



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

Xianfu Wang<sup>1</sup> · Ziyuan Wang<sup>1</sup>

Received: 16 November 2021 / Accepted: 3 March 2022 / Published online: 4 April 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

We extend the Malitsky-Tam forward-reflected-backward (FRB) splitting method for inclusion problems of monotone operators to nonconvex minimization problems. By assuming the generalized concave Kurdyka-Łojasiewicz (KL) property of a quadratic regularization of the objective, we show that the FRB method converges globally to a stationary point of the objective and enjoys the finite length property. Convergence rates are also given. The sharpness of our approach is guaranteed by virtue of the exact modulus associated with the generalized concave KL property. Numerical experiments suggest that FRB is competitive compared to the Douglas-Rachford method and the Boţ-Csetnek inertial Tseng’s method.

**Keywords** Generalized concave Kurdyka-Łojasiewicz property · Proximal mapping · Malitsky-Tam forward-reflected-backward splitting method · Merit function · Global convergence · Nonconvex optimization

**Mathematics Subject Classification** Primary 49J52 · 90C26 · Secondary 26D10

## 1 Introduction

Consider the general inclusion problem of maximally monotone operators:

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

where  $A : \mathbb{R}^n \rightarrow 2^{\mathbb{R}^n}$  and  $B : \mathbb{R}^n \rightarrow \mathbb{R}^n$  are (maximally) monotone operators with  $B$  being Lipschitz continuous. Successful splitting methods for the inclusion problem (1) are numerous; see, e.g., [1, Chapters 26, 28], including the forward-backward

---

✉ Xianfu Wang  
shawn.wang@ubc.ca

Ziyuan Wang  
ziyuan.wang@alumni.ubc.ca

<sup>1</sup> Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada

method with  $B$  being cocoercive [2] and Tseng's method [3]. Recently, Malitsky and Tam [4] proposed a *forward-reflected-backward* (FRB) splitting method for solving (1). Let  $\lambda > 0$  and  $J_{\lambda A} = (\text{Id} + \lambda A)^{-1}$  be the resolvent of operator  $\lambda A$ ; see, e.g., [1]. Given initial points  $x_0, x_{-1} \in \mathbb{R}^n$ , the Malitsky-Tam FRB method with a fixed step-size (see [4, Remark 2.1]) iterates

$$x_{k+1} = J_{\lambda A}(x_k - 2\lambda B(x_k) + \lambda B(x_{k-1})), \quad (2)$$

with only one forward step computation required per iteration compared to Tseng's method. In [4] Malitsky and Tam showed that their FRB given by (2) converges to a solution of (1).

*In this paper, we extend the Malitsky-Tam FRB splitting method to solve the following structured possibly nonconvex minimization problem:*

$$\min_{x \in \mathbb{R}^n} F(x) = f(x) + g(x), \quad (3)$$

where  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}} = (-\infty, \infty]$  is proper lower semicontinuous (lsc), and  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  has a Lipschitz continuous gradient with constant  $L > 0$ . This class of optimization problems is of significantly broad interest in both theory and practical applications; see, e.g., [5, 6].

Because both  $\partial f, \partial g$  are not necessarily monotone, the techniques developed by Malitsky and Tam do not apply. To accomplish our goal, we employ the generalized concave Kurdyka-Łojasiewicz property (see Definition 2.2) recently proposed in [7] as a key assumption in our analysis, under which the FRB method demonstrates pleasant behavior. We show that the sequence produced by FRB converges globally to a stationary point of (3) and has finite length property. The sharpest upper bound on the total length of iterates is described by using the exact modulus associated with the generalized concave KL property. Both sequential and function value convergence rate analysis are carried out. Our approach deviates from the convergence mechanism developed in the original FRB paper [4], but follows a similar pattern as in many work that devotes to obtaining nonconvex extensions of splitting algorithms; see, e.g., [6, 8–13]. Numerical simulations indicate that FRB is competitive compared to the Douglas-Rachford method with a fixed step-size [6] and the Bot-Csetnek inertial Tseng's method [12].

The paper is structured as follows. We introduce notation and basic concepts in Sect. 2. The convergence analysis of FRB is presented in Sect. 3. When the merit function of FRB has KL exponent  $\theta \in [0, 1)$ , we obtain convergence rates for both the iterates and function values. Numerical experiments are implemented in Sect. 4. We end this paper by providing some directions for future research in Sect. 5.

## 2 Notation and preliminaries

Throughout this paper,  $\mathbb{R}^n$  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  $\mathbb{N} = \{0, 1, 2, 3, \dots\}$  and  $\mathbb{N}^* = \{-1\} \cup \mathbb{N}$ . The open ball centered at  $\bar{x}$  with radius

$r$  is denoted by  $\mathbb{B}(\bar{x};r)$ . Associated with a subset  $K \subseteq \mathbb{R}^n$ , its distance function is  $\text{dist}(\cdot, K) : \mathbb{R}^n \rightarrow \overline{\mathbb{R}} = (-\infty, \infty]$ ,

$$x \mapsto \text{dist}(x, K) = \inf\{\|x - y\| : y \in K\},$$

where  $\text{dist}(x, K) \equiv \infty$  if  $K = \emptyset$ . The indicator function of  $K$  is defined as  $\delta_K(x) = 0$  if  $x \in K$  and  $\delta_K(x) = \infty$  otherwise. For a function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  and  $r_1, r_2 \in [-\infty, \infty]$ , we set  $[r_1 < f < r_2] = \{x \in \mathbb{R}^n : r_1 < f(x) < r_2\}$ . We say that  $f$  is coercive if  $\lim_{\|x\| \rightarrow \infty} f(x) = \infty$ .

We will use frequently the following subgradients in the nonconvex setting; see, e.g., [14, 15].

**Definition 2.1** Let  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be a proper function. We say that

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

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(\|x - \bar{x}\|). \tag{4}$$

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

$$v \in \{v \in \mathbb{R}^n : \exists x_k \xrightarrow{f} \bar{x}, \exists v_k \in \hat{\partial}f(x_k), v_k \rightarrow v\}, \tag{5}$$

where  $x_k \xrightarrow{f} \bar{x} \Leftrightarrow x_k \rightarrow \bar{x}$  and  $f(x_k) \rightarrow f(\bar{x})$ . Moreover, we set  $\text{dom}\partial f = \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$ .

(iii) We say that  $\bar{x} \in \text{dom}\partial f$  is a *stationary point*, if  $0 \in \partial f(\bar{x})$ .

Recall that the proximal mapping of a proper and lsc  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  with parameter  $\lambda > 0$  is

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

which is the resolvent  $J_{\lambda A}$  with  $A = \partial f$  when  $f$  is convex; see, e.g., [1].

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

**Definition 2.2** [7] Let  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be proper and lsc. Let  $\bar{x} \in \text{dom}\partial f$  and  $\mu \in \mathbb{R}$ , and let  $V \subseteq \text{dom}\partial f$  be a nonempty subset.

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

$$\varphi'_-(f(x) - f(\bar{x})) \cdot \text{dist}(0, \partial f(x)) \geq 1, \tag{6}$$

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

(ii) Suppose that  $f(x) = \mu$  on  $V$ . We say  $f$  has the setwise<sup>1</sup> 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  $x \in U \cap [0 < f - \mu < \eta]$ ,

$$\varphi'_-(f(x) - \mu) \cdot \text{dist}(0, \partial f(x)) \geq 1. \tag{7}$$

**Remark 2.3** The generalized concave KL property reduces to the celebrated concave KL property when concave desingularizing functions  $\varphi \in \Phi_\eta$  are continuously differentiable, which has been employed by a series of papers to establish convergence of proximal-type algorithms in the nonconvex setting; see, e.g., [6, 8–13 16] and the references therein. The generalized concave KL property allows us to describe the optimal (minimal) concave desingularizing function and obtain sharp convergence results; see [7].

Many classes of functions satisfy the generalized concave KL property, for instance semialgebraic functions; see, e.g., [17, Corollary 18] and [8, Section 4.3]. For more characterizations of the concave KL property, we refer readers to [19]; see also the fundamental work of Łojasiewicz [20] and Kurdyka [21].

**Definition 2.4** (i) A set  $E \subseteq \mathbb{R}^n$  is called semialgebraic if there exist finitely many polynomials  $g_{ij}, h_{ij} : \mathbb{R}^n \rightarrow \mathbb{R}$  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  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is called semialgebraic if its graph  $\text{gph}f = \{(x, y) \in \mathbb{R}^{n+1} : f(x) = y\}$  is semialgebraic.

**Fact 2.5** [17, Corollary 18] *Let  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be a proper and lsc function and let  $\bar{x} \in \text{dom}\partial f$ . If  $f$  is semialgebraic, then it has the concave KL property at  $\bar{x}$  with  $\varphi(t) = c \cdot t^{1-\theta}$  for some  $c > 0$  and  $\theta \in (0, 1)$ .*

### 3 Forward-reflected-backward splitting method for nonconvex problems

In this section, we prove convergence results of the FRB splitting method in the nonconvex setting. For convex minimization problems, the iteration scheme (2) specifies to

$$x_{k+1} = \text{Prox}_{\lambda f}(y_k - \lambda \nabla g(x_k)),$$

where  $y_k = x_k + \lambda(\nabla g(x_{k-1}) - \nabla g(x_k))$  is an intermediate variable introduced for simplicity. Since proximal mapping  $\text{Prox}_{\lambda f}$  might not be single-valued without convexity of  $f$ , we formulate the FRB scheme for solving (3) as follows:

<sup>1</sup> In the remainder, we shall omit adjectives “pointwise” and “setwise” whenever there is no ambiguity.

### Forward-reflected-backward (FRB) splitting method

1. Initialization: Pick  $x_{-1}, x_0 \in \mathbb{R}^n$  and real number  $\lambda > 0$ .
2. For  $k \in \mathbb{N}$ , compute

$$y_k = x_k + \lambda(\nabla g(x_{k-1}) - \nabla g(x_k)), \tag{8}$$

$$x_{k+1} \in \text{Prox}_{\lambda f}(y_k - \lambda \nabla g(x_k)). \tag{9}$$

Recall that a proper, lsc function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is prox-bounded if there exists  $\lambda > 0$  such that  $f + \frac{1}{2\lambda} \|\cdot\|^2$  is bounded below; see, e.g., [15, Exercise 1.24]. The supremum of the set of all such  $\lambda$  is the threshold  $\lambda_f$  of prox-boundedness for  $f$ . In the remainder of this paper, we shall assume that

- $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is proper, lsc, and prox-bounded with threshold  $\lambda_f > 0$ .
- $g : \mathbb{R}^n \rightarrow \mathbb{R}$  has a Lipschitz continuous gradient with constant  $L > 0$ .

For  $0 < \lambda < \lambda_f$ , [15, Theorem 1.25] shows that  $(\forall x \in \mathbb{R}^n) \text{Prox}_{\lambda f}(x)$  is nonempty compact, so the above scheme is well-defined. Moreover, it is easy to see that

$$\text{Prox}_{\lambda f}(y_k - \lambda \nabla g(x_k)) = \underset{x \in \mathbb{R}^n}{\text{argmin}} \left\{ f(x) + \langle x - y_k, \nabla g(x_k) \rangle + \frac{1}{2\lambda} \|x - y_k\|^2 \right\}. \tag{10}$$

### 3.1 Basic properties

After formulating the iterative scheme, we now investigate its basic properties. We begin 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., [6, 12, 13, 22].

**Definition 3.1** (*FRB merit function*) Let  $\lambda > 0$ . Define the FRB merit function  $H : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  by

$$H(x, y) = f(x) + g(x) + \left( \frac{1}{4\lambda} - \frac{L}{4} \right) \|x - y\|^2. \tag{11}$$

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 3.2** 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

$$M_1 \|z_k - z_{k-1}\|^2 \leq H(z_{k-1}) - H(z_k), \tag{12}$$

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

$$(ii) \sum_{k=0}^{\infty} \|z_k - z_{k-1}\|^2 < \infty, \text{ consequently } \lim_{k \rightarrow \infty} \|z_k - z_{k-1}\| = 0.$$

**Proof** (i) The sequence  $(x_k)_{k \in \mathbb{N}^*}$  is well-defined because  $\lambda < \lambda_f$ . The FRB iterative scheme and (10) imply that

$$f(x_{k+1}) \leq f(x_k) + \langle \nabla g(x_k), x_k - x_{k+1} \rangle + \frac{1}{2\lambda} \|x_k - y_k\|^2 - \frac{1}{2\lambda} \|x_{k+1} - y_k\|^2. \tag{13}$$

Invoking the descent lemma [1, Theorem 18.15] to  $g$ , we have

$$g(x_{k+1}) \leq g(x_k) + \langle \nabla g(x_k), x_{k+1} - x_k \rangle + \frac{L}{2} \|x_{k+1} - x_k\|^2. \tag{14}$$

Summing (13) and (14) yields

$$\begin{aligned} f(x_{k+1}) + g(x_{k+1}) &\leq f(x_k) + g(x_k) + \frac{L}{2} \|x_{k+1} - x_k\|^2 + \frac{1}{2\lambda} \|x_k - y_k\|^2 - \frac{1}{2\lambda} \|x_{k+1} - y_k\|^2 \\ &= f(x_k) + g(x_k) + \left(\frac{L}{2} - \frac{1}{2\lambda}\right) \|x_{k+1} - x_k\|^2 + \frac{1}{\lambda} \langle x_k - x_{k+1}, x_k - y_k \rangle \\ &\leq f(x_k) + g(x_k) + \left(\frac{L}{2} - \frac{1}{2\lambda}\right) \|x_{k+1} - x_k\|^2 + \|x_k - x_{k+1}\| \|\nabla g(x_{k-1}) - \nabla g(x_k)\| \\ &\leq f(x_k) + g(x_k) + \left(L - \frac{1}{2\lambda}\right) \|x_{k+1} - x_k\|^2 + \frac{L}{2} \|x_k - x_{k-1}\|^2, \end{aligned}$$

where the first equality follows from the identity  $\|x_k - y_k\|^2 = \|x_{k+1} - y_k\|^2 - \|x_{k+1} - x_k\|^2 - 2\langle x_{k+1} - x_k, x_k - y_k \rangle$ , and the second inequality is implied by (8) and the Cauchy-Schwarz inequality. Moreover, the third inequality holds by using the inequality  $a^2 + b^2 \geq 2ab$  and the Lipschitz continuity of  $\nabla g$ . Recall the product norm identity  $\|z_k - z_{k-1}\|^2 = \|x_{k+1} - x_k\|^2 + \|x_k - x_{k-1}\|^2$ . It then follows from the inequality above that

$$\begin{aligned} &f(x_{k+1}) + g(x_{k+1}) \\ &\leq f(x_k) + g(x_k) + \left(L - \frac{1}{2\lambda}\right) (\|z_k - z_{k-1}\|^2 - \|x_k - x_{k-1}\|^2) + \frac{L}{2} \|x_k - x_{k-1}\|^2 \\ &= f(x_k) + g(x_k) + \left(L - \frac{1}{2\lambda}\right) \|z_k - z_{k-1}\|^2 - \left(\frac{L}{2} - \frac{1}{2\lambda}\right) \|x_k - x_{k-1}\|^2. \end{aligned}$$

Adding  $\left(\frac{1}{4\lambda} - \frac{L}{4}\right) \|x_{k+1} - x_k\|^2$  to both sides of the above inequality and applying again the product norm identity, one gets

$$\begin{aligned}
 H(z_k) &= f(x_{k+1}) + g(x_{k+1}) + \left(\frac{1}{4\lambda} - \frac{L}{4}\right) \|x_{k+1} - x_k\|^2 \\
 &\leq f(x_k) + g(x_k) + \left(L - \frac{1}{2\lambda}\right) \|z_k - z_{k-1}\|^2 - \left(\frac{L}{2} - \frac{1}{2\lambda}\right) \|x_k - x_{k-1}\|^2 \\
 &\quad + \left(\frac{1}{4\lambda} - \frac{L}{4}\right) \|z_k - z_{k-1}\|^2 - \left(\frac{1}{4\lambda} - \frac{L}{4}\right) \|x_k - x_{k-1}\|^2 \\
 &= f(x_k) + g(x_k) + \left(\frac{1}{4\lambda} - \frac{L}{4}\right) \|x_k - x_{k-1}\|^2 + \left(\frac{3L}{4} - \frac{1}{4\lambda}\right) \|z_k - z_{k-1}\|^2 \\
 &= H(z_{k-1}) - M_1 \|z_k - z_{k-1}\|^2,
 \end{aligned}$$

which proves (12). Because  $M_1 > 0$ , we have  $H(z_k) \leq H(z_{k-1})$ . The convergence result then follows from the assumption that  $\inf(f + g) > -\infty$ .

(ii) For integer  $l > 0$ , summing (12) from  $k = 0$  to  $k = l - 1$  yields

$$\sum_{k=0}^{l-1} \|z_k - z_{k-1}\|^2 \leq M_1^{-1} (H(z_{-1}) - H(z_l)) \leq M_1^{-1} (H(z_{-1}) - \inf(f + g)).$$

Passing to the limit, one concludes that  $\sum_{k=0}^{\infty} \|z_k - z_{k-1}\|^2 < \infty$ . □

**Remark 3.3** (*Comparison to a known result*) Boğ and Csetnek also utilized a merit function with similar form for the decreasing property of Tseng’s method [12, Lemma 3.2]. The similarity between their approach and ours is not unexpected. Indeed, it is pointed out in [23, Section 2.3] that FRB is a modified Tseng’s method. The major difference between [12, Lemma 3.2] and Lemma 3.2 is that inequality (12) asserts that  $H$  decreases with respect to a single sequence  $(x_k)$  generated by FRB, while [12, Lemma 3.2] concerns the decreasing property of  $H$  with respect to two different sequences produced by Tseng’s method.

Next we estimate the lower bound on the gap between iterates. To this end, a lemma helps.

**Lemma 3.4** *For every  $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ , we have*

$$\partial H(x, y) = \left\{ \partial f(x) + \nabla g(x) + \left(\frac{1}{2\lambda} - \frac{L}{2}\right)(x - y) \right\} \times \left\{ \left(\frac{1}{2\lambda} - \frac{L}{2}\right)(y - x) \right\}.$$

**Proof** Apply the subdifferential sum and separable sum rules [15, Exercise 8.8, Proposition 10.5]. □

**Lemma 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  $k \in \mathbb{N}^*$ . Let  $M_2 = \sqrt{2} \left( \frac{1}{\lambda} + L + \left| \frac{1}{\lambda} - L \right| \right)$  and define for  $k \in \mathbb{N}$*

$$\begin{aligned}
 p_{k+1} &= \frac{1}{\lambda}(x_k - x_{k+1}) + \nabla g(x_{k+1}) - 2\nabla g(x_k) + \nabla g(x_{k-1}), \\
 A_{k+1} &= p_{k+1} + \left(\frac{1}{2\lambda} - \frac{L}{2}\right)(x_{k+1} - x_k), \quad B_{k+1} = \left(\frac{1}{2\lambda} - \frac{L}{2}\right)(x_k - x_{k+1}).
 \end{aligned}$$

Then  $(A_{k+1}, B_{k+1}) \in \partial H(z_k)$  and  $\|(A_{k+1}, B_{k+1})\| \leq M_2 \|z_k - z_{k-1}\|$ .

**Proof** The FRB scheme and (10) imply that  $0 \in \partial f(x_{k+1}) + \nabla g(x_k) + \frac{1}{\lambda}(x_{k+1} - y_k)$ . Equivalently,  $\frac{1}{\lambda}(x_k - x_{k+1}) + \nabla g(x_{k-1}) - 2\nabla g(x_k) \in \partial f(x_{k+1})$ , which together with the subdifferential sum rule [15, Exercise 8.8] implies that

$$p_{k+1} \in \partial f(x_{k+1}) + \nabla g(x_{k+1}).$$

Applying Lemma 3.4 with  $x = x_{k+1}$  and  $y = x_k$  yields

$$\partial H(z_k) = \left\{ \partial f(x_{k+1}) + \nabla g(x_{k+1}) + \left(\frac{1}{2\lambda} - \frac{L}{2}\right)(x_{k+1} - x_k) \right\} \times \left\{ \left(\frac{1}{2\lambda} - \frac{L}{2}\right)(x_k - x_{k+1}) \right\}.$$

It then follows that  $(A_{k+1}, B_{k+1}) \in \partial H(z_k)$  and

$$\begin{aligned} \|(A_{k+1}, B_{k+1})\| &\leq \|A_{k+1}\| + \|B_{k+1}\| \\ &= \left\| p_{k+1} + \left(\frac{1}{2\lambda} - \frac{L}{2}\right)(x_{k+1} - x_k) \right\| + \left\| \left(\frac{1}{2\lambda} - \frac{L}{2}\right)(x_k - x_{k+1}) \right\| \\ &\leq \|p_{k+1}\| + \left| \frac{1}{\lambda} - L \right| \|x_{k+1} - x_k\| \\ &\leq \left( \frac{1}{\lambda} + L + \left| \frac{1}{\lambda} - L \right| \right) \|x_{k+1} - x_k\| + L \|x_k - x_{k-1}\| \\ &\leq M_2 \|z_k - z_{k-1}\|, \end{aligned}$$

where the third inequality follows from the Lipschitz continuity of  $\nabla g$  with modulus  $L$ . □

We now connect the properties of FRB merit function with the actual objective of (3). Denote by  $\omega(z_{-1})$  the set of limit points of  $(z_k)_{k \in \mathbb{N}^*}$ .

**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 3.2 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  $l \rightarrow \infty$ . Then the following statements hold:*

- (i)  $\lim_{l \rightarrow \infty} H(z_{k_l}) = f(x^*) + g(x^*) = F(x^*)$ . In fact,  $\lim_{k \rightarrow \infty} H(z_k) = F(x^*)$ .
- (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  $\lim_{k \rightarrow \infty} \text{dist}(z_k, \omega(z_{-1})) = 0$ .

**Proof** (i) By the identity (10), for every  $k \in \mathbb{N}^*$



$$f(x_{k+1}) + \langle x_{k+1} - y_k, \nabla g(x_k) \rangle + \frac{1}{2\lambda} \|x_{k+1} - y_k\|^2 \leq f(x^*) + \langle x^* - y_k, \nabla g(x_k) \rangle + \frac{1}{2\lambda} \|x^* - y_k\|^2.$$

Combining the identity  $\|x^* - y_k\|^2 = \|x^* - x_{k+1}\|^2 + \|x_{k+1} - y_k\|^2 + 2\langle x^* - x_{k+1}, x_{k+1} - y_k \rangle$  with the above inequality and using the definition  $y_k = x_k + \lambda(\nabla g(x_{k-1}) - \nabla g(x_k))$ , one gets

$$\begin{aligned} f(x_{k+1}) &\leq f(x^*) + \langle x^* - x_{k+1}, \nabla g(x_k) \rangle + \frac{1}{2\lambda} \|x^* - x_{k+1}\|^2 + \frac{1}{\lambda} \langle x^* - x_{k+1}, x_{k+1} - y_k \rangle \\ &\leq f(x^*) + \langle x^* - x_{k+1}, 2\nabla g(x_k) - \nabla g(x_{k-1}) \rangle + \frac{1}{2\lambda} \|x^* - x_{k+1}\|^2 + \frac{1}{\lambda} \langle x^* - x_{k+1}, x_{k+1} - x_k \rangle. \end{aligned}$$

Replacing  $k$  by  $k_l$ , passing to the limit and applying Lemma 3.2(ii), we have

$$\limsup_{l \rightarrow \infty} f(x_{k_l+1}) \leq f(x^*) \leq \liminf_{l \rightarrow \infty} f(x_{k_l+1}),$$

where the last inequality is implied by the lower semicontinuity of  $f$ , which implies that  $\lim_{l \rightarrow \infty} f(x_{k_l+1}) = f(x^*)$ . Hence we conclude that

$$\lim_{l \rightarrow \infty} H(z_{k_l}) = \lim_{l \rightarrow \infty} \left( f(x_{k_l+1}) + g(x_{k_l+1}) + \left( \frac{1}{4\lambda} - \frac{L}{4} \right) \|x_{k_l+1} - x_{k_l}\|^2 \right) = f(x^*) + g(x^*),$$

where the last equality holds because of Lemma 3.2(ii). Moreover, Lemma 3.2(i) states that the function value sequence  $(H(z_k))_{k \in \mathbb{N}^*}$  converges, from which the rest of the statement readily follows.

(ii) By assumption, we have  $x_{k_l+1} \rightarrow x^*$  and  $x_{k_l} \rightarrow y^*$  as  $l \rightarrow \infty$ . Therefore  $0 \leq \|y^* - x^*\| \leq \|y^* - x_{k_l}\| + \|x_{k_l} - x_{k_l+1}\| + \|x_{k_l+1} - x^*\| \rightarrow 0$  as  $l \rightarrow \infty$ , which means that  $x^* = y^*$ . Combining Lemma 3.2, Lemma 3.5, statement(i) and the outer semicontinuity of  $z \mapsto \partial H(z)$ , one gets

$$0 \in \partial H(x^*, y^*) = \partial H(x^*, x^*) = \{ \partial f(x^*) + \nabla g(x^*) \} \times \{ 0 \} \Rightarrow 0 \in \partial f(x^*) + \nabla g(x^*),$$

as desired.

(iii) Given that  $(z_k)_{k \in \mathbb{N}^*}$  is bounded, it is easy to see that  $\omega(z_{-1}) = \bigcap_{l \in \mathbb{N}} \overline{\bigcup_{k \geq l} z_k}$  is a nonempty compact set. By using Lemma 3.2 and a result by Ostrowski [1, Theorem 1.49], the set  $\omega(z_{-1})$  is connected. Pick  $\tilde{z} \in \omega(z_{-1})$  and suppose that  $z_{k_q} \rightarrow \tilde{z}$  as  $q \rightarrow \infty$ . Then statement(i) implies that  $H(\tilde{z}) = \lim_{q \rightarrow \infty} H(z_{k_q}) = H(z^*)$ . Finally, by the definition of  $\omega(z_{-1})$ , we have  $\lim_{k \rightarrow \infty} \text{dist}(z_k, \omega(z_{-1})) = 0$ . □

Theorem 3.6 requires the sequence  $(z_k)_{k \in \mathbb{N}^*}$  to be bounded. The result below provides a sufficient condition to such an assumption.

**Theorem 3.7** *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 3.2 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 3.2(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 Convergence under the generalized concave KL property

Following basic properties of the FRB method, we now present the main convergence result under the generalized concave KL property. For  $\varepsilon > 0$  and nonempty set  $\Omega \subseteq \mathbb{R}^n$ , we define  $\Omega_\varepsilon = \{x \in \mathbb{R}^n : \text{dist}(x, \Omega) < \varepsilon\}$ . The following lemma will be useful soon.

**Lemma 3.8** [7, Lemma 4.4] *Let  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  be proper lsc and let  $\mu \in \mathbb{R}$ . Let  $\Omega \subseteq \text{dom} \partial f$  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  $h : (0, \eta) \rightarrow \mathbb{R}_+$  by*

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

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

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

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,

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

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

We are now ready for the main results. Our methodology is akin to the concave KL convergence mechanism employed by a vast amount of literature; see, e.g., [6, 8–13]. However, we make use of the generalized concave KL property and the associated exact modulus, which guarantees the sharpness of our results; see Remark 3.10.

**Theorem 3.9** (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  $0 < \lambda < \min\{\frac{1}{3L}, \lambda_f\}$ . 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  $i \geq k_0 + 1$*

$$\sum_{k=i}^{\infty} \|z_{k+1} - z_k\| \leq \|z_i - z_{i-1}\| + M\tilde{\varphi}(H(z_i) - H(z^*)). \tag{15}$$

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  $0 \in \partial F(x^*)$ .*

**Proof** (i) By the boundedness assumption, assume without loss of generality that  $z_k \rightarrow z^* = (x^*, y^*)$  for some  $z^* \in \mathbb{R}^n \times \mathbb{R}^n$ . Then Theorem 3.6 implies that  $x^* = y^*$  and  $H(z_k) \rightarrow F(x^*) = H(z^*)$  as  $k \rightarrow \infty$ . Recall from Lemma 3.2 that we have  $H(z_{k+1}) \leq H(z_k)$  for  $k \in \mathbb{N}^*$ , therefore one needs to consider two cases.

Case 1: Suppose that there exists  $k_0$  such that  $H(z^*) = H(z_{k_0})$ . Then Lemma 3.2(i) implies that  $z_{k_0+1} = z_{k_0}$  and  $x_{k_0+1} = x_{k_0}$ . The desired results then follows from a simple induction.

Case 2: Assume that  $H(z^*) < H(z_k)$  for every  $k$ . By assumption and Lemma 3.8, there exist  $\varepsilon > 0$  and  $\eta \in (0, \infty]$  such that the FRB merit function  $H$  has the setwise generalized concave KL property on  $\omega(z_{-1})$  with respect to  $\varepsilon > 0$  and  $\eta > 0$  and the associated exact modulus  $\tilde{\varphi}$ . On one hand, by the fact that  $H(z_k) \rightarrow H(z^*)$ , there exists  $k_1$  such that  $z_k \in [0 < H - H(z^*) < \eta]$  for  $k > k_1$ . On the other hand, Theorem 3.6(iii) implies that there exists  $k_2$  such that  $\text{dist}(z_k, \omega(z_{-1})) < \varepsilon$  for all  $k > k_2$ . Put  $k_0 = \max(k_1, k_2)$ . Then for  $k > k_0$ , we have  $z_k \in \omega(z_{-1})_\varepsilon \cap [0 < H - H(z^*) < \eta]$ . Hence for  $k > k_0$

$$(\tilde{\varphi})'_-(H(z_k) - H(z^*)) \cdot \text{dist}(0, \partial H(z_k)) \geq 1.$$

Invoking Lemma 3.5 yields

$$M_2(\tilde{\varphi})'_-(H(z_k) - H(z^*)) \|z_k - z_{k-1}\| \geq 1. \tag{16}$$

For simplicity, define for  $k > l$

$$\Delta_{k,k+1} = \tilde{\varphi}(H(z_k) - H(z^*)) - \tilde{\varphi}(H(z_{k+1}) - H(z^*)).$$

By the concavity of  $\tilde{\varphi}$  and (16), one has

$$\begin{aligned}
 \Delta_{k,k+1} &= \tilde{\varphi}(H(z_k) - H(z^*)) - \tilde{\varphi}(H(z_{k+1}) - H(z^*)) \\
 &\geq (\tilde{\varphi})'_-(H(z_k) - H(z^*)) \cdot [H(z_k) - H(z_{k+1})] \\
 &\geq \frac{H(z_k) - H(z_{k+1})}{M_2 \|z_k - z_{k-1}\|}.
 \end{aligned}
 \tag{17}$$

Applying Lemma 3.2 to (17) implies that

$$\Delta_{k,k+1} \geq \frac{\|z_{k+1} - z_k\|^2}{\|z_k - z_{k-1}\|} \cdot \frac{M_1}{M_2}.$$

Put  $M = \frac{M_2}{M_1}$ . Note that  $a^2 + b^2 \geq 2ab$  for  $a, b \geq 0$ . Then the above inequality gives

$$2\|z_{k+1} - z_k\| \leq 2\sqrt{M\Delta_{k,k+1}\|z_k - z_{k-1}\|} \leq M\Delta_{k,k+1} + \|z_k - z_{k-1}\|. \tag{18}$$

Pick  $i \geq k_0 + 1$ . Summing (18) from  $i$  to an arbitrary  $j > i$  yields

$$\begin{aligned}
 2 \sum_{k=i}^j \|z_{k+1} - z_k\| &\leq \sum_{k=i}^j \|z_k - z_{k-1}\| + M \sum_{k=i}^j \Delta_{k,k+1} \\
 &\leq \sum_{k=i}^j \|z_{k+1} - z_k\| + \|z_i - z_{i-1}\| + M\tilde{\varphi}(H(z_i) - H(z^*)) - M\tilde{\varphi}(H(z_{j+1}) - H(z^*)) \\
 &\leq \sum_{k=i}^j \|z_{k+1} - z_k\| + \|z_i - z_{i-1}\| + M\tilde{\varphi}(H(z_i) - H(z^*)),
 \end{aligned}$$

where the second inequality is implied by the definition of  $\Delta_{k,k+1}$ , implying

$$\sum_{k=i}^j \|z_{k+1} - z_k\| \leq \|z_i - z_{i-1}\| + M\tilde{\varphi}(H(z_i) - H(z^*)),$$

from which (15) readily follows. Moreover, one gets from the above inequality that for  $i \geq k_0 + 1$  and  $j > i$ ,

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

Recall from Theorem 3.6(i)&(ii) that  $H(z_i) \rightarrow H(z^*)$  and from Lemma 3.2(ii) that  $\|z_i - z_{i-1}\| \rightarrow 0$  as  $i \rightarrow \infty$ . Passing to the limit, one concludes that  $(z_k)_{k \in \mathbb{N}^*}$  is Cauchy.

(ii) The statement follows from the definition of  $(z_k)_{k \in \mathbb{N}^*}$  and Theorem 3.6(ii).  $\square$

**Remark 3.10** (i) Note that iterates distance was shown to be only square-summable in the original FRB paper [4, Theorem 2.5]. Therefore the finite length property is even new in the convex setting.

(ii) Unlike the usual concave KL convergence analysis, our approach uses the generalized concave KL property and the associated exact modulus to describe the

sharpest upper bound of  $\sum_{k=-1}^{\infty} \|z_{k+1} - z_k\|$ . To see this, note that the usual analysis would yield a similar bound as (15) with  $\tilde{\varphi}$  replaced by a concave desingularizing function associated with the concave KL property of  $H$ . Lemma 3.8 states that  $\tilde{\varphi}$  is the infimum of all associated concave desingularizing functions. Hence the upper bound (15) is the sharpest; see also [7] for a similar sharp result of the celebrated PALM algorithm.

A key assumption of Theorem 3.9 is the concave KL property of the FRB merit function  $H$  on  $\omega(z_{-1})$ . The class of semialgebraic functions provides rich examples of functions satisfying such an assumption.

**Corollary 3.11** *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  $0 < \lambda < \min\{\frac{1}{3L}, \lambda_f\}$ . 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 [8, 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 2.5 implies that the FRB merit function  $H$  is concave KL. Applying Theorem 3.9 then completes the proof. □

Assuming that the FRB merit function admits KL exponent  $\theta \in [0, 1)$ , we establish the following convergence rate result. Our analysis is standard and follows from the usual KL convergence rate methodology; see, e.g., [24, Theorem 2], [16, Remark 6] and [10, Lemma 4]. We provide a proof here for the sake of completeness.

**Theorem 3.12** (Convergence rate 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  $0 < \lambda < \min\{\frac{1}{3L}, \lambda_f\}$ . Let  $M, k_0$  and  $x^*$  be those given in Theorem 3.9. 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  $\theta \in (0, \frac{1}{2}]$ , 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  $\theta \in (\frac{1}{2}, 1)$ , 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.*

**Proof** Let  $z^* = (x^*, x^*)$ . Then Theorem 3.9 shows that  $z_k \rightarrow z^*$  as  $k \rightarrow \infty$ . Assume without loss of generality that  $H(z^*) = 0$  and  $H(z_k) > 0$  for all  $k$ . Before proving the desired statements, we will first develop several inequalities which will be used later. We have shown in Theorem 3.9 that for  $k > k_0$ ,

$$1 \leq (\tilde{\varphi})'_-(H(z_k)) \cdot \text{dist}(0, \partial H(z_k)) \leq \varphi'_-(H(z_k)) \cdot \text{dist}(0, \partial H(z_k)),$$

which by Lemma 3.5 and our assumption further implies that

$$1 \leq (1 - \theta)c(H(z_k))^{-\theta} \|(A_{k+1}, B_{k+1})\| \leq (1 - \theta)cM_2(H(z_k))^{-\theta} \|z_k - z_{k-1}\|. \tag{19}$$

Furthermore, define for  $k \in \mathbb{N}$

$$\sigma_k = \sum_{i=k}^{\infty} \|z_{i+1} - z_i\|,$$

which is well-defined due to Theorem 3.9. Assume again without loss of generality that  $\sigma_{k-1} - \sigma_k = \|z_k - z_{k-1}\| < 1$  for every  $k$  (recall Lemma 3.2). Rearranging (19) gives

$$H(z_k)^{1-\theta} \leq [(1 - \theta)cM_2]^{\frac{1-\theta}{\theta}} (\sigma_{k-1} - \sigma_k)^{\frac{1-\theta}{\theta}}. \tag{20}$$

By (15), we have for  $k > k_0$

$$\sigma_k \leq \sigma_{k-1} - \sigma_k + cM(H(z_k))^{1-\theta} \leq \sigma_{k-1} - \sigma_k + C(\sigma_{k-1} - \sigma_k)^{\frac{1-\theta}{\theta}}, \tag{21}$$

where the last inequality follows from (20) and  $C = cM((1 - \theta)cM_2)^{\frac{1-\theta}{\theta}}$ . Clearly  $\|x_k - x^*\| \leq \|z_k - z^*\| \leq \sigma_k$ . Hence, in order to prove the desired statements, it suffices to estimate  $\sigma_k$ .

(i) Let  $\theta = 0$  and suppose that  $(z_k)_{k \in \mathbb{N}^*}$  converges in infinitely many steps. Then (19) means that for every  $k > k_0$

$$1 \leq cM_2 \|z_k - z_{k-1}\|.$$

Passing to the limit and applying Lemma 3.2, one gets  $1 \leq 0$ , which is absurd.

(ii) If  $\theta \in (0, \frac{1}{2}]$ , then  $\frac{1-\theta}{\theta} \geq 1$  and  $(\sigma_{k-1} - \sigma_k)^{\frac{1-\theta}{\theta}} \leq \sigma_{k-1} - \sigma_k$ . Hence (21) implies that for  $k > k_0$

$$\sigma_k \leq (C + 1)(\sigma_{k-1} - \sigma_k) \Rightarrow \sigma_k \leq \frac{C + 1}{C + 2} \sigma_{k-1} \leq \left(\frac{C + 1}{C + 2}\right)^{k-k_0} \sigma_{k_0},$$

from which the desired result readily follows by setting  $Q_1 = \frac{C+1}{C+2}$  and  $c_1 = \left(\frac{C+1}{C+2}\right)^{-k_0} \sigma_{k_0}$ .

(iii) If  $\frac{1}{2} < \theta < 1$ , then  $0 < \frac{1-\theta}{\theta} < 1$  and  $(\sigma_{k-1} - \sigma_k)^{\frac{1-\theta}{\theta}} \geq \sigma_{k-1} - \sigma_k$ . It follows from (21) that  $\sigma_k \leq (1 + C)(\sigma_{k-1} - \sigma_k)^{\frac{1-\theta}{\theta}}$ . Define  $h(t) = t^{-\frac{\theta}{1-\theta}}$  for  $t > 0$ . Then

$$1 \leq (1 + C)^{\frac{\theta}{1-\theta}} (\sigma_{k-1} - \sigma_k) h(\sigma_k). \tag{22}$$

Let  $R > 1$  be a fixed real number. Next we consider two cases for  $k > k_0$ . If  $h(\sigma_k) \leq Rh(\sigma_{k-1})$  then (22) implies that

$$\begin{aligned} 1 &\leq R(1 + C)^{\frac{\theta}{1-\theta}} (\sigma_{k-1} - \sigma_k) h(\sigma_{k-1}) \leq R(1 + C)^{\frac{\theta}{1-\theta}} \int_{\sigma_k}^{\sigma_{k-1}} h(t) dt \\ &= R(1 + C)^{\frac{\theta}{1-\theta}} \frac{1 - \theta}{1 - 2\theta} \left( \sigma_{k-1}^{\frac{1-2\theta}{1-\theta}} - \sigma_k^{\frac{1-2\theta}{1-\theta}} \right). \end{aligned}$$

Set  $v = \frac{1-2\theta}{1-\theta} < 0$ . Then the above inequality can be rewritten as

$$\sigma_k^v - \sigma_{k-1}^v \geq -\frac{v}{R} (1 + C)^{-\frac{\theta}{1-\theta}}. \tag{23}$$

If  $h(\sigma_k) > Rh(\sigma_{k-1})$  then  $\sigma_k < \left(\frac{1}{R}\right)^{\frac{1-\theta}{\theta}} \sigma_{k-1} = q\sigma_{k-1}$ , where  $q = \left(\frac{1}{R}\right)^{\frac{1-\theta}{\theta}} \in (0, 1)$ . Hence

$$\sigma_k < q\sigma_{k-1} \Rightarrow \sigma_k^v > q^v \sigma_{k-1}^v \Rightarrow \sigma_k^v - \sigma_{k-1}^v > (q^v - 1)\sigma_{k-1}^v > (q^v - 1)\sigma_{k_0}^v. \tag{24}$$

Set  $c_2 = \min \left\{ (q^v - 1)\sigma_{k_0}^v, -\frac{v}{R} (1 + C)^{-\frac{\theta}{1-\theta}} \right\} > 0$ . Combining (23) and (24) yields

$$\sigma_k^v - \sigma_{k-1}^v \geq c_2, \forall k > k_0.$$

Summing the above inequality from  $k_0 + 1$  to any  $k > k_0$ , one gets

$$\sigma_k^v - \sigma_{k_0}^v = \sum_{i=k_0+1}^k (\sigma_i^v - \sigma_{i-1}^v) \geq (k - k_0)c_2 \Rightarrow \sigma_k^v > \frac{k}{k_0 + 1} c_2,$$

where the last inequality holds because  $k - k_0 \geq \frac{k}{k_0+1}$  for  $k > k_0 + 1$ . Hence

$$\sigma_k < k^{\frac{1}{v}} \left( \frac{C_2}{k_0 + 1} \right)^{\frac{1}{v}},$$

from which the desired statement follows by setting  $Q_2 = \left(\frac{c_2}{k_0+1}\right)^{\frac{1}{v}}$ . □

The following corollary asserts that convergence rates can be directly deduced from the KL exponent of the objective under certain conditions, which is a consequence of [25, Theorem 3.6], Theorems 3.6 and 3.12.

**Corollary 3.13** *Let  $(x_k)_{k \in \mathbb{N}^*}$  be a sequence generated by FRB and suppose that all conditions of Theorem 3.9 are satisfied. Let  $x^*$  be as in Theorem 3.9. Assume further that  $F$  has KL exponent  $\theta \in [\frac{1}{2}, 1)$  at  $x^*$ . Then the following statements hold:*

- (i) *If  $\theta = \frac{1}{2}$ , 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.

(ii) If  $\theta \in \left(\frac{1}{2}, 1\right)$ , 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.

**Proof** The FRB merit function  $H$  has KL exponent  $\theta$  at  $(x^*, x^*)$  by our assumption and [25, Theorem 3.6]. Applying Theorem 3.12 completes the proof.  $\square$

Following similar techniques in [18, 26], we now deduce convergence rates of the merit function value sequence.

**Theorem 3.14** (Convergence rate of merit 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 3.9(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  $\theta \in \left(0, \frac{1}{2}\right]$ , 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  $\theta \in \left(\frac{1}{2}, 1\right)$ , 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** We learn from Theorems 3.6 and 3.9 that the sequence  $(H(z_k))_{k \in \mathbb{N}}$  is decreasing,  $H(z^k) \rightarrow H(z^*) = F(x^*)$  and  $z_k \rightarrow z^*$  as  $k \rightarrow \infty$ . Therefore  $(e_k)_{k \in \mathbb{N}}$  converges to 0 and  $e_k \geq 0$  for all  $k$ . By the KL exponent assumption, there exist  $c > 0$  and  $k_0$  such that for  $k \geq k_0$

$$\text{dist}(0, \partial H(z_k)) \geq \frac{(H(z_k) - H(z^*))^\theta}{c(1 - \theta)} = \frac{e_k^\theta}{c(1 - \theta)}.$$

Therefore by using Lemmas 3.2 and 3.5



$$e_{k-1} - e_k = H(z_{k-1}) - H(z_k) \geq M_1 \|z_k - z_{k-1}\|^2 \geq \frac{M_1}{M_2^2} \text{dist}(0, \partial H(z_k))^2 \geq C e_k^{2\theta},$$

where  $C = \frac{M_1}{c^2(1-\theta)^2 M_2^2}$ . Applying [18, Lemma 10] completes the proof. □

**Corollary 3.15** (Convergence rate of objective 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  $0 < \lambda < \min\{\frac{1}{3L}, \lambda_f\}$ . Let  $x^*$  be the limit given in Theorem 3.9(ii). Suppose that the FRB merit function  $H$  has KL exponent  $\theta \in [0, 1)$  at  $z^* = (x^*, x^*)$ . Then the following statements hold:*

(i) *If  $\theta = 0$ , then  $F(x_k)$  converges to  $F(x^*)$  in finite steps.*

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

$$|F(x_k) - F(x^*)| \leq \bar{c}_1 \bar{Q}_1^k.$$

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

$$|F(x_k) - F(x^*)| \leq \bar{c}_2 k^{-\frac{1-\theta}{2\theta-1}}.$$

**Proof** Observe that

$$|F(x_k) - F(x^*)| \leq |H(x_k, x_{k-1}) - F(x^*)| + \left| \frac{1}{4\lambda} - \frac{L}{4} \right| \|x_k - x_{k-1}\|^2 \tag{25}$$

$$\leq |H(x_k, x_{k-1}) - F(x^*)| + \left| \frac{1}{2\lambda} - \frac{L}{2} \right| (\|x_k - x^*\|^2 + \|x_{k-1} - x^*\|^2). \tag{26}$$

It suffices to apply Theorems 3.12 and 3.14. □

Let  $\text{Proj}_C$  be the projection mapping associated with a nonempty closed subset  $C \subseteq \mathbb{R}^n$ .

**Example 3.16** 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 (3) with  $f = \delta_D$  and  $g = \frac{1}{2} \text{dist}^2(\cdot, C)$ . 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  $x_k \rightarrow x^*$  and  $0 \in \partial F(x^*)$ .
- (ii) Suppose, in addition, that

$$N_C(\text{Proj}_C(x^*)) \cap (-N_D(x^*)) = \{0\}. \tag{27}$$

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.

**Proof** (i) By the compactness assumption, the function  $f + g$  is coercive. Hence Theorem 3.7 implies that  $(x_k)_{k \in \mathbb{N}^*}$  is bounded. We assume that sets  $C, D$  are semi-algebraic, then so are functions  $f$  and  $g$ ; see [8, Section 4.3] and [9, Lemma 2.3]. Moreover, note that  $\lambda_f = \infty$  and  $g$  admits a Lipschitz continuous gradient with constant  $L = 1$ ; see [15, Exercise 1.24] and [1, Corollary 12.31]. The desired result then follows from Corollary 3.11.

(ii) Taking the fact that  $0 \in \partial F(x^*)$  into account and applying the subdifferential sum rule [15, Exercise 8.8], one gets

$$x^* - \text{Proj}_C(x^*) \in -N_D(x^*).$$

The constraint qualification (27) then implies that  $x^* = \text{Proj}_C(x^*)$  and consequently  $x^* \in C$ . Moreover, the FRB scheme together with the closedness of  $D$  ensures that  $x^* \in D$ , hence  $x^* \in C \cap D$ . Consequently, notice that the constraint qualification (27) amounts to

$$N_C(x^*) \cap (-N_D(x^*)) = \{0\}.$$

Then a direct application of [27, Theorem 5] guarantees that  $F$  has KL exponent  $\theta = \frac{1}{2}$  at  $x^*$ . The desired result follows immediately from Corollary 3.13.  $\square$

### 4 Numerical experiments

In this section, we apply the FRB splitting method to nonconvex feasibility problems. Let  $C = \{x \in \mathbb{R}^n : Ax = b\}$  for  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $r = \lceil m/5 \rceil$ ,  $l = 10^6$ , and  $D = \{x \in \mathbb{R}^n : \|x\|_0 \leq r, \|x\|_\infty \leq l\}$ . Similarly to [6, Section 5], we consider the minimization problem (3) with

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

Clearly  $C$  is semialgebraic. As for  $D$ , first notice that  $\mathbb{R}^n \times \{i\}$  for  $i \in \mathbb{R}$  and  $\text{gph}\|\cdot\|_0$  are semialgebraic; see [16, Example 3]. Then

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

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 [28, 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 3.16.

We find a projection onto  $D$  by the formula given below, which is a consequence of [29, Proposition 3.1] and was already observed by Li and Pong [6]. We provide a proof for the sake of completeness.

**Proposition 4.1** *Let  $z = (z_1, \dots, z_n) \in \mathbb{R}^n$ . For every  $i$ , set  $\tilde{z}_i^* = \text{Proj}_{[-l,l]}(z_i)$ ,  $v_i^* = |z_i|^2 - |\tilde{z}_i^* - z_i|^2$ , and let  $I^* \subseteq \{1, \dots, n\}$  be the set of indices corresponding to the  $r$  largest elements of  $v_i^*$ ,  $i = 1, \dots, n$ . Define  $z^* \in \mathbb{R}^n$  by  $z_i^* = \tilde{z}_i^*$  if  $i \in I^*$  and  $z_i^* = 0$  otherwise. Then  $z^* \in \text{Proj}_D(z)$ .*

**Proof** Apply [29, Proposition 3.1] with  $\phi_i = |\cdot - z_i|^2$  and  $\mathcal{X}_i = [-l, l]$ . □

We shall benchmark FRB against the Douglas-Rachford method with fixed step-size (DR) [6] by Li and Pong, inertial Tseng’s method (iT seng) [12] by Boř and Csetnek, and DR equipped with step-size heuristics (DRh) [6] by Li and Pong. These splitting algorithms for nonconvex optimization problems are known to converge globally to a stationary point of (3) under appropriate assumptions on the concave KL property of merit functions; see [6, Theorems 1–2, Remark 4, Corollary 1] and [12, Theorem 3.1]. The convergence of DR and iT seng in our setting are already proved in [6, 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  $\{x \in \mathbb{R}^n : Ax = b\}$

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

where the step-size  $\lambda = 0.9999 \cdot \frac{1}{3}$  (recall Example 3.16). 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  $\alpha = \frac{1}{8}$ :

$$\begin{aligned} p_{k+1} &\in \text{Proj}_D(x_k - \lambda' A^\dagger (Ax_k - b) + \alpha(x_k - x_{k-1})), \\ x_{k+1} &= p_k + \lambda' A^\dagger A(x_k - p_k). \end{aligned}$$

As for DR and DRh, we employ the schemes specified by [6, Scheme (7)] and [6, Section 5] with the exact same step-sizes, respectively. All algorithms are initialized at the origin, and we terminate FRB and iT seng when

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

We adopt the termination criteria from [6, Section 5] for DR and DRh, where the same tolerance of  $10^{-8}$  is applied. Similar to [6, 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  $C \cap D \neq \emptyset$ .

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 ( $\text{fval}_{\min}$ ), 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 [4, Remark 2.8].
- DRh has the most number of successes and the best precision at termination (see  $\text{fval}_{\min}$ ), but tend to be slower on "easy" problems (large  $m$ ).

Therefore, FRB is a competitive method for finding a sparse solution of a linear system, at least among the aforementioned algorithms with a fixed step-size.<sup>2</sup>

## 5 Conclusion and future work

We established convergence of the Malitsky-Tam FRB splitting algorithm in the nonconvex setting. Under the generalized concave KL property, we showed that FRB converges globally to a stationary point of (3) and admits finite length property, which is even new in the convex setting. The sharpness of our results is demonstrated by virtue of the exact modulus associated with the generalized concave KL property. We also analyzed both function value and sequential convergence rates of FRB when desingularizing functions associated with the generalized concave KL property have the Łojasiewicz form. Numerical simulations suggest that FRB is a competitive method compared to DR and inertial Tseng's methods.

As for future work, it is tempting to analyze convergence rate using the exact modulus, as the usual KL convergence rate analysis assumes desingularizing functions of the Łojasiewicz form, which could be an overestimation. Another direction is to improve FRB through heuristics or other step-size strategies such as FISTA. Finally, as anonymous referees kindly pointed out, inertia techniques have been widely adopted to improve nonconvex proximal-type algorithms in the KL framework; see, e.g., [12, 30] and the references therein. Therefore, it is also tempting to explore an inertial type FRB for nonconvex optimization problems.

<sup>2</sup> We also performed simulations with much larger problem size ( $n = 4000, 5000, 6000$ ), in which case FRB, DR, and iTseng tend to stuck at stationary points while DRh can still hit global minimizers. We believe that this is due to its heuristics; see also [6, Section 5].

**Table 1** Comparing FRB to DR, iTseng, and DRh

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

**Acknowledgements** The authors thank the editor and the reviewers for helpful suggestions and constructive feedback which helped us to improve the presentation of the results. XW and ZW were partially supported by NSERC Discovery Grants.

**Data availability** The data that support the findings of Table 1 are available from the corresponding author upon request.

## References

1. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, Cham (2017)
2. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simulat.* **4**, 1168–1200 (2005)
3. Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control. Optim.* **38**, 431–446 (2000)
4. Malitsky, Y., Tam, M.K.: A forward-backward splitting method for monotone inclusions without cocoercivity. *SIAM J. Optim.* **30**, 1451–1472 (2020)
5. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.* **2**, 183–202 (2009)
6. Li, G., Pong, T.K.: Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems. *Math. Program.* **159**, 371–401 (2016)
7. Wang, X., Wang, Z.: The exact modulus of the generalized Kurdyka-Łojasiewicz property. *Math. Oper. Res.* (2021). [arXiv:2008.13257](https://arxiv.org/abs/2008.13257), to appear
8. 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**, 438–457 (2010)
9. 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**, 91–129 (2013)
10. Banert, S., Boţ, R.I.: A general double-proximal gradient algorithm for dc programming. *Math. Program.* **178**, 301–326 (2019)
11. 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**, 2131–2151 (2018)
12. Boţ, R.I., Csetnek, E.R.: An inertial Tseng’s type proximal algorithm for nonsmooth and nonconvex optimization problems. *J. Optim. Theory Appl.* **171**, 600–616 (2016)
13. Li, G., Liu, T., Pong, T.K.: Peaceman-Rachford splitting for a class of nonconvex optimization problems. *Comput. Optim. Appl.* **68**, 407–436 (2017)
14. Mordukhovich, B.S.: *Variational Analysis and Generalized Differentiation I: Basic Theory*. Springer-Verlag, Berlin (2006)
15. Rockafellar, R.T., Wets, R.J.-B.: *Variational Analysis*. Springer-Verlag, Berlin (1998)
16. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.* **146**, 459–494 (2014)
17. Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. *SIAM J. Optim.* **18**, 556–572 (2007)
18. Boţ, R.I., Nguyen, D.-K.: The proximal alternating direction method of multipliers in the nonconvex setting: convergence analysis and rates. *Math. Oper. Res.* **45**, 682–712 (2020)
19. Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. *Trans. Am. Math. Soc.* **362**, 3319–3363 (2010)
20. Łojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles* **117**, 87–89 (1963)
21. Kurdyka, K.: On gradients of functions definable in o-minimal structures. *Annales de l’institut Fourier* **48**, 769–783 (1998)
22. Ochs, P., Chen, Y., Brox, T., Pock, T.: iPiano: inertial proximal algorithm for nonconvex optimization. *SIAM J. Imag. Sci.* **7**, 1388–1419 (2014)

23. Böhm, A., Sedlmayer, M., Csetnek, E. R., Boţ, R. I.: Two steps at a time-taking gan training in stride with Tseng's method, [arXiv:2006.09033](https://arxiv.org/abs/2006.09033), (2020)
24. Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Math. Program.* **116**, 5–16 (2009)
25. Li, G., Pong, T.K.: Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods. *Found. Comput. Math.* **18**, 1199–1232 (2018)
26. Boţ, R.I., Csetnek, E.R., Nguyen, D.-K.: A proximal minimization algorithm for structured nonconvex and nonsmooth problems. *SIAM J. Optim.* **29**, 1300–1328 (2019)
27. Chen, C., Pong, T.K., Tan, L., Zeng, L.: A difference-of-convex approach for split feasibility with applications to matrix factorizations and outlier detection. *J. Global Optim.* **78**, 107–136 (2020)
28. Bauschke, H.H., Luke, D.R., Phan, H.M., Wang, X.: Restricted normal cones and sparsity optimization with affine constraints. *Found. Comput. Math.* **14**, 63–83 (2014)
29. Lu, Z., Zhang, Y.: Sparse approximation via penalty decomposition methods. *SIAM J. Optim.* **23**, 2448–2478 (2013)
30. Boţ, R.I., Csetnek, E.R., László, S.C.: An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions. *EURO J. Comput. Optimiz.* **4**, 3–25 (2016)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.