Convergence of adaptive operator extrapolation method for operator inclusions in Banach Spaces

Serhii Denysov, Vladimir Semenov

Taras Shevchenko National University of Kyiv, 64/13 Volodymyrska Street, Kyiv, 01161, Ukraine

Abstract

A novel iterative splitting algorithm for solving operator inclusions problem is considered in a real Banach space setting. The operator is a sum of the multivalued maximal monotone operator and the monotone Lipschitz continuous operator. The proposed algorithm is an adaptive modification of the "forward-reflected-backward algorithm" [14]. Step size update rule not require Lipschitz constant knowledge of the operator. Advantage of the proposed algorithm is a single calculation of the maximal monotone operator resolvent and value of the monotone Lipschitz continuous operator on each iterative step. Weak convergence of the method is proved for operator inclusions in 2-uniformly convex and uniformly smooth real Banach space.

Keywords

Maximal monotone operator, operator inclusion, variational inequality, splitting algorithm, convergence, adaptability, 2-uniformly convex Banach space, uniformly smooth Banach space

1. Introduction

Consider real Banach space E. Denote it's dual space as E^*. We study monotone operator inclusion:

$$\text{find } x \in E: \quad 0 \in (A+B)x, \tag{1}$$

where $A : E \to 2^{E^*}$ is multivalued maximal monotone operator, $B : E \to E^*$ is monotone, single-valued, and Lipschitz continuous operator.

Many actual problems can be written in the form of (1). Among them are variational inequalities and optimization problems from various fields of optimal control, inverse problem theory, machine learning, image processing, operations research, and mathematical physics [1–6]. Prominent example is the saddle problem that plays an important role in mathematical economics:

$$\min_{p \in P} \max_{q \in Q} F(p,q),$$

where $F : P \times Q \to R$ is a smooth convex-concave function, $P \subseteq R^n$, $Q \subseteq R^m$ – closed convex sets, which can be formulated as

$$\text{find } x \text{ such that } 0 \in (A+B)x,$$

where $x = (p,q) \in R^{n+m}$ and

$$Ax = \begin{pmatrix} N_P p \\ N_Q q \end{pmatrix}, \quad Bx = \begin{pmatrix} \nabla_1 F(p,q) \\ -\nabla_2 F(p,q) \end{pmatrix},$$

N_P (N_Q) is a normal cone operator for closed convex set P (Q) [1].

CMIS-2022: The Fifth International Workshop on Computer Modeling and Intelligent Systems, May 12, 2022, Zaporizhzhia, Ukraine
EMAIL: denisov.univ@gmail.com (S. Denysov); semenov.volodya@gmail.com (V. Semenov)
ORCID: 0000-0003-4397-4617 (S. Denysov); 0000-0002-3280-8245 (V. Semenov)

Research and development of algorithms for operator inclusions (1) and related problems is a rapidly growing field of applied modern nonlinear analysis [1–4, 7–25].

The most well-known and popular iterative method for solving monotone operator inclusions (1) in Hilbert space is the "forward-backward algorithm" (FBA) [1, 11, 12]

$$x_{n+1} = J_\lambda^A \left(x_n - \lambda B x_n \right),$$

where $J_\lambda^A = \left(I + \lambda A \right)^{-1}$ is the operator resolvent, $A : H \to 2^H$, $\lambda > 0$.

Note that the FBA scheme includes well-known gradient method and proximal method as special cases [1]. For inverse strongly monotone (cocoercive) operators $B : H \to H$ FBA method is weakly converging [1]. However, FBA may diverge for Lipschitz continuous monotone operators B.

The condition of the inverse strong monotonicity of the operator B is a rather strong assumption. To weaken it, Tseng [13] proposed the following modification of the FBA:

$$\begin{cases} y_n = J_\lambda^A \left(x_n - \lambda B x_n \right), \\ x_{n+1} = y_n - \lambda \left(B y_n - B x_n \right), \end{cases}$$

where $B : H \to H$ – monotone and Lipschitz continuous operator with constant $L > 0$ and $\lambda \in \left(0, L^{-1} \right)$. A main limitation of Tseng method is their two calls of B per iteration. Evolution of this idea resulted in the so-called "forward-reflected-backward algorithm" [14]:

$$x_{n+1} = J_\lambda^A \left(x_n - \lambda B x_n - \lambda \left(B x_n - B x_{n-1} \right) \right), \ \lambda \in \left(0, \frac{1}{2L} \right),$$

and related method [15]:

$$x_{n+1} = J_\lambda^A \left(x_n - \lambda B x_n \right) - \lambda \left(B x_n - B x_{n-1} \right), \ \lambda \in \left(0, \frac{1}{3L} \right).$$

A special case of this algorithm is the algorithm "optimistic gradient descent ascent", which is popular among specialists in the field of Machine Learning [14, 15].

These schemes are naturally called operator extrapolation schemes. Note that in the three above algorithms there is a requirement to know operator's Lipschitz constant – which is often unknown or difficult to estimate. To overcome these difficulties, adaptive rules for updating parameter $\lambda > 0$ on each step have been proposed [16].

Some progress has been achieved recently in the study of splitting algorithms for operator inclusions in Banach spaces [2, 17–25]. This progress is mostly related to careful use of theoretical results and concepts from Banach spaces geometry [2, 26-29]. Book [2] contains an extensive material on this topic.

In the article [17] for the solution of inclusions (1) in the 2-uniformly convex and uniformly smooth Banach space the next algorithm is proposed:

$$x_{n+1} = J_\lambda^A \circ J^{-1} \left(J x_n - \lambda B x_n \right), \tag{2}$$

where $J_\lambda^A = \left(J + \lambda A \right)^{-1} J$ is the resolvent of operator $A : E \to 2^{E^*}$, $\lambda > 0$, J is the normalized duality mapping from E to E^*. For inverse strongly monotone (cocoercive) operators $B : E \to E^*$ method (2) weakly converges [17]. Recently, Shehu [18] extended Tseng's result to 2-uniformly convex and uniformly smooth Banach spaces. He proposed the following weakly converging process for approximating the solution of inclusion (1):

$$\begin{cases} y_n = J_\lambda^A \circ J^{-1} \left(x_n - \lambda_n B x_n \right), \\ x_{n+1} = J^{-1} \left(J y_n - \lambda_n \left(B y_n - B x_n \right) \right), \end{cases} \tag{3}$$

where $\lambda_n > 0$ can be set using knowledge of Lipschitz constant of the operator B or calculated with linear search. Note that two values of operator B need to be calculated at the iteration step.

In this article we study a new splitting algorithm for solving operator inclusion (1) in Banach space. The algorithm is an adaptive modification of the well-known Malitsky–Tam "forward-reflected-backward algorithm" [14], where the step size update rule not require Lipschitz continuous constant knowledge for operator B. It's advantage is a single computation of maximal monotone operator A resolvent and B operator value on each iterative step. The method weak convergence theorem is proved for operator inclusions in 2-uniformly convex and uniformly smooth Banach space.

The article structure is the following. Section 2 provides the necessary information from the areas of geometry of Banach spaces and theory of monotonous operators. The proposed adaptive operator extrapolation algorithm is described in section 3. Proof of weak convergence is presented in section 4. Theoretical applications to nonlinear operator equations, convex minimization problems and variational inequalities are given in sections 5 and 6. Some results of numerical experiments are also presented in section 7.

2. Preliminaries

To formulate and prove our results about convergence of algorithms we need some concepts and important facts about the geometry of real Banach space [26–33].

Consider real Banach space E with norm $\|\cdot\|$. Space E^* is the dual space for E. Let $\langle x^*, x\rangle$ is a value of linear bounded mapping $x^* \in E^*$ on $x \in E$ (i.e. $\langle x^*, x\rangle = x^*(x)$). Let's $\|\cdot\|_*$ be dual norm in dual space E^*.

Let $S_E = \{x \in E : \|x\| = 1\}$. Space E is strictly convex, if $\forall\ x, y \in S_E$, $x \neq y$ we have $\left\|\frac{x+y}{2}\right\| < 1$ [26]. The modulus of convexity of Banach space E is defined as [27]

$$\delta_E(\varepsilon) = \inf\left\{1 - \left\|\frac{x+y}{2}\right\| :\ x, y \in S_E,\ \|x-y\| = \varepsilon\right\} \quad \forall \varepsilon \in (0, 2].$$

Space E is uniformly convex, if $\delta_E(\varepsilon) > 0$ $\forall\ \varepsilon \in (0, 2]$ [27]. Space E is 2-uniformly convex, if there exists such $c > 0$ that $\delta_E(\varepsilon) \geq c\varepsilon^2$ $\forall\ \varepsilon \in (0, 2]$ [27]. It is known that 2-uniform convexity implies uniform convexity. Also the uniform convexity of real Banach space implies its reflexivity [26].

A space E is smooth if the limit

$$\lim_{t \to 0} \frac{\|x + ty\| - \|x\|}{t} \tag{4}$$

exists for all $x, y \in S_E$ [26]. A space E is uniformly smooth if the limit (4) exists uniformly over $x, y \in S_E$ [26].

Convexity type and smoothness type of spaces E and E^* are in duality relationship [26, 27]: dual space E^* is strictly convex $\Rightarrow$ space E is smooth [26]; dual space E^* is smooth $\Rightarrow$ space E is strictly convex [26]; space E is uniformly convex $\Rightarrow$ dual space E^* is uniformly smooth [26]; space E is uniformly smooth $\Rightarrow$ dual space E^* is uniformly convex [27]. We can reverse the first two implications if the space E is reflexive [26].

Widely used functional spaces L_p ($1 < p \leq 2$) and real Hilbert spaces are uniformly smooth (spaces L_p are uniformly smooth for $p \in (1, \infty)$) and 2-uniformly convex [26–29].

Also recall [1, 28, 31, 32] that a multivalued operator $A : E \to 2^{E^*}$ is called monotone if $\forall x, y \in E$

$$\langle u - v, x - y\rangle \geq 0 \quad \forall u \in Ax, v \in Ay.$$

A monotone operator $A : E \to 2^{E^*}$ is called maximal monotone if for any monotone operator $B : E \to 2^{E^*}$ we have that $\Gamma(A) \subseteq \Gamma(B)$ implies $\Gamma(A) = \Gamma(B)$, where

$$\Gamma(A) = \{(x, u) \in E \times E^* :\ u \in Ax\}$$

is a graph of A [1, 28, 31].

Lemma 1 ([1, 28]). Let $A : E \to 2^{E^*}$ be a maximal monotone operator, $x \in E$, $u \in E^*$. Then

$$\langle u-v, x-y\rangle \ge 0 \quad \forall (y,v)\in\Gamma(A) \quad \Leftrightarrow \quad (x,u)\in\Gamma(A).$$

It is known that if $A:E\to 2^{E^*}$ is maximal monotone operator, $B:E\to E^*$ is Lipschitz continuous monotone operator, then $A+B$ is maximal monotone operator [28, 31].

Let us also recall [1] that operator $A:E\to E^*$ is called inverse strongly monotone (cocoercive) if there exists such a number $\alpha>0$ (the constant of inverse strong monotonicity) that

$$\langle Ax-Ay, x-y\rangle \ge \alpha\|Ax-Ay\|^2.$$

Inverse strongly monotone operator is Lipschitz continuous, but not every Lipschitz continuous operator is inverse strongly monotone.

Multivalued mapping $J:E\to 2^{E^*}$, which acts as

$$Jx=\left\{x^*\in E^*:\ \langle x^*,x\rangle = \|x\|^2=\|x^*\|_*^2\right\},$$

is called normalized duality mapping [30]. We also use the next facts [28, 30, 32, 33]: if Banach space E is smooth then operator J is single-valued [30]; if real Banach space E is strongly convex then operator J is one-to-one and strongly monotone [28]; if Banach space E is reflexive then operator J is onto [28]; if Banach space E is uniformly smooth then on all bounded subsets of E operator J is uniformly continuous [30].

Remark 1. For a Hilbert space $J=I$. Explicit form of operator J for Banach spaces ℓ_p, L_p, and W_p^m ($p\in(1,\infty)$) is provided in [28, 32].

Consider reflexive, strictly convex and smooth space E [26]. The maximal monotonicity of operator $A:E\to 2^{E^*}$ is equivalent to equality

$$R(J+\lambda A)=E^*$$

for all $\lambda>0$ [31]. For maximal monotone operator $A:E\to 2^{E^*}$ and $\lambda>0$ resolvent $J_\lambda^A:E\to E$ is defined as follows [31]

$$J_\lambda^A x=(J+\lambda A)^{-1}Jx,\quad x\in E,$$

where J is normalized duality mapping from E to E^*. It is known that

$$A^{-1}0=F\left(J_\lambda^A\right)=\left\{x\in E:\ J_\lambda^A x=x\right\} \quad \forall\lambda>0.$$

It is also known that the set $A^{-1}0$ is closed and convex [1, 31].

Consider smooth Banach space E [26]. Yakov Alber introduced the convenient real-valued functional [32]

$$D(x,y)=\|x\|^2-2\langle Jy,x\rangle+\|y\|^2 \quad \forall x,y\in E.$$

Definition of D implies a useful 3-point identity:

$$D(x,y)-D(x,z)-D(z,y)=2\langle Jz-Jy, x-z\rangle \quad \forall x,y,z\in E.$$

For strictly convex Banach space E and $x,y\in E$ [32]:

$$D(x,y)=0 \Leftrightarrow x=y.$$

Lemma 2 ([31]). Let E be a uniformly convex and uniformly smooth Banach space (x_n), (y_n) – bounded sequences of elements from Banach space E. Then

$$\|x_n-y_n\|\to 0 \quad \Leftrightarrow \quad \|Jx_n-Jy_n\|_*\to 0 \quad \Leftrightarrow \quad D(x_n,y_n)\to 0.$$

Lemma 3 ([24, 25]). Let E be a smooth Banach space and 2-uniformly convex. Then for some $\mu\ge 1$ we have:

$$D(x,y)\ge \frac{1}{\mu}\|x-y\|^2 \quad \forall x,y\in E.$$

Remark 2. For Banach spaces ℓ_p, L_p and W_p^m ($1 < p \le 2$) we have $\mu = \frac{1}{p-1}$ [29]. And for a Hilbert space inequality for Lemma 3 becomes identity.

3. Algorithm

Let real Banach space E be a uniformly smooth and 2-uniformly convex [27]. Let A be a multivalued operator acting from E into 2^{E^*}, and B an operator acting from E into E^*.

Consider the next operator inclusion problem:

$$\text{find } x \in E: \quad 0 \in (A+B)x, \tag{5}$$

Let us denote $(A+B)^{-1}0$ set of solutions of this operator inclusion.

Suppose that the next assumptions hold [20]:

- $A : E \to 2^{E^*}$ is a maximal monotone operator;
- $B : E \to E^*$ is a monotone and Lipschitz continuous operator with Lipschitz constant $L > 0$;
- Set $(A+B)^{-1}0$ is nonempty.

Operator inclusion (5) can be formulated as the problem of finding a fixed point:

$$\text{find } x \in E: \quad x = J_\lambda^A \circ J^{-1}(Jx - \lambda Bx), \tag{6}$$

where $\lambda > 0$. Formulation (6) is useful, as it leads to well-known simple algorithmic idea. Calculation scheme

$$x_{n+1} = J_\lambda^A \circ J^{-1}(Jx_n - \lambda Bx_n)$$

was studied in [17] for inverse strongly monotone operators $B : E \to E^*$. However, the scheme generally does not converge for Lipschitz continuous monotone operators. Let's use the idea of work [14] and consider modified scheme

$$x_{n+1} = J_\lambda^A \circ J^{-1}\left(Jx_n - \lambda Bx_n - \lambda(Bx_n - Bx_{n-1})\right)$$

with extrapolation term

$$-\lambda(Bx_n - Bx_{n-1}).$$

And let's use update rule for $\lambda > 0$ similar to one from [16] to exclude explicit use of Lipschitz constant of operator B.

We will assume that we know constant $\mu \ge 1$ from Lemma 3.

Algorithm 1.

Choose some $x_0 \in E$, $x_1 \in E$, $\tau \in \left(0, \frac{1}{2\mu}\right)$ and $\lambda_0, \lambda_1 > 0$. Set $n \leftarrow 1$.

1. Compute
$$x_{n+1} = J_{\lambda_n}^A \circ J^{-1}\left(Jx_n - \lambda_n Bx_n - \lambda_{n-1}(Bx_n - Bx_{n-1})\right).$$
2. If $x_{n-1} = x_n = x_{n+1}$, then $x_n \in (A+B)^{-1}0$, else return to 3.
3. Compute
$$\lambda_{n+1} = \begin{cases} \min\left\{\lambda_n, \tau \dfrac{\|x_{n+1} - x_n\|}{\|Bx_{n+1} - Bx_n\|_*}\right\}, & \text{if } Bx_{n+1} \ne Bx_n, \\ \lambda_n, & \text{otherwise.} \end{cases}$$

   Set $n \leftarrow n+1$ and return to 1.

Step size sequence (λ_n) which is created by rule on step 3 is non-increasing and bounded from below by

$$\min\left\{\lambda_1, \tau L^{-1}\right\}.$$

So, we have $\lim_{n\to\infty}\lambda_n > 0$.

Let us prove convergence result for proposed Algorithm 1.

4. Proof of convergence

The following lemma contains inequality which is crucial to proof weak convergence of adaptive operator extrapolation method (Algorithm 1).

Lemma 4. The next inequality holds for the sequence (x_n), generated by adaptive operator extrapolation method (Algorithm 1):

$$D(z,x_{n+1}) + 2\lambda_n \langle Bx_n - Bx_{n+1}, x_{n+1} - z\rangle + \tau\mu \frac{\lambda_n}{\lambda_{n+1}} D(x_{n+1}, x_n) \le$$

$$\le D(z,x_n) + 2\lambda_{n-1}\langle Bx_{n-1} - Bx_n, x_n - z\rangle + \tau\mu\frac{\lambda_{n-1}}{\lambda_n} D(x_n, x_{n-1}) -$$

$$-\left(1 - \tau\mu\frac{\lambda_{n-1}}{\lambda_n} - \tau\mu\frac{\lambda_n}{\lambda_{n+1}}\right) D(x_{n+1}, x_n),$$

where $z \in (A+B)^{-1}0$.

Proof. Let $z \in (A+B)^{-1}0$. Then

$$-Bz \in Az.$$

Equality $x_{n+1} = J^A_{\lambda_n} \circ J^{-1}\left(Jx_n - \lambda_n Bx_n - \lambda_{n-1}(Bx_n - Bx_{n-1})\right)$ can be formulated as

$$Jx_n - \lambda_n Bx_n - \lambda_{n-1}(Bx_n - Bx_{n-1}) - Jx_{n+1} \in \lambda_n Ax_{n+1}.$$

From monotonicity of A

$$\langle Jx_{n+1} + \lambda_n(Bx_n - Bz) + \lambda_{n-1}(Bx_n - Bx_{n-1}) - Jx_n, z - x_{n+1}\rangle \ge 0. \tag{7}$$

Let's write 3-point identity

$$D(z,x_n) - D(z,x_{n+1}) - D(x_{n+1},x_n) = 2\langle Jx_{n+1} - Jx_n, z - x_{n+1}\rangle. \tag{8}$$

Using (8) withing (7) we get

$$D(z,x_{n+1}) \le D(z,x_n) - D(x_{n+1},x_n) + 2\langle \lambda_n(Bx_n - Bz) + \lambda_{n-1}(Bx_n - Bx_{n-1}), z - x_{n+1}\rangle. \tag{9}$$

Using monotonicity of operator B. We have

$$\langle \lambda_n(Bx_n - Bz) + \lambda_{n-1}(Bx_n - Bx_{n-1}), z - x_{n+1}\rangle = \lambda_n\langle Bx_n - Bx_{n+1}, z - x_{n+1}\rangle +$$

$$+\lambda_{n-1}\langle Bx_n - Bx_{n-1}, z - x_{n+1}\rangle + \underbrace{\lambda_n\langle Bx_{n+1} - Bz, z - x_{n+1}\rangle}_{\le 0} \le$$

$$\le \lambda_n\langle Bx_n - Bx_{n+1}, z - x_{n+1}\rangle + \lambda_{n-1}\langle Bx_n - Bx_{n-1}, z - x_n\rangle +$$

$$+\lambda_{n-1}\langle Bx_n - Bx_{n-1}, x_n - x_{n+1}\rangle. \tag{10}$$

Using (10) withing (9) we get

$$D(z,x_{n+1}) \le D(z,x_n) - D(x_{n+1},x_n) + 2\lambda_n\langle Bx_n - Bx_{n+1}, z - x_{n+1}\rangle +$$

$$+2\lambda_{n-1}\langle Bx_n - Bx_{n-1}, z - x_n\rangle + 2\lambda_{n-1}\langle Bx_n - Bx_{n-1}, x_n - x_{n+1}\rangle. \tag{11}$$

Using the rule for λ_n recalculation, we can estimate term $2\lambda_{n-1}\langle Bx_n - Bx_{n-1}, x_n - x_{n+1}\rangle$ in (11) from above. We get

$$2\lambda_{n-1}\langle Bx_n - Bx_{n-1}, x_n - x_{n+1}\rangle \le 2\lambda_{n-1}\|Bx_n - Bx_{n-1}\|_* \|x_n - x_{n+1}\| \le 2\tau\frac{\lambda_{n-1}}{\lambda_n}\|x_n - x_{n-1}\|\|x_{n+1} - x_n\| \le$$

$$\le \tau\frac{\lambda_{n-1}}{\lambda_n}\|x_n - x_{n-1}\|^2 + \tau\frac{\lambda_{n-1}}{\lambda_n}\|x_n - x_{n+1}\|^2 \le \tau\mu\frac{\lambda_{n-1}}{\lambda_n}D(x_n, x_{n-1}) + \tau\mu\frac{\lambda_{n-1}}{\lambda_n}D(x_{n+1}, x_n).$$

So, we came to inequality

$$D(z, x_{n+1}) + 2\lambda_n\langle Bx_n - Bx_{n+1}, x_{n+1} - z\rangle + \tau\mu\frac{\lambda_n}{\lambda_{n+1}}D(x_{n+1}, x_n) \le$$

$$\le D(z, x_n) + 2\lambda_{n-1}\langle Bx_{n-1} - Bx_n, x_n - z\rangle + \tau\mu\frac{\lambda_{n-1}}{\lambda_n}D(x_n, x_{n-1}) - \left(1 - \tau\mu\frac{\lambda_{n-1}}{\lambda_n} - \tau\mu\frac{\lambda_n}{\lambda_{n+1}}\right)D(x_{n+1}, x_n),$$

which was required to prove.

The next theorem states our main result.

Theorem 1. Let Banach space E be a uniformly smooth and 2-uniformly convex, $A: E \to 2^{E^*}$ be a multivalued maximal monotone operator, $B: E \to E^*$ be a monotone and Lipschitz continuous operator. Suppose that J (normalized duality map) is sequentially weakly continuous and $(A+B)^{-1}0 \ne \varnothing$. Then we have weak convergence of sequence (x_n) generated by adaptive operator extrapolation method (Algorithm 1) to a point $z \in (A+B)^{-1}0$.

Proof. Let $z' \in (A+B)^{-1}0$. Denote

$$a_n = D(z', x_n) + 2\lambda_{n-1}\langle Bx_{n-1} - Bx_n, x_n - z'\rangle + \tau\mu\frac{\lambda_{n-1}}{\lambda_n}D(x_n, x_{n-1}),$$

$$b_n = \left(1 - \tau\mu\frac{\lambda_{n-1}}{\lambda_n} - \tau\mu\frac{\lambda_n}{\lambda_{n+1}}\right)D(x_{n+1}, x_n)$$

Inequality from Lemma 4 becomes

$$a_{n+1} \le a_n - b_n.$$

As $\lim_{n\to\infty}\lambda_n > 0$ exists, we have

$$1 - \tau\mu\frac{\lambda_{n-1}}{\lambda_n} - \tau\mu\frac{\lambda_n}{\lambda_{n+1}} \to 1 - 2\tau\mu \in (0,1) \text{ when } n \to \infty.$$

Let's show, that $a_n \ge 0$ for all big enough $n \in N$. We have

$$a_n = D(z', x_n) + 2\lambda_{n-1}\langle Bx_{n-1} - Bx_n, x_n - z'\rangle + \tau\mu\frac{\lambda_{n-1}}{\lambda_n}D(x_n, x_{n-1}) \ge$$

$$\ge \frac{1}{\mu}\|x_n - z'\|^2 - 2\lambda_{n-1}\|Bx_{n-1} - Bx_n\|_*\|x_n - z'\| + \tau\frac{\lambda_{n-1}}{\lambda_n}\|x_{n-1} - x_n\|^2 \ge$$

$$\ge \frac{1}{\mu}\|x_n - z'\|^2 - 2\tau\frac{\lambda_{n-1}}{\lambda_n}\|x_n - x_{n-1}\|\|x_n - z'\| + \tau\frac{\lambda_{n-1}}{\lambda_n}\|x_{n-1} - x_n\|^2 \ge \left(\frac{1}{\mu} - \tau\frac{\lambda_{n-1}}{\lambda_n}\right)\|x_n - z'\|^2.$$

We can find such $n_0 \in N$ that

$$\frac{1}{\mu} - \tau\frac{\lambda_{n-1}}{\lambda_n} > 0 \text{ for all } n \ge n_0,$$

so $a_n \ge 0$ for all $n \ge n_0$.

Now we can conclude that the next limit exists

$$\lim_{n\to\infty}\left(D(z', x_n) + 2\lambda_{n-1}\langle Bx_{n-1} - Bx_n, x_n - z'\rangle + \tau\mu\frac{\lambda_{n-1}}{\lambda_n}D(x_n, x_{n-1})\right),$$

and

$$\sum_{n=1}^{\infty}\left(1-\tau\mu\frac{\lambda_{n-1}}{\lambda_n}-\tau\mu\frac{\lambda_n}{\lambda_{n+1}}\right)D\left(x_{n+1},x_n\right)<+\infty .$$

So the generated sequence (x_n) is bounded. Also we have

$$\lim_{n\to\infty}\phi\left(x_{n+1},x_n\right)=\lim_{n\to\infty}\left\|x_{n+1}-x_n\right\|=0 .$$

From the fact

$$\lim_{n\to\infty}\left(2\lambda_{n-1}\left\langle Bx_{n-1}-Bx_n,x_n-z'\right\rangle+\tau\mu\frac{\lambda_{n-1}}{\lambda_n}D\left(x_n,x_{n-1}\right)\right)=0 ,$$

we get convergence of sequences $\left(\phi\left(z',x_n\right)\right)$ for all $z'\in\left(A+B\right)^{-1}0$.

Let us show that all weak partial limits of (x_n) sequence belong to set $\left(A+B\right)^{-1}0$. Consider a subsequence $\left(x_{n_k}\right)$ that weakly converges to some point $z\in E$. Let's show that

$$z\in\left(A+B\right)^{-1}0 .$$

If we take any point $(y,v)\in\Gamma\left(A+B\right)$ then $v-By\in Ay$. We have

$$Jx_{n_k}-\lambda_{n_k}Bx_{n_k}-\lambda_{n_k-1}\left(Bx_{n_k}-Bx_{n_k-1}\right)-Jx_{n_k+1}\in\lambda_{n_k}Ax_{n_k+1} .$$

Using monotonicity of A operator, we get

$$\left\langle\lambda_{n_k}v+Jx_{n_k+1}+\lambda_{n_k}\left(Bx_{n_k}-By\right)+\lambda_{n_k-1}\left(Bx_{n_k}-Bx_{n_k-1}\right)-Jx_{n_k},y-x_{n_k+1}\right\rangle\geq 0 .$$

And then, using monotonicity of B operator, we can obtain the following estimation

$$\begin{aligned}
&\left\langle v,y-x_{n_k}\right\rangle+\left\langle v,x_{n_k}-x_{n_k+1}\right\rangle=\\
&=\left\langle v,y-x_{n_k+1}\right\rangle\geq\left\langle By+\frac{1}{\lambda_{n_k}}\left(Jx_{n_k}-Jx_{n_k+1}-\lambda_{n_k}Bx_{n_k}-\lambda_{n_k-1}\left(Bx_{n_k}-Bx_{n_k-1}\right)\right),y-x_{n_k+1}\right\rangle=\\
&=\frac{1}{\lambda_{n_k}}\left\langle Jx_{n_k}-Jx_{n_k+1},y-x_{n_k+1}\right\rangle-\frac{\lambda_{n_k-1}}{\lambda_{n_k}}\left\langle Bx_{n_k}-Bx_{n_k-1},y-x_{n_k+1}\right\rangle+\\
&\quad+\left\langle By-Bx_{n_k+1},y-x_{n_k+1}\right\rangle+\left\langle Bx_{n_k+1}-Bx_{n_k},y-x_{n_k+1}\right\rangle\geq\\
&\geq\frac{1}{\lambda_{n_k}}\left\langle Jx_{n_k}-Jx_{n_k+1},y-x_{n_k+1}\right\rangle-\frac{\lambda_{n_k-1}}{\lambda_{n_k}}\left\langle Bx_{n_k}-Bx_{n_k-1},y-x_{n_k+1}\right\rangle+\left\langle Bx_{n_k+1}-Bx_{n_k},y-x_{n_k+1}\right\rangle .
\end{aligned}$$

From

$$\lim_{n\to\infty}\left\|x_n-x_{n-1}\right\|=0 ,$$

and the Lipschitz continuity of B we have

$$\lim_{n\to\infty}\left\|Bx_n-Bx_{n-1}\right\|_*=0 .$$

Due to the uniform continuity on bounded sets of the J [30], we obtain

$$\lim_{n\to\infty}\left\|Jx_n-Jx_{n+1}\right\|_*=0 .$$

Thus

$$\left\langle v,y-z\right\rangle=\lim_{k\to\infty}\left\langle v,y-x_{n_k}\right\rangle\geq 0 .$$

The maximal monotonicity of operator $A+B$ and arbitrariness of $(y,v)\in\Gamma\left(A+B\right)$ imply $z\in\left(A+B\right)^{-1}0$ (Lemma 1).

We need to prove that the sequence (x_n) weakly converges to point z. Let's argue by contradiction. Let there exist a subsequence $\left(x_{m_k}\right)$ that weakly converges to z', $z\neq z'$. Obviously that $z'\in\left(A+B\right)^{-1}0$. We have the equality

$$2\langle Jx_n, z - z' \rangle = D(z', x_n) - D(z, x_n) + \|z\|^2 - \|z'\|^2 .$$

So the next limit exists

$$\lim_{n\to\infty} \langle Jx_n, z - z' \rangle .$$

Due to the sequential weak continuity of the normalized duality mapping J , we obtain

$$\langle Jz, z - z' \rangle = \lim_{k\to\infty} \langle Jx_{n_k}, z - z' \rangle = \lim_{k\to\infty} \langle Jx_{m_k}, z - z' \rangle = \langle Jz', z - z' \rangle ,$$

so

$$\langle Jz - Jz', z - z' \rangle = 0 .$$

As a result we get the contradiction – $z = z'$.

5. Algorithm variants

For completeness, let us formulate a modification of the proposed algorithm with a fixed step size parameter $\lambda > 0$.

Algorithm 2.

Select some points $x_0 \in E$, $x_1 \in E$, and $\lambda \in \left(0, 0.5 \frac{1}{\mu} L^{-1} \right)$. Set $n \leftarrow 1$.

1. Compute

$$x_{n+1} = J_\lambda^A \circ J^{-1} \left(Jx_n - 2\lambda Bx_n + \lambda Bx_{n-1} \right).$$

2. If $x_{n-1} = x_n = x_{n+1}$, then $x_n \in (A + B)^{-1} 0$. Else set $n \leftarrow n + 1$ and return to 1.

Consider the problem of finding the zero of a nonlinear monotone Lipschitz continuous operator $B : E \to E^*$:

$$\text{find } x \in E: \quad Bx = 0 .$$

For such case Algorithm 1 becomes

Algorithm 3.

Choose some $x_0 \in E$, $x_1 \in E$, $\tau \in \left(0, (2\mu)^{-1} \right)$ and $\lambda_0, \lambda_1 > 0$. Set $n \leftarrow 1$.

1. Compute

$$x_{n+1} = J^{-1} \left(Jx_n - \lambda_n Bx_n - \lambda_{n-1} \left(Bx_n - Bx_{n-1} \right) \right).$$

2. If $x_{n-1} = x_n = x_{n+1}$, then $x_n \in B^{-1} 0$. Else go to 3.
3. Compute

$$\lambda_{n+1} = \begin{cases} \min\left\{ \lambda_n, \tau \dfrac{\|x_{n+1} - x_n\|}{\|Bx_{n+1} - Bx_n\|_*} \right\}, & \text{if } Bx_{n+1} \neq Bx_n, \\ \lambda_n, & \text{otherwise.} \end{cases}$$

Set $n \leftarrow n + 1$ and return to 1.

Weak convergence of Algorithm 3 follows from Theorem 1.

Theorem 2. Let Banach space E be uniformly smooth and 2-uniformly convex, $B : E \to E^*$ – monotone and Lipschitz continuous operator, $B^{-1} 0 \neq \varnothing$. If J (normalized duality mapping) is sequentially weakly continuous, then the sequence (x_n) converges weakly to some point $z \in B^{-1} 0$.

Remark 3. In [17], the following algorithm was proposed to find the zero of a reverse strongly monotone operator $B : E \to E^*$:

$$x_{n+1} = J^{-1} \left(Jx_n - \lambda_n Bx_n \right), \ x_1 \in E ,$$

where $\lambda_n \in [a,b] \subseteq \left(0, \frac{\alpha}{\mu}\right)$, $\alpha > 0$ – the constant of inverse strong monotonicity of operator B. This algorithm generally does not converge for Lipschitz continuous monotone operators, but it converges weakly ergodic.

Using Theorem 2, let us consider the problem of minimizing a convex continuously Frechet differentiable functional

$$f(x) \to \min_{x \in E} . \tag{12}$$

We can formulate variant of Algorithm 3 for (12), which is a smooth convex minimization problem.

Algorithm 4.

Choose some $x_0 \in E$, $x_1 \in E$, $\tau \in \left(0, \frac{1}{2\mu}\right)$ and $\lambda_0, \lambda_1 > 0$. Set $n \leftarrow 1$.

1. Compute

$$x_{n+1} = J^{-1}\left(Jx_n - \lambda_n \nabla f(x_n) - \lambda_{n-1}\left(\nabla f(x_n) - \nabla f(x_{n-1})\right)\right).$$

2. If $x_{n-1} = x_n = x_{n+1}$, then $x_n \in \arg\min f$. Else return to 3.
3. Compute

$$\lambda_{n+1} = \begin{cases} \min\left\{\lambda_n, \tau \dfrac{\|x_{n+1} - x_n\|}{\|\nabla f(x_{n+1}) - \nabla f(x_n)\|_*}\right\}, & \text{if } \nabla f(x_{n+1}) \neq \nabla f(x_n), \\ \lambda_n, & \text{otherwise.} \end{cases}$$

Set $n \leftarrow n+1$ and return to 1.

For this variant of the algorithm, from Theorem 2 we get

Theorem 3. Let Banach space E be uniformly smooth and 2-uniformly convex, $f : E \to R$ – convex continuously Frechet differentiable functional with Lipschitz continuous derivative, and $\arg\min f$ is non-empty. Assume that J is sequentially weakly continuous. Then sequence (x_n) generated by Algorithm 4 converges weakly to a point $z \in \arg\min f$.

6. Application to variational inequalities

Consider real Hilbert space H. Let C is a non-empty, convex and closed subset of space H, $B : H \to H$ is a monotone and Lipschitz continuous operator. Consider the next variational inequality problem:

$$\text{find } x \in C \text{ such that } (Bx, y - x) \geq 0 \quad \forall y \in C, \tag{13}$$

Let $VI(B,C)$ be a solution set of problem (13). Variational inequality (13) is equivalent to the operator inclusion [1]

$$\text{find } x \in H \text{ such that } \quad 0 \in (A + B)x,$$

where $A = N_C$ is a normal cone for convex and closed set C, i.e.

$$N_C x = \begin{cases} \{w \in H : (w, y - x) \leq 0 \ \ \forall y \in C\}, & x \in C, \\ \varnothing, & \text{otherwise.} \end{cases}$$

It is known that

$$J_\lambda^A = (I + \lambda A)^{-1} = (I + \lambda N_C)^{-1} = P_C,$$

where P_C is projection operator onto the set C [1].

For variational inequality problem (13), adaptive operator extrapolation method (Algorithm 1) takes the following form:

Algorithm 5.

Choose some $x_1 \in H$, $x_2 \in H$, $\tau \in (0, 0.5)$ and $\mu_1, \mu_2 > 0$. Set $n \leftarrow 2$.

1. Compute
$$x_{n+1} = P_C\left(x_n - \mu_n B x_n - \mu_{n-1}\left(B x_n - B x_{n-1}\right)\right).$$
2. If $x_{n-1} = x_n = x_{n+1}$, then STOP and $x_n \in VI(B, C)$. Else go to 3.
3. Compute
$$\mu_{n+1} = \begin{cases} \min\left\{\mu_n, \tau \dfrac{\|x_{n+1} - x_n\|}{\|Bx_{n+1} - Bx_n\|}\right\}, & \text{if } Bx_{n+1} \neq Bx_n, \\ \mu_n, & \text{otherwise.} \end{cases}$$

Set $n \leftarrow n+1$ and return to 1.

For variational inequality case, from Theorem 1 we get

Theorem 4. Let H be a real Hilbert space, C is a non-empty, convex and closed subset of H, $B : H \to H$ is a monotone Lipschitz continuous operator, $VI(B, C) \neq \varnothing$. Then the sequence (x_n) generated by Algorithm 5 converges weakly to a point $z \in VI(B, C)$.

7. Numerical experiments

The numerical experiments are performed in Python 3.8.5 with NumPy 1.19 on a 64-bit PC with an Intel Core i7-1065G7 1.3 – 3.9GHz and 16GB RAM. As the test example we consider a toy variational inequality with pseudo-monotone operator.

Example. Let
$$C = \left\{x \in [-5, 5]^3 : x_1 + x_2 + x_3 = 0\right\} \subseteq R^3,$$
and operator $B : R^3 \to R^3$ be defined as
$$Bx = \left(e^{-\|x\|^2} + 0{,}2\right)\begin{pmatrix} 2 & 0 & -2 \\ 0 & 3 & 0 \\ -2 & 0 & 4 \end{pmatrix} x$$

The variational inequality problem of finding $x \in C$: $(Bx, y - x) \geq 0 \ \forall y \in C$ has unique solution $x^* = (0, 0, 0)$. Also, Lipschitz constant for the operator B is known – $L = 10.136$. We compare adaptive and non-adaptive (marked as “lip” on figures) variants of “Extrapolation from the Past” algorithm [9] and Algorithm 5. For stopping criteria and error estimation we use Euclidian distance to known solution $D_n = \|x_n - x^*\|$. For this example, we use $x_0 = (-4, 3, 5)$ as starting point, $\lambda = 0.9(\sqrt{2} - 1)/L$ for non-adaptive version of “Extrapolation from the Past” and $\lambda = 0.9/2L$ for non-adaptive version of Algorithm 5. Also, we use $\tau = 0.3$ and $\tau = 0.45$ (nearly maximal feasible values) for adaptive versions of “Extrapolation from the Past” and Algorithm 5 accordingly.

Time measurements below are got by averaging 100 runs for each algorithm.

Table 1

Time to reach desired error rate D_n, seconds

error\algorithm	Extra Past (lip)	Extra Past	Alg. 5 (lip)	Alg. 5
$\varepsilon = 1 \cdot 10^{-10}$	0.0308	0.0174	0.0181	0.0087
$\varepsilon = 1 \cdot 10^{-13}$	0.0419	0.0251	0.0245	0.0129
$\varepsilon = 1 \cdot 10^{-16}$	0.0555	0.0352	0.0331	0.0182

As we see from Table 1, for this problem Algorithm 5 outperforms other variants.

On the figure below we can see convergence behavior.

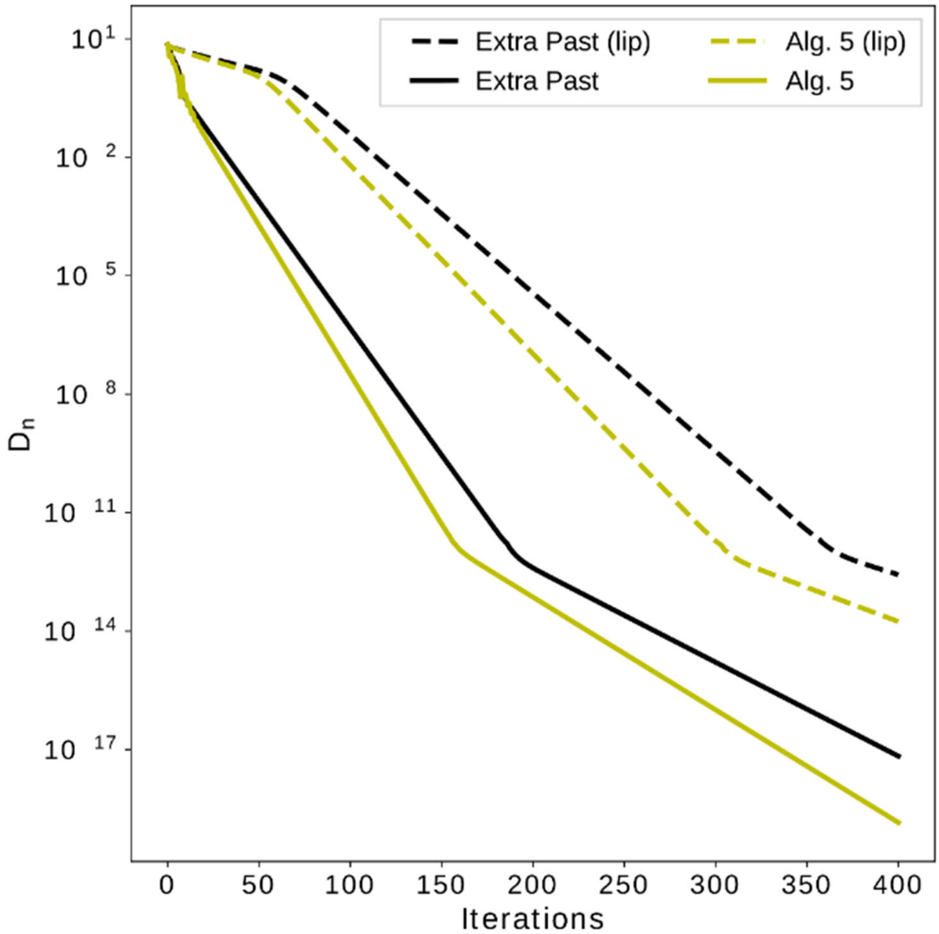

Figure 1: Convergence in terms of iterations

As it can be seen, both adaptive algorithms behave very closely. But it should be noted that Algorithm 5 has only one projection on each iteration instead of two for “Extrapolation from the Past” algorithm [9]

$$\begin{cases} y_n = P_C\left(x_n - \mu_n B y_{n-1}\right), \\ x_{n+1} = P_C\left(x_n - \mu_n B y_n\right). \end{cases}$$

8. Conclusions

In this paper new iterative splitting algorithm for solving an operator inclusion with the sum of a maximal monotone operator and a monotone Lipschitz continuous operator in a real Banach space is investigated.

Algorithm 1 is an improvement of the well-known “forward-reflected-backward algorithm” of Malitsky–Tam [14] with adaptive step size, where the step update rule does not require a priori knowledge of the Lipschitz constant of operator B [16].

The algorithm advantage is a single computation of the resolvent of the maximal monotone operator A and the value of the monotone Lipschitz continuous operator B at the iteration step.

Method weak convergence theorem is proved for operator inclusions in Banach space with 2-uniform convexity and uniform smoothness [27]. Theoretical applications to operator equations, convex minimization problems, and variational inequalities are presented.

An interesting challenge for the future is the development of a strongly convergent modification for Algorithm 1.

In connection with the study we will point out two topical issues. First, all results are obtained for the class 2-uniformly convex and uniformly smooth real Banach spaces [27], which does not contain important for applications spaces L_p and W_p^m ($2 < p < +\infty$). It is highly desirable to get rid of this limitation. Second, fast and robust algorithms for computing the resolvent for a wide range of maximal monotone operators are needed to effectively apply algorithms for nonlinear problems in Banach spaces.

An interesting question is the study of the behavior of Algorithm 1 in the situation $A = I$. Namely, the question of asymptotic behavior of $\|Bx_n\|_*$. Note that an estimate $\|Bx_n\|_* = O\left(\frac{1}{\sqrt{n}}\right)$ is theoretically possible.

Acknowledgements

This work was supported by the MES of Ukraine (project "Computational algorithms and optimization for artificial intelligence, medicine and defense", 0122U002026).

References

[1] H. H. Bauschke, P. L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Berlin; Heidelberg; New York, Springer, 2017.

[2] Y. Alber, I. Ryazantseva, Nonlinear Ill Posed Problems of Monotone Type, Springer, Dordrecht, 2006.

[3] H. Raguet, J. Fadili, G. Peyre, A generalized forward-backward splitting, SIAM J. Imaging Sci. 6 (2013) 1199–1226

[4] D. R. Luke, C. Charitha, R. Shefi, Y. Malitsky, Efficient, Quantitative Numerical Methods for Statistical Image Deconvolution and Denoising. In: Salditt T., Egner A., Luke D.R. (eds.) Nanoscale Photonic Imaging. Topics in Applied Physics, vol 134. Springer, Cham, 2020, pp. 313–338.

[5] R. Tibshirani, Regression shrinkage and selection via the lasso, J. Royal Statist. Soc. 58 (1996) 267–288.

[6] S. I. Lyashko, D. A. Klyushin, D. A. Nomirovsky, V. V. Semenov, Identification of age-structured contamination sources in ground water, in: R. Boucekkine, N. Hritonenko, Y. Yatsenko, Y. (Eds.), Optimal Control of Age-Structured Populations in Economy, Demography, and the Environment, Routledge, London–New York, 2013, pp. 277–292.

[7] V. V. Semenov, A Strongly Convergent Splitting Method for Systems of Operator Inclusions with Monotone Operators, Journal of Automation and Information Sciences 46(5) (2014) 45–56. https://doi.org/10.1615/JAutomatInfScien.v46.i5.40.

[8] D. A. Nomirovskii, B. V. Rublyov, V. V. Semenov, Convergence of Two-Stage Method with Bregman Divergence for Solving Variational Inequalities, Cybernetics and Systems Analysis 55 (2019) 359–368. doi:10.1007/s10559-019-00142-7.

[9] L. Chabak, V. Semenov, Y. Vedel, A New Non-Euclidean Proximal Method for Equilibrium Problems, In: O. Chertov et al. (Eds.), Recent Developments in Data Science and Intelligent Analysis of Information, volume 836 of Advances in Intelligent Systems and Computing, Springer, Cham, 2019, pp. 50–58. doi:10.1007/978-3-319-97885-7_6.

[10] V. V. Semenov, Modified Extragradient Method with Bregman Divergence for Variational Inequalities, Journal of Automation and Information Sciences 50(8) (2018) 26–37. doi:10.1615/JAutomatInfScien.v50.i8.30.

[11] P. L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal. 16 (1979) 964–979.

[12] G. B. Passty, Ergodic Convergence to a Zero of the Sum of Monotone Operators in Hilbert Spaces, Journal of Mathematical Analysis and Applications 72 (1979) 383–390.

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

[14] Y. Malitsky, M. K. Tam, A Forward-Backward Splitting Method for Monotone Inclusions Without Cocoercivity, SIAM Journal on Optimization 30 (2020) 1451–1472. doi:10.1137/18M1207260.
[15] E. R. Csetnek, Y. Malitsky, M. K. Tam, Shadow Douglas-Rachford Splitting for Monotone Inclusions, Appl Math Optim. 80 (2019) 665–678. doi:10.1007/s00245-019-09597-8.
[16] S. Denisov, V. Semenov, Convergence of Adaptive Forward-Reflected-Backward Algorithm for Solving Variational Inequalities, in: CEUR Workshop Proceedings, vol. 3106, 2021, pp. 116–127. URL: http://ceur-ws.org/Vol-3106/Paper_11.pdf.
[17] H. Iiduka, W. Takahashi, Weak convergence of a projection algorithm for variational inequalities in a Banach space, Journal of Mathematical Analysis and Applications 339(1) (2008) 668–679. https://doi.org/10.1016/j.jmaa.2007.07.019.
[18] Y. Shehu, Convergence Results of Forward-Backward Algorithms for Sum of Monotone Operators in Banach Spaces, Results Math. 74 (2019). doi:10.1007/s00025-019-1061-4.
[19] Y. Shehu, Single projection algorithm for variational inequalities in Banach spaces with application to contact problem, Acta Math. Sci. 40 (2020) 1045–1063. doi:10.1007/s10473-020-0412-2.
[20] J. Yang, P. Cholamjiak, P. Sunthrayuth, Modified Tseng's splitting algorithms for the sum of two monotone operators in Banach spaces, AIMS Mathematics 6(5) (2021) 4873–4900. doi:10.3934/math.2021286.
[21] P. Cholamjiak, Y. Shehu, Inertial forward-backward splitting method in Banach spaces with application to compressed sensing, Appl. Math. 64 (2019) 409–435.
[22] H. A. Abass, C. Izuchukwu, O. T. Mewomo, Q. L. Dong, Strong convergence of an inertial forward-backward splitting method for accretive operators in real Banach space, Fixed Point Theory 21 (2020) 397–412.
[23] S. Takahashi, W. Takahashi, Split common null point problem and shrinking projection method for generalized resolvents in two Banach spaces, J. Nonlinear Convex Anal. 17 (2016) 2171–2182.
[24] K. Aoyama, F. Kohsaka, Strongly relatively nonexpansive sequences generated by firmly nonexpansive-like mappings, Fixed Point Theory Appl. 95 (2014). doi:10.1186/1687-1812-2014-95.
[25] T. Bonesky, K. S. Kazimierski, P. Maass, F. Schopfer, T. Schuster, Minimization of Tikhonov Functionals in Banach Spaces, Abstr. Appl. Anal. (2008) Article ID 192679. https://doi.org/10.1155/2008/192679.
[26] J. Diestel, Geometry of Banach Spaces, Springer-Verlag, Berlin–Heidelberg, 1975.
[27] B. Beauzamy, Introduction to Banach Spaces and Their Geometry, North-Holland, Amsterdam, 1985.
[28] I. Cioranescu, Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems, Kluwer Academic, Dordrecht, 1990.
[29] Z. B. Xu, G. F. Roach, Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces, Journal of Mathematical Analysis and Applications 157(1) (1991) 189–210.
[30] M. M. Vainberg, Variational Method and Method of Monotone Operators in the Theory of Nonlinear Equations, Wiley, New York, 1974.
[31] V. Barbu, Nonlinear Semigroups and Differential Equations in Banach Spaces, Springer, Netherlands, 1976.
[32] Y. I. Alber, Metric and generalized projection operators in Banach spaces: properties and applications. in: Theory and Applications of Nonlinear Operators of Accretive and Monotone Type, vol. 178, Dekker, New York, 1996, pp. 15–50.
[33] R. P. Agarwal, D. O'Regan, D. R. Sahu, Fixed Point Theory for Lipschitzian-type Mappings with Applications, Springer, New York, 2009.