Journal of Optimization Theory and Applications
https://doi.org/10.1007/s10957-024-02383-9

A Mirror Inertial Forward–Reflected–Backward Splitting: Convergence Analysis Beyond Convexity and Lipschitz Smoothness

Ziyuan Wang[1] · Andreas Themelis[2] · Hongjia Ou[2] · Xianfu Wang[1]

Received: 26 May 2023 / Accepted: 6 January 2024

Abstract
This work investigates a Bregman and inertial extension of the forward–reflected–backward algorithm (Malitsky and Tam in SIAM J Optim 30:1451–1472, 2020) applied to structured nonconvex minimization problems under relative smoothness. To this end, the proposed algorithm hinges on two key features: taking inertial steps in the *dual* space, and allowing for *possibly negative* inertial values. The interpretation of relative smoothness as a two-sided weak convexity condition proves beneficial in providing tighter stepsize ranges. Our analysis begins with studying an envelope function associated with the algorithm that takes inertial terms into account through a novel product space formulation. Such construction substantially differs from similar objects in the literature and could offer new insights for extensions of splitting algorithms. Global convergence and rates are obtained by appealing to the Kurdyka–Łojasiewicz property.

Communicated by Radu Ioan Boţ.

This work was supported by the NSERC Discovery Grants and JSPS KAKENHI grant number JP21K17710.

✉ Xianfu Wang
shawn.wang@ubc.ca

Ziyuan Wang
ziyuan.wang@alumni.ubc.ca

Andreas Themelis
andreas.themelis@ees.kyushu-u.ac.jp

Hongjia Ou
ou.honjia.069@s.kyushu-u.ac.jp

1 Department of Mathematics, Irving K. Barber Faculty of Science, University of British Columbia, Kelowna, BC V1V 1V7, Canada

2 Faculty of Information Science and Electrical Engineering (ISEE), Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

Published online: 20 February 2024

Keywords Nonsmooth nonconvex optimization · Forward–reflected–backward splitting · Inertia · Bregman distance · Mordukhovich limiting subdifferential · Relative smoothness

Mathematics Subject Classification 90C26 · 49J52 · 49J53

1 Introduction

Consider the following composite minimization problem

$$\underset{x\in\overline{C}}{\text{minimize}}\ \varphi(x) := f(x) + g(x), \tag{P}$$

where $C \subseteq \mathbb{R}^n$ is a nonempty open and convex set with closure $\overline{C}$, $f : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ is proper, lower semicontinuous (lsc), and differentiable on C, and $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper and lsc (we refer to Sect. 2 for a precise statement of the assumptions on the problem). For notational brevity, we denote $\varphi_{\overline{C}} := \varphi + \delta_{\overline{C}}$ where δ_X is the indicator function of set $X \subseteq \mathbb{R}^n$, namely such that $\delta_X(x) = 0$ if $x \in X$ and $+\infty$ otherwise. By doing so, problem (P) can equivalently be cast as the "unconstrained" minimization

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \varphi_{\overline{C}}(x).$$

Note that (P) is beyond the scope of traditional first-order methods that require global Lipschitz continuity of ∇f and the consequential descent lemma [11, Prop. A.24]; see, e.g., [3, 23–25, 28, 35] for such algorithms. To resolve this issue, Lipschitz-like convexity was introduced in the seminal work [5], furnishing a descent lemma beyond the aforementioned setting. This notion was then referred to as relative smoothness (see Definition 3.2) and has played a central role in extending splitting algorithm to the setting of (P); see, e.g., [13, 17, 20, 27, 34, 39].

The goal of this paper is to propose a Bregman inertial forward–reflected–backward method i^*FRB (Algorithm 1) for solving (P), which, roughly speaking, iterates

$$x^{k+1} \in (\nabla h + \gamma\partial g)^{-1}(\nabla h(x^k) + \beta(\nabla h(x^k) - \nabla h(x^{k-1}) - \gamma(2\nabla f(x^k) - \nabla f(x^{k-1})),$$

where $\gamma > 0$ is the stepsize, β is an inertial parameter, and h is the kernel. In the convex case, the above scheme reduces to the inertial forward–reflected–backward (FRB) method proposed in [29] when $h = (1/2)\|\cdot\|^2$, which is not applicable to (P) due to its assumption on Lipschitz continuity of ∇f.

A fundamental tool in our analysis is the i^*FRB-envelope (see Definition 4.4), which is the value function associated with the parametric minimization of a "model" of (P); see Sect. 4.1. The term "envelope" is borrowed from the celebrated Moreau envelope [31] and its relation with the proximal operator. Indeed, there has been a re-emerged interest of employing an associated envelope function to study convergence of splitting methods, such as forward–backward splitting [1, 43], Douglas–Rachford splitting and ADMM [42, 44], alternating minimization algorithm [38], as well as the splitting scheme of Davis and Yin [26]. The aforementioned works share one common theme: regularity properties of the associated envelope function are used for further enhancement and deeper algorithmic insights. Similar conclusions remain valid for the case of i^*FRB, but this direction will not be pursued here and the discussion limited to Remark 5.2.

In this work, we consider an envelope function with two independent variables, allowing us to take inertial terms into account. Although merit functions with two variables have been applied in the literature, see, for instance, [14, 47], to the best of our knowledge *envelopes*, that is, the result of parametric minimizations that enjoy more regularity properties, have only been analyzed and employed as single-variable functions. Continuity properties resulting from marginalization are at the base of line-search extensions such as the one in [1], which also studies Bregman-type proximal algorithms but cannot account for inertial terms. In this regard, we believe that our methodology is appealing in its own right, as it can be instrumental for deriving inertial extensions of other splitting methods. In fact, in accounting for inertial terms, as one shall see in Sect. 4.4 a *nonpositive inertial parameter* is required for the sake of convergence under relative smoothness. This result, although more pessimistic, aligns with the recent work [18] regarding the impossibility of accelerated Bregman forward–backward method under the same assumption; see Remark 4.7 for a detailed discussion. We also note that recent research has shown that negative inertia could contribute to convergence of algorithms; see, for instance, [19, 22]. Another notable feature is that we *express (relative) smoothness of function f equivalently in terms of (relative) weak convexity* of both f and $-f$; see Lemma 3.4. Our motivation stems from the fact that the relative smoothness modulus is a two-sided condition for both f and $-f$, resulting in possibly loose results that fail to capture special structure of these functions. In contrast, treating f and $-f$ separately through their (relative) weak convexity furnishes *tight stepsize results* that better reflect the geometry of the problem; see Sect. 4.4. A similar approach was considered in [41], but to the best of our knowledge the Bregman extension investigated here is novel.

Equipped with the aforementioned novel techniques, we conduct a case study on the forward–reflected–backward splitting. Our work *differs* from the analysis carried out in [45], which also deals with an inertial forward–reflected–backward algorithm using Bregman metrics but is still limited to the Lipschitz smoothness assumption. The game changer that enables us to cope with the relative smoothness is taking the inertial step in the *dual space*, that is, interpolating application of ∇h (cf. step 2 of Algorithm 1), whence the name, inspired by [10], *mirror inertial* forward–reflected–backward splitting (i^*FRB). Despite the fact that there are simpler algorithms for solving (P), the novelty of this work emphasizes on the aforementioned theoretical contribution. Furthermore, we note that the FRB scheme demonstrates its full power when applied to minimax problems (see, e.g., [12]), in which case one shall encounter similar subproblems. In turn, we hope that the i^*FRB-envelope and the operator developed in this work, which are associated with the FRB subproblems, shall again shed light on the convergence analysis.

The rest of the paper is structured as follows. In the next section, we formally define the problem setting and the proposed mirror inertial forward–reflected–backward algorithm (i^*FRB), after providing some preliminary material and notational conventions. In Sect. 3, we revisit the notion of relative smoothness and interpret it as a two-sided relative weak convexity. After introducing the i^*FRB-envelope, these findings are used to construct a merit function for the proposed i^*FRB; the proof of the main result therein is deferred to Appendix A. The convergence analysis of i^*FRB is carried out in Sect. 5. Section 6 draws some concluding remarks.

2 Problem Setting and Proposed Algorithm

2.1 Preliminaries and Notation

We let $\mathbb{R}^n$ be the Euclidean space with norm given by $\|x\| = \sqrt{\langle x, x\rangle}$ for $x \in \mathbb{R}^n$, and $\jmath := (1/2)\|\cdot\|^2$. The extended real line is denoted by $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. The positive and negative part of $r \in \mathbb{R}$ are, respectively, defined as $[r]_+ := \max\{0, r\}$ and $[r]_- := \max\{0, -r\}$, so that $r = [r]_+ - [r]_-$.

The distance of a point $x \in \mathbb{R}^n$ to a nonempty set $S \subseteq \mathbb{R}^n$ is given by $\operatorname{dist}(x, S) = \inf_{z\in S} \|z - x\|$. The interior, closure, and boundary of S are, respectively, denoted as $\operatorname{int} S$, $\overline{S}$, and $\operatorname{bdry} S = \overline{S} \setminus \operatorname{int} S$. The *indicator function* of S is $\delta_S : \mathbb{R}^n \to \overline{\mathbb{R}}$ defined as $\delta_S(x) = 0$ if $x \in S$ and $+\infty$ otherwise.

A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *proper* if $f \not\equiv +\infty$ and $f > -\infty$, in which case its *domain* is defined as the set $\operatorname{dom} f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$. For $\alpha \in \mathbb{R}$, $[f \le \alpha] := \{x \in \mathbb{R}^n : f(x) \le \alpha\}$ denotes the α-sublevel set of f; $[\alpha \le f \le \beta]$ with $\alpha, \beta \in \mathbb{R}$ is defined accordingly. We say that f is *level bounded* (or *coercive*) if $\liminf_{\|x\|\to+\infty} f(x) = +\infty$, and 1-coercive if $\lim_{\|x\|\to+\infty} f(x)/\|x\| = +\infty$. The *Fenchel conjugate* of f is denoted as $f^* := \sup_{z\in\mathbb{R}^n}\{\langle\cdot, z\rangle - f(z)\}$. Given $x \in \operatorname{dom} f$, $\partial f(x)$ denotes the *Mordukhovich (limiting) subdifferential* of f at x, given by

$$\partial f(x) := \{v \in \mathbb{R}^n : \exists (x^k, v^k)_{k\in\mathbb{N}} \text{ s.t. } x^k \to x,\ f(x^k) \to f(x),\ \hat\partial f(x^k) \ni v^k \to v\},$$

and $\hat\partial f(x)$ is the set of *regular subgradients* of f at x, namely vectors $v \in \mathbb{R}^n$ such that $\liminf_{\substack{z\to x\\ z\neq x}} \frac{f(z)-f(x)-\langle v, z-x\rangle}{\|z-x\|} \ge 0$. The notation $\partial^\infty f(x)$ denotes the *horizon subdifferential* of f at x, defined as $\partial f(x)$ up to replacing $v^k \to v$ with $\lambda_k v^k \to v$ for some sequence $\lambda_k \searrow 0$. For $x \notin \operatorname{dom} f$, we set $\partial f(x) = \partial^\infty f(x) = \emptyset$; see, e.g., [30, 37]. $\mathscr{C}^k(\mathcal{U})$ is the set of functions $\mathcal{U} \to \mathbb{R}$ which are k times continuously differentiable, where $\mathcal{U}$ is a nonempty open set. We write $\mathscr{C}^k$ if $\mathcal{U}$ is clear from context. The notation $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ indicates a set-valued mapping, whose *domain* and *graph* are defined as $\operatorname{dom} T = \{x \in \mathbb{R}^n : T(x) \neq \emptyset\}$ and $\operatorname{gph} T = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : y \in T(x)\}$, respectively. T is said to be *outer semicontinuous* (osc) if $\operatorname{gph} T$ is a closed subset of $\mathbb{R}^n \times \mathbb{R}^m$, and *locally bounded* if every $\bar x \in \mathbb{R}^n$ admits a neighborhood $\mathcal{N}_{\bar x}$ such that $\bigcup_{x\in\mathcal{N}_{\bar x}} T(x)$ is a bounded subset of $\mathbb{R}^m$.

Following the terminology of [37, Def. 1.16], we say that a function $F : X \times U \subseteq \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ with values $F(x, u)$ is *level bounded in x locally uniformly in u* if for any $\alpha \in \mathbb{R}$ and $\bar{u} \in U$ there exists a neighborhood $\mathcal{N}_{\bar{u}}$ of $\bar{u}$ in U such that the set $\{(x, u) \in X \times \mathcal{N}_{\bar{u}} : F(x, u) \leq \alpha\}$ is bounded.

2.2 The Mirror Inertial Forward–Reflected–Backward algorithm

Throughout, we fix a 1-coercive Legendre kernel $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ with $\operatorname{dom} \nabla h = \operatorname{int} \operatorname{dom} h = C$. Recall that a proper, convex, and lsc function is said to be Legendre if it is *essentially strictly convex* and *essentially smooth*, i.e., such that h is strictly convex on C and $\|\nabla h(x_k)\| \to +\infty$ for every sequence $(x_k)_{k\in\mathbb{N}} \subset C$ converging to a boundary point of C. We will consider the following iterative scheme for addressing problem (P), where $\mathrm{D}_h : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ denotes the *Bregman distance* induced by h, defined as

$$\mathrm{D}_h(x, y) := \begin{cases} h(x) - h(y) - \langle \nabla h(y), x - y\rangle & \text{if } y \in C, \\ +\infty & \text{otherwise.} \end{cases} \tag{2.1}$$

Algorithm 1 Mirror inertial forward–reflected–backward (i^*FRB)

Choose $x^{-1}, x^0 \in C$, inertial parameter $\beta \in \mathbb{R}$, and stepsize $\gamma > 0$

Iterate for $k = 0, 1, \ldots$ until a termination criterion is met (cf. Lemma 5.4)

1: Set y^k such that $\nabla h(y^k) = \nabla h(x^k) - \gamma\big(\nabla f(x^k) - \nabla f(x^{k-1})\big)$

2: Choose $x^{k+1} \in \operatorname*{argmin}_{w\in\mathbb{R}^n}\{g(w) + \langle w, \nabla f(x^k) - \tfrac{\beta}{\gamma}\big(\nabla h(x^k) - \nabla h(x^{k-1})\big)\rangle + \tfrac{1}{\gamma}\,\mathrm{D}_h(w, y^k)\}$

Note that Algorithm 1 takes inertial step in the dual space, hence the abbreviation i^*FRB. We will work under the following assumptions.

Assumption 1 The following hold in problem (P):

A1 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is smooth relative to h (see Sect. 3).
A2 $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper and lsc.
A3 $\inf \varphi_{\overline{C}} > -\infty$.
A4 For any $v \in \mathbb{R}^n$ and $\gamma > 0$, $\operatorname{argmin}\{\gamma g + h - \langle v, \cdot\rangle\} \subseteq C$.
A5 For any $\gamma > 0$, $\lim_{\|x\|\to\infty} \frac{\gamma g(x)+h(x)}{\|x\|} = \infty$.

Some remarks are in order.

Remark 2.1 (constraint qualifications for Assumption 1.A4) As will be made explicit in Lemma 4.2, Assumptions 1.A4 and 1.A5 are requirements ensuring that Algorithm 1 is well defined. Note that, in general, the minimizers therein are a (possibly empty) subset of $\operatorname{dom} h \cap \operatorname{dom} g$; Assumption 1.A4 thus only excludes points on the boundary of $\operatorname{dom} h$. This standard requirement is trivially satisfied when $\operatorname{dom} h$ is open, or more generally when constraint qualifications enabling a subdifferential calculus rule on the boundary are met, as is the case when g is convex. If g is proper and lsc, Assumption 1.A4 is satisfied if $\partial^\infty g \cap \big(-\partial^\infty h\big) \subseteq \{0\}$ holds everywhere (this condition being automatically guaranteed at all point outside the boundary of C, having $\partial^\infty h$ empty outside $\operatorname{dom} h$ and $\{0\}$ in its interior). Indeed, optimality of $\bar{x} \in \operatorname{argmin}\{\gamma g + h - \langle v, \cdot \rangle\}$ implies that $v \in \partial[\gamma g + h](\bar{x}) \subseteq \gamma \partial g(\bar{x}) + \partial h(\bar{x})$, with inclusion holding by [37, Cor. 10.9] and implying nonemptiness of $\partial h(\bar{x})$; see Sect. 2.1 for definitions of subdifferentials.

Remark 2.2 (Assumption 1.A5 and prox-boundedness) Apparently, Assumption 1.A5 together with lower semicontinuity ensures that minimizers of $\gamma g + h - \langle v, \cdot \rangle$ do exist for any γ and v. (Relative) prox-boundedness [21, Def. 2.3], which amounts to the same condition but only required for γ small enough, would also suffice to our purposes as long as parameters exceeding such “threshold” are excluded from the analysis. Assumption 1.A5 is nevertheless a very mild and standard requirement [13] that enables a simpler exposition at virtually no expense of generality. We additionally remark that this requirement is superfluous whenever f and h are continuous relative to $\operatorname{dom} h$, or when $\operatorname{dom} h$ has bounded intersection with its boundary. We postpone the discussion to Lemma 3.6 for the details.

3 Relative Smoothness and Weak Convexity

Throughout the paper, we will adopt the convention that $1/0 = +\infty$ and $+\infty \cdot 0 = 0$. In order to resolve possible ill definitions of difference of extended real-valued functions, we adopt the extended arithmetics $\dot{+}$ and $\dot{-}$ of [32, §2], defined as the respective $+$ and $-$ whenever the operation makes sense and otherwise evaluating to $+\infty$, namely

$$-\infty \mathbin{\dot{-}} (-\infty) = -\infty \mathbin{\dot{+}} \infty = +\infty \mathbin{\dot{-}} \infty = +\infty \mathbin{\dot{+}} (-\infty) = \infty.$$

Furthermore, we denote by $\dot{\pm}$ and $\dot{\mp}$ the extended arithmetic equivalents of $\pm$ and $\mp$, respectively. Notice in particular that

$$a \mathbin{\dot{-}} b = (-b) \mathbin{\dot{-}} (-a) \tag{3.1}$$

holds for any extended-real pair $(a, b) \in \overline{\mathbb{R}} \times \overline{\mathbb{R}}$.

The following lemma collects other properties of extended arithmetics that will be frequently used throughout.

Lemma 3.1 *Let $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, and $a, b \in \mathbb{R}$ be fixed. Then, the following hold:*

(i) $a\psi \pm b\psi \mathbin{\dot{+}} \delta_E = (a \pm b)\psi \mathbin{\dot{+}} \delta_E$ for any $E \subseteq \operatorname{dom} \psi$.
(ii) If $a > 0$, then $a\psi \mathbin{\dot{\pm}} b\psi = (a \pm b)\psi \mathbin{\dot{+}} \delta_{\operatorname{dom} \psi}$.

Proof In both cases, we shall verify the equivalences pointwise at any $x \in \mathbb{R}^n$, analyzing the cases $x \in \operatorname{dom} \psi$ and $x \notin \operatorname{dom} \psi$ separately.

If $x \in \operatorname{dom} \psi$, one has that $\psi(x) \in \mathbb{R}$ by properness of ψ; hence, the extended arithmetic notation is superfluous in both assertions and the claims are trivially true. If $x \notin \operatorname{dom} \psi$, then $\delta_E(x) = +\infty$ in assertion 3.1.(i) and similarly $a\psi(x) = \infty$

in assertion 3.1.(ii). Therefore, the extended arithmetic convention ensures that all expressions evaluate to $+\infty$ in this case, and thus coincide. □

Definition 3.2 (relative smoothness) We say that a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is smooth relative to h if $\operatorname{dom} f \supseteq \operatorname{dom} h$ and there exists a constant $L_{f,h} \geq 0$ such that

$$L_{f,h}h \dotplus f \dotplus \delta_C \quad \text{and} \quad L_{f,h}h \dot{-} f \dotplus \delta_C \tag{3.2}$$

are proper convex functions. We may alternatively say that f is $L_{f,h}$-smooth relative to h to make the smoothness modulus $L_{f,h}$ explicit.

The addition of δ_C in (3.2) serves the purpose of assessing convexity of $L_{f,h} \pm f$ only on set C, in line with the original "Lipschitz-like convexity" notion of [5, §2.2] as well as the subsequent nonconvex generalization in [13, Def. 2.2], referred to as "$L_{f,h}$-smooth adaptability" of the pair (f, h). Differently from those works, however, we do not impose (continuous) differentiability of f on C in the definition, for this property is automatically guaranteed; see Proposition 3.7 for the details.

Notice further that the constant $L_{f,h}$ may be loose. For instance, if f is convex, then $Lh \dotplus f$ is convex for any $L \geq 0$, indicating that it is only convexity of $L_{f,h}h \dot{-} f \dotplus \delta_C$ that dictates the value of $L_{f,h}$. This motivates us to consider one-sided conditions and treat f and $-f$ separately.

Definition 3.3 (relative weak convexity) We say that a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is weakly convex relative to h if $f \dotplus \delta_{\operatorname{dom} h}$ is proper and there exists a (possibly negative) constant $\sigma_{f,h} \in \mathbb{R}$ such that $f \dot{-} \sigma_{f,h}h \dotplus \delta_C$ is a convex function. We may alternatively say that f is $\sigma_{f,h}$-weakly convex relative to h to make the weak convexity modulus $\sigma_{f,h}$ explicit.

In accordance with the Euclidean case, having $\sigma_{f,h} \geq 0$ implies convexity while $\sigma_{f,h} > 0$ relative strong convexity. Considering possibly improper functions in Definition 3.3 allows us to identify relative smoothness as a two-sided relative weak convexity, as we are about to show, regardless of whether a function has full domain or not. This fact extends the well-known equivalence between Lipschitz differentiability and the combination of weak convexity and weak concavity in the Euclidean setting; see Lemma 3.8 for the details.

Lemma 3.4 *(relative smoothness and relative weak convexity) Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper. Then, f is smooth relative to h iff both f and $-f$ are weakly convex relative to h. More precisely, if f is $L_{f,h}$-smooth relative to h, then both f and $-f$ are $(-L_{f,h})$-relatively weakly convex. Conversely, if f and $-f$ are $\sigma_{f,h}$- and $\sigma_{-f,h}$-weakly convex relative to h, respectively, then f (as well as $-f$) is $L_{f,h}$-smooth relative to h with*

$$L_{f,h} = \max\{|\sigma_{f,h}|, |\sigma_{-f,h}|\} \tag{3.3}$$

(see (3.6) for a simplified expression without absolute values).

Proof That relative smoothness implies relative weak convexity with the given moduli is straightforward (properness of $\pm f \dotplus \delta_{\operatorname{dom} h}$ follows from the inclusion $\operatorname{dom} f \supseteq$

dom h). Suppose that $\pm f$ are $\sigma_{\pm f,h}$-relatively weakly convex. First, observe that properness of $\pm f \dotplus \delta_{\operatorname{dom} h}$ implies the inclusion dom $f \supseteq$ dom h. Then, the convexity of $\pm f \dot{-} \sigma_{\pm f,h} h \dotplus \delta_C$ implies that

$$\begin{aligned} L_{f,h} h \dot{\pm} f \dotplus \delta_C &= (-\sigma_{\pm f,h} + \sigma_{\pm f,h} + L_{f,h}) h \dot{\pm} f \dotplus \delta_C \\ &= -\sigma_{\pm f,h} h \dotplus \sigma_{\pm f,h} h \dotplus L_{f,h} h \dot{\pm} f \dotplus \delta_C \\ &= \left[\pm f \dot{-} \sigma_{\pm f,h} h \dotplus \delta_C\right] + \left[(L_{f,h} + \sigma_{\pm f,h}) h\right] \end{aligned}$$

are proper and convex (since $L_{f,h} + \sigma_{\pm f,h} \geq 0$, cf. (3.3)), where the second identity uses Lemma 3.1.(i) to distribute the coefficients of h. Appealing to Definition 3.2, f is $L_{f,h}$-smooth relative to h. □

The relative weak convexity moduli $\sigma_{\pm f,h}$ will be henceforth adopted when referring to Assumption 1.A1. It will be convenient to *normalize* these quantities into pure numbers

$$p_{\pm f,h} := \tfrac{\sigma_{\pm f,h}}{L_{f,h}} \in [-1, 1]. \tag{3.4}$$

Notice that $L_{f,h} = 0$ only when f is affine on C, and in this case, we conventionally set $p_{\pm f,h} = 0$. The comment below will be instrumental in Sect. 4.4.

Remark 3.5 If f and $-f$ are $\sigma_{f,h}$- and $\sigma_{-f,h}$-weakly convex relative to h, respectively, then invoking (3.3) and (3.4) yields that

$$-2 \leq p_{f,h} + p_{-f,h} \leq 0 \quad \text{and} \quad -1 \in \{p_{f,h}, p_{-f,h}\}, \tag{3.5}$$

where the inclusion holds provided $L_{f,h} \neq 0$. (As said above, the case $L_{f,h} = 0$ amounts to f being affine on C.) The second inequality owes to the fact that, by definition, both $f - \sigma_{f,h} h \dotplus \delta_C$ and $-f - \sigma_{-f,h} h \dotplus \delta_C$ are convex functions, and therefore so is their sum

$$\begin{aligned} \left[f \dot{-} \sigma_{f,h} h\right] \dotplus \left[-f \dot{-} \sigma_{-f,h} h\right] \dotplus \delta_C &= (-\sigma_{f,h} - \sigma_{-f,h}) h \dotplus \delta_C \\ &= L_{f,h}(-p_{f,h} - p_{-f,h}) h \dotplus \delta_C, \end{aligned}$$

where we used Lemma 3.1.(i) together with the fact that $f \dot{-} f = \delta_{\operatorname{dom} f}$ and that dom $f \supseteq$ dom $h \supseteq C$. In turn, the inclusion in (3.5) follows from (3.3) and the definition (3.4): indeed, as long as $L_{f,h} \neq 0$, (3.4) implies that *at least* one among $p_{f,h}$ and $p_{-f,h}$ *must* attain absolute value of one. If the value is 1, then combining inequalities (3.4) and (3.5) entails the desired inclusion. Thus, whenever f is *convex (resp. concave)*, since *one can take* $p_{f,h} = 0$ (resp. $p_{-f,h} = 0$), by virtue of the inclusion in (3.5) it directly follows that $p_{-f,h} = -1$ (resp. $p_{f,h} = -1$) *must* hold.

Notice that the condition $\sigma_{f,h} + \sigma_{-f,h} \leq 0$ shown in the above remark yields a simplification in the expression (3.3), for the absolute values can be resolved to

$$L_{f,h} = \max\{|\sigma_{f,h}|, |\sigma_{-f,h}|\} = \max\{-\sigma_{f,h}, -\sigma_{-f,h}\}. \tag{3.6}$$

We now turn to a lemma that guarantees well definedness of Algorithm 1.

Lemma 3.6 *(relative prox-boundedness) suppose that Assumption 1 holds. Then, the set* $\operatorname{argmin}\{\gamma g + h - \langle v, \cdot \rangle\}$ *as in Assumption 1.A4 is nonempty for any* $v \in \mathbb{R}^n$ *and* $0 < \gamma < 1/[\sigma_{-f,h}]_-$. *In other words, g is prox-bounded relative to h with threshold* $\gamma_{g,h} \geq 1/[\sigma_{-f,h}]_-$ *[21, Def. 2.3].*

In fact, the claim still holds with Assumption 1.A5 being replaced by continuity of f and h relative to $\operatorname{dom} h$, *or by the weaker condition*

$$\lim_{\substack{\|x\|\to\infty \\ x \in \operatorname{bdry} \operatorname{dom} h}} \frac{\gamma g(x) + h(x)}{\|x\|} = \infty \tag{3.7}$$

for any $v \in \mathbb{R}^n$ *and* $\gamma < 1/[\sigma_{-f,h}]_-$.

Proof The claim is obvious if Assumption 1 holds in its entirety, and in fact, the restrictions on γ are superfluous in this case; see the commentary within Remark 2.2 and [13, Lem. 3.1] for a formal proof. We now show the sufficiency of the claimed alternatives to Assumption 1.A5. Recall that a proper convex function admits an affine minorant; see, e.g., [9, Cor. 16.18]. By observing that

$$\begin{aligned} \gamma g + h + \delta_C &= \gamma \varphi_{\overline{C}} \stackrel{\cdot}{-} \gamma f + h + \delta_C \\ &= \gamma \underbrace{\varphi_{\overline{C}}}_{\geq \inf \varphi_{\overline{C}}} + \gamma \underbrace{\left(-f \stackrel{\cdot}{+} [\sigma_{-f,h}]_- h + \delta_C\right)}_{\text{proper and convex}} + \underbrace{\left(1 - \gamma[\sigma_{-f,h}]_-\right)}_{>0} \overbrace{h}^{\text{1-coercive}}, \end{aligned}$$

it follows that $\gamma g + h + \delta_C$ is 1-coercive. Therefore, for any $v \in \mathbb{R}^n$ the function $\gamma g + h + \delta_C - \langle v, \cdot \rangle$ is level bounded. Observe that $\operatorname{argmin}\{\gamma g + h - \langle v, \cdot \rangle\} = \operatorname{argmin}\{\gamma g + h + \delta_C - \langle v, \cdot \rangle\}$ provided that the left-hand side is nonempty, which owes to Assumption 1.A4. To fix a notation, let

$$\psi := \gamma g + h - \langle v, \cdot \rangle \quad \text{and} \quad \mathring{\psi} := \psi + \delta_C .$$

Since ψ is lsc and $\mathring{\psi}$ is 1-coercive, the sets of minimizers are (nonempty and) compact provided that $\operatorname{argmin} \psi \neq \emptyset$. It thus suffices to show that indeed $\operatorname{argmin} \psi$ is nonempty. Let $(x^k)_{k\in\mathbb{N}}$ be a minimizing sequence for ψ, namely such that $\psi(x^k) \to \inf \psi$. Since ψ is lsc, it suffices to show that $(x^k)_{k\in\mathbb{N}}$ is bounded.

If $(x^k)_{k\in\mathbb{N}}$ is unbounded, then coercivity of $\mathring{\psi}$ implies that $x^k \in \operatorname{bdry} C$ for k large enough. This clearly cannot happen under condition (3.7). Suppose instead that h and f are continuous on $\operatorname{dom} h$. Then,

$$h \stackrel{\cdot}{-} \gamma f = \gamma \underbrace{\left(-f \stackrel{\cdot}{-} \sigma_{-f,h} h\right)}_{\text{convex on } C} \stackrel{\cdot}{+} \underbrace{\left(1 + \gamma \sigma_{-f,h}\right)}_{>0} h$$

is both 1-coercive (and convex) on $C = \operatorname{int} \operatorname{dom} h$, in the sense that $h \stackrel{\cdot}{-} \gamma f + \delta_C$ is 1-coercive, and continuous on its domain $\operatorname{dom} h$, and consequently, it is 1-coercive on the entire space.[1] Therefore, also

$$\psi = \gamma \varphi_{\overline{C}} - \langle v, \ \cdot \rangle + (h \dot{-} \gamma f) \geq \gamma \inf \varphi_{\overline{C}} - \langle v, \ \cdot \rangle + (h \dot{-} \gamma f)$$

is 1-coercive, which shows that also in this case the minimizing sequence $(x^k)_{k \in \mathbb{N}}$ cannot be unbounded. □

We now discuss "transitivity" properties of relative smoothness, beginning from continuous differentiability. We point out that the following result is well known; however, the proof is included for completeness.

Proposition 3.7 *Suppose that f is $L_{f,h}$-smooth relative to h. Then f is continuously differentiable on C.*

Proof By assumption, $\hat{\partial}(L_{f,h} h \dot{\pm} f)$ are nonempty on C. (In particular, on C the extended arithmetic notation is redundant.) The subdifferential sum rule yields that $(\forall x \in C)\ \hat{\partial}(L_{f,h} h \pm f)(x) = L_{f,h} \nabla h(x) + \hat{\partial}(\pm f)(x)$, implying that $\hat{\partial}(\pm f)(x)$ must be nonempty. The smoothness of h implies that $\pm f$ are regular through [37, Ex. 8.20]. The proof then follows by invoking [37, Thm. 9.18(a)–(d) and Cor. 9.19(a)–(b)]. □

Next we turn to Lipschitz differentiability. The result below is a (well-known) generalization of the (well-known) equivalence between smoothness relative to the Euclidean kernel $\jmath$ and Lipschitz differentiability, a fact that will be invoked in Sect. 4 and whose proof is given next for the sake of completeness.

Lemma 3.8 *(Lipschitz smoothness from weak convexity) For any $F : \mathbb{R}^n \to \overline{\mathbb{R}}$, the following are equivalent:*

(a) There exist $\sigma_{\pm F} \in \mathbb{R}$ such that both $F - \sigma_F \jmath$ and $-F - \sigma_{-F} \jmath$ are proper, convex, and lsc;
(b) $\operatorname{dom} \partial F = \mathbb{R}^n$, and there exist $\sigma_{\pm F} \in \mathbb{R}$ such that for all $(x_i, v_i) \in \operatorname{gph} \partial F$, $i = 1, 2$, it holds that $\sigma_F \|x_1 - x_2\|^2 \leq \langle v_1 - v_2, x_1 - x_2 \rangle \leq -\sigma_{-F} \|x_1 - x_2\|^2$;
(c) There exists $L_F \geq 0$ such that F is L_F-smooth relative to $\jmath$;
(d) There exists $L_F \geq 0$ such that ∇F is L_F-Lipschitz differentiable.

In particular, assertions 3.8.(a) and/or 3.8.(b) imply assertions 3.8.(c) and 3.8.(d) with $L_F = \max\{-\sigma_F, -\sigma_{-F}\}$, and conversely, 3.8.(c) and/or 3.8.(d) imply 3.8.(a) and 3.8.(b) with $\sigma_{\pm F} = -L_F$.

Proof 3.8.(a) ⇔ 3.8.(c) The equivalence between the statements as well as the relation between the constants follows from Lemma 3.4 and (3.6).

3.8.(a) ⇒ 3.8.(b) Function $\psi := F - \sigma_F \jmath$ is convex, and therefore, its subgradient $\partial \psi = \partial F - \sigma_F \operatorname{id}$ is monotone. This readily shows the first inequality in assertion 3.8.(b). The second inequality will follow from the same argument applied to the convex function $-F - \sigma_{-F} \jmath$ once we show that $\partial(-F) = -\partial F$. Indeed, it follows

[1] Take $x^k \in \operatorname{dom} h \setminus C$ with $\|x^k\| \to \infty$. Since $\operatorname{dom} h$ is convex, its interior C is nonempty, and f is continuous on $\operatorname{dom} h$, for each k there exists $\tilde{x}^k \in C$ with $\|x^k - \tilde{x}^k\| \leq 1$ such that $h(\tilde{x}^k) - \gamma f(\tilde{x}^k) \leq h(x^k) - \gamma f(x^k) + 1$. By 1-coercivity on $C \ni \tilde{x}^k$, $(h(\tilde{x}^k) - \gamma f(\tilde{x}^k)) / \|\tilde{x}^k\| \to \infty$, implying that $(h(x^k) - \gamma f(x^k)) / \|x^k\| \to \infty$ as well.

from [37, Ex. 12.28(b),(c)] that both F and $-F$ are lower-$\mathscr{C}^2$, in the sense of [37, Def. 10.29], hence continuously differentiable by [37, Prop. 10.30]. Thus, invoking [37, Thm. 9.18] (which applies by virtue of [37, Cor. 9.19(a)–(b)]) one has that $\partial F = \{\nabla F\} = \{-\nabla(-F)\} = -\partial(-F)$, as claimed.

3.8.(b) ⇒ 3.8.(d) It follows from [42, Lem. 2.1] that F is continuously differentiable and satisfies $|\langle \nabla F(x) - \nabla F(y), x-y\rangle| \leq L_F\|x-y\|^2$ with $L_F := \max\{|\sigma_F|, |\sigma_{-F}|\}$. In turn, simple algebra yields

$$0 \leq \langle \nabla(F + L_F\jmath)(x) - \nabla(F + L_F\jmath)(y), x-y\rangle \leq 2L_F\|x-y\|^2.$$

By virtue of [33, Thm. 2.1.5], function $F + L_F\jmath$ is convex with $(1/2L_F)$-cocoercive gradient, namely such that

$$\begin{aligned}&\langle \nabla(F + L_F\jmath)(x) - \nabla(F + L_F\jmath)(y), x-y\rangle\\ &\geq \tfrac{1}{2L_F}\|\nabla(F + L_F\jmath)(x) - \nabla(F + L_F\jmath)(y)\|^2\end{aligned}$$

for all $x, y \in \mathbb{R}^n$. Expanding the square and rearranging yields the sought Lipschitz inequality $\|\nabla F(x) - \nabla F(y)\|^2 \leq L_F^2\|x-y\|^2$.

3.8.(d) ⇒ 3.8.(a) From the quadratic upper bound [11, Prop. A.24], it follows that

$$\pm F(x_2) \geq \pm F(x_1) \pm \langle \nabla F(x_1), x_2 - x_1\rangle - \tfrac{L_F}{2}\|x_2 - x_1\|^2,$$

or, equivalently,

$$(L_F\jmath \pm F)(x_2) \geq (L_F\jmath \pm F)(x_1) + \langle \nabla(L_F\jmath \pm F)(x_1), x_2 - x_1\rangle.$$

This proves convexity of $L_F\jmath \pm F$, whence the claim by taking $\sigma_{\pm F} = -L_F$. □

The above lemma can be used to show that a function which is smooth relative to h is Lipschitz differentiable whenever h is, as shown next. The proof hinges on the following more general "transitivity" property of relative smoothness.

Lemma 3.9 *Let $h_1, h_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$ be Legendre kernels, and let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. If f is L_{f,h_1}-smooth relative to h_1 and h_1 is L_{h_1,h_2}-smooth relative to h_2, then f is L_{f,h_2}-smooth relative to h_2 with $L_{f,h_2} = L_{f,h_1}L_{h_1,h_2}$.*

Proof By definition, $\operatorname{dom} h_2 \subseteq \operatorname{dom} h_1 \subseteq \operatorname{dom} f$, and $L_{f,h_1}h_1 \dot{\pm} f + \delta_{C_1}$ and $L_{h_1,h_2}h_2 \dot{\pm} h_1 + \delta_{C_2}$ are all proper convex functions, where $C_i := \operatorname{int}\operatorname{dom} h_1, i = 1, 2$. If $L_{f,h_1} = 0$, then f is affine on C_1 as discussed in Lemma 3.4, and the claim is trivially true. Suppose that $L_{f,h_1} > 0$, and notice that necessarily $L_{h_1,h_2} > 0$ too holds since h_1 is strictly convex. Thus,

$$\begin{aligned}L_{f,h_1}L_{h_1,h_2}h_2 \pm f \dot{+} \delta_{C_2} &= L_{f,h_1}L_{h_1,h_2}h_2 \pm f + \delta_{C_2}\\ &= L_{f,h_1}(L_{h_1,h_2}h_2 - h_1) + (L_{f,h_1}h_1 \pm f) + \delta_{C_2}\end{aligned}$$

are also convex functions, where the second identity uses Lemma 3.1.(i) together with the fact that $\operatorname{dom} h_2 \subseteq \operatorname{dom} h_1$. In fact, they are also proper since the domains include

$\operatorname{dom} h_2 \neq \emptyset$. By Definition 3.2, this means that f is $L_{f,h_1} L_{h_1,h_2}$-smooth relative to h_2. □

When $h_2 = j$, and by appealing to the equivalence between Lipschitz differentiability and smoothness relative to h_2 asserted in Lemma 3.8, the following special case is obtained.

Corollary 3.10 *Suppose that f is $L_{f,h}$-smooth relative to h, and that h is L_h-Lipschitz differentiable. Then f is Lipschitz differentiable with modulus $L_f = L_{f,h} L_h$.*

We conclude the section with a result regarding relative weak convexity and smoothness of linear combinations that will be useful in the next section.

Lemma 3.11 *Suppose that f is smooth relative to h, and let $\sigma_{\pm f,h}$ be the weak hypoconvexity moduli of $\pm f$ relative to h. Then, for every $\alpha, \beta \in \mathbb{R}$ the function $\psi := \alpha f \dot{+} \beta h$ is smooth relative to h with*

$$\sigma_{\psi,h} = \begin{cases} |\alpha| \sigma_{f,h} + \beta & \text{if } \alpha \geq 0, \\ |\alpha| \sigma_{-f,h} + \beta & \text{if } \alpha < 0, \end{cases} \tag{3.8a}$$

$$\sigma_{-\psi,h} = \begin{cases} |\alpha| \sigma_{-f,h} - \beta & \text{if } \alpha \geq 0, \\ |\alpha| \sigma_{f,h} - \beta & \text{if } \alpha < 0, \end{cases} \tag{3.8b}$$

and

$$|\alpha| L_{f,h} + |\beta| \geq L_{\psi,h} = \begin{cases} \max\{-\beta - \alpha\sigma_{f,h}, \beta - \alpha\sigma_{-f,h}\} & \text{if } \alpha \geq 0, \\ \max\{\beta + \alpha\sigma_{f,h}, -\beta + \alpha\sigma_{-f,h}\} & \text{if } \alpha < 0. \end{cases} \tag{3.8c}$$

Proof If $\alpha = 0$ the claim is trivial. If $\alpha > 0$, then for every $\sigma \in \mathbb{R}$ we have

$$\psi \dot{-} \sigma h \dot{+} \delta_C = \alpha\big[(f \dot{-} \sigma_{f,h} h \dot{+} \delta_C) \dot{+} \tfrac{\beta - \sigma + \alpha\sigma_{f,h}}{\alpha} h\big],$$

where Lemma 3.1.(i) was used to distribute the coefficients of h. Since the term in round brackets is (proper and) convex, for any $\sigma \leq \alpha\sigma_{f,h} + \beta$ one has that $\psi \dot{-} \sigma h \dot{+} \delta_C$ is convex. Clearly, it is also proper, with domain agreeing with $\operatorname{dom} h$. If $\alpha < 0$, the same arguments can be used via the identity

$$\psi \dot{-} \sigma h \dot{+} \delta_C = -\alpha\big[(-f \dot{-} \sigma_{-f,h} h \dot{+} \delta_C) \dot{+} \tfrac{\beta - \sigma - \alpha\sigma_{-f,h}}{-\alpha} h\big].$$

The expression for $\sigma_{-\psi,h}$ follows by replacing α and β with $-\alpha$ and $-\beta$; in turn, the expression for $L_{\psi,h}$ follows from (3.6). □

4 Algorithmic Analysis Toolbox

In the literature, convergence analysis for nonconvex splitting algorithms typically revolves around the identification of a "Lyapunov potential," namely a lower bounded function that decreases its value along the iterates. In this section, we will pursue this direction. To simplify the discussion, we introduce

$$\hat{h} := \tfrac{1}{\gamma}h \mathbin{\dot{-}} f \quad \text{and} \quad \hat{f}_\beta := f \mathbin{\dot{-}} \tfrac{\beta}{\gamma}h \mathbin{\dot{+}} \delta_{\operatorname{dom} h}\,. \tag{4.1}$$

Notice that $\hat{f}_\beta$ is a proper function with $\operatorname{dom} \hat{f}_\beta = \operatorname{dom} h$ for any $\beta \in \mathbb{R}$, but for strictly positive values of β it may fail to be lsc at some boundary points of C. This will nevertheless cause no concern in the analysis of Algorithm 1, since, as will be showcased in Lemma 4.2.(iii), its iterates remain confined within the open set C onto which $\hat{f}_\beta$ is continuously differentiable. On the other hand, not only is $\hat{h}$ lsc on the whole $\mathbb{R}^n$, but it is actually a Legendre kernel, for γ small enough.

Lemma 4.1 *([1, Thm. 4.1]) Suppose that Assumption 1.A1 holds. Then, for every* $\gamma < 1/[\sigma_{-f,h}]_-$ *the function* $\hat{h}$ *is a Legendre kernel with* $\operatorname{dom} \hat{h} = \operatorname{dom} h$.[2]

Notice further that, as a linear combination of f and h, we may invoke Lemma 3.11 to infer that $\hat{f}_\beta$ is smooth relative to h with

$$\sigma_{\pm \hat{f}_\beta,h} = \tfrac{\gamma\sigma_{\pm f,h} \mp \beta}{\gamma} \quad \text{and} \quad L_{\hat{f}_\beta,h} = \tfrac{1}{\gamma}\max\{\beta - \gamma\sigma_{f,h},\ -\beta - \gamma\sigma_{-f,h}\}. \tag{4.2}$$

We will also (ab)use the notation D_ψ of the Bregman distance for functions $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ differentiable on C that are not necessarily convex. This notational abuse is justified by the fact that all algebraic identities of the Bregman distance used in the manuscript (e.g., the three-point identity [16, Lem. 3.1]) are valid regardless of whether ψ is convex or not, and will overall yield a major simplification of the math. In particular, for any ψ_1, ψ_2 that are continuously differentiable on C and for any $\lambda \in \mathbb{R}$ we may exploit the identities $\mathrm{D}_{\psi_1 \dot{+} \psi_2} = \mathrm{D}_{\psi_1} \mathbin{\dot{+}} \mathrm{D}_{\psi_2}$, $\mathrm{D}_{\psi_1 \dot{-} \psi_2} = \mathrm{D}_{\psi_1} \mathbin{\dot{-}} \mathrm{D}_{\psi_2}$, and $\mathrm{D}_{\lambda\psi_1} = \lambda\,\mathrm{D}_{\psi_1}$, holding on $\mathbb{R}^n \times C$, with no concern about the sign of λ or whether either function is convex or not.

4.1 Parametric Minimization Model

As a first step toward the desired goals, as well as to considerably simplify the discussion, we begin by observing that the i^*FRB-update is the result of a parametric minimization. To this end, we introduce the "model" $\mathcal{M}^{h\text{-FRB}}_{\gamma,\beta} : \operatorname{dom} h \times C \times C \to \overline{\mathbb{R}}$ defined by

$$\begin{aligned}\mathcal{M}^{h\text{-FRB}}_{\gamma,\beta}(w; x, x^-) &= \varphi(w) + \mathrm{D}_{\hat{h}}(w, x) + \langle w - x, \nabla \hat{f}_\beta(x) - \nabla \hat{f}_\beta(x^-)\rangle && \text{(4.3a)}\\ &= \varphi(w) + \mathrm{D}_{\hat{h} - \hat{f}_\beta}(w, x) + \mathrm{D}_{\hat{f}_\beta}(w, x^-) - \mathrm{D}_{\hat{f}_\beta}(x, x^-), && \text{(4.3b)}\end{aligned}$$

[2] The equivalence of the domains follows from the inclusion $\operatorname{dom} f \supseteq \operatorname{dom} h$.

where the last equality holds due to the well-known three-point identity (see [16, Lem. 3.1]). Notice that no extended arithmetics are necessary in the above formulae due to the restriction $(w, x, x^-) \in \operatorname{dom} h \times C \times C$ which guarantees the finiteness of all quantities involved, except possibly $\varphi(w)$. Then, adding constant terms from the x-update in i^*FRB yields

$$\begin{aligned} x^{k+1} &\in \operatorname*{argmin}_{w\in\mathbb{R}^n}\{g(w) + \langle w, \nabla f(x^k) - \tfrac{\beta}{\gamma}\big(\nabla h(x^k) - \nabla h(x^{k-1})\big)\rangle + \tfrac{1}{\gamma}\,\mathrm{D}_h(w, y^k)\} \\ &= \operatorname*{argmin}_{w\in\operatorname{dom} h}\{\varphi(w) + \hat{h}(w) + \langle w, \nabla \hat{f}_\beta(x^k) + \tfrac{\beta}{\gamma}\nabla h(x^{k-1}) - \tfrac{1}{\gamma}\nabla h(y^k)\rangle\} \\ &= \operatorname*{argmin}_{w\in\operatorname{dom} h}\{\varphi(w) + \hat{h}(w) + \langle w, \nabla \hat{f}_\beta(x^k) - \nabla \hat{f}_\beta(x^{k-1}) - \nabla\hat{h}(x^k)\rangle\} \\ &= \operatorname*{argmin}_{w\in\operatorname{dom} h} \mathcal{M}^{h\text{-FRB}}_{\gamma,\beta}(w; x^k, x^{k-1}), \end{aligned}$$

where the second last equality owes to the relation $\nabla h(y^k) = \nabla h(x^k) - \gamma\big(\nabla f(x^k) - \nabla f(x^{k-1})\big)$ (recall step 1 of i^*FRB). It follows that the x-update in i^*FRB can be compactly expressed as

$$x^{k+1} \in \mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x^k, x^{k-1}), \tag{4.4a}$$

where $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta} : C \times C \rightrightarrows C$ defined by

$$\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x, x^-) := \operatorname*{argmin}_{w\in\operatorname{dom} h} \mathcal{M}^{h\text{-FRB}}_{\gamma,\beta}(w; x, x^-) \tag{4.4b}$$

is the i^*FRB-operator with stepsize γ and inertial parameter β. The fact that $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}$ maps pairs in $C \times C$ to subsets of C is a consequence of Assumption 1.A4, as we are about to formalize in Lemma 4.2.(iii). Note that many models $\mathcal{M}^{h\text{-FRB}}_{\gamma,\beta}$ can be defined whose marginal minimization with respect to the first variable results in the same $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}$, and all these differ by additive terms which are constant with respect to w. Among these, the one given in (4.3b) reflects the tangency condition $\mathcal{M}^{h\text{-FRB}}_{\gamma,\beta}(x; x, x^-) = \varphi(x) = \varphi_{\overline{C}}(x)$ for every $x, x^- \in C$. A consequence of this fact and other basic properties are summarized next.

Lemma 4.2 *(basic properties of the model $\mathcal{M}^{h\text{-FRB}}_{\gamma,\beta}$ and the operator $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}$) Suppose that Assumption 1 holds, and let $\gamma < 1/[\sigma_{-f,h}]_-$ and $\beta \in \mathbb{R}$ be fixed. The following hold:*

(i) $\mathcal{M}^{h\text{-FRB}}_{\gamma,\beta}(x; x, x^-) = \varphi_{\overline{C}}(x)$ *for all* $x, x^- \in C$.

(ii) $\mathcal{M}^{h\text{-FRB}}_{\gamma,\beta}(w; x, x^-)$ *is level bounded in* w *locally uniformly in* (x, x^-).

(iii) $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}$ *is locally bounded and osc,*[3] *and* $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x, x^-)$ *is a nonempty and compact subset of* C *for any* $x, x^- \in C$.

[3] Being $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}$ defined on $C \times C$, osc and local boundedness are meant relative to $C \times C$. Namely, gph $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}$ is closed relative to $C \times C \times \mathbb{R}^n$, and $\bigcup_{(x,x^-)\in\mathcal{V}} \mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x, x^-)$ is bounded for any $\mathcal{V} \subset C \times C$ compact.

(iv) $\nabla\hat{h}(x) - \nabla\hat{h}(\bar{x}) - \nabla\hat{f}_\beta(x) + \nabla\hat{f}_\beta(x^-) \in \hat{\partial}\varphi(\bar{x})$ *for any* $x, x^- \in C$ *and* $\bar{x} \in \mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x, x^-)$.

(v) *If* $x \in \mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x, x)$, *then* $0 \in \hat{\partial}\varphi(x)$ *and* $\mathrm{T}^{h\text{-FRB}}_{\gamma',\beta}(x, x) = \{x\}$ *for every* $\gamma' \in (0, \gamma)$.

Proof We start by observing that Lemma 3.6 ensures that $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x, x^-)$ is nonempty for any $x, x^- \in C$; this follows by considering the expression (4.3a) of the model, by observing that, for any $x \in C$, $\varphi + \mathrm{D}_{\hat{h}}(\,\cdot\,, x) = g + \frac{1}{\gamma}h - \hat{h}(x) - \langle\nabla\hat{h}(x),\ \cdot - x\rangle$. For the same reason, it then follows from Assumption 1.A4 that $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x, x^-) \subset C$.

4.2.(i) Apparent, by considering $w = x$ in (4.3b).

4.2.(ii) & 4.2.(iii) The first assertion owes to the fact that $\hat{h}$ is 1-coercive by Lemma 4.1 and that both $\hat{h}$ and $\nabla\hat{f}_\beta$ are continuous on C, so that for any compact set $\mathcal{V} \subset C \times C$ one has that

$$\lim_{\|w\|\to+\infty} \inf_{(x,x^-)\in\mathcal{V}} \mathcal{M}^{h\text{-FRB}}_{\gamma,\beta}(w; x, x^-) = +\infty,$$

as is apparent from (4.3a). In turn, the second assertion follows from [37, Thm. 1.17].

4.2.(iv) Follows from the optimality conditions of $\bar{x} \in \operatorname{argmin} \mathcal{M}^{h\text{-FRB}}_{\gamma,\beta}(\,\cdot\,; x, x^-)$, having $\bar{x} \in C$ by assertion 4.2.(iii) so that the calculus rule of [37, Ex. 8.8(c)] applies (having $\hat{h}$ smooth around $\bar{x} \in C$).

4.2.(v) That $0 \in \hat{\partial}\varphi(x)$ follows from assertion 4.2.(iv), and the other claim from [1, Lem. 3.6] by observing that $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x, x) = \operatorname{argmin}\{\varphi + \mathrm{D}_{\hat{h}}(\,\cdot\,, x)\}$ for any $\gamma > 0$ and $\beta \in \mathbb{R}$. □

Remark 4.3 (inertial effect) Letting $\tilde{f} = f \dot{+} ch$ and $\tilde{g} = g \dot{-} ch$ for some $c \in \mathbb{R}$, $\tilde{f} + \tilde{g}$ gives an alternative decomposition of φ which still complies with Assumption 1, having $\sigma_{\pm\tilde{f},h} = \sigma_{\pm f,h} \pm c$ by Lemma 3.11. Relative to this decomposition, for any stepsize $\tilde{\gamma}$ and inertial parameter $\tilde{\beta}$, the corresponding model $\tilde{\mathcal{M}}^{h\text{-FRB}}_{\gamma,\beta}$ is given by

$$\begin{aligned}\tilde{\mathcal{M}}^{h\text{-FRB}}_{\gamma,\beta}(w; x, x^-) &\overset{(4.3a)}{=} \varphi(w) + \mathrm{D}_{h/\tilde{\gamma}-\tilde{f}}(w, x) + \langle w - x, \nabla\big[\tilde{f} - \tfrac{\tilde{\beta}}{\tilde{\gamma}}h\big](x) \\ &\qquad - \nabla\big[\tilde{f} - \tfrac{\tilde{\beta}}{\tilde{\gamma}}h\big](x^-)\rangle \\ &= \varphi(w) + \mathrm{D}_{\frac{(1-\tilde{\gamma}c)}{\tilde{\gamma}}h-f}(w, x) \\ &\qquad + \langle w - x, \nabla\big[f - \tfrac{\tilde{\beta}-\tilde{\gamma}c}{\tilde{\gamma}}h\big](x) - \nabla\big[f - \tfrac{\tilde{\beta}-\tilde{\gamma}c}{\tilde{\gamma}}h\big](x^-)\rangle.\end{aligned}$$

Thus,

$$\tilde{\mathcal{M}}^{h\text{-FRB}}_{\gamma,\beta} = \mathcal{M}^{h\text{-FRB}}_{\gamma,\beta} \quad \text{with} \quad \begin{cases}\gamma = \frac{\tilde{\gamma}}{1-\tilde{\gamma}c} \\ \beta = \frac{\tilde{\beta}-\tilde{\gamma}c}{1-\tilde{\gamma}c}\end{cases} \quad \Leftrightarrow \quad \begin{cases}\tilde{\gamma} = \frac{\gamma}{1+\gamma c} \\ \tilde{\beta} = \frac{\beta+\gamma c}{1+\gamma c},\end{cases}$$

and in particular *i**FRB steps with the respective parameters coincide. The effect of inertia can then be explained as a redistribution of multiples of h among f and g in the problem formulation, having $\mathcal{M}^{h\text{-FRB}}_{\gamma,\beta} = \tilde{\mathcal{M}}^{h\text{-FRB}}_{\frac{\gamma}{1-\beta},0}$ for any $\gamma > 0$ and $\beta < 1$.

4.2 The *i**FRB-envelope

Having defined model $\mathscr{M}^{h\text{-FRB}}_{\gamma,\beta}$ and its solution mapping $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}$ resulted from parametric minimization, we now introduce the associated value function, which we name i^*FRB-envelope.

Definition 4.4 (i^*FRB-envelope) The envelope associated with i^*FRB with stepsize $\gamma < 1/[\sigma_{f,h}]_-$ and inertia $\beta \in \mathbb{R}$ is the function $\phi^{h\text{-FRB}}_{\gamma,\beta} : C \times C \to \mathbb{R}$ defined as

$$\phi^{h\text{-FRB}}_{\gamma,\beta}(x, x^-) := \inf_{w \in \operatorname{dom} h} \mathscr{M}^{h\text{-FRB}}_{\gamma,\beta}(w; x, x^-). \tag{4.5}$$

Lemma 4.5 *(basic properties of $\phi^{h\text{-FRB}}_{\gamma,\beta}$) Suppose that Assumption 1 holds. Then, for any $\gamma < 1/[\sigma_{f,h}]_-$ and $\beta \in \mathbb{R}$ the following hold:*

(i) $\phi^{h\text{-FRB}}_{\gamma,\beta}$ is (real-valued and) continuous on $C \times C$; in fact, it is locally Lipschitz provided that $f, h \in \mathscr{C}^2(C)$.
(ii) For any $x, x^- \in C$ and $\bar{x} \in \mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x, x^-)$

$$\begin{aligned}\phi^{h\text{-FRB}}_{\gamma,\beta}(x, x^-) &= \mathscr{M}^{h\text{-FRB}}_{\gamma,\beta}(\bar{x}; x, x^-)\\ &= \varphi(\bar{x}) + \mathrm{D}_{\hat{h}-\hat{f}_\beta}(\bar{x}, x) + \mathrm{D}_{\hat{f}_\beta}(\bar{x}, x^-) - \mathrm{D}_{\hat{f}_\beta}(x, x^-).\end{aligned}$$

(iii) $\phi^{h\text{-FRB}}_{\gamma,\beta}(x, x^-) \le \varphi(x)$ for any $x, x^- \in C$.

Proof 4.5.(i) In light of the uniform level boundedness asserted in Lemma 4.2.(ii), continuity of $\phi^{h\text{-FRB}}_{\gamma,\beta}$ follows from [37, Thm. 1.17(c)] by observing that the mapping $(x, x^-) \mapsto \mathscr{M}^{h\text{-FRB}}_{\gamma,\beta}(w; x, x^-)$ is continuous for every w; in fact, when f and h are both $\mathscr{C}^2$ on C, the gradient $\nabla_{(x,x^-)}\mathscr{M}^{h\text{-FRB}}_{\gamma,\beta}(w, x, x^-) = (\nabla \hat{f}_\beta(x^-) - \nabla \hat{f}_\beta(x) + \big(\nabla^2 \hat{f}_\beta - \nabla^2 \hat{h}\big)(x)(w - x),\ \nabla^2 \hat{f}_\beta(x^-)(x - w))$ exists and is continuous with respect to all its arguments, which together with local boundedness of $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}$, cf. Lemma 4.2.(iii), gives that $-\phi^{h\text{-FRB}}_{\gamma,\beta}$ is a lower-$\mathscr{C}^1$ function in the sense of [37, Def. 10.29], and in particular locally Lipschitz continuous by virtue of [37, Thm.s 10.31 and 9.2].

4.5.(ii) & 4.5.(iii) The identity follows by definition, cf. (4.5) and (4.4b). The inequality follows by considering $w = x$ in (4.5) and (4.3b). □

4.3 Establishing a Merit Function

We now work toward establishing a merit function for i^*FRB, starting from comparing the values of $\phi^{h\text{-FRB}}_{\gamma,\beta}(\bar{x}, x)$ and $\phi^{h\text{-FRB}}_{\gamma,\beta}(x, x^-)$, with $\bar{x} \in \mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x, x^-)$. Owing to Lemma 4.5.(iii), we have

$$\begin{aligned}\phi_{\gamma,\beta}^{h\text{-FRB}}(\bar{x}, x) &\leq \varphi(\bar{x}) = \varphi_{\overline{C}}(\bar{x}) \\ &= \phi_{\gamma,\beta}^{h\text{-FRB}}(x, x^-) - \mathrm{D}_{\hat{h}-\hat{f}_\beta}(\bar{x}, x) - \overbrace{\mathrm{D}_{\hat{f}_\beta}(\bar{x}, x^-)}^{\text{complicating term}} + \mathrm{D}_{\hat{f}_\beta}(x, x^-).\end{aligned} \tag{4.6}$$

From here two separate cases can be considered, each yielding surprisingly different results. The watershed lies in whether the "complicating" term is positive or not: one case will result in a very straightforward convergence analysis in the full generality of Assumption 1, while the other will necessitate an additional Lipschitz differentiability requirement. The convergence analysis in both cases revolves around the identification of a constant $c > 0$ determining a lower bounded merit function

$$\mathcal{L}_{\gamma,\beta}^{h\text{-FRB}} = \phi_{\gamma,\beta}^{h\text{-FRB}} + \tfrac{c}{2\gamma}\,\mathrm{D}_h + \mathrm{D}_\xi\,. \tag{4.7}$$

The difference between the two cases is determined by function ξ appearing in the last Bregman operator D_ξ, having $\xi = \hat{f}_\beta$ in the former case and $\xi = L_{\hat{f}_\beta}\jmath$ in the latter, where $L_{\hat{f}_\beta}$ is a Lipschitz constant for $\nabla \hat{f}_\beta$ and we remind that

$$\jmath := \tfrac{1}{2}\|\cdot\|^2 \tag{4.8}$$

is the squared Euclidean norm. The two cases are stated in the next theorem, which constitutes the main result of this section. Special and worst-case scenarios leading to simplified statements will be given in Sect. 4.4. In what follows, patterning the normalization of $\sigma_{\pm f,h}$ into $p_{\pm f,h}$ detailed in Sect. 3, we also introduce the scaled stepsize

$$\alpha := \gamma L_{f,h}, \tag{4.9}$$

which as a result of the convergence analysis will be confined in the interval (0, 1).

Theorem 4.6 *Let α be given by (4.9). Suppose that Assumption 1 holds and consider one of the following scenarios:*

(A) either $\hat{f}_\beta$ is convex (e.g., when $\alpha p_{f,h} - \beta \geq 0$) and $\beta > -(1 + 3\alpha p_{-f,h})/2$, in which case

$$\xi := \hat{f}_\beta \quad \textit{and} \quad c := 1 + 2\beta + 3\alpha p_{-f,h} > 0,$$

(B) or $\hat{f}_\beta$ is $L_{\hat{f}_\beta}$-Lipschitz differentiable, h is σ_h-strongly convex, and

$$c := (1 + \alpha p_{-f,h}) - \frac{2\gamma L_{\hat{f}_\beta}}{\sigma_h} > 0,$$

in which case $\xi := L_{\hat{f}_\beta}\jmath$.

Then, for $\mathcal{L}_{\gamma,\beta}^{h\text{-FRB}}$ as in (4.7) the following assertions hold:

(i) For every $x, x^- \in C$ and $\bar{x} \in \mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x, x^-)$,

$$\mathcal{L}^{h\text{-FRB}}_{\gamma,\beta}(\bar{x}, x) \leq \mathcal{L}^{h\text{-FRB}}_{\gamma,\beta}(x, x^-) - \tfrac{c}{2\gamma}\,\mathrm{D}_h(\bar{x}, x) - \tfrac{c}{2\gamma}\,\mathrm{D}_h(x, x^-) \tag{4.10a}$$

and

$$\varphi_{\overline{C}}(\bar{x}) \leq \mathcal{L}^{h\text{-FRB}}_{\gamma,\beta}(x, x^-) - \tfrac{c}{\gamma}\,\mathrm{D}_h(\bar{x}, x) - \tfrac{c}{2\gamma}\,\mathrm{D}_h(x, x^-). \tag{4.10b}$$

(ii) $\inf \mathcal{L}^{h\text{-FRB}}_{\gamma,\beta} = \inf_C \varphi$.

(iii) If either h is strongly convex or $\operatorname{dom} h = \mathbb{R}^n$*, then $\mathcal{L}^{h\text{-FRB}}_{\gamma,\beta}$ is level bounded provided that $\varphi_{\overline{C}}$ is.*

The proof of this result is detailed in the dedicated Appendix A; before that, let us draw some comments. As clarified in the statement of Theorem 4.6.(A), convexity of $\hat{f}_\beta$ can be enforced by suitably choosing γ and β without imposing additional requirements on the problem. However, an unusual yet reasonable condition on inertial parameter β may be necessary.

Remark 4.7 In order to furnish Theorem 4.6.(A), we shall see soon that $\beta \leq 0$ may be required; see Sect. 4.4. Such assumption, although more pessimistic, coincides with a recent conjecture by Dragomir et al. [18, §4.5.3], which states that inertial methods with nonadaptive coefficients fail to converge in the relative smoothness setting, and provides an alternative perspective to the same matter through the lens of the convexity of $\hat{f}_\beta$.

Unlike Theorem 4.6.(A), however, additional assumptions are needed for the Lipschitz differentiable case of Theorem 4.6.(B). This is because the requirement is equivalent to smoothness relative to the Euclidean Bregman kernel $\jmath$, while Assumption 1 prescribes bounds only relative to h.

Remark 4.8 Under Assumption 1, one has that $\hat{f}_\beta$ is Lipschitz differentiable with modulus $L_{\hat{f}_\beta}$ under either one of the following conditions:

(B1) either ∇h is L_h-Lipschitz, and $L_{\hat{f}_\beta} = \frac{L_h}{\gamma}\max\{\beta - \alpha p_{f,h}, -\beta - \alpha p_{-f,h}\}$,
(B2) or $\beta = 0$ and ∇f is L_f-Lipschitz, in which case $L_{\hat{f}_\beta} = L_f$.

Recalling that $\hat{f}_\beta = f - \frac{\beta}{\gamma}h$, the second condition is tautological. In case ∇h is L_h-Lipschitz, the claim follows from (4.2) together with Corollary 3.10.

4.4 Simplified Bounds

In this section, we provide bounds that only discern whether f is convex, concave, or neither of the above. As discussed in Remark 3.5, these cases can be recovered by suitable combinations of the coefficients $p_{\pm f,h} \in \{0, \pm 1\}$ and thus lead to easier, though possibly looser, bounds compared to those in Theorem 4.6. We will also avail ourselves of the estimates of $L_{\hat{f}_\beta}$ in Remark 4.8 to discuss the cases in which $\hat{f}_\beta$ is

Lipschitz differentiable. To simplify the exposition, we may provide smaller estimates of the coefficient c in Theorem 4.6, owing to the fact that replacing c with any $c' \in (0, c]$ does not affect the validity of the statement and only causes the inequalities (4.10) to be possibly looser.

Without distinguishing between upper and lower relative bounds, whenever f is $L_{f,h}$-smooth relative to h as in Assumption 1 one can consider $\sigma_{\pm f,h} = -L_{f,h}$ or, equivalently, $p_{f,h} = p_{-f,h} = -1$. Plugging these values into Theorem 4.6 yields the following.

Corollary 4.9 *(worst-case bounds) Suppose that Assumption 1 holds. All the claims of Theorem 4.6 hold when $\gamma > 0$, $\beta \in \mathbb{R}$ and $c > 0$ are such that*

(A) either $-1/2 < \beta < 0$ and $\gamma \leq (1/L_{f,h}) \min\{-\beta, (1 + 2\beta - c)/3\}$, in which case $\xi = \hat{f}_\beta$;

(B1) or h is σ_h-strongly convex and L_h-Lipschitz differentiable, $|\beta| < \sigma_h/2L_h$ and $\gamma \leq (1/L_{f,h})[(\sigma_h(1 - c) - 2L_h|\beta|)/(\sigma_h + 2L_h)]$, in which case $\xi = (L_h/\gamma)(\alpha + |\beta|)\jmath$;

(B2) or h is σ_h-strongly convex, ∇f is L_f-Lipschitz continuous, $\beta = 0$ and $\gamma \leq \sigma_h(1 - c)/(\sigma_h L_{f,h} + 2L_{f,h})$, in which case $\xi = L_f \jmath$.

Proof Setting $p_{\pm f,h} = -1$ in Theorem 4.6, one has:

4.9.(A) The bounds in the statement of Theorem 4.6.(A) read $0 < c = 1 + 2\beta - 3\alpha$ and $\beta \leq -\alpha$. Expressed in terms of $\alpha = \gamma L_{f,h}$, the claimed bounds on γ are obtained. In turn, imposing $\alpha > 0$ results in the claimed bounds on β.

4.9.(B1) & 4.9.(B2) The two subcases refer to the corresponding items in Remark 4.8. We shall show only the first one, as the second one is a trivial adaptation after observing that $L_f = L_{f,h} L_h$ by virtue of Corollary 3.10. The value of $L_{\hat{f}_\beta}$ as in Remark (B1) reduces to $L_{\hat{f}_\beta} = (L_h/\gamma)(\alpha + |\beta|)$. Plugged into Theorem 4.6.(B) yields $0 < c = 1 - \alpha - 2(\alpha + |\beta|)L_h/\sigma_h$, implying that $\gamma = \alpha/L_{f,h}$ is bounded as in assertion 4.9.(B1). Imposing $\alpha > 0$ yields also the claimed bounds on β. □

When f is convex on C, $\sigma_{f,h} = 0$ can be considered resulting in $p_{f,h} = 0$ and $p_{-f,h} = -1$.[4]

Corollary 4.10 *(bounds when f is convex) Suppose that Assumption 1 holds and that f is convex. All the claims of Theorem 4.6 remain valid if $\gamma > 0$, $\beta \in \mathbb{R}$ and $c > 0$ are such that*

(A) either $-1/2 < \beta \leq 0$ and $\gamma \leq (1/L_{f,h})[(1 + 2\beta - c)/3]$, in which case $\xi = \hat{f}_\beta$;

(B1) or h is σ_h-strongly convex and L_h-Lipschitz differentiable,

$$|\beta| < \frac{\sigma_h}{2L_h} \quad \textit{and} \quad \gamma \leq \frac{1}{L_{f,h}} \min\{\frac{\sigma_h(1-c)+2L_h\beta}{\sigma_h+2L_h}, \frac{\sigma_h(1-c)-2L_h\beta}{\sigma_h}\},$$

in which case $\xi = (L_h/\gamma)\max\{\beta, \alpha - \beta\}\jmath$;

[4] This also covers the case in which f is affine on C, although a tighter $p_{-f,h} = 0$ could be considered in this case and improve the range to $\beta \in (-1/2, 0]$ and any $\gamma > 0$.

(B2) or h is σ_h-strongly convex, ∇f is L_f-Lipschitz, $\beta = 0$, and $\gamma \leq (1-c)\sigma_h/(\sigma_h L_{f,h} + 2L_f)$, in which case $\xi = L_f \jmath$.

Proof We will pattern the proof of Corollary 4.9, and omit the proof of assertion 4.10.(B2) which is an easy adaptation of that of assertion 4.10.(B1). Setting $p_{f,h} = 0$ and $p_{-f,h} = -1$ in Theorem 4.6, one has:

4.10.(A) The bounds in the statement of Theorem 4.6.(A) read $0 < c = 1+2\beta-3\alpha$ and $\beta \leq 0$. This readily yields the bound $\gamma L_{f,h} = \alpha \leq \frac{1+2\beta-c}{3}$ after replacing $c = 1+2\beta-3\alpha$ by $c \leq 1+2\beta-3\alpha$ with an abuse of notation on c, under which inequalities in the desired Theorem 4.6.(i) hold with possibly looser bounds. In turn, the condition $\alpha > 0$ then constrains $\beta \in (-1/2, 0]$, as claimed.

4.10.(B1) The value of $L_{\hat{f}_\beta}$ in Remark (B1) reduces to $\frac{L_h}{\gamma}\max\{\beta, \alpha-\beta\}$. Plugged into Theorem 4.6.(B) yields $0 < c = (1-\alpha) - 2L_h/\sigma_h \max\{\beta, \alpha-\beta\}$, and in particular

$$c \leq (1-\alpha) - 2\beta L_h/\sigma_h,$$
$$c \leq (1-\alpha) - 2(\alpha-\beta)L_h/\sigma_h.$$

In terms of $\gamma = \alpha/L_{f,h}$, this results in the bound for γ as in assertion 4.10.(B1). In turn, imposing $\alpha > 0$ results in the claimed bounds on β. □

Similarly, when f is concave (that is, $-f$ is convex) on C, then $\sigma_{-f,h} = 0$ can be considered, resulting in $p_{f,h} = -1$ and $p_{-f,h} = 0$.

Corollary 4.11 *(bounds when f is concave) Suppose that Assumption 1 holds and that f is concave. All the claims of Theorem 4.6 remain valid if $\gamma > 0$, $\beta \in \mathbb{R}$ and $c > 0$ are such that*

(A) either $(c-1)/2 \leq \beta < 0$ and $\gamma \leq -\beta/L_{f,h}$, in which case $\xi = \hat{f}_\beta$;
(B1) or h is σ_h-strongly convex and L_h-Lipschitz differentiable,

$$-\frac{(1-c)\sigma_h}{2L_h} \leq \beta < \frac{\sigma_h}{2L_h} \quad \textit{and} \quad \gamma \leq \frac{1}{L_{f,h}}\frac{\sigma_h(1-c)-2L_h\beta}{2L_h},$$

in which case $\xi = \frac{L_h}{\gamma}\max\{\alpha+\beta, -\beta\}\jmath$;
(B2) or h is σ_h-strongly convex, f is L_f-Lipschitz differentiable, $\beta = 0$ and $\gamma \leq \sigma_h(1-c)/(2L_f)$, in which case $\xi = L_f \jmath$.

Proof Set $p_{f,h} = -1$ and $p_{-f,h} = 0$ in Theorem 4.6. A similar argument as in Corollary 4.10 completes the proof:

4.11.(A) From Theorem 4.6.(A), we obtain $-\alpha-\beta \geq 0$ and $c = 1+2\beta > 0$. Recalling that $\alpha = \gamma L_{f,h}$ must be strictly positive, the bound on γ and on β as in the statement are obtained.

4.11.(B1) Remark (B1) yields the estimate $L_{\hat{f}_\beta} = L_h/\gamma \max\{\alpha+\beta, -\beta\}$, which plugged into the statement of Theorem 4.6.(B) gives $0 < c = 1 - 2L_h/\sigma_h \max\{\alpha+\beta, -\beta\}$. Therefore,

$$c \leq 1 + 2\beta L_h/\sigma_h,$$

$$c \leq 1 - 2(\alpha + \beta)L_h/\sigma_h,$$

which in terms of $\gamma = \alpha/L_{f,h}$ results in the bound on γ and lower bound on β as in assertion 4.11.(B1). The upper bound on β follows from ensuring $\sigma_h - 2L_h\beta > 0$, which is necessary for the bound $\gamma > 0$.

Once again, the case 4.11.(B2) is an easy adaptation of 4.11.(B1). □

5 Convergence Analysis

In this section, we study the behavior of sequences generated by i^*FRB. Although some basic convergence results can be derived in the full generality of Assumption 1, establishing local optimality guarantees of the limit point(s) will ultimately require an additional full domain assumption.

Assumption 2 Function h has full domain, that is, $C = \mathbb{R}^n$.

Assumption 2 is standard for nonconvex splitting algorithms in a relative smooth setting. To the best of our knowledge, the question regarding whether this requirement can be removed remains open; see, e.g., [39] and the references therein.

5.1 Function Value Convergence

We begin with the convergence of merit function value.

Theorem 5.1 *(function value convergence of i^*FRB) Let $(x^k)_{k\in\mathbb{N}}$ be a sequence generated by i^*FRB (Algorithm 1) in the setting of Theorem 4.6. Then,*

(i) It holds that

$$\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}\left(x^{k+1}, x^k\right) \leq \mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}\left(x^k, x^{k-1}\right) - \tfrac{c}{2\gamma}\,\mathrm{D}_h\left(x^{k+1}, x^k\right) - \tfrac{c}{2\gamma}\,\mathrm{D}_h\left(x^k, x^{k-1}\right). \quad (5.1)$$

In particular, $\sum_{k=0}^{\infty} \mathrm{D}_h\left(x^k, x^{k-1}\right) < +\infty$ and $\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}(x^k, x^{k-1}) \to \varphi^\star$ as $k \to +\infty$ for some $\varphi^\star \geq \inf \varphi_{\overline{C}}$.

If Assumption 2 also holds, then:

(ii) If $\varphi_{\overline{C}}$ is level bounded, then $(x^k)_{k\in\mathbb{N}}$ is bounded.
(iii) Let Ω be the set of limit points of $(x^k)_{k\in\mathbb{N}}$. Then, φ is constant on Ω with value $\varphi^\star$, and for every $x^\star \in \Omega$ it holds that $x^\star \in \mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x^\star, x^\star)$ and $0 \in \hat{\partial}\varphi(x^\star)$.

Proof 5.1.(i) Recall from Theorem 4.6 that the inequality (5.1) holds and that $\inf \mathscr{L}^{h\text{-FRB}}_{\gamma,\beta} = \inf \varphi_{\overline{C}} > -\infty$, from which convergence of $(\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}(x^k, x^{k-1}))_{k\in\mathbb{N}}$ readily follows. In turn, telescoping (5.1) shows that $\sum_{k\in\mathbb{N}} \mathrm{D}_h(x^k, x^{k-1})$ is finite.

5.1.(ii) From Theorem 5.1.(i), $\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}\left(x^{k+1}, x^k\right) \leq \mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}(x^0, x^{-1})$ holds for every k. Then boundedness of $(x^k)_{k\in\mathbb{N}}$ is implied by level boundedness of $\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}$; see Theorem 4.6.(iii).

5.1.(iii) Suppose that a subsequence $(x^{k_j})_{j\in\mathbb{N}}$ converges to a point $x^\star$, then so do the subsequences $(x^{k_j\pm1})_{j\in\mathbb{N}}$ by Theorem 5.1.(i) and [8, Prop. 2.2(iii)].[5] Since $x^{k_j+1} \in \mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x^{k_j}, x^{k_j-1})$, by passing to the limit, osc of $\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}$ (Lemma 4.2.(iii)) implies that $x^\star \in \mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x^\star, x^\star)$. Invoking Lemma 4.2.(v) yields the stationarity condition $0 \in \hat{\partial}\varphi(x^\star)$. Moreover, by continuity of $\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}$ one has

$$\varphi^\star \overset{\text{(def)}}{=} \lim_{k\to+\infty} \mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}\left(x^k, x^{k-1}\right) = \mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}(x^\star, x^\star) = \varphi(x^\star),$$

where the last equality follows from Lemma 4.5.(ii), owing to the inclusion $x^\star \in \mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x^\star, x^\star)$ (and the fact that $\mathrm{D}_\psi(x, x) = 0$ for any differentiable function ψ). From the arbitrarity of $x^\star \in \Omega$, we conclude that $\varphi \equiv \varphi^\star$ on Ω. □

The full domain assumption on h in Theorem 5.1.(ii) is stronger than necessary, but suffices to our purposes. The proof invokes the level bounedness of $\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}$ via Theorem 4.6.(iii), which hinges on the property that whenever $(x_k)_{k\in\mathbb{N}} \subset C$ is bounded and $(x_k^-)_{k\in\mathbb{N}} \subset C$ is unbounded, $(\mathrm{D}_h(x_k, x_k^-))_{k\in\mathbb{N}}$ too is unbounded. As such, Theorem 5.1.(ii) remains valid for any h, possibly without full domain, as long as the induced Bregman distance $\mathrm{D}_h(w, x)$ is level bounded in x locally uniformly in w.

Remark 5.2 (i^*FRB as a globalization framework) The "sufficient" decrease property over the merit function $\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}$ assessed in Theorem 5.1.(i) together with the continuity of $\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}$ makes i^*FRB a suitable candidate for the continuous-Lyapunov descent (CLyD) framework [40, §4], enabling the globalization of fast local methods $x^+ = x + d$ by using only i^*FRB operations, with no change of metrics. Indeed, because of continuity, not only is $\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}$ smaller than $\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}(x^k, x^{k-1})$ at (x^{k+1}, x^k), but also at sufficiently close points. This means that the i^*FRB update can be replaced by $(1-\tau_k)(x^{k+1}, x^k) + \tau_k(x^k + d^k, x^{k-1} + d_-^k)$, where (d^k, d_-^k) is the sought update direction at the current iterate pair (x^k, x^{k-1}) and τ_k is a stepsize to be backtracked until a sufficient decrease on $\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}$ is achieved. Under assumptions, suitable choices of (d^k, d_-^k) can yield fast asymptotic rates. We refer the interested reader to the analysis of the Bella algorithm [1, Alg. 5.1], based on Bregman proximal gradient but otherwise very closely related.

It is now possible to demonstrate the necessity of some of the bounds on the stepsize that were discussed in Sect. 4.4, by showing that $\mathrm{D}_h(x^{k+1}, x^k)$ may otherwise fail to vanish. Note that, for $\beta = 0$, the following counterexample constitutes a tightness certificate for the bound $\gamma < 1/3L_f$ derived in [47] in the noninertial Euclidean case.

Example 5.3 The bound $\alpha = \gamma L_{f,h} < (1+2\beta)/3$ is tight even in the Euclidean case. To see this, consider $g = \delta_{\{\pm1\}}$ and for a fixed $L > 0$ let $f(x) = Lh(x) = \frac{L}{2}x^2$. Then, one has $L_{f,h} = \sigma_{f,h} = L$ and $\sigma_{-f,h} = -L$. For $\gamma < 1/L = 1/[\sigma_{-f,h}]_-$, it is easy to see that

$$\mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x, x^-) = -\operatorname{sgn}\bigl(\nabla\hat{f}_\beta(x) - \nabla\hat{f}_\beta(x^-) - \nabla\hat{h}(x)\bigr)$$

[5] [8, Prop. 2.2(iii)] is applicable due to the 1-coercivity assumption on h (recall Sect. 2) and [9, Prop. 14.15].

$$= \operatorname{sgn}\big((1-2\alpha+\beta)x + (\alpha-\beta)x^-\big)$$

(with $\operatorname{sgn} 0 := \{\pm 1\}$), where the first equality follows from (4.3a) and (4.4b). Let $x^{-1} = -1$, $x^0 = 1$. Suppose that $\alpha \geq (1+2\beta)/3$, then $((-1)^k)_{k\in\mathbb{N}}$ is a sequence generated by i^*FRB for which $\mathrm{D}_h(x^{k+1}, x^k) \equiv 2 \not\to 0$.

As a consequence of Theorem 5.1.(i), the condition $\mathrm{D}_h(x^{k+1}, x^k) \leq \varepsilon$ is satisfied in finitely many iterations for any tolerance $\varepsilon > 0$. While this could be used as termination criterion, in the generality of Assumptions 1 and 2 there is no guarantee on the relaxed stationarity measure $\operatorname{dist}(0, \hat{\partial}\varphi(x^{k+1}))$, which through Lemma 4.2.(iv) can only be estimated as

$$\begin{aligned} &\operatorname{dist}(0, \hat{\partial}\varphi(x^{k+1})) \leq \|v^{k+1}\| \\ &\text{with } v^{k+1} := \nabla\hat{h}(x^k) - \nabla\hat{h}(x^{k+1}) - \nabla\hat{f}_\beta(x^k) + \nabla\hat{f}_\beta(x^{k-1}), \end{aligned} \tag{5.2}$$

where $\hat{h}$ and $\hat{f}_\beta$ are as in (4.1). On the other hand, in accounting for possibly unbounded sequences, additional assumptions are needed for the condition $\|v^{k+1}\| \leq \varepsilon$ to be met in finitely many iterations. One such is the so-called *uniform smoothness* of h, which by [4, Thm. 3.8(1)–(2)] can be defined in terms of an inequality

$$\mathrm{D}_h(x, y) \leq \varrho(\|x-y\|) \tag{5.3}$$

holding for every $x, y \in \mathbb{R}^n$, where $\varrho : \mathbb{R}_+ \to \mathbb{R}_+$ is a nondecreasing function such that $\rho(0) = 0$ and $\rho(s)/s \to 0$ as $s \searrow 0$. As shown in [4, Thm. 3.8(1)–(2)], the dual counterpart is given by *uniform convexity*, which amounts to h^* being uniformly smooth.

Lemma 5.4 *(termination criteria) Suppose that Assumption 2 holds, and let $(x^k)_{k\in\mathbb{N}}$ be a sequence generated by i^*FRB (Algorithm 1) in the setting of Theorem 4.6. If*

(A) either φ is level bounded,
(B) or h^* is uniformly convex (equivalently, h is uniformly smooth),

then, for v^{k+1} as in (5.2) it holds that $v^{k+1} \to 0$. Thus, for any $\varepsilon > 0$ the condition $\|v^{k+1}\| \leq \varepsilon$ is satisfied for all k large enough and guarantees $\operatorname{dist}(0, \hat{\partial}\varphi(x^{k+1})) \leq \varepsilon$.

Proof The implication of $\|v^{k+1}\| \leq \varepsilon$ and ε-stationarity of x^{k+1} has already been discussed. If φ is level bounded, then Theorem 5.1 implies that $(x^k)_{k\in\mathbb{N}}$ is contained in a compact subset of $C = \mathbb{R}^n$. Recall from Theorem 5.1.(i) that $\mathrm{D}_h(x^{k+1}, x^k) \to 0$, which implies through [7, Ex. 4.10(ii)], Assumption 2, and the boundedness of $(x^k)_{k\in\mathbb{N}}$ that $x^{k+1} - x^k \to 0$. In turn, $v^{k+1} \to 0$ holds by uniform continuity of $\nabla\hat{h}$ and $\nabla\hat{f}_\beta$ on the aforementioned compact set. In case h^* is uniformly convex, this being equivalent to uniform smoothness of h by [4, Thm. 3.8(1)–(2)], the vanishing of $\mathrm{D}_{h^*}(\nabla h(x^k), \nabla h(x^{k+1})) = \mathrm{D}_h(x^{k+1}, x^k)$ implies through [36, Prop. 4.13(IV)] that $\|\nabla h(x^k) - \nabla h(x^{k+1})\| \to 0$. Since $\mathrm{D}_{L_{f,h}h+f} \leq 2L_{f,h}\mathrm{D}_h$ holds by convexity of $L_{f,h}h - f$, from the characterization (5.3) it is apparent that $L_{f,h}h + f$ too is uniformly smooth. Arguing as above, the vanishing of $\mathrm{D}_{L_{f,h}h+f}(x^{k+1}, x^k)$ implies that of $\|\nabla[L_{f,h}h + f](x^{k+1}) - \nabla[L_{f,h}h + f](x^k)\|$. Note that

$$\hat{f}_\beta \overset{\text{(def)}}{=} f - \frac{\beta}{\gamma}h = L_{f,h}h + f - \left[L_{f,h} + \frac{\beta}{\gamma}\right]h \quad \text{and} \quad \hat{h} \overset{\text{(def)}}{=} \frac{1}{\gamma}h - f = \frac{1-\beta}{\gamma}h - \hat{f}_\beta.$$

Then the vanishing of $\|\nabla[L_{f,h}h + f](x^{k+1}) - \nabla[L_{f,h}h + f](x^k)\|$ and $\|\nabla h(x^k) - \nabla h(x^{k+1})\|$ implies that

$$v^{k+1} \overset{\text{(def)}}{=} \nabla\hat{h}(x^k) - \nabla\hat{h}(x^{k+1}) - \nabla\hat{f}_\beta(x^k) + \nabla\hat{f}_\beta(x^{k-1}) \to 0,$$

as desired. □

5.2 Global Convergence

In this subsection, we work toward the global sequential convergence of i^*FRB. To this end, we introduce a key concept which will be useful soon. For $\eta \in (0, +\infty]$, denote by Ψ_η the class of functions $\psi : [0, \eta) \to \mathbb{R}_+$ satisfying the following: (i) $\psi(t)$ is right-continuous at $t = 0$ with $\psi(0) = 0$; (ii) ψ is strictly increasing on $[0, \eta)$; (iii) ψ is continuously differentiable on $(0, \eta)$.

Definition 5.5 ([2, Def. 3.1]) Let $F : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc, and let ∂F be its Mordukhovich limiting subdifferential. We say that F has the Kurdyka–Łojasiewicz (KL) property at $\bar{x} \in \operatorname{dom} \partial F$, if there exist a neighborhood $U \ni \bar{x}$, $\eta \in (0, +\infty]$ and a concave $\psi \in \Psi_\eta$, such that for all $x \in U \cap [0 < F - F(\bar{x}) < \eta]$,

$$\psi'\big(F(x) - F(\bar{x})\big) \cdot \operatorname{dist}\big(0, \partial F(x)\big) \geq 1.$$

Moreover, F is a KL function if it has the KL property at every $x \in \operatorname{dom} \partial F$.

Now we present our main result on global convergence. As the proof is standard, we defer it to Appendix B for the sake of completeness.

Theorem 5.6 *(sequential convergence of i^*FRB) Suppose that Assumption 2 holds, and let $(x^k)_{k\in\mathbb{N}}$ be a sequence generated by i^*FRB (Algorithm 1) in the setting of Theorem 4.6. Assume in addition the following:*

A1 φ is level bounded.
A2 f, h are twice continuously differentiable and $\nabla^2 h$ is positive definite everywhere.
A3 φ, h are semialgebraic functions (see, e.g., [2, §4.3]).

Then $\sum_{k=0}^\infty \|x^{k+1} - x^k\| < +\infty$ and there exists $x^\star$ with $0 \in \hat{\partial}\varphi(x^\star)$ such that $x^k \to x^\star$ as $k \to +\infty$.

Remark 5.7 We note that a sharp estimation on $\sum_{k=0}^\infty \|x^{k+1} - x^k\|$ can be obtained by replacing Assumption 5.6.A3 in Theorem 5.6 with the notion introduced in [46].

Remark 5.8 Compared to the Lipschitz smooth case considered in [47], the twice continuous differentiability assumption in Theorem 5.6 is a technicality for finding an upper bound on $\|(u^k, v^k)\|$, which consists of difference of gradients, as a multiple of $\|x^k - x^{k-1}\|$; see also [1, Thm. 5.7] for a similar assumption. We delay its relaxation for future research.

5.3 Convergence Rates

Having established convergence of i^*FRB, we now turn to its rate. Recall that a function is said to have KL exponent $\theta \in [0, 1)$ if it satisfies the KL property (recall Definition 5.5) and there exists a desingularizing function of the form $\psi(t) = ct^{1-\theta}$ for some $c > 0$.

Theorem 5.9 *(function value and sequential convergence rate) Suppose that all the assumptions in Theorem 5.6 are satisfied, and follow the notation therein. Define* $(\forall k \in \mathbb{N})$ $e_k = \mathcal{L}^{h\text{-FRB}}_{\gamma,\beta}(x^{k+1}, x^k) - \varphi^\star$ *and*

$$\mathcal{F}^{h\text{-FRB}}_{\gamma,\beta}(\omega, x, x^-) = \mathcal{M}^{h\text{-FRB}}_{\gamma,\beta}(\omega; x, x^-) + \tfrac{c}{2\gamma}\,\mathrm{D}_h(x, x^-) + D_\xi(x, x^-)$$

for all $\omega, x, x^- \in \mathbb{R}^n$. *Assume in addition that* $\mathcal{F}^{h\text{-FRB}}_{\gamma,\beta}$ *has KL exponent* $\theta \in [0, 1)$ *at* $(x^\star, x^\star, x^\star)$. *Then the following hold:*

(i) If $\theta = 0$, *then* $e_k \to 0$ *and* $x^k \to x^\star$ *after finite steps.*

(ii) If $\theta \in (0, 1/2]$, *then there exist* $c_1, \hat{c}_1 > 0$ *and* $Q_1, \hat{Q}_1 \in [0, 1)$ *such that for k sufficiently large,*

$$e_k \le \hat{c}_1 \hat{Q}_1^k \text{ and } \|x^k - x^\star\| \le c_1 Q_1^k.$$

(iii) If $\theta \in (1/2, 1)$, *then there exist* $c_2, \hat{c}_2 > 0$ *such that for k sufficiently large,*

$$e_k \le \hat{c}_2 k^{-\frac{1}{2\theta-1}} \text{ and } \|x^k - x^\star\| \le c_2 k^{-\frac{1-\theta}{2\theta-1}}.$$

Proof See Appendix C. □

6 Conclusions

This work contributes a mirror inertial forward–reflected–backward splitting algorithm (i^*FRB), extending the forward–reflected–backward method proposed in [29] to the nonconvex and relative smooth setting. We have shown that the proposed algorithms enjoy pleasant properties akin to other splitting methods in the same setting. However, our methodology deviates from tradition through the i^*FRB-envelope, an envelope function defined on a product space that takes inertial terms into account, which, to the best of our knowledge, is the first of its kind and thus could be instrumental for future research. This approach also requires the inertial parameter to be negative, which coincides with a recent result [18] regarding the impossibility of accelerated non-Euclidean algorithms under relative smoothness. Thus, it is interesting to see whether an explicit example can be constructed to prove the sharpness of such restrictive assumption. It is also worth applying our technique to other two-stage splitting methods, such as Tseng's method, to obtain similar extensions.

Acknowledgements The authors are deeply thankful to the anonymous reviewers for their thorough reading and many constructive comments that significantly improved the quality and rigor of the manuscript.

A Proof of Theorem 4.6

Throughout this appendix, we remind that the adoption of extended arithmetics is not necessary, since, as a result of Lemma 4.2.(iii), all variables are confined in the open set C onto which both h and f (and consequently $\hat{h}$ and $\hat{f}_\beta$ as well, for any $\beta \in \mathbb{R}$) are finite-valued.

A.1 Proof of Theorem 4.6.(i) and 4.6.(ii)

We begin by proving a technical lemma in the setting of Theorem 4.6.(A).

Lemma A.1 *Suppose that Assumption 1 holds and let $\gamma > 0$ and $\beta \in \mathbb{R}$ be such that $\hat{f}_\beta = f \dot{-} \frac{\beta}{\gamma} h + \delta_{\operatorname{dom} h}$ is a convex function. Then, for every $x, x^- \in C$ and $\bar{x} \in \mathrm{T}_{\gamma,\beta}^{h\text{-FRB}}(x, x^-)$*

$$\left(\phi_{\gamma,\beta}^{h\text{-FRB}} + \mathrm{D}_{\hat{f}_\beta}\right)(\bar{x}, x) \leq \left(\phi_{\gamma,\beta}^{h\text{-FRB}} + \mathrm{D}_{\hat{f}_\beta}\right)(x, x^-) - \mathrm{D}_{\hat{h}-2\hat{f}_\beta}(\bar{x}, x). \tag{A.1}$$

Proof The claimed inequality follows from (4.6) together with the fact that $\mathrm{D}_{\hat{f}_\beta} \geq 0$. □

In the setting of Theorem 4.6.(A), recall that we set $c := 1 + 2\beta + 3\alpha p_{-f,h} > 0$. Then, inequality (A.1) can equivalently be written in terms of $\mathscr{L}_{\gamma,\beta}^{h\text{-FRB}}$ as

$$\begin{aligned} \mathscr{L}_{\gamma,\beta}^{h\text{-FRB}}(\bar{x}, x) &\leq \mathscr{L}_{\gamma,\beta}^{h\text{-FRB}}(x, x^-) - \mathrm{D}_{\hat{h}-2\hat{f}_\beta-\frac{c}{\gamma}h}(\bar{x}, x) - \tfrac{c}{2\gamma}\mathrm{D}_h(\bar{x}, x) - \tfrac{c}{2\gamma}\mathrm{D}_h(x, x^-) \\ &\leq \mathscr{L}_{\gamma,\beta}^{h\text{-FRB}}(x, x^-) - \tfrac{c}{2\gamma}\mathrm{D}_h(\bar{x}, x) - \tfrac{c}{2\gamma}\mathrm{D}_h(x, x^-), \end{aligned}$$

where the second inequality owes to the fact that $\mathrm{D}_{\hat{h}-2\hat{f}_\beta-\frac{c}{\gamma}h} \geq 0$, since $\hat{h} - 2\hat{f}_\beta - \frac{c}{\gamma}h$ is convex, having

$$\hat{h} - 2\hat{f}_\beta - \tfrac{c}{\gamma}h = \tfrac{1+2\beta-c}{\gamma}h - 3f = \overbrace{\tfrac{1+2\beta+3\alpha p_{-f,h}-c}{\gamma}}^{=0}h + 3(-f - \sigma_{-f,h}h),$$

the coefficient of h being null by definition of c, and $-f - \sigma_{-f,h}h$ being convex by definition of the relative weak convexity modulus $\sigma_{-f,h}$, cf. Definition 3.3. This proves (4.10a); inequality (4.10b) follows similarly by observing that

$$0 \le \mathrm{D}_{\hat{h}-2\hat{f}_\beta-\frac{c}{\gamma}h} = \mathrm{D}_{\hat{h}-\hat{f}_\beta} - \mathrm{D}_{\hat{f}_\beta} - \tfrac{c}{\gamma}\,\mathrm{D} \le \mathrm{D}_{\hat{h}-\hat{f}_\beta} - \tfrac{c}{\gamma}\,\mathrm{D}_h,$$

so that

$$\mathcal{L}^{h\text{-FRB}}_{\gamma,\beta}(x,x^-) \overset{4.5.\text{(ii)}}{=} \varphi(\bar{x}) + \overbrace{\mathrm{D}_{\hat{h}-\hat{f}_\beta}}^{\ge \frac{c}{\gamma}\mathrm{D}_h}(\bar{x},x) + \overbrace{\mathrm{D}_{\hat{f}_\beta}(\bar{x},x^-)}^{\ge 0} + \tfrac{c}{2\gamma}\,\mathrm{D}_h(x,x^-) \ge \varphi_{\overline{C}}(\bar{x}).$$

This concludes the proof of Theorem 4.6.(i) and 4.6.(ii) in the setting of Theorem 4.6.(A).

Now we work under the setting of Theorem 4.6.(B), in which case the following lemma will be useful.

Lemma A.2 *Additionally to Assumption 1, suppose that $\hat{f}_\beta$ is $L_{\hat{f}_\beta}$-Lipschitz differentiable for some $L_{\hat{f}_\beta} \ge 0$. Then, for every $x, x^- \in C$ and $\bar{x} \in \mathrm{T}^{h\text{-FRB}}_{\gamma,\beta}(x,x^-)$ we have*

$$\left(\phi^{h\text{-FRB}}_{\gamma,\beta} + \mathrm{D}_{L_{\hat{f}_\beta}j}\right)(\bar{x},x) \le \left(\phi^{h\text{-FRB}}_{\gamma,\beta} + \mathrm{D}_{L_{\hat{f}_\beta}j}\right)(x,x^-) - \mathrm{D}_{\hat{h}-2L_{\hat{f}_\beta}j}(\bar{x},x). \tag{A.2}$$

Proof By means of the three-point identity, that is, by using (4.3a) in place of (4.3b), inequality (4.6) can equivalently be written as

$$\phi^{h\text{-FRB}}_{\gamma,\beta}(\bar{x},x) \le \varphi_{\overline{C}}(\bar{x}) = \phi^{h\text{-FRB}}_{\gamma,\beta}(x,x^-) - \mathrm{D}_{\hat{h}}(\bar{x},x) - \langle \bar{x}-x, \nabla \hat{f}_\beta(x) - \nabla \hat{f}_\beta(x^-)\rangle$$

which by using Young's inequality on the inner product and $L_{\hat{f}_\beta}$-Lipschitz differentiability yields

$$\le \phi^{h\text{-FRB}}_{\gamma,\beta}(x,x^-) - \mathrm{D}_{\hat{h}}(\bar{x},x) + \frac{L_{\hat{f}_\beta}}{2}\|\bar{x}-x\|^2 + \frac{L_{\hat{f}_\beta}}{2}\|x-x^-\|^2. \tag{A.3}$$

Rearranging and using the fact that $\mathrm{D}_j(x,y) = \frac{1}{2}\|x-y\|^2$ yields the claimed inequality. □

Under Theorem 4.6.(B), recall that we define

$$c := (1+\alpha p_{-f,h}) - \frac{2\gamma L_{\hat{f}_\beta}}{\sigma_h} > 0.$$

We will pattern the arguments of the previous case, and observe that inequality (A.2) can equivalently be written in terms of $\mathcal{L}^{h\text{-FRB}}_{\gamma,\beta}$ as

$$\mathcal{L}^{h\text{-FRB}}_{\gamma,\beta}(\bar{x},x) \le \mathcal{L}^{h\text{-FRB}}_{\gamma,\beta}(x,x^-) - \overbrace{\mathrm{D}_{\hat{h}-2L_{\hat{f}_\beta}j-\frac{c}{\gamma}h}}^{\ge 0}(\bar{x},x) - \tfrac{c}{2\gamma}\,\mathrm{D}_h(\bar{x},x) - \tfrac{c}{2\gamma}\,\mathrm{D}_h(x,x^-).$$

Once again, the fact that $\mathrm{D}_{\hat{h}-2L_{\hat{f}_\beta}\jmath-\frac{c}{\gamma}h} \geq 0$ owes to the convexity of $\hat{h}-2L_{\hat{f}_\beta}\jmath-\frac{c}{\gamma}h$ on dom h, having

$$\begin{aligned}\hat{h}-2L_{\hat{f}_\beta}\jmath-\tfrac{c}{\gamma}h &= \tfrac{1-c}{\gamma}h - f - 2L_{\hat{f}_\beta}\jmath \\ &= \overbrace{\tfrac{1+\gamma\sigma_{-f,h}-c}{\gamma}}^{=2L_{\hat{f}_\beta}/\sigma_h\geq 0}\overbrace{(h-\sigma_h\jmath)}^{\text{convex}} + \overbrace{(-f-\sigma_{-f,h}h)}^{\text{convex}} + \overbrace{\left(\tfrac{1+\gamma\sigma_{-f,h}-c}{\gamma}\sigma_h - 2L_{\hat{f}_\beta}\right)}^{=0}\jmath\end{aligned}$$

altogether proving (4.10a). Similarly, inequality (4.10b) follows from (A.3) together with the fact that $\mathrm{D}_{\hat{h}-L_{\hat{f}_\beta}\jmath} \geq \mathrm{D}_{\hat{h}-2L_{\hat{f}_\beta}\jmath} \geq \frac{c}{\gamma}\,\mathrm{D}_h$, as shown above, having

$$\begin{aligned}\varphi_{\overline{C}}(\bar{x}) &\overset{(A.3)}{\leq} \Big(\overbrace{\mathcal{L}^{h\text{-FRB}}_{\gamma,\beta}-\mathrm{D}_{\frac{c}{2\gamma}h+L_{\hat{f}_\beta}\jmath}}^{\phi^{h\text{-FRB}}_{\gamma,\beta}}\Big)(x,x^-)-\mathrm{D}_{\hat{h}}(\bar{x},x)+\tfrac{L_{\hat{f}_\beta}}{2}\|\bar{x}-x\|^2+\tfrac{L_{\hat{f}_\beta}}{2}\|x-x^-\|^2 \\ &= \mathcal{L}^{h\text{-FRB}}_{\gamma,\beta}(x,x^-)-\tfrac{c}{2\gamma}\,\mathrm{D}_h(x,x^-)-\mathrm{D}_{\hat{h}-L_{\hat{f}_\beta}\jmath}(\bar{x},x) \\ &\leq \mathcal{L}^{h\text{-FRB}}_{\gamma,\beta}(x,x^-)-\tfrac{c}{2\gamma}\,\mathrm{D}_h(x,x^-)-\tfrac{c}{\gamma}\,\mathrm{D}_h(\bar{x},x).\end{aligned}$$

This concludes the proof of Theorems 4.6.(i) and 4.6.(ii).

A.2 Proof of Theorem 4.6.(iii)

We first state a property of the Bregman distance D_h that holds when h is as in the assertion of the theorem. The proof is provided for completeness, though part of it is straightforward and the rest is an easy adaptation of [6, Lem. 7.3(viii)].

Lemma A.3 *Let $h:\mathbb{R}^n\to\overline{\mathbb{R}}$ be a 1-coercive Legendre kernel. If either h is strongly convex or* dom $h=\mathbb{R}^n$*, then $\mathrm{D}_h(x,y)$ is level bounded in y locally uniformly in x.*

Proof Let $(x^k)_{k\in\mathbb{N}}$ and $(y^k)_{k\in\mathbb{N}}$ be sequences in $\mathbb{R}^n$ such that $\mathrm{D}_h(x^k,y^k)\leq\ell$ for some $\ell\in\mathbb{R}$. Suppose that $(x^k)_{k\in\mathbb{N}}$ is bounded; then the proof reduces to showing that also $(y^k)_{k\in\mathbb{N}}$ is. If h is, say, σ_h-strongly convex for some $\sigma_h>0$, the claim trivially follows from the fact that $\mathrm{D}_h(x,y)\geq(\sigma_h/2)\|x-y\|^2$ in this case.

Suppose that dom $h=\mathbb{R}^n$, then it follows from [6, Thm. 3.4] that h^* is 1-coercive. Furthermore, observe that

$$\begin{aligned}\ell\geq\mathrm{D}_h(x^k,y^k) &= h(x^k)-h(y^k)-\langle\nabla h(y^k),x^k-y^k\rangle \\ &= h(x^k)+h^*(\nabla h(y^k))-\langle\nabla h(y^k),x^k\rangle \\ &\geq c+h^*(\nabla h(y^k))-\|\nabla h(y^k)\|c',\end{aligned}$$

where $c:=\inf h(x^k)$ and $c':=\sup\|x^k\|$ are finite. Since h^* is 1-coercive, it follows that $(\nabla h(y^k))_{k\in\mathbb{N}}$ is bounded, and therefore so is $(y^k)_{k\in\mathbb{N}}$ by virtue of [6, Thm. 3.3]. □

We now turn to the proof of Theorem 4.6.(iii). By contraposition, suppose that $\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}$ is not level bounded, and consider an unbounded sequence $(x_k, x_k^-)_{k\in\mathbb{N}}$ such that

$$\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}(x_k, x_k^-) \leq \ell,$$

for some $\ell \in \mathbb{R}$. Then, it follows from (4.10b) that

$$\begin{aligned}\inf \varphi_{\overline{C}} + \tfrac{c}{\gamma}\,\mathrm{D}_h(\bar{x}_k, x_k) + \tfrac{c}{2\gamma}\,\mathrm{D}_h(x_k, x_k^-) \leq \varphi_{\overline{C}}(\bar{x}_k) + \tfrac{c}{\gamma}\,\mathrm{D}_h(\bar{x}_k, x_k)\\ + \tfrac{c}{2\gamma}\,\mathrm{D}_h(x_k, x_k^-) \leq \ell,\end{aligned}$$

and in particular both $(\mathrm{D}_h(\bar{x}_k, x_k))_{k\in\mathbb{N}}$ and $(\mathrm{D}_h(x_k, x_k^-))_{k\in\mathbb{N}}$ are bounded. Moreover, it follows from Lemma A.3 that if $(x_k^-)_{k\in\mathbb{N}}$ is unbounded then so is $(x_k)_{k\in\mathbb{N}}$, and similarly unboundedness of $(x_k)_{k\in\mathbb{N}}$ implies that of $(\bar{x}_k)_{k\in\mathbb{N}}$. Since at least one among $(x_k)_{k\in\mathbb{N}}$ and $(x_k^-)_{k\in\mathbb{N}}$ is unbounded, it follows that $(\bar{x}_k)_{k\in\mathbb{N}}$ is unbounded. Noticing that this sequence is contained in $[\varphi_{\overline{C}} \leq \ell]$, we conclude that $\varphi_{\overline{C}}$ is not level bounded.

B Proof of Theorem 5.6

In the remainder of this section, we will make use of the norm $|||\cdot|||$ on the product space $\mathbb{R}^n \times \mathbb{R}^n$ defined as $|||(x, y)||| = \|x\| + \|y\|$. For a set $E \subseteq \mathbb{R}^n$, define $(\forall \varepsilon > 0)$ $E_\varepsilon = \{x \in \mathbb{R}^n : \mathrm{dist}(x, E) < \varepsilon\}$.

Let $(\forall k \in \mathbb{N})$ $z^k = (x^{k+1}, x^k, x^{k-1})$, and let Ω be the set of limit points of $(z_k)_{k\in\mathbb{N}}$. Define

$$(\forall \omega, x, x^- \in \mathbb{R}^n)\ \mathscr{F}^{h\text{-FRB}}_{\gamma,\beta}(\omega, x, x^-) = \mathscr{M}^{h\text{-FRB}}_{\gamma,\beta}(\omega; x, x^-) + \tfrac{c}{2\gamma}\,\mathrm{D}_h(x, x^-) + D_\xi(x, x^-).$$

Set $(\forall k \in \mathbb{N})$ $\delta_k = \mathscr{F}^{h\text{-FRB}}_{\gamma,\beta}(z^k)$ for simplicity. Then $\delta_k = \mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}(x^k, x^{k-1})$, $\delta_k \to \varphi^\star$ decreasingly and $\mathrm{dist}(x^k, \Omega) \to 0$ as $k \to +\infty$ by invoking Theorem 5.1. Assume without loss of generality that $(\forall k \in \mathbb{N})$ $\delta_k > \varphi^\star$, otherwise we would have $(\exists k_0 \in \mathbb{N})$ $x^{k_0} = x^{k_0+1}$ due to Theorem 5.1.(i), from which the desired result readily follows by simple induction. Thus, $(\exists k_0 \in \mathbb{N})$ $(\forall k \geq k_0)$ $z^k \in \Omega_r \cap [0 < \mathscr{F}^{h\text{-FRB}}_{\gamma,\beta} - \varphi^\star < \eta]$. Appealing to Theorem 5.1.(iii) and Lemma 4.2.(i) yields that $\mathscr{F}^{h\text{-FRB}}_{\gamma,\beta}$ is constantly equal to $\varphi^\star$ on the compact set Ω. Note that $\mathscr{F}^{h\text{-FRB}}_{\gamma,\beta}$ satisfies the KL property under Assumption 5.6.A3; see, e.g., [2, §4.3]. In turn, appealing to Assumption 5.6.A3 and a standard uniformizing technique of the KL property, see, e.g., [13, Lem. 6.2], implies that there exists a concave $\psi \in \Psi_\eta$ such that for $k \geq k_0$

$$1 \leq \psi'\left(\delta_k - \varphi^\star\right) \mathrm{dist}\left(0, \partial \mathscr{F}^{h\text{-FRB}}_{\gamma,\beta}(z^k)\right). \tag{B.1}$$

Define $(\forall k \in \mathbb{N})$

$$u^k = \nabla^2 \hat{h}(x^k)(x^k - x^{k+1}) + \nabla^2 \hat{f}_\beta(x^k)(x^{k+1} - x^k) + \tfrac{c}{2\gamma}\left(\nabla h(x^k) - \nabla h(x^{k-1})\right)$$

$$
\begin{aligned}
&\quad + \nabla \hat{f}_\beta(x^{k-1}) - \nabla \hat{f}_\beta(x^k) + \nabla\xi(x^k) - \nabla\xi(x^{k-1}),\\
v^k &= \nabla^2 \hat{f}_\beta(x^{k-1})(x^k - x^{k+1})\\
&\quad + \tfrac{c}{2\gamma}\nabla^2 h(x^{k-1})(x^{k-1} - x^k) + \nabla^2\xi(x^{k-1})(x^{k-1} - x^k).
\end{aligned}
$$

Applying subdifferential calculus to $\partial\mathcal{F}^{h\text{-FRB}}_{\gamma,\beta}(z^k)$ yields that

$$
\begin{aligned}
\partial\mathcal{F}^{h\text{-FRB}}_{\gamma,\beta}(z^k) = \big(\partial\varphi(x^{k+1}) + \nabla\hat{h}(x^{k+1}) - \nabla\hat{h}(x^k) + \nabla \hat{f}_\beta(x^k) - \nabla \hat{f}_\beta(x^{k-1})\big)\\
\times\{u^k\}\times\{v^k\},
\end{aligned}
$$

which together with Lemma 4.2.(iv) entails that $(0, u^k, v^k) \in \partial\mathcal{F}^{h\text{-FRB}}_{\gamma,\beta}(z^k)$. In turn, Assumption 5.6.A2 implies that there exists $M > 0$ such that

$$
\operatorname{dist}\left(0, \partial\mathcal{F}^{h\text{-FRB}}_{\gamma,\beta}(z^k)\right) \le M|||(x^{k+1} - x^k, x^k - x^{k-1})|||. \tag{B.2}
$$

Finally, we show that $(x^k)_{k\in\mathbb{N}}$ is convergent. For simplicity, define $(\forall k, l \in \mathbb{N})$ $\Delta_{k,l} = \psi(\delta_k - \varphi^\star) - \psi(\delta_l - \varphi^\star)$. Then, combining (B.1) and (B.2) yields

$$
\begin{aligned}
1 &\le M\psi'(\delta_k - \varphi^\star)|||(x^{k+1} - x^k, x^k - x^{k-1})||| \le \tfrac{M\Delta_{k,k+1}}{\delta_k - \delta_{k+1}}|||(x^{k+1} - x^k, x^k - x^{k-1})|||\\
&\le \frac{2\gamma M\Delta_{k,k+1}}{c\left(\mathrm{D}_h(x^{k+1}, x^k) + \mathrm{D}_h(x^k, x^{k-1})\right)}|||(x^{k+1} - x^k, x^k - x^{k-1})|||\\
&\le \frac{2\gamma M\Delta_{k,k+1}|||(x^{k+1} - x^k, x^k - x^{k-1})|||}{c\sigma\|x^{k+1} - x^k\|^2 + c\sigma\|x^k - x^{k-1}\|^2} \le \frac{4\gamma M\Delta_{k,k+1}}{c\sigma|||(x^{k+1} - x^k, x^k - x^{k-1})|||},
\end{aligned}
$$

where the second inequality is implied by concavity of ψ, the third one follows from (5.1), and the fourth one holds because $\sigma > 0$ is the strong convexity modulus of h on a convex compact set that contains all the iterates. Hence,

$$
\|x^k - x^{k-1}\| \le |||(x^{k+1} - x^k, x^k - x^{k-1})||| \le \tfrac{4\gamma M}{c\sigma}\Delta_{k,k+1}. \tag{B.3}
$$

Summing (B.3) from $k = k_0$ to an arbitrary $l \ge k_0 + 1$ yields that

$$
\begin{aligned}
\sum_{k=k_0}^{l}\|x^k - x^{k-1}\| \le \tfrac{4\gamma M}{c\sigma}\Delta_{k_0,l} \le \tfrac{4\gamma M}{c\sigma}\psi\left(\delta_{k_0} - \varphi^\star\right)\\
= \tfrac{4\gamma M}{c\sigma}\psi\left(\mathcal{L}^{h\text{-FRB}}_{\gamma,\beta}(x^{k_0}, x^{k_0-1}) - \varphi^\star\right),
\end{aligned}
$$

where the second inequality holds as $\psi \ge 0$, from which one sees that $\sum_{k=0}^{\infty}\|x^{k+1} - x^k\|$ is finite as l is arbitrary. A similar procedure shows that $(x^k)_{k\in\mathbb{N}}$ is Cauchy, which together with Theorem 5.1.(iii) entails the rest of the statement.

C Proof of Theorem 5.9

Assume without loss of generality that $\mathscr{F}^{h\text{-FRB}}_{\gamma,\beta}$ has desingularizing function $\psi(t) = t^{1-\theta}/(1-\theta)$ and let $(\forall k \in \mathbb{N})\ \delta_k = \sum_{i=k}^{\infty} \|x^{i+1} - x^i\|$. We claim that

$$(\forall k \geq k_0 + 1)\ \delta_k \leq \tfrac{4\gamma M}{(1-\theta)c\sigma} e_k^{1-\theta}. \tag{C.1}$$

Indeed, summing (B.3) from every $k \geq k_0$ to $l \geq k+1$ and passing l to infinity give

$$\delta_{k-1} = \sum_{i=k}^{\infty} \|x^i - x^{i-1}\| \leq \tfrac{4\gamma M}{c\sigma} \psi\left(\mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}(x^k, x^{k-1}) - \varphi^\star\right) = \tfrac{4\gamma M}{c\sigma} \psi\left(e_{k-1}\right),$$

from which the desired claim readily follows. It is routine to see that the desired sequential rate can be implied by those of $(e_k)_{k\in\mathbb{N}}$ through (C.1); see, e.g., [45, Thm. 5.3]; therefore, it suffices to prove convergence rate of $(e_k)_{k\in\mathbb{N}}$.

Recall from Theorem 5.1.(i) that $(e_k)_{k\in\mathbb{N}}$ is a decreasing sequence converging to 0. Then invoking the KL exponent assumption yields

$$e_{k-1}^\theta = [\mathscr{F}^{h\text{-FRB}}_{\gamma,\beta}(x^{k+1}, x^k, x^{k-1}) - \varphi^\star]^\theta \leq \operatorname{dist}(0, \partial\mathscr{F}^{h\text{-FRB}}_{\gamma,\beta}(x^{k+1}, x^k, x^{k-1})),$$

where the first equality holds due to Lemma 4.5.(ii), which together with (B.2) implies that

$$e_k^\theta \leq e_{k-1}^\theta \leq M|||(x^{k+1} - x^k, x^k - x^{k-1})|||. \tag{C.2}$$

Appealing again to Theorem 5.1.(i) gives

$$\begin{aligned} e_{k-1} - e_k &= \mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}(x^k, x^{k-1}) - \mathscr{L}^{h\text{-FRB}}_{\gamma,\beta}(x^{k+1}, x^k) \\ &\geq \tfrac{c}{2\gamma}\left[\mathrm{D}_h(x^{k+1}, x^k) + \mathrm{D}_h(x^k, x^{k-1})\right] \\ &\geq \tfrac{c\sigma}{8\gamma}|||(x^{k+1} - x^k, x^k - x^{k-1})|||^2 \geq \tfrac{c\sigma}{8\gamma M^2} e_k^{2\theta}, \end{aligned}$$

where the last inequality is implied by (C.2). Then [15, Lem. 10] justifies the desired rate of $(e_k)_{k\in\mathbb{N}}$.

References

1. Ahookhosh, M., Themelis, A., Patrinos, P.: A Bregman forward-backward linesearch algorithm for nonconvex composite optimization: superlinear convergence to nonisolated local minima. SIAM J. Optim. **31**(1), 653–685 (2021)
2. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality. Math. Oper. Res. **35**(2), 438–457 (2010)
3. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. Math. Program. **137**(1), 91–129 (2013)

4. Azé, D., Penot, J.: Uniformly convex and uniformly smooth convex functions. Annales de la Faculté des sciences de Toulouse: Mathématiques, Ser. 6 **4**(4), 705–730 (1995)
5. Bauschke, H.H., Bolte, J., Teboulle, M.: A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. Math. Oper. Res. **42**(2), 330–348 (2017)
6. Bauschke, H.H., Borwein, J.M., Combettes, P.L.: Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces. Commun. Contemp. Math. **3**(04), 615–647 (2001)
7. Bauschke, H.H., Borwein, J.M., Combettes, P.L.: Bregman monotone optimization algorithms. SIAM J. Control. Optim. **42**(2), 596–636 (2003)
8. Bauschke, H.H., Combettes, P.L.: Iterating Bregman retractions. SIAM J. Optim. **13**(4), 1159–1173 (2003)
9. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. CMS Books in Mathematics, Springer (2017)
10. Beck, A., Teboulle, M.: Mirror descent and nonlinear projected subgradient methods for convex optimization. Oper. Res. Lett. **31**(3), 167–175 (2003)
11. Bertsekas, D.P.: Nonlinear Programming. Athena Scientific (2016)
12. Böhm, A., Sedlmayer, M., Csetnek, E.R., Boţ, R.I.: Two steps at a time-taking GAN training in stride with Tseng's method. SIAM J. Math. Data Sci. **4**(2), 750–771 (2022)
13. Bolte, J., Sabach, S., Teboulle, M., Vaisbourd, Y.: First order methods beyond convexity and Lipschitz gradient continuity with applications to quadratic inverse problems. SIAM J. Optim. **28**(3), 2131–2151 (2018)
14. Boţ, R.I., Dao, M.N., Li, G.: Extrapolated proximal subgradient algorithms for nonconvex and nonsmooth fractional programs. Math. Oper. Res. **47**(3), 2415–2443 (2022)
15. Boţ, R.I., Nguyen, D.: The proximal alternating direction method of multipliers in the nonconvex setting: convergence analysis and rates. Math. Oper. Res. **45**(2), 682–712 (2020)
16. Chen, G., Teboulle, M.: Convergence analysis of a proximal-like minimization algorithm using Bregman functions. SIAM J. Optim. **3**(3), 538–543 (1993)
17. Dragomir, R., d'Aspremont, A., Bolte, J.: Quartic first-order methods for low-rank minimization. J. Optim. Theory Appl. **189**(2), 341–363 (2021)
18. Dragomir, R., Taylor, A.B., d'Aspremont, A., Bolte, J.: Optimal complexity and certification of Bregman first-order methods. Math. Program. **194**(1), 41–83 (2022)
19. Gidel, G., Hemmat, R.A., Pezeshki, M., Le Priol, R., Huang, G., Lacoste-Julien, S., Mitliagkas, I.: Negative momentum for improved game dynamics. In: The 22nd International Conference on Artificial Intelligence and Statistics, pp. 1802–1811. PMLR (2019)
20. Hanzely, F., Richtarik, P., Xiao, L.: Accelerated Bregman proximal gradient methods for relatively smooth convex optimization. Comput. Optim. Appl. **79**(2), 405–440 (2021)
21. Kan, C., Song, W.: The Moreau envelope function and proximal mapping in the sense of the Bregman distance. Nonlinear Anal.: Theory Methods Appl. **75**(3), 1385–1399 (2012)
22. László, S.C.: A forward-backward algorithm with different inertial terms for structured non-convex minimization problems. J. Optim, Theory Appl (2023)
23. Li, G., Liu, T., Pong, T.K.: Peaceman-Rachford splitting for a class of nonconvex optimization problems. Comput. Optim. Appl. **68**(2), 407–436 (2017)
24. Li, G., Pong, T.K.: Global convergence of splitting methods for nonconvex composite optimization. SIAM J. Optim. **25**(4), 2434–2460 (2015)
25. Li, G., Pong, T.K.: Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems. Math. Program. **159**(1), 371–401 (2016)
26. Liu, Y., Yin, W.: An envelope for Davis-Yin splitting and strict saddle-point avoidance. J. Optim. Theory Appl. **181**(2), 567–587 (2019)
27. Lu, H., Freund, R.M., Nesterov, Y.: Relatively smooth convex optimization by first-order methods, and applications. SIAM J. Optim. **28**(1), 333–354 (2018)
28. Mairal, J.: Incremental majorization-minimization optimization with application to large-scale machine learning. SIAM J. Optim. **25**(2), 829–855 (2015)
29. Malitsky, Y., Tam, M.K.: A forward-backward splitting method for monotone inclusions without cocoercivity. SIAM J. Optim. **30**(2), 1451–1472 (2020)
30. Mordukhovich, B.: Variational Analysis and Applications, volume 30. Springer (2018)
31. Moreau, J.: Proximité et dualité dans un espace hilbertien. Bull. Soc. Math. France **93**, 273–299 (1965)
32. Moreau, J.: Fonctionnelles convexes. Séminaire Jean Leray (2):1–108 (1966–1967)
33. Nesterov, Y.: Lectures on Convex Optimization, volume 137. Springer (2018)

34. Nesterov, Y.: Implementable tensor methods in unconstrained convex optimization. Math. Program. **186**, 157–183 (2021)
35. Parikh, N., Boyd, S.: Proximal algorithms. Found. Trends Optim. **1**(3), 127–239 (2014)
36. Reem, D., Reich, S., De Pierro, A.: Re-examination of Bregman functions and new properties of their divergences. Optimization **68**(1), 279–348 (2019)
37. Rockafellar, R.T., Wets, R.J.: Variational Analysis, volume 317. Springer (2011)
38. Stella, L., Themelis, A., Patrinos, P.: Newton-type alternating minimization algorithm for convex optimization. IEEE Trans, Automatic Control (2018)
39. Teboulle, M.: A simplified view of first order methods for optimization. Math. Program. **170**(1), 67–96 (2018)
40. Themelis, A.: Proximal Algorithms for Structured Nonconvex Optimization. PhD thesis, KU Leuven (2018)
41. Themelis, A., Hermans, B., Patrinos, P.: A new envelope function for nonsmooth DC optimization. In: 2020 59th IEEE Conference on Decision and Control (CDC), pp. 4697–4702 (2020)
42. Themelis, A., Patrinos, P.: Douglas-Rachford splitting and ADMM for nonconvex optimization: Tight convergence results. SIAM J. Optim. **30**(1), 149–181 (2020)
43. Themelis, A., Stella, L., Patrinos, P.: Forward-backward envelope for the sum of two nonconvex functions: further properties and nonmonotone linesearch algorithms. SIAM J. Optim. **28**(3), 2274–2303 (2018)
44. Themelis, A., Stella, L., Patrinos, P.: Douglas-Rachford splitting and ADMM for nonconvex optimization: accelerated and Newton-type algorithms. Comput. Optim. Appl. **82**, 395–440 (2022)
45. Wang, X., Wang, Z.: A Bregman inertial forward-reflected-backward method for nonconvex minimization. J. Glob. Optim. (2023). https://doi.org/10.1007/s10898-023-01348-y
46. Wang, X., Wang, Z.: The exact modulus of the generalized concave Kurdyka-Łojasiewicz property. Math. Oper. Res. **47**(4), 2765–2783 (2022)
47. Wang, X., Wang, Z.: Malitsky-Tam forward-reflected-backward splitting method for nonconvex minimization problems. Comput. Optim. Appl. **82**(2), 441–463 (2022)