Some Adaptive First-order Methods for Variational Inequalities with Relatively Strongly Monotone Operators and Generalized Smoothness[*]

Seydamet S. Ablaev[1,2][0000−0002−9927−6503], Alexander A. Titov[2,3][0000−0001−9672−0616], Fedor S. Stonyakin[1,2][0000−0002−9250−4438], Mohammad S. Alkousa[2,3][0000−0001−5470−0182], and Alexander V. Gasnikov[2,3,4][0000−0002−5982−8983]

[1] V. I. Vernadsky Crimean Federal University, Simferopol, Russia
[2] Moscow Institute of Physics and Technology, Moscow, Russia
[3] HSE University, Moscow, Russia
[4] Caucasus Mathematical Center, Adyghe State University, Maikop, Russia
`seydamet.ablaev@yandex.ru, a.a.titov@phystech.edu,`
`mohammad.alkousa@phystech.edu, fedyor@mail.ru, gasnikov@yandex.ru`

Abstract. In this paper, we introduce some adaptive methods for solving variational inequalities with relatively strongly monotone operators. Firstly, we focus on the modification of the recently proposed, in smooth case [18], adaptive numerical method for generalized smooth (with Hölder condition) saddle point problem, which has convergence rate estimates similar to accelerated methods. We provide the motivation for such an approach and obtain theoretical results of the proposed method. Our second focus is the adaptation of widespread recently proposed methods for solving variational inequalities with relatively strongly monotone operators. The key idea in our approach is the refusal of the well-known restart technique, which in some cases causes difficulties in implementing such algorithms for applied problems. Nevertheless, our algorithms show a comparable rate of convergence with respect to algorithms based on the above-mentioned restart technique. Also, we present some numerical experiments, which demonstrate the effectiveness of the proposed methods.

Keywords: Saddle point problem. Hölder continuity. Variational inequality. Restart technique. Strongly convex programming problem.

Introduction

Non-smooth convex optimization plays a key role in solving the vast majority of modern applied problems [6,8,10,30]. In this paper, we focus on non-smooth saddle point problems and variational inequalities, which, as can be easily shown,

---

are closely related to each other [20,31]. Such settings of the optimization problems naturally arise when considering problems in machine learning [12,17], data science [20], economic systems [5], optimal transport [11], network equilibrium [14], game theory [26], general equilibrium theory [9,16], etc.

Remind the problem of solving Minty variational inequality. For a given operator $g : X \to \mathbb{R}^n$, where X is a closed convex subset of some finite-dimensional vector space, we need to find a vector $x_* \in X$, such that

$$\langle g(x), x_* - x \rangle \leqslant 0, \quad \forall x \in X. \tag{1}$$

The operator g is called L-smooth, if for any $x, y, z \in X$ the following inequality holds

$$\langle g(y) - g(z), x - z \rangle \leqslant LV(x, z) + LV(z, y), \tag{2}$$

where $V(\cdot, \cdot)$ is the distance in some generalized sense, namely, Bregman divergence (see (9), below). We also assume, that the operator g is μ–relatively strongly monotone, i.e.

$$\mu V(y, x) + \mu V(x, y) \leqslant \langle g(y) - g(x), y - x \rangle, \quad \forall x, y \in X. \tag{3}$$

The concept of relative strong monotonicity is a natural generalization of the concept of relative strong convexity of the objective functional [23] in optimization problems, for variational inequalities.

We are motivated by the following saddle point problem

$$\min_x \max_y f(x, y). \tag{4}$$

Using the recently proposed new paradigm of convex optimization [3,28], namely, relative smoothness condition, there was proposed a technique [18], which provides the possibility of the acceleration of numerical methods for solving saddle point problems, assuming, that the gradient of the objective function $\nabla f(x, y)$ satisfies Lipschitz condition. Moreover, the proposed method is adaptive with respect to all Lipschitz constants of the objective's gradient.

In this paper, we extend the considered class of saddle point problems and replace the classical Lipschitz continuity condition

$$\|\nabla f(z) - \nabla f(u)\| \leqslant L\|z - u\|, \tag{5}$$

by the following Hölder continuity condition

$$\|\nabla f(z) - \nabla f(u)\| \leqslant L_\nu \|z - u\|^\nu, \tag{6}$$

$z = (x_z, y_z), u = (x_u, y_u)$, with respect to the gradient of the objective, where $0 \leqslant \nu \leqslant 1$. Hereinafter we consider arbitrary non-Euclidean norms, which are defined on the corresponding spaces unless otherwise stated.

Note, that the concept of Hölder continuity is an extremely important generalization of the Lipschitz continuity condition. A huge number of applied problems can be formulated exclusively on the class of minimization of Hölder continuous functionals, e.g. smooth multi-armed bandit problem [21], detecting heart

rate variability [27], etc. Also, if some function is uniformly convex, then its conjugated will necessarily have the Hölder-continuous gradient, according to [29].

Based on the recently proposed restart technique of the Universal Proximal Method for solving variational inequalities [15], we propose algorithms, which ensure the ε–approximate solution of the considered problem (1) after no more than

$$N = \left\lceil \frac{2L\Omega}{\mu} \log_2 \frac{R_0^2}{\varepsilon} \right\rceil \tag{7}$$

iterations, where μ denotes the constant of relative strong monotonicity of g, Ω and R_0 are specified in Algorithm 2 and can be understood as some characteristics of the domain of the operator g.

The paper consists of an introduction and four main sections. In Sect. 1, we discuss approach [7,18] to the accelerated rates of first-order methods for strongly convex-concave saddle point problems basing on relative smoothness and strong monotonicity conditions. Moreover, we consider generalizations of smoothness conditions for saddle point problems. In Sect. 2, we propose adaptive version of restarted Mirror Prox method [32] for generalized smooth problems. Further, in Sect. 3, we propose some adaptive methods, which do not imply the restart technique, but have the similar convergence rate estimates of the proposed algorithm with restart technique. In Sect. 4, we present some numerical experiments for the saddle point problem and Minty variational inequality, which demonstrate the effectiveness of the proposed methods.

The contributions of the paper can be formulated as follows.

- We consider the non-smooth strongly convex-concave saddle point problem and propose a restarted version of the Universal Proximal Method for the corresponding variational inequality, which guarantees the ε–approximate solution of the problem (1) in an optimal rate of $N = \left\lceil \frac{2L\Omega}{\mu} \log_2 \frac{R_0^2}{\varepsilon} \right\rceil$ iterations. These algorithms are of interest in case of a huge value of the condition number $\frac{L}{\mu}$ as well as in the case of considering not strongly convex-concave saddle point problems.
- We propose methods beyond the restart technique and show, that in some cases they may have even better estimates of the convergence rate compared to methods, based on the restart technique. Moreover, the required number of iterations of such algorithms does not exceed $N = \left\lceil \frac{L+\mu}{\mu} \log_2 \frac{R_0^2}{\varepsilon} \right\rceil$.
- We present some numerical experiments, which demonstrate the effectiveness of the proposed methods.

We start with some auxiliaries. Let E be a finite-dimensional vector space and E^* be its dual. Let us choose some norm $\|\cdot\|$ on E. Define the dual norm $\|\cdot\|_*$ as follows

$$\|\phi\|_* = \max_{\|x\| \leqslant 1} \{\langle \phi, x \rangle\}, \tag{8}$$

where $\langle \phi, x \rangle$ denotes the value of the linear function $\phi \in E^*$ at the point $x \in E$.

Let us choose some prox-function $d(x)$, which is continuously differentiable and convex on E, and define the corresponding Bregman divergence as follows

$$V(y,x) = V_d(y,x) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle, \quad \forall x, y \in E. \qquad (9)$$

The Bregman divergence can be understood as some generalization of the distance in the considered set.

1   Towards Adaptive Accelerated Rates for Saddle Point Problems with Generalized Smoothness Condition

Let $Q_x \subset \mathbb{R}^n$ and $Q_y \subset \mathbb{R}^m$ be nonempty, convex and compact sets. Consider the following saddle point problem

$$\min_{x \in Q_x} \max_{y \in Q_y} \{ f(x,y) + h(x) - g(x) \}, \qquad (10)$$

where $f(x,y) : Q_x \times Q_y \to \mathbb{R}$ is convex function for fixed $y \in Q_y$ and concave for fixed $x \in Q_x$, functions $h(x)$ and $g(y)$ are convex on Q_x and Q_y, respectively, and for each $x, x' \in Q_x, y, y' \in Q_y$, we have

$$\|\nabla h(x) - \nabla h(x')\|_* \leqslant L_x \|x - x'\|, \quad \|\nabla g(y) - \nabla g(y')\|_* \leqslant L_y \|y - y'\|,$$

for some $L_x > 0, L_y > 0$.

Remark 1. If $f(x,y) : Q_x \times Q_y \to \mathbb{R}$ is strongly convex function for fixed $y \in Q_y$ and strongly concave for fixed $x \in Q_x$, we can consider the problem (10) with $h(x) = \frac{\|x\|^2}{2}$ and $g(y) = \frac{\|y\|^2}{2}$.

Let us consider the following setting of the problem (10). Suppose, for any $x, x' \in Q_x, \ y, y' \in Q_y, \ L_{xx} > 0, L_{yy} > 0, L_{xy} > 0$, and for some $\nu \in [0,1]$, the following inequalities hold

$$\|\nabla_x f(x,y) - \nabla_x f(x',y')\|_* \leqslant L_{xx} \|x - x'\|^\nu + L_{xy} \|y - y'\|^\nu, \qquad (11)$$

$$\|\nabla_y f(x,y) - \nabla_y f(x',y')\|_* \leqslant L_{xy} \|x - x'\|^\nu + L_{yy} \|y - y'\|^\nu. \qquad (12)$$

Let ω_x, ω_y be some 1-strongly convex function w.r.t. $\|\cdot\|_{Q_x}, \|\cdot\|_{Q_y}$ respectively, $\alpha : Q_x^* \to \mathbb{R}$ be the convex conjugate of $\frac{\|x\|^2}{2} - \mu_x \omega_x$, and $\beta : Q_y^* \to \mathbb{R}$ be the convex conjugate of $\frac{\|y\|^2}{2} - \mu_y \omega_y$. Then we can consider the following saddle point problem

$$\min_{\substack{x \in Q_x, \ y \in Q_y, \\ b \in Q_y^* \ a \in Q_x^*}} \max \left\{ \langle a, x \rangle + \langle b, y \rangle + \mu_x \omega_x(x) - \mu_y \omega_y(y) + f(x,y) - \alpha(a) + \beta(b) \right\}. \qquad (13)$$

It is shown [18], that if (x^*, y^*, a^*, b^*) is the saddle point to (13), then (x^*, y^*) is the saddle point to (10). Also, α is $\frac{1}{L_x}$-strongly convex in $\|\cdot\|_{Q_x,*}$, and β is $\frac{1}{L_y}$-strongly convex in $\|\cdot\|_{Q_y,*}$ [18].

Lemma 1. *Define the following operator $((x, y) \in Q_x \times Q_y := X)$*

$$g(x, y, a, b) = \Big(a + \mu_x \nabla \omega_x(x) + \nabla_x f(x, y), -b + \mu_y \nabla \omega_y(y) - \nabla_y f(x, y),$$
$$-x + \nabla \alpha(a), y + \nabla \beta(b)\Big),$$

where f satisfies (11)-(12). Consider the following prox-function

$$d(x, y, a, b) = \mu_x \omega_x(x) + \mu_y \omega_y(y) + \alpha(a) + \beta(b),$$

and the corresponding Bregman divergence, defined according to (9). Then g is 1-relatively strongly monotone, i.e.

$$\langle g(y) - g(x), y - x \rangle \geqslant V(y, x) + V(x, y), \tag{14}$$

and generalized relatively smooth operator, i.e.

$$\langle g(y) - g(z), y - x \rangle \leqslant LV(y, z) + LV(x, y) + \delta, \tag{15}$$

for some $\delta > 0$, with

$$L = \widetilde{L}(\delta) = \left(\frac{2}{\delta}\right)^{\frac{1-\nu}{1+\nu}} \left(\frac{L_{xx}^{\frac{2}{1+\nu}}}{\mu_x} + \frac{L_{xy}^{\frac{2}{1+\nu}}}{\sqrt{\mu_x \mu_y}} + \frac{L_{yy}^{\frac{2}{1+\nu}}}{\mu_y}\right). \tag{16}$$

Proof. The proof is given in arXiv preprint [1].

Hence, considered variational inequalities with relatively strongly monotone and generalized relatively smooth operators allow one to obtain first-order method complexity estimates for the corresponding class of strongly convex-convex saddle point problems, which are similar to the accelerated methods [18]. Moreover, using the artificial inaccuracy, Lemma 1 extends this approach to saddle point problems with generalized smoothness conditions [31,32].

However, extending the class of problems, one can potentially encounter the problem of a large value of $\widetilde{L}(\delta)$. On the other hand, even while considering the smooth case for saddle point problems, it may be difficult to estimate all the 5 parameters μ_x, μ_y, L_{xx}, L_{xy} and L_{yy}. Motivated by this and starting from the methodology of Y. E. Nesterov's works [13,24,29], we propose methods allowing the adaptively selection of the corresponding values of these parameters.

2 Adaptive Restarted Mirror Prox for Variational Inequalities with Relative Strongly Monotone Operators

Recently [32], there was proposed an adaptive universal algorithm (listed as Algorithm 1, below), which can automatically adjust to the smoothness level of the operator g.

Algorithm 1 Universal Mirror Prox for Variational Inequalities [32].

Require: $\varepsilon > 0$, $\delta > 0$, $x_0 \in X$, initial guess $L_0 > 0$, prox-setup: $d(x)$, $V(x, z)$.
1: Set $k = 0$, $z_0 = \arg\min_{u \in X} d(u)$.
2: **repeat**
3:    Find the smaller $i_k \geqslant 0$, such that

$$\langle g(z_k), z_{k+1} - z_k \rangle \leqslant \langle g(w_k), z_{k+1} - w_k \rangle + \langle g(z_k), w_k - z_k \rangle +$$
$$+ L_{k+1}(V(w_k, z_k) + V(z_{k+1}, w_k)) + \delta,$$

   where $L_{k+1} = 2^{i_k - 1} L_k$, and

$$w_k = \arg\min_{x \in X}\{\langle g(z_k), x - z_k \rangle + L_{k+1} V(x, z_k)\},$$

$$z_{k+1} = \arg\min_{x \in X}\{\langle g(w_k), x - w_k \rangle + L_{k+1} V(x, z_k)\}.$$

4: **until** $S_N := \sum_{k=0}^{N-1} \frac{1}{L_{k+1}} \geqslant \frac{\max_{x \in X} V(x, x_0)}{\varepsilon}$.
Ensure: z_N.

Theorem 1 ([32]). *Let g be a monotone operator, z_N be the output of Algorithm 1 after N iterations. Then the following inequality holds*

$$\langle g(x_*), z_N - x_* \rangle \leqslant -\frac{1}{S_N} \sum_{k=0}^{N-1} \frac{\langle g(w_k), x_* - w_k \rangle}{L_{k+1}} \leqslant \frac{2LV(x_*, z_0)}{N}. \qquad (17)$$

Moreover, the total number of iterations does not exceed

$$N = \left\lceil \frac{2L}{\varepsilon} \cdot \max_{x \in X} V(x_0, x) \right\rceil. \qquad (18)$$

Lemma 2. *Let g be a relatively strongly monotone operator. For Algorithm 1, the following δ-decreasing of Bregman divergence takes place*

$$V(x_*, z_N) \leqslant V(x_*, z_0) + \delta S_N. \qquad (19)$$

Proof. The proof is given in arXiv preprint [1].

The following Algorithm 2 provides the possibility of the acceleration of the proposed Algorithm 1 for solving variational inequality with relatively strongly monotone operator.

Theorem 2. *Let g be a generalized relatively smooth (15) and μ-relatively strongly monotone operator. Then for the output point x_p of the Algorithm 2, it will be hold: $V(x_*, x_p) \leqslant \varepsilon + \frac{2\Omega L\delta}{\mu^2}$. Moreover, the total number of iterations of Algorithm 2 does not exceed*

$$N = \left\lceil \frac{2L\Omega}{\mu} \cdot \log_2 \frac{R_0^2}{\varepsilon} \right\rceil. \qquad (20)$$

Algorithm 2 Restarted version of Algorithm 1.

Require: $\varepsilon > 0$, $\mu > 0$, $\Omega : d(x) \leqslant \frac{\Omega}{2}$ $\forall x \in X : \|x\| \leqslant 1$; x_0, R_0 $: V(x_*, x_0) \leqslant R_0^2$.

1: $p = 0, d_0(x) = R_0^2 d\left(\frac{x - x_0}{R_0}\right)$.

2: **repeat**

3:   x_{p+1} — output of Algorithm 1 with prox function $d_p(\cdot)$ and stopping criterion $S_N := \sum_{i=0}^{N-1} L_{i+1}^{-1} \geqslant \frac{\Omega}{\mu}$.

4:   $R_{p+1}^2 = \frac{\Omega R_0^2}{2^{(p+1)} \mu S_{N_p}}$.

5:   $d_{p+1}(x) \leftarrow R_{p+1}^2 d\left(\frac{x - x_{p+1}}{R_{p+1}}\right)$.

6:   $p = p + 1$.

7: **until** $p > \log_2\left(\frac{2R_0^2}{\varepsilon}\right)$.

Ensure: x_p.

Proof. The proof is given in arXiv preprint [1].

Remark 2. As shown above, Algorithm 2 needs no more than $N = \left\lceil \frac{2L\Omega}{\mu} \cdot \log_2 \frac{R_0^2}{\varepsilon} \right\rceil$ iterations to provide a solution of the problem (10), L is defined according to (16), while the technique, described in [34] has the following worse complexity estimate

$$\inf_{\nu \in [0,1]} \left\lceil \left(\frac{L_\nu}{\mu}\right)^{\frac{2}{1+\nu}} \cdot \frac{2^{\frac{2}{1+\nu}} \Omega}{\varepsilon^{\frac{1-\nu}{1+\nu}}} \cdot \log_2 \frac{2R_0^2}{\varepsilon} \right\rceil, \tag{21}$$

where $L_\nu = \tilde{L} \left(\frac{\tilde{L}}{2\varepsilon} \frac{(1-\nu)(2-\nu)}{2-\nu}\right)^{\frac{(1-\nu)(1+\nu)}{2-\nu}}, \tilde{L} = \left(L_{xy}\left(\frac{2L_{xy}}{\mu_y}\right)^{\frac{\nu}{2-\nu}} + L_{xx}D^{\frac{\nu-\nu^2}{2-\nu}}\right)$, and D is the diameter of the domain of $f(x, \cdot)$.

Remark 3. The obtained estimate

$$N = \left\lceil 2\Omega \left(\frac{L_{xx}}{\mu_x} + \frac{L_{xy}}{\sqrt{\mu_x \mu_y}} + \frac{L_{yy}}{\mu_y}\right) \cdot \log_2 \frac{R_0^2}{\varepsilon} \right\rceil \tag{22}$$

is optimal for saddle point problems (13) with $\nu = 1$.

Remark 4. Note, that Ω may depend on the dimension of the considered space [4].

3   First-order methods for relatively strongly monotone variational inequalities beyond the restart technique

Basing on some recently proposed methods [7,18] for VIs with strongly relatively monotone operators $g : X \to \mathbb{R}^n$, we consider algorithms (see Algorithms 3, 4 and 5) without using the restart technique. Similarly to the previous section we consider the case of operators g with the generalized smoothness condition (15). We improve the quality of the solution, compared to Algorithm 3 by reducing

$O\left(\frac{\delta}{\mu^2}\right)$ to $O\left(\frac{\delta}{\mu}\right)$, which provides the better convergence rate in case of small value of μ. It is also worth noting, that proposed algorithms do not require knowledge of the parameter Ω.

Algorithm 3 Adaptive first-order method for variational inequalities with μ-relatively strongly monotone operators without restarts.

Require: $\varepsilon > 0, \delta > 0, x_0 \in X, L_0 > 0, \mu > 0, d(x), V(x, z)$.

1: Set $z_0 = \arg\min_{u \in X} d(u)$.

2: **for** $k \geqslant 0$ **do**

3:     Find the smallest integer $i_k \geqslant 0$, such that

$$\langle g(z_k) - g(w_k), z_{k+1} - w_k \rangle \leqslant L_{k+1}\left(V(w_k, z_k) + V(z_{k+1}, w_k)\right) + \delta, \qquad (23)$$

where $L_{k+1} = 2^{i_k - 1} L_k$, and

$$w_k = \arg\min_{y \in X}\left\{\left\langle \frac{1}{L_{k+1}} g(z_k), y \right\rangle + V(y, z_k)\right\}, \qquad (24)$$

$$z_{k+1} = \arg\min_{z \in X}\left\{\left\langle \frac{1}{L_{k+1}} g(w_k), z \right\rangle + V(z, z_k) + \frac{\mu}{L_{k+1}} V(z, w_k)\right\}. \qquad (25)$$

4: **end for**

Ensure: z_k.

Theorem 3. *Let g be a μ-relatively strongly monotone operator, and z_* be the exact solution of the variational inequality (1). Then for Algorithm 3, the following inequality holds*

$$V(z_*, z_{k+1}) \leqslant \prod_{i=1}^{k+1}\left(1 + \frac{\mu}{L_i}\right)^{-1} V(z_*, z_0) + \frac{\delta}{L_{k+1} + \mu} +$$
$$+ \sum_{j=1}^{k} \frac{\delta}{L_j + \mu} \prod_{i=j+1}^{k+1}\left(1 + \frac{\mu}{L_i}\right)^{-1}. \qquad (26)$$

Proof. The proof is given in arXiv preprint [1].

Corollary 1. *Let g be a μ-relatively strongly monotone operator, and z_* be the exact solution of the variational inequality (1). Then for Algorithm 4, the following inequalities hold*

$$V(z_*, z_{k+1}) \leqslant \prod_{i=0}^{k}\left(1 + \frac{\mu}{L_{i+1}}\right)^{-1} V(z_*, z_0), \qquad (30)$$

$$V(z_*, z_{k+1}) \leqslant \left(1 + \frac{\mu}{2L}\right)^{-(k+1)} V(z_*, z_0). \qquad (31)$$

Algorithm 4 Adaptive first-order method for variational inequalities with μ-relatively strongly monotone and L-smooth operators.

Require: $\varepsilon > 0, x_0 \in X, L_0 > 0, \mu > 0, d(x), V(x, z)$.
 1: Set $z_0 = \arg\min_{u \in X} d(u)$.
 2: **for** $k \geqslant 0$ **do**
 3:    Find smallest integer $i_k \geqslant 0$, such that

$$\langle g(z_k) - g(w_k), z_{k+1} - w_k \rangle \leqslant L_{k+1} \left(V(w_k, z_k) + V(z_{k+1}, w_k) \right), \qquad (27)$$

   where $L_{k+1} = 2^{i_k - 1} L_k$, and

$$w_k = \arg\min_{y \in X} \left\{ \left\langle \frac{1}{L_{k+1}} g(z_k), y \right\rangle + V(y, z_k) \right\}, \qquad (28)$$

$$z_{k+1} = \arg\min_{z \in X} \left\{ \left\langle \frac{1}{L_{k+1}} g(w_k), z \right\rangle + V(z, z_k) + \frac{\mu}{L_{k+1}} V(z, w_k) \right\}. \qquad (29)$$

 4: **end for**
Ensure: z_k.

Remark 5. Due to (31), the number of iterations of Algorithm 4 for solving the problem (1) does not exceed $N = \left\lceil \frac{2L+\mu}{\mu} \log_2 \frac{R_0^2}{\varepsilon} \right\rceil$, which coincides with (20) up to the multiplication by a constant in the case of $L \sim \mu \sim 1$.

Remark 6. Taking into account (16), we find that the inequality (31) has the following form

$$V(z_*, z_{k+1}) = \left(\frac{2\widetilde{L}}{2\widetilde{L} + \mu} \right)^{k+1} V(z_*, z_0),$$

where $\widetilde{L}$ is given in (16).

The main difference between Algorithms 3 and 4 and the next Algorithm 5 is a modified exit criterion, which leads to decreasing of the coefficient at δ.

Theorem 4. *Let g be a μ-relatively strongly monotone operator, and z_* be the exact solution of the variational inequality (1). Then for Algorithm 5, we have*

$$V(z_*, z_{k+1}) \leqslant \prod_{i=1}^{k+1} \left(1 + \frac{\mu}{L_i} \right)^{-1} V(z_*, z_0) + \delta \left(1 + \sum_{j=1}^{k} \prod_{i=j+1}^{k+1} \left(1 + \frac{\mu}{L_i} \right)^{-1} \right), \tag{35}$$

and

$$V(z_*, z_{k+1}) \leqslant \left(1 + \frac{\mu}{2L} \right)^{-(k+1)} V(z_*, z_0) + \delta \left(1 + \frac{2L}{\mu} \right). \tag{36}$$

Proof. The proof is given in arXiv preprint [1].

Algorithm 5 The third adaptive first-order method for variational inequalities with μ-relatively strongly monotone operators.

Require: $\varepsilon > 0, \delta > 0, x_0 \in X, L_0 > 0, \mu > 0, d(x), V(x, z)$.
1: Set $z_0 = \arg\min_{u \in X} d(u)$.
2: **for** $k \geqslant 0$ **do**
3:     Find the smallest integer $i_k \geqslant 0$, such that

$$\langle g(z_k) - g(w_k), z_{k+1} - w_k \rangle \leqslant L_{k+1}\left(V(w_k, z_k) + V(z_{k+1}, w_k)\right) + L_{k+1}\delta, \quad (32)$$

    where $L_{k+1} = 2^{i_k - 1}L_k$, and

$$w_k = \arg\min_{y \in X}\left\{ \left\langle \frac{1}{L_{k+1}}g(z_k), y \right\rangle + V(y, z_k) \right\}, \quad (33)$$

$$z_{k+1} = \arg\min_{z \in X}\left\{ \left\langle \frac{1}{L_{k+1}}g(w_k), z \right\rangle + V(z, z_k) + \frac{\mu}{L_{k+1}}V(z, w_k) \right\}. \quad (34)$$

4: **end for**
Ensure: z_k.

Remark 7. If we iterate the inequality

$$V(z_*, z_{k+1}) \leqslant \left(1 + \frac{\mu}{L_{k+1}}\right)^{-1} V(z_*, z_k) + \frac{L_{k+1}\delta}{L_{k+1} + \mu},$$

then for the Algorithm 5, the following inequality holds

$$V(z_*, z_{k+1}) \leqslant \prod_{i=1}^{k+1}\left(1 + \frac{\mu}{L_i}\right)^{-1} V(z_*, z_0) + \frac{L_{k+1}\delta}{L_{k+1} + \mu} + \sum_{j=1}^{k}\frac{L_j\delta}{L_j + \mu}\prod_{i=j+1}^{k+1}\left(1 + \frac{\mu}{L_i}\right)^{-1}.$$

Remark 8. For the selected type of saddle point problems (13) with (16) and $\mu = 1$, we find that the inequality

$$V(z_*, z_{k+1}) \leqslant \left(1 + \frac{\mu}{2L}\right)^{-(k+1)} V(z_*, z_0) + \delta\left(1 + \frac{2L}{\mu}\right),$$

has the following form

$$V(z_*, z_{k+1}) \leqslant \left(1 + \frac{1}{2\widetilde{L}}\right)^{-(k+1)} V(z_*, z_0) + \widetilde{L}2^{\frac{2}{1-\nu}}\delta^{\frac{2\nu(1-\nu)}{(1+\nu)^2}} + \delta,$$

where $\widetilde{L}$ is given in (16).

4 Numerical Experiments

4.1 Saddle point problem for the smallest covering ball problem with functional constraints

In this subsection, we consider an example of the Lagrange saddle point problem induced by a problem with geometrical nature, namely, an analogue of the well-known *smallest covering ball problem* with functional constraints. This example
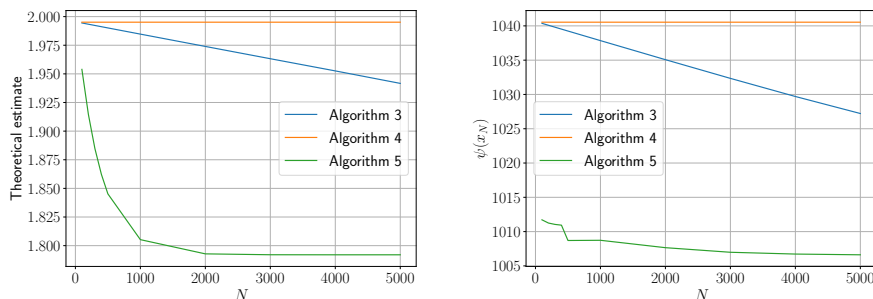
Fig. 1: The results of Algorithms 3, 4 and 5, for case 1.

is equivalent to the following non-smooth convex optimization problem with functional constraints

$$\min_{x \in X} \left\{ \psi(x) := \max_{1 \leqslant k \leqslant s} \|x - A_k\|_2^2; \ \varphi_p(x) \leqslant 0, \ p = 1, ..., m \right\}, \qquad (37)$$

where $A_k \in \mathbb{R}^n, k = 1, ..., s$ are given points and X is a convex compact set. Functional constraints φ_p, for $p = 1, ..., m$, have the following form

$$\varphi_p(x) := \sum_{i=1}^{n} \alpha_{pi} x_i^2 - 5, \quad p = 1, ..., m. \qquad (38)$$

See [1], for more details about the setting of this problem and the setting of its connected conducted experiments with it. The coefficients α_{pi} in (38) are drawn randomly from the following distributions.
**Case 1:** Pareto II or Lomax distribution with shape equalling 10.
Case 2: chi-square distribution, with a number of degrees of freedom, equals 3.

For case 1, the results of the work of Algorithms 3, 4, and 5 are presented in the Fig. 1, below. These results demonstrate the theoretical estimate (26), (30) and (35) for Algorithms 3, 4 and 5, respectively. Also, they demonstrate the value of the objective function in (37) at the output point x_N of the algorithm after performing N iterations.

For case 2, we compare the work of Algorithms 2, 3, 4, and 5. The results are presented in Fig. 2. They demonstrate the theoretical estimate (26), (30) and (35) for Algorithms 3, 4 and 5, respectively. Also, in the comparison with Algorithm 2, these results demonstrate the value of the objective function at the output point x_N of the algorithms after performing N iterations.

From Fig. 1, we can see that the proposed Algorithm 5 provides the best quality solution with respect to the value of the objective function ψ at the output point x_N, although the theoretical estimate of this quality is not very high. Therefore the efficiency of Algorithm 5 is obvious when we look at the value of the objective function at the output point of compared algorithms.
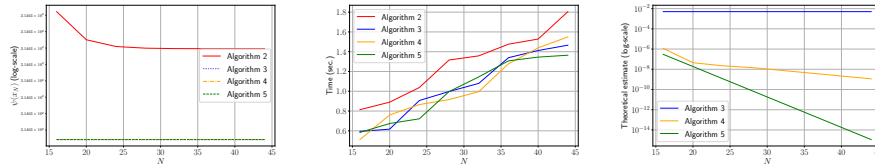
Fig. 2: The results of Algorithms 2, 3, 4 and 5, for case 2.

From Fig. 2, we can see that Algorithm 5 also gives the best estimate of the quality of the solution. Also, Algorithms 3, 4 and 5 give the same objective values at the output points, with approximately the same running time. In this case, we see that Algorithm 4 works better than Algorithm 3 (it gives better estimate of the quality of the solution), not as in case 1.

4.2 One Minty Variational Inequality

In this subsection we consider variational inequality with the following Lipschitz-continuous and strongly monotone operators (See [1], for more details about the setting of this problem and the setting of its connected conducted experiments with it.).

The first operator is (see Example 5.2 in [19])

$$g : X \subset \mathbb{R}^n \to \mathbb{R}^n, \quad g(x) = x. \tag{39}$$

The second operator is

$$g : X \subset \mathbb{R}^n \to \mathbb{R}^n, \quad g(x_1, \ldots, x_n) = \left(x_1, 2^2 x_2, 3^2 x_3, \ldots, n^2 x_n\right). \tag{40}$$

This operator is L-Lipschitz continuous with $L = n^2$ and μ-strongly monotone with $\mu = 1$. The condition number for this operator is $\kappa = L/\mu = n^2$, therefore this operator will be ill-conditioned when n is relatively big.
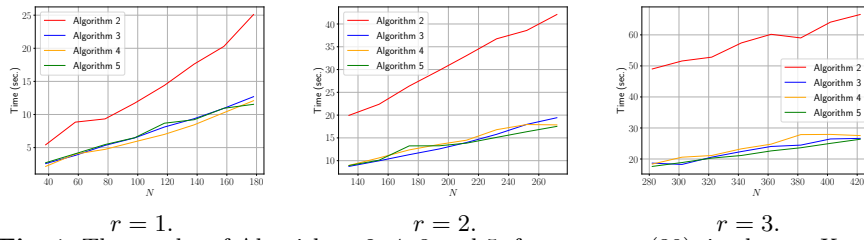
For the experiments with operator (39), the results are presented in Fig. 3 and 4, which illustrate the norm $\|x_{\text{out}} - x_*\|_2$, and the running time in seconds as a function of iterations, where x_{out} is the output of each algorithm.

In Fig. 3, we can not see the graphics of the Algorithms 3 and 5, that indicate the distance $\|x_{\text{out}} - x_*\|_2$, because by these algorithms this distance is equal to zero for all considered number of iterations.

From the conducted experiments (for operator (39)), we can see that the shape of the feasible set very much affects the progress of the work of the proposed algorithms. We note that when we increase r (the radius of the ball X), the corresponding running time of the compared algorithms is also increased. Also, from Fig. 3 and 4, we can see that the proposed Algorithms 3 and 5, for any value of the radius r, are the best, where they give the solution of the problem under consideration with very high quality and at the same (approximately) running time. Also, we can see that Algorithm 4 always works better than Algorithm 2.

$r = 1.$ $r = 2.$ $r = 3.$
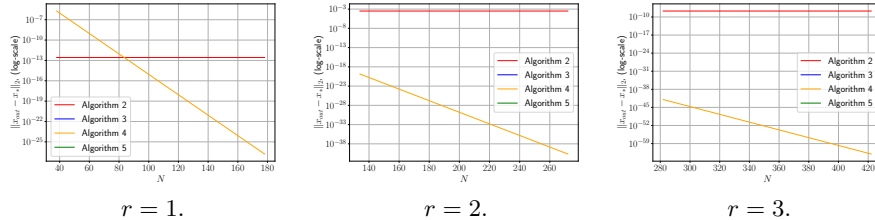
Fig. 3: The results of Algorithms 2, 4, 3 and 5, for operator (39), in the set $X = \{x \in \mathbb{R}^n, \|x\|_2 \leqslant r\}$ with different radii r and $n = 10^6, \varepsilon \in \{10^{-3i}, i = 1, 2, \ldots, 8\}, \delta = 0.01$.

$r = 1.$ $r = 2.$ $r = 3.$

Fig. 4: The results of Algorithms 2, 4, 3 and 5, for operator (39), in the set $X = \{x \in \mathbb{R}^n, \|x\|_2 \leqslant r\}$ with different radii r and $n = 10^6, \varepsilon \in \{10^{-3i}, i = 1, 2, \ldots, 8\}, \delta = 0.01$.

Now, for the experiments with operator (40), the results are presented in Fig. 5, which illustrates the norm $\|x_{\text{out}} - x_*\|_2$, the running time in seconds as a function of iterations, and the theoretical estimates (19), (26) and (35) of the quality of the solution, for Algorithms 1, 3 and 5, respectively.
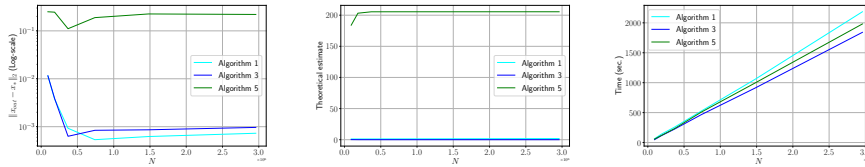
Fig. 5: The results of Algorithms 1, 3 and 5, for operator (40).

From Fig. 5, we can see that Algorithm 5 is the worst. Algorithm 1, with respect to the distance $\|x_{\text{out}} - x_*\|_2$, gives results better than Algorithms 3. Also, the theoretical estimate of the quality of the solution by Algorithm 1 is approximately the same as by Algorithm 3. The difference between the running times of Algorithms 1 and 3 is not big. Note that, Algorithm 1 is applicable to a wider type of problem and can work better than Algorithms 3, 4 and 5 with small μ. At the same time for variational inequalities with strongly monotone operators, it can be restarted in such a way that Algorithm 2 will have similar

convergence rate estimates. Thus, Algorithm 2 will be the best for the problems with an ill-conditioned operator.

Conclusion

In this paper, we study adaptive first-order methods for variational inequalities from the class of relatively strongly monotone operators recently introduced in [33]. Our research is motivated, in particular, by the recently proposed technique [7,18] for strongly convex-concave saddle point problems, which allows one to obtain the complexity estimates of the accelerated methods. First of all, the paper deals with the issue of adaptive tuning of the method to the global smoothness parameters of the saddle point problem. Moreover, an essential feature is the consideration of operators with a generalized condition of relative L-Lipschitz property and the corresponding generalizations of smoothness for the class of saddle point problems under consideration. Based on the methods from [7,18], we proposed algorithms for solving variational inequalities with relatively strongly monotone operators and obtained estimates of their convergence We also presented some numerical experiments, which demonstrate the effectiveness of the proposed methods. We considered an example of the convex optimization problem with functional constraints, and an example of the Minty variational inequality. The conducted experiments showed that the proposed Algorithms 3, 4 and 5 without using the technique of restarts work better than algorithm with restarts (Algorithm 2) and vice versa for an example of the variational inequality with an ill-conditioned operator.

References

1. Ablaev, S. S., Titov, A. A., Alkousa, M. S., Stonyakin, F. S., Gasnikov, A. V.: Some Adaptive First-order Methods for Variational Inequalities with Relatively Strongly Monotone Operators and Generalized Smoothness. arXiv preprint `https://arxiv.org/pdf/2207.09544.pdf` (2022)
2. Alkousa, M.S., Gasnikov, A.V., Dvinskikh, D.M., Kovalev, D.A., Stonyakin, F.S.: Accelerated methods for saddle-point problems. Comput. Math. and Math. Phys., **60**(11), 1843–1866 (2020)
3. Bauschke, H. H., Bolte, J., Teboulle, M.: A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. Mathematics of Operations Research, **42**(2), 330–348 (2017)
4. Ben-Tal, A., Nemirovski, A.: Lectures on modern convex optimization: analysis, algorithms, and engineering applications. Society for industrial and applied mathematics. (2001)
5. Buiter, W. H.: Saddle point Problems in Continuous Time Rational Expectations Models: A General Method and Some Macroeconomic Examples. NBER Technical Working Paper No. 20 (1984)
6. Bubeck, S.: Convex optimization: Algorithms and complexity. Foundations and Trends® in Machine Learning, **8**(3-4), 231–357 (2015)

7. Cohen, M. B., Sidford, A., Tian, K.: Relative Lipschitzness in Extragradient Methods and a Direct Recipe for Acceleration. arXiv preprint `https://arxiv.org/pdf/2011.06572.pdf` (2020)

8. Cheng, L., Hou, Z. G., Lin, Y., Tan, M., Zhang, W. C., Wu, F. X.: Recurrent neural network for non-smooth convex optimization problems with application to the identification of genetic regulatory networks. IEEE Transactions on Neural Networks, **22**(5), 714–726 (2011)

9. Cherukuri, A., Gharesifard, B., Cortes, J.: Saddle-point dynamics: conditions for asymptotic stability of saddle points. SIAM Journal on Control and Optimization, **55**(1), 486–511 (2017)

10. Clarke, F. H.: Method of Dynamic and Nonsmooth Optimization. Society for Industrial and Applied Mathematics. (1989)

11. Dafermos, S.: Traffic equilibrium and variational inequalities. Transportation science, **14**(1), 42–54 (1980)

12. Dauphin, Y. N., Pascanu, R., Gulcehre, C., Cho, K., Ganguli, S., Bengio, Y.: Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. Advances in neural information processing systems, 27. (2014)

13. Devolder, O., Glineur, F., Nesterov, Y.: First-order methods of smooth convex optimization with inexact oracle. Mathematical Programming, **146**(1), 37–75 (2014)

14. Friesz, T. L., Bernstein, D., Smith, T. E., Tobin, R. L., Wie, B. W. A variational inequality formulation of the dynamic network user equilibrium problem. Operations research, **41**(1), 179–191 (1993)

15. Gasnikov, A. V., Dvurechensky, P. E., Stonyakin, F. S., Titov, A. A.: An adaptive proximal method for variational inequalities. Computational Mathematics and Mathematical Physics, **59**(5), 836–841 (2019)

16. Grandmont, J. M.: Temporary general equilibrium theory. Econometrica: Journal of the Econometric Society, 535–572 (1977)

17. Jin, C., Netrapalli, P., Ge, R., Kakade, S. M., Jordan, M. I. On nonconvex optimization for machine learning: Gradients, stochasticity, and saddle points. Journal of the ACM (JACM), **68**(2), 1–29 (2021)

18. Jin, Y., Sidford, A., Tian, K.: Sharper Rates for Separable Minimax and Finite Sum Optimization via Primal-Dual Extragradient Methods. arXiv preprint `https://arxiv.org/pdf/2202.04640.pdf` (2022)

19. Khanh, P. D., Vuong, P. T.: Modified projection method for strongly pseudomonotone variational inequalities. Journal of Global Optimization **58**, 341–350 (2014)

20. Kinderlehrer, D., Stampacchia, G.: An introduction to variational inequalities and their applications. Society for Industrial and Applied Mathematics. (2000)

21. Liu, Y., Wang, Y., Singh, A.: Smooth Bandit Optimization: Generalization to Holder Space. In International Conference on Artificial Intelligence and Statistics, PMLR, 2206–2214 (2021)

22. Lu, H.: relative continuity for non-lipschitz nonsmooth convex optimization using stochastic (or deterministic) mirror descent. INFORMS Journal on Optimization, **1**(4), 288–303 (2019)

23. Lu, H., Freund, R. M., Nesterov, Y.: Relatively smooth convex optimization by first-order methods, and applications. SIAM Journal on Optimization, **28**(1), 333–354 (2018)

24. Nesterov, Y.: Gradient methods for minimizing composite functions. Mathematical programming, **140**(1), 125–161 (2013)

25. Lu, H., Freund, R. M., and Nesterov, Y.: Relatively smooth convex optimization by first-order methods, and applications. SIAM Journal on Optimization, **28**(1), 333-354 (2018)

26. Mertikopoulos, P., Lecouat, B., Zenati, H., Foo, C. S., Chandrasekhar, V., Piliouras, G.: Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. arXiv preprint `https://arxiv.org/pdf/1807.02629.pdf` (2018)

27. Nakamura, T., Horio, H., Chiba, Y.: Local holder exponent analysis of heart rate variability in preterm infants. IEEE Transactions on biomedical engineering, **53**(1), 83–88 (2005)

28. Nesterov, Y.: Relative smoothness: new paradigm in convex optimization. In Conference report, EUSIPCO-2019, A Coruna, Spain (Vol. 4) (2019, September)

29. Nesterov, Y.: Universal gradient methods for convex optimization problems. Mathematical Programming, **152**(1), 381–404 (2015)

30. Nesterov, Y.: Smooth minimization of non-smooth functions. Mathematical programming, **103**(1), 127–152 (2005)

31. Stonyakin, F., Gasnikov, A., Dvurechensky, P., Titov, A., Alkousa, M.: Generalized Mirror Prox Algorithm for Monotone Variational Inequalities: Universality and Inexact Oracle. Journal of Optimization Theory and Applications, 1–26 (2022)

32. Stonyakin, F., Tyurin, A., Gasnikov, A., Dvurechensky, P., Agafonov, A., Dvinskikh, D., Alkousa, M., Pasechnyuk, D., Artamonov, S., Piskunova, V.: Inexact Relative Smoothness and Strong Convexity for Optimization and Variational Inequalities by Inexact Model. Optimization Methods and Software, **36**(6), 1155–1201 (2021)

33. Stonyakin, F.S., Titov, A.A., Makarenko, D.V., Alkousa, M.S.: Some Methods for Relatively Strongly Monotone Variational Inequalities. arXiv preprint `https://arxiv.org/pdf/2109.03314.pdf` (2022)

34. Titov, A., Stonyakin, F., Alkousa, M., Gasnikov, A.: Algorithms for solving variational inequalities and saddle point problems with some generalizations of Lipschitz property for operators. In International Conference on Mathematical Optimization Theory and Operations Research. Springer, Cham, 86–101 (2021)

35. Titov, A.A., Stonyakin, F.S., Alkousa, M.S., Ablaev, S.S., Gasnikov, A.V.: Analogues of Switching Subgradient Schemes for Relatively Lipschitz-Continuous Convex Programming Problems. In: Kochetov Y., Bykadorov I., Gruzdeva T. (eds) Mathematical Optimization Theory and Operations Research. MOTOR 2020. Communications in Computer and Information Science, Springer, Cham. **1275**, 133–149 (2020)