# First Order Nonconvex and Nonsmooth Optimization Algorithms: Regularity Conditions, Convergence Rates and Applications

Thesis Proposal

Ziyuan Wang

Department of Mathematics
University of British Columbia,
Okanagan Campus.

July 6, 2022

# Contents

# 1 Introduction

Consider the inclusion problem of maximally monotone operators:

$$\text{find } x \in \mathbb{R}^n \text{ such that } 0 \in (A+B)(x), \tag{1}$$

where $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ and $B : \mathbb{R}^n \to \mathbb{R}^n$ are (maximally) monotone operators with $B$ being Lipschitz continuous. Behavior of splitting methods for solving (1) is well-understood; see, e.g., the recent forward-reflected-backward method by Malitsky and Tam [26], and [5, Chapters 26, 28], which includes classical algorithms such as the forward-backward method [16] with $B$ being cocoercive, the Douglas-Rachford method [18, 23] and Tseng's method [33].

We are interested in nonconvex problems of the form

$$\min_{x \in \mathbb{R}^n} f(x) + g(x),$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper lsc and $g : \mathbb{R}^n \to \mathbb{R}$ has Lipschitz gradient. When $f$ and $g$ are convex, the above problem corresponds to (1) with $A = \partial f$ and $B = \nabla g$. However, in the absence of convexity, the usual convergence analysis of splitting methods for (1) collapses. This is where various regularity conditions come into play. *In this proposal, we focus on the generalized concave Kurdyka-Łojasiewicz (KL) property (see Definition 2.3).*

*The organization of this proposal now follows.* Beginning with notation and background knowledge in Section 2, *we present main results completed during my PhD program in Sections 3–5*, which are based on published paper [37] and preprints [35, 36]. The convergence of Malitsky-Tam forward-backward-reflected method in the absence of convexity is presented in Section 3 . In Section 4, we propose a Bregman inertial forward-reflected-backward method and analyze its behavior, followed by a sum rule of the generalized concave KL property in Section 5. *Finally, we end this proposal by providing direction for future research in Section 6.*

# 2 Notation and preliminaries

Throughout this proposal,

$\mathbb{R}^n$ is the standard Euclidean space

equipped with inner product $\langle x, y \rangle = x^T y$ and the Euclidean norm $\|x\| = \sqrt{\langle x, x \rangle}$ for $x, y \in \mathbb{R}^n$. Let $\mathbb{N} = \{0, 1, 2, \ldots\}$ and $\mathbb{N}^* = \{-1\} \cup \mathbb{N}$. The open ball centered at $\bar{x}$ with radius $r$ is denoted by $\mathbb{B}(\bar{x};r)$. The distance function of a subset $K \subseteq \mathbb{R}^n$ is $\operatorname{dist}(\cdot,K) : \mathbb{R}^n \to \overline{\mathbb{R}} = (-\infty,\infty] : x \mapsto \operatorname{dist}(x,K) = \inf\{\|x-y\| : y \in K\}$, where $\operatorname{dist}(x,K) \equiv \infty$ if $K = \emptyset$. For $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $r_1, r_2 \in [-\infty,\infty]$, we set $[r_1 < f < r_2] = \{x \in \mathbb{R}^n : r_1 < f(x) < r_2\}$.

## 2.1 Elements from variational analysis

We say a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is coercive if $\lim_{\|x\|\to\infty} f(x) = \infty$. A proper, lsc function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is prox-bounded if there exists $\lambda > 0$ such that $f + \|\cdot\|^2/(2\lambda)$ is bounded below; see, e.g., [32, Exercise 1.24]. The supremum of the set of all such $\lambda$ is the threshold $\lambda_f$ of prox-boundedness for $f$. The indicator function of a set $K \subseteq \mathbb{R}^n$ is $\delta_K(x) = 0$ if $x \in K$; and $\delta_K(x) = \infty$ otherwise. The Fenchel conjugate of $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is $(\forall y \in \mathbb{R}^n)\ f^*(y) = \sup_{x\in\mathbb{R}^n}\{\langle x,y\rangle - f(x)\}$ and the Moreau envelope of $f$ with parameter $\lambda > 0$ is

$$(\forall x \in \mathbb{R}^n)\ M_f^\lambda(x) = \inf_{y\in\mathbb{R}^n}\left\{f(y) + \frac{1}{2\lambda}\|x-y\|^2\right\},$$

and the proximal mapping of $f$ with parameter $\lambda > 0$ is

$$(\forall x \in \mathbb{R}^n)\ \operatorname{Prox}_{\lambda f}(x) = \operatorname*{argmin}_{y\in\mathbb{R}^n}\left\{f(y) + \frac{1}{2\lambda}\|x-y\|^2\right\}.$$

The Bregman distance induced by differentiable $h : \mathbb{R}^n \to \mathbb{R}$ is

$$D_h : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} : (x,y) \mapsto h(x) - h(y) - \langle x-y, \nabla h(y)\rangle.$$

**Lemma 2.1.** *[5, Exercise 17.5, Theorem 18.15]* Let $h : \mathbb{R}^n \to \mathbb{R}$ be $\sigma$-strongly convex and has Lipschitz continuous gradient with constant $L_{\nabla h}$. Then

$$(\forall x,y \in \mathbb{R}^n)\ \frac{\sigma}{2}\|x-y\|^2 \le D_h(x,y) \le \frac{L_{\nabla h}}{2}\|x-y\|^2.$$

We will use frequently the following concepts; see, e.g., [27,32].

**Definition 2.2.** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper function and $K \subseteq \mathbb{R}^n$ a nonempty set. We say that

(i) $v \in \mathbb{R}^n$ is a *Fréchet subgradient* of $f$ at $\bar{x} \in \operatorname{dom} f$, denoted by $v \in \hat{\partial} f(\bar{x})$, if

$$f(x) \ge f(\bar{x}) + \langle v, x-\bar{x}\rangle + o(\|x-\bar{x}\|). \tag{2}$$

(ii) $v \in \mathbb{R}^n$ is a *limiting subgradient* of $f$ at $\bar{x} \in \operatorname{dom} f$, denoted by $v \in \partial f(\bar{x})$, if

$$v \in \{v \in \mathbb{R}^n : \exists x_k \xrightarrow{f} \bar{x}, \exists v_k \in \hat{\partial} f(x_k), v_k \to v\}, \tag{3}$$

where $x_k \xrightarrow{f} \bar{x} \Leftrightarrow x_k \to \bar{x}$ and $f(x_k) \to f(\bar{x})$. Moreover, we set $\operatorname{dom} \partial f = \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$. We say that $\bar{x} \in \operatorname{dom} \partial f$ is a stationary point, if $0 \in \partial f(\bar{x})$.

(iii) The Fréchet and limiting normal cones to $K$ at $\bar{x}$ are $\hat{N}_K(\bar{x}) = \hat{\partial}\delta_K(\bar{x})$ and $N_K(\bar{x}) = \partial\delta_K(\bar{x})$, respectively.

If $K \subseteq \mathbb{R}^n$ is a nonempty convex set, then $N_K(\bar{x}) = \hat{N}_K(\bar{x}) = \{v \in \mathbb{R}^n : (\forall x \in K)\ \langle v, x - \bar{x}\rangle \leq 0\}$; see, e.g., [32].

## 2.2 The generalized concave KL property

This subsection is based on published paper [34], which is partially completed in the master thesis [38]. Here we only collect definitions and facts necessary for the presentation of this proposal.

For $\eta \in (0, \infty]$, denote by $\Phi_\eta$ the class of functions $\varphi : [0, \eta) \to \mathbb{R}_+$ satisfying the following conditions: (i) $\varphi(t)$ is right-continuous at $t = 0$ with $\varphi(0) = 0$; (ii) $\varphi$ is strictly increasing on $[0, \eta)$.

**Definition 2.3.** *[34]* Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc. Let $\bar{x} \in \operatorname{dom} \partial f$ and $\mu \in \mathbb{R}$, and let $V \subseteq \operatorname{dom} \partial f$ be a nonempty subset.

(i) We say that $f$ has the pointwise generalized concave Kurdyka-Łojasiewicz (KL) property at $\bar{x} \in \operatorname{dom} \partial f$, if there exist neighborhood $U \ni \bar{x}$, $\eta \in (0, \infty]$ and concave $\varphi \in \Phi_\eta$, such that for all $x \in U \cap [0 < f - f(\bar{x}) < \eta]$,

$$\varphi'_-\big(f(x) - f(\bar{x})\big) \cdot \operatorname{dist}\big(0, \partial f(x)\big) \geq 1, \tag{4}$$

where $\varphi'_-$ denotes the left derivative. Moreover, $f$ is a generalized concave KL function if it has the generalized concave KL property at every $x \in \operatorname{dom} \partial f$.

(ii) Suppose that $f(x) = \mu$ on $V$. We say $f$ has the setwise[1] generalized concave Kurdyka-Łojasiewicz property on $V$, if there exist $U \supset V$, $\eta \in (0, \infty]$ and concave $\varphi \in \Phi_\eta$ such that for every $x \in U \cap [0 < f - \mu < \eta]$,

$$\varphi'_-\big(f(x) - \mu\big) \cdot \operatorname{dist}\big(0, \partial f(x)\big) \geq 1. \tag{5}$$

**Remark 2.4.** When $\varphi$ in Definition 2.3 is continuously differentiable on $(0, \eta)$, the generalized concave KL property reduces to the concave KL property, which is a celebrated regularity condition for the convergence of various first-order algorithms for nonconvex optimization; see, e.g., [1, 2, 3, 4, 10, 12, 13, 39] and the references therein.

[1] We shall omit adjectives "pointwise" and "setwise" whenever there is no ambiguity.

Generalized concave KL functions are ubiquitous in applications. The class of semialgebraic functions is one particularly rich resource. The interested reader is referred to [2,8,9] for more examples; see also the fundamental work of Łojasiewicz [24] and Kurdyka [19].

**Definition 2.5.** (i) A set $E \subseteq \mathbb{R}^n$ is called semialgebraic if there exist finitely many polynomials $g_{ij}, h_{ij} : \mathbb{R}^n \to \mathbb{R}$ such that $E = \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{x \in \mathbb{R}^n : g_{ij}(x) = 0 \text{ and } h_{ij}(x) < 0\}$.

(ii) A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called semialgebraic if its graph $\operatorname{gph} f = \{(x, y) \in \mathbb{R}^{n+1} : f(x) = y\}$ is semialgebraic.

**Fact 2.6.** *[8, Corollary 16]* Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper and lsc function and let $\bar{x} \in \operatorname{dom} \partial f$. If $f$ is 1/2braic, then it has the concave KL property at $\bar{x}$ with $\varphi(t) = c \cdot t^{1-\theta}$ for some $c > 0$ and $\theta \in (0, 1)$.

**Lemma 2.7.** *[34, Lemma 4.4]* Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper lsc and let $\mu \in \mathbb{R}$. Let $\Omega \subseteq \operatorname{dom} \partial f$ be a nonempty compact set on which $f(x) = \mu$ for all $x \in \Omega$. The following statements hold:

(i) Suppose that $f$ satisfies the pointwise generalized concave KL property at each $x \in \Omega$. Then there exist $\varepsilon > 0, \eta \in (0, \infty]$ and $\varphi \in \Phi_\eta$ such that $f$ has the setwise generalized concave KL property on $\Omega$ with respect to $U = \Omega_\varepsilon$, $\eta$ and $\varphi$.

(ii) Set $U = \Omega_\varepsilon$ and define $h : (0, \eta) \to \mathbb{R}_+$ by

$$h(s) = \sup \left\{ \operatorname{dist}^{-1} \left(0, \partial f(x)\right) : x \in U \cap [0 < f - \mu < \eta], s \leq f(x) - \mu \right\}.$$

Then the function $\tilde{\varphi} : [0, \eta) \to \mathbb{R}_+$,

$$(\forall t \in (0, \eta)) \ t \mapsto \int_0^t h(s) ds,$$

and $\tilde{\varphi}(0) = 0$, is well-defined and belongs to $\Phi_\eta$. The function $f$ has the setwise generalized concave KL property on $\Omega$ with respect to $U$, $\eta$ and $\tilde{\varphi}$. Moreover,

$$\tilde{\varphi} = \inf \left\{ \varphi \in \Phi_\eta : \varphi \text{ is a concave desingularizing function of } f \text{ on } \Omega \text{ with respect to } U \text{ and } \eta \right\}.$$

We say $\tilde{\varphi}$ is the exact modulus of the setwise generalized concave KL property of $f$ on $\Omega$ with respect to $U$ and $\eta$.

# 3 Malitsky-Tam forward-reflected-backward splitting method for nonconvex minimization problems

This section is based on published paper [37]. In the remainder of this section, assume that:

**Assumption 3.1.** (i) $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper, lsc and prox-bounded with threshold $\lambda_f$.

(ii) $g : \mathbb{R}^n \to \mathbb{R}$ has Lipschitz continuous gradient with constant $L_{\nabla g}$.

(iii) The function $F = f + g$ is bounded below.

Malitsky and Tam [26] proposed a *forward-reflected-backward* (FRB) splitting method for solving (1). Let $\lambda > 0$ and $J_{\lambda A} = (\mathrm{Id} + \lambda A)^{-1}$ be the resolvent of operator $\lambda A$; see, e.g., [5]. Given initial points $x_0, x_{-1} \in \mathbb{R}^n$, the Malitsky-Tam FRB method with a fixed step-size (see [26, Remark 2.1]) iterates

$$x_{k+1} = J_{\lambda A}(x_k - 2\lambda B(x_k) + \lambda B(x_{k-1})). \tag{6}$$

*The goal of this section is to establish the convergence of (6) when being applied to the nonconvex composite problem*

$$\min_{x \in \mathbb{R}^n} F(x) = f(x) + g(x). \tag{7}$$

We formulate the FRB scheme for solving (7) as follows

**Algorithm 3.2** (FRB)**.**

1. Initialization: Pick $x_{-1}, x_0 \in \mathbb{R}^n$. Let $\lambda > 0$.
2. For $k \in \mathbb{N}$, compute $y_k = x_k + \lambda\big(\nabla g(x_{k-1}) - \nabla g(x_k)\big)$ and

$$x_{k+1} \in \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ f(x) + \langle x - y_k, \nabla g(x_k) \rangle + \frac{1}{2\lambda} \|x - y_k\|^2 \right\}.$$

For $0 < \lambda < \lambda_f$, [32, Theorem 1.25] shows that the above scheme is well-defined.

## 3.1 Function value convergence

We now present function value convergence of Algorithm 3.2 , beginning with a merit function for the FRB splitting method, which will play a central role in our analysis. Such functions of various forms appear frequently in the convergence analysis of splitting methods in nonconvex settings; see, e.g., [12,20,22,31].

**Definition 3.3** (FRB merit function)**.** Let $\lambda > 0$. Define the FRB merit function $H : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ by

$$H(x, y) = f(x) + g(x) + \left( \frac{1}{4\lambda} - \frac{L}{4} \right) \|x - y\|^2 . \tag{8}$$

We use the Euclidean norm for $\mathbb{R}^n \times \mathbb{R}^n$, i.e., $\|(x,y)\| = \sqrt{\|x\|^2 + \|y\|^2}$. The FRB merit function has the decreasing property under a suitable step-size assumption.

**Lemma 3.4.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by FRB and define $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$. Assume that $0 < \lambda < \min\left\{\frac{1}{3L}, \lambda_f\right\}$ and $\inf(f+g) > -\infty$. Let $M_1 = \frac{1}{4\lambda} - \frac{3L}{4}$. Then the following statements hold:

(i) For $k \in \mathbb{N}$, we have

$$M_1 \|z_k - z_{k-1}\|^2 \leq H(z_{k-1}) - H(z_k), \tag{9}$$

which means that $H(z_k) \leq H(z_{k-1})$. Hence, the sequence $\big(H(z_k)\big)_{k\in\mathbb{N}^*}$ is convergent.

(ii) $\sum\limits_{k=0}^{\infty} \|z_k - z_{k-1}\|^2 < \infty$, consequently $\lim\limits_{k\to\infty} \|z_k - z_{k-1}\| = 0$.

**Theorem 3.5.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by FRB and define $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$. Assume that conditions in Lemma 3.4 are satisfied and $(z_k)_{k\in\mathbb{N}^*}$ is bounded. Suppose that a subsequence $(z_{k_l})_{l\in\mathbb{N}}$ of $(z_k)_{k\in\mathbb{N}^*}$ converges to some $z^* = (x^*, y^*)$ as $l \to \infty$. Then the following statements hold:

(i) $\lim\limits_{l\to\infty} H(z_{k_l}) = f(x^*) + g(x^*) = F(x^*)$. In fact, $\lim\limits_{k\to\infty} H(z_k) = F(x^*)$.

(ii) We have $x^* = y^*$ and $0 \in \partial H(x^*, y^*)$, which implies $0 \in \partial F(x^*) = \partial f(x^*) + \nabla g(x^*)$.

(iii) The set $\omega(z_{-1})$ is nonempty, compact and connected, on which the FRB merit function $H$ is finite and constant. Moreover, we have $\lim\limits_{k\to\infty} \operatorname{dist}(z_k, \omega(z_{-1})) = 0$.

Theorem 3.5 requires the sequence $(z_k)_{k\in\mathbb{N}^*}$ to be bounded. The result below provides a sufficient condition to such an assumption.

**Theorem 3.6.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by FRB and define $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$. Assume that conditions in Lemma 3.4 are satisfied. If $f+g$ is coercive (or level bounded), then the sequence $(x_k)_{k\in\mathbb{N}^*}$ is bounded, so is $(z_k)_{k\in\mathbb{N}^*}$.

*Proof.* Lemma 3.4(i) implies that we have

$$H(x_1, x_0) \geq f(x_{k+1}) + g(x_{k+1}) + \left(\frac{1}{4\lambda} - \frac{L}{4}\right) \|x_{k+1} - x_k\|^2 \geq f(x_{k+1}) + g(x_{k+1}).$$

Suppose that $(x_k)_{k\in\mathbb{N}^*}$ was unbounded. Then we would have a contradiction by the coercivity or level boundedness of $f+g$. ■

## 3.2 Sequential convergence

**Theorem 3.7** (Global convergence of FRB)**.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by FRB and define $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$. Assume that $(z_k)_{k\in\mathbb{N}^*}$ is bounded, $\inf(f+g) > -\infty$, and $0 < \lambda < \min\left\{\frac{1}{3L}, \lambda_f\right\}$. Suppose that the FRB merit function $H$ has the generalized concave KL property on $\omega(z_{-1})$. Then the following statements hold:

(i) The sequence $(z_k)_{k\in\mathbb{N}^*}$ is Cauchy and has finite length. To be specific, there exist $M > 0$, $k_0 \in \mathbb{N}$, $\varepsilon > 0$ and $\eta \in (0, \infty]$ such that for $i \geq k_0 + 1$

$$\sum_{k=i}^{\infty} \|z_{k+1} - z_k\| \leq \|z_i - z_{i-1}\| + M\tilde{\varphi}\left(H(z_i) - H(z^*)\right). \tag{10}$$

where $\tilde{\varphi} \in \Phi_\eta$ is the exact modulus associated with the setwise generalized concave KL property of $H$ on $\omega(z_{-1})$ with respect to $\varepsilon$ and $\eta$.

(ii) The sequence $(x_k)_{k\in\mathbb{N}^*}$ has finite length and converges to some $x^*$ with $0 \in \partial F(x^*)$.

**Corollary 3.8.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by FRB. Assume that $(x_k)_{k\in\mathbb{N}^*}$ is bounded, $\inf(f+g) > -\infty$, and $0 < \lambda < \min\left\{\frac{1}{3L}, \lambda_f\right\}$. Suppose further that $f$ and $g$ are both semialgebraic functions. Then $(x_k)_{k\in\mathbb{N}^*}$ converges to some $x^*$ with $0 \in \partial F(x^*)$ and has the finite length property.

*Proof.* Recall from [2, Section 4.3] that the class of semialgebraic functions is closed under summation and notice that the quadratic function $(x, y) \mapsto (\frac{1}{4\lambda} - \frac{L}{4})\|x - y\|^2$ is semialgebraic. Then Fact 2.6 implies that the FRB merit function $H$ is concave KL. Applying Theorem 4.17 then completes the proof. ∎

## 3.3 Convergence rates

**Theorem 3.9** (Function value)**.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by FRB and define $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$. Assume that $(z_k)_{k\in\mathbb{N}^*}$ is bounded, $\inf(f+g) > -\infty$, and $0 < \lambda < \min\left\{\frac{1}{3L}, \lambda_f\right\}$. Let $x^*$ be the limit given in Theorem 4.17(ii). Suppose that the FRB merit function $H$ has KL exponent $\theta \in [0, 1)$ at $z^* = (x^*, x^*)$. Define $e_k = H(z_k) - F(x^*)$ for $k \in \mathbb{N}$. Then $(e_k)_{k\in\mathbb{N}}$ converges to 0 and the following statements hold:

(i) If $\theta = 0$, then $(e_k)_{k\in\mathbb{N}}$ converges in finite steps.

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

$$e_k \leq \hat{c}_1 \hat{Q}_1^k.$$

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

$$e_k \leq \hat{c}_2 k^{-\frac{1}{2\theta-1}}.$$

**Theorem 3.10** (Actual sequence)**.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by FRB and define $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$. Assume that $(z_k)_{k\in\mathbb{N}^*}$ is bounded, $\inf(f+g) > -\infty$, and $0 < \lambda < \min\left\{\frac{1}{3L}, \lambda_f\right\}$. Let $M, k_0$ and $x^*$ be those given in Theorem 4.17. Suppose that the FRB merit function $H$ has KL exponent $\theta \in [0, 1)$ at $(x^*, x^*)$. Then the following statements hold:

(i) If $\theta = 0$, then $(x_k)_{k\in\mathbb{N}^*}$ converges to $x^*$ in finite steps.

(ii) If $\theta \in (0, 1/2]$, then there exist $Q_1 \in (0, 1)$ and $c_1 > 0$ such that

$$\|x_k - x^*\| \leq c_1 Q_1^k,$$

for $k$ sufficiently large.

(iii) If $\theta \in (1/2, 1)$, then there exists $Q_2 > 0$ such that

$$\|x_k - x^*\| \leq Q_2 k^{-\frac{1-\theta}{2\theta-1}},$$

for $k$ sufficiently large.

## 3.4 Numerical experiments

In this subsection, we apply the FRB splitting method to nonconvex feasibility problems.

### 3.4.1 Convergence guarantee

**Example 3.11.** Let $C \subseteq \mathbb{R}^n$ be a nonempty, closed and convex set and let $D \subseteq \mathbb{R}^n$ be nonempty and closed. Suppose that $C \cap D \neq \varnothing$ and either $C$ or $D$ is compact. Assume further that both $C$ and $D$ are semialgebraic. Consider the minimization problem (12) with $f = \delta_D$ and $g = \frac{1}{2}\operatorname{dist}^2(\cdot, C)$. Let $0 < \lambda < \frac{1}{3}$, and let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by FRB and $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$. Then the following statements hold:

(i) There exists $x^* \in \mathbb{R}^n$ such that $x_k \to x^*$ and $0 \in \partial F(x^*)$.

(ii) Suppose, in addition, that

$$N_C(\operatorname{Proj}_C(x^*)) \cap \big(-N_D(x^*)\big) = \{0\}. \tag{11}$$

Then $x^* \in C \cap D$. Moreover, there exist $Q_1 \in (0,1)$ and $c_1 > 0$ such that

$$\|x_k - x^*\| \leq c_1 Q_1^k,$$

for $k$ sufficiently large.

### 3.4.2 Application: finding sparse solution of linear system

Let $C = \{x \in \mathbb{R}^n : Ax = b\}$ for $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $r = m/5$, $l = 10^6$, and $D = \{x \in \mathbb{R}^n : \|x\|_0 \leq r, \|x\|_\infty \leq l\}$. Similarly to [22, Section 5], we consider the minimization problem (12) with

$$f(x) = \delta_D \text{ and } g(x) = \frac{1}{2}\operatorname{dist}^2(x, C).$$

Clearly $C$ is semialgebraic. As for $D$, first notice that $\mathbb{R}^n \times \{i\}$ for $i \in \mathbb{R}$ and $\operatorname{gph}\|\cdot\|_0$ are semialgebraic; see [10, Example 3]. Then

$$\begin{aligned}\|x\|_0 \leq r &\Leftrightarrow \exists i \in \{0,\dots,r\} \text{ such that } \|x\|_0 = i.\\ &\Leftrightarrow \exists i \in \{0,\dots,r\} \text{ such that } x \in \operatorname{Proj}_{\mathbb{R}^n}\left[\operatorname{gph}\|\cdot\|_0 \cap (\mathbb{R}^n \times \{i\})\right],\end{aligned}$$

which means that $\{x \in \mathbb{R}^n : \|x\|_0 \leq r\}$ is a finite union of intersections of semialgebraic sets, hence semialgebraic; see also [6, Formula 27(d)]. On the other hand, one has $\|x\|_\infty \leq l \Leftrightarrow \max_{1\leq i\leq n}(|x_i| - l) \leq 0$, which means that the box $[-l, l]^n$ is semialgebraic. Altogether, the set $D$, which is intersection of semialgebraic sets, is semialgebraic. Hence, when specified to the problem above, FRB converges to a stationary point thanks to Example 4.30.

We shall benchmark FRB against the Douglas-Rachford method with fixed step-size (DR) [22] by Li and Pong, inertial Tseng's method (iTseng) [12] by Boţ and Csetnek, and DR equipped with step-size heuristics (DRh) [22] by Li and Pong. These splitting algorithms for nonconvex optimization problems are known to converge globally to a stationary point of (7) under appropriate assumptions on the concave KL property of merit functions; see [22, Theorems 1–2, Remark 4, Corollary 1] and [12, Theorem 3.1]. The convergence of DR and iTseng in our setting are already proved in [22, Proposition 2] and [12, Corollary 3.1], respectively.

We implement FRB with the following specified scheme for the problem of finding an $r$-sparse solution of a linear system $\{x \in \mathbb{R}^n : Ax = b\}$

$$x_{k+1} \in \operatorname{Proj}_D\left(x_k - \lambda A^\dagger A(2x_k - x_{k-1}) + \lambda A^\dagger b\right)$$

where the step-size $\lambda = 0.9999 \cdot \frac{1}{3}$ (recall Example 4.30). The inertial type Tseng's method studied in [12, Scheme (6)] is applied with a step-size $\lambda' = 0.1316$ given by [12, Lemma

3.3] and a fixed inertial term $\alpha = \frac{1}{8}$:

$$p_{k+1} \in \mathrm{Proj}_D \left( x_k - \lambda' A^\dagger (Ax_k - b) + \alpha(x_k - x_{k-1}) \right),$$
$$x_{k+1} = p_k + \lambda' A^\dagger A(x_k - p_k).$$

As for DR and DRh, we employ the schemes specified by [22, Scheme (7)] and [22, Section 5] with the exact same step-sizes, respectively. All algorithms are initialized at the origin, and we terminate FRB and iTseng when

$$\frac{\max\left\{ \|x_{k+1} - x_k\|, \|x_k - x_{k-1}\| \right\}}{\max\left\{1, \|x_{k+1}\|, \|x_k\|, \|x_{k-1}\| \right\}} < 10^{-8}.$$

We adopt the termination criteria from [22, Section 5] for DR and DRh, where the same tolerance of $10^{-8}$ is applied. Similar to [22, Section 5], our problem data is generated through creating random matrices $A \in \mathbb{R}^{m\times n}$ with entries following the standard Gaussian distribution. Then we generate a vector $\hat{x} \in \mathbb{R}^r$ randomly with the same distribution, project it onto the box $[-10^6, 10^6]^r$, and create a sparse vector $\tilde{x} \in \mathbb{R}^n$ whose $r$ entries are chosen randomly to be the respective values of the projection of $\hat{x}$ onto $[-10^6, 10^6]^r$. Finally, we set $b = A\tilde{x}$ to guarantee $C \cap D \neq \varnothing$.

Results of our experiments are presented in Table 1 below. For each problem of the size $(m, n)$, we randomly generate 50 instances using the strategy described above, and report ceilings of the average number of iterations (iter), the minimal objective value at termination ($\mathrm{fval}_{\min}$), and the number of successes (succ). Here we say an experiment is "successful", if the objective function value at termination is less than $10^{-12}$, which means that the algorithm actually hits a global minimizer rather than just a stationary point. We observed the following:

- Among algorithms with a fixed step-size (FRB, DR, and iTseng), FRB outperforms the others in terms of both the number of iterations and successes, and it also has the smallest termination values. Moreover, it's worth noting that our simulation results align with Malitsky and Tam's observation that FRB converges faster than Tseng's method on a specific problem [26, Remark 2.8].
- DRh has the most number of successes and the best precision at termination (see $\mathrm{fval}_{\min}$), but tend to be slower on "easy" problems (large $m$).

Table 1: Comparing FRB to DR, iTseng, and DRh.

| Size | | FRB | | | DR | | | iTseng | | | DRh | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| m | n | iter | $\text{fval}_{\min}$ | succ | iter | $\text{fval}_{\min}$ | succ | iter | $\text{fval}_{\min}$ | succ | iter | $\text{fval}_{\min}$ | succ |
| 300 | 600 | 290 | 5.78001e-14 | 49 | 476 | 1.63979e-13 | 43 | 724 | 3.93939e-13 | 23 | 436 | 1.69111e-30 | 50 |
| 300 | 700 | 381 | 7.99399e-14 | 48 | 601 | 1.90854e-13 | 36 | 866 | 4.72653e-13 | 10 | 444 | 1.25125e-30 | 50 |
| 300 | 800 | 473 | 1.00221e-13 | 42 | 743 | 2.17886e-13 | 22 | 1054 | 7.52841e-13 | 7 | 448 | 8.06062e-31 | 50 |
| 300 | 900 | 577 | 1.20731e-13 | 34 | 890 | 2.57981e-13 | 21 | 1226 | 7.01903e-13 | 5 | 453 | 8.60352e-31 | 50 |
| 300 | 1000 | 662 | 1.35447e-13 | 31 | 954 | 3.0211e-13 | 11 | 1326 | 9.27902e-13 | 1 | 459 | 8.77713e-31 | 50 |
| 400 | 600 | 177 | 4.89578e-14 | 50 | 271 | 1.12518e-13 | 50 | 461 | 3.09084e-13 | 45 | 427 | 5.72429e-30 | 50 |
| 400 | 700 | 230 | 7.51594e-14 | 50 | 365 | 1.74449e-13 | 50 | 628 | 4.5344e-13 | 35 | 433 | 3.50651e-30 | 50 |
| 400 | 800 | 294 | 8.46721e-14 | 50 | 487 | 2.0469e-13 | 47 | 810 | 5.37112e-13 | 26 | 441 | 2.05666e-30 | 50 |
| 400 | 900 | 359 | 1.03652e-13 | 49 | 589 | 2.39422e-13 | 41 | 910 | 7.90857e-13 | 18 | 443 | 2.05578e-30 | 50 |
| 400 | 1000 | 431 | 1.00662e-13 | 47 | 712 | 2.40475e-13 | 35 | 1064 | 7.66054e-13 | 8 | 447 | 1.55654e-30 | 50 |
| 500 | 600 | 116 | 4.46448e-14 | 50 | 169 | 8.31312e-14 | 50 | 267 | 2.6074e-13 | 50 | 369 | 2.70253e-29 | 50 |
| 500 | 700 | 158 | 5.70864e-14 | 50 | 239 | 1.33359e-13 | 50 | 407 | 3.81561e-13 | 50 | 419 | 1.11142e-29 | 50 |
| 500 | 800 | 198 | 8.82954e-14 | 50 | 308 | 1.9478e-13 | 50 | 545 | 5.10359e-13 | 46 | 431 | 7.31337e-30 | 50 |
| 500 | 900 | 246 | 9.45053e-14 | 50 | 399 | 2.12284e-13 | 50 | 700 | 5.67092e-13 | 36 | 434 | 5.51452e-30 | 50 |
| 500 | 1000 | 294 | 1.25389e-13 | 50 | 478 | 2.76383e-13 | 47 | 847 | 6.55755e-13 | 17 | 436 | 4.44226e-30 | 50 |

# 4 A Bregman inertial forward-reflected-backward method

This section is based on preprint [36]. In the remainder of this section, suppose that the following hold:

**Assumption 4.1.** (i) $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper, lsc and prox-bounded with threshold $\lambda_f$.

(ii) $g : \mathbb{R}^n \to \mathbb{R}$ has Lipschitz continuous gradient with constant $L_{\nabla g}$.

(iii) $h : \mathbb{R}^n \to \mathbb{R}$ is $\sigma$-strongly convex and has Lipschitz continuous gradient with constant $L_{\nabla h}$.

(iv) The objective $F = f + g$ is bounded below.

*The goal of this section is to propose a Bregman inertial forward-reflected-backward method (Algorithm 4.2) for solving the nonconvex composite problem*

$$\min_{x\in\mathbb{R}^n} F(x) = f(x) + g(x). \tag{12}$$

Informally, the proposed algorithm computes subproblem

$$x_{k+1} \in (\nabla h + \lambda_k \partial f)^{-1}(x_k - 2\lambda_k \nabla g(x_k) + \lambda_k \nabla g(x_{k-1}) + \alpha_k(x_k - x_{k-1})), \tag{13}$$

for some stepsize $\lambda_k > 0$, inertial parameter $\alpha_k \in [0,1)$ and kernel $h : \mathbb{R}^n \to \mathbb{R}$. In particular, (13) reduces to the inertial forward-reflected-backward (FRB) method studied in [26, Corollary 4.4] when $h = \|\cdot\|^2/2$, $f, g$ are convex and inertial parameter is fixed. Before stating our main contribution, let us note that several splitting algorithms have been extended to nonconvex setting for solving (12); see, e.g., the Douglas-Rachford method [22], forward-backward method [3, 11], inertial forward-backward methods [13,31], the Peaceman-Rachford method [20], and inertial Tseng's method [12].

The *main contribution* now follows. Our convergence analysis relies on the generalized concave Kurdyka-Łojasiewicz property (Definition 2.3) and *a novel framework for analyzing a sequence of implicit merit functions (Definition 4.9) through a general condition (Assumption 4.11) on merit function parameters*. In turn, we derive *global sequential convergence to a stationary point of F with explicit stepsize condition that is independent of inertial parameter* whereas stepsize assumption in current literature may be tangled with inertial parameter or even implicit; see Remark 4.20(i). *Such independence result resolves a question of Malitsky and Tam [26, Section 7] regarding whether FRB can be adapted to incorporate a Nesterov-type acceleration; see Remark 4.20(ii). Convergence rate analysis on both function value and the actual sequence* is also carried out. Moreover, we *provide several formulae of the associated Bregman subproblem*, which to the best of our knowledge are the first of its kind among present results devoting to Bregman extension of splitting algorithms with the same kernel; see, e.g., [12, 13].

## 4.1 The proposed algorithm: BiFRB and its variants

We now propose the Bregman inertial forward-reflected-backward (BiFRB) algorithm.

**Algorithm 4.2** (BiFRB)**.** 1. Initialization: Pick $x_{-1}, x_0 \in \mathbb{R}^n$. Let $0 < \underline{\lambda} \leq \overline{\lambda} \leq \lambda_f$. Choose $\underline{\lambda} \leq \lambda_{-1} \leq \overline{\lambda}$.

2. For $k \in \mathbb{N}$, choose $\underline{\lambda} \leq \lambda_k \leq \overline{\lambda}$ and $0 \leq \alpha_k < 1$. Compute

$$y_k = x_k + \lambda_{k-1}\big(\nabla g(x_{k-1}) - \nabla g(x_k)\big), \tag{14}$$

$$x_{k+1} \in \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ f(x) + \langle x - y_k, \nabla g(x_k) + \frac{\alpha_k}{\lambda_k}(x_{k-1} - x_k)\rangle + \frac{1}{\lambda_k} D_h(x, y_k) \right\}. \tag{15}$$

Below is a systemic way to construct kernel $h$ via Fenchel conjugate and Moreau envelope.

**Proposition 4.3.** Let $L, \sigma > 0$ and let $p : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, lsc and convex. Define $q = \|\cdot\|^2/2$. Then the following hold:

(i) [5, Corollary 18.18] $p$ has $L$-Lipschitz gradient if and only if $p$ is the Moreau envelope with parameter $\lambda = 1/L$ of a proper lsc and convex function $(p^* - \lambda q)^*$. Therefore, in particular, if $p^*$ is $1/L$-strongly convex then $p$ has $L$-Lipschitz gradient.

(ii) If $p$ has $L$-Lipschitz gradient, then $h = p + \sigma q$ is $\sigma$-strongly convex with $(L + \sigma)$-Lipschitz gradient.

**Example 4.4.** Let $h(x) = \sqrt{1 + \|x\|^2} + \|x\|^2/2$. Then $h$ satisfies Assumption 4.1(iii) with $\sigma = 1$ and $L_{\nabla h} = 2$.

*Proof.* Define $(\forall x \in \mathbb{R}^n)$ $p(x) = \sqrt{1 + \|x\|^2}$. Then $p^*$ is 1-strongly convex; see, e.g., [7, Section 4.4, Example 5.29]. Applying Proposition 4.3 completes the proof. ∎

We also assume the following throughout this section, under which the proposed algorithm is well-defined.

**Assumption 4.5.** Suppose that one of the following holds:

(i) The kernel $h$ has strong convexity modulus $\sigma > 1$.

(ii) Function $f$ has prox-threshold $\lambda_f = \infty$.

**Remark 4.6.** The threshold $\lambda_f = \infty$ when $f$ is convex or bounded below.

**Proposition 4.7.** Let $u, v \in \mathbb{R}^n$ and suppose that Assumption 4.5 holds. Then for $0 < \lambda < \lambda_f$, the function $x \mapsto f(x) + \langle x - u, v\rangle + D_h(x,u)/\lambda$ is coercive. Consequently, the set

$$\operatorname*{argmin}_{x\in\mathbb{R}^n} \left\{ f(x) + \langle x - u, v\rangle + \frac{1}{\lambda} D_h(x,u) \right\} \neq \emptyset.$$

*Proof.* We claim that $x \mapsto f(x) + \langle x, \omega\rangle$ is prox-bounded with $\lambda_f$ for an arbitrary $\omega \in \mathbb{R}^n$. Indeed,

$$\liminf_{\|x\|\to\infty} \frac{f(x) + \langle x, \omega\rangle}{\|x\|^2} = \liminf_{\|x\|\to\infty} \frac{f(x)}{\|x\|^2} + \lim_{\|x\|\to\infty} \frac{\langle x, \omega\rangle}{\|x\|^2} = \liminf_{\|x\|\to\infty} \frac{f(x)}{\|x\|^2}.$$

Thus invoking [32, Excercise 1.24] proves the claim.

By our claim, $x \mapsto f(x) + \langle x, v - \sigma u/\lambda\rangle$ is prox-bounded with threshold $\lambda_f$. Let $\lambda'$ be such that $\lambda < \lambda' < \lambda_f$. Then $\lambda'/\sigma < \lambda_f$ under Assumption 4.5. Therefore

$$\begin{aligned}
&f(x) + \langle x - u, v\rangle + \frac{1}{\lambda} D_h(x,u) \geq f(x) + \langle x - u, v\rangle + \frac{\sigma}{2\lambda}\|x-u\|^2 - \frac{\sigma}{2\lambda'}\|x\|^2 + \frac{\sigma}{2\lambda'}\|x\|^2\\
&\geq \frac{\sigma}{2}\left(\frac{1}{\lambda} - \frac{1}{\lambda'}\right)\|x\|^2 + f(x) + \langle x, v - \frac{\sigma}{\lambda}u\rangle + \frac{\sigma}{2\lambda'}\|x\|^2 - \langle u, v\rangle\\
&\geq \frac{\sigma}{2}\left(\frac{1}{\lambda} - \frac{1}{\lambda'}\right)\|x\|^2 + \inf_{x\in\mathbb{R}^n}\left\{ f(x) + \langle x, v - \frac{\sigma}{\lambda}u\rangle + \frac{\sigma}{2\lambda'}\|x\|^2 \right\} - \langle u, v\rangle \to \infty
\end{aligned}$$

as $\|x\| \to \infty$, where the infimum in the last inequality is finite thanks to prox-boundedness of $x \mapsto f(x) + \langle x, v - \sigma u/\lambda\rangle$, which completes the proof. ■

Setting the kernel $h = \|\cdot\|/2$ gives Algorithm 4.8, which is an inertial forward-reflected-backward method (iFRB), whose behavior is studied by Malitsky and Tam with fixed stepsize and inertial parameter in the presence of convexity; see [26, Corollary 4.4].

**Algorithm 4.8** (iFRB)**.**

1. Initialization: Pick $x_{-1}, x_0 \in \mathbb{R}^n$. Let $0 < \underline{\lambda} \leq \overline{\lambda}$. Choose $\underline{\lambda} \leq \lambda_{-1} \leq \overline{\lambda}$.

2. For $k \in \mathbb{N}$, choose $\underline{\lambda} \leq \lambda_k \leq \overline{\lambda}$ and $0 \leq \alpha_k < 1$. Compute

$$y_k = x_k + \lambda_{k-1}\big(\nabla g(x_{k-1}) - \nabla g(x_k)\big),$$

$$x_{k+1} \in \operatorname*{argmin}_{x\in\mathbb{R}^n} \left\{ f(x) + \langle x - y_k, \nabla g(x_k) + \frac{\alpha_k}{\lambda_k}(x_{k-1} - x_k)\rangle + \frac{1}{2\lambda_k}\|x - y_k\|^2 \right\}.$$

Choosing $\alpha_k = 0$ and constant stepsize $\lambda_k$ in Algorithm 4.8, one recovers Algorithm 3.2.

## 4.2 Abstract function value convergence

In this subsection, we present function value convergence results of Algorithm 4.2 under an abstract condition (Assumption 4.11). Stepsize rules under which such condition holds are provided in Section 4.4.

In the remainder of this section, unless specified otherwise, let $p_{-1} \in \mathbb{R}$ and define for $k \in \mathbb{N}$

$$p_k = \left(\frac{(L_{\nabla h} - \sigma)L_{\nabla g}^2}{2}\right)\frac{\lambda_{k-1}^2}{\lambda_k} + \frac{1}{2}\left(\frac{\sigma}{\lambda_k} - L_{\nabla g}\right) - p_{k-1}, \tag{16}$$

$$M_{1,k} = p_{k-1} - \frac{\alpha_k + \sigma L_{\nabla g}\lambda_{k-1}}{2\lambda_k} - \frac{(L_{\nabla h} - \sigma)L_{\nabla g}^2\lambda_{k-1}^2}{2\lambda_k}, \tag{17}$$

where $(\lambda_k)_{k\in\mathbb{N}^*}$ is a stepsize sequence of BiFRB; $(\alpha_k)_{k\in\mathbb{N}}$ is a sequence of inertial parameters; and $L_{\nabla h}, L_{\nabla g}, \sigma$ are parameters specified in Assumption 4.1. The following concept plays a central role in our analysis.

**Definition 4.9.** Let $p \in \mathbb{R}$. Define $H_p : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ by

$$(\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n)\ H_p(x, y) = F(x) + p\,\|x - y\|^2 .$$

Then we say that $H_p$ is a quadratic merit function with parameter $p$. Let $(p_k)_{k\in\mathbb{N}^*}$ be given by (16). Then $H_{p_k}$ is the $k^{th}$ BiFRB merit function. If $(\forall k \in \mathbb{N}^*)$ $p_k \leq \bar{p}$ for some $\bar{p} > 0$, then $H_{\bar{p}}$ is a dominating BiFRB merit function.

**Remark 4.10.** Quadratic merit functions of various forms appear frequently in literature that devotes to extending splitting methods to nonconvex setting; see, e.g., [12, 13, 30, 31] and the references therein.

Despite the recursion (16), $(p_k)_{k\in\mathbb{N}^*}$ and $(M_{1,k})_{k\in\mathbb{N}^*}$ are indeed implicit in our analysis. Imposing the following novel condition on $(p_k)_{k\in\mathbb{N}^*}$ and $(M_{1,k})_{k\in\mathbb{N}^*}$, one gets abstract convergence of BiFRB, which in turn yields less restrictive stepsize rules; see Section 4.4.

**Assumption 4.11** (general rules of BiFRB merit function parameter)**.** Let $(p_k)_{k\in\mathbb{N}^*}$ and $(M_{1,k})_{k\in\mathbb{N}}$ be given as in (16) and (17). Suppose that the following hold:

(i) $\liminf_{k\to\infty} p_k \geq 0$ and $\liminf_{k\to\infty} M_{1,k} > 0$.

(ii) There exists $\bar{p}$ such that $p_k \leq \bar{p}$ for all $k$.

**Remark 4.12.** Assumption 4.11 generalizes quadratic merit function properties in the present literature; see, e.g., [12, Lemma 3.2], [13, Lemma 6], [30, Assumption (H1)] and [31, Algorithm 5, Lemma 4.6, Proposition 4.7].

**Lemma 4.13.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by BiFRB and define $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$. Then the following hold:

(i) For all $k \in \mathbb{N}$,

$$H_{p_k}(z_k) \leq H_{p_{k-1}}(z_{k-1}) - M_{1,k} \|z_k - z_{k-1}\|^2. \tag{18}$$

(ii) Suppose that Assumption 4.11(i) holds. Then the sequence $\left(H_{p_k}(z_k)\right)_{k\in\mathbb{N}}$ is decreasing and $\sum_{k=0}^{\infty} M_{1,k} \|z_k - z_{k-1}\|^2 < \infty$. Consequently, $\left(H_{p_k}(z_k)\right)_{k\in\mathbb{N}}$ is convergent and $\lim_{k\to\infty} \|z_k - z_{k-1}\| = 0$.

**Lemma 4.14.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by BiFRB, $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$, and $p \geq 0$. For $k \in \mathbb{N}$, define $u_k = (\nabla h(y_k) - \nabla h(x_{k+1}))/\lambda_k - \nabla g(x_k) + \alpha_k(x_k - x_{k-1})/\lambda_k$,

$$A_k = u_k + \nabla g(x_{k+1}) + 2p(x_{k+1} - x_k), B_k = 2p(x_k - x_{k+1}).$$

Set

$$M_{2,k} = \sqrt{2} \max\left\{\frac{L_{\nabla h}}{\lambda_k} + L_{\nabla g} + 6p, \frac{L_{\nabla h} L_{\nabla g} \lambda_{k-1} + 1}{\lambda_k}\right\},$$

and

$$M_2 = \sqrt{2} \max\left\{\frac{L_{\nabla h}}{\underline{\lambda}} + L_{\nabla g} + 6p, \frac{L_{\nabla h} L_{\nabla g} \overline{\lambda} + 1}{\underline{\lambda}}\right\}.$$

Then $(A_k, B_k) \in \partial H_p(z_k)$ and $\|(A_k, B_k)\| \leq M_{2,k} \|z_k - z_{k-1}\| \leq M_2 \|z_k - z_{k-1}\|$.

Having established basic properties of $\left(H_{p_k}(x_{k+1}, x_k)\right)$, we now show its convergence. Denote by $\omega(z_{-1})$ the set of all limit points of $(z_k)_{k\in\mathbb{N}^*}$.

**Theorem 4.15** (function value convergence)**.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by BiFRB and define $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$. Suppose that Assumption 4.11 holds and $(z_k)_{k\in\mathbb{N}^*}$ is bounded. Let $(z_{k_l})_{l\in\mathbb{N}}$ be a subsequence such that $z_{k_l} \to z^*$ for some $z^* = (x^*, y^*)$ as $l \to \infty$. Then the following hold:

(i) We have $\lim_{l\to\infty} H_{p_{k_l}}(z_{k_l}) = F(x^*)$. In fact, $\lim_{k\to\infty} H_{p_k}(z_k) = F(x^*)$. Consequently $\lim_{k\to\infty} H_{\bar{p}}(z_k) = F(x^*)$, where $\bar{p}$ is given in Assumption 4.11(ii).

(ii) The limit point $z^* = (x^*, y^*)$ satisfies $x^* = y^*$ and $0 \in \partial H_{\bar{p}}(x^*, y^*)$, which implies that $0 \in \partial F(x^*)$.

(iii) The set $\omega(z_{-1})$ is nonempty, connected, and compact, on which the function $H_{\bar{p}}$ is constant $F(x^*)$. Moreover, $\lim_{k\to\infty} \operatorname{dist}(z_k, \omega(z_{-1})) = 0$.

The result below provides a sufficient condition to the critical boundedness assumption in Theorem 4.15.

**Theorem 4.16.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by BiFRB and define $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$. Suppose that Assumption 4.11(i) holds. If $f + g$ is coercive, then the sequence $(x_k)_{k\in\mathbb{N}^*}$ is bounded, so is $(z_k)_{k\in\mathbb{N}^*}$.

## 4.3 Abstract sequential convergence

Now we turn to sequential convergence, which relies on the generalized concave KL property (recall Definition 2.3) and Assumption 4.11. For $\varepsilon > 0$ and nonempty set $\Omega \subseteq \mathbb{R}^n$, we define $\Omega_\varepsilon = \{x \in \mathbb{R}^n : \operatorname{dist}(x, \Omega) < \varepsilon\}$.

**Theorem 4.17** (global sequential convergence)**.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by BiFRB and define $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$. Suppose that Assumption 4.11 holds and $(z_k)_{k\in\mathbb{N}^*}$ is bounded. Suppose further that the dominating BiFRB merit function $H_{\bar{p}}$ (recall Definition 4.9) has the generalized concave KL property on $\omega(z_{-1})$. Then the following hold:

(i) The sequence $(z_k)_{k\in\mathbb{N}^*}$ is Cauchy and has finite length. Thus $z_k \to z^*$ for some $z^*$. To be specific, there exist index $k_0 \in \mathbb{N}$, $\varepsilon > 0$ and $\eta \in (0, \infty]$ such that for $i \geq k_0 + 1$,

$$\sum_{k=i}^{\infty} \|z_{k+1} - z_k\| \leq \|z_i - z_{i-1}\| + C\tilde{\varphi}\left(H_{p_i}(z_i) - F(x^*)\right), \tag{19}$$

where

$$C = \frac{2\sqrt{2}\max\left\{\frac{L_{\nabla h}}{\underline{\lambda}} + L_{\nabla g} + 6\bar{p}, \frac{L_{\nabla h}L_{\nabla g}\overline{\lambda}+1}{\underline{\lambda}}\right\}}{\liminf_{k\to\infty} M_{1,k}},$$

and $\tilde{\varphi}$ is the exact modulus associated with the generalized concave KL property of $H_{\bar{p}}$ at $z^*$ with respect to $\varepsilon$ and $\eta$.

(ii) The sequence $(x_k)_{k\in\mathbb{N}^*}$ has finite length and converges to some $x^*$ with $0 \in \partial F(x^*)$.

*Proof.* (i) See [36, Theorem 3.20(i)]. (ii) Apply the definition of $(z_k)_{k\in\mathbb{N}^*}$ and Theorem 4.15(ii). ∎

The generalized concave KL assumption in Theorem 4.17 can be easily satisfied by semialgebraic functions.

**Corollary 4.18.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated BiFRB. Assume that $(x_k)_{k\in\mathbb{N}^*}$ is bounded and Assumption 4.11 holds. Suppose further that both $f$ and $g$ are semialgebraic (recall Definition 2.5). Then $(x_k)_{k\in\mathbb{N}^*}$ converges to some $x^*$ with $0 \in \partial F(x^*)$ and has finite length property.

*Proof.* Let $\bar{p}$ be as in Assumption 4.11. Then $(x, y) \mapsto \bar{p}\|x - y\|^2$ is semialgebraic. In turn, the dominating BiFRB function $H_{\bar{p}}$ (recall Definition 4.9) is semialgebraic by the semialgebraic functions sum rule; see, e.g., [2, Section 4.3]. The desired result then follows immediately from Theorem 4.17. ■

## 4.4 Stepsize rules

Following abstract convergence properties of BiFRB, we now exploit conditions such that the important Assumption 4.11 holds. To furnish Assumption 4.11, it suffices to establish the existence of suitable $p_{-1}$ under desirable stepsize rules. *Unlike most literature that emphasizes on explicit merit function parameters [12,13,31,37], we shall see soon that our $(p_k)_{k\in\mathbb{N}^*}$ and $(M_{1,k})_{k\in\mathbb{N}}$ are implicit.* This turns out to be instrumental in designing less restrictive stepsize rules.

In the remainder of this subsection, unless specified otherwise,

$$a = (L_{\nabla h} - \sigma)L_{\nabla g}^2, b = \sigma, c = L_{\nabla g}. \tag{20}$$

Consequently, the sequence $(p_k)_{k\in\mathbb{N}^*}$ and $(M_{1,k})_{k\in\mathbb{N}^*}$ given by (16) and (17) satisfy

$$p_k = \frac{a}{2}\left(\frac{\lambda_{k-1}^2}{\lambda_k}\right) + \frac{1}{2}\left(\frac{b}{\lambda_k} - c\right) - p_{k-1}, \; M_{1,k} = p_{k-1} - \frac{\alpha_k + bc\lambda_{k-1}}{2\lambda_k} - \frac{a\lambda_{k-1}^2}{2\lambda_k}. \tag{21}$$

### 4.4.1 Non-Euclidean case

**Proposition 4.19** (fixed stepsize)**.** Let $(\lambda_k)_{k\in\mathbb{N}^*}$ be a constant BiFRB stepsize sequence with $(\forall k \in \mathbb{N}^*)\ \lambda_k = \lambda > 0$. Suppose that $\sigma > 2$, $(L_{\nabla h} - \sigma)\sigma > 1/4$ and $(\forall k \in \mathbb{N}^*)\ 0 \leq \alpha_k < \min(1, \sigma/2) = 1$. Define

$$\lambda^* = \frac{\sqrt{(2bc + c)^2 + 4a(b-2)} - 2bc - c}{2a} > 0,$$

where $a, b, c$ are specified by (20). If $\lambda < \min\{\lambda^*, (\sigma - 1)/[(\sigma + 1)L_{\nabla g}]\}$, then there exists $p_{-1}$ such that $(p_k)_{k\in\mathbb{N}^*}$ and $(M_{1,k})_{k\in\mathbb{N}}$ generated by (16) and (17) satisfy Assumption 4.11.

**Remark 4.20.** (i) (Comparison to known results) Compared to a result of Boţ and Csetnek [12, Lemma 3.3] with explicit merit function parameter but implicit stepsize bounds, our stepsize is explicit and independent of inertial parameter; see [13, Remark 7] for a similar restrictive result and [26, Corollary 4.4] for a result with dependent stepsize bound and inertial parameter in the presence of convexity; see also [29, Table 1].

(ii) (Regarding an open question of Malitsky and Tam) In [26, Section 7], Malitsky and Tam posted a question regarding whether FRB can be adapted to incorporate a Nesterov-type acceleration. With inertial parameter $\alpha_k \in [0,1)$ independent of stepsize $\lambda$ in Proposition 4.19, as opposed to $\alpha_k \in [0,1/3)$ dependent of $\lambda$ in [26, Corollary 4.4], this question is resolved. Indeed, the Nesterov acceleration scheme corresponds to Proposition 4.19 with

$$(\forall k \in \mathbb{N}^*)\ \alpha_k = \frac{t_k - 1}{t_{k+1}},$$

where $(\forall k \in \mathbb{N}^*)\ t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2$ and $t_{-1} = 1$; see, e.g., [28].

(iii) The assumption that $(L_{\nabla h} - \sigma)\sigma > 1/4$ is not demanding. For instance, let $\alpha, \beta > 0$ be such that $\alpha\beta > 1/4$ and let $u : \mathbb{R}^n \to \mathbb{R}$ be a function with 1-Lipschitz gradient. Then $h = \alpha u + \beta \|\cdot\|^2 /2$ satisfies $(L_{\nabla h} - \sigma)\sigma > 1/4$.

### 4.4.2 Euclidean case

**Proposition 4.21** (dynamic stepsize)**.** Let $(\lambda_k)_{k\in\mathbb{N}^*} \subseteq [\underline{\lambda}, \overline{\lambda}]$ be a iFRB stepsize sequence and let $(p_k)_{k\in\mathbb{N}^*}$ be the sequence generated by (21) with $a = 0$, $b = 1$ and $0 \leq \alpha_k < \bar{\alpha}$ for some $\bar{\alpha} < 1/2$. Let $(a_k)_{k\in\mathbb{N}}$ be a sequence of positive real numbers with $\sum_{k=0}^{\infty} a_k < \infty$ and let $0 < \varepsilon < (1 - 2\bar{\alpha})/(3L_{\nabla g})$. Suppose that the following hold:

(i) The stepsize lower bound $\underline{\lambda} = \varepsilon$ and upper bound $\overline{\lambda} < \varepsilon/(2\bar{\alpha} + 3\varepsilon L_{\nabla g})$.

(ii) For all $k \in \mathbb{N}$, $0 \leq (1/\lambda_k) - (1/\lambda_{k-1}) \leq a_k$.

Then there exists $p_{-1}$ such that $(p_k)_{k\in\mathbb{N}^*}$ and $(M_{1,k})_{k\in\mathbb{N}}$ generated by (16) and (17) satisfy Assumption 4.11.

**Corollary 4.22** (fixed stepsize)**.** Let $(\lambda_k)_{k\in\mathbb{N}^*}$ be a constant iFRB stepsize sequence with $\lambda_k = \lambda > 0$ for all $k$ and let $(p_k)_{k\in\mathbb{N}^*}$ be the sequence generated by (21) with $a = 0$, $b = 1$ and $0 \leq \alpha_k < \bar{\alpha}$ for some $\bar{\alpha} < 1/2$. If $\lambda < (1 - 2\bar{\alpha})/(3L_{\nabla g})$, then there exists $p_{-1}$ such that $(p_k)_{k\in\mathbb{N}^*}$ and $(M_{1,k})_{k\in\mathbb{N}}$ generated by (16) and (17) satisfy Assumption 4.11.

## 4.5 Convergence rates

**Lemma 4.23.** *[14, Lemma 10]* Let $(e_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}_+$ be a decreasing sequence converging 0. Assume further that there exist natural numbers $k_0 \geq l_0 \geq 1$ such that for every $k \geq k_0$,

$$e_{k-l_0} - e_k \geq C_e e_k^{2\theta}, \tag{22}$$

where $C_e > 0$ is some constant and $\theta \in [0, 1)$. Then the following hold:

(i) If $\theta = 0$, then $(e_k)_{k \in \mathbb{N}}$ converges in finite steps.

(ii) If $\theta \in (0, 1/2]$, then there exists $C_{e,0} > 0$ such that for every $k \geq k_0$,

$$0 \leq e_k \leq C_{e,0} Q^k.$$

(iii) If $\theta \in (1/2, 1)$, then there exists $C_{e,1} > 0$ such that for every $k \geq k_0 + l_0$,

$$0 \leq e_k \leq C_{e,1}(k - l_0 + 1)^{-\frac{1}{2\theta - 1}}.$$

We say a generalized concave KL function has KL exponent $\theta \in [0, 1)$, if an associated desingularizing function $\varphi(t) = c \cdot t^{1-\theta}$ for some $c > 0$. It turns out that convergence rates of BiFRB are governed by the KL exponent.

**Theorem 4.24** (function value convergence rate)**.** Let $(x_k)_{k \in \mathbb{N}^*}$ be a sequence generated by BiFRB and define $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$. Suppose that all assumptions in Theorem 4.17 are satisfied and let $x^*$ be the limit given in Theorem 4.17(ii). Suppose further that the dominating BiFRB merit function $H_{\bar{p}}$ (recall Definition 4.9) has the generalized concave KL property at $z^* = (x^*, x^*)$ with KL exponent $\theta \in [0, 1)$. Define $e_k = H_{p_k}(z_k) - F(x^*)$ for $k \in \mathbb{N}$. Then $(e_k)_{k \in \mathbb{N}}$ converges to 0 and the following hold:

(i) If $\theta = 0$, then $(e_k)_{k \in \mathbb{N}}$ converges in finite steps.

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

$$e_k \leq \hat{c}_1 \hat{Q}_1^k.$$

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

$$e_k \leq \hat{c}_2 k^{-\frac{1}{2\theta - 1}}.$$

*Proof.* Recall from Lemma 4.13(ii) and Theorem 4.15(i) that the sequence $(e_k)_{k \in \mathbb{N}}$ is decreasing and converges to 0. By the KL exponent assumption, there exist $c > 0$ and $k_0$ such that for $k \geq k_0$

$$\operatorname{dist}\left(0, \partial H_{\bar{p}}(z_k)\right) \geq \frac{\left(H_{p_k}(z_k) - H_{\bar{p}}(z^*)\right)^\theta}{c(1-\theta)} = \frac{e_k^\theta}{c(1-\theta)}.$$

Assume without loss of generality that $M_{1,k} \geq \liminf_{k \to \infty} M_{1,k}/2$. So Lemmas 4.13 and 4.14 imply

$$e_{k-1} - e_k = H_{p_{k-1}}(z_{k-1}) - H_{p_k}(z_k) \geq M_{1,k} \|z_k - z_{k-1}\|^2 \geq \frac{M_{1,k}}{M_{2,k}^2} \operatorname{dist}(0, \partial H_{\bar{p}}(z_k))^2 \geq C_e e_k^{2\theta},$$

where

$$C_e = \frac{\liminf_{k\to\infty} M_{1,k}}{4c^2(1-\theta)^2 \max\left\{ \left(\frac{L_{\nabla h}}{\underline{\lambda}} + L_{\nabla g} + 6\bar{p}\right)^2, \left(\frac{L_{\nabla h} L_{\nabla g}\overline{\lambda}+1}{\underline{\lambda}}\right)^2 \right\}}.$$

Applying Lemma 4.23 completes the proof. ■

**Theorem 4.25** (sequential convergence rate)**.** Let $(x_k)_{k\in\mathbb{N}^*}$ be a sequence generated by BiFRB and define $z_k = (x_{k+1}, x_k)$ for $k \in \mathbb{N}^*$. Suppose that all assumptions in Theorem 4.17 are satisfied and let $x^*$ be the limit given in Theorem 4.17(ii). Suppose further that the dominating BiFRB merit function $H_{\bar{p}}$ (recall Definition 4.9) has the generalized concave KL property at $z^* = (x^*, x^*)$ with KL exponent $\theta \in [0, 1)$. Then the following hold:

(i) If $\theta = 0$, then $z_k$ converges $z^*$ in finite steps.

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

$$\|z_k - z^*\| \le c_1 Q_1^k.$$

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

$$\|z_k - z^*\| \le c_2 k^{-\frac{1-\theta}{2\theta-1}}.$$

*Proof.* For simplicity, define for $k \in \mathbb{N}^*$

$$e_k = H_{p_k}(z_k) - F(x^*), \delta_k = \sum_{i=k}^{\infty} \|z_{i+1} - z_i\|.$$

Then $e_k \to 0$ and $\|z_k - z^*\| \le \delta_k$. Without loss of generality (recall Lemma 4.13), assume that $e_k \in [0, 1)$ and $e_{k+1} \le e_k$ for $k \in \mathbb{N}^*$. To obtain the desired results, it suffices to estimate $\delta_k$.

Assume without loss of generality that $M_{1,k} \ge \liminf_{k\to\infty} M_{1,k}/2$. Invoking Lemma 4.13(i) and Theorem 4.15(i) yields that $(\liminf_{k\to\infty} M_{1,k}/2)\|z_k - z_{k-1}\|^2 \le M_{1,k}\|z_k - z_{k-1}\| \le H_{p_{k-1}}(z_{k-1}) - H_{p_k}(z_k) \le H_{p_{k-1}}(z_{k-1}) - F(x^*) = e_{k-1}$, which in turn implies that

$$\|z_k - z_{k-1}\| \le d\sqrt{e_{k-1}}, \tag{23}$$

where $d = \sqrt{2/\liminf_{k\to\infty} M_{1,k}}$. Moreover, recall from Theorem 4.17(i) that there exists $k_0$ such that for $k \ge k_0 + 1$

$$\delta_k \le \|z_k - z_{k-1}\| + C\tilde{\varphi}(e_k) \le \|z_k - z_{k-1}\| + C\tilde{\varphi}(e_{k-1}), \tag{24}$$

where the last inequality holds because $e_k \le e_{k-1}$. Assume without loss of generality that $H_{\bar{p}}$ is associated with desingularizing $\varphi(t) = C^{-1}t^{1-\theta}$ due to the KL exponent assumption. Thus, combined with (23), inequality (24) yields

$$\delta_k \le d\sqrt{e_{k-1}} + e_{k-1}^{1-\theta}. \tag{25}$$

Case 1 ($\theta = 0$): Appealing to Theorem 4.24, the sequence $e_k$ converges to 0 in finite steps. Thus (25) implies $\delta_k \to 0$ in finite steps, so does $(z_k)_{k\in\mathbb{N}^*}$.

Case 2 ($0 < \theta \le 1/2$): Clearly $1/2 \le 1-\theta < 1$, therefore $\sqrt{e_{k-1}} \ge e_{k-1}^{1-\theta}$. Inequality (25) and Theorem 4.24 imply that for $k$ sufficiently large

$$\delta_k \le d\sqrt{e_{k-1}} + \sqrt{e_{k-1}} = (1+d)\sqrt{e_{k-1}} \le (1+d)\sqrt{\hat{c}_1\hat{Q}_1^{k-1}}.$$

Case 3 ($1/2 < \theta < 1$): Note that $\sqrt{e_{k-1}} \le e_{k-1}^{1-\theta}$. Then appealing to (25) and Theorem 4.24 yields for $k$ sufficiently large

$$\delta_k \le de_{k-1}^{1-\theta} + e_{k-1}^{1-\theta} = (1+d)e_{k-1}^{1-\theta} \le (1+d)(k-1)^{-\frac{1-\theta}{2\theta-1}},$$

from which the desired result readily follows. ■

### 4.6 Bregman proximal subproblem formulae

Set $\alpha \ge 0, \beta > 0$. Fix $\omega \in \mathbb{R}^n$ and $\lambda > 0$. In the remainder of this section, let

$$h(x) = \alpha\sqrt{1+\|x\|^2} + \frac{\beta}{2}\|x\|^2.$$

Define for $u \in \mathbb{R}^n$, $p_\lambda(u) = \lambda\omega - \nabla h(u)$ and

$$T_\lambda(u) = \operatorname*{argmin}_{x\in\mathbb{R}^n}\left\{f(x) + \langle x-u,\omega\rangle + \frac{1}{\lambda}D_h(x,u)\right\} \tag{26}$$

$$= \operatorname*{argmin}_{x\in\mathbb{R}^n}\left\{f(x) + \langle x,\omega\rangle + \frac{1}{\lambda}h(x) - \frac{1}{\lambda}\langle\nabla h(u),x\rangle\right\}$$

$$= \operatorname*{argmin}_{x\in\mathbb{R}^n}\left\{\lambda f(x) + \langle x,p_\lambda(u)\rangle + h(x)\right\}. \tag{27}$$

Clearly Assumption 4.5(ii) holds with $\sigma = \beta$ and $L_{\nabla h} = \alpha + \beta$ and the Bregman proximal subproblem (15) corresponds to (26) with $\omega = \alpha_k(x_{k-1} - x_k)/\lambda_k + \nabla g(x_k)$, $u = y_k$ and $\lambda = \lambda_k$.

*In this subsection, we present formulae of $T_\lambda(u)$ with the kernel h chosen as above, supplementing not only Algorithm 4.2 but also [12, Algorithm 3.1] and [13, Algorithm 1].*

#### 4.6.1 $l_0$-constrained problems

Let $D = \{x \in \mathbb{R}^n : \|x\|_0 \le r, \|x\| \le R\}$ for integer an $r \le n$ and real number $R > 0$, where $(\forall x \in \mathbb{R}^n)$ $\|x\|_0 = \sum_{i=1}^n |\operatorname{sgn}(x_i)|$.[2] In this subsection, consider minimization problem

$$\min_{x \in D} g(x),$$

which corresponds to (12) with $f = \delta_D$. Recall that the Hard-threshold of $x \in \mathbb{R}^n$ with parameter $r$, denoted by $H_r(x)$, is given by

$$(\forall i \in \{1,\dots,n\}) \ \big(H_r(x)\big)_i = \begin{cases} x_i, & \text{if } |x_i| \text{ belongs to the } r \text{ largest among } |x_1|,\dots,|x_n|, \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

Immediately, $H_r(cx) = cH_r(x)$ for $c \neq 0$. The following lemma will be instrumental.

**Lemma 4.26.** *[25, Proposition 4.3]* Given $0 \neq p \in \mathbb{R}^n$ and a positive integer $r < n$, we have

$$\max_{x \in \mathbb{R}^n} \left\{ \langle p, x \rangle : \|x\| = 1, \ \|x\|_0 \le r \right\} = \|H_r(p)\|,$$

with optimal value attained at $x^* = \|H_r(p)\|^{-1} H_r(p)$.

**Proposition 4.27** (subproblem formula)**.** Let $f = \delta_D$, where $D = \{x \in \mathbb{R}^n : \|x\|_0 \le r, \|x\| \le R\}$ for integer an $r \le n$ and real number $R > 0$. Define

$$x^* = \begin{cases} 0, & \text{if } \nabla h(u) = \lambda\omega, \\ -t^* \left\|H_r\big(p_\lambda(u)\big)\right\|^{-1} H_r\big(p_\lambda(u)\big), & \text{if } \nabla h(u) \neq \lambda\omega, \end{cases} \tag{29}$$

where $t^*$ is the unique solution of

$$\min_{0 \le t \le R} \left\{ \varphi(t) = \alpha\sqrt{1+t^2} + \frac{\beta}{2}t^2 - \|H_r(p_\lambda(u)\| \, t \right\}, \tag{30}$$

which satisfies $t^* = R$ if $\|H_r(p_\lambda(u))\| \ge \alpha(1+R^2)^{-1/2} + \beta R$; otherwise $t^* \in (0, R)$ is the unique solution of

$$\alpha(1+t^2)^{-1/2} t + \beta t - \|H_r(p_\lambda(u))\| = 0.$$

Then $x^* \in T_\lambda(u)$.

*Proof.* Set $p = p_\lambda(u)$ for simplicity. If $\nabla h(u) = \lambda\omega$, then $p = 0$ and clearly $T_\lambda(u) = \operatorname*{argmin}_{x \in \mathbb{R}^n}\{h(x) : \|x\|_0 \le r, \|x\| \le R\} = 0$. Now we consider $\nabla h(u) \neq \lambda\omega$, in which case $p \neq 0$. Invoking Lemma 4.26, one has for $t > 0$

$$\min_{x \in \mathbb{R}^n}\{\langle x, p\rangle : \|x\|_0 \le r, \|x\| = t\} = -\max_{x \in \mathbb{R}^n}\{\langle x/t, -tp\rangle : \|x/t\| \le r, \|x/t\| = 1\}$$

[2] We set $\operatorname{sgn}(0) = 0$.

$$= -\left\|H_r(-tp)\right\| = -t\left\|H_r(p)\right\|, \tag{31}$$

attained at

$$\hat{x} = \hat{x}(t) = t\left\|H_r(-tp)\right\|^{-1}H_r(-tp) = -t\left\|H_r(p)\right\|^{-1}H_r(p). \tag{32}$$

It is easy to see that optimal value (31) and solution (32) still hold when $t = 0$.

Next we show that $x^* \in T_\lambda(u)$. By (27), $T_\lambda(u)$ is the solution set to

$$\min_{x\in\mathbb{R}^n}\left\{\langle x, p\rangle + \alpha\sqrt{1+\|x\|^2} + \frac{\beta}{2}\|x\|^2 : \|x\|_0 \le r, \|x\| \le R\right\}. \tag{33}$$

Consider an arbitrary feasible $x$ of (33) and let $t_x = \|x\|$. Then $0 \le t_x \le R$ and

$$\begin{aligned}\langle p, x^*\rangle + \alpha\sqrt{1+\|x^*\|^2} + \frac{\beta}{2}\|x^*\|^2 &= -t^*\left\|H_r(p)\right\| + \alpha\sqrt{1+(t^*)^2} + \frac{\beta}{2}(t^*)^2 \\ &= \varphi(t^*) \le \varphi(t_x) = -\left\|H_r(p)\right\| t_x + \alpha\sqrt{1+t_x^2} + \frac{\beta}{2}t_x^2 \\ &\le \langle p, x\rangle + \alpha\sqrt{1+\|x\|^2} + \frac{\beta}{2}\|x\|^2,\end{aligned}$$

where the first and last inequalities hold due to (30) and (31) respectively.

Finally, we turn to (30). Noticing the strong convexity of $\varphi$ and applying optimality condition, one gets

$$-\alpha(1+t^2)^{-1/2}t - \beta t + \left\|H_r(p)\right\| = -\varphi'(t^*) \in N_{[0,R]}(t^*),$$

So

$$t^* = R \Leftrightarrow -\varphi'(R) \in N_{[0,R]}(R) = [0,\infty) \Leftrightarrow \varphi'(R) \le 0 \Leftrightarrow \left\|H_r(p)\right\| \ge \alpha(1+R^2)^{-1/2} + \beta R,$$

and

$$t^* = 0 \Leftrightarrow -\varphi'(0) \in N_{[0,R]}(0) = (-\infty, 0] \Leftrightarrow \left\|H_r(p)\right\| \le 0 \Leftrightarrow H_r(p) = 0,$$

which never occurs when $p \ne 0$. If $0 < \left\|H_r(p)\right\| < \alpha(1+R^2)^{-1/2} + \beta R$, then $\varphi'(0) < 0$ and $\varphi'(R) > 0$, meaning that there exists unique $t^* \in (0, R)$ such that $\varphi'(t^*) = 0$ due to the strong convexity of $\varphi$. ■

#### 4.6.2 Problems penalized by a positively homogeneous convex function

In this subsection, we provide subproblem formula for problems of the form

$$\min_{x\in\mathbb{R}^n} f(x) + g(x),$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex and positively homogeneous in the sense that $(\forall x \in \mathbb{R}^n)$ $(\forall t > 0)$ $f(tx) = tf(x)$, which is a special case of problem (12). We also suppose that the proximal mapping $f$ is known.

**Proposition 4.28.** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, lsc and convex. Suppose that $f$ is positively homogeneous. Then the following hold:

(i) $(\forall x \in \operatorname{dom} \partial f)$ $(\forall t > 0)$ $\partial f(x) = \partial f(tx)$.

(ii) (subproblem formula)$(\forall u \in \mathbb{R}^n)$ $T_\lambda(u) = t^* \operatorname{Prox}_{\lambda f}(-p_\lambda(u))$, where $t^*$ is the unique root of the equation

$$1 - \alpha t\big(1 + t^2 \left\|\operatorname{Prox}_{\lambda f}(-p_\lambda(u))\right\|\big)^{-1/2} - \beta t = 0.$$

*Proof.* (i) $v \in \partial f(x) \Leftrightarrow (\forall y \in \mathbb{R}^n)$ $f(y) \geq f(x) + \langle v, y - x\rangle \Leftrightarrow (\forall y \in \mathbb{R}^n)$ $(\forall t > 0)$ $f(ty) \geq f(tx) + \langle v, ty - tx\rangle \Leftrightarrow v \in \partial f(tx)$.

(ii) For simplicity, let $p = p_\lambda(u)$. Note that $T_\lambda(p)$ is single-valued by strong convexity. Then invoking optimality condition and subdifferential sum rule yields

$$x^* = T_\lambda(u) \Leftrightarrow 0 \in \lambda\partial f(x^*) + p + \nabla h(x) = \lambda\partial f(x^*) + p + [\alpha(1 + \|x^*\|^2)^{-1/2} + \beta]x^*.$$

Define $t^* = [\alpha(1 + \|x^*\|^2)^{-1/2} + \beta]^{-1}$ and $v = x^*/t^*$. Then $t^* > 0$ and the above inclusion together with statement(i) implies that

$$\begin{aligned} 0 \in \lambda\partial f(x^*) + p + v &= \lambda\partial f(t^* v) + p + v = \lambda\partial f(v) + v + p \\ \Leftrightarrow -p \in (\operatorname{Id} + \lambda\partial f)(v) &\Leftrightarrow v = (\operatorname{Id} + \lambda\partial f)^{-1}(-p) = \operatorname{Prox}_{\lambda f}(-p). \end{aligned}$$

In turn, $t^*$ satisfies

$$v[1 - \alpha t^*(1 + (t^*)^2 \|v\|^2)^{-1/2} - \beta t^*] = v - [\alpha(1 + (t^*)^2 \|v\|^2)^{-1/2} + \beta]t^* v = v - \frac{x^*}{t^*} = v - v = 0.$$

If $v \neq 0$, then clearly $1 - \alpha t^*(1 + (t^*)^2 \|v\|^2)^{-1/2} - \beta t^* = 0$ as claimed; otherwise $x^* = 0$ and $t^* = 1/(\alpha + \beta)$, which is consistent with statement(ii). To see the uniqueness of $t^*$, let $(\forall t \in \mathbb{R})$ $\varphi(t) = \alpha(1 + \|v\|^2 t^2)^{1/2}/\|v\|^2 + \beta t^2/2 - t$, which is strongly convex. Simple calculus shows that $t^*$ is a stationary point of $\varphi$, thus unique. ■

Recall that the Soft-threshold of $x \in \mathbb{R}^n$ with parameter $\lambda$ is $S_\lambda(x) = \operatorname{Prox}_{\lambda\|\cdot\|_1}(x)$, which satisfies

$$(\forall i \in \{1, \ldots, n\})\ (S_\lambda(x))_i = \max(|x_i| - \lambda, 0)\operatorname{sgn}(x_i),$$

see, e.g., [7, Example 6.8].

**Corollary 4.29** ($l_1$-penalized problem)**.** Let $f = \|\cdot\|_1$. Then $T_\lambda(u) = -t^* S_\lambda\left(p_\lambda(u)\right)$, where $t^*$ is the unique solution of $1 - \alpha t\left(1 + t^2 \|S_\lambda\left(p_\lambda(u)\right)\|^2\right)^{-1/2} - \beta t = 0$.

## 4.7 Application to nonconvex feasibility problems

Let $C, D \subseteq \mathbb{R}^n$ be nonempty closed sets such that $C \cap D \neq \emptyset$. Consider

$$\min_{x \in D} \frac{1}{2} \operatorname{dist}^2(x, C), \tag{34}$$

which corresponds to (12) with $f = \delta_D$ and $g = \operatorname{dist}^2(\cdot, C)/2$. A global minimizer $x^*$ of (34) satisfies $x^* \in C \cap D$.

### 4.7.1 Convergence guarantee

**Proposition 4.30** (nonconvex feasibility problem)**.** Let $C \subseteq \mathbb{R}^n$ be a nonempty, closed and convex set and let $D \subseteq \mathbb{R}^n$ be nonempty and closed. Suppose that $C \cap D \neq \emptyset$ and either $C$ or $D$ is compact. Assume further that both $C$ and $D$ are semialgebraic. Consider the minimization problem (12) with $f = \delta_D$ and $g = \operatorname{dist}^2(\cdot, C)/2$. Let $(x_k)_{k \in \mathbb{N}^*}$ be a sequence generated by BiFRB. Suppose that Assumption 4.11 holds. Then the following hold:

(i) There exists $x^* \in \mathbb{R}^n$ such that $x_k \to x^*$ and $0 \in \partial F(x^*)$.

(ii) Suppose additionally that

$$N_C(\operatorname{Proj}_C(x^*)) \cap \big(-N_D(x^*)\big) = \{0\}. \tag{35}$$

Then $x^* \in C \cap D$. Moreover, there exist $Q_1 \in (0, 1)$ and $c_1 > 0$ such that

$$\|x_k - x^*\| \leq c_1 Q_1^k,$$

for $k$ sufficiently large.

### 4.7.2 Application: finding sparse solution of linear system

In the remainder, let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $r \in \mathbb{N} \backslash \{0\}$, and $R > 0$. Define

$$C = \{x \in \mathbb{R}^n : Ax = b\} \text{ and } D = \{x \in \mathbb{R}^n : \|x\|_0 \leq r, \|x\| \leq R\}.$$

Clearly $C$ is semialgebraic. The set $D$ can be represented as

$$D = \{x \in \mathbb{R}^n : \|x\|_0 \leq r\} \cap \{x \in \mathbb{R}^n : \|x\|^2 \leq R^2\},$$

which is an intersection of semialgebraic sets; see, e.g., [6, Formula 27(d)], thus semialgebraic. In turn, Proposition 4.30 ensures that BiFRB and its Euclidean variant iFRB converge to a stationary point of (34) with $C$ and $D$ specified as above.

We shall benchmark BiFRB, iFRB and FRB against the Douglas-Rachford (DR) [22] and inertial Tseng's (iTseng) [12] methods with $R = 1, 1000$ and $r = \lceil m/5 \rceil$. These splitting methods are known to converge globally to a stationary point of (12) under the same assumptions on $F$; see [22, Theorems 1–2, Remark 4, Corollary 1], [37, Theorem 3.9] and [12, Theorem 3.1], respectively.

BiFRB, iFRB and FRB are implemented with stepsize rules given by Proposition 4.19, Corollary 4.22 and [37, Theorem 3.9] respectively, and terminated when

$$\frac{\max\left\{ \|x_{k+1} - x_k\|, \|x_k - x_{k-1}\| \right\}}{\max\left\{1, \|x_k\|, \|x_{k-1}\| \right\}} < 10^{-10}. \tag{36}$$

Moreover, BiFRB uses the kernel $h(x) = 0.1 \cdot \sqrt{1 + \|x\|^2} + 2.51 \cdot \|x\|^2 / 2$. We apply DR with stepsize and termination condition given by [22, Section 5] with the same tolerance $10^{-10}$; while iTseng employs a stepsize given by [12, Lemma 3.3] and termination condition (36). As for inertial parameter, we use $\alpha_1 = 0.9$ for BiFRB and $\alpha_2 = 0.49$ for both of iFRB and iTseng. Finally, each algorithm is initialized at the origin and is equipped with the stepsize heuristic described in [22, Remark 4].

Simulation results are presented in Table 2. Our problem data is generated through creating random matrices $A \in \mathbb{R}^{m\times n}$ with entries following the standard Gaussian distribution. For each problem of the size $(m, n)$, we randomly generate 50 instances and report ceilings of the average number of iterations (iter) and the minimal objective value at termination ($\text{fval}_{\min}$). As problem (34) has optimal value zero, $\text{fval}_{\min}$ indicates whether an algorithm actually hits a global minimizer rather than simply converging to a stationary point. We observed the following:

- BiFRB is the most advantageous method on "bad" problems (small $m$), but its performance is not robust.
- DR tends to have the smallest function value at termination, however, it can converge very slowly.
- In most cases, iFRB has the smallest number of iterations with fair function value at termination.

Table 2: Comparing BiFRB, iFRB, FRB to DR and iTseng.

| Size | $(R = 1)$ | BiFRB | | iFRB | | FRB | | DR | | iTseng | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| m | n | iter | $\text{fval}_{\min}$ | iter | $\text{fval}_{\min}$ | iter | $\text{fval}_{\min}$ | iter | $\text{fval}_{\min}$ | iter | $\text{fval}_{\min}$ |
| 100 | 4000 | 50 | 0.03251 | 93 | 0.03251 | 631 | 0.03929 | 860 | 0.02819 | 1367 | 0.03149 |
| 100 | 5000 | 52 | 0.02413 | 115 | 0.02413 | 733 | 0.02875 | 684 | 0.02192 | 1632 | 0.02268 |
| 100 | 6000 | 57 | 0.01369 | 145 | 0.01369 | 911 | 0.01794 | 683 | 0.01253 | 1970 | 0.0133 |
| 200 | 4000 | 870 | 0.2923 | 39 | 0.2745 | 190 | 0.2769 | 5466 | 0.2711 | 448 | 0.2746 |
| 200 | 5000 | 1198 | 0.2367 | 41 | 0.2095 | 231 | 0.2157 | 5864 | 0.2073 | 546 | 0.2097 |
| 200 | 6000 | 81 | 0.2306 | 43 | 0.2259 | 257 | 0.2292 | 4199 | 0.2236 | 620 | 0.2264 |
| 300 | 4000 | 885 | 1.051 | 32 | 1.036 | 105 | 1.038 | 5497 | 1.036 | 253 | 1.048 |
| 300 | 5000 | 925 | 0.7481 | 34 | 0.7044 | 124 | 0.7071 | 5673 | 0.7032 | 302 | 0.7051 |
| 300 | 6000 | 1407 | 0.583 | 36 | 0.5546 | 144 | 0.555 | 6173 | 0.5541 | 352 | 0.5566 |
| Size | $(R = 1000)$ | BiFRB | | iFRB | | FRB | | DR | | iTseng | |
| m | n | iter | $\text{fval}_{\min}$ | iter | $\text{fval}_{\min}$ | iter | $\text{fval}_{\min}$ | iter | $\text{fval}_{\min}$ | iter | $\text{fval}_{\min}$ |
| 100 | 4000 | 1873 | 0.006609 | 1210 | 0.00365 | 7376 | 0.00816 | 2194 | 4e-21 | 10001 | 0.01671 |
| 100 | 5000 | 574 | 0.00278 | 1427 | 0.002244 | 8675 | 1.098e-05 | 2919 | 1.626e-20 | 10001 | 0.0158 |
| 100 | 6000 | 790 | 0.003827 | 1696 | 0.001261 | 9394 | 0.0002952 | 2344 | 3.438e-20 | 10001 | 0.01066 |
| 200 | 4000 | 5842 | 2.992 | 750 | 2.442e-18 | 8223 | 0.007688 | 1038 | 1.106e-20 | 9797 | 0.1094 |
| 200 | 5000 | 5938 | 0.6547 | 859 | 2.876e-05 | 7257 | 0.02581 | 1242 | 2.082e-20 | 9997 | 0.07125 |
| 200 | 6000 | 5666 | 0.917 | 1086 | 0.0001211 | 7312 | 0.006726 | 1487 | 9.129e-21 | 10001 | 0.07524 |
| 300 | 4000 | 7164 | 3.322 | 619 | 2.196e-18 | 4620 | 1.462e-05 | 753 | 9.795e-21 | 7722 | 0.2417 |
| 300 | 5000 | 5267 | 2.299 | 693 | 2.175e-18 | 6352 | 0.02936 | 880 | 5.446e-20 | 9194 | 0.1702 |
| 300 | 6000 | 5724 | 5.06 | 750 | 4.71e-18 | 8644 | 0.1109 | 1027 | 2.175e-20 | 9823 | 0.1423 |

# 5 A sum rule of the generalized concave KL property

Results in this section are from preprint [35], which is partially completed in the master thesis [38]. Here we only present new results completed during my PhD.

We say a generalized concave KL function has KL exponent $\theta \in [0,1)$, if an associated desingularizing function $\varphi(t) = c \cdot t^{1-\theta}$ for some $c > 0$.

**Theorem 5.1.** Let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper continuous function for each $i \in \{1,\dots,m\}$ with $\cap_{i=1}^m \operatorname{int}\operatorname{dom}\partial f_i \neq \varnothing$, and let $f = \sum_{i=1}^m f_i$. Suppose that at most one of $f_i$ is not locally Lipschitz. Pick $\bar{x} \in \cap_{i=1}^m \operatorname{int}\operatorname{dom}\partial f_i$. Assume that for each $i$, $f_i$ has the generalized concave KL property at $\bar{x}$ with respect to $U_i = \mathbb{B}(\bar{x};\varepsilon_i)$ for some $\varepsilon_i > 0$, $\eta_i \in (0,\infty]$ and $\varphi_i \in \Phi_{\eta_i}$. Suppose that there exist $\varepsilon_0 > 0$ and $\alpha > 0$ such that for every $x_i \in \mathbb{B}(\bar{x};\varepsilon_0)$

$$\left\|\sum_{i=1}^m u_i\right\| \geq \alpha \sum_{i=1}^m \|u_i\|,\ \forall u_i \in \partial f_i(x_i). \tag{37}$$

Set $\varepsilon = \min_{0\leq i\leq m} \varepsilon_i$ and $\eta = \min_{1\leq i\leq m} \eta_i$. Define $\varphi : [0,\eta) \to \mathbb{R}$ by

$$\varphi(t) = \frac{1}{\alpha}\int_0^t \max_{1\leq i\leq m} (\varphi_i)'_-\left(\frac{s}{m}\right) ds, \forall t \in (0,\eta),$$

and $\varphi(0) = 0$. If each $\varphi_i$ is strictly concave, then the sum $f$ has the generalized concave KL property at $\bar{x}$ with respect to $U = \mathbb{B}(\bar{x};\varepsilon)$, $\eta$ and $\varphi$.

**Remark 5.2.** (i) Condition (37) is weaker than the Li-Pong disjoint condition [21, Condition (12)] for a sum rule of two functions. Considering $f = f_1 + f_2$ with associated desingularizing functions having the form $\varphi_i(t) = c_i \cdot t^{1-\theta_i}$ for some $c_i > 0$ and $\theta_i \in [0,1)$, Li and Pong provided a sum rule [21, Theorem 3.4] assuming the following:

$$W_{f_1}(\bar{x}) \cap (-W_{f_2}(\bar{x})) = \varnothing, \tag{38}$$

where

$$W_{f_i}(\bar{x}) = \left\{\lim_{k\to\infty} w_k : w_k = \frac{u_k}{\|u_k\|}, u_k \in \partial f_i(x_k)\backslash\{0\}, x_k \to \bar{x}\right\},$$

which implies (37) with $m = 2$ by the first line of [21, Inequality (16)]. The converse, however, fails even on the real line. For instance, consider $f_1(x) = x$ and $f_2(x) = -\frac{1}{2}x$. On one hand, we have $|f_1'(x) + f_2'(y)| = \frac{1}{3}(|f_1'(x)| + |f_2'(y)|)$ for every $x, y \in \mathbb{R}$. On the other hand, it is easy to see that $W_{f_1}(0) = -W_{f_2}(0) = \{1\}$, which means that (38) fails. Nevertheless, these two conditions are equivalent when $f_i = \delta_{C_i}$, where $C_i$ are nonempty closed sets, see, e.g., [17, Proposition 2.4].

(ii) For $i = \{1, \ldots, m\}$, if $\eta_i < m$ and $\varphi_i(t) = t^{1-\theta_i}/(1-\theta_i)$ for some $\theta_i \in [0, 1)$, then the desingularizing function given by Theorem 5.1 reduces to

$$\varphi(t) = \frac{1}{\alpha}\int_0^t \max_{1\leq i\leq m}\left(\frac{s}{m}\right)^{-\theta_i} ds = \frac{1}{\alpha}\int_0^t \left(\frac{s}{m}\right)^{-\theta} ds = \frac{m^\theta}{(1-\theta)\alpha} t^{1-\theta}, \forall t \in (0, \eta) \subseteq (0, m),$$

where $\theta = \max_{1\leq i\leq m} \theta_i$ and the second equality holds because $s/m < 1$.

# 6 Future work

In this section, we discuss topics for future research.

## 6.1 An open question of Malitsky and Tam

Consider the inclusion problem of maximally monotone operators:

$$\text{find } x \in \mathbb{R}^n \text{ such that } 0 \in (A + B)(x), \tag{39}$$

where $A : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is maximally monotone and $B = \sum_{i=1}^n B_i$ with $(\forall i \in \{1, \ldots, n\})$ $B_i$ being monotone and Lipschitz continuous. In [26, Section 6], Malitsky and Tam proved almost sure convergence of the following stochastic algorithm:

$$\begin{aligned}&\text{choose } i_k \text{ uniformly at random from } \{1,\ldots,\text{n}\},\\ &x_{k+1} = J_{\lambda A}\big(x_k - \lambda B(x_k) - \lambda(B_{i_k}(x_k) - B_{i_k}(x_{k-1}))\big),\end{aligned}$$

and posted *an open question regarding a block coordinate extension of the above algorithm with full stochastic approximation of operator B; see [26, Section 7].*

Thus, with the generalized concave KL property, *it is tempting to resolve the above open question and push it to nonconvex setting*. To be specific, we are interested in an stochastic extension of Algorithm 4.2 for

$$\min_{x\in\mathbb{R}^N} f(x) + \sum_{i=1}^n g_i(x_i),$$

where $(\forall i \in \{1, \ldots, n\})$ $x_i \in \mathbb{R}^{n_i}, n_i \in \mathbb{N}$, $N = \sum_i n_i$, $f : \mathbb{R}^N \to \overline{\mathbb{R}}$ is proper lsc and $g_i : \mathbb{R}^{n_i} \to \mathbb{R}$ has Lipschitz gradient.

## 6.2 An open question of Yu, Li, and Pong

In [40, Remark 3.1], Yu, Li, and Pong posted an open question regarding the behavior of KL exponents under supremum operation: *"It would be of interest to see, under what additional conditions, the supremum operation can preserve the KL exponents. "*. The KL exponent corresponds to a special case of the generalized concave KL property. Therefore, one direction for our future research is *to study a calculus rule of the generalized concave KL property under supremum operation*, which could shed lights on the Yu-Li-Pong open question.

# References


[1] H. ATTOUCH AND J. BOLTE, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Mathematical Programming, 116 (2009), pp. 5–16.

[2] H. ATTOUCH, J. BOLTE, P. REDONT, AND A. SOUBEYRAN, *Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality*, Mathematics of Operations Research, 35 (2010), pp. 438–457.

[3] H. ATTOUCH, J. BOLTE, AND B. F. SVAITER, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods*, Mathematical Programming, 137 (2013), pp. 91–129.

[4] S. BANERT AND R. I. BOŢ, *A general double-proximal gradient algorithm for d.c. programming*, Mathematical Programming, 178 (2019), pp. 301–326.

[5] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, Cham, 2017.

[6] H. H. BAUSCHKE, D. R. LUKE, H. M. PHAN, AND X. WANG, *Restricted normal cones and sparsity optimization with affine constraints*, Foundations of Computational Mathematics, 14 (2014), pp. 63–83.

[7] A. BECK, *First-order Methods in Optimization*, SIAM, 2017.

[8] J. BOLTE, A. DANIILIDIS, A. LEWIS, AND M. SHIOTA, *Clarke subgradients of stratifiable functions*, SIAM Journal on Optimization, 18 (2007), pp. 556–572.

[9] J. BOLTE, A. DANIILIDIS, O. LEY, AND L. MAZET, *Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity*, Transactions of the American Mathematical Society, 362 (2010), pp. 3319–3363.

[10] J. BOLTE, S. SABACH, AND M. TEBOULLE, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming, 146 (2014), pp. 459–494.

[11] J. BOLTE, S. SABACH, M. TEBOULLE, AND Y. VAISBOURD, *First order methods beyond convexity and Lipschitz gradient continuity with applications to quadratic inverse problems*, SIAM Journal on Optimization, 28 (2018), pp. 2131–2151.

[12] R. I. BOŢ AND E. R. CSETNEK, *An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems*, Journal of Optimization Theory and Applications, 171 (2016), pp. 600–616.

[13] R. I. BOŢ, E. R. CSETNEK, AND S. C. LÁSZLÓ, *An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions*, EURO Journal on Computational Optimization, 4 (2016), pp. 3–25.

[14] R. I. BOŢ AND D.-K. NGUYEN, *The proximal alternating direction method of multipliers in the nonconvex setting: convergence analysis and rates*, Mathematics of Operations Research, 45 (2020), pp. 682–712.

[15] C. CHEN, T. K. PONG, L. TAN, AND L. ZENG, *A difference-of-convex approach for split feasibility with applications to matrix factorizations and outlier detection*, Journal of Global Optimization, 78 (2020), pp. 107–136.

[16] P. L. COMBETTES AND V. R. WAJS, *Signal recovery by proximal forward-backward splitting*, Multiscale Modeling & Simulation, 4 (2005), pp. 1168–1200.

[17] M. N. DAO AND H. M. PHAN, *Linear convergence of projection algorithms*, Mathematics of Operations Research, 44 (2019), pp. 715–738.

[18] J. DOUGLAS AND H. H. RACHFORD, *On the numerical solution of heat conduction problems in two and three space variables*, Transactions of the American Mathematical Society, 82 (1956), pp. 421–439.

[19] K. KURDYKA, *On gradients of functions definable in o-minimal structures*, Annales de l'institut Fourier, 48 (1998), pp. 769–783.

[20] G. LI, T. LIU, AND T. K. PONG, *Peaceman–Rachford splitting for a class of nonconvex optimization problems*, Computational Optimization and Applications, 68 (2017), pp. 407–436.

[21] G. LI AND T. K. PONG, *Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods*, arXiv:1602.02915v1, (2016).

[22] G. LI AND T. K. PONG, *Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems*, Mathematical Programming, 159 (2016), pp. 371–401.

[23] P.-L. LIONS AND B. MERCIER, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis, 16 (1979), pp. 964–979.

[24] S. ŁOJASIEWICZ, *Une propriété topologique des sous-ensembles analytiques réels*, Les équations aux dérivées partielles, 117 (1963), pp. 87–89.

[25] R. LUSS AND M. TEBOULLE, *Conditional gradient algorithms for rank-one matrix approximations with a sparsity constraint*, SIAM Review, 55 (2013), pp. 65–98.

[26] Y. MALITSKY AND M. K. TAM, *A forward-backward splitting method for monotone inclusions without cocoercivity*, SIAM Journal on Optimization, 30 (2020), pp. 1451–1472.

[27] B. S. MORDUKHOVICH, *Variational Analysis and Generalized Differentiation I: Basic Theory*, Springer-Verlag, Berlin, 2006.

[28] Y. NESTEROV, *A method for solving the convex programming problem with convergence rate $O(1/k^2)$*, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543–547.

[29] P. OCHS, *Local convergence of the heavy-ball method and iPiano for non-convex optimization*, Journal of Optimization Theory and Applications, 177 (2018), pp. 153–180.

[30] P. OCHS, *Unifying abstract inexact convergence theorems and block coordinate variable metric iPiano*, SIAM Journal on Optimization, 29 (2019), pp. 541–570.

[31] P. OCHS, Y. CHEN, T. BROX, AND T. POCK, *iPiano: Inertial proximal algorithm for nonconvex optimization*, SIAM Journal on Imaging Sciences, 7 (2014), pp. 1388–1419.

[32] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

[33] P. TSENG, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM Journal on Control and Optimization, 38 (2000), pp. 431–446.

[34] X. WANG AND Z. WANG, *The exact modulus of the generalized Kurdyka-Łojasiewicz property*, Mathematics of Operations Research (2022), https://doi.org/10.1287/moor.2021.1227.

[35] ——, *Calculus rules of the generalized concave Kurdyka-Łojasiewicz property*, preprint, arXiv:2110.03795, (2021).

[36] ——, *A Bregman inertial forward-reflected-backward method for nonconvex minimization*, preprint, (2022).

[37] ——, *Malitsky-Tam forward-reflected-backward splitting method for nonconvex minimization problems*, Computational Optimization and Applications, 82 (2022), pp. 441–463.

[38] Z. WANG, *The exact modulus of the generalized Kurdyka-Łojasiewicz property*, Master's thesis, University of British Columbia, 2020.

[39] J.-H. WON, J. XU, AND K. LANGE, *Projection onto Minkowski sums with application to constrained learning*, in Proceedings of the 36th International Conference on Machine Learning, vol. 97 of Proceedings of Machine Learning Research, PMLR, 2019, pp. 3642–3651.

[40] P. YU, G. LI, AND T. K. PONG, *Kurdyka–Łojasiewicz exponent via inf-projection*, Foundations of Computational Mathematics, (2021), pp. 1–47.