On the convergence of the proximal forward-backward splitting methods with linesearches

J.Y. Bello Cruz* T.T.A. Nghia†

January 10, 2015

Abstract

In this paper we focus on the convergence analysis of the proximal forward-backward splitting method for solving nonsmooth optimization problems in Hilbert spaces when the objective function is the sum of two convex functions. Assuming that one of the functions is Fréchet differentiable and using two new linesearches, the weak convergence is established without any Lipschitz continuity assumption. We obtain some complexity of the iterates when the stepsizes are bounded below by a positive constant. A fast version with linesearch is also provided.

Keywords: Armijo-type linesearch; Iteration complexity; Nonsmooth and convex optimization problems; Proximal forward-backward splitting method.

Mathematical Subject Classification (2010): 65K05, 90C25, 90C30.

1 Introduction

We are interested in solving problems of the following form:

$$\min\ f(x)+g(x)\ \text{ s.t. }\ x\in\mathcal{H}, \tag{1}$$

where $\mathcal{H}$ is a real Hilbert space with the inner product $\langle\cdot,\cdot\rangle$, and $f,g:\mathcal{H}\to\overline{\mathbb{R}}:=\mathbb{R}\cup\{+\infty\}$ are two proper lower semicontinuous convex functions in which f is Fréchet differentiable on the domain of g. The optimal solution set of this problem will be denoted by S_*. Recently problem (1) together with many variants of it has been received much attention from optimization community due to its broad applications to many disciplines such as optimal control, signal processing, system identification, machine learning, and image analysis; see, *e.g.*, [11,12,20] and the references therein. Many effective methods have been proposed to solve problem (1). Most of them keep using the idea of splitting f and g separately and taking the advantage of some Lipschitz assumption on the

*Institute of Mathematics and Statistics, Federal University of Goiás, Goiânia, GO. 74001-970, Brazil. E-mails: yunier@impa.br & yunier@ufg.br. This author was partially supported by CNPq grants 303492/2013-9, 474160/2013-0 and 202677/2013-3 and by projects CAPES-MES-CUBA 226/2012 and UNIVERSAL FAPEG/CNPq.

†Department of Mathematics and Statistics, Oakland University, Rochester, MI. 48309, USA. E-mail: nttran@oakland.edu.

derivative of f at each iteration. Here we focus our attention on the so-called proximal forward-backward splitting method, which contains a forward gradient step of f (an explicit step) followed by a backward proximal step of g (an implicit step) for problem (1). In this work linesearches are used to eliminate the undesired Lipschitz assumption mostly imposed in the literature.

To describe and motivate our methods, let us recall here the so-called proximal operator $\mathrm{prox}_g := (\partial g + \mathrm{Id})^{-1}$, where ∂g is the classical convex subdifferential of g and Id is the identity operator in $\mathcal{H}$. Among many important properties of proximal operators, it is well-known that prox_g is well-defined with full domain, single-valued, and even nonexpansive; see, *e.g.*, [1, 11, 12]. Furthermore, for any $\alpha > 0$, x is an optimal solution to problem (1) if and only if $x = \mathrm{prox}_{\alpha g}(x - \alpha \nabla f(x))$. This indeed motivates the construction of the iterative sequence forming the proximal forward-backward iteration as following:

$$x^{k+1} := \mathrm{prox}_{\alpha_k g}(x^k - \alpha_k \nabla f(x^k)) \tag{2}$$

with positive stepsize α_k. The iteration presented in (2) has been attracted extensive interests due to its simplicity and several important advantages. It is well-known that this method uses little storage, readily exploits the separable structure of problem (1), and is easily implemented to practical applications; see [20]. Moreover, scheme (2) may reduce to many popular optimization methods as particular cases including the projected gradient method for smooth constrained minimization; the proximal point method; the CQ algorithm for the split feasibility problem; the projected Landweber algorithm for constrained least squares; the iterative soft thresholding algorithm for linear inverse problems; decomposition methods for solving variational inequalities; and the simultaneous orthogonal projection algorithm for the convex feasibility problem; see, *e.g.*, [2, 9, 13, 14, 26–28] and the references therein.

The convergence of the iteration (2) to an optimal solution of (1) is usually established under the assumption that the gradient of f is Lipschitz continuous and the stepsize α_k is taken less than some constant related with the Lipschitz constant; see, *e.g.*, [12, Theorem 3.4(i)]. In this case, the main machinery to prove the convergence and its complexity is based on the renowned *Baillon-Haddad Theorem* [1, Corollary 18.16]. When ∇f is still Lipschitz continuous but somehow the Lipschitz constant is not known, choosing the stepsize α_k that guarantees the convergence of (2) would be a challenge. However, the following linesearch proposed in [3] overcome this inconvenience: choosing the stepsize α_k in (2) as the largest $\alpha \in \{\sigma, \sigma\theta, \sigma\theta^2, \ldots, \}$ with constants $\sigma > 0$ and $\theta \in (0, 1)$ such that:

$$f(J(x^k, \alpha)) \le f(x^k) + \langle \nabla f(x^k), J(x^k, \alpha) - x^k \rangle + \frac{1}{2\alpha} \|x^k - J(x^k, \alpha)\|^2, \tag{3}$$

where $J(x^k, \alpha) := \mathrm{prox}_{\alpha g}(x^k - \alpha \nabla f(x^k))$ and $\| \cdot \|$ is the norm induced by the inner product in $\mathcal{H}$. Unfortunately, this linesearch is well-defined by taking the advantage of the Lipschitz assumption for ∇f again via the so-called *Descent Lemma* [1, Theorem 18.15(iii)]. As far as we observe, the theory of convergence and complexity for the proximal forward-backward is almost complete under such a Lipschitz assumption. However, that Lipschitz condition fails in many natural circumstance; see, *e.g.*, [10]. It is quite interesting to question the convergence of the method and its complexity without the Lipschitz assumption aforementioned. In [29] Tseng provided an evidence of positive answer even for more general problems with maximal monotone operators. This motivates us to implement his main idea for functional problems (1) in our **Method 1** to release the Lipschitz continuity assumption of ∇f. Working on functionals rather than just operators actually gives us many advantages. Indeed, we completely relax an (expensive) extra projection step from Tseng's scheme and omit several unnatural assumptions in the main theorem [29, Theorem 3.4]. Further information for the convergence of the cost value sequence is also obtained. Moreover, in the spirit of linesearch on functionals like (3), we also introduce a new linesearch mainly used in our

Method 3. Both **Method 1** and **Method 3** not only relax the Lipschitz assumption on ∇f but also guarantee weak convergence of the generated sequences to optimal solutions.

Another achievement of our work is the study on complexity of cost values at generated sequences. To obtain the rate $\mathcal{O}(k^{-1})$ to the optimal cost, the gradient ∇f is usually supposed to be globally Lipschitz continuous; see, e.g., [2, 3, 12, 20, 22]. Here we achieve the same rate with strictly weaker assumptions, for instance, ∇f only needs to be *locally* Lipschitz continuous at any optimal solution. In order to accelerate the method to the rate $\mathcal{O}(k^{-2})$, we also introduce a modification of the scheme in [2] for our **Method 1**. Again, global Lipschitz continuity on ∇f is not supposed.

The paper is organized as follows. The next section presents some preliminary results that will be used throughout the paper. We also discuss here our standing assumptions for the problem which is somewhat natural for the lack of Lipschitz assumption aforementioned. Section 3 devotes to the two different linesearches for the proximal forward-backward methods used in Sections 4 and 5. Weak convergence and complexity of the proximal forward-backward method with the first linesearch are analyzed in Section 4. Moreover, we also consider its accelerated version here. Section 5 provides a similar study for a variant of the proximal forward-backward method with the second linesearch. We complete the paper with the conclusion for further research.

2 Preliminary results

In this section we present some definitions and results needed for our paper. Let $h : \mathcal{H} \to \overline{\mathbb{R}}$ be a proper, lower semicontinuous (*l.s.c.*), and convex function. We denote the domain of h by $\operatorname{dom} h := \{x \in \mathcal{H} \mid h(x) < +\infty\}$. For any $x \in \operatorname{dom} h$, the directional derivative of h at x in the direction d is

$$h'(x; d) := \lim_{t \to 0^+} \frac{h(x + td) - h(x)}{t},$$

which always exists (although it may be infinite). The subdifferential of h at x is defined by

$$\partial h(x) := \{v \in \mathcal{H} \mid \langle v, y - x\rangle \le h(y) - h(x),\ y \in \mathcal{H}\}. \tag{4}$$

Fact 2.1 ([1, Proposition 17.2]). *For $x \in \operatorname{dom} h$ and $y \in \mathcal{H}$, the following hold:*

(i) *The function $\phi : \mathbb{R}_{++} \to \overline{\mathbb{R}}$ with $\phi(\beta) = \dfrac{h(x + \beta y) - h(x)}{\beta}$ is increasing*

(ii) *$h'(x; y)$ exists and $h'(x; y) = \inf_{\beta \in \mathbb{R}_{++}} \dfrac{h(x + \beta y) - h(x)}{\beta}$.*

(iii) *$h'(x; y - x) + h(x) \le h(y)$.*

Fact 2.2 ([7, Theorem 4.7.1 and Proposition 4.2.1(i)]). *The subdifferential operator ∂h is maximal monotone,* i.e.*, it has no proper monotone extension in the graph inclusion sense. Moreover, the graph of ∂h, $\operatorname{Gph}(\partial h) := \{(x, v) \in \mathcal{H} \times \mathcal{H} \mid v \in \partial h(x)\}$ is demiclosed,* i.e.*, if the sequence $(x^k, v^k)_{k\in\mathbb{N}} \subset \operatorname{Gph}(\partial h)$ satisfies that $(x^k)_{k\in\mathbb{N}}$ converges weakly to x and $(v^k)_{k\in\mathbb{N}}$ converges strongly to v, then $(x, v) \in \operatorname{Gph}(\partial h)$.*

Next we set the standing assumptions on the data of problem (1) used throughout the paper as follows:

A1 $f, g : \mathcal{H} \to \overline{\mathbb{R}}$ are two proper *l.s.c.* convex functions with $\operatorname{dom} g \subseteq \operatorname{dom} f$;

A2 The function f is Fréchet differentiable on $\operatorname{dom} g$. The gradient ∇f is uniformly continuous on any bounded subset of $\operatorname{dom} g$ and maps any bounded subset of $\operatorname{dom} g$ to a bounded set in $\mathcal{H}$.

Assumption **A1** is popular and crucial for the well-definedness of the proximal forward-backward iteration (2). Our Assumption **A2** is not restrictive. Indeed, it is easy to check that **A2** is strictly weaker than the one usually used in the literature that is ∇f is globally Lipschitz continuous. For instance, the gradient of the convex functions $f(x) = \|x\|^p$ $(1 < p < +\infty)$ with $p \neq 2$, $x \in \mathcal{H}$ satisfies all the conditions in Assumption **A2** but it is not globally Lipschitz continuous. When $\mathcal{H}$ is a finite-dimensional space and the domain of g is closed, Assumption **A2** actually means that f is continuously differentiable on $\operatorname{dom} g$ as proved below. It is worth noting that the closedness of $\operatorname{dom} g$ is broadly assumed in the literature for problem (1) especially including the case of optimization problems with geometric constraint which can be written as (1) when g is an indicator function; see, e.g., [20].

Proposition 2.3. *Let $\mathcal{H}$ be a finite-dimensional space and let $f, g : \mathcal{H} \to \overline{\mathbb{R}}$ be two functions satisfying* **A1**. *Suppose that there is a closed set $X \subseteq \mathcal{H}$ such that $\operatorname{dom} g \subseteq X \subseteq \operatorname{dom} f$ and that f is continuously differentiable on X. Then Assumption* **A2** *is satisfied.*

Consequently, if $\operatorname{dom} g$ is closed then the validity of Assumption **A2** *is equivalent to the statement that f is continuously differentiable on $\operatorname{dom} g$.*

Proof. To justify, suppose that $\dim \mathcal{H} < +\infty$ and that f is continuously differentiable on X with $\operatorname{dom} g \subseteq X \subseteq \operatorname{dom} f$. Take any bounded set A of $\operatorname{dom} g$. Hence ∇f is uniformly continuous on the compact set $\operatorname{cl} A \subseteq X$ and thus on A due to the classical *Heine-Cantor Theorem*.

Furthermore, since ∇f is continuous on X, it maps the compact set $\operatorname{cl} A \subseteq X$ to a compact set in $\mathcal{H}$. This verifies that $\nabla f(A)$ is bounded and completes the first part of the proposition.

Now suppose that $\operatorname{dom} g$ is closed. It is easy to see that the validity of Assumption **A2** implies that f is continuously differentiable on $\operatorname{dom} g$. Conversely, if f is continuously differentiable on $\operatorname{dom} g$, Assumption **A2** is also satisfied by taking $X = \operatorname{dom} g$ in the first part. The proof is completed. □

Fact 2.4 ([7, Theorem 3.5.7])**.** *Let $f, g : \mathcal{H} \to \overline{\mathbb{R}}$ be two functions satisfying* **A1** *and* **A2**. *Then we have $\partial(f+g)(x) = \nabla f(x) + \partial g(x)$ for all $x \in \operatorname{dom} g$.*

Let us recall the proximal operator mentioned in Section 1 that $\operatorname{prox}_g : \mathcal{H} \to \operatorname{dom} g$ with $\operatorname{prox}_g(z) = (\operatorname{Id} + \partial g)^{-1}(z)$, $z \in \mathcal{H}$. It is worth noting that the proximal operator has full domain. Furthermore, observe that

$$\frac{z - \operatorname{prox}_{\alpha g}(z)}{\alpha} \in \partial g(\operatorname{prox}_{\alpha g}(z)) \quad \text{for all} \quad z \in \mathcal{H},\, \alpha \in \mathbb{R}_{++} := \{r \in \mathbb{R} |\, r > 0\}. \tag{5}$$

We also denote the *proximal forward-backward operator* $J : \operatorname{dom} f \times \mathbb{R}_{++} \to \operatorname{dom} g \subset \mathcal{H}$ by

$$J(x, \alpha) := \operatorname{prox}_{\alpha g}(x - \alpha \nabla f(x)) \quad \text{for all} \quad x \in \operatorname{dom} f,\, \alpha > 0. \tag{6}$$

The following lemma is very useful for our further study.

Lemma 2.5. *Let $f, g : \mathcal{H} \to \overline{\mathbb{R}}$ be two functions satisfying Assumption* **A1**. *Then for any $x \in \operatorname{dom} f$ and $\alpha_2 \geq \alpha_1 > 0$, we have*

$$\frac{\alpha_2}{\alpha_1}\|x - J(x, \alpha_1)\| \geq \|x - J(x, \alpha_2)\| \geq \|x - J(x, \alpha_1)\|. \tag{7}$$

Proof. By using (5) with $z = x - \alpha\nabla f(x)$ and $x \in \operatorname{dom} f$, we have

$$\frac{x - \alpha\nabla f(x) - J(x,\alpha)}{\alpha} \in \partial g(J(x,\alpha)) \tag{8}$$

for all $\alpha > 0$. Take any $\alpha_2 \geq \alpha_1 > 0$, it follows from the monotonicity of ∂g and (8) that

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

which easily implies the following expression

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

Since $\dfrac{\alpha_2}{\alpha_1} \geq 1$, we derive (7) and thus complete the proof of the lemma. □

Let us end the section by recalling the well-known concepts so-called quasi-Fejér and Fejér convergence. The definition originates in [15] and has been elaborated further in [8, 17].

Definition 2.1. *Let S be a nonempty subset of $\mathcal{H}$. A sequence $(x^k)_{k\in\mathbb{N}}$ in $\mathcal{H}$ is said to be quasi-Fejér convergent to S if and only if for all $x \in S$ there exists a sequence $(\epsilon_k)_{k\in\mathbb{N}}$ in $\mathbb{R}_+$ such that $\sum_{k=0}^{\infty} \epsilon_k < +\infty$ and $\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + \epsilon_k$ for all $k \in \mathbb{N}$. When $(\epsilon_k)_{k\in\mathbb{N}}$ is a null sequence, we say that $(x^k)_{k\in\mathbb{N}}$ is Fejér convergent to S.*

Fact 2.6 ([17, Theorem 4.1])**.** *If $(x^k)_{k\in\mathbb{N}}$ is quasi-Fejér convergent to S, then one has*

(i) *The sequence $(x^k)_{k\in\mathbb{N}}$ is bounded;*

(ii) *If all accumulation points of $(x^k)_{k\in\mathbb{N}}$ belong to S, then $(x^k)_{k\in\mathbb{N}}$ is weakly convergent to a point in S.*

3 The linesearches

In this section we present two different linesearches mainly used in the proximal forward-backward methods proposed in Sections 4 and 5. The first one contains a backtracking procedure which computes *at least one* backward step (implicit step) inside the updating inner loop for finding the steplength. This linesearch is a particular case of the one proposed in [29] for solving inclusion problems. It will be used in **Method 1** and **Method 2** of Section 4.

Linesearch 1. *Given x, $\sigma > 0$, $\theta \in (0,1)$ and $\delta \in (0, 1/2)$.*
Input. *Set $\alpha = \sigma$ and $J(x,\alpha) := \operatorname{prox}_{\alpha g}(x - \alpha\nabla f(x))$ with $x \in \operatorname{dom} g$.*
  While $\alpha \left\|\nabla f\big(J(x,\alpha)\big) - \nabla f(x)\right\| > \delta \|J(x,\alpha) - x\|$ **do**
  $\alpha = \theta\alpha$.
  End While
Output. α.

The well-definedness of **Linesearch 1** follows from [29, Theorem 3.4(a)]. For the reader's convenience, we provide a different proof revealing that the convexity of f is not necessary.

Lemma 3.1. *If $x \in \operatorname{dom} g$ then* **Linesearch 1** *stops after finitely many steps.*

Proof. If $x \in S_*$ then $x = J(x, \sigma)$. Thus the linesearch stops with zero step and gives us the output σ. If $x \notin S_*$, by contradiction suppose that for all $\alpha \in \mathcal{P} := \{\sigma, \sigma\theta, \sigma\theta^2, \ldots\}$,

$$\alpha \left\| \nabla f\big(J(x,\alpha)\big) - \nabla f(x) \right\| > \delta \left\| J(x,\alpha) - x \right\|. \tag{9}$$

When $\alpha \in \mathcal{P}$ is sufficiently closed to 0, it follows from Lemma 2.5 that $J(x,\alpha)$ is uniformly bounded. Thus we get from (9) that $\|x - J(x,\alpha)\| \to 0$ as $\alpha \downarrow 0$ thanks to Assumption **A2**. The latter implies $\|\nabla f\big(J(x,\alpha)\big) - \nabla f(x)\| \to 0$ when $\alpha \downarrow 0$ by Assumption **A2** again. Thus we get from (9) that

$$\lim_{\alpha \downarrow 0} \frac{\|x - J(x,\alpha)\|}{\alpha} = 0. \tag{10}$$

Employing (5) with $z = x - \alpha \nabla f(x)$ gives us that

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

By letting $\alpha \downarrow 0$ in the above inclusion and using (10), we get from the demiclosedness of $\operatorname{Gph}(\partial g)$ from Fact 2.2 that $0 \in \nabla f(x) + \partial g(x)$. This contradicts the assumption that x is not an optimal solution to problem (1) and completes the proof of the lemma. $\square$

Next we propose the second backtracking procedure. In contrast to **Linesearch 1**, this linesearch demands *only one* evaluation of the backward step and uses it in all possible iterations. This is somehow an advantage of this linesearch, since in many practical problems computing the proximal operator many times may be very expensive. The linesearch is indeed a generalization of the one studied in [5] for solving the nonlinear constrained optimization problem ($g = \delta_C$). We will employ it in **Method 3** in Section 5.

Linesearch 2. *Given x and $\theta \in (0,1)$.*
Input. *Set $\beta = 1$, $J_x := J(x,1) = \operatorname{prox}_g(x - \nabla f(x))$ with $x \in \operatorname{dom} g$.*

While $(f+g)\,(x - \beta(x - J_x)) > (f+g)(x) - \beta\,[g(x) - g(J_x)] - \beta\langle \nabla f(x), x - J_x\rangle + \frac{\beta}{2}\|x - J_x\|^2$ **do**

$\beta = \theta\beta$.

End While

Output. β.

Similarly to **Linesearch 1**, we also have finite termination for **Linesearch 2**. It is important to note that the well-definedness analysis is done without assuming the second part of **A2** (uniformly continuity and boundedness).

Lemma 3.2. *If $x \in \operatorname{dom} g$ then* **Linesearch 2** *stops after finitely many steps.*

Proof. If $x \in S_*$ we have $x = J_x$. Thus the linesearch immediately gives us the output 1 without proceeding any step. If $x \notin S_*$, by contradiction let us assume that **Linesearch 2** does not stop after finitely many steps. Thus for all $\beta \in \mathcal{Q} := \{1, \theta, \theta^2, \ldots\}$, we have

$$(f+g)(x - \beta(x - J_x)) > (f+g)(x) - \beta\,[g(x) - g(J_x)] - \beta\langle \nabla f(x), x - J_x\rangle + \frac{\beta}{2}\|x - J_x\|^2.$$

It follows that

$$\frac{(f+g)(x-\beta(x-J_x)) - (f+g)(x)}{\beta} + g(x) - g(J_x) + \langle \nabla f(x), x - J_x\rangle > \frac{1}{2}\|x - J_x\|^2.$$

Taking $\beta \downarrow 0$ and using the Fréchet differentiability of f and the convexity of g give us that

$$\begin{aligned}\frac{1}{2}\|x - J_x\|^2 \leq & \langle \nabla f(x), J_x - x\rangle + g'(x; J_x - x) + g(x) - g(J_x) + \langle \nabla f(x), x - J_x\rangle \\ = & g'(x; J_x - x) + g(x) - g(J_x) \leq 0,\end{aligned}$$

where the last inequality follows from Fact 2.1(iii). Hence we have $x = J_x$, which readily implies that $x - \nabla f(x) \in \partial g(x) + x$, i.e., $0 \in \nabla f(x) + \partial g(x)$. This is a contradiction due to the assumption $x \notin S_*$. □

4 The proximal forward-backward method with Linesearch 1

This section devotes to the study of the proximal forward-backward splitting method with **Linesearch 1**. We mainly derive the weak convergence of the generated sequences from this method and also obtain the same complexity of [3, Theorem 1.1] for the sequences generated by the forward-backward iteration under a weaker assumption than the Lipschitz one on ∇f usually imposed in the literature.

The following method has some similarities to the one proposed in [29] for maximal monotone operators. However, it completely relaxes an extra expensive projection step [29, Equation (2.3)] and seems to be more natural in comparison with the classical proximal forward-backward splitting method (2).

Method 1.

Initialization Step. *Take* $x^0 \in \operatorname{dom} g$, $\sigma > 0$, $\theta \in (0,1)$ *and* $\delta \in (0, 1/2)$.

Iterative Step. *Given* x^k *set*

$$x^{k+1} = J(x^k, \alpha_k) := \operatorname{prox}_{\alpha_k g}(x^k - \alpha_k \nabla f(x^k)), \tag{11}$$

where $\alpha_k :=$ **Linesearch 1**$(x^k, \sigma, \theta, \delta)$.

Stop Criteria. *If* $x^{k+1} = x^k$, *then stop.*

First note that from Lemma 3.1 that **Linesearch 1** for finding the stepsize α_k in the above scheme is finite. Hence the choice of sequence $(x^k)_{k\in\mathbb{N}}$ in **Method 1** is well-defined. Another important feature from the definition of **Linesearch 1** useful for our analysis is the following inequality

$$\alpha_k \left\|\nabla f(x^{k+1}) - \nabla f(x^k)\right\| \leq \delta \left\|x^{k+1} - x^k\right\|. \tag{12}$$

Note further that if **Method 1** stops at iteration k then we have $x^k = \text{prox}_{\alpha_k g}(x^k - \alpha_k \nabla f(x^k))$ and consequently $x^k \in S_*$. Otherwise, we will mainly show that the sequence $(x^k)_{k\in\mathbb{N}}$ generated by this method is converging weakly to some optimal solution. To verify this claim, we need some auxiliary results as follows.

Proposition 4.1. *Let $\alpha_k =$ **Linesearch 1**$(x^k, \sigma, \theta, \delta)$. For all $k \in \mathbb{N}$ and $x \in \text{dom}\, g$, we have*

(i) $\|x^k - x\|^2 - \|x^{k+1} - x\|^2 \geq 2\alpha_k \left[(f+g)(x^{k+1}) - (f+g)(x)\right] + (1-2\delta)\|x^{k+1} - x^k\|^2$;

(ii) $(f+g)(x^{k+1}) - (f+g)(x^k) \leq -\dfrac{(1-\delta)}{\alpha_k}\|x^{k+1} - x^k\|^2$.

Proof. First let us justify **(i)** by noting from (5) and (11) that

$$\frac{x^k - x^{k+1}}{\alpha_k} - \nabla f(x^k) = \frac{x^k - J(x^k, \alpha_k)}{\alpha_k} - \nabla f(x^k) \in \partial g(J(x^k, \alpha_k)) = \partial g(x^{k+1}).$$

It follows from the convexity of g that

$$g(x) - g(x^{k+1}) \geq \left\langle \frac{x^k - x^{k+1}}{\alpha_k} - \nabla f(x^k), x - x^{k+1} \right\rangle \quad \text{for all} \quad x \in \text{dom}\, g. \tag{13}$$

Since f is convex, we also have

$$f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle \quad \text{for all} \quad x, y \in \text{dom}\, f. \tag{14}$$

Summing (13) and (14) with any $x \in \text{dom}\, g \subseteq \text{dom}\, f$ and $y = x^k \in \text{dom}\, g$ gives us the following expressions

$$\begin{aligned}
(f+g)(x) \geq & f(x^k) + g(x^{k+1}) + \left\langle \frac{x^k - x^{k+1}}{\alpha_k} - \nabla f(x^k), x - x^{k+1} \right\rangle + \langle \nabla f(x^k), x - x^k \rangle \\
= & f(x^k) + g(x^{k+1}) + \frac{1}{\alpha_k}\left\langle x^k - x^{k+1}, x - x^{k+1} \right\rangle + \langle \nabla f(x^k), x^{k+1} - x^k \rangle \\
= & f(x^k) + g(x^{k+1}) + \frac{1}{\alpha_k}\left\langle x^k - x^{k+1}, x - x^{k+1} \right\rangle + \langle \nabla f(x^k) - \nabla f(x^{k+1}), x^{k+1} - x^k \rangle \\
& + \langle \nabla f(x^{k+1}), x^{k+1} - x^k \rangle \\
\geq & f(x^k) + g(x^{k+1}) + \frac{1}{\alpha_k}\left\langle x^k - x^{k+1}, x - x^{k+1} \right\rangle - \|\nabla f(x^k) - \nabla f(x^{k+1})\| \cdot \|x^{k+1} - x^k\| \\
& + \langle \nabla f(x^{k+1}), x^{k+1} - x^k \rangle \\
\geq & f(x^k) + g(x^{k+1}) + \frac{1}{\alpha_k}\left\langle x^k - x^{k+1}, x - x^{k+1} \right\rangle - \frac{\delta}{\alpha_k}\|x^{k+1} - x^k\|^2 + \langle \nabla f(x^{k+1}), x^{k+1} - x^k \rangle,
\end{aligned}$$

where the last inequality follows from (12) again. After rearrangement we get

$$\langle x^k - x^{k+1}, x^{k+1} - x \rangle \geq \alpha_k[f(x^k) + g(x^{k+1}) - (f+g)(x) + \langle \nabla f(x^{k+1}), x^{k+1} - x^k \rangle] - \delta\|x^{k+1} - x^k\|^2. \tag{15}$$

Observe that $2\langle x^k - x^{k+1}, x^{k+1} - x \rangle = \|x^k - x\|^2 - \|x^{k+1} - x\|^2 - \|x^k - x^{k+1}\|^2$. Combining this with (15) implies

$$\begin{aligned}
\|x^k - x\|^2 - \|x^{k+1} - x\|^2 \geq & 2\alpha_k[f(x^k) + g(x^{k+1}) - (f+g)(x)] \\
& + 2\alpha_k \langle \nabla f(x^{k+1}), x^{k+1} - x^k \rangle + (1-2\delta)\|x^k - x^{k+1}\|^2.
\end{aligned} \tag{16}$$

By using (14) with $x = x^k$ and $y = x^{k+1}$, we have $f(x^k) - f(x^{k+1}) \geq \langle \nabla f(x^{k+1}), x^k - x^{k+1} \rangle$. This together with (16) gives us that

$$\|x^k - x\|^2 - \|x^{k+1} - x\|^2 \geq 2\alpha_k[(f+g)(x^{k+1}) - (f+g)(x)] + (1-2\delta)\|x^k - x^{k+1}\|^2,$$

which verifies **(i)**. Note further that **(ii)** is a consequence of **(i)** when $x = x^k$. The proof of the proposition is complete. □

Proposition 4.1(ii) shows that **Method 1** is a descent method in the sense that the value of the cost function $f+g$ at each iteration is decreasing. Furthermore, it is easy to check from Proposition 4.1(i) that the generated sequence of **Method 1** is Fejér convergent to the optimal solution set S_* whenever $S_* \neq \emptyset$. This observation is indeed the center of the following main result of this section.

Theorem 4.2. *Let* $(x^k)_{k\in\mathbb{N}}$ *and* $(\alpha_k)_{k\in\mathbb{N}}$ *be the sequence generated by* **Method 1**. *The following statements hold:*

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

$$\lim_{k\to\infty} (f+g)(x^k) = \min_{x\in\mathcal{H}} (f+g)(x). \tag{17}$$

(ii) *If* $S_* = \emptyset$ *then we have*

$$\lim_{k\to\infty} \|x^k\| = +\infty \quad \textit{and} \quad \lim_{k\to\infty} (f+g)(x^k) = \inf_{x\in\mathcal{H}} (f+g)(x). \tag{18}$$

Proof. First let us justify **(i)** by supposing that $S_* \neq \emptyset$. By applying Proposition 4.1(ii) at any $x \in S_*$, we have

$$\begin{aligned}\|x^k - x\|^2 - \|x^{k+1} - x\|^2 &\geq 2\alpha_k[(f+g)(x^{k+1}) - (f+g)(x)] + (1-2\delta)\|x^k - x^{k+1}\|^2 \\ &\geq (1-2\delta)\|x^k - x^{k+1}\|^2 \geq 0. \end{aligned} \tag{19}$$

It follows that the sequence $(x^k)_{k\in\mathbb{N}}$ is Féjer convergent to S_* and thus is bounded by Fact 2.6(i). By using (19) at $x_* \in S_*$, we get

$$\begin{aligned} 0 \leq 2\alpha_k[(f+g)(x^{k+1}) - (f+g)(x_*)] &\leq \|x^k - x_*\|^2 - \|x^{k+1} - x_*\|^2 \\ &= (\|x^k - x_*\| - \|x^{k+1} - x_*\|) \cdot (\|x^k - x_*\| + \|x^{k+1} - x_*\|) \\ &\leq 2M(\|x^k - x_*\| - \|x^{k+1} - x_*\|) \\ &\leq 2M\|x^k - x^{k+1}\|, \end{aligned}$$

where $M := \sup\{\|x^k - x_*\| \mid k \in \mathbb{N}\} < +\infty$. Hence the above inequality leads us to

$$(f+g)(x^{k+1}) - (f+g)(x_*) \leq \frac{1}{M} \cdot \frac{\|x^k - x^{k+1}\|}{\alpha_k}. \tag{20}$$

Furthermore, observe from (19) that

$$(1-2\delta)\sum_{k=0}^{\infty} \|x^k - x^{k+1}\|^2 \leq \sum_{k=0}^{\infty} \left(\|x^k - x\|^2 - \|x^{k+1} - x\|^2\right) \leq \|x^0 - x\|^2,$$

which tells us that $\|x^k - x^{k+1}\| \to 0$ as $k \to \infty$.

Since $(x^k)_{k\in\mathbb{N}}$ is bounded, the set of its weak accumulation points is nonempty. Take any weak accumulation point $\bar{x}$ of $(x^k)_{k\in\mathbb{N}}$, we find a subsequence $(x^{n_k})_{k\in\mathbb{N}}$ weakly converging to $\bar{x}$. Now let us split our further analysis into two distinct cases.

Case 1. Suppose that the sequence $(\alpha_{n_k})_{k\in\mathbb{N}}$ defined in **Method 1** does not converge to 0. Hence there exist a subsequence (without relabeling) of $(\alpha_{n_k})_{k\in\mathbb{N}}$ and $\alpha > 0$ such that

$$\alpha_{n_k} \geq \alpha \quad \text{for sufficiently large } k. \tag{21}$$

Since $(x^k)_{k\in\mathbb{N}}$ is bounded and $\|x^k - x^{k+1}\| \to 0$ as claimed above, we get from Assumption **A2** that

$$\lim_{k\to\infty} \|\nabla f(x^{n_k}) - \nabla f(x^{n_k+1})\| = 0. \tag{22}$$

Since $x^{n_k+1} = J(x^{n_k}, \alpha_{n_k})$, it follows from (5), (11), and Fact 2.4 that

$$\frac{x^{n_k} - \alpha_{n_k}\nabla f(x^{n_k}) - x^{n_k+1}}{\alpha_{n_k}} + \nabla f(x^{n_k+1}) \in \partial g(x^{n_k+1}) + \nabla f(x^{n_k+1}) = \partial(f+g)(x^{n_k+1}),$$

which implies in turn the expression

$$[x^{n_k} - x^{n_k+1}]/\alpha_{n_k} + \nabla f(x^{n_k+1}) - \nabla f(x^{n_k}) \in \partial(f+g)(x^{n_k+1}). \tag{23}$$

Note further that x^{n_k+1} also converges weakly to $\bar{x}$ due to the fact that $\|x^{n_k} - x^{n_k+1}\| \to 0$ as $k \to \infty$. By passing $k \to \infty$ in (23), we get from (21), (22), and Fact 2.2 that $0 \in \partial(f+g)(\bar{x})$, which implies that $\bar{x} \in S_*$. Furthermore, since $(f+g)(x^k)$ is decreasing due to Proposition 4.1(ii), (17) is a consequence of (20) and (21).

Case 2. Suppose now that $\lim_{k\to\infty} \alpha_{n_k} = 0$. Define $\hat{\alpha}_{n_k} := \dfrac{\alpha_{n_k}}{\theta} > \alpha_{n_k} > 0$ and $\hat{x}^{n_k} := J\left(x^{n_k}, \hat{\alpha}_{n_k}\right)$. Due to Lemma 2.5 we have

$$\|x^{n_k} - \hat{x}^{n_k}\| = \|x^{n_k} - J(x^{n_k}, \hat{\alpha}_{n_k})\| \leq \frac{\hat{\alpha}_{n_k}}{\alpha_{n_k}}\|x^{n_k} - J(x^{n_k}, \alpha_{n_k})\| = \frac{1}{\theta}\|x^{n_k} - x^{n_k+1}\|,$$

which combines with the boundedness of $(x^{n_k})_{k\in\mathbb{N}}$ to show that the sequence $(\hat{x}^{n_k})_{k\in\mathbb{N}}$ is also bounded. It follows from the definition of **Linesearch 1** that

$$\hat{\alpha}_{n_k} \left\|\nabla f\left(\hat{x}^{n_k}\right) - \nabla f(x^{n_k})\right\| > \delta \left\|\hat{x}^{n_k} - x^{n_k}\right\|. \tag{24}$$

Since $\hat{\alpha}_{n_k} \downarrow 0$ and both $(x^{n_k})_{k\in\mathbb{N}}$ and $(\hat{x}^{n_k})_{k\in\mathbb{N}}$ are bounded, (24) together with Assumption **A2** tells us that $\lim_{k\to 0} \|\hat{x}^{n_k} - x^{n_k}\| = 0$ and thus $(\hat{x}^{n_k})_{k\in\mathbb{N}}$ also weakly converges to $\bar{x}$. Thanks to **A2** we obtain

$$\lim_{k\to 0} \left\|\nabla f\left(\hat{x}^{n_k}\right) - \nabla f(x^{n_k})\right\| = 0. \tag{25}$$

This and (24) imply that

$$\lim_{k\to 0} \frac{1}{\hat{\alpha}_{n_k}} \|\hat{x}^{n_k} - x^{n_k}\| = 0. \tag{26}$$

Using (5) with $z = x^{n_k} - \hat{\alpha}_{n_k}\nabla f(x^{n_k})$ and Fact 2.4 gives us that

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

By letting $k \to \infty$, we get from (25), (26), and Fact 2.2 that $0 \in \partial(f+g)(\bar{x})$, which means $\bar{x} \in S_*$.

From both cases above, we have that any weak accumulation point of $(x^k)_{k\in\mathbb{N}}$ is an element of S_*. Thanks to Fact 2.6(ii), the sequence $(x^k)_{k\in\mathbb{N}}$ weakly converges to some point in S_*. It remains to verify (17) in this case. Indeed, we get from Lemma 2.5 that

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

This together with (26) yields $\frac{\|x^{n_k}-x^{n_k+1}\|}{\alpha_{n_k}} \to 0$ as $k \to \infty$. Since $(f+g)(x^k)$ is decreasing due to Proposition 4.1(ii), we derive from the latter and (20) that

$$0 \geq \lim_{k\to\infty} (f+g)(x^{n_k}) - (f+g)(x_*) = \lim_{k\to\infty} (f+g)(x^k) - (f+g)(x^*) \geq 0,$$

which clearly ensures (17). This verifies **(i)** of the theorem.

To justify **(ii)**, suppose that $S_* = \emptyset$. Observe from the proof of **(i)** (without regarding (17) and (20)) that if $(x^k)_{k\in\mathbb{N}}$ has any weak accumulation point then this point is an optimal solution. Since $S_* = \emptyset$, any subsequence of $(x^k)_{k\in\mathbb{N}}$ is unbounded and thus $\|x^k\| \to +\infty$ as $k \to \infty$. Furthermore, note that

$$s := \lim_{k\to\infty} (f+g)(x^k) \geq \inf_{x\in\mathcal{H}} (f+g)(x) := s_*,$$

where s exists due to fact that $(f+g)(x^k)$ is decreasing as $k \to \infty$ by Proposition 4.1(ii). If $s > s_*$ then the following auxiliary set

$$S_{\text{lev}}(x^0) := \Big\{x \in \operatorname{dom} g\colon \quad (f+g)(x) \leq (f+g)(x^k),\ \forall k \in \mathbb{N}\Big\}$$

is nonempty. By applying Proposition 4.1(i) at any $x \in S_{\text{lev}}(x^0)$, similarly to (19) we also have $(x^k)_{k\in\mathbb{N}}$ is Féjer convergent to $S_{\text{lev}}(x^0)$. It follows from Fact 2.6(i) that the sequence $(x^k)_{k\in\mathbb{N}}$ is bounded, which is a contradiction. Hence we have $s = s_*$ and complete the proof of the theorem. □

As discussed before **Method 1**, our method improves the scheme in [29] for the case that the maximal monotone operators considered there are ∇f and ∂g by completely relaxing an additional step. Furthermore, our Theorem 4.2 also loosens some unnatural assumptions imposed in [29, Theorem 3.4(b)]. Working on functionals (f and g) instead of only operators (∇f and ∂g) indeed gives us many further advantages.

4.1 Complexity analysis of Method 1

In this subsection we present complexity analysis of the iterates in **Method 1**. When the stepsizes generated by **Linesearch 1** are bounded below by a positive number. Our analysis shows that the expected error from the cost value at the k-th iteration to the optimal value is $\mathcal{O}(k^{-1})$, which is the same rate of the first-order algorithm presented in [3, Theorem 1.1]. It is worth noting further that the global Lipschitz continuity assumption on the gradient ∇f is sufficient but not necessary for the boundedness from below of the stepsizes aforementioned; see our Proposition 4.5 below. Since $\alpha_k > 0$ for any $k \in \mathbb{N}$, this condition actually means that $\liminf_{k\to\infty} \alpha_k > 0$, which was used before in [29] for different purposes.

Theorem 4.3. *Let $(x^k)_{k\in\mathbb{N}}$ and $(\alpha_k)_{k\in\mathbb{N}}$ be the sequences generated in* **Method 1**. *Suppose that $S_* \neq \emptyset$ and there exists $\alpha > 0$ such that $\alpha_k \geq \alpha > 0$ for all $k \in \mathbb{N}$. Then,*

$$(f+g)(x^k) - \min_{x\in\mathcal{H}} (f+g)(x) \leq \frac{1}{2\alpha} \frac{[\operatorname{dist}(x^0, S_*)]^2}{k} \quad \textit{for all} \quad k \in \mathbb{N}. \tag{27}$$

Proof. Using Proposition 4.1(i) at $\ell \in \mathbb{N}$ gives us that

$$0 \le \ell \Big[(f+g)(x^\ell) - (f+g)(x^{\ell+1})\Big] = \ell(f+g)(x^\ell) - (\ell+1)(f+g)(x^{\ell+1}) + (f+g)(x^{\ell+1}).$$

By summing the above inequalities over $\ell = 0, 1, \ldots, k-1$, we have

$$-k(f+g)(x^k) + \sum_{\ell=0}^{k-1}(f+g)(x^{\ell+1}) \ge 0 \quad \forall\, k \in \mathbb{N}. \tag{28}$$

Moreover, pick any $x_* \in S_*$, then Proposition 4.1(ii) tells us that

$$\begin{aligned} 0 \ge (f+g)(x_*) - (f+g)(x^{\ell+1}) \ge& \frac{1}{2\alpha_\ell}\Big(\|x^{\ell+1} - x_*\|^2 - \|x^\ell - x_*\|^2 + (1-2\delta)\|x^\ell - x^{\ell+1}\|^2\Big) \\ \ge& \frac{1}{2\alpha_\ell}(\|x^{\ell+1} - x_*\|^2 - \|x^\ell - x_*\|^2) \end{aligned} \tag{29}$$

for any $\ell \in \mathbb{N}$. It follows that $\|x^{\ell+1} - x_*\|^2 - \|x^\ell - x_*\|^2 \le 0$. Since $\alpha_\ell \ge \alpha$, we get from the latter and (29) that

$$0 \ge (f+g)(x_*) - (f+g)(x^{\ell+1}) \ge \frac{1}{2\alpha}(\|x^{\ell+1} - x_*\|^2 - \|x^\ell - x_*\|^2).$$

Summing the above inequality over $\ell = 0, 1, \ldots, k-1$ implies that

$$k(f+g)(x_*) - \sum_{\ell=0}^{k-1}(f+g)(x^{\ell+1}) \ge \frac{1}{2\alpha}(\|x^k - x_*\|^2 - \|x^0 - x_*\|^2).$$

This together with (28) yields

$$(f+g)(x^k) - (f+g)(x_*) \le \frac{1}{2\alpha}\frac{\|x_* - x^0\|^2 - \|x^k - x_*\|^2}{k} \le \frac{1}{2\alpha}\frac{\|x_* - x^0\|^2}{k}. \tag{30}$$

Note further that no matter how we choose $x_* \in S$, the optimal value $(f+g)(x_*) = \min_{x\in\mathcal{H}}(f+g)(x)$ is fixed. Hence we get from (30) that

$$(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x) \le \inf_{y\in S_*} \frac{1}{2\alpha}\frac{\|y - x^0\|^2}{k} = \frac{1}{2\alpha}\frac{[\operatorname{dist}(x^0, S_*)]^2}{k},$$

which verifies (27) and completes the proof. ◻

We obtain linear convergence when the stepsizes are bounded below by a positive number and either f or g is strongly convex. Recall that $h : \mathcal{H} \to \overline{\mathbb{R}}$ is strongly convex with constant $\mu > 0$ if,

$$h(x) \ge h(y) + \langle v, x - y\rangle + \frac{\mu}{2}\|x - y\|^2 \quad \text{for all} \quad x \in \mathcal{H},\ (y, v) \in \operatorname{Gph}\partial h.$$

Theorem 4.4. *Let $(x^k)_{k\in\mathbb{N}}$ and $(\alpha_k)_{k\in\mathbb{N}}$ be the sequences generated in* **Method 1***. Suppose that $S_* \ne \emptyset$, that there exists $\alpha > 0$ satisfying $\alpha_k \ge \alpha > 0$ for all $k \in \mathbb{N}$, and that either f or g is strongly convex with constant $\mu > 0$. Then $S_* = \{x_*\}$ is singleton and*

$$\|x^{k+1} - x_*\| \le \frac{1}{\sqrt{1+\alpha\mu}} \cdot \|x^k - x_*\| \le \left(\frac{1}{\sqrt{1+\alpha\mu}}\right)^{k+1} \|x^0 - x_*\| \quad \forall\, k \in \mathbb{N}, \tag{31}$$

i.e., *the sequence $(x^k)_{k\in\mathbb{N}}$ converges to x_* with the linear rate* $\dfrac{1}{\sqrt{1+\alpha\mu}} < 1$.

Proof. Since either f or g is μ-strongly convex, we have $f+g$ is also μ-strongly convex. It follows that S_* is singleton (*i.e.*, $S_* = \{x_*\}$). Moreover, using Proposition 4.1(i) with $x = x_*$ and the strong convexity of $f+g$ gives us that

$$\begin{aligned}\|x^k - x_*\|^2 \geq & \|x^{k+1} - x_*\|^2 + 2\alpha_k[(f+g)(x^{k+1}) - (f+g)(x_*)] \\ \geq & \|x^{k+1} - x_*\|^2 + \alpha_k \mu \|x^{k+1} - x_*\|^2 \\ \geq & (1+\alpha\mu)\|x^{k+1} - x_*\|^2.\end{aligned}$$

It follows that

$$\|x^{k+1} - x_*\| \leq \frac{1}{\sqrt{1+\alpha\mu}} \cdot \|x^k - x_*\| \leq \left(\frac{1}{\sqrt{1+\alpha\mu}}\right)^{k+1} \|x^0 - x_*\|,$$

which verifies (31) and thus completes the proof of the theorem. □

Now we study the behavior of the stepsizes when the gradient of f satisfies some Lipschitz continuity. The first part of the following result is not much surprising due to the similar achievement in [29, Theorem 3.4(a)]. However, the second part is a huge improvement when we replace the global Lipschitz continuity by the local one in finite dimensions.

Proposition 4.5. *Let $(\alpha_k)_{k\in\mathbb{N}}$ be the sequence generated by* **Linesearch 1** *on* **Method 1**. *The following statements hold:*

(i) *If the gradient of f is globally Lipschitz continuous on $\operatorname{dom} g$ with constant $L > 0$, then $\alpha_k \geq \min\left\{\sigma, \frac{\delta\theta}{L}\right\}$ for all $k \in \mathbb{N}$.*

(ii) *Suppose that $\dim \mathcal{H} < +\infty$ and $S_* \neq \emptyset$. If ∇f is locally Lipschitz continuous at any $x \in S_*$ then there exists $x_* \in S_*$ such that*

$$\liminf_{k\to\infty} \alpha_k \geq \min\left\{\sigma, \frac{\delta\theta}{\mathcal{L}}\right\}, \tag{32}$$

where $\mathcal{L} > 0$ is a Lipschitz constant of ∇f around x_*. Consequently, there exists $\alpha > 0$ such that $\alpha_k \geq \alpha$ for all $k \in \mathbb{N}$.

Proof. To justify **(i)**, suppose that ∇f is globally Lipschitz continuous with constant $L > 0$. If $\alpha_k < \sigma$, define $\hat{\alpha}_k := \dfrac{\alpha_k}{\theta} > 0$ and $\hat{x}^k := J(x^k, \hat{\alpha}_k)$. It follows from the definition of **Linesearch 1** that

$$\hat{\alpha}_k \left\|\nabla f(\hat{x}^k) - \nabla f(x^k)\right\| > \delta \left\|\hat{x}^k - x^k\right\|, \tag{33}$$

which yields $\|\hat{x}^k - x^k\| \neq 0$ for all $k \in \mathbb{N}$. Moreover, due to Lipschitz assumption on ∇f, we get $\|\nabla f(x^k) - \nabla f(\hat{x}^k)\| \leq L\|x^k - \hat{x}^k\|$ for all $k \in \mathbb{N}$. Combining the latter inequality with (33) gives us that $\alpha_k L > \delta\theta$, i.e., $\alpha_k \geq \frac{\delta\theta}{L}$ when $\alpha_k < \sigma$. Thus we complete the first part of the proposition.

To verify the second part, we suppose that $\dim \mathcal{H} < +\infty$, that $S_* \neq \emptyset$, and that f is locally Lipschitz continuous at any $x_* \in S_*$. Similarly to the first part, we define $\hat{\alpha}_k := \frac{\alpha_k}{\theta} > 0$ and $\hat{x}^k := J(x^k, \hat{\alpha}_k)$. If $\liminf_{k\to\infty} \alpha_k < \sigma$, there exist $K \in \mathbb{N}$ such that $\alpha_k < \sigma$ for all $k > K$. Thus we also have (33). Thanks to Theorem 4.2, $(x^k)_{k\in\mathbb{N}}$ converges (strongly) to some $x_* \in S_*$. Moreover, it follows from Lemma 2.5 that

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

Thus the sequence $(\hat{x}^k)_{k\in\mathbb{N}}$ is also converging to $x_* \in S_*$. Since ∇f is Lipschitz continuous around x_* with some constant $\mathcal{L} > 0$, we may find $K_1 > K$ such that

$$\|\nabla f(x^k) - \nabla f(\hat{x}^k)\| \le \mathcal{L}\|x^k - \hat{x}^k\| \quad \text{for all} \quad k > K_1.$$

This together with (33) gives us that $\mathcal{L}\hat{\alpha}_k \ge \delta$, i.e., $\alpha_k \ge \frac{\delta\theta}{\mathcal{L}}$ for $k > K_1$. It follows that $\liminf_{k\to\infty} \alpha_k \ge \frac{\delta\theta}{\mathcal{L}}$ whenever $\liminf_{k\to\infty} \alpha_k < \sigma$, which clearly verifies (32).

Finally, since $\alpha_k > 0$ for $k \in \mathbb{N}$, we obtain that

$$\alpha_k \ge \alpha := \min\left\{\alpha_1, \ldots, \alpha_{K_1}, \frac{\delta\theta}{\mathcal{L}}, \sigma\right\} > 0$$

and complete the proof of the proposition. $\square$

It is worth recalling that the assumption of Proposition 4.5(i) that ∇f is globally Lipschitz continuous on $\operatorname{dom} g$ is also sufficient for **A2**. Assumptions of Proposition 4.5(ii) are certainly not enough to guarantee **A2**. However, there are many broad classes of functions satisfying both **A2** and those of Proposition 4.5(ii). For instance, when $\dim \mathcal{H} < +\infty$ and $\operatorname{dom} g$ is closed, a function f, which is differentiable and has gradient locally Lipschitz continuous on $\operatorname{dom} g$ satisfies all the requirements; see also Proposition 2.3.

Theorem 4.3 together with Proposition 4.5 and Theorems 4.4 leads us to the following result.

Corollary 4.6. *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated by* **Method 1**. *Suppose that $S_* \neq \emptyset$. If either*

(i) *the gradient of f is globally Lipschitz continuous on* $\operatorname{dom} g$*;*

or

(ii) $\dim \mathcal{H} < +\infty$ *and the gradient of f is locally Lipschitz continuous on S_*,*

then we have

$$(f+g)(x^k) - \min_{x\in\mathcal{H}} (f+g)(x) = \mathcal{O}(k^{-1}).$$

Furthermore, if either f or g is strongly convex then $(x^k)_{k\in\mathbb{N}}$ converges linearly to the unique optimal solution.

Since the condition $x = J(x, \alpha)$ for $\alpha > 0$ is necessary and sufficient for x to be an optimal solution to problem (1), it is interesting to study the complexity of $\|x^k - J(x^k, \alpha_k)\|$ in our **Method 1**. The velocity of the convergence obtained below is not affected by the behavior of the stepsizes α_k.

Theorem 4.7. *Let $(x^k)_{k\in\mathbb{N}}$ and $(\alpha_k)_{k\in\mathbb{N}}$ be the sequences generated from* **Method 1**. *Then we have*

$$\liminf_{k\to\infty} \sqrt{k+1} \cdot \|x^k - J(x^k, \alpha_k)\| = 0. \tag{34}$$

Proof. If (34) does not hold, then we may find a number $\varepsilon > 0$ such that for $\bar{k}$ large enough, we have $\|x^k - J(x^k, \alpha_k)\| \ge \dfrac{\varepsilon}{\sqrt{k+1}}$ for all $k \ge \bar{k}$. Thus,

$$\sum_{k=\bar{k}}^{\infty} \|x^k - J(x^k, \alpha_k)\|^2 \ge \varepsilon^2 \sum_{k=\bar{k}}^{\infty} \frac{1}{k+1} = +\infty. \tag{35}$$

On the other hand, using (11) and Lemma 4.1(i), we get, for all $k \geq \bar{k}$,

$$\|x^k - J(x^k, \alpha_k)\|^2 = \|x^k - x^{k+1}\|^2 \leq \frac{\alpha_k}{1-\delta}\left[(f+g)(x^k) - (f+g)(x^{k+1})\right]$$
$$\leq \frac{\sigma}{1-\delta}\left[(f+g)(x^k) - (f+g)(x^{k+1})\right],$$

where we have used in the last inequality the fact that $\alpha_k \leq \sigma$ for all $k \in \mathbb{N}$, which is direct consequence of **Linesearch 1**. Thus,

$$\sum_{k=\bar{k}}^{\infty} \|x^k - J(x^k, \alpha_k)\|^2 \leq \frac{\sigma}{1-\delta}\left[(f+g)(x^{\bar{k}}) - (f+g)(x_*)\right] < +\infty,$$

which contradicts (35). The proof is complete. □

4.2 A fast multistep proximal forward-backward method with Linesearch 1

In the spirit of the classical work of Nesterov [24] many accelerated multistep versions have been proposed in the literature for the proximal forward-backward iteration, but to the best of our knowledge all of them have to employ the global Lipschitz continuity assumption on ∇f; see, *e.g.*, [2, 3, 22]. In this subsection, by following these ideas and assuming no Lipschitz continuity on ∇f, we present a fast version of the proximal forward-backward method with **Linesearch 1**, improving the convergence result of Theorem 4.3 for **Method 1**. In [2,3,22] this kind of fast versions usually demands full domain of the objective function ($\operatorname{dom} g = \mathcal{H}$) and Lipschitz assumption over ∇f to establish convergence of this method. Here we modify the method by adding a linesearch and an extra projection step in (37) below to avoid the requirements aforementioned. For simplicity, we suppose $\Omega := \operatorname{dom} g$ is closed in this section.

Method 2.

Initialization Step. *Take* $x^{-1} = x^0 \in \operatorname{dom} g$, $t_0 = 1$, $\theta \in (0,1)$ *and* $\delta \in (0, 1/2)$.

Iterative Step. *Given* t_k *and* x^k, *set*

$$t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}. \tag{36}$$

$$y^k = x^k + \left(\frac{t_k - 1}{t_{k+1}}\right)(x^k - x^{k-1}), \quad \tilde{y}^k = P_\Omega(y^k). \tag{37}$$

$$x^{k+1} = J(\tilde{y}^k, \alpha_k) := \operatorname{prox}_{\alpha_k g}(\tilde{y}^k - \alpha_k \nabla f(\tilde{y}^k)), \tag{38}$$

where $\alpha_k :=$ **Linesearch 1**$(\tilde{y}^k, \sigma, \theta, \delta)$.

Stop Criteria. *If* $x^{k+1} = \tilde{y}^k$, *then stop.*

Note that from (37) and (38), $\tilde{y}^k$ and x^k belong to $\operatorname{dom} g$ for all $k \in \mathbb{N}$ and as a direct consequence of Lemma 3.1, α_k satisfying (39) is always positive. Moreover, it is similar to **Method 1** that if $x^{k+1} = \tilde{y}^k$ then x^{k+1} is an optimal solution. To avoid finite termination of this method, we suppose from now on that $x^{k+1} \neq \tilde{y}^k$ for any $k \in \mathbb{N}$. An important inequality for our further study from **Linesearch 1** is

$$\alpha_k \left\|\nabla f(x^{k+1}) - \nabla f(\tilde{y}^k)\right\| \leq \delta \left\|x^{k+1} - \tilde{y}^k\right\| \tag{39}$$

with $\delta \in (0, 1/2)$. We also need some auxiliary results before establishing the convergence results.

Lemma 4.8. *The positive sequence $(t_k)_{k\in\mathbb{N}}$ generated by* **Method 2** *via* (36) *satisfies, for all $k \in \mathbb{N}$,*

(i) $\frac{1}{t_k} \leq \frac{2}{k+1}$;

(ii) $t_{k+1}^2 - t_{k+1} = t_k^2$.

Proof. The proof easily follows by induction argument. □

Proposition 4.9. *Let α_k be defined in* **Method 2** *and $x \in \operatorname{dom} g$. Then we have*

$$(f+g)(x) - (f+g)(x^{k+1}) \geq \frac{1}{2\alpha_k}\left(\|x^{k+1} - x\|^2 - \|y^k - x\|^2\right) \quad \textit{for all} \quad k \in \mathbb{N}. \tag{40}$$

Proof. First note from (5) with $z = \tilde{y}^k - \alpha_k \nabla f(\tilde{y}^k)$ that $\frac{\tilde{y}^k - x^{k+1}}{\alpha_k} - \nabla f(\tilde{y}^k) \in \partial g(x^{k+1})$. Then,

$$g(x) - g(x^{k+1}) \geq \left\langle \frac{\tilde{y}^k - x^{k+1}}{\alpha_k} - \nabla f(\tilde{y}^k), x - x^{k+1} \right\rangle \tag{41}$$

for all $x \in \operatorname{dom} g$. The convexity of f implies that

$$f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle \quad \text{for all} \quad x, y \in \operatorname{dom} f. \tag{42}$$

By summing (41) and (42) with $y = \tilde{y}^k$, we obtain that

$$\begin{aligned}
(f+g)(x) \geq& f(\tilde{y}^k) + g(x^{k+1}) + \left\langle \frac{\tilde{y}^k - x^{k+1}}{\alpha_k} - \nabla f(\tilde{y}^k), x - x^{k+1} \right\rangle + \langle \nabla f(\tilde{y}^k), x - \tilde{y}^k \rangle \\
=& f(\tilde{y}^k) + g(x^{k+1}) + \frac{1}{\alpha_k}\left\langle \tilde{y}^k - x^{k+1}, x - x^{k+1} \right\rangle + \langle \nabla f(\tilde{y}^k), x^{k+1} - \tilde{y}^k \rangle \\
=& f(\tilde{y}^k) + g(x^{k+1}) + \frac{1}{\alpha_k}\left\langle \tilde{y}^k - x^{k+1}, x - x^{k+1} \right\rangle + \langle \nabla f(\tilde{y}^k) - \nabla f(x^{k+1}), x^{k+1} - \tilde{y}^k \rangle \\
&+ \langle \nabla f(x^{k+1}), x^{k+1} - \tilde{y}^k \rangle \\
\geq& f(\tilde{y}^k) + g(x^{k+1}) + \frac{1}{\alpha_k}\left\langle \tilde{y}^k - x^{k+1}, x - x^{k+1} \right\rangle - \frac{\delta}{\alpha_k}\|x^{k+1} - \tilde{y}^k\|^2 + \langle \nabla f(x^{k+1}), x^{k+1} - \tilde{y}^k \rangle.
\end{aligned}$$

Rearranging the inequality gives us that

$$\begin{aligned}
\langle \tilde{y}^k - x^{k+1}, x^{k+1} - x \rangle \geq& \alpha_k[f(\tilde{y}^k) + g(x^{k+1}) - (f+g)(x)] - \delta\|x^{k+1} - \tilde{y}^k\|^2 \\
&+ \alpha_k \langle \nabla f(x^{k+1}), x^{k+1} - \tilde{y}^k \rangle.
\end{aligned} \tag{43}$$

Observe that

$$2\langle \tilde{y}^k - x^{k+1}, x^{k+1} - x \rangle = \|\tilde{y}^k - x\|^2 - \|x^{k+1} - x\|^2 - \|\tilde{y}^k - x^{k+1}\|^2.$$

By combining the above equality with (43), we have

$$\begin{aligned}
\|\tilde{y}^k - x\|^2 - \|x^{k+1} - x\|^2 \geq& 2\alpha_k\left[f(\tilde{y}^k) + g(x^{k+1}) - (f+g)(x) + \langle \nabla f(x^{k+1}), x^{k+1} - \tilde{y}^k \rangle\right] \\
&+ (1 - 2\delta)\|\tilde{y}^k - x^{k+1}\|^2 \\
\geq& 2\alpha_k\left[f(\tilde{y}^k) + g(x^{k+1}) - (f+g)(x) + \langle \nabla f(x^{k+1}), x^{k+1} - \tilde{y}^k \rangle\right].
\end{aligned} \tag{44}$$

It follows from (42) with $x = \tilde{y}^k$ and $y = x^{k+1}$ that $f(\tilde{y}^k) - f(x^{k+1}) \geq \langle \nabla f(x^{k+1}), \tilde{y}^k - x^{k+1} \rangle$, which together with (44) implies

$$\begin{aligned}\|\tilde{y}^k - x\|^2 - \|x^{k+1} - x\|^2 \geq & 2\alpha_k \left[f(\tilde{y}^k) + g(x^{k+1}) - (f+g)(x) + f(x^{k+1}) - f(\tilde{y}^k) \right] \\ = & 2\alpha_k \left[(f+g)(x^{k+1}) - (f+g)(x) \right].\end{aligned}$$

Since $\|\tilde{y}^k - x\| \leq \|y^k - x\|$ for all $x \in \operatorname{dom} g$ due to (37), we get from the latter (40) and complete the proof of the proposition. □

In the next result we establish a better complexity for **Method 2** than Theorem 4.3 under the same assumption set.

Theorem 4.10. *Let $(x^k)_{k\in\mathbb{N}}$ and $(\alpha_k)_{k\in\mathbb{N}}$ be the sequences generated in* **Method 2**. *Suppose that $S_* \neq \emptyset$ and there is $\alpha > 0$ such that $\alpha_k \geq \alpha > 0$ for all $k \in \mathbb{N}$. Then we have*

$$(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x) \leq \frac{(2/\alpha)[\operatorname{dist}(x^0, S_*)]^2 + 4\left[(f+g)(x^0) - \min_{x\in\mathcal{H}}(f+g)(x)\right]}{(k+1)^2} \quad \text{for all} \quad k \in \mathbb{N}.$$

Proof. To justify, pick any $x_* \in S_*$. By Lemma 4.8(i) we have $x = t_{k+1}^{-1} x_* + (1 - t_{k+1}^{-1}) x^k \in \operatorname{dom} g$. Applying Proposition 4.9 for this x gives us that

$$\begin{aligned}&\frac{1}{2\alpha_k}\left(\left\| x^{k+1} - \left(t_{k+1}^{-1} x_* + \left(1 - t_{k+1}^{-1}\right) x^k\right) \right\|^2 - \left\| y^k - \left(t_{k+1}^{-1} x_* + \left(1 - t_{k+1}^{-1}\right) x^k\right) \right\|^2 \right) \\ &\leq (f+g)(t_{k+1}^{-1} x_* + \left(1 - t_{k+1}^{-1}\right) x^k) - (f+g)(x^{k+1}) \\ &\leq t_{k+1}^{-1}(f+g)(x_*) + (1 - t_{k+1}^{-1})(f+g)(x^k) - (f+g)(x^{k+1}).\end{aligned}$$

After rearrangement, we obtain

$$\begin{aligned}&(1 - t_{k+1}^{-1}) \left[(f+g)(x^k) - (f+g)(x_*) \right] - \left[(f+g)(x^{k+1}) - (f+g)(x_*) \right] \\ &\geq \frac{1}{2\alpha_k t_{k+1}^2} \left(\left\| t_{k+1} x^{k+1} - (x_* + (t_{k+1} - 1) x^k) \right\|^2 - \left\| t_{k+1} y^k - (x_* + (t_{k+1} - 1) x^k) \right\|^2 \right).\end{aligned}$$

By multiplying by t_{k+1}^2 to the above inequality and using (37) and Lemma 4.8(ii), we have

$$\begin{aligned}&\frac{1}{2\alpha_k}\left(\|t_{k+1} x^{k+1} - (x_* + (t_{k+1} - 1)x^k)\|^2 - \|t_{k+1} y^k - (x_* + (t_{k+1} - 1)x^k)\|^2 \right) \\ &= \frac{1}{2\alpha_k}\left(\|t_{k+1} x^{k+1} - (t_{k+1} - 1)x^k - x_*\|^2 - \|t_k x^k - (t_k - 1)x^{k-1} - x_*\|^2 \right) \\ &\leq (t_{k+1}^2 - t_{k+1}) \left[(f+g)(x^k) - (f+g)(x_*) \right] - t_{k+1}^2 \left[(f+g)(x^{k+1}) - (f+g)(x_*) \right] \\ &= t_k^2 \left[(f+g)(x^k) - (f+g)(x_*) \right] - t_{k+1}^2 \left[(f+g)(x^{k+1}) - (f+g)(x_*) \right].\end{aligned}$$

It follows that

$$\begin{aligned}&\|t_k x^k - (t_k - 1)x^{k-1} - x_*\|^2 - \|t_{k+1} x^{k+1} - (t_{k+1} - 1)x^k - x_*\|^2 \\ &\geq 2\alpha_k \left(t_{k+1}^2 \left[(f+g)(x^{k+1}) - (f+g)(x_*) \right] - t_k^2 \left[(f+g)(x^k) - (f+g)(x_*) \right] \right) \\ &\geq 2\alpha t_{k+1}^2 \left[(f+g)(x^{k+1}) - (f+g)(x_*) \right] - 2\alpha t_k^2 \left[(f+g)(x^k) - (f+g)(x_*) \right].\end{aligned}$$

Reordering the above inequality and applying it inductively yield

$$\begin{aligned}
&2\alpha t_{k+1}^2 \left[(f+g)(x^{k+1}) - (f+g)(x_*)\right] \\
&\leq \|t_{k+1}x^{k+1} - (t_{k+1}-1)x^k - x_*\|^2 + 2\alpha t_{k+1}^2 \left[(f+g)(x^{k+1}) - (f+g)(x_*)\right] \\
&\leq \|t_k x^k - (t_k-1)x^{k-1} - x_*\|^2 + 2\alpha t_k^2 \left[(f+g)(x^k) - (f+g)(x_*)\right] \\
&\leq \ldots \leq \|t_0 x^0 - (t_0-1)x^{-1} - x_*\|^2 + 2\alpha t_0^2 \left[(f+g)(x^0) - (f+g)(x_*)\right] \\
&= \|x^0 - x_*\|^2 + 2\alpha \left[(f+g)(x^0) - (f+g)(x_*)\right],
\end{aligned}$$

which readily implies

$$2\alpha t_k^2[(f+g)(x^k) - (f+g)(x_*)] \leq \|x^0 - x_*\|^2 + 2\alpha \left[(f+g)(x^0) - (f+g)(x_*)\right].$$

Using this inequality together with Lemma 4.8(i) gives us that

$$\begin{aligned}
(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x) \leq& \frac{1}{2\alpha t_k^2}\left(\|x^0 - x_*\|^2 + 2\alpha\left[(f+g)(x^0) - \min_{x\in\mathcal{H}}(f+g)(x)\right]\right) \\
\leq& \frac{(2/\alpha)\|x^0 - x_*\|^2 + 4\left[(f+g)(x^0) - \min_{x\in\mathcal{H}}(f+g)(x)\right]}{(k+1)^2}
\end{aligned}$$

and thus completes the proof of this theorem. □

This theorem shows that the expected error of the iterates generated by **Method 2** after k iterations is $\mathcal{O}((k+1)^{-2})$ when the stepsizes are bounded below by a positive constant. Similarly to Proposition 4.5, we prove in the next result that such a requirement is satisfied under global Lipschitz assumption on the gradient of f. Unfortunately, obtaining the similar conclusion under the local Lipschitz one like Proposition 4.5(ii) is out of our reach in the current work.

Proposition 4.11. *Let $(\alpha_k)_{k\in\mathbb{N}}$ be the sequence generated by* **Linesearch 1** *on* **Method 2**. *If the gradient of f is globally Lipschitz continuous on $\operatorname{dom} g$ with constant $L > 0$ then $\alpha_k \geq \min\left\{\sigma, \frac{\delta\theta}{L}\right\}$ for all $k \in \mathbb{N}$.*

Proof. Define $\hat{\alpha}_k := \dfrac{\alpha_k}{\theta} > 0$, and $\hat{y}^k := J(\tilde{y}^k, \hat{\alpha}_k) = \operatorname{prox}_{\hat{\alpha}_k g}(\tilde{y}^k - \hat{\alpha}_k \nabla f(\tilde{y}^k)) \in \operatorname{dom} g$. If $\alpha_k < \sigma$, it follows from the definition of **Linesearch 1** that

$$\hat{\alpha}_k \left\|\nabla f(\hat{y}^k) - \nabla f(\tilde{y}^k)\right\| > \delta \left\|\hat{y}^k - \tilde{y}^k\right\|. \tag{45}$$

Due to the fact ∇f is Lipschitz continuous on $\operatorname{dom} g$ with constant L, we get from (45) gives us that $\alpha_k L\|\tilde{y}^k - \hat{y}^k\| > \delta\theta\|\tilde{y}^k - \hat{y}^k\|$. Thus $\alpha_k \geq \frac{\delta\theta}{L}$ whenever $\alpha_k < \sigma$, which verifies the desired inequality. □

Let us complete the section with a direct consequence of the above proposition and Theorem 4.10.

Corollary 4.12. *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated by* **Method 2**. *Suppose that $S_* \neq \emptyset$ and the gradient of f is Lipschitz continuous on $\operatorname{dom} g$ with constant $L > 0$. Then we have*

$$(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x) \leq \frac{2\max\left\{\dfrac{1}{\sigma}, \dfrac{L}{\theta\delta}\right\}[\operatorname{dist}(x^0, S_*)]^2 + 4\left[(f+g)(x^0) - \min_{x\in\mathcal{H}}(f+g)(x)\right]}{(k+1)^2},$$

for all $k \in \mathbb{N}$.

5 The proximal forward-backward method with Linesearch 2

Method 1 requires to evaluate the resolvent of ∂g inside **Linesearch 1** at each step of the iteration. When the proximal step is not easy to compute, **Method 1** may be inefficient. To overcome this drawback, we propose here a modification of the proximal forward-backward method by using **Linesearch 2**, which involves only one computation of the resolvent of ∂g for all steps of this linesearch. We also prove that the sequence generated by this method is weakly convergent to a solution of problem (1).

Method 3.

Initialization Step. *Take* $x^0 \in \operatorname{dom} g$ *and* $\theta \in (0,1)$.

Iterative Step. *Set*

$$J_k = \operatorname{prox}_g(x^k - \nabla f(x^k)), \tag{46}$$

$$x^{k+1} = x^k - \beta_k(x^k - J_k) \tag{47}$$

where $\beta_k :=$ **Linesearch 2**(x^k, θ).

Stop Criteria. *If* $x^{k+1} = x^k$*, then stop.*

Thanks to Lemma 3.2 and the convexity of g, we note that $x^k \in \operatorname{dom} g$ inductively. Moreover, it follows from **Linesearch 2** that

$$(f+g)(x^{k+1}) \le (f+g)(x^k) - \beta_k\Big[g(x^k) - g(J_k)\Big] - \beta_k\langle \nabla f(x^k), x^k - J_k\rangle + \frac{\beta_k}{2}\|x^k - J_k\|^2. \tag{48}$$

Next we obtain some similar results for **Method 3** to the ones in Section 3 for **Method 1**. The following proposition is corresponding to Proposition 4.1.

Proposition 5.1. *Let* $x \in \operatorname{dom} g$*. Then we have*

$$\|x^{k+1}-x\|^2 \le \|x^k-x\|^2+2\Big[(f+g)(x^k) - (f+g)(x^{k+1})\Big]+2\beta_k\Big[(f+g)(x) - (f+g)(x^k)\Big], \quad \forall\, k \in \mathbb{N}.$$

Proof. Fix any $x \in \operatorname{dom} g$ and set $A_k := \|x^{k+1}-x^k\|^2+\|x^k-x\|^2-\|x^{k+1}-x\|^2 = 2\langle x^k-x^{k+1}, x^k-x\rangle$. Moreover, we get from (47) and (46) that

$$\begin{aligned}\frac{A_k}{2\beta_k} &= \langle x^k - J_k, x^k - x\rangle = \langle \nabla f(x^k), x^k - x\rangle + \langle x^k - J_k - \nabla f(x^k), x^k - x\rangle\\ &= \langle \nabla f(x^k), x^k - x\rangle + \langle x^k - J_k - \nabla f(x^k), J_k - x\rangle + \langle x^k - J_k - \nabla f(x^k), x^k - J_k\rangle\\ &= \langle \nabla f(x^k), x^k - x\rangle + \langle x^k - J_k - \nabla f(x^k), J_k - x\rangle - \langle \nabla f(x^k), x^k - J_k\rangle + \|x^k - J_k\|^2.\end{aligned}$$

Observe from (46) that $x^k - \nabla f(x^k) - J_k \in \partial g(J_k)$. By applying (4) and (48) to the above expression, we have

$$\begin{aligned}\frac{A_k}{2\beta_k} &\ge f(x^k) - f(x) + g(J_k) - g(x) - \langle \nabla f(x^k), x^k - J_k\rangle + \|x^k - J_k\|^2\\ &\ge f(x^k) + g(J_k) - (f+g)(x) + \frac{1}{\beta_k}\Big[(f+g)(x^{k+1}) - (f+g)(x^k)\Big] + g(x^k) - g(J_k) + \frac{1}{2}\|x^k - J_k\|^2\\ &= \Big[(f+g)(x^k) - (f+g)(x)\Big] + \frac{1}{\beta_k}\Big[(f+g)(x^{k+1}) - (f+g)(x^k)\Big] + \frac{1}{2}\|x^k - J_k\|^2.\end{aligned}$$

It follows that

$$\|x^{k+1}-x\|^2 \leq \|x^k-x\|^2 + \|x^{k+1}-x^k\|^2 - \beta_k\|x^k-J_k\|^2 + 2\left[(f+g)(x^k)-(f+g)(x^{k+1})\right] + 2\beta_k\left[(f+g)(x)-(f+g)(x^k)\right].$$

Since $x^{k+1}-x^k=\beta_k(J_k-x^k)$ by (47) and $\beta_k^2 \leq \beta_k$, we conclude that

$$\begin{aligned}\|x^{k+1}-x\|^2 \leq& \|x^k-x\|^2 + (\beta_k^2-\beta_k)\|x^k-J_k\|^2 + 2\left[(f+g)(x^k)-(f+g)(x^{k+1})\right] \\ &+ 2\beta_k\left[(f+g)(x)-(f+g)(x^k)\right] \\ \leq& \|x^k-x\|^2 + 2\left[(f+g)(x^k)-(f+g)(x^{k+1})\right] + 2\beta_k\left[(f+g)(x)-(f+g)(x^k)\right]\end{aligned}$$

as desired. The proof is complete. □

It is worth noting that using Proposition 5.1 with $x = x^k \in \operatorname{dom} g$ gives us that

$$(f+g)(x^k)-(f+g)(x^{k+1}) \geq \frac{1}{2}\|x^{k+1}-x^k\|^2 \geq 0, \tag{49}$$

which shows that **Method 3** is also a descent method in which the cost function is decreasing after each iteration. Next we establish the main result of this section whose statement is similar to Theorem 4.2.

Theorem 5.2. *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated by* **Method 3**. *The following statements hold:*

(i) *If $S_* \neq \emptyset$ then the sequence $(x^k)_{k\in\mathbb{N}}$ is quasi-Fejér convergent to S_* and weakly converges to a point in S_*.*

(ii) *If $S_* = \emptyset$ then we have*

$$\lim_{k\to\infty}\|x^k\| = +\infty \quad \text{and} \quad \lim_{k\to\infty}(f+g)(x^k) = \inf_{x\in\mathcal{H}}(f+g)(x). \tag{50}$$

Proof. To justify **(i)**, suppose that $S_* \neq \emptyset$. By employing Proposition 5.1 at $x = x_* \in S_* \subseteq \operatorname{dom} g$, we have

$$\|x^{k+1}-x_*\|^2 \leq \|x^k-x_*\|^2 + 2\left[(f+g)(x^k)-(f+g)(x^{k+1})\right] \quad \text{for all} \quad k\in\mathbb{N}. \tag{51}$$

It follows from (49) that $\epsilon_k := 2\left[(f+g)(x^k)-(f+g)(x^{k+1})\right] \geq 0$. Moreover, observe that

$$\begin{aligned}\sum_{k=0}^{\infty}\epsilon_k =& 2\sum_{k=0}^{\infty}\left[(f+g)(x^k)-(f+g)(x^{k+1})\right] \leq 2\left[(f+g)(x^0)-\lim_{k\to\infty}(f+g)(x^{k+1})\right] \\ \leq& 2\left[(f+g)(x^0)-(f+g)(x_*)\right] < +\infty.\end{aligned}$$

This together with (51) tells us that the sequence $(x^k)_{k\in\mathbb{N}}$ is quasi-Fejér convergent to S_* via Definition 2.1. By Fact 2.6(i), this sequence is bounded and hence it has weak accumulation points. Let $\bar{x}$ be a weak accumulation point of $(x^k)_{k\in\mathbb{N}}$. Hence there exists a subsequence $(x^{n_k})_{k\in\mathbb{N}}$ of $(x^k)_{k\in\mathbb{N}}$ converging weakly to $\bar{x}$. Now we distinguish our analysis into two cases.

Case 1. The sequence $(\beta_{n_k})_{k\in\mathbb{N}}$ does not converge to 0, *i.e.*, there exist some $\beta > 0$ and a subsequence of $(\beta_{n_k})_{k\in\mathbb{N}}$ (without relabeling) such that

$$\beta_{n_k} \geq \beta, \quad \forall\, k \in \mathbb{N}. \tag{52}$$

By using Proposition 5.1 with $x = x_* \in S_*$, we get

$$\beta_k \left[(f+g)(x^k) - (f+g)(x_*)\right] \leq \frac{1}{2}(\|x^k - x_*\|^2 - \|x^{k+1} - x_*\|^2) + (f+g)(x^k) - (f+g)(x^{k+1}).$$

Summing from $k = 0$ to m in the above inequality implies

$$\begin{aligned}\sum_{k=0}^{m} \beta_k \left[(f+g)(x^k) - (f+g)(x_*)\right] &\leq \frac{1}{2}(\|x^0 - x_*\|^2 - \|x^{m+1} - x_*\|^2) + (f+g)(x^0) - (f+g)(x^{m+1}) \\ &\leq \frac{1}{2}\|x^0 - x_*\|^2 + (f+g)(x^0) - (f+g)(x_*).\end{aligned}$$

By taking $m \to \infty$ and using the fact that $(f+g)(x^k) \geq (f+g)(x_*)$, we obtain that

$$\sum_{k=0}^{\infty} \beta_{n_k} \left[(f+g)(x^{n_k}) - (f+g)(x_*)\right] \leq \sum_{k=0}^{\infty} \beta_k \left[(f+g)(x^k) - (f+g)(x_*)\right] < \infty,$$

which together with (52) establishes that

$$\lim_{k\to\infty} \left[(f+g)(x^{n_k}) - (f+g)(x_*)\right] = 0. \tag{53}$$

Since $f+g$ is lower semicontinuous on $\operatorname{dom} g$, it is also weakly l.s.c. due to the convexity of $f+g$. It follows from (53) that

$$(f+g)(x_*) \leq (f+g)(\bar{x}) \leq \liminf_{k\to\infty} (f+g)(x^{n_k}) = \lim_{k\to\infty} (f+g)(x^{n_k}) = (f+g)(x_*),$$

which yields $(f+g)(\bar{x}) = (f+g)(x_*)$ and thus $\bar{x} \in S_*$.

Case 2. $\lim_{k\to\infty} \beta_k = 0$. Define $\hat{\beta}_k := \frac{\beta_k}{\theta} > 0$ and

$$\hat{y}^k := x^k - \hat{\beta}_k(x^k - J_k) = (1 - \hat{\beta}_k)x^k + \hat{\beta}_k J_k. \tag{54}$$

It follows from the definition of **Linesearch 2** that

$$(f+g)(\hat{y}^k) > (f+g)(x^k) - \hat{\beta}_k[g(x^k) - g(J_k)] - \hat{\beta}_k\langle \nabla f(x^k), x^k - J_k\rangle + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2.$$

This together with (4) and (54) gives us that

$$\begin{aligned} 0 >& -\hat{\beta}_k\langle \nabla f(x^k), x^k - J_k\rangle + (f+g)(x^k) - (f+g)(\hat{y}^k) - \hat{\beta}_k[g(x^k) - g(J_k)] + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2 \\ =& -\hat{\beta}_k\langle \nabla f(x^k), x^k - J_k\rangle + f(x^k) - f(\hat{y}^k) + g(x^k) - g(\hat{y}^k) - \hat{\beta}_k[g(x^k) - g(J_k)] + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2 \\ \geq& -\hat{\beta}_k\langle \nabla f(x^k), x^k - J_k\rangle + \langle \nabla f(\hat{y}^k), x^k - \hat{y}^k\rangle + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2 \\ &+ g(x^k) - (1 - \hat{\beta}_k)g(x^k) - \hat{\beta}_k g(J_k) - \hat{\beta}_k[g(x^k) - g(J_k)] \\ =& \,\hat{\beta}_k\langle \nabla f(\hat{y}^k) - \nabla f(x^k), x^k - J_k\rangle + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2. \end{aligned}$$

We obtain from the latter that

$$\frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2 \le \hat{\beta}_k \|\nabla f(\hat{y}^k) - \nabla f(x^k)\| \cdot \|x^k - J_k\|,$$

which yields

$$\frac{1}{2}\|x^k - J_k\| \le \|\nabla f(\hat{y}^k) - \nabla f(x^k)\|. \tag{55}$$

Since $\text{prox}_g(\cdot)$ is Lipschitz continuous with constant 1, we get from (46) that

$$\|J_k - J_0\| \le \|x^k - x^0\| + \|\nabla f(x^k) - \nabla f(x^0)\|.$$

Due to Assumption **A2** and the boundedness of $(x^k)_{k\in\mathbb{N}}$, the latter tells us that $(J_k)_{k\in\mathbb{N}}$ is also bounded. This together with (54) and the fact $\beta_k \to 0$ implies that $\|\hat{y}^k - x^k\| \to 0$ as $k \to \infty$. Since ∇f is uniformly continuous on bounded sets, we get $\|\nabla f(\hat{y}^k) - \nabla f(x^k)\| \to 0$ as $k \to \infty$ and derive from (55) that

$$\lim_{k\to\infty} \|x^k - J_k\| = 0, \tag{56}$$

Since ∇f is uniformly continuous on bounded sets, (56) implies

$$\lim_{k\to\infty} \|\nabla f(x^k) - \nabla f(J_k)\| = 0. \tag{57}$$

Using (5) with $z = x^k - \nabla f(x^k)$ and Fact 2.4 gives us that

$$x^k - J_k + \nabla f(J_k) - \nabla f(x^k) \in \nabla f(J_k) + \partial g(J_k) = \partial(f+g)(J_k).$$

By passing to the limit over the subsequence $(n_k)_{k\in\mathbb{N}}$ in the above inclusion, we get from Fact 2.2, (56), and (57) that $0 \in \partial(f+g)(\bar{x})$, which implies $\bar{x} \in S_*$.

In all possible cases above, any the weak accumulation point of $(x^k)_{k\in\mathbb{N}}$ belongs to S_*. Fact 2.6(ii) tells us that $(x^k)_{k\in\mathbb{N}}$ converges weakly to an optimal solution in S_* and thus completely verifies **(i)**. Moreover, the proof of part **(ii)** is quite similar to the arguments used to prove Theorem 4.2(ii). We omit the detail and complete the proof. □

From the view of (50) and also our Theorem 4.2, it is natural to question that whether

$$\lim_{k\to\infty} (f+g)(x^k) = \min_{x\in\mathcal{H}} (f+g)(x) \tag{58}$$

in the case $S_* \ne \emptyset$. We do not know the answer in general, but when the sequence $(\beta_k)_{k\in\mathbb{N}}$ is bounded below by a positive constant, the equality (58) is true with some further complexity discussed in the next section.

5.1 Complexity analysis of Method 3

In this section we are able to establish the complexity of **Method 3** with a similar rate to Theorem 4.3 as follows.

Theorem 5.3. *Let $(x^k)_{k\in\mathbb{N}}$ and $(\beta_k)_{k\in\mathbb{N}}$ be the sequences generated in* **Method 3**. *Suppose that $S_* \ne \emptyset$ and there is some $\beta > 0$ satisfying $\beta_k \ge \beta > 0$ for all $k \in \mathbb{N}$. Then we have*

$$(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x) \le \frac{1}{2\beta} \frac{[\text{dist}(x^0, S_*)]^2 + 2\left[(f+g)(x^0) - \min_{x\in\mathcal{H}}(f+g)(x)\right]}{k} \quad \textit{for all} \quad k \in \mathbb{N}.$$

Proof. Thanks to (49), we get

$$0 \le \ell \Big[(f+g)(x^\ell) - (f+g)(x^{\ell+1})\Big] = \ell(f+g)(x^\ell) - (\ell+1)(f+g)(x^{\ell+1}) + (f+g)(x^{\ell+1}),$$

for all $\ell \in \mathbb{N}$. Summing the above inequality over $\ell = 0, 1, \ldots, k-1$, we have

$$-k(f+g)(x^k) + \sum_{\ell=0}^{k-1}(f+g)(x^{\ell+1}) \ge 0. \tag{59}$$

By using Proposition 5.1, at $\ell \in \mathbb{N}$ and $x_* \in S_*$, we get

$$\begin{aligned} 0 \ge &(f+g)(x_*) - (f+g)(x^{\ell+1}) \ge \frac{1}{2\beta_\ell}\left(\|x^{\ell+1} - x_*\|^2 - \|x^\ell - x_*\|^2 + 2\Big[(f+g)(x^{\ell+1}) - (f+g)(x^\ell)\Big]\right) \\ \ge &\frac{1}{2\beta}\left(\|x^{\ell+1} - x_*\|^2 - \|x^\ell - x_*\|^2 + 2\Big[(f+g)(x^{\ell+1}) - (f+g)(x^\ell)\Big]\right), \end{aligned} \tag{60}$$

for all $\ell \in \mathbb{N}$. Summing the above inequality (60), over $\ell = 0, 1, \ldots, k-1$, we have

$$\begin{aligned} &k(f+g)(x_*) - \sum_{\ell=0}^{k-1}(f+g)(x^{\ell+1}) \ge \frac{1}{2\beta}\left(\|x^k - x_*\|^2 - \|x^0 - x_*\|^2 + 2[(f+g)(x^k) - (f+g)(x^0)]\right) \\ &\ge \frac{1}{2\beta}\left(\|x^k - x_*\|^2 - \|x^0 - x_*\|^2 + 2\left[(f+g)(x_*) - (f+g)(x^0)\right]\right). \end{aligned} \tag{61}$$

Adding (59) and (61), we obtain

$$(f+g)(x^k) - (f+g)(x_*) \le \frac{1}{2\beta}\,\frac{\|x_* - x^0\|^2 - \|x^k - x_*\|^2 + 2\left[(f+g)(x^0) - (f+g)(x_*)\right]}{k}$$

establishing the result. □

Similarly to Lemma 4.5, we present next some sufficient conditions for the below boundedness by a positive constant of the stepsize generated by **Linesearch 2**.

Proposition 5.4. *Let* $(\beta_k)_{k\in\mathbb{N}}$ *be the sequence generated by* **Linesearch 2** *on* **Method 3**. *The following statements hold:*

(i) *If the gradient of* f *is globally Lipschitz continuous on* $\operatorname{dom} g$ *with constant* $L > 0$, *then* $\beta_k \ge \min\left\{1, \frac{\theta}{2L}\right\}$ *for all* $k \in \mathbb{N}$.

(ii) *Suppose that* $\dim \mathcal{H} < +\infty$ *and* $S_* \ne \emptyset$. *If* ∇f *is locally Lipschitz continuous at any* $x \in S_*$ *then there exists* $x_* \in S_*$ *such that*

$$\liminf_{k\to\infty} \beta_k \ge \min\left\{1, \frac{\theta}{2\mathcal{L}}\right\}, \tag{62}$$

where $\mathcal{L} > 0$ *is a Lipschitz constant of* ∇f *around* x_*. *Consequently, there exists* $\beta > 0$ *such that* $\beta_k \ge \beta$ *for all* $k \in \mathbb{N}$.

Proof. First let us verify **(i)** by supposing that the gradient of f is globally Lipschitz continuous on $\operatorname{dom} g$ with constant $L > 0$. Define $\hat\beta_k := \dfrac{\beta_k}{\theta} > 0$ and

$$\hat y^k := \hat\beta_k J_k + (1 - \hat\beta_k)x^k = x^k - \hat\beta_k(x^k - J_k). \tag{63}$$

If $\beta_k < 1$, we get from **Linesearch 2** that

$$(f+g)(\hat{y}^k) > (f+g)(x^k) - \hat{\beta}_k[g(x^k) - g(J_k)] - \hat{\beta}_k\langle \nabla f(x^k), x^k - J_k\rangle + \frac{\hat{\beta}_k}{2}\|x^k - J_k\|^2,$$

which together with (63) and that $x^k - J_k \neq 0$ implies that $\hat{y}^k \neq x^k$. Furthermore, it is similar to (55) in the proof of Theorem 5.2 that $\frac{1}{2}\|x^k - J_k\| \le \|\nabla f(\hat{y}^k) - \nabla f(x^k)\|$. Due to the Lipschitz continuity with constant L of ∇f, we get from the latter and (63) that

$$\frac{1}{2}\|x^k - J_k\| \le L\|x^k - \hat{y}^k\| = L\hat{\beta}_k\|x^k - J_k\|.$$

Since $x^k - J_k \neq 0$, the inequality above yields $\hat{\beta}_k \ge \frac{1}{2L}$ and thus $\beta_k \ge \frac{\theta}{2L}$ when $\beta_k < 1$. It follows that $\beta_k \ge \min\{1, \frac{\theta}{2L}\}$ as desired.

To verify the second part, suppose that $\dim \mathcal{H} < +\infty$, $S_* \neq \emptyset$, and that ∇f is locally Lipschitz continuous at any $x \in S_*$. Define also $\hat{\beta}_k = \dfrac{\beta_k}{\theta} > 0$ and $\hat{y}^k = \hat{\beta}_k J_k + (1-\hat{\beta}_k)x^k$. If $\liminf_{k\to\infty} \beta_k < 1$ then there exists $K \in \mathbb{N}$ such that $\beta_k < 1$ for all $k > K$. Similarly to the above argument of the first part, we have $x^k - J_k \neq 0$ and

$$\frac{1}{2}\|x^k - J_k\| \le \|\nabla f(\hat{y}^k) - \nabla f(x^k)\|. \tag{64}$$

Thanks to Theorem 5.2, the sequence $(x^k)_{k\in\mathbb{N}}$ converges (strongly) to some $x_* \in S_*$. Note further from Lemma 2.5 that

$$\|x^k - J_k\| = \|x^k - J(x^k, 1)\| \le \|x^k - J(x^k, \beta_k)\| = \|x^k - x^{k+1}\| \quad \text{for all} \quad k > K,$$

which tells us that $(J_k)_{k\in\mathbb{N}}$ converges to x_*. Thus we get from (63) that $\hat{y}^k$ also converges to x_*. Since ∇f is locally Lipschitz continuous around x_* with some constant $\mathcal{L} > 0$, there exist $K_1 \in \mathbb{N}$ such that $K_1 > K$ and $\|\nabla f(\hat{y}^k) - \nabla f(x^k)\| \le \mathcal{L}\|\hat{y}^k - x^k\|$ for all $k > K_1$. This together with (64) and (63) gives us that

$$\frac{1}{2}\|x^k - J_k\| \le \mathcal{L}\|\hat{y}^k - x^k\| = \mathcal{L}\hat{\beta}_k\|x^k - J_k\| \quad \text{for all} \quad k > K_1.$$

Since $\|x^k - J_k\| \neq 0$, we have $\frac{1}{2} \le \mathcal{L}\hat{\beta}_k$, i.e., $\beta_k \ge \frac{\theta}{2\mathcal{L}}$ for all $k > K_1$. It follows that $\liminf_{k\to\infty} \beta_k \ge \frac{\theta}{2\mathcal{L}}$ when $\liminf_{k\to\infty} \beta_k < 1$. This verifies (62) and completes the proof of the proposition. □

Let us complete the section by presenting a corresponding corollary to Corollary 4.6, which is easily derived from Theorem 5.3 and Proposition 5.4.

Corollary 5.5. *Let $(x^k)_{k\in\mathbb{N}}$ be the sequence generated by* **Method 3**. *Suppose that $S_* \neq \emptyset$. If either*

(i) *the gradient of f is globally Lipschitz continuous on* dom g*;*

or

(ii) $\dim \mathcal{H} < +\infty$ *and the gradient of f is locally Lipschitz continuous on S_*,*

then we have $(f+g)(x^k) - \min_{x\in\mathcal{H}}(f+g)(x) = \mathcal{O}(k^{-1})$ for all $k \in \mathbb{N}$.

6 Conclusion

In Hilbert spaces, it is well-known that convexity on both functions and global Lipschitz continuity on the gradient of f are sufficient for providing convergence of the sequence generated by the proximal forward-backward splitting methods for solving problem (1). However, the Lipschitz assumption is usually a restriction in many particular circumstances. In this work we dealt with weak convergence of the proximal forward-backward splitting Method for convex optimization problems by taking the advantage of the linesearches. This not only eliminates the serious drawback of estimating the Lipschitz constant to choose the stepsize in (2) but also establishes many complexity results without imposing the Lipschitz assumption. Our schemes through the linesearches provide rigorous and implementable ways of updating the iterates, which can be easily adapted for applications.

We hope that this study will serve as a basis for future research on other efficient variants of the the forward-backward splitting methods. In particular we find possibility to develop our methods to the descent coordinate gradient method [19, 23, 25] for solving structured convex optimization problems. Moreover, we discuss in separate papers the cases when f or g are nonconvex following the ideas exposed in [6] and even removing the differentiability of f and adding dynamic choices of the stepsizes with conditional and deflected techniques combining the ideas in [4, 16, 18]. We are also looking to the incremental (sub)gradient method like [21] for problem (1), when f is the sum of a large number of functions.

ACKNOWLEDGMENTS

This work was partially completed while the authors were visiting University of British Columbia (UBC). The authors are grateful to the Irving K. Barber School of Arts and Sciences at UBC and particularly to Heinz Bauschke and Shawn Wang for the generous hospitality.

References

[1] Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer, New York (2011).

[2] Beck, A., Teboulle, M.: A fast iterative shrinkage–thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* **2** (2009) 183–202.

[3] Beck, A., Teboulle, M.: *Gradient-Based Algorithms with Applications to Signal Recovery Problems.* in *Convex Optimization in Signal Processing and Communications*, (D. Palomar and Y. Eldar, eds.) 42–88 University Press, Cambribge (2010).

[4] Bello Cruz, J.Y.: *On proximal subgradient splitting method for minimizing the sum of two nonsmooth convex functions.* `http://arxiv.org/pdf/1410.5477.pdf` (2014).

[5] Bello Cruz, J.Y., de Oliveira, W.: *On weak and strong convergence of the projected gradient method for convex optimization in Hilbert spaces.* `http://arxiv.org/pdf/1402.5884.pdf` (2014).

[6] Bot, R.I., Csetnek, E.R., László, S.: *An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions.* `http://arxiv.org/pdf/1410.0641.pdf` (2014).

[7] Burachik, R.S., Iusem, A.N.: *Set-Valued Mappings and Enlargements of Monotone Operators*, Springer, Berlin (2008).

[8] Combettes, P.L.: Quasi-Fejérian analysis of some optimization algorithms. *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications. Studies in Computational Mathematics* **8** 115–152 North-Holland, Amsterdam (2001).

[9] Combettes, P.L.: Inconsistent signal feasibility problems: Least-squares solutions in a product space. *IEEE Transaction Signal Processing* **42** (1994) 2955–2966.

[10] Combettes, P.L., Pesquet, J.-C. A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of selected topics in signal precessing* **1** (2007) 564–574.

[11] 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* **49** 185–212 Springer, New York (2011).

[12] Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation* **4** (2005) 1168–1200.

[13] Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics* **57** (2004) 1413–1457.

[14] Eicke, B.: Iteration methods for convexly constrained ill-posed problems in Hilbert space. *Numerical Functional Analysis and Optimization* **13** (1992) 413–429.

[15] Ermoliev, Yu.M.: On the method of generalized stochastic gradients and quasi-Fejér sequences. *Cybernetics* **5** (1969) 208–220.

[16] d'Antonio, G., Frangioni, A.: Convergence analysis of deflected conditional approximate subgradient methods. *SIAM Journal on Optimization* **20** (2009) 357–386.

[17] Iusem, A.N., Svaiter, B.F., Teboulle, M.: Entropy-like proximal methods in convex programming. *Mathematics of Operations Research* **19** (1994) 790–814.

[18] Larson, T., Patriksson, M., Stromberg, A-B.: Conditional subgradient optimization - Theory and application. *European Journal of Operational Research* **88** (1996) 382–403.

[19] Lu, Z., Xiao, L.: On the complexity analysis of randomized block-coordinate descent methods. *Mathematical Programming.* DOI: 10.1007/s10107-014-0800-2 (2014).

[20] Neal, P., Boyd, S.: Proximal Algorithms. *Foundations and Trends in Optimization* **1** (2014) 127–239.

[21] Nedic, A., Bertsekas, D.P.: Incremental subgradient methods for nondifferentiable optimization. *SIAM Journal on Optimization* **12** (2001) 109–138.

[22] Nesterov, Yu.: Gradient methods for minimizing composite functions. *Mathematical Programming* **140** (2013) 125–161.

[23] Nesterov, Yu.: Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM Journal on Optimization* **22** (2012) 341–362.

[24] Nesterov, Yu.: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Mathematics Doklady* **27** (1983) 372–376.

[25] Richtárik, P., Takáč, M.: Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Mathematical Programming* **144** (2014) 1–38.

[26] Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research* **1** (1976) 97–116.

[27] Tseng, P.: Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM Journal on Control Optimization* **29** (1991) 119–138.

[28] Tseng, P.: Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming. *Mathematical Programming* **48** (1990) 249–263.

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