DOI 10.1007/s10559-019-00142-7
Cybernetics and Systems Analysis, Vol. 55, No. 3, May, 2019

SYSTEMS ANALYSIS

CONVERGENCE OF TWO-STAGE METHOD WITH BREGMAN DIVERGENCE FOR SOLVING VARIATIONAL INEQUALITIES*

D. A. Nomirovskii,[1†] **B. V. Rublyov,**[1‡] **and V. V. Semenov**[1††]

UDC 517.988

Abstract. *A new two-stage method is proposed for the approximate solution of variational inequalities with pseudo-monotone and Lipschitz-continuous operators acting in a finite-dimensional linear normed space. This method is a modification of several well-known two-stage algorithms using the Bregman divergence instead of the Euclidean distance. Like other schemes using Bregman divergence, the proposed method can sometimes efficiently take into account the structure of the feasible set of the problem. A theorem on the convergence of the method is proved and, in the case of a monotone operator and convex compact feasible set, non-asymptotic estimates of the efficiency of the method are obtained.*

Keywords: *variational inequality, pseudo-monotonicity, monotonicity, Lipschitz condition, two-stage method, Bregman divergence, convergence.*

INTRODUCTION

A lot of important problems in operations research and mathematical physics can be written in the form of variational inequalities [1, 2]. Such inequalities are applied especially often in mathematical economics, mathematical modeling of transportation flows, and game theory [1, 2]. Many methods are proposed for their solution, in particular, projection methods (which use the operation of metric projection onto feasible set) [1–17]. The Korpelevich extra-gradient method [3] is the most well-known analog of the gradient projection method for variational inequalities. Many publications are devoted to its generalization and analysis [4–6, 9–14]. In particular, modifications of the Korpelevich algorithm with one metric projection onto the feasible set are proposed [9–14]. In so-called subgradient extra-gradient algorithms [9, 10, 13, 14] and the Korpelevich algorithm, the first stages of iteration coincide, and to obtain the next approximation, projection is carried out not onto the feasible set but on some half-space being the support for the feasible set. In the early 1980s, Popov proposed an interesting modification of the Arrow–Hurwitz algorithm of finding saddle points of convex–concave functions [15]. A modification of the Popov method for solution of variational inequalities with monotone operators is analyzed in [16]. A two-stage proximal algorithm is proposed in [18] for solution of the equilibrium programming problem, which is an adaptation of the method [15] to general Ky Fan inequalities.

The majority of the above methods use Euclidean distance and projection. In certain cases, this does not allow application of the structure of feasible sets and efficient problem solution. A possible way out is a more flexible selection

*The study was partially supported by the Ministry of Education and Science of Ukraine (Project "Development of the algorithms for modeling and optimization of dynamic systems for defense, medicine and ecology," 0116U004777).

[1]Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, †*kashpir74@gmail.com*; ‡*rublyovbv@gmail.com*; ††*semenov.volodya@gmail.com.* Translated from Kibernetika i Sistemnyi Analiz, No. 3, May–June, 2019, pp. 17–27. Original article submitted July 6, 2018.

of the distance for projection onto the feasible set. Bregman was the first to successfully implement such strategy in [19], where he proposed a method of the type of cyclic projection for finding the common point of convex sets. This paper has initiated a new field in mathematical programming and nonlinear analysis. To solve convex optimization problems, Nemirovskii and Yudin proposed the mirror descent method in the late 1970s [20], which became widely used to solve high-dimensional problems. In case of constrained problems, it can be interpreted as a version of the subgradient projection method, where projection is understood in the sense of Bregman divergence [21]. The mirror descent method allows taking into account the structure of feasible set of the optimization problem. For example, the Kullback–Leibler divergence can be used as the distance for simplex and an explicitly calculated operator of projection onto simplex can be obtained [21]. Nemirovskii's proximal mirror method [4] is one of the modern variants of the extra-gradient method for variational inequalities. Such methods are considered in detail in [5]. An interesting method of dual extrapolation for solution of variational inequalities is proposed in [6]. The paper [17] analyzes the two-stage proximal mirror method, which is a modification of the two-stage proximal algorithm [18] with the use of the Bregman divergence instead of Euclidean distance.

The present paper continues the study [22] and analyzes a new two-stage method for approximate solution of variational inequalities with pseudo-monotone and Lipschitz-continuous operators defined in finite-dimensional linear normed space. The method is a modification of the two-stage algorithms described earlier [16, 18]. The proposed scheme can also be obtained by replacing the feasible set with special support half-spaces at the first stage of the proximal mirror method [17].

PROBLEM STATEMENT AND ALGORITHM DESCRIPTION

Let E be a finite-dimensional real linear space with norm $||\cdot||$ (not necessarily Euclidean one). Denote the dual space by E^*. For $a \in E^*$ and $b \in E$, denote by (a, b) the value of the linear function a at point b. Define dual norm $||\cdot||_*$ on E^* in a standard way: $||a||_* = \max\{(a, b): ||b|| = 1\}$, ensuring the Schwarz inequality $(a, b) \le ||a||_* \, ||b||$ for all $a \in E^*$ and $b \in E$.

Let C be a nonempty subset of space E and A be an operator acting from E into E^*. Let us consider the variational inequality: find

$$x \in C: \ (Ax, y - x) \ge 0 \ \ \forall \, y \in C, \tag{1}$$

denote its set of solutions by S.

Assume that the following conditions are satisfied:

- set $C \subseteq E$ is convex and closed;
- operator $A: E \to E^*$ is pseudo-monotone and Lipschitz-continuous with constant $L > 0$ on C;
- set S is not empty.

Remark 1. Let us remind that pseudo-monotonicity of operator A on set C implies that for all x, $y \in C$ from $(Ax, y - x) \ge 0$ it follows that $(Ay, y - x) \ge 0$.

Let us consider the dual variational inequality: find

$$x \in C: (Ay, x - y) \le 0 \ \ \forall \, y \in C. \tag{2}$$

Denote the set of solutions of (2) by S^d. Note that set S^d is convex and closed. Inequality (2) is sometimes called weak or dual statement of (1), solutions (2) are called weak solutions of (1) [1]. Indeed, for pseudo-monotonicity of operator A, we have $S \subseteq S^d$. Under the considered conditions, $S^d = S$ [1].

Let us introduce the structures necessary to formulate the algorithm. Let function $\varphi: E \to \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ satisfy the following conditions:

- $\operatorname{int} \operatorname{dom} \varphi \subseteq E$ is a nonempty convex set;
- φ is continuously differentiable on $\operatorname{int} \operatorname{dom} \varphi$;

- if $\operatorname{int}\operatorname{dom}\varphi \ni x_n \to x \in \operatorname{bd}\operatorname{dom}\varphi$, then $||\nabla\varphi(x_n)||_* \to +\infty$;
- φ is strongly convex with respect to the norm $||\cdot||$ with the strong convexity constant $\sigma > 0$:

$$\varphi(a) \geq \varphi(b) - (\nabla\varphi(b), a-b) + \frac{\sigma}{2}||a-b||^2 \quad \forall\, a \in \operatorname{dom}\varphi,\ b \in \operatorname{int}\operatorname{dom}\varphi.$$

Remark 2. Functions φ are called distance generating functions.

The Bregman divergence corresponding to the function φ is defined by the formula [21]

$$V(a,b) = \varphi(a) - \varphi(b) - (\nabla\varphi(b), a-b) \quad \forall\, a \in \operatorname{dom}\varphi,\ b \in \operatorname{int}\operatorname{dom}\varphi.$$

Remark 3. Bregman divergence is sometimes called a distance [4, 21, 23], but it is not a correct definition: out of the metrics axioms for V, only $V(x, y) = 0 \Leftrightarrow x = y$ holds in the general case.

Examples of practically important Bregman divergences are presented in [21]. We will consider here two main examples. For $\varphi(\cdot) = \frac{1}{2}||\cdot||_2^2$, where $||\cdot||_2$ is a Euclidean norm, we have $V(x, y) = \frac{1}{2}||x-y||_2^2$. For the nonnegative orthant $\mathbb{R}_+^m = \{x \in \mathbb{R}^m : x_i \geq 0\}$ and function of negative entropy $\varphi(x) = \sum_{i=1}^{m} x_i \ln x_i$ (strongly convex with constant 1 with respect to ℓ_1-norm on simplex $S_m = \left\{x \in \mathbb{R}^m : x_i \geq 0, \sum_{i=1}^m x_i = 1\right\}$), we obtain the Kullback–Leibler divergence (distance)

$$V(x, y) = \sum_{i=1}^{m} x_i \ln(x_i / y_i) - \sum_{i=1}^{m}(x_i - y_i),\ x \in \mathbb{R}_+^m,\ y \in \mathbb{R}_{++}^m = \operatorname{int}(\mathbb{R}_+^m).$$

The useful three-point identity takes place [21]:

$$V(a,c) = V(a,b) + V(b,c) + (\nabla\varphi(b) - \nabla\varphi(c), a-b). \tag{3}$$

From the strong convexity of function φ, the estimate follows

$$V(a,b) \geq \frac{\sigma}{2}||a-b||^2 \quad \forall\, a \in \operatorname{dom}\varphi,\ b \in \operatorname{int}\operatorname{dom}\varphi. \tag{4}$$

Let $K \subseteq \operatorname{dom}\varphi$ be a nonempty closed convex set and $K \cap \operatorname{int}\operatorname{dom}\varphi \neq \varnothing$. Let us consider strongly convex minimization problems of the form

$$P_x^K(a) = \arg\min_{y \in K}\{-(a, y-x) + V(y,x)\} \quad \forall\, a \in E^*,\ x \in \operatorname{int}\operatorname{dom}\varphi. \tag{5}$$

Problem (5) is known [4, 21] to have a unique solution $z \in K \cap \operatorname{int}\operatorname{dom}\varphi$ and

$$-(a, y-z) + (\nabla\varphi(z) - \nabla\varphi(x), y-z) \geq 0 \quad \forall\, y \in K. \tag{6}$$

Remark 4. In the Euclidean case, point $P_x^K(a)$ coincides with the Euclidean metric projection $P_K(x+a) = \arg\min_{y \in K}||y-(x+a)||_2$.

Remark 5. For the simplex $S_m = \left\{x \in \mathbb{R}^m : x_i \geq 0, \sum_{i=1}^m x_i = 1\right\}$ and the Kullback–Leibler divergence, we get [21]

$$P_x^{S_m}(a) = \left(\frac{x_1 e^{a_1}}{\sum_{j=1}^m x_j e^{a_j}}, \frac{x_2 e^{a_2}}{\sum_{j=1}^m x_j e^{a_j}}, \ldots, \frac{x_m e^{a_m}}{\sum_{j=1}^m x_j e^{a_j}}\right),\ a \in \mathbb{R}^m,\ x \in \operatorname{ri}(S_m).$$

For the case of half-space $H_{\leq}(b,\beta) = \{y : (b, y) \leq \beta\}$, where $b \in E^* \setminus \{0\}$, $\beta \in \mathbb{R}$, we get [23]

$$P_x^{H_{\leq}(b,\beta)}(a) = (\nabla\varphi)^{-1}(\nabla\varphi(x) + a)$$

if $(\nabla\varphi)^{-1}(\nabla\varphi(x)+a)\in H_{\le}(b,\beta)$; otherwise,

$$P_x^{H_{\le}(b,\beta)}(a)=(\nabla\varphi)^{-1}(\nabla\varphi(x)+a-\tau b),$$

where $\tau=\arg\min_{t>0}\varphi^*(\nabla\varphi(x)+a-tb)+t\beta$, φ^* is a function conjugate to φ, i.e.,

$$\varphi^*(y)=\sup_{x\in\operatorname{dom}\varphi}((y,x)-\varphi(x)).$$

Let us describe the algorithm for solution of variational inequality (1).

Algorithm 1. The two-stage method with Bregman divergence. Choose elements x_0, $y_0\in C$ and positive number λ. Put $n=1$.

Step 0. Calculate

$$x_1=P_{x_0}^C(-\lambda Ay_0),\ \ y_1=P_{x_1}^C(-\lambda Ay_0).$$

Step 1. Calculate $x_{n+1}=P_{x_n}^{T_n}(-\lambda Ay_n)$ and $y_{n+1}=P_{x_{n+1}}^C(-\lambda Ay_n)$,

where

$$T_n=\{z\in E:\ (\nabla\varphi(x_n)-\lambda Ay_{n-1}-\nabla\varphi(y_n),z-y_n)\le 0\}.$$

Step 2. If $x_{n+1}=x_n$ and $y_{n+1}=y_n=y_{n-1}$, then STOP and $y_n\in S$; otherwise, put $n:=n+1$ and go to Step 1.

Remark 6. We get $C\subseteq T_n$. Indeed, if we assume that point $w\in C\setminus T_n$ exists, then the inequality

$$(\nabla\varphi(x_n)-\lambda Ay_{n-1}-\nabla\varphi(y_n),w-y_n)>0$$

contradicts the equality $y_n=P_{x_n}^C(-\lambda Ay_{n-1})$.

Remark 7. If $\varphi(\cdot)=\frac{1}{2}||\cdot||_2^2$, then Algorithm 1 takes the form of the method proposed in [16]:

$$\begin{cases}T_n=\{z\in H:\ (x_n-\lambda Ay_{n-1}-y_n,z-y_n)\le 0\},\\ x_{n+1}=P_{T_n}(x_n-\lambda Ay_n),\\ y_{n+1}=P_C(x_{n+1}-\lambda Ay_n).\end{cases}$$

The following lemma takes place.

LEMMA 1. If for some $n\in\mathbb{N}$ in Algorithm 1 we have $x_{n+1}=x_n$ and $y_{n+1}=y_n=y_{n-1}$, then $y_n\in S$.

Proof. Due to (6), equality $x_{n+1}=P_{x_n}^{T_n}(-\lambda Ay_n)$ is equivalent to the inequality

$$(Ay_n,y-x_{n+1})+\frac{(\nabla\varphi(x_{n+1})-\nabla\varphi(x_n),y-y_n)}{\lambda}\ge 0\ \ \forall\, y\in T_n.$$

From the equality $x_{n+1}=x_n$ it follows that

$$(Ay_n,y-x_n)\ge 0\ \ \forall\, y\in T_n. \tag{7}$$

Considering that $x_{n+1}\in T_n$ and $y_n=y_{n-1}$, we obtain $(\nabla\varphi(x_n)-\lambda Ay_n-\nabla\varphi(y_n),x_n-y_n)\le 0$, whence $(Ay_n,x_n-y_n)\ge 0$. Let us represent (7) as

$$(Ay_n,y-y_n)-(Ay_n,x_n-y_n)\ge 0\ \ \forall\, y\in T_n.$$

Hence,

$$(Ay_n,y-y_n)\ge(Ay_n,x_n-y_n)\ge 0\ \ \forall\, y\in T_n.$$

Since $y_n\in C\subseteq T_n$, we get $y_n\in S$. ■

In what follows, we will assume that for all numbers $n \in \mathbb{N}$ the stopping condition at Step 2 of Algorithm 1 does not hold and go to substantiation of the convergence of Algorithm 1.

THE MAIN INEQUALITY FOR POINTS GENERATED BY THE ALGORITHM

First, let us prove an important estimate that relates Bregman divergence between point x_n generated by the two-stage Algorithm 1 and arbitrary element from the solution set S.

LEMMA 2. For sequences (x_n) and (y_n) generated by Algorithm 1, the inequality

$$V(z,x_{n+1}) \le V(z,x_n) - \left(1-(1+\sqrt{2})\frac{\lambda L}{\sigma}\right)V(y_n,x_n) - \left(1-\sqrt{2}\frac{\lambda L}{\sigma}\right)V(x_{n+1},y_n) + \frac{\lambda L}{\sigma}V(x_n,y_{n-1}) \tag{8}$$

holds, where $z \in S$.

Proof. Let $z \in S$. Let us write the three-point identity (3) as

$$V(z,x_{n+1}) = V(z,x_n) - V(x_{n+1},x_n) + (\nabla\varphi(x_{n+1}) - \nabla\varphi(x_n), x_{n+1} - z). \tag{9}$$

From the definition of point x_{n+1} and $z \in S \subseteq T_n$ it follows that

$$\lambda(Ay_n, z - x_{n+1}) + (\nabla\varphi(x_{n+1}) - \nabla\varphi(x_n), z - x_{n+1}) \ge 0. \tag{10}$$

Using inequalities (10) for the estimate of the scalar product in (9), we obtain

$$V(z,x_{n+1}) \le V(z,x_n) - V(x_{n+1},x_n) + \lambda(Ay_n, z - x_{n+1}). \tag{11}$$

From pseudo-monotonicity of A and $y_n \in C$ it follows that $(Ay_n, z - y_n) \le 0$. Adding the term $\lambda(Ay_n, y_n - z)$ to the right-hand side of inequality (11) yields

$$\begin{aligned} V(z,x_{n+1}) &\le V(z,x_n) - V(x_{n+1},x_n) + \lambda(Ay_n, y_n - x_{n+1}) \\ &= V(z,x_n) - V(x_{n+1},x_n) + \lambda(Ay_{n-1}, y_n - x_{n+1}) + \lambda(Ay_n - Ay_{n-1}, y_n - x_{n+1}). \end{aligned} \tag{12}$$

Let us write the term $\lambda(Ay_{n-1}, y_n - x_{n+1})$ as

$$\begin{aligned} \lambda(Ay_{n-1}, y_n - x_{n+1}) &= (\nabla\varphi(x_n) - \lambda Ay_{n-1} - \nabla\varphi(y_n), x_{n+1} - y_n) \\ &\quad + (\nabla\varphi(y_n) - \nabla\varphi(x_n), x_{n+1} - y_n). \end{aligned}$$

Inclusion $x_{n+1} \in T_n$ yields the inequality

$$(\nabla\varphi(x_n) - \lambda Ay_{n-1} - \nabla\varphi(y_n), x_{n+1} - y_n) \le 0.$$

Hence, we get

$$\lambda(Ay_{n-1}, y_n - x_{n+1}) \le (\nabla\varphi(y_n) - \nabla\varphi(x_n), x_{n+1} - y_n).$$

Using the three-point identity (3), we obtain

$$\lambda(Ay_{n-1}, y_n - x_{n+1}) \le V(x_{n+1},x_n) - V(x_{n+1},y_n) - V(y_n,x_n). \tag{13}$$

Estimating the right-hand side of (12) by means of (13), we obtain the inequality

$$V(z,x_{n+1}) \le V(z,x_n) - V(x_{n+1},y_n) - V(y_n,x_n) + \lambda(Ay_{n-1} - Ay_n, x_{n+1} - y_n). \tag{14}$$

Let us now estimate the term $\lambda(Ay_{n-1} - Ay_n, x_{n+1} - y_n)$. We get

$$\lambda(Ay_{n-1} - Ay_n, x_{n+1} - y_n) \le \lambda\|Ay_{n-1} - Ay_n\|_* \|x_{n+1} - y_n\|$$

$$\leq \lambda L \| y_{n-1} - y_n \| \| x_{n+1} - y_n \| \leq \lambda L \left\{ \frac{1}{2\sqrt{2}} \| y_{n-1} - y_n \|^2 + \frac{1}{\sqrt{2}} \| x_{n+1} - y_n \|^2 \right\}$$

$$\leq \frac{\lambda L}{2\sqrt{2}} \{ \sqrt{2} \| y_{n-1} - x_n \|^2 + (2+\sqrt{2}) \| x_n - y_n \|^2 \} + \frac{\lambda L}{\sqrt{2}} \| x_{n+1} - y_n \|^2$$

$$= \frac{\lambda L}{2} \| y_{n-1} - x_n \|^2 + \lambda L \frac{1+\sqrt{2}}{2} \| x_n - y_n \|^2 + \frac{\lambda L}{\sqrt{2}} \| x_{n+1} - y_n \|^2. \tag{15}$$

We have used the elementary inequalities

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

Estimating the norms in (15) by means of inequality (4), we get

$$\lambda (A y_n - A y_{n-1}, y_n - x_{n+1}) \leq \frac{\lambda L}{\sigma} V(x_n, y_{n-1})$$

$$+ \frac{\lambda L}{\sigma} (1+\sqrt{2}) V(y_n, x_n) + \frac{\lambda L}{\sigma} \sqrt{2} V(x_{n+1}, y_n). \tag{16}$$

Applying (16) in (14) yields

$$V(z, x_{n+1}) \leq V(z, x_n) - V(x_{n+1}, y_n) - V(y_n, x_n)$$

$$+ \frac{\lambda L}{\sigma} V(x_n, y_{n-1}) + \frac{\lambda L}{\sigma} (1+\sqrt{2}) V(y_n, x_n) + \frac{\lambda L}{\sigma} \sqrt{2} V(x_{n+1}, y_n)$$

$$\leq V(z, x_n) - \left(1 - \frac{\lambda L}{\sigma} \sqrt{2} \right) V(x_{n+1}, y_n) - \left(1 - \frac{\lambda L}{\sigma} (1+\sqrt{2}) \right) V(y_n, x_n) + \frac{\lambda L}{\sigma} V(x_n, y_{n-1}),$$

as was to be shown. ■

Let us prove the convergence of Algorithm 1.

CONVERGENCE OF ALGORITHM 1

To prove the convergence of the algorithm, we will need the elementary lemma about numerical sequences.

LEMMA 3. Let (a_n) and (b_n) be sequences of nonnegative numbers satisfying the inequality $a_{n+1} \leq a_n - b_n$ for all $n \in \mathbb{N}$. Then there exists a finite limit $\lim_{n \to \infty} a_n$ and $\sum_{n=1}^{\infty} b_n < +\infty$.

Let us formulate one of the main results of the present study.

THEOREM 1. Let set $C \subseteq E$ be convex and closed, operator $A: E \to E^*$ be pseudo-monotone and Lipschitz-continuous with the constant $L > 0$, $S \neq \varnothing$, and $\lambda \in \left(0, (\sqrt{2} - 1) \frac{\sigma}{L} \right)$. Then sequences (x_n) and (y_n) generated by Algorithm 1 converge to some point $\bar{z} \in S$.

Proof. Let $z \in S$. Assume that

$$a_n = V(z, x_n) + \frac{\lambda L}{\sigma} V(x_n, y_{n-1}),$$

$$b_n = \left(1 - \frac{\lambda L}{\sigma} (1+\sqrt{2}) \right) (V(y_n, x_n) + V(x_{n+1}, y_n)).$$

Inequality (8) becomes $a_{n+1} \le a_n - b_n$. Then from Lemma 3 about numerical sequences there follows the existence of the finite limit

$$\lim_{n\to\infty}\left(V(z,x_n)+\frac{\lambda L}{\sigma}V(x_n, y_{n-1})\right),$$

$$\sum_{n=1}^{\infty}\left(1-\frac{\lambda L}{\sigma}(1+\sqrt{2})\right)(V(y_n,x_n)+V(x_{n+1}, y_n)) < +\infty.$$

From here we obtain

$$\lim_{n\to\infty} V(y_n,x_n)=\lim_{n\to\infty} V(x_{n+1}, y_n)=0 \tag{17}$$

and convergence of the numerical sequence $(V(z,x_n))$ for all $z\in S$. From (17) it follows that

$$\lim_{n\to\infty}\|y_n-x_n\|=\lim_{n\to\infty}\|x_{n+1}-y_n\|=0. \tag{18}$$

Hence,

$$\lim_{n\to\infty}\|x_{n+1}-x_n\|=0. \tag{19}$$

The inequality $V(z,x_n)\ge\frac{\sigma}{2}\|z-x_n\|^2$ and (18) yield boundedness of the sequences (x_n) and (y_n).

Let us consider a subsequence (x_{n_k}) converging to some point $\bar{z}\in E$. Then from (18) it follows that $y_{n_k}\to\bar{z}$ and $x_{n_k+1}\to\bar{z}$ and $\bar{z}\in C$. Let us show that $\bar{z}\in S$. We have

$$(Ay_{n_k}, y-x_{n_k+1})+\frac{1}{\lambda}(\nabla\varphi(x_{n_k+1})-\nabla\varphi(x_{n_k}), y-x_{n_k+1})\ge 0 \ \ \forall\, y\in C\subseteq T_n. \tag{20}$$

Passing to the limit in (20) and taking into account (18) and (19), we obtain $(A\bar{z}, y-\bar{z})\ge 0 \ \ \forall\, y\in C$, i.e., $\bar{z}\in C$.

Let us show that $x_n\to\bar{z}$ (then from $\|x_n-y_n\|\to 0$ it follows that $y_n\to\bar{z}$ as well). It is known that there exists the limit

$$\lim_{n\to\infty} V(\bar{z},x_n)=\lim_{n\to\infty}(\varphi(\bar{z})-\varphi(x_n)-(\nabla\varphi(x_n),\bar{z}-x_n)).$$

Since $\lim_{n\to\infty} V(\bar{z},x_{n_k})=0$, we also have $\lim_{n\to\infty} V(\bar{z},x_n)=0$, whence $\|x_n-\bar{z}\|\to 0$. ■

EFFICIENCY ESTIMATE OF ALGORITHM 1

Let us consider variational inequality (1) with the monotone Lipschitz-continuous operator A and convex compact set C. For this case, let us obtain non-asymptotic estimates of the efficiency of Algorithm 1.

Let us remind one important concept. Gap function is a function of the form

$$G(x)=\max_{y\in C}(Ay, x-y), \ \ x\in C.$$

Gap function is convex, nonnegative, and takes zero value at point $x\in C$ if and only if this point belongs to the set S [1]. Gap function is often used to estimate the quality of approximate solution of variational inequality (1) [4, 6, 22].

The following theorem is true.

THEOREM 2. Let set $C\subseteq E$ be convex and compact, operator $A: E\to E^*$ be monotone and Lipschitz-continuous with constant $L>0$, and $\lambda\in\left(0,(\sqrt{2}-1)\frac{\sigma}{L}\right)$. Then the inequality

$$G(z_N)=\max_{y\in C}(Ay, z_N-y)\le\frac{\frac{1}{\lambda}R_C(x_1)+\frac{L}{\sigma}V(x_1, y_0)}{N}$$

holds, where $z_N=\frac{\sum_{n=1}^{N} y_n}{N}$ is an averaged output of Algorithm 1, $R_C(x_1)=\max_{y\in C} V(y,x_1)$.

Proof. For an arbitrary element $y \in C$, the inequality holds

$$V(y, x_{n+1}) \le V(y, x_n) - V(x_{n+1}, x_n) + \lambda(Ay_n, y - x_{n+1}).$$

From the monotonicity of operator A it follows that

$$(Ay_n, y - x_{n+1}) = (Ay_n, y - y_n) + (Ay_n, y_n - x_{n+1}) \le (Ay, y - y_n) + (Ay_n, y_n - x_{n+1}).$$

Thus,

$$V(y, x_{n+1}) \le V(y, x_n) - V(x_{n+1}, x_n) + \lambda(Ay_n, y_n - x_{n+1}) + \lambda(Ay, y - y_n)$$

$$= V(y, x_n) - V(x_{n+1}, x_n) + \lambda(Ay_{n-1}, y_n - x_{n+1}) + \lambda(Ay_n - Ay_{n-1}, y_n - x_{n+1}) + \lambda(Ay, y - y_n). \tag{21}$$

Like in the proof of Lemma 2, from (21) we obtain the inequality

$$V(y, x_{n+1}) \le V(y, x_n) - \left(1 - (1+\sqrt{2})\frac{\lambda L}{\sigma}\right) V(y_n, x_n)$$

$$- \left(1 - \sqrt{2}\frac{\lambda L}{\sigma}\right) V(x_{n+1}, y_n) + \frac{\lambda L}{\sigma} V(x_n, y_{n-1}) + \lambda(Ay, y - y_n). \tag{22}$$

Let us rearrange (22) as

$$\lambda(Ay, y_n - y) \le \left(V(y, x_n) + \frac{\lambda L}{\sigma} V(x_n, y_{n-1}) \right) - \left(V(y, x_{n+1}) + \frac{\lambda L}{\sigma} V(x_{n+1}, y_n) \right)$$

$$- \left(1 - \frac{\lambda L}{\sigma}(1+\sqrt{2})\right)(V(y_n, x_n) + V(x_{n+1}, y_n)). \tag{23}$$

Summing (23) over n from 1 to N yields

$$\lambda \sum_{n=1}^{N} (Ay, y_n - y) \le \left(V(y, x_1) + \frac{\lambda L}{\sigma} V(x_1, y_0) \right) - \left(V(y, x_{N+1}) + \frac{\lambda L}{\sigma} V(x_{N+1}, y_N) \right)$$

$$- \left(1 - \frac{\lambda L}{\sigma}(1+\sqrt{2})\right) \sum_{n=1}^{N} (V(y_n, x_n) + V(x_{n+1}, y_n)).$$

From here it follows that

$$(Ay, z_N - y) \le \frac{\frac{1}{\lambda} V(y, x_1) + \frac{L}{\sigma} V(x_1, y_0)}{N}, \tag{24}$$

where $z_N = \frac{\sum_{n=1}^{N} y_n}{N}$. Passing to the maximum with respect to $y \in C$ in (24), we obtain

$$G(z_N) = \max_{y \in C} (Ay, z_N - y) \le \frac{\frac{1}{\lambda} R_C(x_1) + \frac{L}{\sigma} V(x_1, y_0)}{N},$$

where $R_C(x_1) = \max_{y \in C} V(y, x_1)$. ■

The corollary below follows from Theorem 2.

COROLLARY 1. Let it be required to solve problem (1) by means of Algorithm 1 for $\lambda = \dfrac{\sigma}{3L}$ and $\varepsilon > 0$. Then the following estimate holds after

$$N = \left\lceil \frac{1}{\varepsilon} \frac{L}{\sigma} (3R_C(x_1) + V(x_1, y_0)) \right\rceil$$

iterations:

$$G(z_N) = \max_{y \in C} (Ay, z_N - y) \le \varepsilon,$$

where $z_N = \dfrac{\sum_{n=1}^{N} y_n}{N}$ is an averaged output of Algorithm 1 after N iterations.

CONCLUSIONS

We have proposed a new two-stage method for approximate solution of variational inequalities with pseudo-monotone and Lipschitz-continuous operators defined in a finite-dimensional linear normed space. This method is a modification of several two-stage algorithms described earlier [16, 18] with application of Bregman divergence instead of Euclidean distance. As well as other schemes that use Bregman divergence [4–6, 17, 19–23], in certain cases the method allows efficient application of the structure of feasible set of the problem, i.e., construction of the sequence of approximations by means of explicitly calculated operator of projection onto the feasible set.

The theorem about the convergence of the method has been proved. For the case of monotone operator and convex compact feasible set, we have obtained non-asymptotic estimates of the efficiency of the method.

Of interest is to construct an adaptive analog of the considered method, which would allow us to obtain approximating sequences with unknown exact value of the Lipschitz constant of the operator. Note also that advent of generative adversarial networks (GAN) has caused an interest in adaptive algorithms of solution of variational inequalities among machine training experts.

REFERENCES

1. I. V. Konnov, Combined Relaxation Methods for Variational Inequalities, Springer-Verlag, Berlin–Heidelberg–New York (2001).
2. A. Nagurney, Network Economics: A Variational Inequality Approach, Kluwer Academic Publishers, Dordrecht (1999).
3. G. M. Korpelevich, "The extra-gradient method for finding saddle points and other problems," Ekonomika i Mat. Metody, Vol. 12, No. 4, 747–756 (1976).
4. 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. on Optimization, Vol. 15, 229–251 (2004).
5. A. Auslender and M. Teboulle, "Interior projection-like methods for monotone variational inequalities," Mathematical Programming, Vol. 104, Iss. 1, 39–68 (2005).
6. Yu. Nesterov, "Dual extrapolation and its applications to solving variational inequalities and related problems," Mathematical Programming, Vol. 109, Iss. 2–3, 319–344 (2007).
7. E. A. Nurminski, "The use of additional diminishing disturbances in Fejer models of iterative algorithms," J. Comp. Math. Math. Physics, Vol. 48, No. 12, 2154–2161 (2008).
8. V. V. Semenov, "On the parallel proximal decomposition method for solving the problems of convex optimization," J. Autom. Inform. Sci., Vol. 42, Iss. 4, 13–18 (2010).

9. Y. Censor, A. Gibali, and S. Reich, “The subgradient extragradient method for solving variational inequalities in Hilbert space,” J. of Optimization Theory and Applications, Vol. 148, 318–335 (2011).
10. S. I. Lyashko, V. V. Semenov, and T. A. Voitova, “Low-cost modification of Korpelevich’s methods for monotone equilibrium problems,” Cybern. Syst. Analysis, Vol. 47, No. 4, 631–639 (2011).
11. V. V. Semenov, “A strongly convergent splitting method for the system of operator inclusions with monotone operators,” J. Autom. Inform. Sci., Vol. 46, Iss. 5, 45–56 (2014).
12. V. V. Semenov, “Hybrid splitting metods for the system of operator inclusions with monotone operators,” Cybern. Syst. Analysis, Vol. 50, No. 5, 741–749 (2014).
13. D. A. Verlan, V. V. Semenov, and L. M. Chabak, “A strongly convergent modified extragradient method for variational inequalities with non-Lipschitz operators,” J. Autom. Inform. Sci., Vol. 47, Iss. 7, 31–46 (2015).
14. S. V. Denisov, V. V. Semenov, and L. M. Chabak, “Convergence of the modified extragradient method for variational inequalities with non-Lipschitz operators,” Cybern. Syst. Analysis, Vol. 51, No. 5, 757–765 (2015).
15. L. D. Popov, “A modification of the of Arrow–Hurwitz method for finding saddle points,” Mat. Zametki, Vol. 28, No. 5, 777–784 (1980).
16. Yu. V. Malitsky and V. V. Semenov, “An extragradient algorithm for monotone variational inequalities,” Cybern. Syst. Analysis, Vol. 50, No. 2, 271–277 (2014).
17. V. V. Semenov, “A version of the mirror descent method to solve variational inequalities,” Cybern. Syst. Analysis, Vol. 53, No. 2, 234–243 (2017).
18. S. I. Lyashko and V. V. Semenov, “A new two-step proximal algorithm of solving the problem of equilibrium programming,” in: B. Goldengorin (ed.), Optimization and Its Applications in Control and Data Sciences, Springer Optimization and Its Applications, Springer, Cham, Vol. 115 (2016), 315–325.
19. L. M. Bregman, “A relaxation method to find a common point of convex sets and its application to solve convex programming problems,” Zhurn. Vych. Mat. Mat. Fiziki, Vol. 7, No. 3, 620–631 (1967).
20. A. S. Nemirovskii and D. B. Yudin, Problem Complexity and Efficiency of Optimization Methods [in Russian], Nauka, Moscow (1979).
21. A. Beck, First-Order Methods in Optimization, Society for Industrial and Applied Mathematics, Philadelphia (2017).
22. V. V. Semenov, “Modified extragradient method with Bregman divergence for variational inequalities,” J. Autom. Inform. Sci., Vol. 50, Iss. 8, 26–37 (2018).
23. D. A. Lorenz, F. Schopfer, and S. Wenger, “The linearized Bregman method via split feasibility problems,” Analysis and Generalizations. SIAM J. on Imaging Sciences, Vol. 7, Iss. 2, 1237–1262 (2014).