Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 8(5): 12195–12216.
DOI:10.3934/math.2023615
Received: 28 January 2023
Revised: 12 March 2023
Accepted: 14 March 2023
Published: 22 March 2023

Research article

Convergence rates of the modified forward reflected backward splitting algorithm in Banach spaces

Weibo Guan and Wen Song*

School of Mathematical Sciences, Harbin Normal University, Harbin 150025, China

* **Correspondence:** Email: wsong@hrbnu.edu.cn.

Abstract: Consider the problem of minimizing the sum of two convex functions, one being smooth and the other non-smooth in Banach space. In this paper, we introduce a non-traditional forward-backward splitting method for solving such minimization problem. We establish different convergence estimates under different stepsize assumptions.

Keywords: convergence rate; forward-backward splitting algorithm; Banach spaces
Mathematics Subject Classification: 65K05, 90C25, 90C30

1. Introduction

Let X be a reflexive, strictly convex and smooth Banach space with dual space X^*. Consider the optimization problem:

$$\min_{x\in X} \Phi(x) =: f(x) + g(x), \tag{1.1}$$

where, the following assumptions are made throughout the paper:

- $f : X \to (-\infty, +\infty]$ is proper, lower semicontinuous and convex.
- $g : X \to R$ is convex Gâteaux differentiable, and its gradient is Lipschitz continuous with constant L :

$$\|\nabla g(x) - \nabla g(y)\| \le L\|x - y\|, \forall x, y \in X.$$

- The set of solutions to Problem (1.1), denote by Sol(P), is nonempty. The optimal value is denoted by Φ^*.

Recently problem (1.1) 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., [1, 2] and the references therein. It is known that problem (1.1) is characterized by the fixed point equation:

$$x = \mathrm{prox}_f^t(x - t\nabla g(x)).$$

where, $t > 0$ and

$$\mathrm{prox}_f^t(x) = \mathrm{argmin}_{y\in H}\{tf(y) + \frac{1}{2}\|x-y\|^2\}.$$

This equation suggests the possibility of iterating(see [3]):

$$x_{n+1} = \mathrm{prox}_f^{t_n}(x_n - t_n\nabla g(x_n)).$$

This method is called the forward-backward splitting method and includes, in particular, the proximal point method and the gradient method. The forward-backward splitting method is an effective method to solve (1.1), which allows to decouple the contributions of the functions f and g in a gradient descent step determined by f and in a backward implicit step induced by g taking the advantage of some Lipschitz assumption on the derivative of g at each iteration. Forward-backward methods belong to the class of proximal splitting methods. These methods require the computation of the proximity operator and the approximation of proximal points (see [4]).

In 2020, Malitsky and Tam [5] introduced the forward-reflected-backward algorithm. Given $\lambda_0 > 0, \delta \in (0,1), \gamma \in \{1, \frac{1}{\beta}\}$ and $\beta \in (0,1)$. Compute

$$x_{n+1} = \mathrm{prox}_f^{\lambda_n}(x_n - \lambda_n\nabla g(x_n) - \lambda_{n-1}(\nabla g(x_n) - \nabla g(x_{n-1}))) \tag{1.2}$$

where the stepsize $\lambda_n = \gamma\lambda_{n-1}\beta^i$ with i being the smallest nonnegative integer satisfying $\lambda_n\|\nabla g(x_{n+1}) - \nabla g(x_n))\| \le \frac{\delta}{2}\|x_{n+1} - x_n\|$. Very recently, Padcharoen et al. [6] proposed the modified forward-backward splitting method. Give $\{\lambda_n\} \subset (0, \frac{1}{L}), \{\alpha_n\} \subset [0,\alpha] \subset [0,1)$. Let $x_0, x_1 \in H$ and compute

$$\begin{cases} \omega_n = x_n + \theta_n(x_n - x_{n-1}), \\ y_n = \mathrm{prox}_f^{\lambda_n}(\omega_n - \lambda_n\nabla g(\omega_n)), \\ x_{n+1} = y_n - \lambda_n(\nabla g(y_n) - \nabla g(\omega_n)). \end{cases}$$

They established weak convergence of the proposed method.

Generalization of this method from Hilbert space to Banach space is not immediate. In [7], the following generalization of the forward-backward iteration procedure was proposed in reflexive Banach spaces X:

$$x_0 \in X, x_{n+1} \in \mathrm{argmin}_{y\in X}\{\frac{1}{p}\|y - x_n\|^p + t_n(\langle\nabla g(x_n), y\rangle + f(y))\}, \tag{1.3}$$

where, the gradient operator ∇g is $(p-1)$ Hölder-continuous on X, i.e., there exists a constant L such that

$$\|\nabla g(x) - \nabla g(y)\| \le L\|x - y\|^{p-1}, \ \forall\ x, y \in X.$$

In [8], Guan and Song replace the square of the norm distance with Bregman distance proposed another type generalization of the forward-backward method in Banach spaces

$$x_{n+1} = \mathrm{argmin}_{y\in X}\{\frac{1}{q}\|x_n\|^p + \frac{1}{p}\|y\|^p - \langle J_p(x_n), y\rangle + t_n(\langle\nabla g(x_n) + J_p(z_n), y\rangle + f(y))\},$$

where, J_p is the p-duality mapping and q is the dual exponent.

In [9], Guan and Song further extended forward-backward splitting method (1.3) to more general case, i.e., by taking a convex combination of the current step and the previous step:

$$\begin{cases} y_n = \operatorname{argmin}_{y\in X}\{\frac{1}{p}\|x_n - y\|^p + t_n(\langle \nabla g(x_n) + J_p(z_n), y\rangle + f(y))\}, \\ x_{n+1} = (1-\lambda_n)x_n + \lambda_n y_n. \end{cases}$$

In [7–9], they respectively proved that the sequence of functional values converges with the convergence rate n^{1-p} to the optimal value of Problem (1.1).

Inspired and motivated by previous works, we replace x_{n-1} with x_{n+1} and the square of the norm distance with Bregman distance in (1.2) to study non-traditional schemes of the forward-backward splitting algorithm for minimization problems with nonsmooth convex functionals in a Banach space. This non-traditional algorithm is an implicit algorithm. The traditional forward backward splitting algorithm is explicit algorithm, and the forward step and the backward step are completely separated. After simple calculation (see Remark 3.1), the implicit algorithm can be converted into an explicit algorithm, but the forward step and the backward step of the explicit algorithm are partially separated. Our main goal is to prove the criterial convergence of these algorithms and obtain estimates of the convergence rate under different stepsize assumptions.

2. Mathematical toolbox

Let $f : X \to (-\infty, +\infty]$ be an extended real valued function. The subdifferential of f is the set valued operator $\partial f : X \to X^*$ the value of which at $x \in X$ is

$$\partial f(x) = \{x^* \in X^* : \langle x^*, y - x\rangle \leq f(y) - f(x),\ \forall y \in X\}.$$

Let D be a closed convex set. For every $x \in D$, we define the set of normals to D at x by

$$N_D(x) = \{x^* \in X^* : \langle x^*, y - x\rangle \leq 0, \forall y \in D\}.$$

The function $\tau_D : X \to (-\infty, +\infty]$ defined by

$$\tau_D(x) = \begin{cases} 0, & \text{if } x \in \text{D}, \\ +\infty, & \text{otherwise}, \end{cases}$$

is called the indicator function of D. Clearly, τ_D is a proper convex function and for every $x \in D$, we have $\partial\tau_D(x) = N_D(x)$.

The duality mapping $J : X \to X^*$ is defined by

$$J(x) = \{x^* \in X^* | \langle x^*, x\rangle = \|x^*\|\|x\|, \|x^*\| = \|x\|\}, \quad \forall x \in X.$$

The Hahn-Banach theorem guarantees that $J(x) \neq \emptyset$ for every $x \in X$. It is clear that $J(x) = \partial(\frac{1}{2}\|\cdot\|^2)(x)$ for all $x \in X$. It is well known that if X is smooth, then J is single valued and is norm-to-weak star continuous. X is reflexive if and only if J is surjective. X is strictly convex if and only if J is injective. In particular, J is a monotone operator in any Banach space, that is,

$$\langle J(x) - J(y), x - y\rangle \geq 0, \forall x, y \in X.$$

The duality mapping J is said to be weakly continuous on a smooth Banach space if $x_n \rightharpoonup x$ implies $J(x_n) \rightharpoonup J(x)$. This happens, for example, if X is a Hilbert space, or finite dimensional and smooth, or $l^p, 1 < p < +\infty$. This property of Banach spaces was introduced by Browder [10]. It is also well known that if a Banach space X is reflexive, strictly convex and smooth, then the duality mapping J^* from X^* into X is the inverse of J, that is, $J^{-1} = J^*$. Properties of the duality mapping have been given in [11–14].

Let X be a smooth Banach space. Alber ([15]) considered the following function:

$$W(x, y) = \|x\|^2 - 2\langle J(x), y\rangle + \|y\|^2, \forall x, y \in X.$$

It is obvious from the definition of W that

$$(\|x\| - \|y\|)^2 \leq W(x, y) \leq (\|x\| + \|y\|)^2, \ \forall \ x, y \in X.$$

Consider the Moreau envelope $\mathrm{env}_{tf}(x)$ and the set-valued proximal mapping $\pi_{tf}(x)$ defined by

$$\mathrm{env}_{tf} = \inf_{y\in X}\{tf(y) + \frac{1}{2}W(x, y)\}, \tag{2.1}$$

$$\pi_{tf}(x) = \mathrm{argmin}_{y\in X}\{tf(y) + \frac{1}{2}W(x, y)\}.$$

The operator π_{tf} is called the proximity operator. For every $x \in X$, the infimum in (2.1) is achieved at a unique point $\pi_{tf}(x)$ which is characterized by the inclusion

$$J(x) - J(\pi_{tf}(x)) \in \partial(tf)(\pi_{tf}(x)). \tag{2.2}$$

Let us introduce some concepts and characteristics of Banach space geometry [15, 16]. Denote the modulus of convexity of X by $\delta_X(\epsilon)$ and modulus of smoothness of X by $\rho_X(\tau)$ and set

$$h_X(\tau) = \frac{\rho_X(\tau)}{\tau}. \tag{2.3}$$

We recall that a Banach space X is uniformly convex if and only if $\delta_X(\epsilon) > 0$ for all $\epsilon > 0$ and it is uniformly smooth if and only if $\lim_{\tau\to 0} h_X(\tau) = 0$. The space X is said to be 2-uniformly convex (resp. 2-uniformly smooth) if there is a constant $c > 0$ such that $\delta_X(\epsilon) \geq c\epsilon^2$ (resp. $\rho_X(\tau) \leq c\tau^2$).

If $\|x\| \leq M$ and $\|y\| \leq M$, then J is uniformly monotone in a uniformly convex Banach space in the form

$$\langle J(x) - J(y), x - y\rangle \geq (2\mu)^{-1}M^2\delta_X(\frac{\|x - y\|}{2M}), \tag{2.4}$$

where $1 < \mu < 1.7$ is Figiel constant, it is uniformly continuous in a uniformly smooth Banach space in the form

$$\|J(x) - J(y)\| \leq 8Mh_X(\frac{16\mu\|x - y\|}{M}), \tag{2.5}$$

Next we present some auxiliary lemmas on the recursive numerical inequalities which are often used in the proofs below.

Lemma 2.1. *[17] If X is a reflexive, strictly convex and smooth Banach space and A is a maximal monotone operator, then for each* $\lambda > 0$ *and* $x \in X$*, there is a unique element* $\bar{x}$ *satisfying* $J(x) \in J(\bar{x}) + \lambda A(\bar{x})$.

Lemma 2.2. *[18] Assume that*

(i) $z(t) : [1, \infty) \to [0, \infty)$ *is a continuous decreasing function satisfying* $\int_1^\infty z(t)dt = +\infty$;

(ii) $\varphi(t) : [0, \infty) \to [0, \infty), \varphi(0) = 0$ *is continuous and strictly increasing;*

(iii) $\{\mu(n)\}$ *is a sequence of nonnegative real numbers such that the implicit recursive inequality*

$$\mu_{n+1} \leq \mu_n - \alpha_n \varphi(\mu_{n+1}) \tag{2.6}$$

holds, where $\alpha_n = z(n)$*, then* $\lim_{n\to\infty} \mu_n = 0$*, and there exists* $\bar{c} \geq 1$ *such that for all n=1,2,*$\cdots$

$$\mu_n \leq \varphi^{-1}\left(\frac{\bar{c}}{\sum_{i=1}^n \alpha_i}\right).$$

Lemma 2.3. *[18] Assume that* $\{\mu_n\}$ *is a sequence of non-negative real numbers such that* $\mu_{n+1} \leq \mu_n - c\mu_{n+1}^2$ *with a constant* $c > 0$*, then there exists* $\bar{c} > 0$ *such that*

$$\mu_n \leq \frac{\bar{c}}{n}.$$

Lemma 2.4. *[19] Let* $\{a_n\}, \{b_n\}$ *and* $\{\epsilon_n\}$ *be real sequences. Assume that* $\{a_n\}$ *is bounded from below,* $\{b_n\}$ *is nonnegative,* $\sum_{n=1}^\infty |\epsilon_n| < +\infty$ *and* $a_{n+1} - a_n + b_n \leq \epsilon_n$*. Then* $\{a_n\}$ *converges and* $\sum_{n=1}^\infty b_n < +\infty$.

Lemma 2.5. *[20] Let X be a uniformly convex Banach space. If* $\lim_{n\to\infty} W(x_n, x_{n+1}) \to 0$*, then* $\lim_{n\to\infty} \|x_n - x_{n+1}\| \to 0$.

Lemma 2.6. *[21] If* $g : X \to R$ *is convex Gâteaux differentiable, and its gradient is Lipschitz continuous with constant L, then we have*

$$\langle \nabla g(x) - \nabla g(y), x - y\rangle \geq \frac{1}{L}\|\nabla g(x) - \nabla g(y)\|^2, \forall x, y \in X.$$

Lemma 2.7. *[22] (Descent lemma) Let* $g : X \to R$ *be a continuously differentiable function whose gradient is Lipschitz continuous with constant L. Then, for any* $x, y \in X$*, we have*

$$g(x) \leq g(y) + \langle \nabla g(y), x - y\rangle + \frac{L}{2}\|x - y\|^2.$$

Lemma 2.8. *[23] Let X be a 2-uniformly convex and 2-uniformly smooth Banach space. Then there exists two constants* $\alpha > 0$ *and* $\beta > 0$ *such that*

$$\alpha\|x - y\|^2 \leq W(x, y) \leq \beta\|x - y\|^2, \ \forall\ x, y \in X.$$

Remark 2.1. *If X is a 2-uniformly convex and 2-uniformly smooth Banach space, then X is also a 2-uniformly convex and 2-uniformly smooth Banach space. Hence, there exists a constant* $c > 0$ *such that*

$$\langle J(x) - J(y), x - y\rangle \geq c \max\{\|J(x) - J(y)\|^2, \|x - y\|^2\}, \forall x, y \in X.$$

3. Main results

Algorithm 3.1. ***Modified forward reflected backward splitting algorithm***
Take $x_0 \in X$. Given x_n, define x_{n+1} by the inclusion

$$x_{n+1} = \pi_{t_n f} J^{-1}(J(x_n) - t_n \nabla g(x_n) - \lambda_n(\nabla g(x_{n+1}) - \nabla g(x_n))), \tag{3.1}$$

where, $\{t_n\}$ and $\{\lambda_n\}$ are two sequences of nonnegative real numbers.

Remark 3.1. *By (2.2), we have*

$$\begin{aligned}
& x_{n+1} = \pi_{t_n f} J^{-1}(J(x_n) - t_n \nabla g(x_n) - \lambda_n(\nabla g(x_{n+1}) - \nabla g(x_n))) \\
\Leftrightarrow\ & J(x_n) - t_n \nabla g(x_n) - \lambda_n(\nabla g(x_{n+1}) - \nabla g(x_n)) - J(x_{n+1}) \in t_n \partial f(x_{n+1}) \\
\Leftrightarrow\ & J(x_n) - t_n \nabla g(x_n) + \lambda_n \nabla g(x_n) \in J(x_{n+1}) + \lambda_n \nabla g(x_{n+1}) + t_n \partial f(x_{n+1}) \\
\Leftrightarrow\ & JJ^{-1}(J(x_n) - t_n \nabla g(x_n) + \lambda_n \nabla g(x_n)) \in J(x_{n+1}) + (\lambda_n \nabla g + t_n \partial f)(x_{n+1}) \\
\Leftrightarrow\ & x_{n+1} = \pi_{(\lambda_n g + t_n f)} J^{-1}(J(x_n) - t_n \nabla g(x_n) + \lambda_n \nabla g(x_n)).
\end{aligned}$$

Hence, the iterative sequence $\{x_n\}$ defined by Algorithm 3.1 is well-defined, that is, for each x_n, there is a unique element x_{n+1} satisfying Eq (3.1).

Remark 3.2. *Notice that*

$$\begin{aligned}
& x_{n+1} = \pi_{t_n f} J^{-1}(J(x_n) - t_n \nabla g(x_n) - \lambda_n(\nabla g(x_{n+1}) - \nabla g(x_n))) \\
\Leftrightarrow\ & JJ^{-1}(J(x_n) - t_n \nabla g(x_n) + \lambda_n \nabla g(x_n)) \in J(x_{n+1}) + (\lambda_n \nabla g + t_n \partial f)(x_{n+1}) \\
\Leftrightarrow\ & x_{n+1} \in (J + \lambda_n \nabla g + t_n \partial f)^{-1}(J(x_n) - t_n \nabla g(x_n) + \lambda_n \nabla g(x_n)).
\end{aligned}$$

Hence, by Lemma 2.1, we also know that the iterative sequence $\{x_n\}$ defined by Algorithm 3.1 is well-defined.

Proposition 3.1. *Let $\{x_n\}$ be a sequence generated by Algorithm 3.1 and define*

$$h(x_n) :\ = f(x_n) - f(x_{n+1}) + \langle \nabla g(x_n), x_n - x_{n+1} \rangle.$$

Assume that $\langle J(x) - J(y), x - y \rangle \geq c\|x - y\|^2,\ \forall x, y \in X$ with some constant $c > 0$. If $0 < \bar{t} < t_n \leq \frac{2c - 2L\lambda_n}{L}$, then we have

(i) $\|x_n - x_{n+1}\|^2 \leq \frac{t_n}{c - L\lambda_n} h(x_n)$.

(ii) $\Phi(x_{n+1}) \leq \Phi(x_n) - (1 - \frac{t_n L}{2c - 2L\lambda_n}) h(x_n)$.

Proof. (i). By Remark 3.1, we have that

$$J(x_n) - t_n \nabla g(x_n) - \lambda_n(\nabla g(x_{n+1}) - \nabla g(x_n)) - J(x_{n+1}) \in t_n \partial f(x_{n+1}) \tag{3.2}$$

and so that

$$\langle J(x_n) - t_n \nabla g(x_n) - \lambda_n(\nabla g(x_{n+1}) - \nabla g(x_n)) - J(x_{n+1}), x_n - x_{n+1} \rangle \leq t_n(f(x_n) - f(x_{n+1})).$$

Hence, by the Lipschitz continuity of $\nabla g(\cdot)$, we have

$$\begin{aligned}
&\frac{1}{t_n}\langle J(x_n) - J(x_{n+1}), x_n - x_{n+1}\rangle \\
\leq\ & f(x_n) - f(x_{n+1}) + \langle \nabla g(x_n), x_n - x_{n+1}\rangle + \frac{\lambda_n}{t_n}\langle \nabla g(x_{n+1}) - \nabla g(x_n), x_n - x_{n+1}\rangle \\
\leq\ & f(x_n) - f(x_{n+1}) + \langle \nabla g(x_n), x_n - x_{n+1}\rangle + \frac{\lambda_n}{t_n}\|\nabla g(x_{n+1}) - \nabla g(x_n)\|\|x_n - x_{n+1}\| \\
\leq\ & h(x_n) + \frac{L\lambda_n}{t_n}\|x_n - x_{n+1}\|^2.
\end{aligned}$$

By condition $\langle J(x) - J(y), x - y\rangle \geq c\|x - y\|^2,\ \forall x, y \in X$, we can get

$$\frac{1}{t_n}\langle J(x_n) - J(x_{n+1}), x_n - x_{n+1}\rangle \geq \frac{c}{t_n}\|x_n - x_{n+1}\|^2.$$

Hence we have

$$\|x_n - x_{n+1}\|^2 \leq \frac{t_n}{c - L\lambda_n}h(x_n).$$

(ii). Using the definition of $h(x_n)$, we have

$$\Phi(x_n) - \Phi(x_{n+1}) = h(x_n) + g(x_n) - g(x_{n+1}) - \langle \nabla g(x_n), x_n - x_{n+1}\rangle. \tag{3.3}$$

By the Lipschitz continuity of $\nabla g(\cdot)$ and Lemma 2.7, we get that

$$|g(x_n) - g(x_{n+1}) - \langle \nabla g(x_n), x_n - x_{n+1}\rangle| \leq \frac{L}{2}\|x_n - x_{n+1}\|^2 \leq \frac{t_n L}{2c - 2L\lambda_n}h(x_n). \tag{3.4}$$

It follows from (3.3) and (3.4) that

$$\Phi(x_{n+1}) \leq \Phi(x_n) - (1 - \frac{t_n L}{2c - 2L\lambda_n})h(x_n).$$

□

Proposition 3.2. *Let* $\{x_n\}$ *be a sequence generated by Algorithm 3.1. Assume that*

$$\langle J(x) - J(y), x - y\rangle \geq c\max\{\|x - y\|^2, \|J(x) - J(y)\|^2\},\ \forall x, y \in X$$

with some constant $c > 0$ *and the sequence* $\{x_n\}$ *is bounded. If* $0 < \bar{t} < t_n \leq \frac{2c - 2L\lambda_n}{L}$, *then the following hold.*

(i) $\lim_{n\to\infty} \Phi(x_n) = \Phi^*$.

(ii) *all weak accumulation points of* $\{x_n\}$ *belong to* $\mathrm{Sol}(P)$.

(iii) $\Phi(x_n) - \Phi^* \leq \beta n^{-1}$ *for some* $\beta > 0$.

Proof. (i). Let $\hat{x} \in \mathrm{Sol}(P)$. Since

$$g(x_n) - g(\hat{x}) \le \langle \nabla g(x_n), x_n - \hat{x} \rangle$$

and

$$\Phi(x_n) - \Phi^* = f(x_n) - f(\hat{x}) + g(x_n) - g(\hat{x}),$$

we have that

$$\begin{aligned} \Phi(x_n) - \Phi^* &\le f(x_n) - f(\hat{x}) + \langle \nabla g(x_n), x_n - \hat{x} \rangle \\ &= h(x_n) + f(x_{n+1}) - f(\hat{x}) + \langle \nabla g(x_n), x_{n+1} - \hat{x} \rangle. \end{aligned} \tag{3.5}$$

Then, by (3.5) and (3.2), we have

$$\begin{aligned} & \Phi(x_n) - \Phi^* \\ \le\ & h(x_n) + \langle \nabla g(x_n), x_{n+1} - \hat{x} \rangle \\ & - \left\langle \frac{J(x_n) - J(x_{n+1})}{t_n} - \nabla g(x_n) - \frac{\lambda_n}{t_n}(\nabla g(x_{n+1}) - \nabla g(x_n)), \hat{x} - x_{n+1} \right\rangle \\ =\ & h(x_n) - \left\langle \frac{J(x_n) - J(x_{n+1})}{t_n} - \frac{\lambda_n}{t_n}(\nabla g(x_{n+1}) - \nabla g(x_n)), \hat{x} - x_{n+1} \right\rangle \\ =\ & h(x_n) + \|\frac{J(x_n) - J(x_{n+1})}{t_n}\|\|\hat{x} - x_{n+1}\| + \frac{\lambda_n}{t_n}\|\nabla g(x_{n+1}) - \nabla g(x_n)\|\|\hat{x} - x_{n+1}\| \\ \le\ & h(x_n) + \frac{1}{ct_n}\|x_n - x_{n+1}\|\|\hat{x} - x_{n+1}\| + \frac{L\lambda_n}{t_n}\|x_{n+1} - x_n\|\|\hat{x} - x_{n+1}\| \\ \le\ & h(x_n) + c_1\|x_{n+1} - x_n\|, \end{aligned}$$

where, $c_1 \ge \frac{1}{ct_n}\|\hat{x} - x_{n+1}\| + \frac{L\lambda_n}{t_n}\|\hat{x} - x_{n+1}\|$. Hence, by Proposition 3.1, we have

$$\begin{aligned} \Phi(x_n) - \Phi^* &\le h(x_n) + c_1\sqrt{\frac{t_n}{c - L\lambda_n}h(x_n)} \\ &\le \frac{\Phi(x_n) - \Phi(x_{n+1})}{(1 - \frac{t_n L}{2c - 2L\lambda_n})} + c_1\sqrt{\frac{t_n}{c - L\lambda_n}\frac{\Phi(x_n) - \Phi(x_{n+1})}{(1 - \frac{Lt_n}{2c - 2L\lambda_n})}} \\ &\le (\Phi(x_n) - \Phi(x_{n+1}))^{\frac{1}{2}}(\frac{1}{c_2})^{\frac{1}{2}}, \end{aligned} \tag{3.6}$$

where, $(\frac{1}{c_2})^{\frac{1}{2}} \ge \frac{\sqrt{\Phi(x_n) - \Phi(x_{n+1})}}{(1 - \frac{t_n L}{2c - 2L\lambda_n})} + c_1\sqrt{\frac{2t_n}{2c - 2L\lambda_n - Lt_n}}$. Then we have

$$c_2(\Phi(x_n) - \Phi^*)^2 \le (\Phi(x_n) - \Phi^*) - (\Phi(x_{n+1}) - \Phi^*). \tag{3.7}$$

By Lemma 2.4 and (3.7), we obtain, $\lim\limits_{n\to\infty} \Phi(x_n) = \Phi^*$.

(ii). Since $\lim_{n\to\infty} \Phi(x_n) = \Phi^*$, the sequence $\{x_n\}$ is a minimizing sequence, thus, due to the weak lower semicontinuity of Φ, all weak accumulation points of $\{x_n\}$ belong to $\mathrm{Sol}(P)$.

(iii). By (3.7), we have

$$c_2(\Phi(x_{n+1}) - \Phi^*)^2 \le c_2(\Phi(x_n) - \Phi^*)^2 \le (\Phi(x_n) - \Phi^*) - (\Phi(x_{n+1}) - \Phi^*).$$

Then, by Lemma 2.3, there exists $\beta > 0$, such that $\Phi(x_n) - \Phi^* \le \beta n^{-1}$. □

Proposition 3.3. *Let $\{x_n\}$ be a sequence generated by Algorithm 3.1. If $\lambda_n = t_n \geq \bar{t} > 0$, then we have*

(i) *the sequence $\{x_n\}$ is bounded.*

(ii) $\lim_{n\to\infty} \Phi(x_n) = \Phi^*$.

(iii) *all weak accumulation points of $\{x_n\}$ belong to* $\mathrm{Sol}(P)$.

(iv) *if the duality mapping J is weakly continuous, then $\{x_n\}$ convergence weakly as $n \to +\infty$ to a point in* $\mathrm{Sol}(P)$.

Proof. (i). For all $\bar{x} \in \mathrm{Sol}(P)$,

$$\begin{aligned} W(x_{n+1}, \bar{x}) &\leq W(x_n, \bar{x}) + 2\langle J(x_{n+1}) - J(x_n), x_{n+1} - \bar{x}\rangle \\ &= W(x_n, \bar{x}) - 2t_n\langle \partial\Phi(x_{n+1}), x_{n+1} - \bar{x}\rangle \\ &\leq W(x_n, \bar{x}) - 2t_n(\Phi(x_{n+1}) - \Phi(\bar{x})). \end{aligned} \tag{3.8}$$

Then we have

$$W(x_{n+1}, \bar{x}) \leq W(x_n, \bar{x}) \leq W(x_1, \bar{x}).$$

Hence, the sequence $\{x_n\}$ is bounded.

(ii). By (3.8) and $t_n \geq \bar{t}$, we have $\lim_{n\to\infty} \Phi(x_n) = \Phi^*$.

(iii). Since $\lim_{n\to\infty} \Phi(x_n) = \Phi^*$, the sequence $\{x_n\}$ is a minimizing sequence, thus, due to the weak lower semicontinuity of Φ, all weak accumulation points of $\{x_n\}$ belong to $\mathrm{Sol}(P)$.

(iv). The space being reflexive, it suffices to prove that $\{x_n\}$ has only one weak cluster point as $n \to +\infty$. Suppose otherwise that $x_{n,k} \rightharpoonup \tilde{x}_1 \in \mathrm{Sol}(P)$ and $x_{n,s} \rightharpoonup \tilde{x}_2 \in \mathrm{Sol}(P)$. By part (i), we know that there exists nonnegative numbers n_1 and n_2 such that

$$\lim_{n\to\infty} W(x_n, \tilde{x}_1) \to n_1 \ \ \textit{and} \ \ \lim_{n\to\infty} W(x_n, \tilde{x}_2) \to n_2.$$

Hence

$$\begin{aligned} \lim_{n\to\infty}(W(x_n, \tilde{x}_1) - W(x_n, \tilde{x}_2)) &= \|\tilde{x}_1\|^2 - \|\tilde{x}_2\|^2 + 2\lim_{n\to\infty}\langle J(x_n), \tilde{x}_2 - \tilde{x}_1\rangle \\ &= n_1 - n_2. \end{aligned}$$

Since the duality mapping J is weakly continuous, we have

$$\|\tilde{x}_1\|^2 - \|\tilde{x}_2\|^2 + 2\langle J(\tilde{x}_1), \tilde{x}_2 - \tilde{x}_1\rangle = n_1 - n_2$$

and

$$\|\tilde{x}_1\|^2 - \|\tilde{x}_2\|^2 + 2\langle J(\tilde{x}_2), \tilde{x}_2 - \tilde{x}_1\rangle = n_1 - n_2.$$

Hence

$$\langle J(\tilde{x}_2) - J(\tilde{x}_1), \tilde{x}_2 - \tilde{x}_1\rangle = 0.$$

Since X is strictly convex, according to the properties of the duality mapping J, we conclude that

$$\tilde{x}_2 = \tilde{x}_1,$$

which establishes the uniqueness of the weak accumulation point. □

Proposition 3.4. *Let $\{x_n\}$ be defined by Algorithm 3.1. Assume that $t_n \le t_{n+1}$ and $h_n(x) = \frac{1}{2t_{n-1}}\|x\|^2 - g(x)$, $n = 1, 2, \cdots$ is a convex function. Then we have the following:*

(i) *if the sequence $\{x_n\}$ is bounded and $\sum_{i=1}^{\infty} \lambda_i < +\infty$, then $\lim_{n\to\infty} \Phi(x_n) = \Phi^*$ and all weak accumulation points of $\{x_n\}$ belong to* Sol(P).

(ii) *if $\lambda_n = 0$ and $t_{n+1} \le \bar{t}$, then the sequence $\{x_n\}$ is bounded.*

Proof. Let

$$G_n(x) = \Phi(x) + \frac{1}{2t_{n-1}} W(x_{n-1}, x) - D_g(x, x_{n-1}) + \frac{\lambda_{n-1}}{t_{n-1}} \langle \nabla g(x_n) - \nabla g(x_{n-1}), x\rangle$$

and

$$g_n(x) = \frac{1}{2t_{n-1}} W(x_{n-1}, x) - D_g(x, x_{n-1}) + \frac{\lambda_{n-1}}{t_{n-1}} \langle \nabla g(x_n) - \nabla g(x_{n-1}), x\rangle,$$

where, $D_g(x, x_{n-1}) = g(x) - g(x_{n-1}) - \langle \nabla g(x_{n-1}), x - x_{n-1}\rangle$ is a Bregman distance function. Since $h_n(x) = \frac{1}{2t_{n-1}}\|x\|^2 - g(x)$ is a convex function, we have that

$$g_n(x) = D_{h_n}(x, x_{n-1}) + \frac{\lambda_{n-1}}{t_{n-1}} \langle \nabla g(x_n) - \nabla g(x_{n-1}), x\rangle \ge \frac{\lambda_{n-1}}{t_{n-1}} \langle \nabla g(x_n) - \nabla g(x_{n-1}), x\rangle$$

and so that

$$G_n(x) \ge \Phi(x) + \frac{\lambda_{n-1}}{t_{n-1}} \langle \nabla g(x_n) - \nabla g(x_{n-1}), x\rangle \ge \Phi^* + \frac{\lambda_{n-1}}{t_{n-1}} \langle \nabla g(x_n) - \nabla g(x_{n-1}), x\rangle.$$

Then we have,

$$\begin{aligned}
& G_n(x) - G_n(x_n) \\
= \; & \Phi(x) + \frac{1}{2t_{n-1}} W(x_{n-1}, x) - D_g(x, x_{n-1}) + \frac{\lambda_{n-1}}{t_{n-1}} \langle \nabla g(x_n) - \nabla g(x_{n-1}), x\rangle \\
& -\Phi(x_n) - \frac{1}{2t_{n-1}} W(x_{n-1}, x_n) + D_g(x_n, x_{n-1}) - \frac{\lambda_{n-1}}{t_{n-1}} \langle \nabla g(x_n) - \nabla g(x_{n-1}), x_n\rangle \\
= \; & \frac{1}{2t_{n-1}} W(x_n, x) + f(x) - f(x_n) \\
& - \left\langle \frac{J(x_{n-1}) - J(x_n)}{t_{n-1}} - \nabla g(x_{n-1}) - \frac{\lambda_{n-1}}{t_{n-1}} (\nabla g(x_n) - \nabla g(x_{n-1})), x - x_n \right\rangle.
\end{aligned}$$

Since $\frac{J(x_{n-1}) - J(x_n)}{t_{n-1}} - \nabla g(x_{n-1}) - \frac{\lambda_{n-1}}{t_{n-1}} (\nabla g(x_n) - \nabla g(x_{n-1})) \in \partial f(x_n)$, then we have

$$f(x) - f(x_n) - \left\langle \frac{J(x_{n-1}) - J(x_n)}{t_{n-1}} - \nabla g(x_{n-1}) - \frac{\lambda_{n-1}}{t_{n-1}} (\nabla g(x_n) - \nabla g(x_{n-1})), x - x_n \right\rangle \ge 0.$$

Hence,

$$\begin{aligned}
G_n(x) - G_n(x_n) &\ge \frac{1}{2t_{n-1}} W(x_n, x) \ge \frac{1}{2t_n} W(x_n, x) \\
&\ge g_{n+1}(x) - \frac{\lambda_n}{t_n} \langle \nabla g(x_{n+1}) - \nabla g(x_n), x\rangle. \qquad (3.9)
\end{aligned}$$

Let $\hat{x} \in \mathrm{Sol}(P)$, then we have

$$\begin{aligned} & G_n(\hat{x}) - G_n(x_n) \\ = \;& \Phi(\hat{x}) + g_n(\hat{x}) - \Phi(x_n) - g_n(x_n) \\ \leq \;& \Phi(\hat{x}) - \Phi(x_n) - g_n(x_n) + G_{n-1}(\hat{x}) - G_{n-1}(x_{n-1}) + \frac{\lambda_{n-1}}{t_{n-1}}\langle \nabla g(x_n) - \nabla g(x_{n-1}), \hat{x}\rangle. \end{aligned}$$

Hence, by $g_n(x_n) \geq \frac{\lambda_{n-1}}{t_{n-1}}\langle \nabla g(x_n) - \nabla g(x_{n-1}), x_n\rangle$ and $\Phi(\hat{x}) \leq \Phi(x_n)$, we have

$$\begin{aligned} & (G_n(\hat{x}) - G_n(x_n)) - (G_{n-1}(\hat{x}) - G_{n-1}(x_{n-1})) \\ \leq \;& \Phi(\hat{x}) - \Phi(x_n) - g_n(x_n) + \frac{\lambda_{n-1}}{t_{n-1}}\langle \nabla g(x_n) - \nabla g(x_{n-1}), \hat{x}\rangle \\ \leq \;& -\frac{\lambda_{n-1}}{t_{n-1}}\langle \nabla g(x_n) - \nabla g(x_{n-1}), x_n\rangle + \frac{\lambda_{n-1}}{t_{n-1}}\langle \nabla g(x_n) - \nabla g(x_{n-1}), \hat{x}\rangle. \end{aligned} \tag{3.10}$$

By (3.9), we have that the sequence $\{G_n(\hat{x}) - G_n(x_n)\}$ is bounded from below. Since $\{x_n\}$ is bounded, then by Lemma 2.4 and (3.10), we obtain that $\{G_n(\hat{x}) - G_n(x_n)\}$ is converges. Hence, by (3.10), we have that $\lim_{n\to\infty} \Phi(x_n) = \Phi^*$, and so that, all weak accumulation points of $\{x_n\}$ belong to $\mathrm{Sol}(P)$.

(ii) If $\lambda_n = 0$, then $g_n(x) = \frac{1}{2t_{n-1}}W(x_{n-1}, x) - D_g(x, x_{n-1}) \geq 0$. Then, once again from inequalities (3.9) and (3.10), we get that

$$G_n(x) - G_n(x_n) \geq \frac{1}{2t_{n-1}}W(x_n, x)$$

and

$$(G_n(\hat{x}) - G_n(x_n)) - (G_n(\hat{x}) - G_n(x_{n-1})) \leq \Phi(\hat{x}) - \Phi(x_n).$$

Hence, $\{x_n\}$ is bounded. □

Remark 3.3. *If there exist $c > 0$ such that $\langle Jx - Jy, x - y\rangle \geq c\|x - y\|^2$ and ∇g is Lipschitz continuous with constant L, then for $0 < t_n \leq \frac{c}{L}$ we have*

$$\frac{1}{t_n}\langle J(x) - J(y), x - y\rangle \geq \frac{c}{t_n}\|x - y\|^2 \geq L\|x - y\|^2 \geq \langle \nabla g(x) - \nabla g(y), x - y\rangle.$$

Then by

$$\frac{1}{t_n}\langle J(x) - J(y), x - y\rangle \geq \langle \nabla g(x) - \nabla g(y), x - y\rangle \Leftrightarrow \langle \nabla h(x) - \nabla h(y), x - y\rangle \geq 0,$$

we have that, $h_n(x) = \frac{1}{2t_n}\|x\|^2 - g(x), n = 1, 2, \cdots$ is a convex function. □

We have proved that existence of solutions of Problem (1.1) is sufficient to guarantee convergence of the sequence generated by Algorithm 3.1. The next proposition shows that it is also a necessary condition.

Proposition 3.5. *Let $\{x_n\}$ be a sequence generated by Algorithm 3.1. Assume that*

$$h_n(x) = \frac{1}{2t_{n-1}}\|x\|^2 - g(x), n = 1, 2, \cdots$$

is convex function. If the sequence $\{x_n\}$ is bounded and $t_n \leq t_{n+1}$ and $\sum_{i=1}^{\infty} \lambda_i < +\infty$, then $\mathrm{Sol}(P)$ is nonempty.

Proof. Since $\{x_n\}$ is bounded, then weak closure $\overline{\{x_n\}}^w$ of $\{x_n\}$ is also bounded and there exists a bounded closed convex set $D \subset X$ such that

$$\overline{\{x_n\}}^w \subset \text{int}D, \tag{3.11}$$

where int$D$ is the interior of D. It follows that any weak accumulation points of $\{x_n\}$ belong to int$D$. Let $\tilde{f} = f + \tau_D$ and

$$\tilde{x}_0 \in X, \quad \tilde{x}_{n+1} = \pi_{t_n\tilde{f}} J^{-1}(J(\tilde{x}_n) - t_n\nabla g(\tilde{x}_n) - \lambda_n(\nabla g(\tilde{x}_{n+1}) - \nabla g(\tilde{x}_n))).$$

We will first prove by induction that the sequence $\{\tilde{x}_n\}$ coincides with the sequence $\{x_n\}$ when $\tilde{x}_0 = x_0$. Suppose that $\tilde{x}_n = x_n$. By (3.11), we have that $x_{n+1} \in \text{int}D$, and so that,

$$N_D(x_{n+1}) = 0.$$

Hence, we have that

$$\begin{aligned} & \frac{J(x_n) - J(x_{n+1})}{t_n} - \nabla g(x_n) - \frac{\lambda_n}{t_n}(\nabla g(x_{n+1}) - \nabla g(x_n)) \\ \in \quad & \partial f(x_{n+1}) = \partial f(x_{n+1}) + N_D(x_{n+1}) = \partial \tilde{f}(x_{n+1}) \end{aligned}$$

and so that

$$x_{n+1} = \pi_{t_n\tilde{f}} J^{-1}(J(x_n) - t_n\nabla g(x_n)) - \lambda_n(\nabla g(x_{n+1}) - \nabla g(x_n))).$$

Then, by Remark 3.1, we have

$$\tilde{x}_{n+1} = x_{n+1}.$$

Consider the optimization problem:

$$\min_{x\in X} \tilde{f}(x) + g(x). \tag{3.12}$$

Since D is a bounded closed convex set and X is a reflexive Banach space, the solution set of problem (3.12) is nonempty. Then by Proposition 3.4, all weak accumulation points of $\{\tilde{x}_n\}$ are solutions to problem (3.12). Since the sequence $\{\tilde{x}_n\}$ and the sequence $\{x_n\}$ are coincidences, the weak accumulation point $\hat{x}$ of $\{x_n\}$ is also solution to problem (3.12).

We prove next that $\hat{x} \in \text{Sol}(P)$. Since $\hat{x} \in \overline{\{x_n\}}^w \subset \text{int}D$, we obtain $N_D(\hat{x}) = 0$. Hence,

$$0 \in \partial(\tilde{f}(\hat{x}) + g(\hat{x})) = \partial f(\hat{x}) + N_D(\hat{x}) + \nabla g(\hat{x}) = \partial\Phi(\hat{x}),$$

that is, $\hat{x} \in \text{Sol}(P)$. □

Proposition 3.6. *Let X be a uniformly convex and uniformly smooth Banach space. Let $\{x_n\}$ be a sequence generated by Algorithm 3.1. Assume that $\langle J(x) - J(y), x - y\rangle \geq c\|x - y\|^2$, $\forall x, y \in X$ with some constant $c > 0$ and assume that the sequence $\{x_n\}$ is bounded. Assume that $0 \leq \lambda \leq \lambda_n \leq t_n \leq t$ and $(\frac{1}{t} - (1 - \frac{\lambda}{t})\frac{L}{c}) > 0$. Then the following estimate holds*

$$\Phi(x_n) - \Phi^* \leq \varphi^{-1}(\frac{\bar{c}}{n}), \tag{3.13}$$

where $\bar{c}$ is some positive constant and φ is defined by (3.17).

Proof. Using

$$g(x_{n+1}) - g(x_n) \leq \langle \nabla g(x_{n+1}), x_{n+1} - x_n \rangle$$

and

$$J(x_n) - t_n \nabla g(x_n) - \lambda_n(\nabla g(x_{n+1}) - \nabla g(x_n)) - J(x_{n+1}) \in t_n \partial f(x_{n+1}),$$

we calculate

$$\begin{aligned}
& \Phi(x_{n+1}) - \Phi(x_n) \\
= \;& g(x_{n+1}) - g(x_n) + f(x_{n+1}) - f(x_n) \\
\leq \;& \langle \nabla g(x_{n+1}), x_{n+1} - x_n \rangle \\
& + \langle \frac{1}{t_n}(J(x_n) - J(x_{n+1})) - \nabla g(x_n) - \frac{\lambda_n}{t_n}(\nabla g(x_{n+1}) - \nabla g(x_n)), x_{n+1} - x_n \rangle \\
= \;& \frac{1}{t_n}\langle J(x_n) - J(x_{n+1}), x_{n+1} - x_n \rangle + (1 - \frac{\lambda_n}{t_n})\langle \nabla g(x_{n+1}) - \nabla g(x_n), x_{n+1} - x_n \rangle \\
\leq \;& \frac{1}{t_n}\langle J(x_n) - J(x_{n+1}), x_{n+1} - x_n \rangle + (1 - \frac{\lambda_n}{t_n})\|\nabla g(x_{n+1}) - \nabla g(x_n)\|\|x_{n+1} - x_n\| \\
\leq \;& \frac{1}{t_n}\langle J(x_n) - J(x_{n+1}), x_{n+1} - x_n \rangle + (1 - \frac{\lambda_n}{t_n})L\|x_{n+1} - x_n\|^2. \qquad (3.14)
\end{aligned}$$

Then by condition $\langle J(x) - J(y), x - y \rangle \geq c\|x - y\|^2, \; \forall x, y \in X$, we have

$$\begin{aligned}
& \Phi(x_{n+1}) - \Phi(x_n) \\
\leq \;& \frac{1}{t_n}\langle J(x_n) - J(x_{n+1}), x_{n+1} - x_n \rangle + (1 - \frac{\lambda_n}{t_n})L\|x_{n+1} - x_n\|^2 \\
\leq \;& \frac{1}{t_n}\langle J(x_n) - J(x_{n+1}), x_{n+1} - x_n \rangle + (1 - \frac{\lambda_n}{t_n})\frac{L}{c}\langle J(x_n) - J(x_{n+1}), x_n - x_{n+1} \rangle \\
\leq \;& (\frac{1}{t_n} - (1 - \frac{\lambda_n}{t_n})\frac{L}{c})\langle J(x_n) - J(x_{n+1}), x_{n+1} - x_n \rangle \\
\leq \;& \alpha\langle J(x_n) - J(x_{n+1}), x_{n+1} - x_n \rangle. \qquad (3.15)
\end{aligned}$$

where, $0 < \alpha = (\frac{1}{t} - (1 - \frac{\lambda}{t})\frac{L}{c})$. Due to the monotonicity property of J,

$$\langle J(x_n) - J(x_{n+1}), x_{n+1} - x_n \rangle \leq 0.$$

Thus, from (3.15) we obtain the inequality $\Phi(x_{n+1}) \leq \Phi(x_n)$ and $\Phi^* \leq \Phi(x_n) \leq \Phi(x_1)$. Therefore, the sequence $\{\Phi(x_n)\}$ has a limit. Since the sequence $\{x_n\}$ is bounded, there exists $M > 0$ such that $\|x_n\| \leq M$. By (2.4)

$$\langle J(x_n) - J(x_{n+1}), x_n - x_{n+1} \rangle \geq (2\mu)^{-1} M^2 \delta_X(\frac{\|x_n - x_{n+1}\|}{2M}).$$

This implies the estimate

$$\Phi(x_{n+1}) - \Phi(x_n) \leq -\alpha(2\mu)^{-1} M^2 \delta_X(\frac{\|x_n - x_{n+1}\|}{2M}). \qquad (3.16)$$

By analogy with (3.14), we can write down

$$
\begin{aligned}
& \Phi(x_{n+1}) - \Phi(\bar{x}) \\
\leq\ & \langle \nabla g(x_{n+1}), x_{n+1} - \bar{x} \rangle \\
& + \langle \frac{1}{t_n}(J(x_n) - J(x_{n+1})) - \nabla g(x_n) - \frac{\lambda_n}{t_n}(\nabla g(x_{n+1}) - \nabla g(x_n)), x_{n+1} - \bar{x} \rangle \\
\leq\ & \frac{1}{t_n}\|J(x_n) - J(x_{n+1})\|\|x_{n+1} - \bar{x}\| + (1 - \frac{\lambda_n}{t_n})L\|x_{n+1} - x_n\|\|x_{n+1} - \bar{x}\|.
\end{aligned}
$$

By condition $\langle J(x) - J(y), x - y \rangle \geq c\|x - y\|^2,\ \forall x, y \in X$, we have

$$\frac{1}{c}\|J(x) - J(y)\| \geq \|x - y\|,\ \forall x, y \in X.$$

Hence,

$$
\begin{aligned}
& \Phi(x_{n+1}) - \Phi(\bar{x}) \\
\leq\ & \frac{1}{t_n}\|J(x_n) - J(x_{n+1})\|\|x_{n+1} - \bar{x}\| + (1 - \frac{\lambda_n}{t_n})\frac{L}{c}\|J(x_{n+1}) - J(x_n)\|\|x_{n+1} - \bar{x}\| \\
=\ & (\frac{1}{t_n} + (1 - \frac{\lambda_n}{t_n})\frac{L}{c})\|J(x_n) - J(x_{n+1})\|\|x_{n+1} - \bar{x}\|. \\
\leq\ & \beta\|J(x_n) - J(x_{n+1})\|\|x_{n+1} - \bar{x}\|.
\end{aligned}
$$

where, $\beta = (\frac{1}{\lambda} + (1 - \frac{\lambda}{t})\frac{L}{c})$. Hence, we obtain

$$\Phi(x_{n+1}) - \Phi(\bar{x}) \leq C_2 h_X(C_3\|x_{n+1} - x_n\|),$$

where, $C_2 = 8\beta M(M + \|\bar{x}\|),\ C_3 = 16\mu M^{-1}$ and $h_X(t)$ is defined by (2.3).

From (3.16), we now deduce

$$\Phi(x_{n+1}) - \Phi(x_n) \leq -C_1\delta_X(\frac{h_X^{-1}(C_2^{-1}(\Phi(x_{n+1}) - \Phi(\bar{x})))}{2MC_3}).$$

where $C_1 = \alpha(2\mu)^{-1}M^2$. If we denote $\mu_n = \Phi(x_n) - \Phi(\bar{x})$ and

$$\varphi(t) = \delta_X(\frac{h_X^{-1}(C_2^{-1}t)}{2MC_3}), \tag{3.17}$$

then we obtain

$$\mu_{n+1} \leq \mu_n - C_1\varphi(\mu_{n+1}).$$

By Lemma 2.2, we conclude that $\Phi(x_n) \to \Phi^*$ with the estimate (3.13). □

Proposition 3.7. *Let $\{x_n\}$ be a sequence generated by Algorithm 3.1. Assume that there exists $c > 0$ such that*

$$\langle J(x) - J(y), x - y \rangle \geq c\max\{\|J(x) - J(y)\|^2, \|x - y\|^2\}, \forall x, y \in X \tag{3.18}$$

and assume that $L|t_n - \lambda_n| \leq \eta < c$. Then there exists a constant $C > 0$ such that the estimate

$$\Phi(x_n) - \Phi^* \leq \left(\frac{C}{\sum_{i=1}^n t_i}\right)^{\frac{1}{2}}$$

holds. If $0 < \bar{t} \leq t_n$ *in the Algorithm 3.1, then there exists a constant* $\bar{c} > 0$ *such that the following the estimate of the convergence rate holds:*

$$\Phi(x_n) - \Phi^* \leq \frac{\bar{c}}{n}.$$

Proof. Using

$$J(x_n) - J(x_{n+1}) - t_n\nabla g(x_n) - \lambda_n(\nabla g(x_{n+1}) - \nabla g(x_n)) \in t_n\partial f(x_{n+1}),$$

we have

$$J(x_n) - J(x_{n+1}) + (t_n - \lambda_n)(\nabla g(x_{n+1}) - \nabla g(x_n)) \in t_n\nabla g(x_{n+1}) + t_n\partial f(x_{n+1}) = t_n\partial\Phi(x_{n+1}).$$

Hence,

$$\Phi(x_{n+1}) - \Phi(x_n) \leq \left\langle \frac{J(x_n) - J(x_{n+1})}{t_n} + \frac{(t_n - \lambda_n)}{t_n}(\nabla g(x_{n+1}) - \nabla g(x_n)), x_{n+1} - x_n \right\rangle.$$

By condition (3.18), there exists $c > 0$, such that

$$\begin{aligned}
&\Phi(x_{n+1}) - \Phi(x_n) \\
\leq\ & \frac{1}{t_n}\left\langle J(x_n) - J(x_{n+1}), x_{n+1} - x_n\rangle + \frac{(t_n - \lambda_n)}{t_n}\langle\nabla g(x_{n+1}) - \nabla g(x_n), x_{n+1} - x_n \right\rangle \\
\leq\ & -\frac{c}{t_n}\|x_n - x_{n+1}\|^2 + |\frac{t_n - \lambda_n}{t_n}|\|\nabla g(x_{n+1}) - \nabla g(x_n)\|\|x_{n+1} - x_n\| \\
\leq\ & -(\frac{c}{t_n} - L|\frac{t_n - \lambda_n}{t_n}|)\|x_{n+1} - x_n\|^2. \qquad (3.19)
\end{aligned}$$

By Remark 3.1 and the definition of subdifferential , we have

$$\begin{aligned}
&\Phi(x_{n+1}) - \Phi(\bar{x}) \\
\leq\ & \left\langle \frac{J(x_n) - J(x_{n+1})}{t_n} + \frac{(t_n - \lambda_n)}{t_n}(\nabla g(x_{n+1}) - \nabla g(x_n)), x_{n+1} - \bar{x} \right\rangle \\
\leq\ & \frac{1}{t_n}\|J(x_n) - J(x_{n+1})\|\|x_{n+1} - \bar{x}\| + |\frac{t_n - \lambda_n}{t_n}|\|\nabla g(x_{n+1}) - \nabla g(x_n)\|\|x_{n+1} - \bar{x}\| \\
\leq\ & \frac{1}{ct_n}\|x_n - x_{n+1}\|\|x_{n+1} - \bar{x}\| + L|\frac{t_n - \lambda_n}{t_n}|\|x_{n+1} - x_n\|\|x_{n+1} - \bar{x}\| \\
\leq\ & (\frac{1}{ct_n} + L|\frac{t_n - \lambda_n}{t_n}|)(M + \|\bar{x}\|)\|x_n - x_{n+1}\|. \qquad (3.20)
\end{aligned}$$

Then by (3.19) and (3.20), there exists a constant $\lambda > 0$, such that

$$\Phi(x_{n+1}) - \Phi(x_n) \leq -\lambda t_n(\Phi(x_{n+1}) - \Phi(\bar{x}))^2.$$

Let us to denote $\mu_n = \Phi(x_n) - \Phi(\bar{x})$. Then

$$\mu_{n+1} \leq \mu_n - \lambda t_n\mu_{n+1}^2. \qquad (3.21)$$

We now use Lemma 2.2. In the case of (3.21), we have $\varphi(t) = t^2$, $\alpha_n = \lambda t_n$, $z(t) = t$ (up to a constant) and then we conclude that there exists a constant $C > 0$ such that

$$\mu_n \leq \left(\frac{C}{\sum_{i=1}^{n} t_i} \right)^{\frac{1}{2}}.$$

If $0 < \bar{t} \leq t_n$, then by (3.21) we have

$$\mu_{n+1} \leq \mu_n - \lambda t_{n+1} \mu_{n+1}^2 \leq \mu_n - \lambda \bar{t} \mu_{n+1}^2.$$

Then by Lemma 2.3, we obtain

$$\Phi(x_n) - \Phi^* \leq \frac{\bar{c}}{n}$$

with some constant $\bar{c} > 0$. □

Proposition 3.8. *Let $\{x_n\}$ be a sequence generated by Algorithm 3.1. Assume that there exists $c > 0$ such that*

$$\langle J(x) - J(y), x - y \rangle \geq c \max\{\|J(x) - J(y)\|^2, \|x - y\|^2\}, \forall x, y \in X. \tag{3.22}$$

If $0 < \bar{t} \leq t_n \leq \bar{\bar{t}}$ and $\lambda_n = t_n$ in the Algorithm 3.1, then there exists a constant $C > 0$ such that the following the estimate of the convergence rate holds:

$$\Phi(x_n) - \Phi^* \leq \left(\frac{C}{\ln(n+1)} \right)^{\frac{1}{2}}. \tag{3.23}$$

Proof. As in Proposition 3.7, we assert that $\Phi(x_n) \leq \Phi(x_1)$. Moreover,

$$\Phi(x_{n+1}) - \Phi(x_n) \leq \frac{1}{t_n} \langle J(x_n) - J(x_{n+1}), x_{n+1} - x_n \rangle \leq -\frac{c}{t_n} \|J(x_n) - J(x_{n+1})\|^2.$$

Hence,

$$(\Phi(x_{n+1}) - \Phi^*) - (\Phi(x_n) - \Phi^*) \leq -\frac{c}{t_n} \|J(x_n) - J(x_{n+1})\|^2. \tag{3.24}$$

Therefore, the sequence $\{\Phi(x_n) - \Phi^*\}$ has a limit and there exists a constant C_1 such that

$$\sum_{i=1}^{n} \|J(x_i) - J(x_{i+1})\|^2 \leq C_1.$$

Letting $\bar{x} \in \mathrm{Sol}(P)$. Assume that $x_n \neq \bar{x}$, otherwise the Algorithm 3.1 terminates at finite steps. We calculate

$$\begin{aligned} \|J(x_{n+1}) - J(\bar{x})\|^2 &\leq (\|J(x_n) - J(\bar{x})\| + \|J(x_n) - J(x_{n+1})\|)^2 \\ &\leq (\|J(x_1) - J(\bar{x})\| + \sum_{i=1}^{n} \|J(x_i) - J(x_{i+1})\|)^2 \\ &\leq 2\|J(x_1) - J(\bar{x})\|^2 + 2(\sum_{i=1}^{n} \|J(x_i) - J(x_{i+1})\|)^2. \end{aligned} \tag{3.25}$$

Since

$$(\sum_{i=1}^{n} \|J(x_i) - J(x_{i+1})\|)^2 \leq (n+1) \sum_{i=1}^{n} \|J(x_i) - J(x_{i+1})\|^2,$$

we deduce

$$\|J(x_{n+1}) - J(\bar{x})\|^2 \leq 2\|J(x_1) - J(\bar{x})\|^2 + 2C_1(n+1).$$

By making use of the inequalities

$$\Phi(x_{n+1}) - \Phi^* \leq \langle \partial\Phi(x_{n+1}), x_{n+1} - \bar{x} \rangle \leq \frac{1}{t_n}\|J(x_n) - J(x_{n+1})\|\|x_{n+1} - \bar{x}\|.$$

and (3.24), one gets

$$\begin{aligned}
\Phi(x_{n+1}) - \Phi(x_n) &\leq -ct_n \frac{(\Phi(x_{n+1}) - \Phi^*)^2}{\|x_{n+1} - \bar{x}\|^2} \\
&\leq -c^3 t_n \frac{(\Phi(x_{n+1}) - \Phi^*)^2}{\|J(x_{n+1}) - J(\bar{x})\|^2} \\
&\leq -\frac{c^3 t_n (\Phi(x_{n+1}) - \Phi^*)^2}{2\|J(x_1) - J(\bar{x})\|^2 + 2C_1(n+1)}.
\end{aligned} \tag{3.26}$$

It is not difficult to verify that there exists a constant $C_2 > 0$ such that

$$\frac{c^3 t_n}{2\|J(x_1) - J(\bar{x})\|^2 + 2C_1(n+1)} \geq \frac{C_2}{n+1}.$$

It is clear that

$$C_2 \leq \frac{c^3 t_n}{2\|J(x_1) - J(\bar{x})\|^2 + 2C_1}.$$

The following recurrence inequality is established from (3.26):

$$\Phi(x_{n+1}) - \Phi^* \leq \Phi(x_n) - \Phi^* - \frac{C_2}{n+1}(\Phi(x_{n+1}) - \Phi^*)^2.$$

This is the particular case of (2.6) with $\alpha_n = \frac{C_2}{n+1}$ and $\varphi(t) = t^2$. Thus, there exists a constant $C > 0$ such that

$$\Phi(x_n) - \Phi^* \leq \left(\frac{C}{\sum_{i=1}^{n} (i+1)^{-1}} \right)^{\frac{1}{2}}.$$

Since $\xi > \ln(\xi + 1),\ \forall \xi > 0$,

$$\sum_{i=1}^{n} \frac{1}{i} > \sum_{i=1}^{n} \ln(1 + \frac{1}{i}) = \ln \prod_{i=1}^{n} (1 + \frac{1}{i}) = \ln(n+1).$$

Hence, (3.23) holds and the proof is accomplished. □

Proposition 3.9. *Let $\{x_n\}$ be a sequence generated by Algorithm 3.1. Assume that there exists $c > 0$ such that*

$$\langle J(x) - J(y), x - y\rangle \geq c \max\{\|J(x) - J(y)\|^2, \|x - y\|^2\}, \forall x, y \in X. \tag{3.27}$$

If $0 < \bar{t} \leq t_n \leq \bar{\bar{t}} < \frac{c}{L}$ and $\sum_{n=1}^{\infty} \lambda_n < +\infty$, then, for every $k \geq 1$, one has

$$\min_{1\leq n\leq k} \Phi(x_{n+1}) - \Phi^* = O\left(\frac{1}{k^{\frac{1}{2}}}\right).$$

In particular, if $\lambda_n = 0$ for all n in the Algorithm 3.1, then there exists a constant $\bar{c} > 0$ such that the following the estimate of the convergence rate holds:

$$\Phi(x_n) - \Phi^* \leq \frac{\bar{c}}{n}.$$

Proof. Using

$$J(x_n) - J(x_{n+1}) - t_n\nabla g(x_n) - \lambda_n(\nabla g(x_{n+1}) - \nabla g(x_n)) \in t_n\partial f(x_{n+1}),$$

we have

$$J(x_n) - J(x_{n+1}) + (t_n - \lambda_n)(\nabla g(x_{n+1}) - \nabla g(x_n)) \in t_n\nabla g(x_{n+1}) + t_n\partial f(x_{n+1}) = t_n\partial\Phi(x_{n+1}).$$

Hence,

$$\Phi(x_{n+1}) - \Phi(x_n) \leq \langle\frac{J(x_n) - J(x_{n+1})}{t_n} + \frac{t_n - \lambda_n}{t_n}(\nabla g(x_{n+1}) - \nabla g(x_n)), x_{n+1} - x_n\rangle.$$

By condition (3.27), there exists $c > 0$, such that

$$\begin{aligned}
&\Phi(x_{n+1}) - \Phi(x_n)\\
\leq\ & \frac{1}{t_n}\langle J(x_n) - J(x_{n+1}), x_{n+1} - x_n\rangle + \langle\nabla g(x_{n+1}) - \nabla g(x_n), x_{n+1} - x_n\rangle\\
&-\frac{\lambda_n}{t_n}\langle\nabla g(x_{n+1}) - \nabla g(x_n), x_{n+1} - x_n\rangle\\
\leq\ & -\frac{c}{t_n}\|x_n - x_{n+1}\|^2 + L\|x_{n+1} - x_n\|^2 - \frac{\lambda_n}{t_n}\langle\nabla g(x_{n+1}) - \nabla g(x_n), x_{n+1} - x_n\rangle\\
\leq\ & (L - \frac{c}{t_n})\|x_n - x_{n+1}\|^2 - \frac{\lambda_n}{t_n}\langle\nabla g(x_{n+1}) - \nabla g(x_n), x_{n+1} - x_n\rangle.
\end{aligned} \tag{3.28}$$

By Remark 3.1 and the definition of subdifferential, we have

$$\begin{aligned}
&\Phi(x_{n+1}) - \Phi(\bar{x})\\
\leq\ & \langle\frac{J(x_n) - J(x_{n+1})}{t_n}, x_{n+1} - \bar{x}\rangle + \langle\nabla g(x_{n+1}) - \nabla g(x_n), x_{n+1} - \bar{x}\rangle\\
&-\frac{\lambda_n}{t_n}\langle\nabla g(x_{n+1}) - \nabla g(x_n), x_{n+1} - \bar{x}\rangle\\
\leq\ & \frac{1}{ct_n}\|x_n - x_{n+1}\|\|x_{n+1} - \bar{x}\| + L\|x_{n+1} - x_n\|\|x_{n+1} - \bar{x}\|\\
&-\frac{\lambda_n}{t_n}\langle\nabla g(x_{n+1}) - \nabla g(x_n), x_{n+1} - \bar{x}\rangle
\end{aligned}$$

$$\leq \quad (\frac{1}{ct_n} + L)\|x_n - x_{n+1}\|\|x_{n+1} - \bar{x}\| - \frac{\lambda_n}{t_n}\langle \nabla g(x_{n+1}) - \nabla g(x_n), x_{n+1} - \bar{x}\rangle.$$

Then by (3.28), we have

$$\begin{aligned}
& \Phi(x_{n+1}) - \Phi(\bar{x}) \\
\leq \; & (\frac{1}{ct_n} + L)\|x_n - x_{n+1}\|\|x_{n+1} - \bar{x}\| - \frac{\lambda_n}{t_n}\langle \nabla g(x_{n+1}) - \nabla g(x_n), x_{n+1} - \bar{x}\rangle \\
\leq \; & (\frac{1}{ct_n} + L)\left(\frac{\Phi(x_n) - \Phi(x_{n+1}) - \frac{\lambda_n}{t_n}\langle \nabla g(x_{n+1}) - \nabla g(x_n), x_{n+1} - x_n\rangle}{\frac{c}{t_n} - L}\right)^{\frac{1}{2}} \|x_{n+1} - \bar{x}\| \\
& - \frac{\lambda_n}{t_n}\langle \nabla g(x_{n+1}) - \nabla g(x_n), x_{n+1} - \bar{x}\rangle.
\end{aligned}$$

Since $\{x_n\}$ is bounded and $0 < \bar{t} \leq t_n \leq \bar{\bar{t}} < \frac{c}{L}$, $\sum_{n=1}^{\infty} \lambda_n < +\infty$, there exists $\theta > 0$ and $\alpha_n \geq 0$ with $\sum_{n=1}^{\infty} \alpha_n < +\infty$ such that

$$(\Phi(x_{n+1}) - \Phi(\bar{x}))^2 \leq \theta(\Phi(x_n) - \Phi(\bar{x})) - \theta(\Phi(x_{n+1}) - \Phi(\bar{x})) + \alpha_n. \tag{3.29}$$

Hence, we get that

$$\sum_{n=1}^{k}(\Phi(x_{n+1}) - \Phi(\bar{x}))^2 \leq \theta(\Phi(x_1) - \Phi(\bar{x})) - \theta(\Phi(x_{k+1}) - \Phi(\bar{x})) + \sum_{n=1}^{k} \alpha_n,$$

from which we obtain that

$$k \min_{1\leq n\leq k}(\Phi(x_{n+1}) - \Phi(\bar{x}))^2 \leq \theta(\Phi(x_1) - \Phi(\bar{x})) + \sum_{n=1}^{k} \alpha_n < +\infty.$$

This implies that

$$\min_{1\leq n\leq k} \Phi(x_{n+1}) - \Phi^* = O\left(\frac{1}{k^{\frac{1}{2}}}\right).$$

In particular, if $\lambda_n = 0$, then we can take $\alpha_n = 0$. Hence by Lemma 2.3 and (3.29), there exists $\bar{c} > 0$ such that

$$\Phi(x_n) - \Phi^* \leq \frac{\bar{c}}{n},$$

which proves the desired result. □

4. Numerical experiment

Let $X = \{x = (x_1, x_2) : \|x\| = (|x_1|^{1.5} + |x_2|^{1.5})^{\frac{1}{1.5}}\}$. Then we know that X is finite dimensional, but $\|\cdot\|$ is not an euclidean norm, and $J(x) = \|x\|^{0.5}(x_1|x_1|^{-0.5}, x_2|x_2|^{-0.5})$ (see [14]). For simplicity, consider the following Tikhonov-type optimization problem:

$$\min_{(x_1,x_2)\in X} \frac{\|(x_1, x_2) - (2, 3)\|^{1.5}}{1.5} + |x_1| + |x_2|. \tag{4.1}$$

Simple computations show that

$$\text{argmin}_{(x_1,x_2)\in X}\frac{\|(x_1,x_2)-(2,3)\|^{1.5}}{1.5}+|x_1|+|x_2|=\{(1,2)\}.$$

Let $t_n = 0.5,\ \lambda_n = 0.25$, take $x^0 = (0,0)$, then Algorithm 3.1 becomes

$$x^0=(0,0),\ (x_1^{n+1},x_2^{n+1})=(0.6x_1^n+0.4,0.6x_2^n+0.8).$$

We denote by Iter the number of iterations and Iter-Sequ the iterative sequence.

The data in Table 1 shows that Algorithm 3.1 after 21 iterations converges to the optimal solution within a 0.0001 error, and the function value converge to the optimal value after only 11 iterations.

Table 1. Numerical results for solving the problem (4.1).

Iter(n)	Iter-Sequ(x^n)	$\Phi(x^n)$	*Iter*(n)	Iter-Sequ(x^n)	$\Phi(x^n)$
0	(0, 0)	5.3497	11	(0.9960, 1.9927)	4.3333
1	(0.4000, 0.8000)	4.7247	12	(0.9978, 1.9956)	4.3333
2	(0.6400, 1.2800)	4.4812	13	(0.9987, 1.9974)	4.3333
3	(0.7840, 1.5680)	4.3884	14	(0.9992, 1.9984)	4.3333
4	(0.8704, 1.7408)	4.3536	15	(0.9995, 1.9991)	4.3333
5	(0.9222, 1.8445)	4.3407	16	(0.9997, 1.9994)	4.3333
6	(0.9533, 1.9067)	4.3360	17	(0.9998, 1.9997)	4.3333
7	(0.9720, 1.9440)	4.3343	18	(0.9999, 1.9998)	4.3333
8	(0.9832, 1.9664)	4.3337	19	(0.9999, 1.9999)	4.3333
9	(0.9899, 1.9798)	4.3335	20	(1.0000, 1.9999)	4.3333
10	(0.9940, 1.9879)	4.3334	21	(1.0000, 2.0000)	4.3333

Acknowledgments

Weibo Guan was partially supported by the Heilongjiang Provincial Natural Sciences Foundation (Grant No. LH2022A017). Wen Song was partially supported by the National Natural Sciences Foundation (Grant No. 11871182).

Conflict of interest

All authors declare no conflict of interest regarding this study.

References

1. P. L. Combettes, D. Dũng, B. C. Vũ, Dualization of signal recovery problems, *Set-Valued Var. Anal.*, **18** (2010), 373–404. https://doi.org/10.1007/s11228-010-0147-7

2. P. L. Combettes, Inconsistent signal feasibility problems: least-squares solutions in a product space, *IEEE Trans. Signal Process.*, **42** (1994), 2955–2966. https://doi.org/10.1109/78.330356

3. P. L. Combettes, V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Sim.*, **4** (2005), 1168–1200. https://doi.org/10.1137/050626090
4. P. L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.*, **16** (1979), 964–979. https://doi.org/10.1137/0716071
5. Y. Malitsky, M. K. Tam, A forward-backward splitting method for monotone inclusions without cocoercivity, *SIAM J. Optim.*, **30** (2020), 1451–1472. https://doi.org/10.1137/18M1207260
6. A. Padcharoen, D. Kitkuan, W. Kumam, P. Kumam, Tseng methods with inertial for solving inclusion problems and application to image deblurring and image recovery problems, *Comput. Math. Methods*, **3** (2021), 2577-7408 https://doi.org/10.1002/cmm4.1088
7. K. Bredies, A forward-backward splitting algorithm for the minimization of nonsmooth convex functionals in Banach space, *Inverse Probl.*, **25** (2008), 14228341. https://doi.org/10.1088/0266-5611/25/1/015005
8. W. B. Guan, W. Song, The generalized forward-backward splitting method for the minimization of the sum of two functions in Banach spaces, *Numer. Func. Anal. Opt.*, **36** (2015), 867–886. https://doi.org/10.1080/01630563.2015.1037591
9. W. B. Guan, W. Song, The forward-backward splitting method and its convergence rate for the minimization of the sum of two functions in Banach space. *Optim. Lett.*, **15** (2021), 1735–1758. https://doi.org/10.1007/s11590-020-01544-9
10. F. E. Browder, Fixed point therems for nonlinear semicontractive mappings in Banach spaces, *Arch. Rational Mech. Anal.*, **21** (1966), 259–269. https://doi.org/10.1007/BF00282247
11. J. P. Gossez, E. L. Dozo, Some geometric properties related to the fixed point theory for nonexpansive mappings, *Pac. J. Math.*, **40** (1972), 565–573. https://doi.org/10.2140/PJM.1972.40.565
12. I. Cioranescu, *Geometry of Banach spaces, duality mappings and nonlinear problems*, Dordrecht: Springer, 1990.
13. W. Takahashi, *Convex analysis and approximation of fixed points*, Japan: Yokohama Publishers, 2000.
14. Y. I. Alber, R. S. Burachik, A. N. Iusem, A proximal point method for nonsmooth convex optimization problems in Banach spaces, *Abstr. Appl. Anal.*, **2** (1996), 614871. https://doi.org/10.1155/S1085337597000298
15. Y. I. Alber, Metric and generalized projection operators in Banach spaces: properties and applications, *Mathematics*, **1** (1993), 1109–1120. https://doi.org/10.48550/arXiv.funct-an/9311001
16. Y. I. Albert, I. Ryazantseva, *Nonlinear ill-posed problems of monotone type*, Dordrecht: Springer, 2006. https://doi.org/10.1007/1-4020-4396-1
17. Y. Yao, N. Shahzad, Strong convergence of aproximal point algorithm with general errors, *Optim. Lett.*, **6** (2012), 621–628. https://doi.org/10.1007/s11590-011-0286-2
18. Y. I. Alber, S. Reich, D. Shoikhet, Iterative approximations of null points of uniformly accretive operators with estimates of convergence rate, *Communications in Applied Analysis*, **6** (2002), 89–104.

19. H. Attouch, M. O. Czarnecki, J. Peypouquet, Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities, *SIAM J. Optim.*, **21** (2011), 1251–1274. https://doi.org/10.1137/110820300

20. S. Kamimura, W. Takahashi, Strong convergence of a proximal type algorithm in a Banach space, *SIAM J. Optim.*, **13** (2002), 938–945. https://doi.org/10.1137/S105262340139611X

21. S. Sabach, S. Shtern, A first order method for solving convex bilevel optimization problems, *SIAM J. Optim.*, **27** (2017), 640–660. https://doi.org/10.1137/16M105592X

22. D. P. Bertsekas, *Nonlinear programming*, 2 Eds., MA: Athena Scientific, 1999.

23. M. Bounkhel, R. Al-Yusof, Proximal analysis in reflexive smooth Banach spaces, *Nonlinear Analysis*, **73** (2010), 1921–1939. https://doi.org/10.1016/j.na.2010.04.077