FEDERAL STATE AUTONOMOUS EDUCATIONAL INSTITUTION OF HIGHER EDUCATION MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY (NATIONAL RESEARCH UNIVERSITY)

as a manuscript

Titov Alexander Alexandrovich

OPTIMIZATION METHODS FOR NON-SMOOTH PROBLEMS IN LARGE DIMENSIONAL SPACES

PhD Dissertation Summary
for the purpose of obtaining academic degree Doctor of Philosophy in Computer Science

Moscow — 2023

The PhD dissertation was prepared at Moscow Institute of Physics and Technology.

Supervisor: Doctor of Sciences **Gasnikov Alexander Vladimirovich**

Introduction

The necessity of developing numerical methods of non-smooth convex optimization in recent years is caused by significant progress in various fields of science, including biology, economics, chemistry, applied mathematics, theoretical physics, and many others. A particular difficulty, primarily in machine learning and data analysis problems, is the large size of the initial sample, the importance of obtaining as accurate solution as possible and minimizing the error, and the difficulty in calculating the value of a function or its derivatives that describe a particular mathematical model. The latter aspect is especially relevant due to the impossibility of performing an accurate numerical evaluation of the various characteristics of a function in many applied problems. Thus, many classical optimization algorithms turn out to be inapplicable, for example, in case when the objective function is non-smooth. It is worth noting that today the vast majority of applied problems generate optimization problems with non-smooth functions [3; 17; 26].

It is well known that both convex [32] and Lipschitz continuous [31] functions are differentiable in their domain almost everywhere. Nevertheless, optimization methods for the smooth case are not applicable to many applied problems with similar properties of the objective functional. It is not difficult to show [6] that for a function whose gradient satisfies the Lipschitz condition (hereafter — for a smooth function)

$$\|\nabla f(x) - \nabla f(y)\|_* \leq L\|x - y\|, \tag{1}$$

the following inequality is satisfied

$$f(y) - f(x) \leq \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|x - y\|^2 \quad \forall x{,}y \in dom f. \tag{2}$$

A natural attempt to generalize the inequality (2) is to replace the second summand of the right-hand side of (2), by distance in some generalized sense. The motivation for this approach is both the difficulty of using Euclidean distance in many applied problems and the deliberate replacement of $\|x - y\|^2$ by a distance more adapted to the particular formulation of the problem. Such an idea was proposed in [16] with replacing the norm of the difference of variables x, y by distance in a generalized sense, namely — by Bregman divergence. An important point of such a generalization is to preserve the optimal convergence rate of first-order methods [1; 15; 16].

Thus, one of the main results in this direction was the recently proposed concept of relative smoothness [16], which allowed the application of the gradient-type method to solve the problem of constructing an optimal ellipsoid covering a set of given points. This problem plays a crucial role in [1; 16] statistics and data analysis. Further, in [15], there was proposed a concept of relative Lipschitz continuity, which allowed us to take a new look at many applied problems, among which there is mentioned the Support Vector Machine for the binary classification problem and the problem of finding a common point of n of given ellipsoids.

The second problem that arises when considering the inequality (2) is the impossibility of calculating the exact gradient of the function $f(x)$. Recently, in [30; 33], it was shown that various modifications of the Mirror Descent Algorithm are applicable in the case of using the so-called δ–subgradient, it also was shown that there is no accumulation of error value in the final estimates of the convergence rate. However, in many applied problems, it is difficult not only to calculate the gradient with some error but also the value of the objective function itself. One of the most important results in this direction was the concept (δ,L)–oracle proposed in [4], which deteriorates the quality of the solution of the optimization problem by only $O(\delta)$.

The first part of the dissertation extends the class of problems, including an abstract model of a function replacing the first term $\langle \nabla f(x), y-x \rangle$ with some error, while admitting that the function itself is represented in abstract form using the so-called model, that allows one to work with composite optimization problems. It is worth noting that the optimal convergence rate estimates for the proposed methods are preserved.

The second focus of the dissertation is variational inequalities and saddle point problems with corresponding levels of smoothness of operators. It is worth noting, that variational inequalities play a key role in solving many applications in fluid dynamics [12], dynamical system design [5; 20] economics, particularly in modeling the network effect [19], finding general economic equilibrium [5; 8; 11], Nash equilibria [28], matrix games [22] etc.

One of the most notable numerical methods for solving variational inequalities was the extragradient method of G. M. Korpelevich, proposed in 1976 [13]. Recently, combining some ideas of [22] and [24], there was proposed a universal numerical algorithm [7] that is capable of making automatic adjustments to the smoothness level of the problem (namely, the ν parameter, see (3) below). In the thesis, this method is extended to the strong monotonicity condition of the operator under the assumptions of the smoothness classes introduced earlier. Also, there was considered a variant of the recently proposed accelerated method for solving saddle point problems in the non-smooth formulation. In particular, there is considered a generalization of the Lipschitz condition of the gradient of the objective function (1) to the following Holder condition for saddle point problems

$$\|\nabla f(z) - \nabla f(u)\|_* \leqslant L_\nu \|z - u\|^\nu, \tag{3}$$

which plays an important role in solving many applied problems, such as the multi-armed bandit problem [14], the heart rate variability problem [21], etc.

This thesis **aims** to develop optimal numerical methods for solving multidimensional non-smooth convex optimization problems with functional constraints. The well-known Mirror Descent Algorithm and its modifications are used as a base for developing the methods. To be able to apply the proposed methods in a broader class of functional The modification of the concept of inexact model of the objective functional and functional constraint. The applicability of the methods for functionals satisfying a relaxed version of the Lipschitz condition, namely the condition of relative Lipschitz continuity for classical optimization problems and the relative smoothness concept for variational inequalities. Also, the goal of the thesis includes the development of modifications of methods for solving variational inequalities and saddle point problems with corresponding smoothness classes. In particular, it is planned to propose for the first time a restart technique of Adaptive Proximal method for strongly monotone variational inequalities as well as an accelerated method for the non-smooth (Holder) saddle point problem.

The following **goals** were proposed to achieve the aim:

1. Develop an analogue of the Mirror Descent method with switchings to solve the problem of minimization of quasi-convex non-Lipschitz continuous functions with quasi-convex Lipschitz continuous inequality

constraints; substantiate the corresponding theoretical estimates of the convergence rate.

2. Extend the applicability of Mirror Descent Methods to a class of relatively Lipschitz problems, extending the proposed modifications of the Mirror Descent method to minimize functions admitting representation in abstract model generality, and investigating the theoretical characteristics of the proposed methods in the case of online and stochastic settings of relatively Lipschitz optimization problems.
3. Develop a modification of the Mirror Descent algorithm for variational inequalities with monotone and relatively bounded operator.
4. Propose a restart technique for the adaptive proximal mirror method for strongly monotone variational inequalities with a Holder continuous operator.
5. Develop an accelerated algorithm for solving strongly convex-concave saddle point problems with a decreased level of smoothness of the functional.

Relevance.

The relevance of this direction is primarily due to the sharp development of related disciplines, requiring solving multidimensional optimization problems with minimal errors. The issue of optimization of high-dimensional functions plays a crucial role in such sciences such as machine learning and data analysis. The online formulation of the problem is used in financial markets, social networks, and decision-making problems. The introduced concept of relative Lipschitz continuity helps to solve reinforcement learning problems. Variational inequalities, in turn, are an essential tool for solving problems of general economics, market equilibrium search, and complementary problems.

The obtained results:

1. An analogue of the Mirror Descent Algorithm has been developed for optimization problems of quasi-convex functions satisfying the condition of non-standard growth in the presence of quasi-convex functional inequalities.
2. A variant of the Mirror Descent method was proposed for convex programming problems on a class of relatively Lipschitz problems, including online optimization problems as well as problems in the

stochastic setting; theoretical estimates of the convergence rate of Mirror Descent algorithms for solving optimization problems with functions admitting representation in abstract model generality were obtained.

3. A modification of the Mirror Descent Algorithm for variational inequalities with a monotone and relatively bounded operator was proposed, an estimate of the convergence rate that can be considered optimal was proved.
4. A restarted version of the Adaptive Proximal Mirror for variational inequalities with a strongly monotone Holder continuous operator was proposed, an estimate of the convergence rate that is optimal at $\nu = 0$ and $\nu = 1$ was proved.
5. A technique for accelerating an algorithm for solving strongly convex-concave saddle point problems with decreased smoothness has been described.

Novelties:

1. For the first time, an analogue of the Mirror Descent Method with switchings was proposed for minimization problems with quasi-convex objective functional with quasi-convex inequality constraints.
2. The restart technique of Adaptive Proximal Mirror Method for strongly monotone variational inequalities with Holder continuous operators was proposed for the first time.
3. An accelerated method for the saddle point problems with decreased smoothness was proposed for the first time.

Reliability of the obtained results is due to the publication of 12 articles indexed by Scopus and Web of Science. Below is a list of publications related to the materials of the thesis.

First-tier publications

1. Bayandina, A., Dvurechensky, P., Gasnikov, A., Stonyakin, F., Titov, A. Mirror descent and convex optimization problems with non-smooth inequality constraints //Large-scale and distributed optimization. – Springer, Cham, 2018. – C. 181-213.
2. Gasnikov, A. V., Dvurechensky, P. E., Stonyakin, F. S., Titov, A. A. An adaptive proximal method for variational inequalities //Computational

Mathematics and Mathematical Physics. – 2019. – T. 59. – №. 5. – C. 836-841.
3. 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. – 2022. – C. 1-26.
4. Ablaev, S. S., Titov, A. A., Stonyakin, F. S., Alkousa, M. S., Gasnikov, A. (2022). Some Adaptive First-Order Methods for Variational Inequalities with Relatively Strongly Monotone Operators and Generalized Smoothness. In International Conference on Optimization and Applications (pp. 135-150). Springer, Cham.

Second-tier publications

1. Titov, A. A., Stonyakin, F. S., Gasnikov, A. V., Alkousa, M. S. Mirror descent and constrained online optimization problems //International Conference on Optimization and Applications. – Springer, Cham, 2018. – C. 64-78.
2. Stonyakin, F. S., Alkousa, M. S., Titov, A. A., Piskunova, V. V. On some methods for strongly convex optimization problems with one functional constraint //International Conference on Mathematical Optimization Theory and Operations Research. – Springer, Cham, 2019. – C. 82-96.
3. Stonyakin, F. S., Alkousa, M., Stepanov, A. N., Titov, A. A. Adaptive mirror descent algorithms for convex and strongly convex optimization problems with functional constraints //Journal of Applied and Industrial Mathematics. – 2019. – T. 13. – №. 3. – C. 557-574.
4. Stonyakin F.S., Stepanov A.N., Gasnikov A.V., Titov A.A. Mirror descent for constrained optimization problems with large subgradient values of functional constraints // Computer Research and Modeling, 2020, vol. 12, no. 2, pp. 301-317
5. 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 //International Conference on Mathematical Optimization Theory and Operations Research. – Springer, Cham, 2020. – C. 133-149.

6. Titov, A. A., Stonyakin, F. S., Alkousa, M. S., Gasnikov, A. V. Algorithms for solving variational inequalities and saddle point problems with some generalizations of Lipschitz property for operators //International Conference on Mathematical Optimization Theory and Operations Research. – Springer, Cham, 2021. – C. 86-101.
7. Savchuk O.S., Titov A.A., Stonyakin F.S., Alkousa M.S. Adaptive first-order methods for relatively strongly convex optimization problems // Computer Research and Modeling, 2022, vol. 14, no. 2, pp. 445-472.

Other publications

1. F. S. Stonyakin, A. A. Titov. One Mirror Descent algorithm for convex constrained optimization problems with non-standard growth properties.// SchoolSeminar on Optimization Problems and their Applications, OPTA-SCL 2018. CEUR-WS 2018, Vol. 2098, P. 372–384.

Reports at conferences and seminars.

1. MIPT Scientific Conference, 2018, 2019.
2. 23rd International Symposium on Mathematical Programming (ISMP 2018), Bordeaux, France.
3. International Conference Optimization and Applications (OPTIMA), 2019, 2020 Petrovac, Montenegro.
4. Mathematical Optimization Theory and Operations Research (MOTOR), 2019, 2020, 2021. Novosibirsk, Irkutsk, Russia
5. Quasilinear Equations, Inverse Problems and Their Applications (QIPA), 2018, 2019, 2021, Moscow, Russia.
6. International Symposium on Application of Numerical Optimization Methods for Solving Inverse Problems, 2021, Moscow, Russia.
7. Moscow Conference on Combinatorics and Applications, 2021, Dolgoprudny, Russia.
8. International Conference “Optimization without Borders”, 2021, Sochi, Russia.

Contents

In the first chapter of the dissertation, the formal statement of the problem is given, and some modifications of the Mirror Descent method of the minimization problem of functions with non-standard growth conditions are proposed. In item 1.1 the general statement of the optimization problem as well as basic definitions used in the dissertation, are presented. As noted earlier, an essential focus of the thesis is the abandonment of the classical distance work in favor of distance in a more general sense. Let us introduce some basic concepts concerning the so-called Bragman distance.

Let $(E, ||\cdot||)$ be some normed finite-dimensional vector space, E^* — the space of continuous linear functionals, defined in E — its conjugated. Let the norm of the conjugate space be defined as follows

$$\|y\|_{E,*} = \|y\|_* = \max_x \Big\{ \langle y,x \rangle, ||x|| \leq 1 \Big\}, \tag{4}$$

where $\langle y,x \rangle$ denotes the value of a continuous linear functional y at $x \in E$. Consider a convex compact subset $X \subset E$, and two convex subdifferentiable functions $f(x) : X \to \mathbb{R}$ и $g(x) : X \to \mathbb{R}$.

Definition 1. *Let $d(x) : X \to \mathbb{R}^+$ be some continuously differentiable and 1-strongly convex function with respect to the norm $\|\cdot\|$, i.e.*

$$\langle \nabla d(x) - \nabla d(y), x - y \rangle \geq \|x - y\|^2 \quad \forall x, y \in X. \tag{5}$$

Let us call the function $d(x)$ a prox or distance generating function.

Definition 2. *Let us say that $V_d(y, x) = V(y, x)$ — Bregman distance, generated by the prox function $d(\cdot)$, if the following equality is satisfied*

$$V(y, x) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle. \tag{6}$$

In Section 1.2 an adaptive modification of the Mirror Descent method [18] for solving the optimization problem with a functional constraint is considered

$$f(x) \to \min_{x \in dom f}, \tag{7}$$

$$\text{s.t.} \quad g(x) \leq 0. \tag{8}$$

The convergence rate (the number of iterations sufficient to obtain ε-accuracy of the problem in question) is $O\left(\frac{1}{\varepsilon^2}\right)$, assuming that the objective function and functional constraint satisfy the Lipschitz condition, that is, the following inequalities are satisfied for all $x, y \in X$

$$|f(x) - f(y)| \leq M_f ||x - y||, \tag{9}$$

$$|g(x) - g(y)| \leq M_g ||x - y||. \tag{10}$$

Hereinafter ε will be understood as the accuracy of the solution of the problem.

Definition 3. *Let us say that the point z is an ε–solution of the problem (7-8), if the following inequalities hold*

$$f(z) - f(x_*) \leq \varepsilon, \tag{11}$$

$$g(z) \leq \varepsilon. \tag{12}$$

If the objective function $f(x)$ does not satisfy the Lipschitz condition, but has a Lipschitz continuous gradient, a corresponding modification of the Mirror Descent method [18] with a similar convergence rate is considered. A natural example of the emergence of a problem statement with such a smoothness class is quadratic functions.

Section 1.3 is devoted to the case of non-convex functions, including both simultaneous quasi-convexity of the objective function and the functional constraint and the case of quasi-convexity of the objective function alone. The convergence rate of the proposed methods is also $O\left(\frac{1}{\varepsilon^2}\right)$. Methods of minimization of quasi-convex functions found a lot of applications in many applied problems, among which the problem of search for internal rate of return was considered in the framework of this dissertation.

Definition 4 ([10])**.** *Function $f : X \to \mathbb{R}$ is called quasi-convex if the following inequality holds*

$$f\Big((1-\alpha)x + \alpha y\Big) \leqslant \max\{f(x), f(y)\} \quad \forall \alpha \in [0;1] \quad \forall x, y \in X. \tag{13}$$

When working with quasi-convex functions instead of the classical (sub)gradient the following set is often considered [25]

$$\hat{D}f(x) = \{p \mid \langle p, x - y\rangle \geqslant 0 \quad \forall y \in X : \ f(y) \leqslant f(x)\}. \tag{14}$$

Hereinafter $Df(x)$ will be understood as an arbitrary vector of $\hat{D}f(x)$

$$Df(x) \in \hat{D}f(x). \tag{15}$$

For a given function $f(x)$ and each subgradient $\nabla f(x)$ at the point $y \in X$, define the following function that will be used to characterize the complexity of the Algorithm 1

$$v_f(x,y) = \begin{cases} \left\langle \frac{\nabla f(x)}{\|\nabla f(x)\|_*}, x - y \right\rangle, & \nabla f(x) \neq 0 \\ 0 & \nabla f(x) = 0 \end{cases}, \quad x \in X. \tag{16}$$

Let us introduce the following definition of a proximal operator.

Definition 5. *For any $x \in X$ and $p \in E^*$ define a proximal operator* $\mathrm{Mirr}_x(p)$ *as follows*

$$\mathrm{Mirr}_x(p) = \arg\min_{y \in X} \Big\{ \langle p, y \rangle + V(y,x) \Big\}. \tag{17}$$

Algorithm 1 Modification of Adaptive Mirror Descent for quasi-convex functions

---

Require: $\varepsilon > 0$; Θ_0, such that $d(x_*) \leq \Theta_0^2$, C_f,C_g
$x_0 = \arg\min_{x \in X} d(x)$
Define $I = \emptyset$, $k = 0$
repeat
  if $g(x_k) \leq \varepsilon M_g$ **then**
    $h_k^f = \frac{C_f}{\|Df(x_k)\|_*}$
    $x_{k+1} = \mathrm{Mirr}_{x_k}\Big(h_k^f Df(x_k)\Big)$ "productive step"
    $I = I \cup \{k\}$.
  else
    $h_k^g = \frac{C_g}{\|Dg(x_k)\|_*}$
    $x_{k+1} = \mathrm{Mirr}_{x_k}\Big(h_k^g Dg(x_k)\Big)$ "non-productive step"
  end if
  $k = k + 1$
until $\frac{2\Theta_0^2}{\varepsilon^2} \leq N$
Ensure: $\widetilde{x} := \arg\min_{i \in I} f(x_i)$

Theorem 1. *Let $f(x)$ be a quasi-convex function, $g(x)$ be a quasi-convex function satisfying the Lipschitz condition with constant M_g. Then after $N = \left\lceil \frac{2\Theta_0^2}{\varepsilon^2} \right\rceil$ steps of Algorithm 1 the following inequalities are satisfied*

$$\min_{k\in I} v_f(x_k,x_*) \leqslant \varepsilon, \quad \max_{k\in I} g(x_k) \leqslant \varepsilon M_g. \tag{18}$$

The second chapter is primarily devoted to a generalization of the Lipschitz condition (2) in the case of replacing the difference norm by a distance in some generalized sense, more precisely, by a Bregman divergence.

Item 2.1 is devoted to the motivation in considering classes of relatively smooth and relatively Lipschitz continuous functions. An important feature of these concepts is the relaxation of the requirements for the prox-function (1), namely, the replacement of the 1-strong convexity condition by ordinary convexity. The thesis describes how the following definition of the relative smoothness of the function has found application in solving the optimal design problem.

Definition 6 ([16]). *Let us say that the function $f(x)$ satisfies the condition of relative smoothness with constant L (or is L-relatively smooth), if for all $x, y \in X$ the following inequality is satisfied*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + LV_d(y, x). \tag{19}$$

The main focus of chapter 2, however, is relatively Lipchitz continuous functions, which allowed us to take a fresh look at many well-known applied problems, among which the Support Vector Machine for the binary classification problem and the Intersection of Ellipsoids problem were considered.

Definition 7 ([15]). *Let us say that a function $f(x)$ satisfies the condition of relative Lipschitz continuity with constant M_f (or is M_f-relatively Lipschitz continuous) if for all $x, y \in X$ the following inequality is satisfied*

$$\|\nabla f(x)\|_* \leq \frac{M_f\sqrt{2V(y,x)}}{\|y-x\|} \quad \forall x, y \in X,\ y \neq x. \tag{20}$$

Example 1 (Support Vector Machine). *Consider the optimization setting of the binary classification problem solved by Support Vector Machine with l_2-*

regularization [27; 29]

$$f(x) := \frac{1}{n}\sum_{i=1}^{n} f_j(x) \to \min_x, \qquad (21)$$
$$f_j(x) := \max\left\{0{,}1 - y_i x^T w_i\right\} + \frac{\lambda}{2}\|x\|_2^2,$$

where w_i is the feature vector of the sample element, and $y_i \in \{-1{,}1\}$ is the class label. Obviously, $f(x)$ is neither differentiable nor Lipschitz continuous (due to regularization), so it is difficult to use classical (sub)gradient methods to solve (21). In [15] it is shown that the considered function $f(x)$ is 1-relatively Lipshitz continuous for the following prox-function

$$d(x) := \frac{\lambda^2}{4}\|x\|_2^4 + \frac{2\lambda}{3n}\left(\sum_{i=1}^{n}\|w_i\|_2\right)\|x\|_2^3 + \frac{1}{2n}\left(\sum_{i=1}^{n}\|w_i\|_2^2\right)\|x\|_2^2. \qquad (22)$$

Thus, using the approach with stochastic approximation of the (sub)gradient of the objective function $f(x)$ and a prox-function according to (22), the classical Mirror Descent Algorithm in the stochastic setting

$$x_{k+1} = \arg\min_{x\in X}\left\{f(x_k) + \langle\nabla f(x_k), x - x_k\rangle + \frac{1}{\varepsilon}V_d(x, x_k)\right\}, \qquad (23)$$
$$\tilde{x} := \frac{1}{k+1}\sum_{i=0}^{k} x^i$$

guarantees the point $\tilde{x}$, which is a stochastic ε–solution of the problem (21)

$$\mathbb{E}f(z) - f(x_*) \le \varepsilon, \qquad (24)$$

$$g(z) \le \varepsilon. \qquad (25)$$

Moreover, the number of steps of the algorithm is $O(\frac{1}{\varepsilon^2})$. Note that $\nabla f(x)$ in (23) is a stochastic (sub)gradient, satisfying

$$\mathbb{E}\left[\nabla f(x,\xi)\right] = \nabla f(x) \in \partial f(x), \quad \mathbb{E}\left[\nabla g(x,\zeta)\right] = \nabla g(x) \in \partial g(x), \qquad (26)$$

and

$$\|\nabla f(x,\xi)\|_* \leq M_f, \quad \|\nabla g(x,\zeta)\|_* \leq M_g \quad \textit{almost surely.} \tag{27}$$

Example 2 (Intersection of Ellipsoids)**.** *Consider n ellipsoids, each defined as follows*

$$\Upsilon_i = \left\{ x \in \mathbb{R}^m : \frac{1}{2} x^T A_i x + b_i x + c_i \leq 0 \right\}, \tag{28}$$

где $A_i \in \mathbb{S}^m_{++}$, $i = 1, \ldots, n$. The problem is to find such a point $x \in \mathbb{R}^m$, that

$$x \in \bigcap_{i=1}^{n} \Upsilon_i. \tag{29}$$

It is worth noting that the inner point methods common for solving such problems are applicable only in the case of relatively small dimension m, n. Consider the problem of finding the intersection of ellipsoids in the following form

$$f(x) := \max_{0 \leq i \leq n} \left\{ \frac{1}{2} x^T A_i x + b_i^T x + c_i \right\} \to \min_x \tag{30}$$

$f(x)$ is neither differentiable nor Lipschitz continuous. Let $\sigma := \max_{0 \leq i \leq n} \|A_i\|_2^2$, where $\|A_i\|_2$ is spectral radius A_i; $\rho := 2 \max_{0 \leq i \leq n} \|A_i b_i\|_2$, $\gamma := \max_{0 \leq i \leq n} \|b_i\|_2^2$.

[15] showed that the function $f(x)$ is 1-relative Lipschitz continuous for the following prox function

$$h(x) := \frac{\sigma}{4} \|x\|_2^4 + \frac{\rho}{3} \|x\|_2^3 + \frac{\gamma}{2} \|x\|_2^2. \tag{31}$$

Moreover, the classical Mirror Descent Algorithm (23) with the classical (sub)gradient $\nabla f(x)$ instead of the stochastic one guarantees ε–solution of the problem (30) after no more than $O(\frac{1}{\varepsilon^2})$ iterations.

Thus, various adaptive modifications of the Mirror Descent method are already proposed in section 2.2 within the framework of the relative Lipschitz continuous of functions. An important feature of the proposed methods is the assumption of function representation in some abstract generality, which is a natural generalization of the abovementioned concept of (δ, L)-model of the function [4]. Generalizations of the proposed methods in the case of multiple function constraints are also considered. More precisely, the functions are assumed to admit the following representation in model generality.

Definition 8. *Let $\delta > 0$. Let us say that functions $f(x)$ and $g(x)$ admit a relatively Lipschitz continuous (δ, ϕ, V)-model at the point $y \in X$ if*

$$f(x) + \psi_f(y,x) \le f(y), \quad -\psi_f(y,x) \le \phi_f^{-1}\Big(V(y,x)\Big) + \delta, \tag{32}$$

$$g(x) + \psi_g(y,x) \le g(y), \quad -\psi_g(y,x) \le \phi_g^{-1}\Big(V(y,x)\Big) + \delta, \tag{33}$$

where $\psi_f(y, x)$ and $\psi_g(y, x)$ are convex functions with respect to the first variable and $\psi_f(x, x) = \psi_g(x, x) = 0$ for all $x \in X$.

Let us motivate the introduced Definition 8 by the following examples [26], while omitting the problem statement in model generality, that is, let us put

$$\psi_f(y,x) = \langle \nabla f(x), y - x\rangle, \ \ \psi_g(y,x) = \langle \nabla g(x), y - x\rangle. \tag{34}$$

Example 3. *Let prox-function $d(x)$ again be 1-strongly convex and the (sub)gradient $f(x)$ be bounded. Under these assumptions, it is known that the Bregman divergence satisfies the following inequality*

$$V(y,x) \ge \frac{1}{2}\|x - y\|^2 \quad \forall x,y \in X. \tag{35}$$

Then

$$\langle \nabla f(x), y - x\rangle \le M_f\|x - y\| \le M_f\sqrt{2V(y,x)}, \tag{36}$$

so one can consider

$$\phi_f(t) = \frac{t^2}{2M_f^2}. \tag{37}$$

Example 4. *Consider the problem of maximizing a positive concave function $q(x) : X \to \mathbb{R}_+$:*

$$q(x) \to \max_{x\in X}. \tag{38}$$

Then the function $f(x) := -\log q(x)$ will satisfy the inequality (36) with $M_f = 1$. In this case, the maximization problem (38) can be solved by standard minimization of the function $f(x)$.

Example 5 (Composite optimization problem [2; 16; 23])**.** *The method proposed below, as well as its modifications considered in the dissertation, are applicable to* *the composite optimization problem*

$$\min\{f(x)+r(x): \quad x \in X,\ g(x)+\eta(x) \leq 0\}, \tag{39}$$

where $r(x), \eta(x): Q \to \mathbb{R}$ *are simple convex functions. Then for all* $x, y \in X$

$$\psi_f(y,x) = \langle \nabla f(x), y-x \rangle + r(y) - r(x), \tag{40}$$

$$\psi_g(y,x) = \langle \nabla g(x), y-x \rangle + \eta(y) - \eta(x). \tag{41}$$

Let us define the proximal operator for the step $h > 0$ as follows

$$\mathrm{Mirr}_h(x,\psi) = \arg\min_{y \in X} \left\{ \psi(y,x) + \frac{1}{h} V(y,x) \right\}. \tag{42}$$

Algorithm 2 Mirror Descent Method in Model Generality.

Require: $\varepsilon > 0, \delta > 0, h^f > 0, h^g > 0, \Theta_0,\ d(x_*) \leq \Theta_0^2$

$x_0 = \arg\min_{x \in X} d(x)$

Set $I = \emptyset$, $J =: \emptyset$

$k = 0$

repeat

  if $g(x_k) \leq \varepsilon + \delta$ **then**

    $x_{k+1} = \mathrm{Mirr}_{h^f}(x_k, \psi_f)$ "productive step"

    $I = I \cup \{k\}$

  else

    $x_{k+1} = \mathrm{Mirr}_{h^g}(x_k, \psi_g)$ "non-productive step"

    $J = J \cup \{k\}$

  end if

  $k = k + 1$

until $\Theta_0^2 \leq \varepsilon\left(|J|h^g + |I|h^f\right) - |J|\phi_g^*(h^g) - |I|\phi_f^*(h^f)$

Ensure: $\widetilde{x} := \frac{1}{|I|} \sum_{k \in I} x_k$

Theorem 2. *Let* $f(x)$ *and* $g(x)$ *be convex functions admitting the model generality representation* (32), (33), *respectively,* $\varepsilon > 0, \delta > 0$. *As before, suppose that there exists a constant* $\Theta_0 > 0$ *such that* $d(x_*) \leq \Theta_0^2$. *Suppose that at a certain step of the Algorithm* 2 *the stopping criterion is satisfied, then the* *following inequality holds*

$$f(\widetilde{x}) - f(x_*) \leq \varepsilon + \delta, \qquad g(\widetilde{x}) \leq \varepsilon + \delta. \tag{43}$$

Further, various options for specifying the type of function model are considered and there are proposed corresponding optimal algorithms.

Section 2.3 is devoted to the online optimization problem with a functional constraint

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^{N} f_i(x) \to \min_x \\ &s.t. \;\; g(x) \leq 0 \end{aligned} \tag{44}$$

There were proposed algorithms to solve the problem both in the classical formulation and in the case when functions admit a representation in model generality. The proposed methods are also optimal [9]. More precisely, the number of non-productive steps is proportional to the total number of iterations of algorithms N. Also, in the dissertation, the case of negative regret was analyzed, and corresponding theoretical estimates of the number of non-productive steps, in this case, were obtained.

Section 2.4 is devoted to the stochastic setting of the optimization problem with preserving assumptions about the smoothness of the objective function and functional constraint; moreover, the proposed methods also have optimal estimates of the convergence rate $O\left(\frac{1}{\varepsilon^2}\right)$.

Section 2.5 considers the minimization problem of the relatively strongly convex

$$f(x) - f(y) + \mu V(y,x) \leq \langle \nabla f(x), x - y \rangle \quad \forall x, y \in dom f \tag{45}$$

of the objective function with functional constraint and proposes a restart procedure of the previously introduced Mirror Descent Algorithm with convergence rate $O\left(\frac{M^2}{\mu\varepsilon}\right)$, where M — the maximum constant of the relative Lipschitz continuity of the objective function M_f and functional constraint M_g

$$M = \max\{M_f, M_g\}. \tag{46}$$

<u>Secton 3</u> is devoted to variational inequality problems for some operator $g : Q \longrightarrow \mathbb{R}^n$, where Q is some convex closed subset of $\mathbb{R}^n$

$$\max_{x} \langle g(x), x_* - x\rangle \leq 0, \tag{47}$$

and saddle point problem

$$f^* = \min_{x} \max_{y} f(x,y), \tag{48}$$

where $f(x,y) : Q_x \times Q_y \to \mathbb{R}$ is convex with respect to x and concave with respect to y, $Q_x \subset E_1$, $Q_y \subset E_2$ are convex compact subsets of some normalized finite-dimensional vector spaces with given norms $\|\cdot\|\|_1$, $\|\cdot\|_2$, respectively.

In Section 3.1, there is considered an Adaptive Proximal method for solving variational inequalities with Lipschitz continuous operator that guarantees an ε–solution after no more than $O\left(\frac{1}{\varepsilon}\right)$ iterations, which is an optimal estimate.

Section 3.2 proposes an analogue of the Mirror Descent method for variational inequalities with relatively bounded

$$\langle g(x),y - x\rangle \leq M\sqrt{2V(y,x)}, \tag{49}$$

and monotone

$$\langle g(y) - g(x),y - x\rangle \geq 0, \tag{50}$$

operator, which guarantees an ε–solution of (47) after no more than $O\left(\frac{1}{\varepsilon^2}\right)$ iterations.

Further, in Section 3.3 there is proposed an accelerated method for solving the saddle point problem (48) assuming that the gradient of the objective function partially satisfies the Hodler condition, $\nu \in [0,1]$, while being smooth in one variable (the gradient satisfies the Lipschitz condition)

$$\|\nabla_x f(x,y) - \nabla_x f(x',y)\|_2 \leq L_{xx}\|x - x'\|_2^{\nu}, \tag{51}$$

$$\|\nabla_x f(x,y) - \nabla_x f(x,y')\|_2 \leq L_{xy}\|y - y'\|_2^{\nu}, \tag{52}$$

$$\|\nabla_y f(x,y) - \nabla_y f(x',y)\|_2 \leq L_{xy}\|x - x'\|_2^{\nu}, \tag{53}$$

$$\|\nabla_y f(x,y) - \nabla_y f(x,y')\|_2 \leq L_{yy}\|y - y'\|_2. \tag{54}$$

The number of iterations to obtain the ε–solution is given as follows

$$\mathcal{O}\left(\sqrt{\frac{L}{\mu_x}} \cdot \log \sqrt{\frac{L_{yy}}{\mu_y}} \cdot \log \frac{2LR^2}{\varepsilon}\right), \tag{55}$$

where

$$L = \tilde{L}\left(\frac{\tilde{L}}{2\varepsilon}\frac{(1-\nu)(2-\nu)}{2-\nu}\right)^{\frac{(1-\nu)(1+\nu)}{2-\nu}}, \tilde{L} = \left(L_{xy}\left(\frac{2L_{xy}}{\mu_y}\right)^{\frac{\nu}{2-\nu}} + L_{xx}D^{\frac{\nu-\nu^2}{2-\nu}}\right), \tag{56}$$

where D is diameter of the domain of $f(x,\cdot)$.

Section 3.4 first proposes a restart technique for variational inequalities with Holder continuous and strongly monotone operator

$$\langle g(x) - g(y), x - y\rangle \geq \mu\|x-y\|^2, \tag{57}$$

it is assumed that the calculation of the operator satisfying the Holder condition with constant L_ν is acceptable with some inaccuracy. The convergence rate of the algorithm in this case is

$$O\left(\left(\frac{L_\nu}{\mu}\right)^{\frac{2}{1+\nu}} \frac{2^{\frac{2}{1+\nu}}\Omega}{\varepsilon^{\frac{1-\nu}{1+\nu}}} \cdot \log_2 \frac{2R_0^2}{\varepsilon}\right), \tag{58}$$

where R_0, Ω are some characteristics of the considered space. It is worth noting that for $\nu = 0$ the convergence rates of the proposed algorithm and the accelerated method, proposed in Section 3.3, coincide, while for $\nu > 0$ the asymptotic convergence rate of the accelerated method is better.

Definition 9. *Suppose that for some $\delta_u > 0$ (uncontrollable error) and for any $\delta_c > 0$ (controllable error) there exists a constant $L(\delta_c) \in (0, +\infty)$ that $\forall x,y \in Q$ one can calculate such $\tilde{g}(x,\delta_c,\delta_u)$ and $\tilde{g}(y,\delta_c,\delta_u) \in E^*$ that the following inequalities are satisfied*

$$\langle \tilde{g}(y,\delta_c,\delta_u) - \tilde{g}(x,\delta_c,\delta_u), y - z\rangle \leq \frac{L(\delta_c)}{2}\left(\|y-x\|^2 + \|y-z\|^2\right) + \delta_c + \delta_u, \tag{59}$$

$$\langle \tilde{g}(y,\delta_c,\delta_u) - g(y), y - z\rangle \geq -\delta_u, \quad \forall z \in Q. \tag{60}$$

Then let us call the operator $\tilde{g}(\cdot,\delta_c,\delta_u)$ an inexact oracle of g.

Algorithm 3 Adaptive Proximal Method for variational inequalities with inexact oracle

Require: $\varepsilon > 0,\ \delta_u > 0,\ \delta_{pu} > 0,\ M_{-1},\ L(\delta_c),\ d(x)$

$k = 0,\ z_0 = \arg\min_{u\in Q} d(u)$

for $k = 0,1,\ldots$ **do**

  $i_k = 0,\ \delta_{c,k} = \frac{\varepsilon}{4},\ \delta_{pc,k} = \frac{\varepsilon}{8}$

  repeat

    $M_k = 2^{i_k - 1} M_{k-1}$

    Calculate

$$w_k = \arg\min_{x\in Q}{}^{\delta_{pc,k}+\delta_{pu}} \left\{ \langle \tilde{g}(z_k,\delta_{c,k},\delta_u),x\rangle + M_k V(x,z_k) \right\} \tag{61}$$

$$z_{k+1} = \arg\min_{x\in Q}{}^{\delta_{pc,k}+\delta_{pu}} \left\{ \langle \tilde{g}(w_k,\delta_{c,k},\delta_u),x\rangle + M_k V(x,z_k) \right\} \tag{62}$$

    $i_k = i_k + 1$

  until

$$\langle \tilde{g}(w_k,\delta_c,\delta_u) - \tilde{g}(z_k,\delta_c,\delta_u), w_k - z_{k+1}\rangle \le \frac{M_k}{2}\left(\|w_k - z_k\|^2 + \|w_k - z_{k+1}\|^2\right) + \delta_{c,k} + \delta_u \tag{63}$$

  $k = k + 1$

end for

Ensure: $\widehat{w}_k = \frac{1}{\sum_{i=0}^{k-1} M_i^{-1}} \sum_{i=0}^{k-1} M_i^{-1} w_i$

Theorem 3 ([7]). *Suppose that $g(\cdot)$ and $\tilde{g}(\cdot,\delta_c,\delta_u)$ satisfy (59) and (60). Then, $\forall k \ge 1$ and any $u \in Q$*

$$\begin{aligned} &\frac{1}{\sum_{i=0}^{k-1} M_i^{-1}} \sum_{i=0}^{k-1} M_i^{-1} \langle g(w_i), w_i - u\rangle \\ &\le \frac{1}{\sum_{i=0}^{k-1} M_i^{-1}} (V(u,z_0) - V(u,z_k)) + \frac{\varepsilon}{2} + \delta_u + 2\delta_{pu}. \end{aligned} \tag{64}$$

Moreover, the total number of oracle calls does not exceed

$$\begin{aligned} \inf_{\nu\in[0,1]} \left(16 \left(\frac{L_\nu}{\varepsilon}\right)^{\frac{2}{1+\nu}} \cdot \max_{u\in C} V(u,z_0) + 2\log_2 2 \left(\left(\frac{1}{\varepsilon}\right)^{\frac{1-\nu}{1+\nu}} L_\nu^{\frac{2}{1+\nu}} \right) \right) \\ -2\log_2(M_{-1}). \end{aligned} \tag{65}$$

Algorithm 4 Restarts of Adaptive Proximal Method for Variational Inequalities with Inexact Oracle

Require: $\varepsilon > 0$, $\delta_u > 0$, $\delta_{pu} > 0$, $\mu > 0$, Ω, такое что $d(x) \leq \frac{\Omega}{2}$ $\forall x \in Q : \|x\| \leq 1$; x_0, R_0, такое что $\|x_0 - x_*\|^2 \leq R_0^2$

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

2: **repeat**

3: x_{p+1} — result of 3 with accuracy $\frac{\mu\varepsilon}{2}$, δ_u, δ_{pu}, prox-function $d_p(\cdot)$ and stopping criterion $\sum_{i=0}^{k-1} M_i^{-1} \geq \frac{\Omega}{\mu}$

4: $R_{p+1}^2 = R_0^2 \cdot 2^{-(p+1)} + 2(1 - 2^{-(p+1)})(\frac{\varepsilon}{4} + \delta_u + 2\delta_{pu})$

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** $p > \log_2 \frac{2R_0^2}{\varepsilon}$.

Ensure: x_p.

Theorem 4. *Suppose that operator $g(x)$ is $\mu > 0$–strongly monotone. Also, suppose that prox-function $d(x)$ satisfies $d(x) \leq \frac{\Omega}{2}$ $\forall x \in Q :$ $\|x\||leq1$, and initial point $x_0 \in Q$ and $R_0 > 0$ are such that $\|x_0 - x_*\|^2 \leq R_0^2$. Then for $p \geq 0$, the sequence x_p, generated by Algorithm 4, satisfies*

$$\|x_p - x_*\|^2 \leq R_0^2 \cdot 2^{-p} + \frac{\varepsilon}{2} + \frac{2\delta_u + 4\delta_{pu}}{\mu}, \tag{66}$$

and the point x_p, which is the result of Algorithm 4, satisfies

$$\|x_p - x_*\|^2 \leq \varepsilon + \frac{2\delta_u + 4\delta_{pu}}{\mu}. \tag{67}$$

Conclusion contains the main results of the work, which are as follows.

Within the framework of the dissertation for the first time there were proposed

1. an analogue of the Mirror Descent method with switchings for problems of minimization a quasi-convex objective functional with quasi-convex inequality constraint;
2. the restart technique of the Adaptive Proximal Mirror method for strongly monotone variational inequalities with Holder continuous operators;
3. an accelerated method for the (Holder) saddle point problem with decreased smoothness level.

Within the framework of the given problems, an analogue of the mirror descent method with switchings was developed to solve the problem of minimization of quasi-convex non-Lipschitz continuous functions with quasi-convex Lipchitz continuous inequality constraint with convergence rate $O\left(\frac{1}{\varepsilon^2}\right)$. An algorithm was proposed to solve the problem of minimizing a quasi-convex objective function that does not satisfy the Lipschitz condition but has a Lipschitz continuous gradient with a quasi-convex constraint. Optimal Mirror Descent methods for relatively Lipschitz optimization problem with a functional constraint in the case of an online and stochastic problem setting were proposed. In doing so, the concept of model generality of the function of the corresponding smoothness class was considered. Appropriate modifications of the Mirror Descent method for relatively Lipschitz problems in the case of the stochastic setting of the optimization problem were proposed.

Список литературы

1. *Bauschke H. H.*, *Bolte J.*, *Teboulle M.* A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications // Mathematics of Operations Research. — 2017. — Т. 42, № 2. — С. 330—348.
2. *Beck A.*, *Teboulle M.* A fast iterative shrinkage-thresholding algorithm for linear inverse problems // SIAM journal on imaging sciences. — 2009. — Т. 2, № 1. — С. 183—202.
3. *Clarke F. H.* Method of Dynamic and Nonsmooth Optimization. — SIAM, 1989.
4. *Devolder O.*, *Glineur F.*, *Nesterov Y.* First-order methods of smooth convex optimization with inexact oracle // Mathematical Programming. — 2014. — Т. 146, № 1. — С. 37—75.
5. *Dupuis P.*, *Nagurney A.* Dynamical systems and variational inequalities // Annals of Operations Research. — 1993. — Т. 44, № 1. — С. 7—42.
6. *Gasnikov A.* Modern numerical optimization methods // The method of universal gradient descent. Moscow: MIPT. — 2018.
7. Generalized Mirror Prox Algorithm for Monotone Variational Inequalities: Universality and Inexact Oracle / F. Stonyakin [и др.] // Journal of Optimization Theory and Applications. — 2022. — С. 1—26.
8. *Giannessi F.*, *Maugeri A.* Variational inequalities and network equilibrium problems. — Springer, 1995.
9. *Hazan E.*, *Kale S.* Beyond the regret minimization barrier: an optimal algorithm for stochastic strongly-convex optimization // Proceedings of the 24th Annual Conference on Learning Theory. — JMLR Workshop, Conference Proceedings. 2011. — С. 421—436.
10. *Hazan E.*, *Levy K.*, *Shalev-Shwartz S.* Beyond convexity: Stochastic quasi-convex optimization // Advances in neural information processing systems. — 2015. — Т. 28.

11. *Jofré A.*, *Rockafellar R. T.*, *Wets R. J.* Variational inequalities and economic equilibrium // Mathematics of Operations Research. — 2007. — Т. 32, № 1. — С. 32—50.

12. *Kinderlehrer D.*, *Stampacchia G.* An introduction to variational inequalities and their applications. — SIAM, 2000.

13. *Korpelevich G. M.* The extragradient method for finding saddle points and other problems // Matecon. — 1976. — Т. 12. — С. 747—756.

14. *Liu Y.*, *Wang Y.*, *Singh A.* Smooth Bandit Optimization: Generalization to Holder Space // International Conference on Artificial Intelligence and Statistics. — PMLR. 2021. — С. 2206—2214.

15. *Lu H.* "Relative continuity" for non-lipschitz nonsmooth convex optimization using stochastic (or deterministic) mirror descent // INFORMS Journal on Optimization. — 2019. — Т. 1, № 4. — С. 288—303.

16. *Lu H.*, *Freund R. M.*, *Nesterov Y.* Relatively smooth convex optimization by first-order methods, and applications // SIAM Journal on Optimization. — 2018. — Т. 28, № 1. — С. 333—354.

17. *Makela M. M.*, *Neittaanmaki P.* Nonsmooth optimization: analysis and algorithms with applications to optimal control. — World Scientific, 1992.

18. Mirror Descent and Convex Optimization Problems With Non-Smooth Inequality Constraints / A. Bayandina [et al.] // LCCC Focus Period on Large-Scale and Distributed Optimization. — 2018. — P. 181–213.

19. *Nagurney A.* Network economics: A variational inequality approach. Т. 10. — Springer Science & Business Media, 1998.

20. *Nagurney A.*, *Zhang D.* Projected dynamical systems and variational inequalities with applications. Т. 2. — Springer Science & Business Media, 1995.

21. *Nakamura T.*, *Horio H.*, *Chiba Y.* Local holder exponent analysis of heart rate variability in preterm infants // IEEE Transactions on biomedical engineering. — 2005. — Т. 53, № 1. — С. 83—88.

22. *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. — T. 15, № 1. — C. 229—251.

23. *Nesterov Y.* Gradient methods for minimizing composite functions // Mathematical programming. — 2013. — T. 140, № 1. — C. 125—161.

24. *Nesterov Y.* Universal gradient methods for convex optimization problems // Mathematical Programming. — 2015. — T. 152, № 1. — C. 381—404.

25. *Nesterov Y. E.* Effective methods in nonlinear programming // Moscow, Radio i Svyaz. — 1989.

26. *Nesterov Y.* Relative smoothness: new paradigm in convex optimization // Conference report, EUSIPCO-2019, A Coruna, Spain. T. 4. — 2019.

27. *Oneto L.*, *Ridella S.*, *Anguita D.* Tikhonov, Ivanov and Morozov regularization for support vector machine learning // Machine Learning. — 2016. — T. 103, № 1. — C. 103—136.

28. *Pang J.-S.*, *Fukushima M.* Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games // Computational Management Science. — 2005. — T. 2, № 1. — C. 21—56.

29. *Platt J.* Sequential minimal optimization: A fast algorithm for training support vector machines. — 1998.

30. *Polyak B. T.* Introduction to optimization // Science. — 1987. — Vol. 1, no. 32.

31. *Rademacher H.* Über partielle und totale differenzierbarkeit von Funktionen mehrerer Variabeln und über die Transformation der Doppelintegrale // Mathematische Annalen. — 1919. — T. 79, № 4. — C. 340—359.

32. *Schwartz L.* Analyse mathématique. — 1967.

33. *Stonyakin F. S.* Adaptive Mirror Descent Methods for Convex Programming Problems with delta-subgradients // arXiv preprint arXiv:2012.12856. — 2020.