

Theoretical Bound of the Complexity of Some Extragradient-Type Algorithms for Variational Inequalities in Banach Spaces

Serhii Denysov and Vladimir Semenov

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

Abstract

The work presents a study of three new extragradient-type algorithms for solving variational inequalities in a Banach space. Two algorithms are natural modifications of Tseng method and “Extrapolation from the Past” method for problems in Banach spaces, based on the generalized Alber projection. The third algorithm, called the operator extrapolation method, is a variant of forward-reflected-backward algorithm, where the generalized Alber projection is also used instead of the metric projection onto the admissible set. Advantage of the latter algorithm is only one calculation of the operator value and the generalized projection onto the feasible set on each iteration. For variational inequalities with monotone Lipschitz operators acting in a 2-uniformly convex and uniformly smooth Banach space $O\left(\frac{1}{\varepsilon}\right)$ estimates for the complexity in terms of the gap function are proved.

Keywords ¹

variational inequality, monotone operator, extragradient algorithm, Extrapolation from the Past, gap function, complexity, 2-uniformly convex Banach space, uniformly smooth Banach space

1. Introduction

Areas of operations research, data analysis, and mathematical physics produces many problems, which can be written in the form of variational inequalities [1–5]. Prominent example is the saddle problem that plays an important role in mathematical economics:

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

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

$$\text{find } \bar{x} \in C : \langle A\bar{x}, x - \bar{x} \rangle \geq 0 \quad \forall x \in C,$$

where $x = (p, q) \in R^{n+m}$, $C = P \times Q \subseteq R^{n+m}$, and mapping $A: P \times Q \rightarrow R^{n+m}$ has the next form:

$$Ax = \begin{pmatrix} \nabla_1 L(p, q) \\ -\nabla_2 L(p, q) \end{pmatrix}.$$

Let's recall another important example related to ranking and search on the Web. The problem of finding the PageRank vector can be rewritten as follows

$$\min_{x \in \Delta^n} \|Px - x\|_\infty,$$

where P is an $n \times n$ column-stochastic matrix, $\Delta^n = \left\{ x \in R^n : \sum_{i=1}^n x_i = 1, x_i \geq 0 \right\}$, and $\|p\|_\infty = \max_{i=1, \dots, n} |p_i|$.

And the minimization problem can be transformed into a saddle point problem

$$\min_{x \in \Delta^n} \max_{\|y\|_\infty \leq 1} (y, Px - x),$$

Information Technology and Implementation (IT&I-2021), December 01–03, 2021, Kyiv, Ukraine

EMAIL: denisov.univ@gmail.com (S. Denysov); volodya.semenov@gmail.com (V. Semenov)

ORCID: 0000-0003-4397-4617 (S. Denysov); 0000-0002-3280-8245 (V. Semenov)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

where $\|p\|_1 = \sum_{i=1}^n |p_i|$. The last saddle point problem can be rewritten as a bilinear minmax problem on the product of standard simplexes

$$\min_{x \in \Delta^n} \max_{v \in \Delta^{2n}} (Jv, Px - x),$$

where J is an $n \times 2n$ -matrix of the form $(I \quad -I)$, where I is the identity matrix. We got the following variational inequality problem

$$\text{find } \bar{x} \in \Delta^n, \bar{v} \in \Delta^{2n} : \left((P^* - I)J\bar{v}, x - \bar{x} \right) + \left(J^*(I - P)\bar{x}, v - \bar{v} \right) \geq 0 \quad \forall x \in \Delta^n \quad \forall v \in \Delta^{2n}.$$

Note that often nonsmooth optimization problems can be effectively solved with algorithms for variational inequalities, if former are reformulated as saddle point problems [6]. With the growing popularity of generative adversarial networks (GANs) and other adversarial learning models, a steady interest in algorithms for solving variational inequalities has arisen among specialists in the field of machine learning [7, 8].

The most widely known method for solving variational inequalities is so called Korpelevich extragradient algorithm [9]. Many publications are devoted to the study of this algorithm and its modifications [6, 10–16].

An effective modern version of the extragradient method is Nemirovski mirror-prox method [6]. This method can be interpreted as a variant of the extragradient method with projection understood in the sense of Bregman divergence. One more interesting method of dual extrapolation for solving variational inequalities was proposed by Nesterov [10]. Adaptive variants of the Nemirovski proximal mirror method were studied in [11–13].

In the early 1980s, Popov proposed a modification of the classical Arrow-Hurwitz algorithm for finding saddle points of convex-concave functions [17]. Recently Popov's algorithm for variational inequalities has become well known among machine learning specialists under the name “Extrapolation from the Past” [7, 8].

A modification of Popov's method for solving variational inequalities with monotone operators was studied in [18]. And in the article [19], a two-stage proximal algorithm for solving the equilibrium programming problem is proposed, which is an adaptation of the method [17] to the general Ky Fan inequalities. The algorithm from [19] uses Bregman divergence instead of Euclidean distance. Further development of this circle of ideas led to the emergence of the so-called forward-reflected-backward algorithm [20]:

$$x_{n+1} = P_C \left(x_n - \lambda_n A x_n - \lambda_{n-1} (A x_n - A x_{n-1}) \right),$$

and similar method [21]:

$$x_{n+1} = P_C \left(x_n - \lambda_n A x_n \right) - \lambda_{n-1} (A x_n - A x_{n-1}).$$

Recently, using theory of Banach spaces and constructions of their geometry [26–30], progress has been achieved in the research of algorithms for problems above in Banach spaces [3, 23–25]. Extensive material on this topic is contained in the book [3]. The next algorithm for solving variational inequalities in a 2-uniformly convex and uniformly smooth Banach space was proposed in [22]:

$$x_{n+1} = \Pi_C J^{-1} (J x_n - \lambda A x_n),$$

where Π_C is Alber generalized projection operator [23], $\lambda > 0$, J is normalized dual mapping from E to E^* . This method weakly converges for inversely strongly monotone (cocoercive) operators $A: E \rightarrow E^*$. Shehu [24] has recently extended Tseng's result to 2-uniformly convex and uniformly smooth Banach spaces. He proposed the next weakly converging process:

$$\begin{cases} y_n = \Pi_C J^{-1} (x_n - \lambda_n A x_n), \\ x_{n+1} = J^{-1} (J y_n - \lambda_n (A y_n - A x_n)), \end{cases}$$

where $\lambda_n > 0$ is either chosen based on operator A Lipschitz constant value or calculated with a kind of linear search procedure.

It should be noted that the early research on algorithms for solving variational inequalities was usually concentrated on the study of convergence of algorithms and related questions of an asymptotic nature [13–15, 17, 19–25].

More recent studies are focused on estimating the number of iterations of algorithms required to obtain an approximate solution of a given quality [6, 8, 10–12, 16, 18]. This direction of research was initiated by the work of Nemirovski [6], where mentioned earlier mirror-prox algorithm was proposed and $O(\frac{1}{\varepsilon})$ complexity estimate in terms of the gap function was obtained for the class of problems with monotone Lipschitz continuous operator.

A fundamental question arose about the construction of an algorithm with $O(\frac{1}{\varepsilon})$ complexity estimation and single computation of the operator's value and projection onto the feasible set at the iteration step. Algorithm [20] answers this question in the case of a Hilbert space.

This work is devoted to the study of three new extragradient type algorithms for solving monotone variational inequalities in a Banach space. The first two algorithms are natural modifications of Tseng's method [14] and “Extrapolation from the Past” method [18] for problems in Banach spaces using the generalized Alber projection. Iteration of each of these algorithms is more economical than iteration of the extragradient method, because the first one uses single projection on iteration, and the second one needs only one operator calculation. The third algorithm, called operator extrapolation method, is a variant of the forward-reflected-backward algorithm, proposed by Malitsky and Tam [20]. Operator extrapolation method also uses generalized Alber projection onto the feasible set. An attractive feature of the algorithm is only one computation at the iterative step of the operator value and the generalized projection. The $O(\frac{1}{\varepsilon})$ complexity estimations are proved in terms of the gap function for variational inequalities with monotone Lipschitz operators acting in a 2-uniformly convex and uniformly smooth Banach space.

The article has the following structure. Section 2 contains the necessary information on the geometry of Banach spaces. Section 3 is devoted to variational inequalities. The algorithms are described in Section 4. The formulations and proofs of complexity estimations are presented in Section 5.

2. Preliminaries

Let us remind some basic terms and results from Banach spaces geometry, which are needed for us to formulate and prove our results [23, 25–30].

Everywhere later E is a real Banach space with norm $\|\cdot\|$, E^* – dual space for E , $\langle x^*, x \rangle$ – value of functional $x^* \in E^*$ on element $x \in E$. Let's also denote $\|\cdot\|_*$ norm in E^* .

Let $S_E = \{x \in E : \|x\| = 1\}$. Banach space E is called strictly convex, if for all $x, y \in S_E$ and $x \neq y$ we have $\left\| \frac{x+y}{2} \right\| < 1$. Convexity module of E is defined as

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

Banach space E is called uniformly convex, if $\delta_E(\varepsilon) > 0 \quad \forall \varepsilon \in (0, 2]$ [26, 27]. Banach space E is called 2-uniformly convex, if exists $c > 0$ such that $\delta_E(\varepsilon) \geq c\varepsilon^2$ for all $\varepsilon \in (0, 2]$ [27]. Obviously 2-uniformly convex space is uniformly convex. It's known that uniformly convex Banach space is reflexive [26-28].

Banach space E is called smooth, if

$$\lim_{t \rightarrow 0} \frac{\|x+ty\| - \|x\|}{t} \quad (1)$$

exists for all $x, y \in S_E$ [26]. Banach space E is called uniformly smooth if limit (1) exists uniformly for $x, y \in S_E$ [26]. There is a duality between convexity and smoothness of E and it's dual space E^*

[26,27]: E^* – strictly convex $\Rightarrow E$ – smooth; E^* – smooth $\Rightarrow E$ – strictly convex; E – uniformly convex $\Leftrightarrow E^*$ – uniformly smooth; E – uniformly smooth $\Leftrightarrow E^*$ – uniformly convex. The first two implications can be reversed for reflexive space E .

Smoothness module of space E is defined as

$$\rho_E(t) = \sup \left\{ \frac{\|x+ty\| + \|x-ty\|}{2} - 1 : x, y \in S_E \right\} \quad \forall t > 0.$$

Uniform smoothness of Banach space E is equivalent to the relation $\lim_{t \rightarrow +0} \rho_E(t)t^{-1} = 0$ [27, 28]. Banach space E is called 2-uniformly smooth, if exists $c > 0$ such that $\rho_E(t) \leq ct^2$ for all $t > 0$ [27, 28]. Banach space E is 2-uniformly convex if and only if E^* is 2-uniformly smooth [27–29].

It is known, that Hilbert spaces and spaces L_p ($1 < p \leq 2$) are 2-uniformly convex and uniformly smooth (L_p are uniformly smooth for $p \in (1, +\infty)$ and 2-uniformly smooth for $p \in [2, +\infty)$) [27, 28].

Multivalued mapping $J : E \rightarrow 2^{E^*}$, which has the form

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

is called normalized dual mapping [27, 28].

For Hilbert space $J = I$ (identity). It is known [23, 27, 28] that: if space E is smooth, then mapping J is single-valued; if E is strictly convex, then J is injective and strictly monotone; if space E is reflexive, then mapping J is surjective; if E is uniformly smooth, then J is uniformly continuous on bounded subsets of E . The explicit form of the mapping J for spaces ℓ_p , L_p and W_p^m ($p \in (1, \infty)$) is given in [3, 23, 27, 28].

Let E be a smooth Banach space. Let's consider functional, introduced by Y. Alber in [23]:

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

The next useful 3-point identity follows from the definition above:

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

If space E is strictly convex, then for $x, y \in E$ we have

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

Lemma 1 ([25]). Let E be a 2-uniformly convex smooth Banach space. Then for some $\mu \geq 1$

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

For Banach spaces ℓ_p , L_p and W_p^m ($1 < p \leq 2$) we have $\mu = \frac{1}{p-1}$ [29, 30]. For a Hilbert space, inequality from Lemma 1 becomes identity with $\mu = 1$.

Lemma 2 ([29]). Let E be a 2-uniformly smooth Banach space. Then for some $\kappa > 0$

$$\|x + y\|^2 \leq \|x\|^2 + 2\langle Jx, y \rangle + \kappa \|y\|^2 \quad \forall x, y \in E.$$

For Banach spaces ℓ_p , L_p and W_p^m ($2 < p < +\infty$) we have $\kappa = p - 1$ [29, 30]. For a Hilbert space, inequality from Lemma 2 becomes identity with $\kappa = 1$.

Let K be a non-empty closed and convex subset of a reflexive, strictly convex and smooth space E . It is known [23] that for each $x \in E$ there exists unique point $z \in K$ such that

$$d(z, x) = \inf_{y \in K} d(y, x).$$

This point z is denoted by $\Pi_K x$, and the corresponding operator $\Pi_K : E \rightarrow K$ is called generalized projection of E onto K (generalized Alber projection) [23]. Note that for a Hilbert space Π_K coincides with the metric projection onto the set K .

Lemma 3 ([23]). Let K – closed and convex subset of a reflexive, strictly convex and smooth space E , $x \in E$, $z \in K$. Then

$$z = \Pi_K x \Leftrightarrow \langle Jz - Jx, y - z \rangle \geq 0 \quad \forall y \in K.$$

Inequality from Lemma 3 is equivalent to the next one [23]:

$$d(y, \Pi_K x) + d(\Pi_K x, x) \leq d(y, x) \quad \forall y \in K.$$

Remark 1. The main element of the algorithms studied below is the calculation of a new point

$$x^+ = \Pi_K J^{-1}(Jx - x^*)$$

by known $x \in E$ and $x^* \in E^*$. From Lemma 3 and mentioned 3-point identity follows the inequality fundamental for the analysis of algorithms.

$$d(y, x^+) \leq d(y, x) - d(x^+, x) + 2\langle x^*, y - x^+ \rangle \quad \forall y \in K.$$

Basic information about monotone operators and variational inequalities in Banach spaces can be found in [1, 3, 23, 28].

3. Variational inequalities

Let E be a 2-uniformly convex and uniformly smooth Banach space, C is a non-empty subset of space E , A is an operator, acting from E to E^* . Let's consider the variational inequality:

$$\text{find } x \in C : \langle Ax, y - x \rangle \geq 0 \quad \forall y \in C. \quad (2)$$

Let's denote set of solutions of (2) as S .

We need the next assumptions:

- set $C \subseteq E$ is convex and closed;
- operator $A : E \rightarrow E^*$ is a monotone and Lipschitz continuous with constant $L > 0$ on C ;
- set S is non-empty.

Let's consider dual variational inequality:

$$\text{find } x \in C : \langle Ay, x - y \rangle \leq 0 \quad \forall y \in C. \quad (3)$$

We will denote set of solutions of (3) as S^d . Note that set S^d is convex and closed. Inequality (3) is sometimes called weak or dual formulation of (2) (or Minty inequality), and solutions of (3) are called weak solutions for (2). For monotone operator A we always have $S \subseteq S^d$. In our setting we have $S^d = S$ [1].

Variational inequality (2) can be formulated as a fixed-point problem [23]:

$$x = \Pi_C J^{-1}(Jx - \lambda Ax), \quad (4)$$

with $\lambda > 0$. Formulation (4) is useful, as it leads to obvious algorithmic idea. Procedure

$$x_{n+1} = \Pi_C J^{-1}(Jx_n - \lambda Ax_n) \quad (5)$$

was studied in [22] for inversely strongly monotone operators $A : E \rightarrow E^*$. However, for Lipschitz continuous monotone operators, algorithm (5) generally does not converge. Numerous modifications of the extragradient algorithm [6, 9–16, 24] can be used for such conditions.

In this paper, we focus on three algorithms: the Tseng method [14], the extrapolation from the past method [18] and more recent forward-reflected-backward algorithm [20]. We consider their natural

modifications for problems in Banach spaces using Alber generalized projection instead of the metric projection.

The goal is to estimate the number of iterations of algorithms necessary to obtain an approximate solution of a given quality. The quality of approximate solution $x \in C$ of variational inequality (2) will be measured using the non-negative gap function [6]

$$\text{Gap}(x) = \sup_{y \in C} \langle Ay, x - y \rangle. \quad (6)$$

Obviously, for definition (6) to be correct, feasible set C should be bounded.

The next lemma holds:

Lemma 4. Let operator A be monotone. If $x \in C$ is a solution of (2), then $\text{Gap}(x) = 0$. And vice versa, if for some $x \in C$ we have $\text{Gap}(x) = 0$, then x is a solution of (2).

Everywhere below we assume, that set $C \subseteq E$ is bounded.

4. Algorithms

Let us study the next iterative extragradient-type algorithms for finding solutions of variational inequality problem (2).

Algorithm 1. Modified Tseng method ([24]).

Select $x_1 \in E$, $\lambda_n > 0$. Set $n = 1$.

1. Calculate

$$y_n = \Pi_C J^{-1}(Jx_n - \lambda_n Ax_n).$$

2. If $y_n = x_n$, then STOP. Else calculate

$$x_{n+1} = J^{-1}(Jy_n - \lambda_n (Ay_n - Ax_n)).$$

3. Set $n := n + 1$ and go to step 1.

Algorithm 1 is a modification of forward-backward-forward method by P. Tseng [14] for problems in Banach spaces, where generalized Alber projection is used instead of the metric one.

The weak convergence of the Algorithm 1 in 2-uniformly convex and uniformly smooth Banach space is proved in [24].

Note that in the case of a Hilbert space and without constraints, the Algorithm 1 coincides with the Korpelevich extragradient method.

Algorithm 2. Extrapolation from the Past.

Select $x_1 = y_0 \in E$, $\lambda_n > 0$. Set $n = 1$.

1. Calculate

$$y_n = \Pi_C J^{-1}(Jx_n - \lambda_n Ay_{n-1}).$$

2. Calculate

$$x_{n+1} = \Pi_C J^{-1}(Jx_n - \lambda_n Ay_n),$$

if $x_{n+1} = y_n = x_n$, then STOP. Else set $n := n + 1$ and go to step 1.

Algorithm 2 is a modification of L.D. Popov algorithm [17] for problems in Banach spaces using generalized Alber projection operator instead of the metric one.

The convergence of Algorithm 2 in a Hilbert space and in Euclidean space with Bregman divergence instead of Euclidean distance is proved in [18, 19].

Algorithm 3. Operator extrapolation.

Select $x_0 = x_1 \in E$, $\lambda_n > 0$. Set $n = 1$.

1. Calculate

$$x_{n+1} = \Pi_C J^{-1}(Jx_n - \lambda_n Ax_n - \lambda_{n-1}(Ax_n - Ax_{n-1})).$$

2. If $x_{n-1} = x_n = x_{n+1}$, then STOP. Else set $n := n + 1$ and go to step 1.

Algorithm 3 is a modification of modern "forward-reflected-backward algorithm" [20] for variational inequalities in Banach spaces.

Remark 2. Some text Algorithm 3 can be represented in form, which is similar to Algorithm 2:

$$\begin{cases} x_n = \Pi_C J^{-1}(Jy_n - \lambda_{n-1}Ax_{n-1}), \\ y_{n+1} = J^{-1}(Jx_n - \lambda_{n-1}(Ax_n - Ax_{n-1})). \end{cases}$$

This formulation indicates conceptual similarity of Algorithms 1 and 3. More precisely, the operator extrapolation algorithm is obtained from Algorithm 1 in the same way, as Algorithm 2 can be obtained from the analogue of the extragradient algorithm

$$\begin{cases} y_n = \Pi_C J^{-1}(Jx_n - \lambda_n Ax_n), \\ x_{n+1} = \Pi_C J^{-1}(Jx_n - \lambda_n Ay_n). \end{cases}$$

Now let us turn to the analysis of the algorithms, namely, to the estimation of the number of iterations required to obtain an approximate solution of the variational inequality (2) with a given value of the gap function.

5. Analysis

We will prove, that Algorithms 1 – 3, mentioned above, require $O\left(\frac{LD}{\varepsilon}\right)$ iterations to obtain feasible point $x \in C$, for which $\text{Gap}(x) \leq \varepsilon$, where $\varepsilon > 0$ and $D = \sup_{a,b \in C} d(a,b) < +\infty$.

Theorem 1. Let (x_n) , (y_n) are sequences, generated by Algorithm 1. Suppose that $\lambda_n \in \left(0, \frac{1}{\sqrt{\mu\kappa L}}\right]$.

Then for the sequence of Cesaro means $z_N = \frac{\sum_{n=1}^N \lambda_n y_n}{\sum_{n=1}^N \lambda_n}$ the next inequality holds

$$\text{Gap}(z_N) \leq \frac{1}{2\sum_{n=1}^N \lambda_n} \sup_{y \in C} d(y, x_1).$$

Proof. For arbitrary $y \in C$ we have

$$\begin{aligned} d(y, x_{n+1}) &= d\left(y, J^{-1}\left(Jy_n - \lambda_n(Ay_n - Ax_n)\right)\right) = \\ &= \|y\|^2 - 2\langle Jy_n - \lambda_n(Ay_n - Ax_n), y \rangle + \|Jy_n - \lambda_n(Ay_n - Ax_n)\|_*^2 = \\ &= \|y\|^2 - 2\langle Jy_n, y \rangle + 2\lambda_n \langle Ay_n - Ax_n, y \rangle + \|Jy_n - \lambda_n(Ay_n - Ax_n)\|_*^2 = \\ &= d(y, y_n) - \|y_n\|^2 + 2\lambda_n \langle Ay_n - Ax_n, y \rangle + \|Jy_n - \lambda_n(Ay_n - Ax_n)\|_*^2. \end{aligned}$$

Using Lemma 2, we get an estimation

$$d(y, x_{n+1}) \leq d(y, y_n) - 2\lambda_n \langle Ay_n - Ax_n, y_n - y \rangle + \kappa\lambda_n^2 \|Ay_n - Ax_n\|_*^2. \quad (7)$$

Let's use 3-point identity to transform $d(y, y_n)$

$$d(y, y_n) = d(y, x_n) + d(x_n, y_n) + 2\langle Jx_n - Jy_n, y - x_n \rangle.$$

and now use this in (7)

$$\begin{aligned} d(y, x_{n+1}) &\leq d(y, x_n) + d(x_n, y_n) + \\ &\quad + 2\langle Jx_n - Jy_n, y - x_n \rangle - 2\lambda_n \langle Ay_n - Ax_n, y_n - y \rangle + \kappa\lambda_n^2 \|Ay_n - Ax_n\|_*^2. \end{aligned}$$

From Lipschitz continuity of A , equality $d(x_n, y_n) = 2\langle Jy_n - Jx_n, y_n - x_n \rangle - d(y_n, x_n)$ and Lemma 1 we have

$$\begin{aligned} d(y, x_{n+1}) &\leq d(y, x_n) - (1 - \kappa\mu\lambda_n^2 L^2) d(y_n, x_n) - \\ &\quad - 2\langle Jx_n - Jy_n - \lambda_n Ax_n + \lambda_n Ay_n, y_n - y \rangle. \end{aligned}$$

Identity

$$y_n = \Pi_C J^{-1}(Jx_n - \lambda_n Ax_n)$$

is equivalent to inequality

$$\langle Jx_n - \lambda_n Ax_n - Jy_n, y_n - y \rangle \geq 0.$$

So, we have

$$d(y, x_{n+1}) \leq d(y, x_n) - (1 - \kappa\mu\lambda_n^2 L^2) d(y_n, x_n) - 2\lambda_n \langle Ay_n, y_n - y \rangle.$$

From monotonicity of A operator it follows

$$d(y, x_{n+1}) \leq d(y, x_n) - (1 - \kappa\mu\lambda_n^2 L^2) d(y_n, x_n) - 2\lambda_n \langle Ay, y_n - y \rangle. \quad (8)$$

Let's transform (8) as

$$2\lambda_n \langle Ay, y_n - y \rangle \leq d(y, x_n) - d(y, x_{n+1}) - (1 - \kappa\mu\lambda_n^2 L^2) d(y_n, x_n). \quad (9)$$

Summing (9) over n from 1 to N , we get

$$2 \sum_{n=1}^N \lambda_n \langle Ay, y_n - y \rangle \leq d(y, x_1),$$

which leads to

$$\langle Ay, z_N - y \rangle \leq \frac{1}{2 \sum_{n=1}^N \lambda_n} d(y, x_1), \quad (10)$$

where $z_N = \frac{\sum_{n=1}^N \lambda_n y_n}{\sum_{n=1}^N \lambda_n}$. Passing in (10) to the supremum over $y \in C$, we obtain

$$\text{Gap}(z_N) \leq \frac{1}{2 \sum_{n=1}^N \lambda_n} \sup_{y \in C} d(y, x_1),$$

which was required to prove. ■

Corollary 1. Let (x_n) , (y_n) are the sequences, generated by Algorithm 1 with $\lambda_n = \frac{1}{2\sqrt{\mu\kappa L}}$. Then for the sequence of means $z_N = \frac{1}{N} \sum_{n=1}^N y_n$ the next inequality holds:

$$\text{Gap}(z_N) \leq \frac{L\sqrt{\mu\kappa} \sup_{y \in C} d(y, x_1)}{N}.$$

For algorithm 2 we have the next result.

Theorem 2. Let (x_n) , (y_n) are the sequences, generated by Algorithm 2. Let $\lambda_n \in \left(0, \frac{\sqrt{2}-1}{\mu L}\right]$. Then

for the Cesaro means sequence $z_N = \frac{\sum_{n=1}^N \lambda_n y_n}{\sum_{n=1}^N \lambda_n}$ the next inequality holds:

$$\text{Gap}(z_N) \leq \frac{\sup_{y \in C} d(y, x_1) + \mu\lambda_1 L d(x_1, y_0)}{2 \sum_{n=1}^N \lambda_n}.$$

Proof. For arbitrary $y \in C$ we have

$$d(y, x_{n+1}) \leq d(y, x_n) - d(x_{n+1}, x_n) + 2\lambda_n \langle Ay_n, y - x_{n+1} \rangle.$$

From monotonicity of A follows

$$\langle Ay_n, y - x_{n+1} \rangle = \langle Ay_n, y - y_n \rangle + \langle Ay_n, y_n - x_{n+1} \rangle \leq \langle Ay, y - y_n \rangle + \langle Ay_n, y_n - x_{n+1} \rangle.$$

So, we have

$$\begin{aligned} d(y, x_{n+1}) &\leq d(y, x_n) - d(x_{n+1}, x_n) + 2\lambda_n \langle Ay_n, y_n - x_{n+1} \rangle + 2\lambda_n \langle Ay, y - y_n \rangle = \\ &= d(y, x_n) - d(x_{n+1}, x_n) + 2\lambda_n \langle Ay_{n-1}, y_n - x_{n+1} \rangle + \\ &\quad + 2\lambda_n \langle Ay_n - Ay_{n-1}, y_n - x_{n+1} \rangle + 2\lambda_n \langle Ay, y - y_n \rangle. \end{aligned} \quad (11)$$

Let's write $2\lambda_n \langle Ay_{n-1}, y_n - x_{n+1} \rangle$ as

$$2\lambda_n \langle Ay_{n-1}, y_n - x_{n+1} \rangle = 2 \langle Jx_n - \lambda_n Ay_{n-1} - Jy_n, x_{n+1} - y_n \rangle + 2 \langle Jy_n - Jx_n, x_{n+1} - y_n \rangle.$$

From the inclusion $x_{n+1} \in C$ and Lemma 3 follows inequality

$$\langle Jx_n - \lambda_n Ay_{n-1} - Jy_n, x_{n+1} - y_n \rangle \leq 0.$$

As a result, we have

$$2\lambda_n \langle Ay_{n-1}, y_n - x_{n+1} \rangle \leq 2 \langle Jy_n - Jx_n, x_{n+1} - y_n \rangle = d(x_{n+1}, x_n) - d(x_{n+1}, y_n) - d(y_n, x_n). \quad (12)$$

Estimating the right side of (11) with (12), we come to the next inequality:

$$d(y, x_{n+1}) \leq d(y, x_n) - d(x_{n+1}, y_n) - d(y_n, x_n) + 2\lambda_n \langle Ay_n - Ay_{n-1}, y_n - x_{n+1} \rangle + 2\lambda_n \langle Ay, y - y_n \rangle. \quad (13)$$

Now let's estimate term $2\lambda_n \langle Ay_n - Ay_{n-1}, y_n - x_{n+1} \rangle$. We get

$$\begin{aligned} 2\lambda_n \langle Ay_n - Ay_{n-1}, y_n - x_{n+1} \rangle &\leq 2\lambda_n \|Ay_{n-1} - Ay_n\|_* \|x_{n+1} - y_n\| \leq 2\lambda_n L \|y_{n-1} - y_n\| \|x_{n+1} - y_n\| \leq \\ &\leq 2\lambda_n L \left\{ \frac{1}{2\sqrt{2}} \|y_{n-1} - y_n\|^2 + \frac{1}{\sqrt{2}} \|x_{n+1} - y_n\|^2 \right\} \leq \\ &\leq \frac{\lambda_n L}{\sqrt{2}} \left\{ \sqrt{2} \|y_{n-1} - x_n\|^2 + (2 + \sqrt{2}) \|x_n - y_n\|^2 \right\} + \frac{2\lambda_n L}{\sqrt{2}} \|x_{n+1} - y_n\|^2 = \\ &= \lambda_n L \left(\|y_{n-1} - x_n\|^2 + (1 + \sqrt{2}) \|x_n - y_n\|^2 + \sqrt{2} \|x_{n+1} - y_n\|^2 \right). \end{aligned} \quad (14)$$

Here we have used auxiliary inequalities

$$ab \leq \frac{\epsilon^2}{2} a^2 + \frac{1}{2\epsilon^2} b^2, \quad (a+b)^2 \leq \sqrt{2} a^2 + (2 + \sqrt{2}) b^2.$$

Estimating norms in (14) using inequality from Lemma 1, we get

$$\begin{aligned} 2\lambda_n \langle Ay_n - Ay_{n-1}, y_n - x_{n+1} \rangle &\leq \mu\lambda_n L d(x_n, y_{n-1}) + \\ &+ \mu\lambda_n L (1 + \sqrt{2}) d(y_n, x_n) + \mu\lambda_n L \sqrt{2} d(x_{n+1}, y_n). \end{aligned} \quad (15)$$

Using (15) in (13), we get

$$\begin{aligned} \phi(y, x_{n+1}) &\leq d(y, x_n) - (1 - \mu\lambda_n L \sqrt{2}) d(x_{n+1}, y_n) - (1 - \mu\lambda_n L (1 + \sqrt{2})) d(y_n, x_n) + \\ &+ \mu\lambda_n L d(x_n, y_{n-1}) + 2\lambda_n \langle Ay, y - y_n \rangle. \end{aligned} \quad (16)$$

Let's rewrite (16) as

$$\begin{aligned} 2\lambda_n \langle Ay, y_n - y \rangle &\leq (d(y, x_n) + \mu\lambda_n L d(x_n, y_{n-1})) - (d(y, x_{n+1}) + \mu\lambda_{n+1} L d(x_{n+1}, y_n)) - \\ &- (1 - \mu L (\lambda_{n+1} + \sqrt{2}\lambda_n)) d(x_{n+1}, y_n) - (1 - \mu\lambda_n L (1 + \sqrt{2})) d(y_n, x_n). \end{aligned} \quad (17)$$

Summing (17) over n from 1 to N , we get

$$2 \sum_{n=1}^N \lambda_n \langle Ay, y_n - y \rangle \leq d(y, x_1) + \mu\lambda_1 L d(x_1, y_0),$$

and so

$$\langle Ay, z_N - y \rangle \leq \frac{d(y, x_1) + \mu\lambda_1 L d(x_1, y_0)}{2 \sum_{n=1}^N \lambda_n}, \quad (18)$$

where $z_N = \frac{\sum_{n=1}^N \lambda_n y_n}{\sum_{n=1}^N \lambda_n}$. Passing in (18) to the supremum over $y \in C$, we obtain

$$\text{Gap}(z_N) \leq \frac{\sup_{y \in C} d(y, x_1) + \mu\lambda_1 L d(x_1, y_0)}{2 \sum_{n=1}^N \lambda_n},$$

which was required to prove. ■

Corollary 2. Let $(x_n), (y_n)$ are sequences, generated by Algorithm 2 with $\lambda_n = \frac{1}{3\mu L}$. Then for the means sequence $z_N = \frac{1}{N} \sum_{n=1}^N y_n$ the next inequality holds:

$$\text{Gap}(z_N) \leq \frac{\mu L \frac{3}{2} \sup_{y \in C} d(y, x_1) + \frac{1}{6} d(x_1, y_0)}{N}.$$

Let's study Algorithm 3.

Theorem 3. Let (x_n) be a sequence, generated by Algorithm 3. Let $\lambda_n \in \left(0, \frac{1}{2\mu L}\right]$. Then for the

sequence of Cesaro means $z_{N+1} = \frac{\sum_{n=1}^N \lambda_n x_{n+1}}{\sum_{n=1}^N \lambda_n}$ the next inequality holds

$$\text{Gap}(z_{N+1}) \leq \frac{1}{2\sum_{n=1}^N \lambda_n} \sup_{y \in C} d(y, x_1).$$

Proof. For sequence (x_n) , generated by algorithm 3, the next inequality holds

$$-2\langle \lambda_n Ax_n + \lambda_{n-1}(Ax_n - Ax_{n-1}), y - x_{n+1} \rangle \leq d(y, x_n) - d(x_{n+1}, x_n) - d(y, x_{n+1}) \quad \forall y \in C. \quad (19)$$

Let's rewrite (19) the next way:

$$d(y, x_n) - d(y, x_{n+1}) \geq 2\lambda_n \langle Ax_{n+1}, x_{n+1} - y \rangle - 2\lambda_n \langle Ax_{n+1} - Ax_n, x_{n+1} - y \rangle + \\ + 2\lambda_{n-1} \langle Ax_n - Ax_{n-1}, x_n - y \rangle + 2\lambda_{n-1} \langle Ax_n - Ax_{n-1}, x_{n+1} - x_n \rangle + d(x_{n+1}, x_n). \quad (20)$$

Summing (20) over n from 1 to N , we get

$$d(y, x_1) - d(y, x_{N+1}) \geq \\ \geq 2\sum_{n=1}^N \lambda_n \langle Ax_{n+1}, x_{n+1} - y \rangle - 2\lambda_N \langle Ax_{N+1} - Ax_N, x_{N+1} - y \rangle + \\ + \sum_{n=1}^N (2\lambda_{n-1} \langle Ax_n - Ax_{n-1}, x_{n+1} - x_n \rangle + d(x_{n+1}, x_n)). \quad (21)$$

Using Lipschitz continuity of A and Lemma 1 we get

$$\sum_{n=1}^N (-2\lambda_{n-1}L \langle Ax_n - Ax_{n-1}, x_{n+1} - x_n \rangle + d(x_{n+1}, x_n)) \geq \\ \geq \sum_{n=1}^N \left(-2\lambda_{n-1}L \|x_n - x_{n-1}\| \|x_{n+1} - x_n\| + \frac{1}{2\mu} \|x_{n+1} - x_n\|^2 + \frac{1}{2\mu} \|x_n - x_{n-1}\|^2 \right) + \frac{1}{2\mu} \|x_{N+1} - x_N\|^2 \geq \\ \geq \frac{1}{2\mu} \|x_{N+1} - x_N\|^2.$$

Using this estimation in (21), we get

$$d(y, x_1) - d(y, x_{N+1}) \geq 2\sum_{n=1}^N \lambda_n \langle Ax_{n+1}, x_{n+1} - y \rangle - 2\lambda_N \langle Ax_{N+1} - Ax_N, x_{N+1} - y \rangle + \\ + \frac{1}{2\mu} \|x_{N+1} - x_N\|^2 \geq \\ \geq 2\sum_{n=1}^N \lambda_n \langle Ax_{n+1}, x_{n+1} - y \rangle - 2\lambda_N L \|x_{N+1} - x_N\| \|x_{N+1} - y\| + \frac{1}{2\mu} \|x_{N+1} - x_N\|^2 \geq \\ \geq 2\sum_{n=1}^N \lambda_n \langle Ax_{n+1}, x_{n+1} - y \rangle - \lambda_N L \|x_{N+1} - y\|^2.$$

So, we can come to inequality

$$2\sum_{n=1}^N \lambda_n \langle Ax_{n+1}, x_{n+1} - y \rangle - \lambda_N L \|x_{N+1} - y\|^2 + d(y, x_{N+1}) \leq d(y, x_1) \quad \forall y \in C. \quad (22)$$

Using monotonicity of A , we get

$$\sum_{n=1}^N \lambda_n \langle Ax_{n+1}, x_{n+1} - y \rangle \geq \sum_{n=1}^N \lambda_n \langle Ay, x_{n+1} - y \rangle = \left(\sum_{n=1}^N \lambda_n \right) \langle Ay, z_{N+1} - y \rangle, \quad (23)$$

where $z_{N+1} = \frac{\sum_{n=1}^N \lambda_n x_{n+1}}{\sum_{n=1}^N \lambda_n}$. Using estimation (22) in (23), we come to inequality

$$2\left(\sum_{n=1}^N \lambda_n \right) \langle Ay, z_{N+1} - y \rangle + \left(\frac{1}{\mu} - \lambda_N L \right) \|x_{N+1} - y\|^2 \leq d(y, x_1) \quad \forall y \in C,$$

from which follows

$$\text{Gap}(z_{N+1}) = \sup_{y \in C} \langle Ay, z_{N+1} - y \rangle \leq \left(2 \sum_{n=1}^N \lambda_n \right)^{-1} \sup_{y \in C} d(y, x_1),$$

which was required to prove. ■

Corollary 3. Let (x_n) be a sequence, generated by Algorithm 3 with $\lambda_n = \frac{1}{2\mu L}$. Then for the means sequence $z_{N+1} = \frac{1}{N} \sum_{n=1}^N x_{n+1}$ the next inequality holds

$$\text{Gap}(z_{N+1}) \leq \frac{L\mu}{N} \sup_{y \in C} d(y, x_1).$$

6. Conclusions

Three new extragradient-type algorithms for solving monotone variational inequalities in a Banach space are studied in the paper.

The first two algorithms are natural modifications of Tseng's method [14] and “Extrapolation from the Past” method [18] for problems in Banach spaces using generalized Alber projection. Iteration of each of these algorithms is more economical than iteration of the extragradient method. The first algorithm has less projections, the second has less operator calculations.

The third algorithm, called the method of operator extrapolation, is a variant of the Malitsky–Tam forward-reflected-backward algorithm [20]. Generalized Alber projection is used instead of the metric projection onto feasible set. An attractive feature of the algorithm is only one computation of the operator value and of the generalized projection onto the feasible set at the iterative step. For variational inequalities with monotone Lipschitz operators acting in a 2-uniformly convex and uniformly smooth Banach space, $O\left(\frac{1}{\varepsilon}\right)$ complexity bounds are proved in terms of the gap function.

Let us point out two actual questions, related to the current research. First, fast and stable algorithms for computing generalized Alber projection for a wide range of sets are needed to efficiently apply algorithms to nonlinear problems in Banach spaces. Second, all results were obtained for the class of 2-uniformly convex and uniformly smooth Banach spaces, which does not contain spaces L_p and W_p^m ($2 < p < +\infty$), which are important for applications. It is highly desirable to get rid of this limitation.

Acknowledgements

This work was supported by the Ministry of Education and Science of Ukraine (project “Mathematical modeling and optimization of dynamic systems for defense, medicine and ecology”, state registration number 0119U100337) and the National Academy of Sciences of Ukraine (project “New methods of research of correctness and solution search for discrete optimization problems, variational inequalities and their applications”, state registration number 0119U101608).

References

- [1] D. Kinderlehrer, G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications*, Society for Industrial and Applied Mathematics, Philadelphia, 2000.
- [2] A. Nagurny, *Network economics: A variational inequality approach*, Kluwer Academic Publishers, Dordrecht, 1999.
- [3] Y. Alber, I. Ryazantseva, *Nonlinear Ill Posed Problems of Monotone Type*, Springer, Dordrecht, 2006.
- [4] F. Facchinei, J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer Series in Operations Research, vol. I, Springer, New York, 2003.
- [5] S. I. Lyashko, D. A. Klyushin, D. A. Nomirovsky, V. V. Semenov, Identification of age-structured contamination sources in ground water, in: R. Boucekine, 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.
- [6] A. Nemirovski, Prox-method with rate of convergence $O(1/T)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems, *SIAM J. Optim.* 15 (2004) 229–251. doi:10.1137/S1052623403425629.

- [7] G. Gidel, H. Berard, P. Vincent, S. Lacoste-Julien, A Variational Inequality Perspective on Generative Adversarial Networks, arXiv preprint arXiv:1802.10551. (2018).
- [8] M. Liu, W. Zhang, Y. Mroueh, X. Cui, J. Ross, T. Yang, P. Das, A decentralized parallel algorithm for training generative adversarial nets, arXiv preprint arXiv:1910.12999. (2020).
- [9] G. M. Korpelevich, An extragradient method for finding saddle points and for other problems, *Matecon.* 12 (1976) 747–756.
- [10] Yu. Nesterov, Dual extrapolation and its applications to solving variational inequalities and related problems, *Mathematical Programming* 109 (2007), 319–344.
- [11] F. S. Stonyakin, E. A. Vorontsova, M. S. Alkousa, New Version of Mirror Prox for Variational Inequalities with Adaptation to Inexactness, in: Jacimovic M., Khachay M., Malkova V., Posypkin M. (Eds.), *Optimization and Applications, OPTIMA 2019*, volume 1145 of *Communications in Computer and Information Science*, Springer, Cham, 2020, 427–442. doi:10.1007/978-3-030-38603-0_31.
- [12] F. Bach, K. Y. Levy, A Universal Algorithm for Variational Inequalities Adaptive to Smoothness and Noise, arXiv preprint arXiv:1902.01637. (2019).
- [13] S. V. Denisov, D. A. Nomirovskii, B. V. Rublyov, V. V. Semenov, Convergence of Extragradient Algorithm with Monotone Step Size Strategy for Variational Inequalities and Operator Equations, *Journal of Automation and Information Sciences.* 51 (2019) 12–24. doi:10.1615/JAutomatInfScien.v51.i6.20.
- [14] P. Tseng, A modified forward-backward splitting method for maximal monotone mappings, *SIAM Journal on Control and Optimization* 38 (2000) 431–446.
- [15] 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>.
- [16] 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.
- [17] L. D. Popov, A modification of the Arrow-Hurwicz method for search of saddle points, *Mathematical notes of the Academy of Sciences of the USSR.* 28 (1980) 845–848.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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>.
- [23] 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.
- [24] 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.
- [25] 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.
- [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] H. K. Xu, Inequalities in Banach spaces with applications, *Nonlinear Anal.* 16(12) (1991) 1127–1138.
- [30] 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.