Journal of Optimization Theory and Applications
https://doi.org/10.1007/s10957-020-01787-7

On the Linear Convergence of Forward–Backward Splitting Method: Part I—Convergence Analysis

Yunier Bello-Cruz[1] · **Guoyin Li**[2] · **Tran T. A. Nghia**[3]

Received: 19 March 2019 / Accepted: 9 November 2020

Abstract

In this paper, we study the complexity of the forward–backward splitting method with Beck–Teboulle's line search for solving convex optimization problems, where the objective function can be split into the sum of a differentiable function and a nonsmooth function. We show that the method converges weakly to an optimal solution in Hilbert spaces, under mild standing assumptions without the global Lipschitz continuity of the gradient of the differentiable function involved. Our standing assumptions is weaker than the corresponding conditions in the paper of Salzo (SIAM J Optim 27:2153–2181, 2017). The conventional complexity of sublinear convergence for the functional value is also obtained under the local Lipschitz continuity of the gradient of the differentiable function. Our main results are about the linear convergence of this method (in the quotient type), in terms of both the function value sequence and the iterative sequence, under only the quadratic growth condition. Our proof technique is direct from the quadratic growth conditions and some properties of the forward–backward splitting method without using error bounds or Kurdya-Łojasiewicz inequality as in other publications in this direction.

Keywords Nonsmooth and convex optimization problems · Forward–Backward splitting method · Linear convergence · Quadratic growth condition

Mathematics Subject Classification 65K05 · 90C25 · 90C30

✉ Guoyin Li
g.li@unsw.edu.au

Yunier Bello-Cruz
yunierbello@niu.edu

Tran T. A. Nghia
nttran@oakland.edu

[1] Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115, USA

[2] Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia

[3] Department of Mathematics and Statistics, Oakland University, Rochester, MI 48309, USA

Published online: 28 November 2020

1 Introduction

In this paper, we mainly consider an optimization problem of minimizing the sum of two convex functions, one of which is differentiable and the other is nondifferentiable. Many problems in this format have appeared in different fields of science and engineering including machine learning, compressed sensing, and image processing. A particular class of this problem is the so-called ℓ_1-regularized problem that usually provides sparse solutions with many applications in signal processing and statistics [1–3].

Among many methods solving the aforementioned optimization problems, the forward–backward splitting method (FBS in brief, known also as the proximal gradient method) [2,4–9] is very popular due to its simplicity and efficiency. It is well-known that this method is globally convergent to an optimal solution with the complexity $\mathscr{O}(k^{-1})$ on the iterative cost values under the assumption that the gradient of the differentiable function is globally Lipschitz continuous [6]. An important advance was due to [10], where the sublinear convergence was improved to $o(k^{-1})$ in Hilbert spaces. Independently, by using some line searches motivated from the work of Tseng [11] and under a more restricted step size rule, Bello-Cruz and Nghia [12] achieve the same complexity with the milder assumption that the aforementioned gradient is only *locally* Lipschitz continuous in finite-dimensional spaces. The results in [12] were further extended in [13] to different line searches for the FBS method and more general classes of optimization problems even in infinite-dimensional spaces. A recent work of Bauschke, Bolte, and Teboulle [14] also tackles the absence of Lipschitz continuous gradient by introducing the so-called NoLips algorithm close to the FBS method with the involvement of Bregman distance. Their algorithm provides the sublinear complexity $\mathscr{O}(\frac{1}{k})$ and guarantees the global convergence with an additional hypothesis about the closedness of the domain of an auxiliary Legendre function defined there. Unfortunately, the latter assumption as well as those in [12] are not satisfied for the Poisson inverse regularized problems with Kullback–Leibler divergence [15,16], one of the main applications in [14]. This situation was overcome in [13] by using the FBS method and would be revisited again in our sequence [17] with further advanced achievements on the linear convergence in the quotient type (often referred as Q-linear convergence). Indeed, Salzo in [13, Section 4] proposed some new hypotheses to avoid the standard assumptions [6,10,12] on the domain of cost functions. Although these hypotheses are valid in many natural situations of optimization problems, some nontrivial parts of them can be further relaxed. In Sect. 3, by reanalyzing the theory of the FBS method in [12,13], we relax and weaken several conditions assumed in [13] to acquire the same global convergence and sublinear complexity in Hilbert spaces.[1] The proof of many results in this section modifies our corresponding ones in [12] under the simplified standing assumptions. This section is not a major part of the paper, but the corresponding results will be used intensively in our main Sect. 4.

The central part of our paper, Sect. 4 devotes to the linear convergence of the FBS method without the global Lipschitz continuity of gradient mentioned above in finite

[1] We are indebted to some remarks from one referee that allow us to extend the results from finite dimensions to Hilbert spaces in the current version.

dimensions. Despite the popularity of the FBS method, the linear convergence of this method has been established recently by using Kurdya-Łojasiewicz inequality [18–22] (even for nonconvex optimization problems) or some error bound conditions [23–27] with the base from [28]. Those conditions are somehow equivalent in convex settings; see, e.g., [18,19,25]. Linear convergence of the FBS method was also obtained in Hilbert spaces [5] for Lasso problems under some nontrivial conditions such as *finite basis injectivity* and *strict sparsity pattern*. These conditions were fully relaxed in the papers [19,20], which also work in infinite dimensions for more general optimization problems. Our results can be also extended to infinite dimensions by following some ideas of [19], but for simplicity, we restrict ourselves on finite dimensions in this section. Our approach is close to the works of Drusvyatskiy and Lewis [25] and Garrigos, Rosasco, and Villa [19,20] by using the so-called *quadratic growth condition* known also as *2-conditioned property*. However, our proof of linear convergence is more direct just from quadratic growth condition without using KŁ inequality as in [18–21] or the error bound [28] as in [23,25] and reveals the Q-linear convergence for the FBS sequence rather than the R-one obtained in all the aforementioned works. Some of our linear rates are sharper than those in [19,25]. The property of quadratic growth condition is indeed automatic in many classes of optimization problems including Poisson inverse regularized problem [14–16], least-square ℓ_1 regularized problems [3,5,6,19], group Lasso [23] or under mild second-order conditions on initial data [2,30–34]; see our second part [17] for further studies in this direction.

2 Preliminaries

Throughout the paper, $\mathscr{H}$ is a Hilbert space, where $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ denote the corresponding norm and inner product in $\mathscr{H}$. We use $\Gamma_0(\mathscr{H})$ to denote the set of proper, lower semicontinuous, and convex functions on $\mathscr{H}$. Let $h \in \Gamma_0(\mathscr{H})$, we write $\operatorname{dom} h := \{x \in \mathscr{H} : h(x) < +\infty\}$. The subdifferential of h at $\bar{x} \in \operatorname{dom} h$ is defined by

$$\partial h(\bar{x}) := \{v \in \mathscr{H} : \langle v, x - \bar{x}\rangle \le h(x) - h(\bar{x}),\ x \in \mathscr{H}\}. \tag{1}$$

We say h satisfies the *quadratic growth condition* at $\bar{x}$ with modulus $\kappa > 0$ if there exists $\varepsilon > 0$ such that

$$h(x) \ge h(\bar{x}) + \frac{\kappa}{2} d^2(x; (\partial h)^{-1}(0)) \quad \text{for all} \quad x \in \mathbb{B}_\varepsilon(\bar{x}). \tag{2}$$

Moreover, if in additionally $(\partial h)^{-1}(0) = \{\bar{x}\}$, h is said to satisfy the *strong* quadratic growth condition at $\bar{x}$ with modulus κ.

Some relationship between the quadratic growth condition and the so-called metric subregularity of the subdifferential could be found in [18,29–31,34] even for the case of nonconvex functions. The quadratic growth condition (2) is also called *quadratic functional growth* property in [26] when h is continuously differentiable over a closed convex set. In [19,20], h is said to be 2-conditioned on $\mathbb{B}_\varepsilon(\bar{x})$ if it satisfies the quadratic growth condition (2).

The quadratic growth condition is slightly different in [25] as follows:

$$h(x) \geq h(\bar{x}) + \frac{c}{2} d^2(x; (\partial h)^{-1}(0)) \quad \text{for all} \quad x \in [h < h^* + \nu] \tag{3}$$

for some constants $c, \nu > 0$, where $[h < h^* + \nu] := \{x \in \mathscr{H} : h(x) < h^* + \nu\}$ and $h^* := \inf h(x) = h(\bar{x})$. It is easy to check that this growth condition implies (2). Indeed, suppose that (3) is satisfied for some $c, \nu > 0$ Define $\eta := \sqrt{\frac{2\nu}{c}}$ and note that $\bar{x} \in [h < h^* + \nu]$. Take any $x \in \mathbb{B}_\eta(\bar{x})$, if $x \in [h < h^* + \nu]$, inequality (2) is trivial. If $x \notin [h < h^* + \nu]$, it follows that

$$h(x) \geq h(\bar{x}) + \nu = h(\bar{x}) + \frac{c}{2}\eta^2 \geq h(\bar{x}) + \frac{c}{2}\|x - \bar{x}\|^2 \geq h(\bar{x}) + \frac{c}{2} d^2(x; (\partial h)^{-1}(0)),$$

which clearly verifies (2). Thus, (2) is weaker than (3), but it is equivalent to the local version of (3):

$$h(x) \geq h(\bar{x}) + \frac{c}{2} d^2(x; (\partial h)^{-1}(0)) \quad \text{for all} \quad x \in [h < h^* + \nu] \cap \mathbb{B}_\varepsilon(\bar{x})$$

for some constants $c, \nu, \varepsilon > 0$. This property has been showed recently in [18, Theorem 5] to be equivalent to the fact that h satisfies the Kurdyka-Łojasiewicz inequality with order $\frac{1}{2}$.

To complete this section, we recall two important notions of linear convergence in our study. Let μ be a real number such that $0 < \mu < 1$. A sequence $(x^k)_{k \in \mathbb{N}} \subset \mathscr{H}$ is said to be R-linearly convergent to x^* with rate μ if there exists $M > 0$ such that

$$\|x^k - x^*\| \leq M\mu^k = \mathscr{O}(\mu^k) \quad \text{for all} \quad k \in \mathbb{N}.$$

We say $(x^k)_{k \in \mathbb{N}}$ is *Q-linearly convergent* to x^* with rate μ if there exists $K \in \mathbb{N}$ with

$$\|x^{k+1} - x^*\| \leq \mu \|x^k - x^*\| \quad \text{for all} \quad k \geq K.$$

From the definitions, it can be directly verified that Q-linear convergence implies R-linear convergence with the same rate. On the other hand, in general, R-linear convergence does not imply Q-linear convergence.

3 Global Convergence of Forward–Backward Splitting Methods in Hilbert Spaces

Throughout the paper, we consider the following optimization problem

$$\min_{x \in \mathscr{H}} \quad F(x) := f(x) + g(x), \tag{4}$$

where $f, g \in \Gamma_0(\mathscr{H})$ and f is differentiable on $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$; see our standing assumptions below. This section reanalyzes the theory for forward–backward splitting

(FBS) methods:

$$x^{k+1} = \operatorname{prox}_{\alpha_k g}(x^k - \alpha_k \nabla f(x^k)) \tag{5}$$

with the proximal operator defined later in (8) and the stepsize $\alpha_k > 0$, when the gradient ∇f is not globally Lipschitz continuous [12,13]. The results here will be employed in our main topic about the linear convergence of the FBS method in Sect. 4.

The convergence of the FBS method without Lipschitz continuity of ∇f was studied in [12] via some line searches with the standard assumption $\operatorname{dom} g \subset \operatorname{dom} f$ together with some assumptions on the continuity of ∇f. In [13], Salzo simplified these assumptions and significantly extended it to more general frameworks via many line searches. It is worth noting that the standard assumption $\operatorname{dom} g \subset \operatorname{dom} f$, used in [12], is not satisfied by several important optimization problems including the Poisson inverse regularized problems with Kullback–Leibler divergence [15,16]. This situation was overcome in [13, Section 4 and 5] with the following assumptions on f and g:

H1. $f, g \in \Gamma_0(\mathscr{H})$ are bounded from below with $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g \neq \emptyset$.

H2. f is differentiable on $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$, ∇f is uniformly continuous on any weakly compact subset of $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$, and ∇f is bounded on any sublevel sets of F.

H3. For $x \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$, $\{F \leq F(x)\} \subset \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$, and

$$d(\{F \leq F(x)\}; \mathscr{H} \setminus \operatorname{int}(\operatorname{dom} f)) > 0. \tag{6}$$

These assumptions hold under some mild conditions; see [13, Proposition 4.2] for details. However, they are not trivial even in finite dimensions. Throughout the paper, we focus on the FBS method with the popular Beck–Teboulle line search [6] that was also studied in [13]. Next, following the techniques used in [12,13], we show that the global convergence of the FBS method holds true under the below simplified standing assumptions that relax some conditions from **H1–H3**:

A1. $f, g \in \Gamma_0(\mathscr{H})$ and $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g \neq \emptyset$.

A2. f is differentiable at any point in $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$, ∇f is uniformly continuous on any weakly compact subset of $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$.

A3. For any $x \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$, $\{F \leq F(x)\} \subset \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$ and for any weakly compact subset $\mathscr{X}$ of $\{F \leq F(x)\}$ we have

$$d(\mathscr{X}; \mathscr{H} \setminus \operatorname{int}(\operatorname{dom} f)) > 0. \tag{7}$$

Remark 3.1 (The standing assumptions in finite dimensions) When $\dim \mathscr{H} < \infty$, **A2** means that f is continuously differentiable on $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$. Moreover, in **A3** the distance gap requirement (7) is automatically true as $\mathscr{X} \cap (\mathscr{H} \setminus \operatorname{int}(\operatorname{dom} f)) = \emptyset$ and $\mathscr{X}$ is compact. On the other hand, the condition (6) is still nontrivial; see the example below.

Clearly, some boundedness assumptions in **H1–H2** are relaxed in our **A1–A2**. Moreover, the positive distance gap requirement (6) in (**H3**) is slightly stronger than (7) in **(A3)**. But, it should be noted that condition (6) can, in general, be easier to verify in infinite-dimensional spaces. Even in finite dimensions, **(A1–A3)** are strictly weaker than (**H1**–**H3**). To see this, consider $p \in \{1, 2\}$,

$$C_p := \{x = (x_1, x_2) \in \mathbb{R}^2 \colon x_1 x_2 \geq p, x_1 > 0, x_2 > 0\},$$

and $f, g : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$ be defined as $f(x) = -\log x_1 - \log x_2$ if $x \in \mathbb{R}^2_{++}$ and $+\infty$ otherwise, and $g(x) = \delta_{C_1}(x)$ where δ_{C_1} is the indicator function of the set C_1 defined above. It can be directly verified that the assumptions (**A1**–**A3**) are satisfied. On the other hand, **H1** fails because f is unbounded on

$$\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g = \mathbb{R}^2_{++} \cap C_1 = C_1 = \{(x_1, x_2) \\ \in \mathbb{R}^2 \colon x_1 x_2 \geq 1, x_1 > 0, x_2 > 0\}.$$

Moreover, $(2, 1) \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$ and since $F(2, 1) = (f + g)(2, 1) = -\log 2$, we get

$$\{x \colon F(x) \leq -\log 2\} = \{(x_1, x_2) \in \mathbb{R}^2 \colon x_1 x_2 \geq 2, x_1 > 0, x_2 > 0\} = C_2.$$

So, clearly **H2** also fails because $\nabla f(x_1, x_2) = (-\frac{1}{x_1}, -\frac{1}{x_2})$ is unbounded on the sublevel set of F above. Finally, observing that $(\frac{2}{k}, k) \in C_2$ and $(0, k) \in \mathbb{R}^2 \backslash \operatorname{int}(\operatorname{dom} f) = \mathbb{R}^2 \backslash \mathbb{R}^2_{++}$, we see that

$$d\big(\{F \leq -\log 2\}; \mathbb{R}^2 \backslash \operatorname{int}(\operatorname{dom} f)\big) \\ = d\big(C_2; \mathbb{R}^2 \backslash \mathbb{R}^2_{++}\big) \leq \lim_{k \to \infty} \left\| (\frac{2}{k}, k) - (0, k) \right\| = \lim_{k \to \infty} \frac{2}{k} = 0.$$

This shows that **H3** fails in this case.

Next, let us recall the proximal operator $\operatorname{prox}_g : \mathscr{H} \to \operatorname{dom} g$ given by

$$\operatorname{prox}_g(z) := (\operatorname{Id} + \partial g)^{-1}(z) \quad \text{for all} \quad z \in \mathscr{H}, \tag{8}$$

which is well-known to be a single-valued mapping with full domain. With $\alpha > 0$, it is easy to check that

$$\frac{z - \operatorname{prox}_{\alpha g}(z)}{\alpha} \in \partial g(\operatorname{prox}_{\alpha g}(z)) \quad \text{for all} \quad z \in \mathscr{H}. \tag{9}$$

Let S^* be the optimal solution set to problem (4) and x^* be in $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$ due to our assumption **A3**. Then, $x^* \in S^*$ iff

$$0 \in \partial(f + g)(x^*) = \nabla f(x^*) + \partial g(x^*). \tag{10}$$

The following lemma, a consequence of [4, Theorem 23.47] is helpful in our proof of the finite termination of Beck–Teboulle's line search under our standing assumptions.

Lemma 3.1 *Let $g \in \Gamma_0(\mathscr{H})$ and let $\alpha > 0$. Then, for every $x \in \operatorname{dom} g$, we have*

$$\operatorname{prox}_{\alpha g}(x) \to x \quad as \quad \alpha \downarrow 0. \tag{11}$$

Under our standing assumptions, we define the *proximal forward–backward operator*

$$J : \left[\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g\right] \times \mathbb{R}_{++} \to \operatorname{dom} g$$

by

$$J(x,\alpha) := \operatorname{prox}_{\alpha g}(x - \alpha \nabla f(x)) \quad \text{for all} \quad x \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g,\ \alpha > 0. \tag{12}$$

The following result is essentially from [12, Lemma 2.4]. Even if the standing assumptions are different, the proof is quite similar. A more general variant of this result could be found in [35, Lemma 3].

Lemma 3.2 *For any $x \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$, we have*

$$\frac{\alpha_2}{\alpha_1}\|x - J(x,\alpha_1)\| \geq \|x - J(x,\alpha_2)\| \geq \|x - J(x,\alpha_1)\| \quad \textit{for all} \quad \alpha_2 \geq \alpha_1 > 0. \tag{13}$$

Proof By using (9) and (12) with $z = x - \alpha \nabla f(x)$, we have

$$\frac{x - \alpha \nabla f(x) - J(x,\alpha)}{\alpha} \in \partial g(J(x,\alpha)) \tag{14}$$

for all $\alpha > 0$. For any $\alpha_2 \geq \alpha_1 > 0$, it follows from the monotonicity of ∂g and (14) that

$$\begin{aligned}
0 \leq & \left\langle \frac{x - \alpha_2 \nabla f(x) - J(x,\alpha_2)}{\alpha_2} - \frac{x - \alpha_1 \nabla f(x) - J(x,\alpha_1)}{\alpha_1}, J(x,\alpha_2) - J(x,\alpha_1) \right\rangle \\
= & \left\langle \frac{x - J(x,\alpha_2)}{\alpha_2} - \frac{x - J(x,\alpha_1)}{\alpha_1}, (x - J(x,\alpha_1)) - (x - J(x,\alpha_2)) \right\rangle \\
= & -\frac{\|x - J(x,\alpha_2)\|^2}{\alpha_2} - \frac{\|x - J(x,\alpha_1)\|^2}{\alpha_1} \\
& + \left(\frac{1}{\alpha_2} + \frac{1}{\alpha_1}\right) \langle x - J(x,\alpha_2), x - J(x,\alpha_1) \rangle \\
\leq & -\frac{\|x - J(x,\alpha_2)\|^2}{\alpha_2} - \frac{\|x - J(x,\alpha_1)\|^2}{\alpha_1} \\
& + \left(\frac{1}{\alpha_2} + \frac{1}{\alpha_1}\right) \|x - J(x,\alpha_2)\| \cdot \|x - J(x,\alpha_1)\|.
\end{aligned}$$

This gives us that

$$\left(\|x - J(x,\alpha_2)\| - \|x - J(x,\alpha_1)\|\right) \cdot \left(\|x - J(x,\alpha_2)\| - \frac{\alpha_2}{\alpha_1}\|x - J(x,\alpha_1)\|\right) \leq 0.$$

Since $\dfrac{\alpha_2}{\alpha_1} \geq 1$, we derive (13) and thus complete the proof of the lemma. □

Next, let us present Beck–Teboulle's backtracking line search [6], which is specifically useful for forward–backward methods when the Lipschitz constant of ∇f is not known or hard to estimate.

Linesearch BT (Beck–Teboulle's Line Search)
Given $x \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$, $\sigma > 0$ and $0 < \theta < 1$.
Input. Set $\alpha = \sigma$ and $J(x, \alpha) = \operatorname{prox}_{\alpha g}(x - \alpha \nabla f(x))$ with $x \in \operatorname{dom} g$.
While $f(J(x, \alpha)) > f(x) + \langle \nabla f(x), J(x, \alpha) - x \rangle + \dfrac{1}{2\alpha}\|x - J(x, \alpha)\|^2$, **do**
$\alpha = \theta\alpha$.
End While
Output. α.

The output α in this line search will be denoted by $LS(x, \sigma, \theta)$. Let us show the well-definedness and finite termination of this line search under the standing assumptions **A1–A3** below.

Proposition 3.1 (Finite Termination of Beck–Teboulle's Line Search) *Suppose that assumptions* **A1–A2** *hold. Then, for any $x \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$, we have*

(i) *The above line search terminates after finitely many iterations with the positive output $\bar{\alpha} = LS(x, \sigma, \theta)$.*
(ii) $\|x - u\|^2 - \|J(x, \bar{\alpha}) - u\|^2 \geq 2\bar{\alpha}[F(J(x, \bar{\alpha})) - F(u)]$ *for any $u \in \mathscr{H}$.*
(iii) $F(J(x, \bar{\alpha})) - F(x) \leq -\dfrac{1}{2\bar{\alpha}}\|J(x, \bar{\alpha}) - x\|^2 \leq 0$. *Consequently, by* **A3**, $J(x, \bar{\alpha}) \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$.

Proof Take any $x \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$. Let us justify (i) first. Note that $J(x, \alpha)$ is well-defined for any $\alpha > 0$ because $\nabla f(x)$ exists due to assumption **A2**. If $x \in S^*$, where S^* is the optimal solution set to problem (4), then $x = J(x, \sigma)$ due to (10) and (9). Thus, the line search stops with zero step and gives us the output σ and $x = J(x, \sigma) \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$. If $x \notin S^*$, suppose by contradiction that the line search does not terminate after finitely many steps. Hence, for all $\alpha \in \mathscr{P} := \{\sigma, \sigma\theta, \sigma\theta^2, \ldots\}$ it follows that

$$\langle \nabla f(x), J(x, \alpha) - x \rangle + \frac{1}{2\alpha}\|x - J(x, \alpha)\|^2 < f(J(x, \alpha)) - f(x). \tag{15}$$

Since $\operatorname{prox}_{\alpha g}$ is non-expansive, we have

$$\begin{aligned} \|J(x, \alpha) - x\| &\leq \|\operatorname{prox}_{\alpha g}(x - \alpha \nabla f(x)) - \operatorname{prox}_{\alpha g}(x)\| + \|\operatorname{prox}_{\alpha g}(x) - x\| \\ &\leq \alpha\|\nabla f(x)\| + \|\operatorname{prox}_{\alpha g}(x) - x\|. \end{aligned} \tag{16}$$

Lemma 3.1 tells us that

$$\|J(x, \alpha) - x\| \to 0 \quad \text{as} \quad \alpha \downarrow 0. \tag{17}$$

Since $x \in \operatorname{int}(\operatorname{dom} f)$, there exists $\ell \in \mathbb{N}$ such that $J(x, \alpha) \in \operatorname{int}(\operatorname{dom} f)$ for all

$$\alpha \in \mathscr{P}' := \{\sigma\theta^{\ell}, \sigma\theta^{\ell+1}, \ldots\} \subseteq \mathscr{P}.$$

Thanks to the convexity of f, we have

$$f(x) - f(J(x, \alpha)) \geq \langle \nabla f(J(x, \alpha)), x - J(x, \alpha) \rangle \quad \text{for} \quad \alpha \in \mathscr{P}'. \tag{18}$$

This inequality together with (15) implies

$$\begin{aligned}\frac{1}{2\alpha} \|J(x, \alpha) - x\|^2 &< \langle \nabla f\big(J(x, \alpha)\big) - \nabla f(x), J(x, \alpha) - x \rangle \\ &\leq \big\|\nabla f\big(J(x, \alpha)\big) - \nabla f(x)\big\| \cdot \|J(x, \alpha) - x\|,\end{aligned}$$

which yields $J(x, \alpha) \neq x$ and

$$0 < \frac{\|J(x, \alpha) - x\|}{\alpha} < 2 \big\|\nabla f\big(J(x, \alpha)\big) - \nabla f(x)\big\| \quad \text{for all} \quad \alpha \in \mathscr{P}'. \tag{19}$$

Since $\|x - J(x, \alpha)\| \to 0$ as $\alpha \to 0$ by (17), $J(x, \alpha) \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$ for sufficiently small $\alpha > 0$. Due to the continuity of ∇f at $x \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$ by assumption **A2**, we obtain from (19) that

$$\lim_{\alpha \to 0, \alpha \in \mathscr{P}'} \frac{\|x - J(x, \alpha)\|}{\alpha} = 0. \tag{20}$$

Applying (9) with $z = x - \alpha \nabla f(x)$ gives us that $\dfrac{x - J(x, \alpha)}{\alpha} - \nabla f(x) \in \partial g(J(x, \alpha))$. Taking $\alpha \to 0$, we have $-\nabla f(x) \in \partial g(x)$ due to the demiclosedness of the subdifferentials; see, e.g., [38, Theorem 4.7.1 and Proposition 4.2.1(i)]. It follows that $0 \in \nabla f(x) + \partial g(x)$, which contradicts the hypothesis that $x \notin S^*$ by (10). Hence, the line search terminates after finitely many steps with the output $\bar{\alpha}$.

To proceed the proof of (ii), note that

$$f(J(x, \bar{\alpha})) \leq f(x) + \langle \nabla f(x), J(x, \bar{\alpha}) - x \rangle + \frac{1}{2\bar{\alpha}} \|x - J(x, \bar{\alpha})\|^2. \tag{21}$$

Moreover, by (9) , we have

$$\frac{x - J(x, \bar{\alpha})}{\bar{\alpha}} - \nabla f(x) \in \partial g(J(x, \bar{\alpha})).$$

Pick any $u \in \mathscr{H}$, we get from the latter that

$$g(u) - g(J(x, \bar{\alpha})) \geq \left\langle \frac{x - J(x, \bar{\alpha})}{\bar{\alpha}} - \nabla f(x), u - J(x, \bar{\alpha}) \right\rangle. \tag{22}$$

Note also that $f(u) - f(x) \geq \langle \nabla f(x), u - x \rangle$.This together with (22) and (21) gives us that

$$\begin{aligned} F(u) = (f+g)(u) &\geq f(x) + g(J(x,\bar{\alpha})) \\ &+ \left\langle \frac{x - J(x,\bar{\alpha})}{\bar{\alpha}} - \nabla f(x), u - J(x,\bar{\alpha}) \right\rangle + \langle \nabla f(x), u - x \rangle \\ &= f(x) + g(J(x,\bar{\alpha})) + \frac{1}{\bar{\alpha}} \langle x - J(x,\bar{\alpha}), u - J(x,\bar{\alpha}) \rangle + \langle \nabla f(x), J(x,\bar{\alpha}) - x \rangle \\ &\geq f(J(x,\bar{\alpha})) + g(J(x,\bar{\alpha})) + \frac{1}{\bar{\alpha}} \langle x - J(x,\bar{\alpha}), u - J(x,\bar{\alpha}) \rangle - \frac{1}{2\bar{\alpha}} \|J(x,\bar{\alpha}) - x\|^2. \end{aligned}$$

It follows that

$$\langle x - J(x,\bar{\alpha}), J(x,\bar{\alpha}) - u \rangle \geq \bar{\alpha}[F(J(x,\bar{\alpha})) - F(u)] - \frac{1}{2} \|J(x,\bar{\alpha}) - x\|^2.$$

Since $2\langle x - J(x,\bar{\alpha}), J(x,\bar{\alpha}) - u \rangle = \|x - u\|^2 - \|J(x,\bar{\alpha}) - x\|^2 - \|J(x,\bar{\alpha}) - u\|^2$, the latter implies that

$$\|x - u\|^2 - \|J(x,\bar{\alpha}) - u\|^2 \geq 2\bar{\alpha}[F(J(x,\bar{\alpha})) - F(u)],$$

which clearly ensures (ii).

Finally, (iii) is a direct consequence of (ii) with $u = x$. It follows that $J(x,\bar{\alpha})$ belongs to the sublevel set $\{F \leq F(x)\}$. By assumption **A3**, $J(x,\bar{\alpha}) \in \mathrm{int}(\mathrm{dom}\, f)$. The proof is complete. □

Now, we recall the forward–backward splitting method with line search proposed by [6] as following.

Forward–Backward Splitting Method with Backtracking Line Search (FBS method)

Step 0. Take $x^0 \in \mathrm{int}(\mathrm{dom}\, f) \cap \mathrm{dom}\, g$, $\sigma > 0$ and $0 < \theta < 1$.
Step k. Set

$$x^{k+1} := J(x^k, \alpha_k) = \mathrm{prox}_{\alpha_k g}(x^k - \alpha_k \nabla f(x^k)) \tag{23}$$

with $\alpha_{-1} := \sigma$ and

$$\alpha_k := LS(x^k, \alpha_{k-1}, \theta). \tag{24}$$

The following result which is a direct consequence of Proposition 3.1 plays the central role in our study.

Corollary 3.1 (Well-definedness of the FBS method) *Let $x^0 \in \mathrm{int}(\mathrm{dom}\, f) \cap \mathrm{dom}\, g$. The sequence $(x^k)_{k\in\mathbb{N}}$ from the FBS method is well-defined, $x^k \in \mathrm{int}\,(\mathrm{dom}\, f) \cap g$, and f is differentiable at any x^k for all $k \in \mathbb{N}$. Moreover, for any $x \in \mathscr{H}$, we have*

(i) $\|x^k - x\|^2 - \|x^{k+1} - x\|^2 \geq 2\alpha_k \left[F(x^{k+1}) - F(x)\right]$.

(i) $F(x^{k+1}) - F(x^k) \le -\frac{1}{2\alpha_k}\|x^{k+1} - x^k\|^2$.

Proof Thanks to Proposition 3.1(iii), $x^k \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$ and f is differentiable at any x^k by assumption **A2**. This verifies the well-definedness of $(x^k)_{k\in\mathbb{N}}$. Moreover, both (i) and (ii) are consequence of (ii) and (iii) from Proposition 3.1 by replacing $u = x$, $x = x^k$, $\bar{\alpha} = \alpha_k$, and $J(x, \bar{\alpha}) = J(x^k, \alpha_k) = x^{k+1}$. □

The following result recovers the global convergence of the FBS method without supposing the Lipschitz continuity on the gradient ∇f in [13, Theorem 3.18] and [12, Theorem 4.2] with our simplified assumptions (**A1** − **A3**). The proof is similar to that of [12, Theorem 4.2] with some modifications and ideas from [13].

Theorem 3.1 (Global Convergence of the FBS Method) *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated from the FBS method. The following statements hold:*

(i) *If $S^* \ne \emptyset$, then $(x^k)_{k\in\mathbb{N}}$ weakly converges to a point in S^*. Moreover,*

$$\lim_{k\to\infty} F(x^k) = \min_{x\in\mathcal{H}} F(x). \tag{25}$$

(ii) *If $S^* = \emptyset$, then we have*

$$\lim_{k\to\infty} \|x^k\| = +\infty \quad \textit{and} \quad \lim_{k\to\infty} F(x^k) = \inf_{x\in\mathcal{H}} F(x).$$

Proof Let us justify (i) by supposing that $S^* \ne \emptyset$. By Corollary 3.1(i), for any $x^* \in S^*$ we have

$$\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 \ge 2\alpha_k[F(x^{k+1}) - F(x^*)] \ge 0. \tag{26}$$

It follows that the sequence $(x^k)_{k\in\mathbb{N}}$ is Fejér monotone with respect to S^*. Thus, it is bounded according to [4, Proposition 5.4]. Define $M := \sup\{\|x^k - x^*\| : k \in \mathbb{N}\} < +\infty$ for some $x^* \in S^*$, we get from (26) that

$$\begin{aligned}0 &\le 2\alpha_k[F(x^{k+1}) - F(x^*)] \le \|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2\\ &= (\|x^k - x^*\| + \|x^{k+1} - x^*\|) \cdot (\|x^k - x^*\| - \|x^{k+1} - x^*\|)\\ &\le 2M\|x^k - x^{k+1}\|.\end{aligned}$$

This yields that

$$0 \le F(x^{k+1}) - F(x^*) \le M\,\frac{\|x^k - x^{k+1}\|}{\alpha_k}. \tag{27}$$

Note further from Corollary 3.1(ii) that

$$2\sigma[F(x^k) - F(x^{k+1})] \ge 2\alpha_k[F(x^k) - F(x^{k+1})] \ge \|x^k - x^{k+1}\|^2.$$

As $(F(x^k))_{k\in\mathbb{N}}$ is decreasing and bounded below by $F(x^*)$, the latter implies

$$\infty > 2\sigma[F(x^0) - F(x^*)] \geq \sum_{k=0}^{\infty} 2\sigma[F(x^k) - F(x^{k+1})] \geq \sum_{k=0}^{\infty} \|x^k - x^{k+1}\|^2. \quad (28)$$

Thus, $\|x^k - x^{k+1}\| \to 0$ as $k \to \infty$. Let $\bar{x}$ be a weak accumulation point of $(x^k)_{k\in\mathbb{N}}$. Let us find a subsequence $(x^{n_k})_{k\in\mathbb{N}}$ weakly converging to $\bar{x}$. As F is l.s.c. and convex, it is also weakly l.s.c.. Moreover, since $(F(x^k))_{k\in\mathbb{N}}$ is decreasing, we have $F(x^0) \geq F(\bar{x})$. According to assumptions **A2** and **A3**, f is differentiable at $\bar{x} \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$. As $(\alpha_k)_{k\in\mathbb{N}}$ is decreasing by the construction of the line search in the FBS method, $\alpha := \lim_{k\to\infty} \alpha_k = \lim_{k\to\infty} \alpha_{n_k}$ exists.

Case 1: $\alpha > 0$. Note from (9) and (23 that $\dfrac{x^{n_k} - \alpha_{n_k}\nabla f(x^{n_k}) - x^{n_k+1}}{\alpha_{n_k}} \in \partial g(x^{n_k+1})$, which implies

$$\frac{x^{n_k} - x^{n_k+1}}{\alpha_{n_k}} + \nabla f(x^{n_k+1}) - \nabla f(x^{n_k}) \in \nabla f(x^{n_k+1}) + \partial g(x^{n_k+1}). \quad (29)$$

Applying Corollary 3.1(ii), we have $\dfrac{\|x^{n_k} - x^{n_k+1}\|^2}{\alpha_{n_k}} \to 0$ as the sequence $(F(x^k))_{k\in\mathbb{N}}$ is decreasing and thus converging. Since $\alpha_{n_k} \geq \alpha > 0$, $\dfrac{\|x^{n_k} - x^{n_k+1}\|}{\alpha_{n_k}}$ also converges to 0. Moreover, note from (28) that $\|x^{n_k} - x^{n_k+1}\| \to 0$, $\|\nabla f(x^{n_k}) - \nabla f(x^{n_k+1})\| \to 0$ due to the uniform continuity of ∇f on the weakly compact set $\left\{\bar{x}, (x^{n_k})_{k\in\mathbb{N}}, (x^{n_k+1})_{k\in\mathbb{N}}\right\}$ of $[F \leq F(x^0)]$ as in assumption **A2**.

Taking $k \to \infty$ in (29) gives us that $0 \in \nabla f(\bar{x}) + \partial g(\bar{x})$ due to the demiclosedness of the subdifferentials, which implies $\bar{x} \in S^*$. Furthermore, since $(F(x^k))_{k\in\mathbb{N}}$ is decreasing, (25) is a consequence of (27).

Case 2: $\alpha = 0$. Define $\hat{\alpha}_{n_k} := \dfrac{\alpha_{n_k}}{\theta} > \alpha_{n_k} > 0$ and $\hat{x}^{n_k} := J\left(x^{n_k}, \hat{\alpha}_{n_k}\right) \in \operatorname{dom} g$. It follows from the definition of **Linesearch BT** that

$$f(\hat{x}^{n_k}) > f(x^{n_k}) + \langle \nabla f(x^{n_k}), \hat{x}^{n_k} - x^{n_k}\rangle + \frac{1}{2\hat{\alpha}_{n_k}}\|x^{n_k} - \hat{x}^{n_k}\|^2. \quad (30)$$

Due to Lemma 3.2, we have

$$\begin{aligned} \|x^{n_k} - \hat{x}^{n_k}\| = \|x^{n_k} - J(x^{n_k}, \hat{\alpha}_{n_k})\| &\leq \frac{\hat{\alpha}_{n_k}}{\alpha_{n_k}}\|x^{n_k} - J(x^{n_k}, \alpha_{n_k})\| \\ &= \frac{1}{\theta}\|x^{n_k} - x^{n_k+1}\| \to 0, \end{aligned}$$

which tells us that $(\hat{x}^{n_k})_{k\in\mathbb{N}}$ weakly converges to $\bar{x} \in \operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$. Define $\mathscr{A} := \{\bar{x}, (x^{n_k})_{k\in\mathbb{N}}\}$. It follows that $\mathscr{A} \subseteq [F \leq F(x^0)]$, which is weakly compact.

By assumption **A3**, $d(\mathscr{A}, \mathscr{H} \setminus \text{int}(\text{dom } f)) > 0$. As $\|x^{n_k} - \hat{x}^{n_k}\| \to 0$, we find some $K > 0$ such that $\hat{x}^{n_k} \in \text{int}(\text{dom } f)$ and thus $\hat{x}^{n_k} \in \text{int}(\text{dom } f) \cap \text{dom } g$ for any $k > K$. Hence, $\nabla f(\hat{x}^{n_k})$ is well-defined by assumption **A2**. It follows that

$$f(x^{n_k}) \geq f(\hat{x}^{n_k}) + \langle \nabla f(\hat{x}^{n_k}), x^{n_k} - \hat{x}^{n_k} \rangle.$$

This together with (30) gives us that

$$\begin{aligned} \frac{1}{2\hat{\alpha}_{n_k}} \|x^{n_k} - \hat{x}^{n_k}\|^2 &< \langle \nabla f(x^{n_k}) - \nabla f(\hat{x}^{n_k}), x^{n_k} - \hat{x}^{n_k} \rangle \\ &\leq \|\nabla f(x^{n_k}) - \nabla f(\hat{x}^{n_k})\| \cdot \|x^{n_k} - \hat{x}^{n_k}\|. \end{aligned}$$

Hence, we have

$$\|\nabla f(x^{n_k}) - \nabla f(\hat{x}^{n_k})\| \geq \frac{1}{2\hat{\alpha}_{n_k}} \|x^{n_k} - \hat{x}^{n_k}\|. \tag{31}$$

As $\left\{\bar{x}, (x^{n_k})_{k\in\mathbb{N}}, (\hat{x}^{n_k})_{k\in\mathbb{N}}\right\}$ is a weakly compact subset of $\text{int}(\text{dom } f) \cap \text{dom } g$ and $\|\hat{x}^{n_k} - x^{n_k}\| \to 0$, assumption **A2** tells us that $\|\nabla f(x^{n_k}) - \nabla f(\hat{x}^{n_k})\| \to 0$ as $k \to \infty$. We derive from (31) that

$$\lim_{k\to\infty} \frac{1}{\hat{\alpha}_{n_k}} \|x^{n_k} - \hat{x}^{n_k}\| = 0. \tag{32}$$

Applying (9) with $z = x^{n_k} - \hat{\alpha}_{n_k} \nabla f(x^{n_k})$ gives us that

$$\frac{x^{n_k} - \hat{x}^{n_k}}{\hat{\alpha}_{n_k}} = \nabla f(\hat{x}^{n_k}) + \frac{x^{n_k} - \hat{\alpha}_{n_k} \nabla f(x^{n_k}) - \hat{x}^{n_k}}{\hat{\alpha}_{n_k}} \in \nabla f(\hat{x}^{n_k}) + \partial g(\hat{x}^{n_k}).$$

By letting $k \to \infty$, it is similar to the conclusion after (29) that $0 \in \nabla f(\bar{x}) + \partial g(\bar{x})$ due to the demiclosedness of the subdifferentials, i.e., $\bar{x} \in S^*$. To proceed the proof of (25) in this case, we observe from Lemma 3.2 that

$$\|x^{n_k} - \hat{x}^{n_k}\| = \left\|x^{n_k} - J(x^{n_k}, \frac{\alpha_{n_k}}{\theta})\right\| \geq \|x^{n_k} - J(x^{n_k}, \alpha_{n_k})\| = \|x^{n_k} - x^{n_k+1}\|.$$

This together with (32) yield $\dfrac{\|x^{n_k} - x^{n_k+1}\|}{\alpha_{n_k}} \to 0$ as $k \to \infty$. Since $(F(x^k))_{k\in\mathbb{N}}$ is decreasing, we derive from the latter and (27) that

$$0 = \lim_{k\to\infty} M \frac{\|x^{n_k} - x^{n_k+1}\|}{\alpha_{n_k}} \geq \lim_{k\to\infty} F(x^{n_k+1}) - F(x^*) = \lim_{k\to\infty} F(x^k) - F(x^*) \geq 0,$$

which ensures (25).

From the above two cases, we have (25) and the fact that any weak accumulation point of the Fejér sequence $(x^k)_{k\in\mathbb{N}}$ belongs to S^*. The classical Fejér theorem [4, Theorem 5.5] tells us that $(x^k)_{k\in\mathbb{N}}$ weakly converges to some point of S^*. This verifies (i) of the theorem.

The proof for the second part (ii) follows the same lines in the proof of [12, Theorem 4.2]. □

The following proposition shows that when f is locally Lipschitz continuous, the step size α_k is bounded below by a positive number. The second part of this result coincides with [6, Remark 1.2].

Proposition 3.2 (Boundedness from Below for the Step Sizes) *Let $(x^k)_{k\in\mathbb{N}}$ and $(\alpha_k)_{k\in\mathbb{N}}$ be the two sequences generated from the FBS method. Suppose that $S^* \neq \emptyset$ and that the sequence $(x^k)_{k\in\mathbb{N}}$ is converging to some $x^* \in S^*$. If ∇f is locally Lipschitz continuous around x^* with modulus L, then there exists some $K \in \mathbb{N}$ such that*

$$\alpha_k \geq \min\left\{\alpha_K, \frac{\theta}{L}\right\} > 0 \quad \text{for all} \quad k > K. \tag{33}$$

Consequently, for $k > K$ the line search $LS(x^k, \alpha_{k-1}, \theta)$ needs at most $\log_\theta \left(\min\left\{1, \frac{\theta}{\alpha_K L}\right\}\right)$ steps.

Furthermore, if ∇f is globally Lipschitz continuous on $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$ with uniform modulus L then $\alpha_k \geq \min\{\sigma, \frac{\theta}{L}\}$ for all $k \in \mathbb{N}$. In this case, line search $LS(x^k, \alpha_{k-1}, \theta)$ needs at most $\log_\theta \left(\min\{1, \frac{\theta}{\sigma L}\}\right)$ steps for any k.

Proof To justify, suppose that $S^* \neq \emptyset$, the sequence $(x^k)_{k\in\mathbb{N}}$ is converging to $x^* \in S^*$, and that ∇f is locally Lipschitz continuous around x^* with constant $L > 0$. We find some $\varepsilon > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \text{for all} \quad x, y \in \mathbb{B}_\varepsilon(x^*), \tag{34}$$

where $\mathbb{B}_\varepsilon(x^*)$ is the closed ball in $\mathscr{H}$ with center x^* and radius ε. Since $(x^k)_{k\in\mathbb{N}}$ is converging to x^*, there exists some $K \in \mathbb{N}$ such that

$$\|x^k - x^*\| \leq \frac{\theta\varepsilon}{2+\theta} < \varepsilon \quad \text{for all} \quad k > K \tag{35}$$

with $0 < \theta < 1$ defined in **Linesearch BT**. We claim that

$$\alpha_k \geq \min\left\{\alpha_{k-1}, \frac{\theta}{L}\right\} \quad \text{for all} \quad k > K. \tag{36}$$

Suppose by contradiction that $\alpha_k < \min\{\alpha_{k-1}, \frac{\theta}{L}\}$. Then, $\alpha_k < \alpha_{k-1}$, and so, the loop in **Linesearch BT** at (x^k, α_{k-1}) needs more than one iteration. Define $\hat{\alpha}_k := \frac{\alpha_k}{\theta} > 0$ and $\hat{x}^k := J(x^k, \hat{\alpha}_k)$, we have

$$f(\hat{x}^k) > f(x^k) + \langle \nabla f(x^k), \hat{x}^k - x^k\rangle + \frac{1}{2\hat{\alpha}_k}\|x^k - \hat{x}^k\|^2. \tag{37}$$

Furthermore, it follows from Lemma 3.2 that

$$\|x^k - \hat{x}^k\| = \|x^k - J(x^k, \hat{\alpha}_k)\| \le \frac{\hat{\alpha}_k}{\alpha_k}\|x^k - J(x^k, \alpha_k)\| = \frac{1}{\theta}\|x^k - x^{k+1}\|.$$

This together with (35) yields

$$\begin{aligned}\|\hat{x}^k - x^*\| &\le \|\hat{x}^k - x^k\| + \|x^k - x^*\| \\ &\le \frac{1}{\theta}\|x^k - x^{k+1}\| + \|x^k - x^*\| \\ &\le \frac{1}{\theta}\cdot\frac{2\theta\varepsilon}{2+\theta} + \frac{\theta\varepsilon}{2+\theta} = \varepsilon. \end{aligned} \tag{38}$$

Since $x^k, \hat{x}^k \in \mathbb{B}_\varepsilon(x^*)$ by (35) and (38), we get from (34) that

$$\begin{aligned} &f(\hat{x}^k) - f(x^k) - \langle \nabla f(x^k), \hat{x}^k - x^k\rangle \\ &= \int_0^1 \langle \nabla f(x^k + t(\hat{x}^k - x^k)) - \nabla f(x^k), \hat{x}^k - x^k\rangle dt \\ &\le \int_0^1 tL\|\hat{x}^k - x^k\|^2 dt = \frac{L}{2}\|\hat{x}^k - x^k\|^2. \end{aligned}$$

Combining this with (37) yields $\hat{\alpha}_k \ge \frac{1}{L}$ and thus $\alpha_k \ge \frac{\theta}{L}$. This is a contradiction.

If there is some $H > K$ with $H \in \mathbb{N}$ such that $\alpha_H > \frac{\theta}{L}$, we get from (36) that $\alpha_k \ge \frac{\theta}{L}$ for all $k \ge H$. Otherwise, $\alpha_k < \frac{\theta}{L}$ for any $k > K$, which implies that $\alpha_k = \alpha_{k-1} = \alpha_K$ for all $k > K$ due to (36) and the decreasing property of $(\alpha_k)_{k\in\mathbb{N}}$. In both cases, we have (33).

Now suppose that the line search $LS(x^k, \alpha_{k-1}, \theta)$ needs d repetitions. It follows from (33) that

$$\alpha_K\theta^d \ge \alpha_k = \alpha_{k-1}\theta^d \ge \min\left\{\alpha_K, \frac{\theta}{L}\right\},$$

which tells us that $d \le \log_\theta\left(\min\left\{1, \frac{\theta}{\alpha_K L}\right\}\right)$.

Finally, suppose that ∇f is globally Lipschitz continuous with modulus L on $\operatorname{int}(\operatorname{dom} f)\cap\operatorname{dom} g$ subset of $\operatorname{int}(\operatorname{dom} f)$. By using Proposition 3.1(iii), we can repeat the above proof without concerning ε, K and replace (36) by $\alpha_k \ge \min\left\{\sigma, \frac{\theta}{L}\right\}$. □

Next, we present the conventional complexity $o(k^{-1})$ of the FBS method [10, Theorem 3] but under our standing assumptions and that ∇f is *locally* Lipschitz continuous. The complexity remains valid when replacing the local Lipschitz condition there by the weaker one that the step size $\{\alpha^k\}$ is bounded below by a positive number; see Proposition 3.2. This idea was initiated in [12, Theorem 4.3] in finite dimensions and extended to different kinds of line searches in Hilbert spaces in [13], e.g., [13, Corollary 3.20]. For the completeness, we provide a short proof [2], which could be obtained by following the method of proof as in [13, Theorem 3.18(iii)(d)] as well.

Theorem 3.2 (Sublinear Convergence of the FBS Method) *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated in the FBS method. Suppose that $S^* \neq \emptyset$ and that ∇f is locally Lipschitz continuous around any point in S^*. Then,*

$$\lim_{k\to\infty} k\big[F(x^k) - \min_{x\in\mathscr{H}} F(x)\big] = 0. \tag{39}$$

Proof By Proposition 3.2, α_k is bounded below by some $\alpha > 0$. Take any $x^* \in S^*$, we obtain from Corollary 3.1(i) that

$$\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 \geq 2\alpha[F(x^{k+1}) - F(x^*)] \geq 0. \tag{40}$$

As $\|x^k - x^*\|^2$ is decreasing, $(\|x^k - x^*\|^2)_{k\in\mathbb{N}}$ is converging. Given $\varepsilon > 0$, there exists $K \in \mathbb{N}$ (large enough) such that $|\|x^{K+k} - x^*\|^2 - \|x^K - x^*\|^2| \leq \varepsilon$, for any $k \in \mathbb{N}$. It follows from (40) that

$$\begin{aligned}\varepsilon \geq \|x^K - x^*\|^2 - \|x^{k+K} - x^*\|^2 &= \sum_{\ell=K}^{k+K-1} (\|x^\ell - x^*\|^2 - \|x^{\ell+1} - x^*\|^2)\\ &\geq 2\alpha \sum_{\ell=K}^{k+K-1} [F(x^{\ell+1}) - F(x^*)].\end{aligned}$$

Since $(F(x^k))$ is decreasing, we get that

$$\varepsilon \geq 2\alpha k(F(x^{K+k}) - F(x^*)) = 2\alpha \frac{k}{K+k}(K+k)(F(x^{K+k}) - F(x^*)),$$

which implies that

$$\begin{aligned}0 \leq \liminf_{k\to\infty} k(F(x^k) - F(x^*)) &\leq \limsup_{k\to\infty} k(F(x^k) - F(x^*))\\ &= \limsup_{k\to\infty} (K+k)(F(x^{K+k}) - F(x^*)) \leq \frac{\varepsilon}{2\alpha}.\end{aligned}$$

This verifies (39) and completes the proof of the theorem. □

4 Local Linear Convergence of Forward–Backward Splitting Methods

In this section, we obtain the local Q-linear convergence for the FBS method under the quadratic growth condition and local Lipschitz continuity of ∇f, which is automatic in many problems including Lasso problem and Poisson linear inverse regularized

[2] We are grateful to one of the referees whose remarks lead us to this simple proof.

problem. R-linear convergence of the FBS method has been established under some different assumptions such as Kurdya-Łojasiewicz inequality with order $\frac{1}{2}$; see, e.g., [18]. Q-linear convergence of the iterative sequence $F(x^k)$ are also obtained in [18–21]. Our results are close to [19, Theorem 4.5(iii)] and [25, Theorem 3.2 and Corollary 3.7]. However, we focus on the local linear convergence; the proof also suggests a direct way to obtain linear convergence of the FBS method from the quadratic growth condition without going through the error bound or Kurdya-Łojasiewicz inequality. Convexity does not appear explicitly in the proof, but it is hidden in the properties of the FBS method in Sect. 3. This could be important for further study in linear convergence of the FBS method for solving nonconvex optimization problems under the quadratic growth condition. In order to concentrate on the idea and avoid the complication between weak convergence and strong convergence of the sequence $(x^k)_{k\in\mathbb{N}}$ as in Theorem 3.1, we suppose throughout this section that

$$\mathscr{H} \text{ is a finite-dimensional space.}$$

As explained in Remark 3.1, in this case, the standing assumptions **A1–A3** can be simplified as assumptions **A1–A2** and assumption **A3** is superfluous. The traditional FBS method with global Lipschitz property on ∇f also attains linear convergence in infinite dimensions [19, Theorem 4.5(iii)] under the so-called *2-Łojasiewicz property* on a nonempty set $\Omega \subset \mathscr{H}$ and the assumption that the whole sequence $(x^k)_{k\in\mathbb{N}}$ belong to Ω; see [19, Section 4.2] for some choices of Ω. Remark 4.2 below also notes that our results could be extended to infinite-dimensional Hilbert spaces.

Our first result about the R-linear convergence of the FBS method is not quite new [18–20,23,25], but it does not require the global Lipschitz assumption on ∇f and will be improved later in our main Theorem 4.1. Moreover, the R-rates here are sharper (or smaller) than those obtained in [19,25] under the quadratic growth conditions; see Remark 4.1 below for further details.

Proposition 4.1 (R-linear Convergence under Quadratic Growth Condition) *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated from the FBS method. Suppose that S^* is not empty and let $x^* \in S^*$ be the convergent point of $(x^k)_{k\in\mathbb{N}}$ as in Theorem 3.1. Suppose further that ∇f is locally Lipschitz continuous around x^* with constant $L > 0$. If F satisfies the quadratic growth condition at x^* with modulus $\kappa > 0$, then there exists some $K \in \mathbb{N}$ such that*

$$d(x^{k+1}; S^*) \le \frac{1}{\sqrt{1+\alpha\kappa}} d(x^k; S^*) \quad \text{for all} \quad k > K, \tag{41}$$

where $\alpha := \min\left\{\alpha_K, \frac{\theta}{L}\right\}$. Consequently, we have

$$F(x^{k+1}) - \min_{x\in\mathscr{H}} F(x) = \mathscr{O}((1+\alpha\kappa)^{-k}), \tag{42}$$

$$\|x^{k+1} - x^*\| = \mathscr{O}((1+\alpha\kappa)^{-\frac{k}{2}}). \tag{43}$$

If, in addition, ∇f is globally Lipschitz continuous on $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$ with constant L, α could be chosen as $\min\left\{\sigma, \frac{\theta}{L}\right\}$, which is independent of K.

Proof Since F satisfies the quadratic growth condition at x^* with modulus $\kappa > 0$, there exists $\varepsilon > 0$:

$$F(x) - F(x^*) \geq \frac{\kappa}{2} d^2(x; S^*) \quad \text{for all} \quad x \in \mathbb{B}_\varepsilon(x^*). \tag{44}$$

Since $(x^k)_{k\in\mathbb{N}}$ converges to x^* and ∇f is locally Lipschitz continuous around x^*, we find from Proposition 3.2 some constant $K \in \mathbb{N}$ such that $\alpha_k \geq \alpha$ and $x^k \in \mathbb{B}_\varepsilon(x^*)$ for any $k > K$. Denote the projection from x onto the set S^* by $\Pi_{S^*}(x)$. Combining (44) with Corollary 3.1(i) implies that

$$\begin{aligned} d^2(x^k; S^*) - d^2(x^{k+1}; S^*) &\geq \|x^k - \Pi_{S^*}(x^k)\|^2 - \|x^{k+1} - \Pi_{S^*}(x^k)\|^2 \\ &\geq 2\alpha_k[F(x^{k+1}) - F(\Pi_{S^*}(x^k))] \\ &\geq 2\alpha[F(x^{k+1}) - F(x^*)] \geq \alpha\kappa d^2(x^{k+1}; S^*) \end{aligned} \tag{45}$$

for all $k > K$. This clearly verifies (41).

To justify (42), note from (41) that $d(x^k; S^*) = \mathscr{O}((1+\alpha\kappa)^{-\frac{k}{2}})$. This together with (45) allows us to find some $M > 0$ such that

$$0 \leq F(x^{k+1}) - F(x^*) \leq \frac{1}{2\alpha} d^2(x^k; S^*) \leq M(1+\alpha\kappa)^{-k} \quad \text{for all} \quad k \in \mathbb{N},$$

which clearly ensures (42). To verify (43), we derive from Corollary 3.1(ii) that

$$\begin{aligned} \|x^k - x^{k+1}\| &\leq \sqrt{2\alpha_k[F(x^k) - F(x^{k+1})]} \leq \sqrt{2\sigma[F(x^k) - F(x^*)]} \\ &\leq \sqrt{2\sigma M}(1+\alpha\kappa)^{-\frac{k-1}{2}}. \end{aligned}$$

Since $(x^k)_{k\in\mathbb{N}}$ converges to x^*, it follows from the latter inequality that

$$\begin{aligned} \|x^{k+1} - x^*\| &= \sum_{j=k+1}^{\infty} \left(\|x^j - x^*\| - \|x^{j+1} - x^*\|\right) \leq \sum_{j=k+1}^{\infty} \|x^j - x^{j+1}\| \\ &\leq \sqrt{2\sigma M}(1+\alpha\kappa)^{-\frac{k}{2}} \sum_{j=0}^{\infty} (1+\alpha\kappa)^{\frac{-j}{2}} \\ &= \sqrt{2\sigma M}(1+\alpha\kappa)^{-\frac{k}{2}}[1 - (1+\alpha\kappa)^{-\frac{1}{2}}]^{-1}, \end{aligned}$$

which verifies (43). To complete, we repeat the above proof with the note from Proposition 3.2 that $\alpha_k \geq \min\{\sigma, \frac{\theta}{L}\}$ when ∇f is globally Lipschitz continuous on $\text{int}(\text{dom } f) \cap \text{dom } g$ with constant L. □

For the special case $g(x) = \delta_X(x)$, the indicator function to a closed convex set $X \subset \mathbb{R}^n$, the obtained linear convergence of $(d(x^k; S^*))_{k\in\mathbb{N}}$ in (41) is close to the [26, Theorem 12].

Next, we present the promised Q-linear convergence of the FBS method for both the objective value sequence $(F(x^k))_{k\in\mathbb{N}}$ and the iterative sequence $(x^k)_{k\in\mathbb{N}}$. We also point out that Q-linear convergence on the cost sequence $(F(x^k))_{k\in\mathbb{N}}$ has been discovered in the recent papers [18–21] under the standard assumption on global Lipschitz continuity of the gradient ∇f. In our result, ∇f is locally Lipschitz continuous; moreover, the Q-linear convergence of $(x^k)_{k\in\mathbb{N}}$ obtained here is new.

Theorem 4.1 (Q-linear Convergence under Quadratic Growth Condition) *Let $(x^k)_{k\in\mathbb{N}}$ be the sequences generated from the FBS method. Suppose that $S^* \neq \emptyset$ and let $x^* \in S^*$ be the convergent point of $(x^k)_{k\in\mathbb{N}}$ as in Theorem 3.1. Suppose further that ∇f is locally Lipschitz continuous around x^* with constant $L > 0$. If F satisfies the quadratic growth condition at x^* with modulus $\kappa > 0$, there exists $K \in \mathbb{N}$ such that*

$$\|x^{k+1} - x^*\| \le \frac{1}{\sqrt{1+\frac{\alpha\kappa}{4}}}\|x^k - x^*\| \tag{46}$$

$$|F(x^{k+1}) - F(x^*)| \le \frac{\sqrt{1+\alpha\kappa}+1}{2\sqrt{1+\alpha\kappa}}|F(x^k) - F(x^*)| \tag{47}$$

for any $k > K$, where $\alpha := \min\left\{\alpha_K, \frac{\theta}{L}\right\}$.

If, in addition, ∇f is globally Lipschitz continuous on $\mathrm{int}(\mathrm{dom}\, f) \cap \mathrm{dom}\, g$ with constant $L > 0$, α could be chosen as $\min\left\{\sigma, \frac{\theta}{L}\right\}$.

Proof Since F satisfies the quadratic growth condition at x^* with modulus $\kappa > 0$, we also have (44). This together with Corollary 3.1(i) gives us that

$$\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 \ge 2\alpha[F(x^{k+1}) - F(x^*)] \ge \alpha\kappa d^2(x^{k+1}; S^*) \text{ for all } x \in S^*, \tag{48}$$

when $k > K$ for some large $K \in \mathbb{N}$. Moreover, for any $r > k > K$, we get that

$$\|x^r - \Pi_{S^*}(x^{k+1})\| \le \|x^{k+1} - \Pi_{S^*}(x^{k+1})\| = d(x^{k+1}; S^*).$$

Taking $r \to \infty$ gives us that $\|x^* - \Pi_{S^*}(x^{k+1})\| \le d(x^{k+1}; S^*)$. It follows that

$$\|x^{k+1} - x^*\| \le \|x^{k+1} - \Pi_{S^*}(x^{k+1})\| + \|\Pi_{S^*}(x^{k+1}) - x^*\| \le 2d(x^{k+1}; S^*).$$

This together with (48) implies that

$$\|x^k - x^*\|^2 \ge \|x^{k+1} - x^*\|^2 + \frac{\alpha\kappa}{4}\|x^{k+1} - x^*\|^2 = \left(1 + \frac{\alpha\kappa}{4}\right)\|x^{k+1} - x^*\|^2,$$

which clearly verifies (46).

To see the second conclusion, we note from (41) that

$$\begin{aligned}\|x^{k+1} - x^k\| &\ge \|x^k - \Pi_{S^*}(x^{k+1})\| - \|x^{k+1} - \Pi_{S^*}(x^{k+1})\| \\ &\ge d(x^k; S^*) - d(x^{k+1}; S^*) \\ &\ge \beta\left(d(x^{k+1}; S^*) + d(x^k; S^*)\right)\end{aligned} \tag{49}$$

with $\beta := \frac{\sqrt{1+\alpha\kappa}-1}{\sqrt{1+\alpha\kappa}+1}$ for $k > K$ sufficiently large. We derive from this, Corollary 3.1(ii), and (49) that

$$\begin{aligned} F(x^k) - F(x^{k+1}) &\geq \frac{1}{2\alpha_k}\|x^{k+1} - x^k\|^2 \geq \frac{1}{2\alpha_k}\left(d(x^k; S^*) - d(x^{k+1}; S^*)\right)^2 \\ &\geq \frac{\beta}{2\alpha_k}\left(d(x^k; S^*) - d(x^{k+1}; S^*)\right)\left(d(x^k; S^*) + d(x^{k+1}; S^*)\right) \\ &\geq \frac{\beta}{2\alpha_k}\left(d(x^k; S^*)^2 - d(x^{k+1}; S^*)^2\right). \end{aligned} \tag{50}$$

On the other hand, we get from Corollary 3.1(i) that

$$\begin{aligned} \frac{1}{2\alpha_k}\left(d(x^k; S^*)^2 - d(x^{k+1}; S^*)^2\right) &\geq \frac{1}{2\alpha_k}\left(\|x^k - \Pi_{S^*}(x^k)\|^2 - \|x^{k+1} - \Pi_{S^*}(x^k)\|^2\right) \\ &\geq F(x^{k+1}) - F(\Pi_{S^*}(x^k)). \end{aligned}$$

This together with (50) imply

$$F(x^k) - F(x^{k+1}) \geq \beta\left[F(x^{k+1}) - F(\Pi_{S^*}(x^k))\right] = \beta\left[F(x^{k+1}) - F(x^*)\right].$$

It follows that $F(x^k) - F(x^*) \geq (1+\beta)[F(x^{k+1}) - F(x^*)]$, which clarifies (47). □

It is worth noting that Proposition 4.1 and Theorem 4.1 remain valid if we replace the local Lipschitz continuity of ∇f by the weaker one that all the step sizes α_k are bounded below by $\alpha > 0$.

Remark 4.1 (Linear convergence rate comparisons) If ∇f is globally Lipschitz continuous with constant L and $\sigma = \frac{1}{L}$, **Linesearch BT** does not proceed new step due to the Descent Lemma, i.e., $\alpha_k = \sigma$ for any k.[3] The involvement of θ is not necessary, hence α can be chosen as $1/L$ in Theorem 4.1. For this special case, we now summarize the comparison of the derived linear convergence rate with the ones in the literature as below:

– For the convergence rate of $(x^k)_{k\in\mathbb{N}}$: Our derived R-linear convergence rate $\frac{1}{\sqrt{1+\alpha\kappa}}$ is sharper (or smaller) than the corresponding R-linear convergence rate $\frac{1}{\sqrt{1+\frac{\alpha\kappa}{4}}}$ in [19, Theorem 4.2(iii)]. Moreover, our derived R-linear convergence rate $\frac{1}{\sqrt{1+\alpha\kappa}}$ and Q-linear convergence rate $\frac{1}{\sqrt{1+\frac{\alpha\kappa}{4}}}$ are also sharper than the R-rate obtained in [25, Theorem 3.2 and Corollary 3.6], which is $\sqrt{1-\frac{1}{2L\gamma}}$ with $\gamma = (2\kappa^{-1}+t)(1+Lt)$ and $t = L^{-1}$. Indeed, with some simple algebras, we have

$$\left(1+\frac{\alpha\kappa}{4}\right)\left(1-\frac{1}{2L\gamma}\right) = \left(1+\frac{\alpha\kappa}{4}\right)\left(1-\frac{1}{8(\alpha\kappa)^{-1}+4}\right) = 1 + \frac{4+3\alpha\kappa}{4(8(\alpha\kappa)^{-1}+4)} > 1.$$

[3] This observation comes from one of the referees.

So, $\sqrt{1-\frac{1}{2L\gamma}} > \frac{1}{\sqrt{1+\frac{\alpha\kappa}{4}}} \geq \frac{1}{\sqrt{1+\alpha\kappa}}$, and hence, our derived rate is sharper here.

– For the convergence rate of $(F(x^k))_{k\in\mathbb{N}}$: Our Q-linear convergence rate for $F(x^k)$ is bigger than the one $(1+\frac{\alpha\kappa}{4})^{-1}$ in [19, Theorem 4.2(iii)]. This can be seen by noticing

$$\frac{\sqrt{1+\alpha\kappa}+1}{2\sqrt{1+\alpha\kappa}} = \frac{1}{1+\frac{\sqrt{1+\alpha\kappa}-1}{\sqrt{1+\alpha\kappa}+1}} = \frac{1}{1+\frac{\alpha\kappa}{(\sqrt{1+\alpha\kappa}+1)^2}} > \frac{1}{1+\frac{\alpha\kappa}{4}}.$$

At this moment, we are not sure that whether our Q-linear convergence rate for the function value could be improved as in [19] without passing to the 2-Lojasiewicz property and global Lipschitz continuity. This would be a future direction of research.

Remark 4.2 (Extension to Hilbert spaces) Our results above could be extended to infinite-dimensional Hilbert spaces $\mathscr{H}$ by further assuming that the initial point x^0 belongs to the ball $\mathbb{B}_\varepsilon(x^*)$ in the quadratic growth condition (44), where x^* is the *weak* convergence limit point of $(x^k)_{k\in\mathbb{N}}$ as in Theorem 3.1. To see this, since the distances $\|x^k - x^*\|$ are decreasing as in Corollary 3.1(i), $x^k \in \mathbb{B}_\varepsilon(x^*)$ for all $k \in \mathbb{N}$. The inequalities (45) do not change. Hence, $d(x^k; S^*)$ linearly converges to 0. As $(x^k)_{k\in\mathbb{N}}$ is Fejér monotone with respect to S^* by Corollary 3.1(i), $(x^k)_{k\in\mathbb{N}}$ strongly converges to x^*; see [4, Theorem 5.11]. This strong convergence of $(x^k)_{k\in\mathbb{N}}$ in Hilbert space is similar to [19, Theorem 4.1 and first part of Example 4.11] with different explanation. The rest of the proofs of Proposition 4.1 and Theorem 4.1 remain.

Next, we obtain a sharper Q-linear convergence rate of $(x^k)_{k\in\mathbb{N}}$ under a stronger assumption: strong quadratic growth condition, which will be used in our subsequent part [17].

Corollary 4.1 (Sharper Q-linear convergence rate under strong quadratic growth condition) *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated from the FBS method. Suppose that the solution set S^* is not empty and let $x^* \in S^*$ be the convergent point of $(x^k)_{k\in\mathbb{N}}$ as in Theorem 3.1. Suppose further that ∇f is locally Lipschitz continuous around x^* with constant $L > 0$. If F satisfies the strong quadratic growth condition at x^* with modulus $\kappa > 0$, then there exists some $K \in \mathbb{N}$ such that for any $k > K$ we have*

$$\|x^{k+1} - x^*\| \leq \frac{1}{\sqrt{1+\alpha\kappa}}\|x^k - x^*\| \quad \textit{with} \quad \alpha := \min\left\{\alpha_K, \frac{\theta}{L}\right\}.$$

Proof This is a direct consequence of Proposition 4.1 with $S^* = \{x^*\}$. □

The assumption that F satisfies the quadratic growth condition in above results is automatic for a broad class of so-called *piecewise linear-quadratic functions* [36, Definition 10.20] defined below. Some other class of optimization problems satisfying this property will be discussed further in [17].

Definition 4.1 (Convex Piecewise Linear-Quadratic Functions) A function $h \in \Gamma_0(\mathscr{H})$ is called *convex piecewise linear-quadratic* if $\operatorname{dom} h$ is a union of finitely many polyhedral sets, relative to each of which $h(x)$ is given the expression of the form $\frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle + c$ for some scalar $c \in \mathbb{R}$, vector $b \in \mathbb{R}^n$ and a symmetric positive semi-definite matrix $A \in \mathbb{R}^{n\times n}$.

To end this section, we show next that the FBS method (5) has local uniform Q-linear rate when solving optimization problem, where the cost function is a convex piecewise linear-quadratic function.

Corollary 4.2 (Local Linear Convergence for Piecewise Linear-Quadratic Functions) *Let $(x^k)_{k\in\mathbb{N}}$ and $(\alpha_k)_{k\in\mathbb{N}}$ be the sequences generated from the FBS method. Suppose that $F = f + g$ is a convex piecewise linear-quadratic function, the solution set S^* is nonempty, and that ∇f is locally Lipschitz continuous around any point in S^*. Then, the sequences $(x^k)_{k\in\mathbb{N}}$ and $(F(x^k))_{k\in\mathbb{N}}$ are globally convergent to some optimal solution and optimal value respectively with local Q-linear rates. Furthermore, if ∇f is globally Lipschitz continuous on* $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$, *$(x^k)_{k\in\mathbb{N}}$ and $(F(x^k))_{k\in\mathbb{N}}$ are globally convergent with uniform local linear rates.*

Proof Suppose that the sequence $(x^k)_{k\in\mathbb{N}}$ converges to some $x^* \in S^*$ by Theorem 3.1. Since F is a convex piecewise linear-quadratic function, the graph of ∂F is polyhedral. By combining [37, Theorem 3H.1] and [30, Theorem 3.3], we obtain that F satisfies the quadratic growth condition at any minimizer with uniform modulus. It follows from Theorem 4.1 that $(x^k)_{k\in\mathbb{N}}$ and $(F(x^k))_{k\in\mathbb{N}}$ are locally convergent to some x^* and the optimal value $F(x^*)$, respectively, with Q-linear rate.

To complete the proof, suppose that ∇f is globally Lipschitz continuous on $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{dom} g$ with constant L. It follows from the last part of Theorem 4.1 that α could be chosen as $\min\{\sigma, \frac{\theta}{L}\}$. Since the modulus of quadratic growth condition of F is uniform as discussed above, the linear rate in Theorem 4.1 is independent of the choice of initial points. □

5 Conclusions

In this paper, we reanalyze the theory of FBS methods in Hilbert spaces and mainly study the Q-linear convergence of this method for solving nonsmooth convex optimization problems without the global Lipchitz continuity on ∇f. The quadratic growth condition plays significant roles in our analysis. It is well-recognized that KŁ-inequality with order $\frac{1}{2}$, which is a stronger condition than quadratic growth condition at minimizer in nonconvex settings, is a very useful tool to guarantee the convergence of many proximal-type algorithms. In future research, we intend to study the connection of quadratic growth condition with KŁ inequality and their effects on the convergence of proximal algorithms for nonconvex optimization problems.

Acknowledgements This work was partially supported by the National Science Foundation (NSF) Grants DMS—1816386 and 1816449, and a Discovery Project from Australian Research Council (ARC), DP190100555. The authors are deeply grateful to both anonymous referees for their careful readings and thoughtful suggestions that allowed us to improve the original presentation significantly. Many insightful remarks from one referee allow us to rewrite Section 3 in infinite dimensional spaces.

References

1. Daubechies, I., Defrise, M., De Mol, D.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun. Pure Appl. Math. **57**, 1413–1457 (2004)
2. Hale, E.T., Yin, W., Zhang, Y.: Fixed-point continuation for ℓ_1-minimization: methodology and convergence. SIAM J. Optim. **19**, 1107–1130 (2008)
3. Tibshirani, R.: Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. **58**, 267–288 (1996)
4. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)
5. Bredies, K., Lorenz, D.A.: Linear convergence of iterative soft-thresholding. J. Fourier Anal. Appl. **14**, 813–837 (2008)
6. Beck, A., Teboulle, M.: Gradient-Based Algorithms with Applications to Signal Recovery Problems. In: Palomar, D., Eldar, Y. (Eds.), Convex Optimization in Signal Processing and Communications, pp. 42–88. University Press, Cambribge (2010)
7. Combettes, P.L., Pesquet, J.-C.: Proximal splitting methods in signal processing. In: Fixed-Point Algorithms for Inverse Problems. Science and Engineering. Springer Optimization and Its Applications, vol. 49, pp. 185–212. Springer, New York (2011)
8. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward–backward splitting. Multiscale Model. Simul. **4**, 1168–1200 (2005)
9. Neal, P., Boyd, S.: Proximal algorithms. Found. Trends Optim. **1**, 127–239 (2014)
10. Davis, D., Yin, W.: Convergence rate analysis of several splitting schemes. In: Splitting Methods in Communications, Image Science, and Engineering. Scientific Computation, Springer, Cham (2016)
11. Tseng, P.: A modified forward–backward splitting method for maximal monotone mappings. SIAM J. Control Optim. **38**, 431–446 (2000)
12. Bello Cruz, J.Y., Nghia, T.T.A.: On the convergence of the proximal forward–backward splitting method with linesearches. Optim. Method Softw. **31**, 1209–1238 (2016)
13. Salzo, S.: The variable metric forward–backward splitting algorithm under mild differentiability assumptions. SIAM J. Optim. **27**, 2153–2181 (2017)
14. Bauschke, H.H., Bolte, J., Teboulle, M.: A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. Math. Oper. Res. **42**, 330–348 (2017)
15. Csiszár, I.: Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. Ann. Statist. **19**, 2032–2066 (1991)
16. Vardi, Y., Shepp, L.A., Kaufman, L.: A statistical model for positron emission tomography. J. Amer. Statist. Assoc. **80**, 8–37 (1985)
17. Bello-Cruz, J.Y., Li, G., Nghia, T.T.A.: On the linear convergence of forward-backward splitting method. Quadratic growth condition and uniqueness of optimal solution to Lasso, preprint, Part II (2020)
18. Bolte, J., Nguyen, T.P., Peypouquet, J., Suter, B.W.: From error bounds to the complexity of first-order descent methods for convex functions. Math. Program. **165**, 471–507 (2017)
19. Garrigos, G., Rosasco, L., Villa, S.: Convergence of the forward-backward algorithm: beyond the worst case with the help of geometry, arXiv:1703.09477 (2017)
20. Garrigos, G., Rosasco, L., Villa, S.: Thresholding gradient methods in Hilbert spaces: support identification and linear convergence, arXiv:1712.00357 (2017)
21. Li, G., Pong, T.K.: Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods. Found. Comp. Math. **18**, 1199–1232 (2018)
22. Li, G., Mordukhovich, B.S., Nghia, T.T.A., Pham, T.S.: Error bounds for parametric polynomial systems with applications to higher-order stability analysis and convergence rates. Math. Program. **168**, 313–346 (2018)
23. Zhou, Z., So, A.M.-C.: A unified approach to error bounds for structured convex optimization. Math. Program. **165**, 689–728 (2017)
24. Li, G.: Global error bounds for piecewise convex polynomials. Math. Program. **137**(1–2), 37–64 (2013)
25. Drusvyatskiy, D., Lewis, A.: Error bounds, quadratic growth, and linear convergence of proximal methods. Math. Oper. Res. **43**, 693–1050 (2018)
26. Necoara, I., Nesterov, Yu., Glineur, F.: Linear convergence of first order methods for non-strongly convex optimization. Math. Program. **175**, 69–107 (2019)
27. Tao, S., Boley, D., Zhang, S.: Local linear convergence of ISTA and FISTA on the Lasso problem. SIAM J. Optim. **26**, 313–336 (2016)

28. Luo, Z.-Q., Tseng, P.: Error bounds and convergence analysis of feasible descent methods: a general approach. Ann. Oper. Res. **46**, 157–178 (1993)
29. Azé, D., Corvellec, J.-N.: Nonlinear local error bounds via a change of metric. J. Fixed Point Theory Appl. **16**, 251–372 (2014)
30. Aragón Artacho, F.J., Geoffroy, M.H.: Characterizations of metric regularity of subdifferentials. J. Convex Anal. **15**, 365–380 (2008)
31. Drusvyatskiy, D., Mordukhovich, B.S., Nghia, T.T.A.: Second-order growth, tilt stability, and metric regularity of the subdifferential. J. Convex Anal. **21**, 1165–1192 (2014)
32. Liang, J., Fadili, J., Peyré, G.: Local linear convergence of forward-backward under partial smoothness. Adv. Neural Inf. Process Syst. (2014)
33. Liang, J., Fadili, J., Peyré, G.: Activity identification and local linear convergence of forward–backward type methods. SIAM J. Optim. **27**, 408–437 (2017)
34. Aragón Artacho, F.J., Geoffroy, M.H.: Metric subregularity of the convex subdifferential in Banach spaces. J. Nonlinear Convex Anal. **15**, 35–47 (2014)
35. Tseng, P., Yun, S.: A coordinate gradient descent method for nonsmooth separable minimization. Math. Program. **117**, 387–423 (2000)
36. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
37. Dontchev, A.L., Rockafellar, R.T.: Implicit Functions and Solution Mappings. A View from Variational Analysis. Springer, Dordrecht (2009)
38. Burachik, R.S., Iusem, A.N.: Set-Valued Mappings and Enlargements of Monotone Operators. Springer, Berlin (2008)