \in \mathbb{N})\ \ i \in \bigcup_{k=n}^{n+M_i-1} I_k$.

As will be seen subsequently, the above set of assumptions defines a broad framework which covers numerous practical situations. Note that, by virtue of (5.1), the quotient in (5.12) is well defined since ∇f is injective on $\operatorname{int}\operatorname{dom} f$ [8, Thm. 5.10]. Situations in which Condition 5.3(ii) is satisfied are detailed below. Note also that Condition 5.4(iii) imposes that every index i be activated at least once within any M_i consecutive iterations. This control rule, which has already been used in metric projection algorithms in Hilbert spaces [35, 37, 38, 60], provides great flexibility in the management of the constraints and the implementation of the algorithm. Condition 5.4(i) provides added flexibility by offering the possibility of setting $\omega_{i,n} = 0$ if the corresponding step size $\|\nabla f(u_{i,n}) - \nabla f(x_n)\|$ is not maximal. It is thereby possible to meet the control condition Condition 5.4(iii) without actually using the ith constraint in the construction of x_{n+1}.

Recall that an operator T from a Banach space $\mathcal{Y}$ to its dual $\mathcal{Y}^*$ is said to be uniformly monotone on $U \subset \operatorname{dom} T$ with modulus c if [86, section 25.3]

$$(\forall x \in U)(\forall y \in U) \quad \langle x-y, Tx - Ty\rangle \geq \|x-y\| \cdot c(\|x-y\|), \tag{5.13}$$

where $c\colon\ [0,+\infty[\ \to [0,+\infty[$ is a strictly increasing function such that $c(0) = 0$. In particular, T is said to be strongly monotone on U with constant $\alpha \in\ ]0,+\infty[$ if it is uniformly monotone on U with modulus $c\colon t \mapsto \alpha t$.

PROPOSITION 5.5. *Let z and C be as in Condition* 5.3(i). *Then Condition* 5.3(ii) *is satisfied in each of the following cases:*

(i) *∇f^* is uniformly monotone on $\nabla f(C)$.*
(ii) *∇f is Lipschitz-continuous on $\operatorname{dom} f = \mathcal{X}$.*
(iii) *$\mathcal{X} = \mathbb{R}^N$ and $\overline{C} \subset \operatorname{int}\operatorname{dom} f$.*
(iv) *$\mathcal{X} = \mathbb{R}^N$ and $z \in \operatorname{int}\operatorname{dom} f$.*

Proof. Let $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ be two sequences in C such that $(\forall n \in \mathbb{N})$ $u_n \neq v_n$. (i) Let c be the modulus of uniform monotonicity of ∇f^* on $\nabla f(C)$. Since ∇f is a bijection from $\operatorname{int}\operatorname{dom} f$ to $\operatorname{int}\operatorname{dom} f^*$ with inverse ∇f^* [8, Thm. 5.10] and since $C \subset \operatorname{int}\operatorname{dom} f$, we have $(\forall u \in C)(\forall v \in C)$ $\langle u - v, \nabla f(u) - \nabla f(v)\rangle \geq \|\nabla f(u) - \nabla f(v)\| \cdot c\,(\|\nabla f(u) - \nabla f(v)\|)$. Hence, since c is strictly increasing and $c(0) = 0$,

$$\begin{aligned} \frac{\langle u_n - v_n, \nabla f(u_n) - \nabla f(v_n)\rangle}{\|\nabla f(u_n) - \nabla f(v_n)\|} \to 0 \;&\Rightarrow\; c\big(\|\nabla f(u_n) - \nabla f(v_n)\|\big) \to 0 \\ &\Rightarrow\; \nabla f(u_n) - \nabla f(v_n) \to 0. \end{aligned} \tag{5.14}$$

(ii) $\Rightarrow$ (i) If ∇f is κ-Lipschitz-continuous on $\mathcal{X}$, then it follows from the Baillon–Haddad theorem [4, Cor. 10] that $(\forall x \in \mathcal{X})(\forall y \in \mathcal{X})$ $\langle x - y, \nabla f(x) - \nabla f(y)\rangle \geq \|\nabla f(x) - \nabla f(y)\|^2/\kappa$, i.e., ∇f^* is strongly monotone with constant $1/\kappa$. Consequently, ∇f^* is uniformly monotone on $\nabla f(C)$. (iii) Suppose

$$\frac{\langle u_n - v_n, \nabla f(u_n) - \nabla f(v_n)\rangle}{\|\nabla f(u_n) - \nabla f(v_n)\|} \to 0 \quad \text{and} \quad \nabla f(u_n) - \nabla f(v_n) \not\to 0. \tag{5.15}$$

Then there exists a strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ and $\varepsilon \in\ ]0,+\infty[$ such that $\inf_{n\in\mathbb{N}} \|\nabla f(u_{k_n}) - \nabla f(v_{k_n})\| \geq \varepsilon$. Since $(u_{k_n})_{n\in\mathbb{N}}$ lies in C, it is bounded and therefore possesses a convergent subsequence, say $u_{k_{l_n}} \to u$. As $(v_{k_{l_n}})_{n\in\mathbb{N}}$ is also bounded, we can assume (passing to a subsequence if necessary) that it converges, say $v_{k_{l_n}} \to v$. Since $\{u, v\} \subset \overline{C} \subset \operatorname{int}\operatorname{dom} f$ and ∇f is continuous at every point in $\operatorname{int}\operatorname{dom} f$ by [77, Thm. 25.5], taking the limit yields $\|\nabla f(u) - \nabla f(v)\| \geq \varepsilon$ and, by injectivity of ∇f on $\operatorname{int}\operatorname{dom} f$ [8, Thm. 5.10], $u \neq v$. On the other hand, (5.15) yields

$$\frac{\big\langle u_{k_{l_n}} - v_{k_{l_n}}, \nabla f(u_{k_{l_n}}) - \nabla f(v_{k_{l_n}})\big\rangle}{\|\nabla f(u_{k_{l_n}}) - \nabla f(v_{k_{l_n}})\|} \to 0, \tag{5.16}$$

and, since $\|\nabla f(u) - \nabla f(v)\| \neq 0$, taking the limit yields $\langle u - v, \nabla f(u) - \nabla f(v)\rangle = 0$. However, $f|_{\operatorname{int dom} f}$ is strictly convex and therefore ∇f is strictly monotone on $\operatorname{int dom} f \supset \{u, v\}$. This forces $u = v$ and we reach a contradiction. (iv) In view of (iii), it is enough to show that $\overline{C} \subset \operatorname{int dom} f$. If the inclusion does not hold, then we can find $y \in \operatorname{bdry dom} f$ and $(y_n)_{n\in\mathbb{N}}$ in C such that $y_n \to y$. Thus $\sup_{n\in\mathbb{N}} D(z, y_n) \leq D(z, x_0) < +\infty$, and, at the same time, since f is essentially smooth, [7, Thm. 3.8(i)] yields $D(z, y_n) \to +\infty$, which is absurd. ☐

Remark 5.6. A careful analysis of [85, Corollary 3.4.4"(iii)⇔(iv)"], [85, Proposition 3.5.1], and [85, Proposition 3.6.2] shows that Proposition 5.5(i) holds as soon as ∇f is Lipschitz on bounded sets. In turn, this condition is satisfied in L_p spaces for $f = \|\cdot\|_p^s$, where $\{p, s\} \subset [2, +\infty[$. (The proof relies on the case when $s = 2$; see also Example 5.11 below.)

Examples of Legendre functions f which satisfy Conditions 4.3, 4.4, and 5.3(i)–(ii) will be supplied in section 5.4. Our main convergence result can now be stated and proved.

THEOREM 5.7. *Suppose that Conditions* 4.3*,* 4.4*,* 5.3*, and* 5.4 *are satisfied, and let* $(x_n)_{n\in\mathbb{N}}$ *be an arbitrary orbit of Algorithm* 5.1*. Then, for every* $n \in \mathbb{N}$*,* x_n *and* $(u_{i,n})_{i\in I_n}$ *lie in the bounded set* C*. If, in addition, Algorithm* 5.1 *is focusing, then the following statements hold true:*

(i) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point* $x \in S$*.*

(ii) *If the weak limit* x *from* (i) *belongs to* $\operatorname{int dom} f$ *and the algorithm is demicompactly regular, then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly.*

Proof. For every $n \in \mathbb{N}$, set $T_n = P_{H_n}$ and $I_n^+ = \{i \in I_n \mid \omega_{i,n} > 0\}$. Since $x_0 \in \operatorname{int dom} f$ and, by Proposition 3.39(vi), $T_n \in \mathfrak{B}$, we recognize that

(5.17) Algorithm 5.1 is a special case of Algorithm 4.12.

Our goal is to apply Theorem 4.14 and we must start by verifying (4.9). First, considering (5.1), Algorithm 5.1②, and Proposition 3.39(iii), we obtain

(5.18)
$$(\forall n \in \mathbb{N})\ \ \varnothing \neq (\operatorname{int dom} f) \cap \bigcap_{i\in I} S_i \subset \bigcap_{i\in I_n^+} (S_i \cap \operatorname{int dom} f) \subset \bigcap_{i\in I_n^+} \operatorname{Fix} T_{i,n} = \operatorname{Fix} T_n.$$

Hence $\bigcap_{n\in\mathbb{N}} \operatorname{Fix} T_n \neq \varnothing$. In addition, (5.1), Lemma 3.2, and (5.18) yield

$$(\forall n \in \mathbb{N})\ \ S = \overline{\operatorname{dom}}\, f \cap \bigcap_{i\in I} S_i \subset \overline{\operatorname{Fix}}\, T_n. \tag{5.19}$$

Consequently, $S \subset \bigcap_{n\in\mathbb{N}} \overline{\operatorname{Fix}}\, T_n$. Next, we derive from (5.1) that

$$\varnothing \neq (\operatorname{int dom} f) \cap \bigcap_{i\in I} S_i \subset \operatorname{dom} f \cap \overline{\operatorname{dom}}\, f \cap \bigcap_{i\in I} S_i = \operatorname{dom} f \cap S. \tag{5.20}$$

Thus, (4.9) holds. Now, let z and C be as in Condition 5.3(i). It follows from (5.17) and Proposition 4.13(i) that the sequences $(x_n)_{n\in\mathbb{N}}$ and $(T_n x_n)_{n\in\mathbb{N}}$ are contained in C, which is bounded. In order to verify (4.11), some key facts must be established. Let us temporarily fix $n \in \mathbb{N}$. The first fact is supplied by the inclusion $x_{n+1} = P_{H_n} x_n \in H_n$, which yields

$$\|x_{n+1} - x_n\| \geq d_{H_n}(x_n). \tag{5.21}$$

Next, it follows from Condition 5.3(i), (5.1), Lemma 3.2, and Algorithm 5.1② that

$$(\forall i \in I_n) \;\; z \in S_i \cap \overline{\operatorname{dom}}\, f = \overline{S_i \cap \operatorname{int} \operatorname{dom} f} \subset \overline{\operatorname{Fix}}\, T_{i,n}. \tag{5.22}$$

Hence, for every $i \in I_n$, Algorithm 5.1③ and Proposition 3.3(viii) yield $D(z, u_{i,n}) \leq D(z, x_n) - D(u_{i,n}, x_n) \leq D(z, x_n)$. Therefore,

$$(\forall i \in I_n) \;\; u_{i,n} \in C. \tag{5.23}$$

Now, per Condition 5.4(ii), pick $j_n \in I_n$ such that

$$\|\nabla f(u_{j_n,n}) - \nabla f(x_n)\| = \max_{i \in I_n} \|\nabla f(u_{i,n}) - \nabla f(x_n)\| \quad \text{and} \quad \omega_{j_n,n} \geq \delta_1. \tag{5.24}$$

We claim that

$$\begin{cases} x_n \in \bigcap_{i \in I_n^+} \operatorname{Fix} T_{i,n} \quad \Leftrightarrow \quad u_{j_n,n} = x_n \quad \Leftrightarrow \quad \|\nabla f(u_{j_n,n}) - \nabla f(x_n)\| = 0, \\ x_n \notin \bigcap_{i \in I_n^+} \operatorname{Fix} T_{i,n} \quad \Rightarrow \quad d_{H_n}(x_n) \geq \delta_1 \delta_2 \dfrac{\langle u_{j_n,n} - x_n, \nabla f(u_{j_n,n}) - \nabla f(x_n)\rangle}{\|\nabla f(u_{j_n,n}) - \nabla f(x_n)\|}. \end{cases} \tag{5.25}$$

On the one hand, using Proposition 3.3(vi) and the injectivity of ∇f on $\operatorname{int} \operatorname{dom} f$ [8, Thm. 5.10], since (5.24) forces $j_n \in I_n^+$, we get $x_n \in \bigcap_{i \in I_n^+} \operatorname{Fix} T_{i,n} \Leftrightarrow (\forall i \in I_n^+)$ $u_{i,n} = x_n \Rightarrow u_{j_n,n} = x_n \Rightarrow \|\nabla f(u_{j_n,n}) - \nabla f(x_n)\| = 0 \Rightarrow (\forall i \in I_n)$ $\|\nabla f(u_{i,n}) - \nabla f(x_n)\| = 0 \Leftrightarrow (\forall i \in I_n)$ $u_{i,n} = x_n \Rightarrow (\forall i \in I_n^+)$ $u_{i,n} = x_n$. On the other hand, if $x_n \notin \bigcap_{i \in I_n^+} \operatorname{Fix} T_{i,n}$, then Proposition 3.39(ii) asserts that $x_n \notin H_n$ and $x_n^* \neq 0$, so that

$$\begin{aligned} d_{H_n}(x_n) &= \frac{\langle x_n, x_n^*\rangle - \eta_n}{\|x_n^*\|} \qquad (5.26) \\ &= \lambda_n \frac{\sum_{i \in I_n} \omega_{i,n} \langle u_{i,n} - x_n, \nabla f(u_{i,n}) - \nabla f(x_n)\rangle}{\|\sum_{i \in I_n} \omega_{i,n} (\nabla f(u_{i,n}) - \nabla f(x_n))\|} \\ &\geq \delta_2 \frac{\sum_{i \in I_n} \omega_{i,n} \langle u_{i,n} - x_n, \nabla f(u_{i,n}) - \nabla f(x_n)\rangle}{\sum_{i \in I_n} \omega_{i,n} \|\nabla f(u_{i,n}) - \nabla f(x_n)\|} \qquad (5.27) \\ &\geq \delta_1 \delta_2 \frac{\langle u_{j_n,n} - x_n, \nabla f(u_{j_n,n}) - \nabla f(x_n)\rangle}{\|\nabla f(u_{j_n,n}) - \nabla f(x_n)\|}, \qquad (5.28) \end{aligned}$$

where (5.26) follows from [80, Lemma I.1.2] and (5.27) from Condition 5.4(ii). Altogether, (5.25) is verified. The third key fact is derived from (5.23) and Proposition 2.3(i) as follows:

$$\begin{aligned} (\forall i \in I_n) \;\; \operatorname{diam}(C) \|\nabla f(u_{i,n}) - \nabla f(x_n)\| &\geq \langle u_{i,n} - x_n, \nabla f(u_{i,n}) - \nabla f(x_n)\rangle \\ &= D(u_{i,n}, x_n) + D(x_n, u_{i,n}) \\ &\geq D(u_{i,n}, x_n). \qquad (5.29) \end{aligned}$$

Let us now verify (4.11). To this end, let us fix $i \in I$ and $x \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x$. Because $x \in \overline{\operatorname{dom}} f$, it is sufficient to show

$$D(x_{n+1}, x_n) \to 0 \quad \Rightarrow \quad x \in S_i. \tag{5.30}$$

Let M_i be as in Condition 5.4(iii). After passing to a subsequence of $(x_{k_n})_{n\in\mathbb{N}}$ if necessary, we assume that, for every $n \in \mathbb{N}$, $k_{n+1} \geq k_n + M_i$. This guarantees the existence of a sequence $(p_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that

$$(\forall n \in \mathbb{N}) \;\; k_n \leq p_n \leq k_n + M_i - 1 < k_{n+1} \leq p_{n+1} \quad \text{and} \quad i \in I_{p_n}. \tag{5.31}$$

Now consider the subsequence $(x_{p_n})_{n\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$. The triangle inequality yields

(5.32)

$$(\forall n \in \mathbb{N}) \;\; \|x_{p_n} - x_{k_n}\| \leq \sum_{l=k_n}^{k_n+M_i-2} \|x_{l+1} - x_l\| \leq (M_i - 1) \max_{k_n \leq l \leq k_n+M_i-2} \|x_{l+1} - x_l\|.$$

Now suppose $D(x_{n+1}, x_n) \to 0$. Then (4.4) yields

$$x_{n+1} - x_n \to 0 \tag{5.33}$$

and it follows from (5.21) that $d_{H_n}(x_n) \to 0$. Consequently, we derive from (5.25), (5.23), and Condition 5.3(ii) that $\max_{j\in I_n} \|\nabla f(u_{j,n}) - \nabla f(x_n)\| \to 0$. In turn, (5.29) implies that $D(u_{i,p_n}, x_{p_n}) \to 0$ and, invoking (4.4) again, we obtain

$$u_{i,p_n} - x_{p_n} \to 0. \tag{5.34}$$

We also derive from (5.32) and (5.33) that $x_{p_n} - x_{k_n} \to 0$, whence $x_{p_n} \rightharpoonup x$. However, since the algorithm is focusing, (5.10) yields $x \in S_i$. Thus (5.30) holds and, consequently, the following conclusions can be drawn:

(i) Theorem 4.14(i) asserts that $(x_n)_{n\in\mathbb{N}}$ converges weakly to $x \in S$.

(ii) Suppose that $x \in \operatorname{int}\operatorname{dom} f$, $i \in I$ is an index of demicompact regularity, and $D(x_{n+1}, x_n) \to 0$. Then it results from (5.34) and (5.11) that (4.12) holds. In view of Condition 4.4, the strong convergence claim therefore follows from Theorem 4.14(ii). ☐

5.4. When all the assumptions hold. In this subsection, we describe scenarios in which all the assumptions required in Theorem 5.7 on f and on the constraint sets $(S_i)_{i\in I}$ are satisfied.

As a preamble to our first example, recall that if $\mathcal{X}$ is a Hilbert space, the Moreau–Yosida regularization of a proper lower semicontinuous convex function $\varphi\colon \mathcal{X} \to]-\infty, +\infty]$ with parameter $\gamma \in \,]0, +\infty[$ is the finite continuous convex function ${}_{\gamma}\varphi = \varphi \,\square\, \big(\|\cdot\|^2/(2\gamma)\big)$. Moreover, Moreau's classic proximal operator associated with φ and γ is given by Definition 3.16 for $f = \|\cdot\|^2/2$ and will be denoted by $\operatorname{Prox}_{\gamma}^{\varphi}$. It follows from Proposition 3.21(v) that $\operatorname{Prox}_{\gamma}^{\varphi}$ is defined everywhere and, from Proposition 3.22(ii)(d) and (c), that it is single-valued and firmly nonexpansive. Moreover [67, Prop. 7.d],

$$\nabla\, {}_{\gamma}\varphi = \frac{\operatorname{Id} - \operatorname{Prox}_{\gamma}^{\varphi}}{\gamma}. \tag{5.35}$$

Example 5.8 (Moreau–Yosida regularization). *Let $\mathcal{X}$ be a Hilbert space, set $w = \|\cdot\|^2/2$, and define $f\colon \mathcal{X} \to \mathbb{R}$ by*

$$f = (1+\gamma)w - {}_1\varphi, \tag{5.36}$$

where $\varphi\colon \mathcal{X} \to \left]-\infty,+\infty\right]$ is a proper lower semicontinuous convex function and $\gamma \in \left]0,+\infty\right[$. Then

$$\begin{aligned} D\colon (x,y) \mapsto\ & \gamma w(x-y) + w(x - \operatorname{Prox}_1^\varphi y) + \varphi(\operatorname{Prox}_1^\varphi y) \\ & - \big(w(x - \operatorname{Prox}_1^\varphi x) + \varphi(\operatorname{Prox}_1^\varphi x)\big) \end{aligned} \tag{5.37}$$

and Conditions 4.4 *and* 5.3 *are satisfied. If* $\operatorname{Prox}^\varphi_{\frac{\gamma}{1+\gamma}}$ *is affine or S is a singleton, then Condition* 4.3 *is also satisfied.*

Proof. The expression (5.37) is derived from (1.5) by simple algebra. Now set $\psi = w - {}_1\varphi$. Then

$$\psi = w - \inf_{x \in \operatorname{dom}\varphi} \varphi(x) + w(\cdot - x) = \sup_{x\in\operatorname{dom}\varphi} \langle x,\cdot\rangle - \varphi(x) - w(x) = (\varphi + w)^*. \tag{5.38}$$

Hence, ψ is a proper lower semicontinuous convex function as the conjugate of one such function. Since ψ is convex, $f = \psi + \gamma w$ is strongly (hence uniformly) convex and, in view of Example 4.10(i), Condition 4.4 is therefore satisfied. On the other hand, (5.35) yields $\operatorname{dom}\nabla f = \mathcal{X}$ and $\nabla f = \operatorname{Prox}_1^\varphi + \gamma\operatorname{Id}$. Hence f is essentially smooth by [8, Thm. 5.6]. Furthermore, since $\operatorname{Prox}_1^\varphi$ is firmly nonexpansive, it is 1-Lipschitz and therefore ∇f is $(1+\gamma)$-Lipschitz. Accordingly, Proposition 5.5(ii) asserts that Condition 5.3(ii) is satisfied. Next, using standard Hilbertian convex calculus, we obtain

$$\begin{aligned} f^* &= \big(\psi + \gamma w\big)^* = \psi^* \,\square\, (w/\gamma) = (\varphi + w) \,\square\, (w/\gamma) = {}_\gamma\big(\varphi + w\big) \\ &= \big({}_{\gamma/(1+\gamma)}\varphi\big)\big(\cdot/(1+\gamma)\big) + w/(1+\gamma). \end{aligned} \tag{5.39}$$

It therefore follows from (5.35) that

$$\operatorname{dom}\nabla f^* = \mathcal{X} \quad \text{and} \quad \nabla f^* = \frac{\operatorname{Id} - \operatorname{Prox}^\varphi_{\gamma/(1+\gamma)}\big(\cdot/(1+\gamma)\big)}{\gamma}. \tag{5.40}$$

Consequently, f^* is also essentially smooth and it follows from [8, Thm. 5.4] that f is Legendre. Moreover, since $\mathcal{X}$ is a Hilbert space, it is reflexive. We also derive from (5.40) that, since $\operatorname{Id} - \operatorname{Prox}^\varphi_{\gamma/(1+\gamma)}$ is (firmly) nonexpansive, ∇f^* is $1/\gamma$-Lipschitz and, thereby, maps bounded sets to bounded sets. It then follows from [8, Thm. 3.3] that f is supercoercive, and Proposition 4.1(v)(a) asserts that Condition 5.3(i) is satisfied. Finally, since ∇f is continuous, it will be weakly continuous when it is affine, i.e., when $\operatorname{Prox}^\varphi_{\gamma/(1+\gamma)}$ is. In turn, Example 4.6 implies that Condition 4.3 is satisfied. On the other hand, if S is a singleton, the claim follows from Example 4.5. ∎

If we let φ be the indicator function of a nonempty closed convex set in (5.36), then we obtain the Legendre function studied in [8, Example 7.2]. Specializing even further, we obtain the following examples.

Example 5.9 (distance). *In the previous example, set $\varphi = \iota_M$, where M is a closed affine subspace of $\mathcal{X}$, and let P_M be the metric projector onto M. Then Conditions* 4.3, 4.4, *and* 5.3 *are satisfied, $f = (1+\gamma)w - d_M^2/2$, and $D\colon (x,y) \mapsto \gamma w(x-y) + w(x - P_M y) - w(x - P_M x)$.*

EXAMPLE 5.10 (energy). *In the previous example, set $M = \{0\}$ and $\gamma = 1$. Then $f = \|\cdot\|^2/2$, $\nabla f = \mathrm{Id}$, $D\colon (x,y) \mapsto \|x-y\|^2/2$, and we recover the usual Fejér monotonicity framework.*

The next example shows that the function $f = \|\cdot\|^2/2$ can also be used outside Hilbert spaces.

EXAMPLE 5.11 (L_p spaces). *Let $(\Omega, \mathcal{F}, \mu)$ be a positive measure space and let $p \in [2, +\infty[$. Let $\mathcal{X} = L_p(\Omega, \mathcal{F}, \mu)$, equipped with its canonical norm, and set $f = \|\cdot\|^2/2$. Then Conditions 4.4 and 5.3 are satisfied. If S is a singleton, then Condition 4.3 is also satisfied.*

Proof. By [8, Example 6.5], f is Legendre and uniformly convex on closed balls. Hence Condition 4.4 holds by Example 4.10(i). Since f is supercoercive, Condition 5.3(i) follows from [8, Lemma 7.3(viii)]. We now establish Condition 5.3(ii). As $p \in [2, +\infty[$, [45, Corollary V.1.2] implies that $\rho_{\|\cdot\|}$, the modulus of smoothness of $\mathcal{X}$, is of power type 2. We thus obtain $\kappa \in]0, +\infty[$ so that (see [45, section IV.4])

$$(\forall t \in [0, +\infty[) \quad \rho_{\|\cdot\|}(t) \leq \kappa t^2. \tag{5.41}$$

Recall that $\nabla f = J$ and define $j(x) = J(x)/\|x\| = \nabla\|x\|$ for all nonzero $x \in \mathcal{X}$. Now (5.41) and [45, Lemma IV.5.1] yield

$$(\forall u \in S_{\mathcal{X}})(\forall v \in S_{\mathcal{X}}) \quad \|j(u) - j(v)\| \leq \kappa\|u - v\|. \tag{5.42}$$

Fix two nonzero points x and y in $\mathcal{X}$ and assume, without loss of generality, that $\|x\| \geq \|y\|$. Then, using the triangle inequality,

$$\left\|\frac{x}{\|x\|} - \frac{y}{\|y\|}\right\| = \left\|\left(\frac{x}{\|x\|} - \frac{y}{\|x\|}\right) + \frac{\|y\|\cdot y - \|x\|\cdot y}{\|x\|\cdot\|y\|}\right\| \leq \frac{2}{\|x\|}\|x - y\|. \tag{5.43}$$

Thus

$$\|j(x) - j(y)\| = \left\|j\left(\frac{x}{\|x\|}\right) - j\left(\frac{y}{\|y\|}\right)\right\| \leq \kappa \cdot \left\|\frac{x}{\|x\|} - \frac{y}{\|y\|}\right\| \leq \frac{2\kappa}{\|x\|}\|x - y\|, \tag{5.44}$$

where we have used the definition of j for the equality, (5.42) for the first inequality, and (5.43) for the second. Furthermore,

$$\begin{aligned} \|J(x) - J(y)\| &= \big\|\|x\|\cdot j(x) - \|y\|\cdot j(y)\big\| \\ &= \big\|\big(\|x\|\cdot j(x) - \|x\|\cdot j(y)\big) + \big(\|x\|\cdot j(y) - \|y\|\cdot j(y)\big)\big\| \\ &\leq \|x\|\cdot\|j(x) - j(y)\| + \|j(y)\|\cdot\big|\|x\| - \|y\|\big| \\ &\leq (2\kappa + 1)\cdot\|x - y\|, \end{aligned} \tag{5.45}$$

where the last inequality follows from (5.44) and the fact that $\|j(y)\| = 1$. Now (5.45) implies that $J = \nabla f$ is Lipschitz-continuous on $\operatorname{dom} f = \mathcal{X}$, with constant $2\kappa + 1$ (for $x = 0$ or $y = 0$, argue directly). We apply Proposition 5.5(ii) and conclude that Condition 5.3(ii) is satisfied. Finally, if S is a singleton, we employ Example 4.5. ∎

Guaranteeing Condition 4.3 requires some care.

Remark 5.12. As already discussed in Remark 5.6, Proposition 5.5(i) holds as soon as ∇f is Lipschitz on bounded sets. Thus, the assertions of Example 5.11 remain true for $f = \|\cdot\|^s/s$, where $s \in [2, +\infty[$. The case when $s = p$ is particularly interesting because then ∇f becomes J_φ, the duality mapping corresponding to the weight $\varphi\colon t \mapsto t^{p-1}$ (see [34]). If we specialize this further to the space $\ell_p(\mathbb{N})$, then

J_φ is known to be sequentially weakly continuous (see [34, Prop. II.4.14]) and thus Example 4.6 is applicable. To sum up,

let $\mathcal{X} = \ell_p(\mathbb{N})$ and $f = \|\cdot\|^p/p$, for $p \in [2, +\infty[$;
then Conditions 4.3, 4.4, and 5.3 are satisfied.

Additional examples can be generated in suitable product spaces such as $\ell_{p_1}(\mathbb{N}) \times \ell_{p_2}(\mathbb{N})$, equipped with the Euclidean product norm and with $\{p_1, p_2\} \subset [2, +\infty[$, or in certain spaces of power type 2. (See [45] for further information about such spaces.)

Example 5.13 (closed domain Bregman/Legendre functions). *Let $\mathcal{X} = \mathbb{R}^N$ and let f be a Bregman/Legendre function with closed domain. Then Conditions* 4.3*,* 4.4*, and* 5.3 *are satisfied.*

Proof. Example 4.9 implies that Condition 4.3 holds. Condition 4.4 follows from [7, Def. 5.2.BL3 and Thm. 3.9(iii)]. It remains to check items (i) and (ii) in Condition 5.3: since $D(z, \cdot)$ is coercive for every $z \in \operatorname{dom} f$ [7, Remark 5.3], (i) holds, whereas (ii) follows from Proposition 5.5(iv). ☐

The class of Bregman/Legendre functions (see Definition 4.8) is large enough to contain many functions important in convex optimization and it is related to the Bregman functions of [30, 33], which require closed domains. We refer the reader to [7] for further information. The following example gives conditions that are easy to verify in practice.

Example 5.14 (separable Bregman/Legendre functions). *Let $(\varphi_k)_{1 \leq k \leq N} \colon \mathbb{R} \to]-\infty, +\infty]$ be a family of Legendre functions such that $(\operatorname{dom} \varphi_k^*)_{1 \leq k \leq N}$ are open. Let $\mathcal{X} = \mathbb{R}^N$, and let $f \colon (\xi_k)_{1 \leq k \leq N} \mapsto \sum_{k=1}^N \varphi_k(\xi_k)$. Then Conditions* 4.3*,* 4.4*, and* 5.3 *are satisfied.*

Proof. By [7, Corollary 5.13], f is Bregman/Legendre. Mimicking the proof of the previous example, we note that it remains to check Condition 4.4. For every $k \in \{1, \ldots, m\}$, since $\varphi_k|_{\operatorname{int} \operatorname{dom} \varphi_k}$ is strictly convex by Legendreness and $\varphi_k|_{\operatorname{dom} \varphi_k}$ is continuous by (3.5), $\varphi_k|_{\operatorname{dom} \varphi_k}$ is strictly convex. Hence, it follows from Example 4.10(iii) that the Bregman distance D_k induced by φ_k on $\mathbb{R}$ satisfies Condition 4.4 and, in turn, so does $D \colon \big((\xi_k)_{1 \leq k \leq N}, (\chi_k)_{1 \leq k \leq N}\big) \mapsto \sum_{k=1}^N D_k(\xi_k, \chi_k)$. ☐

Unlike the previous examples, the following example does not require that $\mathcal{X}$ be finite-dimensional or that f have full domain.

Example 5.15. *Let $\mathcal{X}$ be the Hilbert space $\ell_2(\mathbb{N}) \times \mathbb{R}$ and define*

$$f \colon \mathcal{X} \to]-\infty, +\infty] : (x, \xi) \mapsto \begin{cases} \frac{1}{2}\|x\|^2 + \xi \ln(\xi) - \xi & \text{if } \xi > 0, \\ \frac{1}{2}\|x\|^2 & \text{if } \xi = 0, \\ +\infty & \text{if } \xi < 0. \end{cases} \tag{5.46}$$

Let $(\forall i \in I)$ $S_i = S = \ell_2(\mathbb{N}) \times [1, +\infty[$. Fix $(z, \zeta) \in S$, $\eta > 0$, and set $C = \operatorname{lev}_{\leq \eta} D\big((z, \zeta), \cdot\big)$. Then Conditions 4.3*,* 4.4*, and* 5.3 *are satisfied.*

Proof. Let $g = f(\cdot, 0)$ and $h = f(0, \cdot)$. Hence, $(\forall (x, \xi) \in \mathcal{X})$ $f(x, \xi) = g(x) + h(\xi)$. Note that g and h are Legendre, and so is f, with $\operatorname{dom} f = \ell_2(\mathbb{N}) \times [0, +\infty[$. Now, let D_g and D_h be the Bregman distances induced by g on $\ell_2(\mathbb{N})$ and h on $\mathbb{R}$, respectively. Take $(y, \chi) \in \mathcal{X}$ with $D\big((z, \zeta), (y, \chi)\big) = D_g(z, y) + D_h(\zeta, \chi) \leq \eta$. In particular, $D_g(z, y) \leq \eta$ and $D_h(\zeta, \chi) \leq \eta$. Since $D_g(z, \cdot)$ and $D_h(\zeta, \cdot)$ are coercive by Proposition 4.1(v)(a) and (b), C is bounded. Condition 4.3 is a consequence of Example 4.6. Since $D \colon (x, \xi) \mapsto D_g(x) + D_h(\xi)$, Condition 4.4 is immediate by Examples 5.10 and 5.13. Applying [7, Thm. 3.8.(i)] to h and $\zeta \in \operatorname{int} \operatorname{dom} h$, we obtain $\varepsilon \in]0, +\infty[$ such that $(\forall (y, \chi) \in C)$ $\chi \geq \varepsilon$. A straightforward computation shows that ∇f^* is strongly monotone with constant $\min\{1, \varepsilon\}$. Therefore, using Proposition 5.5(iv), Condition 5.3(ii) holds as well and the proof is complete. □

5.5. Applications. A broad class of problems in convex optimization and nonlinear analysis are captured by the mixed convex feasibility problem

$$\text{Find } x \in \overline{\operatorname{dom}} f \text{ such that } \begin{cases} (\forall i \in I^{(1)}) & g_i(x) \leq 0, \\ (\forall i \in I^{(2)}) & 0 \in A_i x, \\ (\forall i \in I^{(3)}) & \varphi_i(x) = \inf \varphi_i(\mathcal{X}), \\ (\forall i \in I^{(4)}) & T_i x = x, \end{cases} \tag{5.47}$$

where $(g_i)_{i \in I^{(1)}}$ and $(\varphi_i)_{i \in I^{(3)}}$ are families of proper lower semicontinuous convex functions from $\mathcal{X}$ into $]-\infty, +\infty]$, $(A_i)_{i \in I^{(2)}}$ is a family of maximal monotone operators from $\mathcal{X}$ into $2^{\mathcal{X}^*}$, and $(T_i)_{i \in I^{(4)}}$ is a family of D-firm operators from $\mathcal{X}$ into $\mathcal{X}$. Here, $I^{(1)}$, $I^{(2)}$, $I^{(3)}$, and $I^{(4)}$ are pairwise disjoint, possibly empty, countable index sets such that $I = \bigcup_{k=1}^4 I^{(k)} \neq \varnothing$. Now let us define

$$(\forall i \in I) \;\; S_i = \begin{cases} \operatorname{lev}_{\leq 0} g_i & \text{if } i \in I^{(1)}, \\ A_i^{-1} 0 & \text{if } i \in I^{(2)}, \\ \operatorname{Argmin} \varphi_i & \text{if } i \in I^{(3)}, \\ \overline{\operatorname{Fix}}\, T_i & \text{if } i \in I^{(4)}. \end{cases} \tag{5.48}$$

Throughout this section, the following set of assumptions will be made.

CONDITION 5.16.

(i) *Conditions* 4.3, 4.4, 5.3, *and* 5.4 *are satisfied.*

(ii) *For every* $i \in I^{(1)}$, $\partial g_i(C)$ *is bounded and* $\operatorname{dom} f \subset \operatorname{dom} g_i$.

(iii) *For every* $i \in I^{(2)}$, *one of the following conditions holds:*

  (a) $\operatorname{dom} A_i \subset \operatorname{int} \operatorname{dom} f$,

  (b) A_i *is* 3^**-monotone.*

(iv) *For every* $i \in I^{(4)}$, $\operatorname{dom} T_i = \operatorname{int} \operatorname{dom} f$ *and* $T_i - \operatorname{Id}$ *is demiclosed at* 0 *in the sense that for every sequence* $(y_n)_{n \in \mathbb{N}}$ *in* $\operatorname{dom} T_i$

$$\begin{cases} y_n \rightharpoonup y, \\ (\forall n \in \mathbb{N}) \;\; u_n \in T_i y_n, \\ u_n - y_n \to 0 \end{cases} \Rightarrow \;\; y \in \overline{\operatorname{Fix}}\, T_i. \tag{5.49}$$

Let us observe that the sets $(S_i)_{i \in I}$ are closed and convex. For $i \in I^{(1)} \cup I^{(2)} \cup I^{(3)}$, this follows from well-known facts; for $i \in I^{(4)}$, this follows from Condition 5.16(iv), Propositions 3.5(ii), the essential strict convexity of f, and Proposition 3.3(v). Accordingly, (5.47) is a special case of the convex feasibility problem (5.2) and it can therefore be solved by Algorithm 5.1.

ALGORITHM 5.17 (specific implementation of Algorithm 5.1). *Fix* $(\varepsilon_i)_{i \in I^{(2)}}$ *and* $(\varepsilon_i)_{i \in I^{(3)}}$ *in* $]0, +\infty[$. *Implement Algorithm* 5.1② *by choosing for every* $i \in I_n$

$$T_{i,n} = \begin{cases} Q_{g_i} & \textit{if } i \in I^{(1)} \; (\textit{see Definition } 3.37), \\ R_{\gamma_{i,n} A_i}, \; \textit{where } \gamma_{i,n} \in [\varepsilon_i, +\infty[& \textit{if } i \in I^{(2)} \; (\textit{see Definition } 3.7), \\ \operatorname{prox}_{\gamma_{i,n}}^{\varphi_i}, \; \textit{where } \gamma_{i,n} \in [\varepsilon_i, +\infty[& \textit{if } i \in I^{(3)} \; (\textit{see Definition } 3.16), \\ T_i & \textit{if } i \in I^{(4)} \; (\textit{see Definition } 3.4). \end{cases} \tag{5.50}$$

Thanks to Condition 5.16, (5.50) meets the requirements of Algorithm 5.1② since in each case we have the following:

- $T_{i,n} \in \mathfrak{B}$. This follows from Proposition 3.38(ii) if $i \in I^{(1)}$, from Corollary 3.14(ii) and (iii) if $i \in I^{(2)}$ (since $A_i^{-1}0 \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$, $\operatorname{dom} A_i \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$), from Corollary 3.25(i) if $i \in I^{(3)}$ (since φ_i is proper and $\operatorname{Argmin}\varphi_i \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$, φ_i is bounded below and $\operatorname{dom}\varphi_i \cap \operatorname{int}\operatorname{dom} f \neq \varnothing$), and from Proposition 3.5(ii) if $i \in I^{(4)}$.
- $S_i \cap \operatorname{int}\operatorname{dom} f \subset \operatorname{Fix} T_{i,n}$. (See Proposition 3.38(i), Proposition 3.8(iii), Proposition 3.22(ii)(b), and Proposition 3.3(iv) and (vii), respectively.)

THEOREM 5.18. *Suppose that Condition* 5.16 *is in force and let* $(x_n)_{n\in\mathbb{N}}$ *be an arbitrary orbit of Algorithm* 5.17. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point* $x \in S$. *The convergence is strong if* $x \in \operatorname{int}\operatorname{dom} f$ *and any of the following assumptions is added:*

(i) *For some* $i \in I^{(1)}$ *and some* $\eta \in \,]0,+\infty[$, $C \cap \operatorname{lev}_{\leq\eta} g_i$ *is relatively compact.*
(ii) *For some* $i \in I^{(2)}$, $C \cap \operatorname{dom} A_i$ *is relatively compact.*
(iii) *For some* $i \in I^{(3)}$, $C \cap \operatorname{dom} \partial\varphi_i$ *is relatively compact.*
(iv) *For some* $i \in I^{(4)}$, T_i *is demicompact at* 0 *in the sense that for every sequence* $(y_n)_{n\in\mathbb{N}}$ *in* $\operatorname{dom} T_i$

$$\begin{cases} (y_n)_{n\in\mathbb{N}} \text{ bounded,} \\ (\forall n \in \mathbb{N})\ \ u_n \in T_i y_n, \\ u_n - y_n \to 0 \end{cases} \Rightarrow \quad \mathfrak{S}(y_n)_{n\in\mathbb{N}} \neq \varnothing. \tag{5.51}$$

Proof. As seen above, (5.47) is a special case of (5.2), whereas Algorithm 5.17 is a special case of Algorithm 5.1. Invoking Theorem 5.7, we shall prove that Algorithm 5.17 is focusing to establish the weak convergence claim and then that it is demicompactly regular to establish the strong convergence claim. It is recalled that Theorem 5.7 asserts that $(x_n)_{n\in\mathbb{N}}$ and $\big((u_{i,n})_{i\in I_n}\big)_{n\in\mathbb{N}}$ lie in the bounded set C.

To show that Algorithm 5.17 is focusing, let us fix $i \in I$ and a suborbit $(x_{k_n})_{n\in\mathbb{N}}$ such that $i \in \bigcap_{n\in\mathbb{N}} I_{k_n}$, $x_{k_n} \rightharpoonup x$, and $u_{i,k_n} - x_{k_n} \to 0$. According to (5.10), we must show $x \in S_i$. Four cases will be considered:

(1) $i \in I^{(1)}$. We must show $g_i(x) \leq 0$. In view of (5.50), for every $n \in \mathbb{N}$, u_{i,k_n} is the D-projection of x_n onto $G_i(x_{k_n}, x_n^*) = \big\{y \in \mathcal{X} \mid \langle x_{k_n} - y, x_n^*\rangle \geq g_i(x_{k_n})\big\}$ for some $x_n^* \in \partial g_i(x_{k_n})$. Since $u_{i,k_n} \in G_i(x_{k_n}, x_n^*)$, we have

$$\|u_{i,k_n} - x_{k_n}\| \geq d_{G_i(x_{k_n},x_n^*)}(x_{k_n}) = \begin{cases} g_i^+(x_{k_n})/\|x_n^*\| & \text{if } x_n^* \neq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{5.52}$$

where $g_i^+ = \max\{0, g_i\}$ and the last equality follows from [80, Lemma I.1.2]. Since $(x_{k_n})_{n\in\mathbb{N}}$ lies in C, $(x_n^*)_{n\in\mathbb{N}}$ is bounded by Condition 5.16(ii). Therefore, $u_{i,k_n} - x_{k_n} \to 0$ implies $g_i^+(x_{k_n}) \to 0$. However, as g_i^+ is convex and lower semicontinuous, it is weak lower semicontinuous and thus $g_i^+(x) \leq \underline{\lim}\, g_i^+(x_{k_n}) = 0$. We conclude $g_i(x) \leq 0$.

(2) $i \in I^{(2)}$. We must show $(x,0) \in \operatorname{gr} A_i$. For every $n \in \mathbb{N}$, (5.50) yields $u_{i,k_n} \in (\nabla f + \gamma_{i,k_n} A_i)^{-1}\big(\nabla f(x_{k_n})\big)$ and we define

$$u_n^* = \frac{\nabla f(x_{k_n}) - \nabla f(u_{i,k_n})}{\gamma_{i,k_n}}. \tag{5.53}$$

Therefore $\big((u_{i,k_n}, u_n^*)\big)_{n\in\mathbb{N}}$ lies in $\operatorname{gr} A_i$ and $u_{i,k_n} - x_{k_n} \to 0 \Rightarrow u_{i,k_n} \rightharpoonup x$. If for all n sufficiently large we have $x_{k_n} = u_{i,k_n}$, then by Proposition 3.8(iii) the tail of $(x_{k_n})_{n\in\mathbb{N}}$ is in the weakly closed set $A_i^{-1}0$ and therefore $(x,0)\in\operatorname{gr}A_i$. Otherwise, we can extract a subsequence $(x_{k_{l_n}})_{n\in\mathbb{N}}$ such that, for all $n\in\mathbb{N}$, $x_{k_{l_n}}\neq u_{i,k_{l_n}}$. Since, on the one hand, $(x_{k_{l_n}})_{n\in\mathbb{N}}$ and $(u_{i,k_{l_n}})_{n\in\mathbb{N}}$ lie in C and, on the other hand,

$$(5.54)\quad (\forall n\in\mathbb{N})\ \ \|u_{i,k_{l_n}}-x_{k_{l_n}}\|\geq\frac{\langle u_{i,k_{l_n}}-x_{k_{l_n}},\nabla f(u_{i,k_{l_n}})-\nabla f(x_{k_{l_n}})\rangle}{\|\nabla f(u_{i,k_{l_n}})-\nabla f(x_{k_{l_n}})\|},$$

it follows from Condition 5.3(ii), (5.53), and the inequality $\inf_{n\in\mathbb{N}}\gamma_{i,k_{l_n}}\geq\varepsilon_i$ that $u_{i,k_{l_n}}-x_{k_{l_n}}\to 0\Rightarrow\nabla f(u_{i,k_{l_n}})-\nabla f(x_{k_{l_n}})\to 0\Rightarrow u^*_{l_n}\to 0$. Finally, since A_i is maximal monotone, $\operatorname{gr}A_i$ is sequentially closed in the weak $\times$ strong topology of $\mathcal{X}\times\mathcal{X}^*$ and we conclude that $(x,0)\in\operatorname{gr}A_i$, as required.

(3) $i\in I^{(3)}$. We must show $\varphi_i(x)=\inf\varphi_i(\mathcal{X})$, i.e., $(x,0)\in\operatorname{gr}\partial\varphi_i$. Since φ is a proper lower semicontinuous convex function, $A_i=\partial\varphi_i$ is maximal monotone [79, section 29] and 3*-monotone by Lemma 3.10(iv), and, in view of Propositions 3.22(ii)(a) and 3.23(v)(b), the claim follows from case (2).

(4) $i\in I^{(4)}$. We must show $x\in\overline{\operatorname{Fix}}\,T_i$. This follows at once from (5.49).

It remains to show that in each instance (i)–(iv), i is an index of demicompact regularity. Henceforth, $(x_{k_n})_{n\in\mathbb{N}}$ is a suborbit such that $i\in\bigcap_{n\in\mathbb{N}}I_{k_n}$ and $u_{i,k_n}-x_{k_n}\to 0$. By (5.11), we must show $\mathfrak{S}(x_{k_n})_{n\in\mathbb{N}}\neq\varnothing$. (i) Arguing as in case (1), we obtain $\overline{\lim}\,g_i(x_{k_n})\leq 0$. Therefore, the tail of $(x_{k_n})_{n\in\mathbb{N}}$ lies in the compact set $\overline{C\cap\operatorname{lev}_{\leq\eta}g_i}$, whence $\mathfrak{S}(x_{k_n})_{n\in\mathbb{N}}\neq\varnothing$. (ii) It follows from (3.16) that for every $n\in\mathbb{N}$

$$\begin{cases}u_{i,k_n}\in C\subset\operatorname{int}\operatorname{dom}f,\\ u_{i,k_n}\in\operatorname{ran}(\nabla f+\gamma_{i,k_n}A_i)^{-1}\circ\nabla f\subset\operatorname{dom}\nabla f\cap\operatorname{dom}A_i=\operatorname{int}\operatorname{dom}f\cap\operatorname{dom}\,A_i.\end{cases}$$

Therefore, $(u_{i,k_n})_{n\in\mathbb{N}}$ lies in the compact set $\overline{C\cap\operatorname{dom}\,A_i}$, whence $\mathfrak{S}(u_{i,k_n})_{n\in\mathbb{N}}\neq\varnothing$. Since $u_{i,k_n}-x_{k_n}\to 0$, we conclude $\mathfrak{S}(x_{k_n})_{n\in\mathbb{N}}\neq\varnothing$. (iii) As in case (3), this is a special case of (ii). (iv) This is clear from (5.51). ◻

Theorem 5.18 produces convergence results for various new block-iterative parallel schemes for solving problems, including solving convex inequalities ($I^{(2)}=I^{(3)}=I^{(4)}=\varnothing$), finding common zeros ($I^{(1)}=I^{(3)}=I^{(4)}=\varnothing$), solving systems of variational inequalities ($I^{(1)}=I^{(2)}=I^{(4)}=\varnothing$), finding common fixed points ($I^{(1)}=I^{(2)}=I^{(3)}=\varnothing$), and combinations of these. Note that D-projection methods are also captured by Theorem 5.18 since, in view of Proposition 3.32(ii)(c), one can take, for instance, T_i to be the D-projector onto S_i if $i\in I^{(4)}$ in (5.50).

Naturally, our framework also encompasses relaxed sequential algorithms, which are obtained by taking $(I_n)_{n\in\mathbb{N}}$ to be a sequence of singletons, as in the following example.

EXAMPLE 5.19. *Suppose $\mathcal{X}=\mathbb{R}^N$, $(S_i)_{1\leq i\leq m}$ is a (finite) family of half-spaces with D-projectors $(P_i)_{1\leq i\leq m}$, and, for every $n\in\mathbb{N}$, $I_n=\{n\ (\mathrm{mod}\ m)+1\}$ and $T_{i,n}=P_i$. Then Algorithm* 5.1 *reduces to the relaxed D-projection method of* [44].

In the case of unrelaxed sequential algorithms, our working assumptions can be loosened. This is discussed next.

5.6. Unrelaxed sequential algorithms. Algorithm 5.1 can be specialized to an unrelaxed sequential algorithm for solving the convex feasibility problem (5.2). Indeed, suppose that at each iteration n only one index, say i(n), is retained and $\lambda_n=1$. Then Algorithm 5.1⑧ becomes

$$(5.55)\qquad H_n=\big\{y\in\mathcal{X}\mid\langle y-u_n,\nabla f(x_n)-\nabla f(u_n)\rangle\leq 0\big\},$$

where $u_n \in T_n x_n$ for some $T_n \in \mathfrak{B}$ such that $S_{\mathrm{i}(n)} \cap \operatorname{int} \operatorname{dom} f \subset \operatorname{Fix} T_n$. Consequently, since by Corollary 3.35(ii) $P_{H_n} x_n = u_n$, Algorithm 5.1 can be rewritten as follows.

ALGORITHM 5.20. *Starting with* $x_0 \in \operatorname{int} \operatorname{dom} f$, *take at every iteration* n

① *an index* $\mathrm{i}(n) \in I$,

② *an operator* T_n *in* $\mathfrak{B}$ *such that* $S_{\mathrm{i}(n)} \cap \operatorname{int} \operatorname{dom} f \subset \operatorname{Fix} T_n$.

Then select $x_{n+1} \in T_n x_n$.

In this context, Definition 5.2 takes the following form.

DEFINITION 5.21. *Algorithm* 5.20 *is*

- focusing *if for every bounded suborbit* $(x_{k_n})_{n\in\mathbb{N}}$ *it generates and every index* $i \in I$,

$$\begin{cases} (\forall n \in \mathbb{N}) \ \ i = \mathrm{i}(k_n), \\ x_{k_n} \rightharpoonup x, \\ x_{k_n+1} - x_{k_n} \to 0 \end{cases} \quad \Rightarrow \quad x \in S_i; \tag{5.56}$$

- demicompactly regular *if there exists* $i \in I$, *called an* index of demicompact regularity, *such that for every bounded suborbit* $(x_{k_n})_{n\in\mathbb{N}}$ *it generates,*

$$\begin{cases} (\forall n \in \mathbb{N}) \ \ i = \mathrm{i}(k_n), \\ x_{k_n+1} - x_{k_n} \to 0 \end{cases} \quad \Rightarrow \quad \mathfrak{S}(x_{k_n})_{n\in\mathbb{N}} \neq \varnothing. \tag{5.57}$$

Removing item (ii) from Condition 5.3 yields the following set of assumptions for the unrelaxed sequential case.

CONDITION 5.22. *For some* $z \in \operatorname{dom} f \cap \bigcap_{i\in I} S_i$, $C = \operatorname{lev}_{\leq D(z,x_0)} D(z,\cdot)$ *is bounded.*

CONDITION 5.23. $(\forall i \in I)(\exists\, M_i \in \mathbb{N}\smallsetminus\{0\})(\forall n \in \mathbb{N}) \ \ i \in \big\{\mathrm{i}(n), \ldots, \mathrm{i}(n{+}M_i{-}1)\big\}$.

We now show that Algorithm 5.20 converges under this reduced set of assumptions.

THEOREM 5.24. *Suppose that Conditions* 4.3, 4.4, 5.22, *and* 5.23 *are satisfied and that Algorithm* 5.20 *is focusing. Then the following statements hold true for every orbit* $(x_n)_{n\in\mathbb{N}}$ *generated by Algorithm* 5.20:

(i) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point* $x \in S$.

(ii) *If the weak limit* x *from* (i) *belongs to* $\operatorname{int} \operatorname{dom} f$ *and the algorithm is demicompactly regular, then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly.*

Proof. In the proof of Theorem 5.7, note that Condition 5.3(ii) is used only to obtain (5.34), i.e., in the present context, $x_{p_n+1} - x_{p_n} \to 0$. However, this property follows directly from(5.33). □

As an example, we revisit Bregman's original cyclic projection method (1.2). (See [1, Thm. 3.1] for a special case.)

COROLLARY 5.25. *Suppose that Conditions* 4.3 *and* 4.4 *are satisfied, that* $I = \{1, \ldots, m\}$, *and that* $C = \operatorname{lev}_{\leq D(z,x_0)} D(z,\cdot)$ *is bounded for some* $z \in \operatorname{dom} f \cap \bigcap_{1\leq i\leq m} S_i$. *Let* $(P_i)_{1\leq i\leq m}$ *be the D-projectors of* $(S_i)_{1\leq i\leq m}$. *Then the following statements hold true for every orbit* $(x_n)_{n\in\mathbb{N}}$ *generated by* (1.2):

(i) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point* $x \in \overline{\operatorname{dom}}\, f \cap \bigcap_{1\leq i\leq m} S_i$.

(ii) *If the weak limit* x *from* (i) *belongs to* $\operatorname{int} \operatorname{dom} f$ *and* $C \cap S_i$ *is relatively compact (e.g.,* S_i *is boundedly compact) for some* $i \in \{1, \ldots, m\}$, *then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly.*

Proof. In view of Corollary 3.35(i), (1.2) is a special realization of Algorithm 5.20 with $(\forall n \in \mathbb{N})$ $T_n = P_{n\,(\mathrm{mod}\ m)+1}$ (single-valued) and $\lambda_n = 1$. In addition, the index control rule $i\colon n \mapsto n \,(\mathrm{mod}\, m) + 1$ complies with Condition 5.23. On the other hand, algorithm (1.2) is focusing, as a direct consequence of the weak closedness of the sets $(S_j)_{1 \le j \le m}$. Finally, i is an index of demicompact regularity since $(x_{nm+i})_{n\in\mathbb{N}}$ lies in $C \cap S_i$. The announced results therefore follow from Theorem 5.24. □

Remark 5.26. Throughout section 5, Legendreness has been imposed on f. This property has been shown to provide a rich and convenient framework in which our results could be derived in a unified manner. Further results can nonetheless be obtained from the analysis of sections 3 and 4 for functions which are not Legendre at the expense of more technical assumptions.

Acknowledgments. We wish to thank Dan Butnariu and Yair Censor for sending us [22, 25, 26, 29], Constantin Zălinescu for sending us [85], and especially Jon Vanderwerff for his help in the derivation of Example 5.11. Two anonymous referees made several helpful comments and suggestions, which led to improvements over the originally submitted version.

REFERENCES

[1] Y. Alber and D. Butnariu, *Convergence of Bregman projection methods for solving consistent convex feasibility problems in reflexive Banach spaces*, J. Optim. Theory Appl., 92 (1997), pp. 33–61.

[2] H. Attouch and H. Brézis, *Duality for the sum of convex functions in general Banach spaces*, in Aspects of Mathematics and Its Applications, North–Holland, Amsterdam, 1986, pp. 125–133.

[3] J.-P. Aubin, *Viability Theory*, Birkhäuser, Boston, 1991.

[4] J.-B. Baillon and G. Haddad, *Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones*, Israel J. Math., 26 (1977), pp. 137–150.

[5] H. H. Bauschke, *Projection Algorithms and Monotone Operators*, Ph.D. thesis, Simon Fraser University, Canada, 1996; available online as preprint 96:080 from http://www.cecm.sfu.ca/preprints.

[6] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., 38 (1996), pp. 367–426.

[7] H. H. Bauschke and J. M. Borwein, *Legendre functions and the method of random Bregman projections*, J. Convex Anal., 4 (1997), pp. 27–67.

[8] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, *Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces*, Commun. Contemp. Math., 3 (2001), pp. 615–647.

[9] H. H. Bauschke and P. L. Combettes, *A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces*, Math. Oper. Res., 26 (2001), pp. 248–264.

[10] H. H. Bauschke and P. L. Combettes, *Construction of best Bregman approximations in reflexive Banach spaces*, Proc. Amer. Math. Soc., published electronically April 24, 2003.

[11] H. H. Bauschke and P. L. Combettes, *Iterating Bregman retractions*, SIAM J. Optim., 13 (2003), pp. 1159–1173.

[12] H. H. Bauschke, P. L. Combettes, and D. R. Luke, *Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization*, J. Opt. Soc. Amer. A, 19 (2002), pp. 1334–1345.

[13] J. M. Borwein and J. D. Vanderwerff, *Convex functions of Legendre type in general Banach spaces*, J. Convex Anal., 8 (2001), pp. 569–581.

[14] L. M. Bregman, *The method of successive projection for finding a common point of convex sets*, Soviet Math. Dokl., 6 (1965), pp. 688–692.

[15] 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*, USSR Comput. Math. Math. Phys., 7 (1967), pp. 200–217.

[16] H. Brézis and A. Haraux, *Image d'une somme d'opérateurs monotones et applications*, Israel J. Math., 23 (1976), pp. 165–186.

[17] H. Brézis and P. L. Lions, *Produits infinis de résolvantes*, Israel J. Math., 29 (1978), pp. 329–345.

[18] M. Brohe and P. Tossings, *Perturbed proximal point algorithm with nonquadratic kernel*, Serdica Math. J., 26 (2000), pp. 177–206.
[19] R. E. Bruck and S. Reich, *Nonexpansive projections and resolvents of accretive operators in Banach spaces*, Houston J. Math., 3 (1977), pp. 459–470.
[20] R. S. Burachik and A. N. Iusem, *A generalized proximal point algorithm for the variational inequality problem in a Hilbert space*, SIAM J. Optim., 8 (1998), pp. 197–216.
[21] R. S. Burachik and S. Scheimberg, *A proximal point method for the variational inequality problem in Banach spaces*, SIAM J. Control Optim., 39 (2001), pp. 1633–1649.
[22] D. Butnariu, C. Byrne, and Y. Censor, *Redundant axioms in the definition of Bregman functions,* J. Convex Anal., to appear in (2003).
[23] D. Butnariu and A. N. Iusem, *On a proximal point method for convex optimization in Banach spaces,* Numer. Funct. Anal. Optim., 18 (1997), pp. 723–744.
[24] D. Butnariu and A. N. Iusem, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization,* Kluwer Academic Publishers, Boston, MA, 2000.
[25] D. Butnariu, A. N. Iusem, and C. Zălinescu, *On uniform convexity, total convexity and convergence of the proximal point and outer Bregman projection algorithms in Banach spaces*, J. Convex Anal., to appear in (2003).
[26] D. Butnariu, S. Reich, and A. J. Zaslavski, *Asymptotic behavior of relatively nonexpansive operators in Banach spaces*, J. Appl. Anal., 7 (2001), pp. 151–174.
[27] D. Butnariu and E. Resmerita, *The outer Bregman projection method for stochastic feasibility problems in Banach spaces*, in Inherently Parallel Algorithms for Feasibility and Optimization, D. Butnariu, Y. Censor, and S. Reich, eds., Elsevier, New York, 2001, pp. 69–86.
[28] J. A. Cadzow, *Signal enhancement – A composite property mapping algorithm*, IEEE Trans. Acoust. Speech Signal Process., 36 (1988), pp. 49–62.
[29] Y. Censor and G. T. Herman, *Block-iterative algorithms with underrelaxed Bregman projections*, SIAM J. Optim., 13 (2002), pp. 283–297.
[30] Y. Censor and A. Lent, *An iterative row-action method for interval convex programming*, J. Optim. Theory Appl., 34 (1981), pp. 321–353.
[31] Y. Censor and S. Reich, *Iterations of paracontractions and firmly nonexpansive operators with applications to feasibility and optimization*, Optimization, 37 (1996), pp. 323–339.
[32] Y. Censor and S. A. Zenios, *Proximal minimization algorithm with D-functions,* J. Optim. Theory Appl., 73 (1992), pp. 451–464.
[33] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications,* Oxford University Press, New York, 1997.
[34] I. Ciorănescu, *Geometry of Banach Spaces, Duality Mappings, and Nonlinear Problems,* Kluwer Academic Publishers, Boston, MA, 1990.
[35] P. L. Combettes, *Construction d'un point fixe commun à une famille de contractions fermes*, C. R. Acad. Sci. Paris Sér. I Math., 320 (1995), pp. 1385–1390.
[36] P. L. Combettes, *The convex feasibility problem in image recovery*, in Advances in Imaging and Electron Physics, Vol. 95, P. Hawkes, ed., Academic Press, New York, 1996, pp. 155–270.
[37] P. L. Combettes, *Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections*, IEEE Trans. Image Process., 6 (1997), pp. 493–506.
[38] P. L. Combettes, *Hilbertian convex feasibility problem: Convergence of projection methods*, Appl. Math. Optim., 35 (1997), pp. 311–330.
[39] P. L. Combettes, *Strong convergence of block-iterative outer approximation methods for convex optimization*, SIAM J. Control Optim., 38 (2000), pp. 538–565.
[40] P. L. Combettes, *Fejér monotonicity in convex optimization*, in Encyclopedia of Optimization, Vol. 2, C. A. Floudas and P. M. Pardalos, eds., Kluwer Academic Publishers, Boston, MA, 2001, pp. 106–114.
[41] P. L. Combettes, *Quasi-Fejérian analysis of some optimization algorithms*, in Inherently Parallel Algorithms for Feasibility and Optimization, D. Butnariu, Y. Censor, and S. Reich, eds., Elsevier, New York, 2001, pp. 115–152.
[42] P. L. Combettes and T. Pennanen, *Generalized Mann iterates for constructing fixed points in Hilbert spaces*, J. Math. Anal. Appl., 275 (2002), pp. 521–536.
[43] P. L. Combettes and H. J. Trussell, *Method of successive projections for finding a common point of sets in metric spaces*, J. Optim. Theory Appl., 67 (1990), pp. 487–507.
[44] A. R. de Pierro and A. N. Iusem, *A relaxed version of Bregman's method for convex programming*, J. Optim. Theory Appl., 51 (1986), pp. 421–440.
[45] R. Deville, G. Godefroy, and V. Zizler, *Smoothness and Renormings in Banach Spaces*, John Wiley, New York, 1993.

[46] J. Eckstein, *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming*, Math. Oper. Res., 18 (1993), pp. 202–226.

[47] J. Eckstein, *Approximate iterations in Bregman-function-based proximal algorithms*, Math. Program., 83 (1998), pp. 113–123.

[48] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems,* Classics Appl. Math. 28, SIAM, Philadelphia, PA, 1999.

[49] I. I. Eremin and V. D. Mazurov, *Nonstationary Processes of Mathematical Programming,* Nauka, Moscow, 1979.

[50] U. M. García-Palomares and F. J. González-Castaño, *Incomplete projection algorithms for solving the convex feasibility problem,* Numer. Algorithms, 18 (1998), pp. 177–193.

[51] K. Goebel and S. Reich, *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings,* Marcel Dekker, New York, 1984.

[52] K. M. Grigoriadis and R. E. Skelton, *Low-order control design for LMI problems using alternating projection methods*, Automatica J. IFAC, 32 (1996), pp. 1117–1125.

[53] L. G. Gubin, B. T. Polyak, and E. V. Raik, *The method of projections for finding the common point of convex sets*, USSR Comput. Math. Math. Phys., 7 (1967), pp. 1–24.

[54] C. D. Ha, *A generalization of the proximal point algorithm*, SIAM J. Control Optim., 28 (1990), pp. 503–512.

[55] A. Iusem and R. G. Otero, *Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces*, Numer. Funct. Anal. Optim., 22 (2001), pp. 609–640.

[56] A. Iusem and R. G. Otero, *Erratum: "Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces"*, Numer. Funct. Anal. Optim., 23 (2002), pp. 227–228.

[57] G. Kassay, *The proximal points algorithm for reflexive Banach spaces,* Studia Univ. Babeş-Bolyai Math., 30 (1985), pp. 9–17.

[58] K. C. Kiwiel, *Proximal minimization methods with generalized Bregman functions*, SIAM J. Control Optim., 35 (1997), pp. 1142–1168.

[59] K. C. Kiwiel, *Generalized Bregman projections in convex feasibility problems*, J. Optim. Theory Appl., 96 (1998), pp. 139–157.

[60] K. C. Kiwiel and B. Łopuch, *Surrogate projection methods for finding fixed points of firmly nonexpansive mappings*, SIAM J. Optim., 7 (1997), pp. 1084–1102.

[61] B. Lemaire, *On the convergence of some iterative methods for convex minimization*, in Recent Developments in Optimization, Lect. Notes Econ. Math. Syst. 429, Springer-Verlag, Berlin, 1995, pp. 252–268.

[62] A. Levi and H. Stark, *Image restoration by the method of generalized projections with application to restoration from magnitude*, J. Opt. Soc. Amer. A, 1 (1984), pp. 932–943.

[63] Y. I. Merzlyakov, *On a relaxation method of solving systems of linear inequalities*, USSR Comput. Math. Math. Phys., 2 (1963), pp. 504–510.

[64] J.-J. Moreau, *Fonctions convexes duales et points proximaux dans un espace hilbertien*, C. R. Acad. Sci. Paris Sér. A Math., 255 (1962), pp. 2897–2899.

[65] J.-J. Moreau, *Propriétés des applications prox*, C. R. Acad. Sci. Paris Sér. A Math., 256 (1963), pp. 1069–1071.

[66] J.-J. Moreau, *Sur la fonction polaire d'une fonction semi-continue supérieurement*, C. R. Acad. Sci. Paris Sér. A Math., 258 (1964), pp. 1128–1130.

[67] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.

[68] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc., 73 (1967), pp. 591–597.

[69] N. Ottavy, *Strong convergence of projection-like methods in Hilbert spaces*, J. Optim. Theory Appl., 56 (1988), pp. 433–461.

[70] W. V. Petryshyn, *Construction of fixed points of demicompact mappings in Hilbert space*, J. Math. Anal. Appl., 14 (1966), pp. 276–284.

[71] R. R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, 2nd ed., Lecture Notes in Math. 1364, Springer-Verlag, New York, 1993.

[72] G. Pierra, *Decomposition through formalization in a product space*, Math. Program., 28 (1984), pp. 96–115.

[73] E. Raik, *Fejér type methods in Hilbert space*, Esti NSV Tead. Akad. Toimetised Füüs.-Mat., 16 (1967), pp. 286–293.

[74] S. Reich, *The range of sums of accretive and monotone operators*, J. Math. Anal. Appl., 68 (1979), pp. 310–317.

[75] S. Reich, *A weak convergence theorem for the alternating method with Bregman distances*, in Theory and Applications of Nonlinear Operators of Accretive and Monotone Type, A. G.

Kartsatos, ed., Marcel Dekker, New York, 1996, pp. 313–318.
[76] R. T. Rockafellar, *Level sets and continuity of conjugate convex functions*, Trans. Amer. Math. Soc., 123 (1966), pp. 46–63.
[77] R. T. Rockafellar, *Convex Analysis,* Princeton University Press, Princeton, NJ, 1970.
[78] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.
[79] S. Simons, *Minimax and Monotonicity,* Lecture Notes in Math. 1693, Springer-Verlag, New York, 1998.
[80] I. Singer, *Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces,* Springer-Verlag, New York, 1970.
[81] I. Singer, *The Theory of Best Approximation and Functional Analysis,* CBMS-NSF Regional Conf. Ser. in Appl. Math. 13, SIAM, Philadelphia, PA, 1974.
[82] M. V. Solodov and B. F. Svaiter, *An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions,* Math. Oper. Res., 25 (2000), pp. 214–230.
[83] M. Teboulle, *Entropic proximal mappings with applications to nonlinear programming*, Math. Oper. Res., 17 (1992), pp. 670–690.
[84] C. Zălinescu, *On uniformly convex functions*, J. Math. Anal. Appl., 95 (1983), pp. 344–374.
[85] C. Zălinescu, *Convex Analysis in General Vector Spaces,* World Scientific Publishing, River Edge, NJ, 2002.
[86] E. Zeidler, *Nonlinear Functional Analysis and Its Applications.* II/*B. Nonlinear Monotone Operators,* Springer-Verlag, New York, 1990.