УДК 517.9
MSC 47J20, 49J40, 65K15, 90C25

A NEW MIRROR-PROX ALGORITHM FOR VARIATIONAL INEQUALITIES

S. V. Denisov, V. V. Dudar, V. V. Semenov, Ya. I. Vedel
Faculty of Computer Science and Cybernetics, Taras Shevchenko Kiev National University, Kiev, Ukraine, E-mail: sireukr@gmail.com, slavko123@ukr.net, volodya.semenov@gmail.com, yandx@ukr.net.

НОВИЙ ДЗЕРКАЛЬНО-ПРОКСИМАЛЬНИЙ АЛГОРИТМ ДЛЯ ВАРІАЦІЙНИХ НЕРІВНОСТЕЙ

С. В. Денисов, В. В. Дудар, В. В. Семенов, Я. І. Ведель
Факультет комп'ютерних наук та кібернетики, Київський національний університет імені Тараса Шевченка, Київ, Україна, E-mail: sireukr@gmail.com, slavko123@ukr.net, volodya.semenov@gmail.com, yandx@ukr.net.

Abstract. In this paper, we consider a variational inequalities with Lipschitz continuous pseudo-monotone operators. Quite a number of operational research problems in applications can be stated in this form. We propose new variant of mirror descent method (mirror-prox algorithm) for solving the variational inequalities. This method can be interpreted as the modification of two-step L. D. Popov algorithm with the projection onto the feasible set in the sense of Bregman divergence. Our method, like other mirror descent schemes, can effectively take into account the structure of the feasible set of the problem. The main theoretical result is the proof of the theorem about the convergence of the method. Several preliminary numerical experiments have been also performed to illustrate the convergence of the method.
Keywords: Variational inequality, Bregman divergence, Mirror-Prox Algorithm, Convergence.

Резюме. В статті розглядаються варіаційні нерівності з ліпшицевими та псевдомонотонними операторами. Велика кількість важливих прикладних задач дослідження операцій може бути сформульована у цій формі. Для розв'язання варіаційних нерівностей пропонується новий метод типу дзеркального спуску (дзеркально-проксимальний алгоритм). Метод можна проінтерпретувати як модифікацію алгоритму Попова з використанням проектування на допустиму множину у розумінні відстані Брегмана. Основний теоретичний результат — теорема про збіжність методу. Також наведено результати декількох чисельних експериментів.
Ключові слова: варіаційна нерівність, відстань Брегмана, дзеркально-проксимальний алгоритм, збіжність.

1. Introduction

There are a lot of interesting and actual problems in operation research that can be written in the form of variational inequalities. The solving of the last is the actively developing field of applied nonlinear analysis [1–9]. There are currently a lot of methods to solve variational inequalities, including projection type methods, i.e. using a metric projection onto the feasible set [1, 3, 10–13]. It's known that in the saddle point search, Nash equilibrium problems the convergence of the most simple projection method requires strengthened monotonicity conditions [1]. In the case of non-compliance there are several approaches. One of them is a regularization of the original problem in order to give it the desired property. In extra-gradient type methods first proposed by G. M. Korpelevich [13] the convergence is achieved without the modification of problems. The study of these methods was performed in many papers [14–24]. In 2011, the authors in [16, 17] have replaced the second projection onto any closed convex set in the extra-gradient method by one onto a half-space and proposed the subgradient extra-gradient method for variational inequalities in Hilbert spaces, see also [23, 24].

In 1980, L. D. Popov [25] proposed very interesting modification of Arrow-Hurwicz scheme for approximation of saddle points of convex-concave functions in Euclidean space. Let X and Y are closed convex subset of Euclidean spaces $\mathbb{R}^d$ and $\mathbb{R}^p$, respectively, and $L : X \times Y \to \mathbb{R}$ be a differentiable convex-concave function. Then, the method [25] approximation of saddle points of L on $X \times Y$ can be written as

$$\begin{cases} x_1\,, \;\; \bar{x}_1 \in X\,, \;\; y_1\,, \;\; \bar{y}_1 \in Y\,, \;\; \lambda > 0, \\ x_{n+1} = P_X\left(x_n - \lambda L'_1(\bar{x}_n, \bar{y}_n)\right), \\ y_{n+1} = P_Y\left(y_n + \lambda L'_2(\bar{x}_n, \bar{y}_n)\right), \\ \bar{x}_{n+1} = P_X\left(x_{n+1} - \lambda L'_1(\bar{x}_n, \bar{y}_n)\right), \\ \bar{y}_{n+1} = P_Y\left(y_{n+1} + \lambda L'_2(\bar{x}_n, \bar{y}_n)\right), \end{cases}$$

where P_X and P_Y are metric projection onto X and Y, respectively, L'_1 and L'_2 are partial derivatives. Under some suitable assumptions, L.D. Popov proved the convergence of this method. In recent works [26, 27] proved the convergence of this algorithm for variational inequalities with monotone and Lipschitz operators in infinite-dimensional Hilbert space, and proposed some modifications of this algorithm.

Euclidean distance and projection were used in all these methods. And often this does not allow to take into account the structure of feasible sets and solve problems effectively. A possible solution to the situation is a more flexible selection of the distance for projection onto the feasible set. One of the first successful implementations of this strategy is the work of L. M. Bregman [28] proposed a cyclic non-Euclidean projection method for finding a common point of convex sets. This work has opened the wide scientific field in mathematical programming and nonlinear analysis.

The mirror descent method was proposed in the late 70-ies of the last century by A. S. Nemirovski and D. B. Yudin for solving convex optimization problems [29]. Since then the method has been widely used for solving large-scale problems

[30–32]. For problems with constraints this method can be interpreted as a variant of the subgradient projection method when projecting is understood in the sense of Bregman divergence (Bregman distance) [32]. The mirror descent method allows to take into account the structure of feasible set of optimization problems. For example, for the probability simplex we can use the Kullback-Leibler divergence that is the Bregman divergence built on negative entropy. And then we have explicitly calculated projection operator on the simplex [32]. Versions of the mirror descent method for solving variational inequalities and saddle problems based on the Korpelevich extra-gradient algorithm are studied in [2, 30, 33–35]. These includes also stochastic methods [30, 34].

In this paper we study a new version of the mirror descent method for solving variational inequalities with Lipschitz continuous and pseudo-monotone operators based on the two-step L. D. Popov algorithm [25–27].

The remainder of the paper is organized as follows. In Sect. 2 we formulate the problem and introduce all necessary constructions. In Sect. 3 we propose a new variant of mirror descent method (mirror-prox algorithm) for the variational inequalities and consider several versions for solving more specific problems. The convergence behavior of the proposed algorithm is studied in Sect. 4. In Sect. 5 we perform several numerical experiments to illustrate the computational performance of the proposed algorithm. Finally, Sect. 6 contains concluding remark.

2. Preliminaries

For any finite-dimensional real vector space E, we denote by E^* its dual. We denote the value of a linear function $a \in E^*$ at $b \in E$ by (a, b). Let $\|\cdot\|$ denote some norm on E (not necessary Euclidean) and $\|\cdot\|_*$ denote the norm on E^*, which is dual to $\|\cdot\|$

$$\|a\|_* = \max\{(a, b) : \ \|b\| = 1\} \ .$$

Let C be a nonempty subset of space E, A be a operator, that acts from E to E^*. Consider the variational inequality problem:

$$\text{find } \ x \in C \ \text{ such that } \ \ (Ax, y - x) \geq 0 \quad \forall\, y \in C. \tag{1}$$

The set of solutions of the problem (1) is denoted S.

Assume that the following conditions are satisfied:

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

Remark 1. Recall, that operator A on the set C is called pseudo-monotone if for all x, $y \in C$ from $(Ax, y - x) \geq 0$ follows $(Ay, y - x) \geq 0$ [1].

Consider, so-called, dual variational inequality [1]:

$$\text{find } \ x \in C \ \text{ such that } \ \ (Ay, y - x) \geq 0 \quad \forall\, y \in C. \tag{2}$$

The set of solutions of (2) we will denote as S^d. Inequality (2) sometimes is called weak or dual formulation of (1), and solutions of (2) – weak solutions of (1) [1]. Indeed, if A is pseudo-monotone we have that $S \subseteq S^d$. With our conditions we have that $S^d = S$. Particularly, the set S is convex and closed [1].

We will set the construction necessary for algorithm formulation. Let function $\varphi : E \to \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ satisfies the condition [34]:

- φ is continuous and convex on C. Particularly, the set
$$C^o = \{x \in C : \ \ \partial\varphi(x) \neq \emptyset\}$$
is nonempty;
- φ is regular on C^o, i.e. subdifferential $\partial\varphi$ on the set C^o has continuous selector $\nabla\varphi$;
- function φ is strongly convex with respect to the chosen norm $\|\cdot\|$ with constant of strong convexity $\sigma > 0$:
$$\varphi(a) \geq \varphi(b) - (\nabla\varphi(b), a - b) + \frac{\sigma}{2}\|a - b\|^2 \quad \forall a \in C, \ \ b \in C^o.$$

Remark 2. Such functions are called «distance generating functions» [34].

Remark 3. The minimization problem
$$(a, y) + \varphi(y) \to \min_{y \in C}, \quad a \in E^*,$$
has only one solution that lies in C^o.

The Bregman divergence associated with φ is defined as
$$d(a, b) = \varphi(a) - \varphi(b) - (\nabla\varphi(b), a - b) \quad \forall a \in C, \ \ b \in C^o.$$

Remark 4. Consider two main examples. If $\varphi(\cdot) = \frac{1}{2}\|\cdot\|_2^2$, where $\|\cdot\|_2$ is Euclidean norm, we will have $d(x, y) = \frac{1}{2}\|x - y\|_2^2$. For probability simplex
$$S_m = \left\{x \in \mathbb{R}^m : \ \ x_i \geq 0, \ \ \sum_{i=1}^{m} x_i = 1\right\}$$
and negative Boltzmann-Shannon entropy $\varphi(x) = \sum_{i=1}^{m} x_i \ln x_i$ (it is strongly convex with respect to the ℓ_1-norm on S_m) we obtain Kullback-Leibler divergence (KL-divergence)
$$d(x, y) = \sum_{i=1}^{m} x_i \ln\frac{x_i}{y_i}, \quad x \in S_m, \ \ y \in \operatorname{ri}(S_m).$$

Also it is performed useful 3-point identity [32]:
$$d(a, c) = d(a, b) + d(b, c) + (\nabla\varphi(b) - \nabla\varphi(c), a - b). \tag{3}$$

From strong convexity φ we can estimate
$$d(a, b) \geq \frac{\sigma}{2}\|a - b\|^2 \quad \forall a \in C, \ \ b \in C^o. \tag{4}$$

Suppose, that we have an ability to solve effectively following strongly convex minimization problems:
$$\pi_x(a) = \arg\min_{y \in C}\{-(a, y - x) + d(y, x)\} \quad \forall a \in E^*, \ \ x \in C^o.$$

The point $\pi_x(a)$ in Euclidean case coincides with Euclidean metric projection

$$P_C(x+a) = \arg\min_{y\in C} \|y-(x+a)\|_2 .$$

For probability simplex case S_m and KL-divergence we have [32]

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

Operator $\pi_x : E^* \to C^o$ is called prox mapping.

3. The Algorithm

Let us describe the Mirror-Prox Algorithm for problems (1).

Algorithm 1. Mirror-Prox Algorithm for Variational Inequalities

---

Choose initial points $x_1 \in C^o$, $y_1 \in C$, and number $\lambda > 0$. Generate the sequence of elements x_n, y_n using iterative scheme

$$x_{n+1} = \pi_{x_n}(-\lambda A y_n),$$
$$y_{n+1} = \pi_{x_{n+1}}(-\lambda A y_n).$$

The rule how to choose the parameter λ we will specify in the next section.

Remark 5. If $\varphi(\cdot) = \frac{1}{2}\|\cdot\|_2^2$, then Algorithm 1 takes the form [25, 26, 40, 42]:

$$\begin{cases} x_{n+1} = P_C(x_n - \lambda A y_n), \\ y_{n+1} = P_C(x_{n+1} - \lambda A y_n). \end{cases}$$

We will show several specific versions of Algorithm 1.

Consider the variational inequality on the probability simplex:

$$\text{find } \ x\in S_m \ \text{ such that } \ (Ax, y-x) \ge 0 \quad \forall\, y\in S_m.$$

If we choose KL-divergence we obtain the next version of Algorithm 1:

$$\begin{cases} x_i^{n+1} = \frac{x_i^n \exp\left(-\lambda (A y_n)_i\right)}{\sum_{j=1}^m x_j^n \exp\left(-\lambda (A y_n)_j\right)}, & i = 1,\dots,m, \\ y_i^{n+1} = \frac{x_i^{n+1} \exp\left(-\lambda (A y_n)_i\right)}{\sum_{j=1}^m x_j^{n+1} \exp\left(-\lambda (A y_n)_j\right)}, & i = 1,\dots,m, \end{cases}$$

where $(Ay_n)_i \in \mathbb{R}$ is i-th coordinate of vector $Ay_n \in \mathbb{R}^m$, $\lambda > 0$.

In network equilibrium problems, machine learning and game theory we have to work with variational inequalities with direct products of scaled simplex's

$$C = \prod_{k=1}^p r_k S_{m_k} \subseteq \mathbb{R}^{\sum_{k=1}^p m_k},$$

where $r_k S_{m_k} = \{x\in\mathbb{R}^{m_k} :\ x_i \ge 0,\ \sum_{i=1}^{m_k} x_i = r_k\}$, $r_k > 0$, i.e. with problems

$$\text{find } \ x \in \prod_{k=1}^p r_k S_{m_k} \ \text{ such that } \ (Ax, y-x) \ge 0 \quad \forall\, y\in \prod_{k=1}^p r_k S_{m_k}. \tag{5}$$

From separable function

$$\varphi(x) = \sum_{k=1}^{p} \varphi_k(x_k) = \sum_{k=1}^{p} \sum_{i=1}^{m_k} \frac{x_{k,i}}{r_k} \ln \frac{x_{k,i}}{r_k},$$

where $x = (x_1, \ldots, x_p) = \left(\underbrace{x_{1,1}, \ldots, x_{1,m_1}}_{x_1}, \ldots, \underbrace{x_{p,1}, \ldots, x_{p,m_p}}_{x_p} \right) \in \mathbb{R}^{\sum_{k=1}^{p} m_k}$,
we build Bregman divergence on $\prod_{k=1}^{p} r_k S_{m_k}$:

$$d(x,y) = \sum_{k=1}^{p} d_k(x_k, y_k) = \sum_{k=1}^{p} \sum_{i=1}^{m_k} \frac{x_{k,i}}{r_k} \ln \frac{x_{k,i}}{y_{k,i}}.$$

Algorithm 1 for variational inequality (5) with such choose of Bregman divergence takes the form:

$$\begin{cases} x_{k,i}^{n+1} = r_k \frac{x_{k,i}^{n} \exp\left(-\lambda r_k (Ay_n)_{k,i}\right)}{\sum_{j=1}^{m_k} x_{k,j}^{n} \exp\left(-\lambda r_k (Ay_n)_{k,j}\right)}, & k = 1, \ldots, p,\ i = 1, \ldots, m_k, \\ y_{k,i}^{n+1} = r_k \frac{x_{k,i}^{n+1} \exp\left(-\lambda r_k (Ay_n)_{k,i}\right)}{\sum_{j=1}^{m_k} x_{k,j}^{n+1} \exp\left(-\lambda r_k (Ay_n)_{k,j}\right)}, & k = 1, \ldots, p,\ i = 1, \ldots, m_k, \end{cases}$$

where $(Ay_n)_{k,i}$ – $\left(\sum_{t=1}^{k-1} m_t + i\right)$-th coordinate of vector $Ay_n \in \mathbb{R}^{\sum_{k=1}^{p} m_k}$, $\lambda > 0$.

Notice, that if for some $n \in \mathbb{N}$ the equality is fulfilled

$$x_{n+1} = x_n = y_n \tag{6}$$

then $y_n \in S$ and the following stationarity condition holds $x_k = y_k = y_n$ for all $k \geq n$. Indeed, the equality $x_{n+1} = \pi_{x_n}(-\lambda A y_n)$ means that

$$(Ay_n, y - x_{n+1}) + \lambda^{-1} (\nabla\varphi(x_{n+1}) - \nabla\varphi(x_n), y - x_{n+1}) \geq 0 \quad \forall y \in C.$$

From (6) we have that $(Ay_n, y - y_n) \geq 0$ $\forall y \in C$, i.e. $y_n \in S$.

Further, we assume that for all numbers $n \in \mathbb{N}$ the condition (6) doesn't hold. In the following section the convergence of the sequences (x_n), (y_n) generated by the Algorithm 1 is proved.

4. Main Results

We start the analysis of the convergence with the proof of important inequality for sequences (x_n) and (y_n), generated by the Algorithm 1.

Lemma 1. *Let sequences (x_n), (y_n) be generated by the Algorithm 1, and let $z \in S$. Then, we have*

$$d(z, x_{n+1}) \leq d(z, x_n) - \left(1 - \left(1 + \sqrt{2}\right) \frac{\lambda L}{\sigma}\right) d(y_n, x_n) - \\ - \left(1 - \sqrt{2}\frac{\lambda L}{\sigma}\right) d(x_{n+1}, y_n) + \frac{\lambda L}{\sigma} d(x_n, y_{n-1}). \tag{7}$$

Proof. We have (using twice the identity (3))

$$d(z,x_{n+1}) = d(z,x_n) - d(x_{n+1},x_n) + (\nabla\varphi(x_{n+1}) - \nabla\varphi(x_n), x_{n+1} - z) = \\ = d(z,x_n) - d(x_{n+1},y_n) - d(y_n,x_n) - \\ - (\nabla\varphi(y_n) - \nabla\varphi(x_n), x_{n+1} - y_n) + (\nabla\varphi(x_{n+1}) - \nabla\varphi(x_n), x_{n+1} - z). \quad (8)$$

From definition of points x_{n+1} and y_n it follows that

$$\lambda (Ay_n, z - x_{n+1}) + (\nabla\varphi(x_{n+1}) - \nabla\varphi(x_n), z - x_{n+1}) \geq 0, \quad (9)$$

$$\lambda (Ay_{n-1}, x_{n+1} - y_n) + (\nabla\varphi(y_n) - \nabla\varphi(x_n), x_{n+1} - y_n) \geq 0. \quad (10)$$

Using inequalities (9), (10) for estimation in (8), we obtain

$$d(z,x_{n+1}) \leq d(z,x_n) - d(x_{n+1},y_n) - d(y_n,x_n) + \\ +\lambda \{(Ay_{n-1}, x_{n+1} - y_n) + (Ay_n, z - x_{n+1})\} = \\ = d(z,x_n) - d(x_{n+1},y_n) - d(y_n,x_n) + \\ +\lambda \{(Ay_{n-1} - Ay_n, x_{n+1} - y_n) + (Ay_n, z - y_n)\}. \quad (11)$$

Operator A is pseudo-monotone, so $(Ay_n, z - y_n) \leq 0$. Using this inequality in (11), we get

$$d(z,x_{n+1}) \leq d(z,x_n) - d(x_{n+1},y_n) - d(y_n,x_n) + \\ +\lambda (Ay_{n-1} - Ay_n, x_{n+1} - y_n). \quad (12)$$

Now we will estimate the term $\lambda (Ay_{n-1} - Ay_n, x_{n+1} - y_n)$. We have

$$\lambda (Ay_{n-1} - Ay_n, x_{n+1} - y_n) \leq \lambda \|Ay_{n-1} - Ay_n\|_* \|x_{n+1} - y_n\| \leq \\ \leq \lambda L \|y_{n-1} - y_n\| \|x_{n+1} - y_n\| \leq \\ \leq \lambda L \left\{ \frac{1}{2\sqrt{2}} \|y_{n-1} - y_n\|^2 + \frac{1}{\sqrt{2}} \|x_{n+1} - y_n\|^2 \right\} \leq \\ \leq \frac{\lambda L}{2\sqrt{2}} \left\{ \sqrt{2} \|y_{n-1} - x_n\|^2 + \left(2 + \sqrt{2}\right) \|x_n - y_n\|^2 \right\} + \frac{\lambda L}{\sqrt{2}} \|x_{n+1} - y_n\|^2 = \\ \leq \frac{\lambda L}{2} \|y_{n-1} - x_n\|^2 + \lambda L \frac{1+\sqrt{2}}{2} \|x_n - y_n\|^2 + \frac{\lambda L}{\sqrt{2}} \|x_{n+1} - y_n\|^2. \quad (13)$$

Here we used elementary inequalities

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

After estimation the norms in (13) using inequality (4), we obtain

$$\lambda (Ay_{n-1} - Ay_n, x_{n+1} - y_n) \leq \frac{\lambda L}{\sigma} d(x_n, y_{n-1}) + \\ + \frac{\lambda L}{\sigma} \left(1 + \sqrt{2}\right) d(y_n, x_n) + \frac{\lambda L}{\sigma} \sqrt{2} d(x_{n+1}, y_n). \quad (14)$$

Applying (14) in (12), we have

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

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

i.e. the inequality (7). □

To prove the convergence we need next elementary fact.

Lemma 2. *Let non-negative sequences* (a_n), (b_n) *such that*

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

Then exists the limit $\lim_{n\to\infty}a_n\in\mathbb{R}$ *and* $\sum_{n=1}^{\infty}b_n<+\infty$.

Now we can formulate the main result.

Theorem 1. *Let* $C\subseteq E$ *is nonempty convex closed set, operator* $A:E\to E^*$ *is pseudo-monotone and Lipschitz continuous with a constant* $L>0$ *and* $S\neq\emptyset$. *Assume that* $\lambda\in\left(0,\left(\sqrt{2}-1\right)\frac{\sigma}{L}\right)$. *Then sequences* (x_n), (y_n), *that generated by the Algorithm 1, converge to the solution* $\bar{z}\in C$ *of the problem (1).*

Proof. Let $z\in S$. Assume

$$a_n=d\left(z,x_n\right)+\lambda L\sigma^{-1}d\left(x_n,y_{n-1}\right),$$
$$b_n=\left(1-\lambda L\sigma^{-1}\left(1+\sqrt{2}\right)\right)\left(d\left(y_n,x_n\right)+d\left(x_{n+1},y_n\right)\right).$$

The inequality (7) takes the form

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

Then from Lemma 2 we can conclude, that it exists the limit

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

and

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

Wherefrom we obtain

$$\lim_{n\to\infty}d\left(y_n,x_n\right)=\lim_{n\to\infty}d\left(x_{n+1},y_n\right)=0 \tag{15}$$

and convergence of sequence $\left(d\left(z,x_n\right)\right)$ for all $z\in S$. From (15) follows

$$\lim_{n\to\infty}\left\|y_n-x_n\right\|=\lim_{n\to\infty}\left\|x_{n+1}-y_n\right\|=0 \tag{16}$$

and naturally

$$\lim_{n\to\infty}\left\|x_{n+1}-x_n\right\|=0. \tag{17}$$

From inequality

$$d\left(z,x_n\right)\geq\frac{\sigma}{2}\left\|z-x_n\right\|^2$$

and (16) follows that sequences (x_n), (y_n) are bounded.

Consider the subsequence (x_{n_k}), which converges to some point $\bar{z}\in C$. Then from (16) follows that $y_{n_k}\to\bar{z}$ and $x_{n_k+1}\to\bar{z}$. Show that $\bar{z}\in S$. We have

$$\left(Ay_{n_k},y-x_{n_k+1}\right)+\frac{1}{\lambda}\left(\nabla\varphi\left(x_{n_k+1}\right)-\nabla\varphi\left(x_{n_k}\right),y-x_{n_k+1}\right)\geq 0\quad\forall y\in C. \tag{18}$$

Passing to the limit (18) taking into account (17), we get

$$(A\bar{z}, y - \bar{z}) \geq 0 \quad \forall y \in C,$$

i.e. $\bar{z} \in C$.

Now we show that $x_n \to \bar{z}$ (then from $\|x_n - y_n\| \to 0$ it will follow that also $y_n \to \bar{z}$). It is known, that the limit

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

exists. Because the $\lim_{n\to\infty} d(\bar{z}, x_{n_k}) = 0$, so $\lim_{n\to\infty} d(\bar{z}, x_n) = 0$. Wherefrom we have $\|x_n - \bar{z}\| \to 0$. □

Remark 6. If $\sigma = 1$, so we can use the scheme:

$$\begin{cases} x_{n+1} = \pi_{x_n}\left(-\frac{1}{3L} A y_n\right), \\ y_{n+1} = \pi_{x_{n+1}}\left(-\frac{1}{3L} A y_n\right). \end{cases}$$

5. Computational experiments

This section studies the numerical behavior of Algorithm 1 on a test problem which is related to the PageRank computation.

Consider the optimization problem on the probability simplex $S_N \subseteq \mathbb{R}^N$:

$$\text{find } \ x \in S_N \ \text{ such that } \ \|Ax - x\|_\infty = \min_{\zeta\in S_N} \|A\zeta - \zeta\|_\infty, \tag{19}$$

with a $N \times N$ column-stochastic matrix A and the ℓ_∞-norm $\|\cdot\|_\infty$.

We use game approach proposed in [36, 37] for original PageRank problem. Using representation

$$\|Ax - x\|_\infty = \max_{y\in B_1} (y, Ax - x), \ \ B_1 = \left\{y \in \mathbb{R}^N : \|x\|_1 \leq 1\right\},$$

we transform the optimization problem (19) to the form of a saddle point problem:

$$\min_{x\in S_N} \max_{y\in B_1} (y, Ax - x) = \max_{y\in B_1} \min_{x\in S_N} (y, Ax - x). \tag{20}$$

Saddle point problem (20) is equivalent to the variational inequality

$$\begin{aligned} &\text{find } \ x \in S_N, \ y \in B_1 \ \text{ such that } \ (A^* y - y, \zeta - x) + \\ &\quad + (x - Ax, \eta - y) \geq 0 \quad \forall\, \zeta \in S_N \ \forall\, \eta \in B_1. \end{aligned} \tag{21}$$

For solving the problem (21) we apply the Algorithm 1. In this case it takes the form

$$\begin{cases} x_{n+1} = \pi_{x_n}^{S_N}(\lambda(E - A^*)\eta_n), \\ y_{n+1} = \pi_{y_n}^{B_1}(\lambda(A - E)\zeta_n), \\ \zeta_{n+1} = \pi_{x_{n+1}}^{S_N}(\lambda(E - A^*)\eta_n), \\ \eta_{n+1} = \pi_{y_{n+1}}^{B_1}(\lambda(A - E)\zeta_n), \end{cases}$$

where $\pi_x^{S_N} : \mathbb{R}^N \to S_N$, $\pi_y^{B_1} : \mathbb{R}^N \to B_1$ are suitable prox mappings, $\lambda > 0$.

For the ℓ_1-ball B_1 we used only the Euclidean distance (Euclidean setting). For probability simplex S_N we used Euclidean distance or KL-divergence (Entropy setting). Column-stochastic matrices dimensionality 100×100, 1000×1000 and 2000×2000 are generated randomly. The projections onto

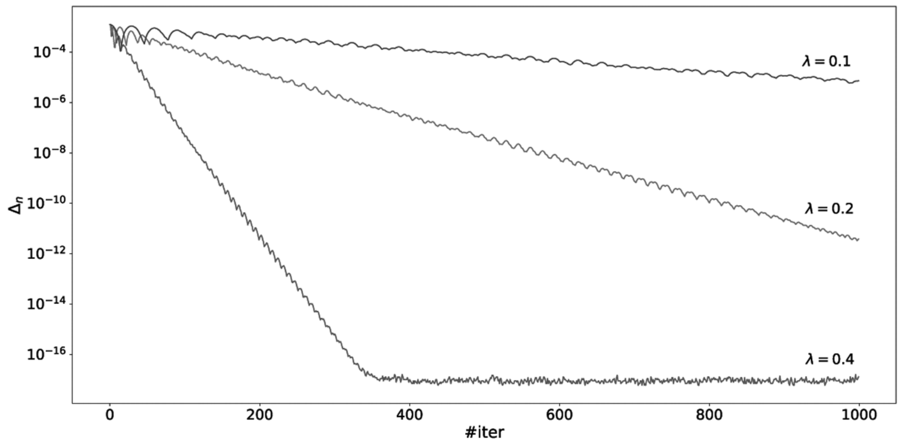

FIG. 1. Δ_n and # iter for $N = 100$, Euclidean–Euclidean, elapsed time 0.26 sec

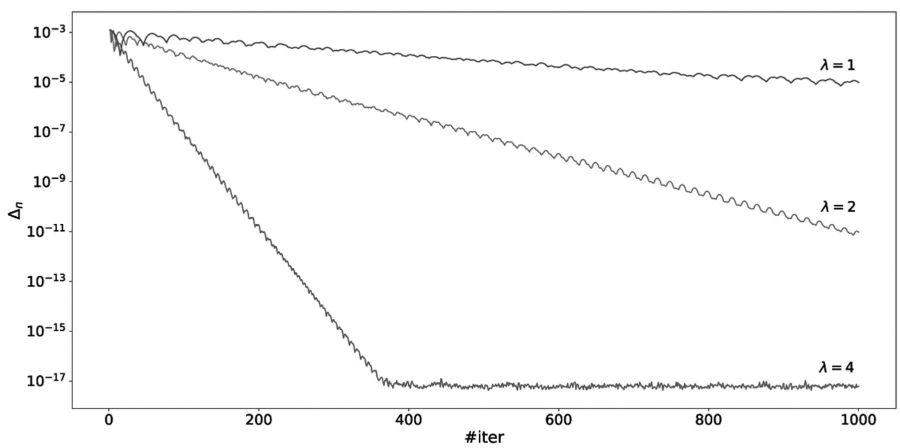

FIG. 2. Δ_n and # iter for $N = 100$, Entropy–Euclidean, elapsed time 0.25 sec

the simplex and ℓ_1-ball are implemented by the efficient algorithm [38]. The starting points x_1, y_1, ζ_1 and η_1 are chosen as $(1/N, 1/N, ..., 1/N)$.

To illustrate the numerical behavior of Algorithm 1, we have performed experiments for number of iterations (# iter). Figs. 1–6 describe the behavior of

$$\Delta_n = \|A\zeta_n - \zeta_n\|_\infty$$

generated by Algorithm 1 for various stepsizes λ. In these figures, the y-axes represent for value of Δ_n while the x-axes are for number of iterations.

All programs are implemented on a Asus Laptop Intel(R) Pentium(R) CPU B980 @ 2.40GHz 2.40 GHz, RAM 4.00 GB (using Code::Blocks environment on C++ language).

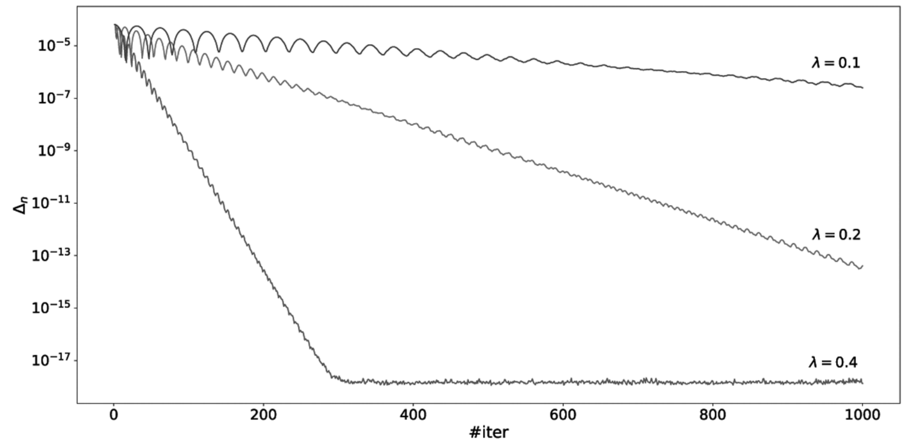

FIG. 3. Δ_n and # iter for $N = 1000$, Euclidean–Euclidean, elapsed time 26.682 sec

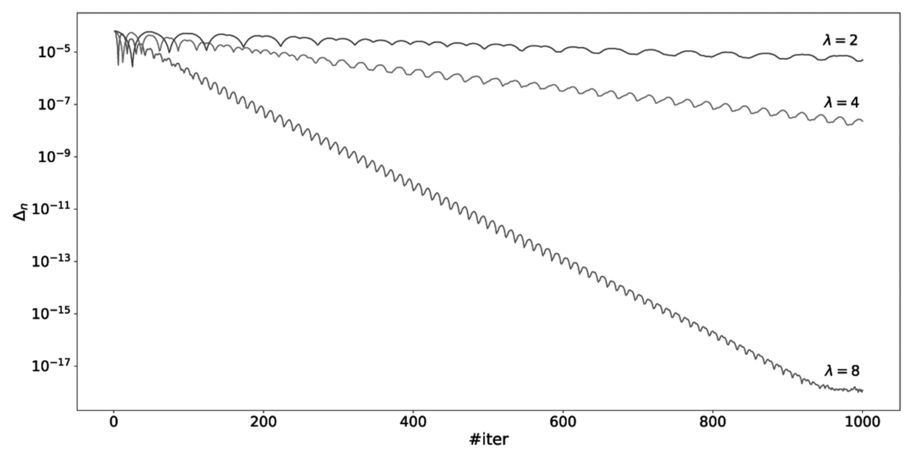

FIG. 4. Δ_n and # iter for $N = 1000$, Entropy–Euclidean, elapsed time 26.238 sec

6. Conclusion

In this paper we propose new variant of mirror descent method (mirror-prox algorithm) for solving the variational inequalities with pseudo-monotone operators. This method can be interpreted as the modification of two-step L. D. Popov algorithm with the projection on the feasible set in the sense of Bregman divergence. Our method, like other mirror descent schemes, can effectively take into account the structure of the feasible set of the problem. The main theoretical result is the proof of the theorem about the convergence of the method. Several preliminary numerical experiments have been also performed to illustrate the convergence of the method.

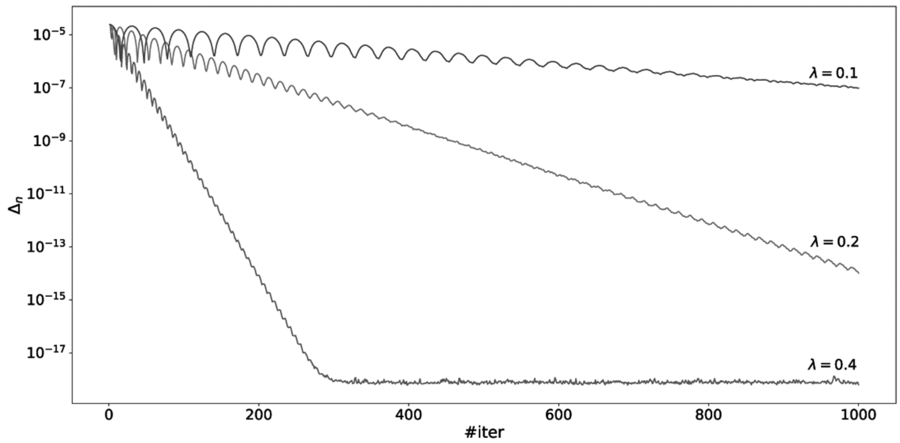

FIG. 5. Δ_n and # iter for $N = 2000$, Euclidean–Euclidean, elapsed time 104.461 sec

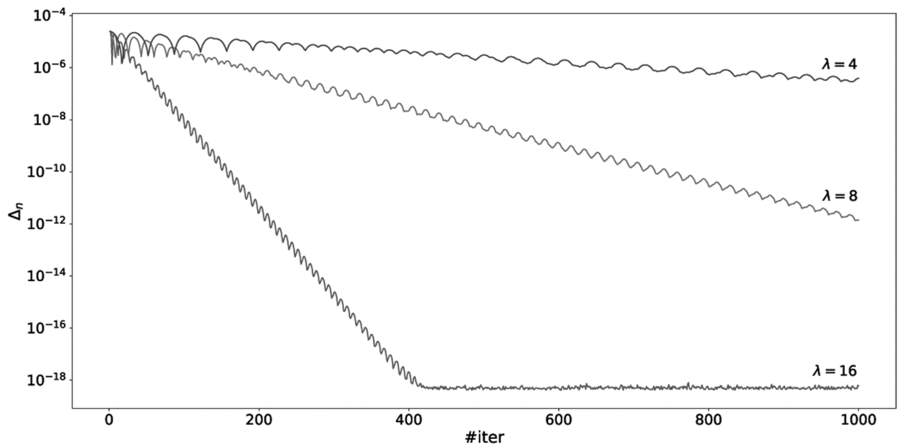

FIG. 6. Δ_n and # iter for $N = 2000$, Entropy–Euclidean, elapsed time 103.510 sec

In one of the future work we plan to consider a randomized version of Algorithm 1 and carry out the corresponding convergence analysis. It will help to have a progress in using this variant the mirror descent method for solving variational inequalities of huge size. Randomized versions of the mirror descent method, based on the extra-gradient algorithm are studied in [30, 34].

Also it is interesting to obtain similar results for the equilibrium programming problems [39–42].

In conclusion we note that, in our opinion, the proposed Algorithm is promising for the further investigation and can be used in practical applications.

This research is supported by the Ministry of Education and Science of Ukraine (project 0116U004777) and grant of the State Fund For Fundamental Research (President's of Ukraine grant, project F74/24921).

References

1. Konnov I. V. Combined relaxation methods for variational inequalities. — New York: Springer-Verlag, 2001.
2. Nesterov Yu. Dual extrapolation and its applications to solving variational inequalities and related problems // Mathematical Programming. — 2007. — Vol. 109. — P. 319–344.
3. Nurminski E. A. The use of additional diminishing disturbances in Fejer models of iterative algorithms // Comput. Math. Math. Phys. — 2008. — Vol. 48. — P. 2154–2161.
4. Stetsyuk P. I. An Approximate Method of Ellipsoids // Cybernetics and Systems Analysis. — 2003. — Vol. 39. — P. 435–439.
5. Semenov V. V. On the Parallel Proximal Decomposition Method for Solving the Problems of Convex Optimization // Journal of Automation and Information Sciences. — 2010. — Vol. 42, № 4. — P. 13–18.
6. Semenov V. V. On the convergence of methods for solving two-level variational inequalities with monotone operators // Journal of Computational & Applied Mathematics. — 2010. — № 1 (110). — P. 121–129.
7. Semenov V. V. Two methods for approximating a fixed point of a Fejer operator // Journal of Computational & Applied Mathematics. — 2013. — № 1 (111). — P. 46–56.
8. Chabak L. M., Dudar V. V., Semenov V. V., Vedel Ya. I. Outer approximations method for variational inequalities // Journal of Computational & Applied Mathematics. — 2015. — № 3 (120). — P. 77–84.
9. Denisov S. V., Semenov V. V. Proximal algorithm for two-level variational inequalities: strong convergence // Journal of Computational & Applied Mathematics. — 2011. — № 3 (106). — P. 27–32.
10. Nadezhkina N., Takahashi W. Strong convergence theorem by a hybrid method for nonexpansive mappings and Lipschitz-continuous monotone mappings // SIAM Journal on Optimization. — 2006. — Vol. 16. — P. 1230–1241.
11. Semenov V. V. An explicit splitting algorithm for variational inequalities with monotone operators // Journal of Computational & Applied Mathematics. — 2013. — № 2 (112). — P. 42–52.
12. Lyashko N. I., Semenov V. V., Chabak L. M. Splitting algorithm for variational inequalities with maximal monotone operators // Journal of Computational & Applied Mathematics. — 2014. — № 3 (117). — P. 131–139.
13. Korpelevich G. M. The extragradient method for finding saddle points and other problems // Matecon. — 1976. — Vol. 12, № 4. — P. 747–756.
14. Khobotov E. N. Modification of the extra-gradient method for solving variational inequalities and certain optimization problems // USSR Comput. Math. Math. Phys. — 1987. — Vol. 27, № 5. — P. 120–127.
15. Tseng P. A Modified Forward-Backward Splitting Method for Maximal Monotone Mappings // SIAM J. Control Optim. — 2000. — Vol. 38. — P. 431–446.
16. Censor Y., Gibali A., Reich S. The Subgradient Extragradient Method for Solving Variational Inequalities in Hilbert Space // Journal of Optimization Theory and Applications. — 2011. — Vol. 148. — P. 318–335.
17. Lyashko S. I., Semenov V. V., Voitova T.A. Low-cost modification of Korpelevich's methods for monotone equilibrium problems // Cybernetics and Systems Analysis. — 2011. — Vol. 47. — P. 631–639.

18. Voitova T. A., Denisov S. V., Semenov V. V. Strongly convergent modification of Korpelevich's method for equilibrium programming problems // Journal of Computational & Applied Mathematics. — 2011. — № 1 (104). — P. 10–23.
19. Apostol R. Ya., Grynenko A. A., Semenov V. V. Iterative algorithms for monotone two-level variational inequalities // Journal of Computational & Applied Mathematics. — 2012. — № 1 (107). — P. 3–14.
20. Semenov V. V. Strongly Convergent Algorithms for Variational Inequality Problem Over the Set of Solutions the Equilibrium Problems // In: Zgurovsky, M.Z., Sadovnichiy, V.A. (eds.) Continuous and Distributed Systems. SMIA, Vol. 211, Springer International Publishing Switzerland, 2014. — P. 131–146.
21. Semenov V. V. Hybrid Splitting Methods for the System of Operator Inclusions with Monotone Operators // Cybernetics and Systems Analysis. — 2014. — Vol. 50. — P. 741–749.
22. Semenov V. V. A Strongly Convergent Splitting Method for Systems of Operator Inclusions with Monotone Operators // Journal of Automation and Information Sciences. — 2014. — Vol. 46, № 5. — P. 45–56.
23. Denisov S. V., Semenov V. V., Chabak L.M. Convergence of the Modified Extragradient Method for Variational Inequalities with Non-Lipschitz Operators // Cybernetics and Systems Analysis. — 2015. — Vol. 51. — P. 757–765.
24. Verlan D. A., Semenov V. V., Chabak L. M. A Strongly Convergent Modified Extragradient Method for Variational Inequalities with Non-Lipschitz Operators // Journal of Automation and Information Sciences. — 2015. — Vol. 47, № 7. — P. 31–46.
25. Popov L. D. A modification of the Arrow-Hurwicz method for search of saddle points // Mathematical notes of the Academy of Sciences of the USSR. — 1980. — Vol. 28, № 5. — P. 845–848.
26. Malitsky Yu. V., Semenov V. V. An Extragradient Algorithm for Monotone Variational Inequalities // Cybernetics and Systems Analysis. — 2014. — Vol. 50. — P. 271–277.
27. Malitsky Yu. V., Semenov V. V. A hybrid method without extrapolation step for solving variational inequality problems // Journal of Global Optimization. — 2015. — Vol. 61. — P. 193–202.
28. Bregman L. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming // USSR Comput. Math. Math. Phys. — 1967. — Vol. 7. — P. 200–217.
29. Nemirovski A., Yudin D. Problem complexity and method efficiency in optimization. — New York: Wiley. — 1983.
30. Baes M., Burgisser M., Nemirovski A. A randomized mirror-prox method for solving structured large-scale matrix saddle-point problems // SIAM Journal on Optimization. — 2013. — Vol. 23. — P. 934–962.
31. Ben-Tal A., Margalit T., Nemirovski A. The ordered subsets mirror descent optimization method with applications to tomography // SIAM Journal on Optimization. — 2001. — Vol. 12. — P. 79–108.
32. Beck A., Teboulle M. Mirror Descent and Nonlinear Projected Subgradient Methods for Convex Optimization // Operations Research Letters. — 2003. — Vol. 31. — P. 167–175.
33. Auslender A., Teboulle M. Interior projection-like methods for monotone variational inequalities // Mathematical Programming. — 2005. — Vol. 104. — P. 39–68.

34. Juditsky A., Nemirovski A., Tauvel C. Solving variational inequalities with Stochastic Mirror-Prox algorithm. Stochastic Systems. — 2011. — Vol. 1, № 1. — P. 17–58.
35. Nemirovski A. 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 Journal on Optimization. — 2004. — Vol. 15. — P. 229–251.
36. Nazin A. V. Estimating the principal eigenvector of a stochastic matrix: Mirror Descent Algorithms via game approach with application to PageRank problem // In: 49th IEEE Conference on Decision and Control. — 2010. — P. 792–297.
37. Tremba A., Nazin A. Extension of a saddle point mirror descent algorithm with application to robust PageRank // In: 52nd IEEE Conference on Decision and Control. — 2013. — P. 3691–3696.
38. Duchi J., Shalev-Shwartz S., Singer Y., Chandra T. Efficient projections onto the ℓ_1-ball for learning in high dimensions // In: Proceedings of the 25th international conference on Machine learning. — 2008. — P. 272–279.
39. Denisov S. V. Parallel decomposition scheme for finding a saddle point, and the Nash equilibrium // Journal of Computational & Applied Mathematics. — 2010. — № 3 (102). — P. 40–48.
40. Vedel Ya. I., Semenov V. V. A new two-phase proximal method of solving the problem of equilibrium programming // Journal of Computational & Applied Mathematics. — 2015. — № 1 (118). — P. 15–23.
41. Vartuzova M. V., Semenov V. V., Chabak L. M. A new algorithm with Bregman distance for solving equilibrium problem // Journal of Computational & Applied Mathematics. — 2016. — № 3 (123). — P. 9–21.
42. Lyashko S. I., Semenov V. V. A New Two-Step Proximal Algorithm of Solving the Problem of Equilibrium Programming // In: Boris Goldengorin (ed.) Optimization and Its Applications in Control and Data Sciences. SOIA, Vol. 115, Springer International Publishing Switzerland, 2016. — P. 315–325.

Received May 4, 2017