ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»

На правах рукописи

Титов Александр Александрович

МЕТОДЫ ОПТИМИЗАЦИИ ДЛЯ НЕГЛАДКИХ ЗАДАЧ В ПРОСТРАНСТВАХ БОЛЬШИХ РАЗМЕРНОСТЕЙ

Резюме

диссертации на соискание учёной степени кандидата
компьютерных наук

Москва — 2023

Работа выполнена в Московском физико-техническом институте.

Научный руководитель: д. ф.-м. н. **Гасников Александр Владимирович**

Общая характеристика работы

Необходимость развития численных методов негладкой выпуклой оптимизации в последние годы обусловлена значительным прогрессом в различных областях науки, в том числе, биологии, экономики, химии, прикладной математики, теоретической физики и многих других. Особенной трудностью, прежде всего, в задачах машинного обучения и анализа данных, является большой объем исходной выборки, важность получения как можно более точного решения и минимизации погрешности, а также сложность в вычислении значения функции или ее производных, которые описывают конкретную математическую модель. Последний аспект является особенно актуальным в связи с невозможностью осуществления точной численной оценки разнообразных характеристик функции во многих прикладных задачах. Таким образом, многие классические алгоритмы оптимизации оказываются неприменимыми, например, в случае, если целевая функция является негладкой. Стоит отметить, что на сегодняшний день подавляющее большинство прикладных задач порождают задачи оптимизации именно негладких функций [3; 17; 26].

Хорошо известно, что как выпуклая [31], так и липшицевая [30] функция, является дифференцируемой на своей области определения почти всюду. Тем не менее, для многих прикладных задач с подобными свойствами целевого функционала, методы оптимизации для гладкого случая являются не применимыми. Нетрудно показать [6], что для функции, градиент которой удовлетворяет условию Липшица (далее — для гладкой функции):

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

выполнено следующее неравенство:

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

Естественной попыткой обобщения неравенства (2) является замена второго слагаемого правой части (2), на расстояние в некотором обобщенном смысле. Мотивацией данного подхода являются как трудности с использованием евклидового расстояния во многих прикладных задачах, так и преднамеренная замена $\|x - y\|^2$ на расстояние, более адаптированное к конкретной постановке рассматриваемой задачи. Такая идея была предложена в [16] с заменой нормы разности переменных x, y на расстояние в обобщенном смысле, а именно — на дивергенцию Брэгмана. Важным моментом подобного обобщения является сохранение оптимальной скорости сходимости методов первого порядка [1; 15; 16].

Так, одним из основных результатов в данном направлении была недавно предложенная концепция относительной гладкости [16], позволившая впервые применить метод градиентного типа для решения задачи о построении оптимального эллипсоида, покрывающего набор заданных точек. Эта задача играет важнейшую роль в статистике и анализе данных [1; 16]. Далее, в [15] была предложена концепция относительной липшицевости, позволившая по новому взглянуть на многие прикладные задачи, среди которых будет рассмотрен метод опорных векторов для задачи бинарной классификации, а также задача о поиске общей точки n заданных эллипсоидов.

Второй проблемой, возникающей при рассмотрении неравенства (2), является невозможность точного вычисления градиента функции $f(x)$. Недавно в [32; 33] было показано, что различные модификации алгоритма зеркального спуска, применимы в случае использования так называемого δ–субградиента, при этом показано, что накопления величины погрешности в итоговых оценках скорости сходимости не происходит. Однако, во многих прикладных задачах затруднительно не только вычисление градиента с некоторой погрешностью, но и само значение целевой функции. Одним из наиболее важных результатов в данном направлении была предложенная в [4] концепция (δ,L)–оракула, ухудшающая качество решения задачи оптимизации по функции лишь на $O(\delta)$.

В первой части диссертации расширяется класс рассмотренных задач, в том числе для абстрактной модели функции, замещающей первое слагаемое $\langle\nabla f(x), y - x\rangle$ с некоторой погрешностью, при этом допускается, что сама функция представима в абстрактном виде с помощью так называемой модели, позволяющей работать с задачами композитной оптимизации. Стоит отметить, что при этом сохраняются оптимальные оценки скорости сходимости для предложенных методов.

Вторым фокусом диссертации являются вариационные неравенства и седловые задачи с соответствующими уровнями гладкости операторов. Стоит отметить, что вариационные неравенства играют ключевую роль в решении многих прикладных в области гидродинамики [12], проектирования динамических систем [5; 20] экономики, в частности при моделировании сетевого эффекта [19], поиска общего экономического равновесия [5; 8; 11], равновесия Нэша [28], матричных играх [22] и т. д.

Одним из наиболее заметных численных методов решения вариационных неравенств был экстраградиентный метод Г. М. Корпелевич, предложенный в 1976 году [13]. Недавно, объединяя подходы, описанные в [22] и [24] был предложен универсальный численный алгоритм [7], который способен производить автоматическую настройку на уровень гладкости задачи (а именно на параметр ν, см. (3) далее). В диссертации этот метод распространяется на условие сильной монотонности оператора в предположениях введенных ранее классах гладкости. Также, рассматривается вариант недавно предложенного ускоренного метода решения седловых задач в негладкой постановке. В частности, для седловых задач предлагается обобщение условия Липшица градиента целевой функции (1) на следующее условие Гельдера:

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

играющее важную роль в решении многих прикладных задач, таких как задача о многоруком бандите (multi armed bandit problem) [14], задача определения вариабельности сердечного ритма (heart rate variability problem) [21] и др.

Целью данной работы является разработка оптимальных численных методов решения многомерных задач негладкой выпуклой оптимизации с функциональными ограничениями. В качестве базы для разработки методов используются известные методы зеркального спуска и их модификации. Для возможности применения предложенных методов на более широком классе функционалов вводится модификация концепции неточной модели целевого функционала и функционального ограничения. Также исследуется применимость методов для функционалов, удовлетворяющих ослабленному варианту условия Липшица, а именно условию относительно липшицевости функции классической задачи оптимизации и относительной гладкости оператора вариационного неравенства. Также цель диссертации включает в себя разработку модификаций методов для решения вариационных неравенств и седловых задач соответствующих классов гладкости. В частности, планируется впервые предложить технику рестартов адаптивного проксимального метода для сильно монотонных вариационных неравенств, а также ускоренный метод для негладкой (гельдеровой) седловой задачи.

Для достижения поставленной цели были предложены следующие **задачи** диссертации:

1. Разработать аналог метода зеркального спуска с переключениями для решения задачи минимизации квазивыпуклых нелипшицевых функций с квазивыпуклыми липшицевыми ограничениями типа неравенств; обосновать соответствующие теоретические оценки скорости сходимости.
2. Расширить применимость методов зеркального спуска на класс относительно липшицевых задач, распространить предложенные модификации метода зеркального спуска для минимизации функций, допускающих представление в абстрактной модельной общности, а также исследовать теоретические характеристики предложенных методов в случае онлайн и стохастической постановки задачи оптимизации относительно липшицевых функционалов.
3. Разработать модификацию алгоритма зеркального спуска для вариационных неравенств с монотонным относительно ограниченным оператором.
4. Предложить технику рестартов адаптивного проксимального зеркального метода для сильно монотонных вариационных неравенств с гельдеровым оператором.
5. Разработать ускоренный алгоритм решения сильно выпукло-вогнутых седловых задач с пониженным уровнем гладкости функционала.

Актуальность темы.

Актуальность данного направления прежде всего обусловлена резким развитием смежных дисциплин, требующих решения задач многомерной оптимизации с минимальными погрешностями. Вопрос оптимизации функционалов большой размерности играет ключевую роль в таких отраслях науки как машинное обучение и анализ данных. Онлайн постановка задачи оптимизации используется в работе с финансовыми рынками, социальными сетями, а также в задачах принятия решений. Введенная концепция относительной липшицевости полезна при решении задач обучения с подкреплением. Вариационные неравенства, в свою очередь, являются основным инструментом решения задач общей экономики, поиска рыночного равновесия, а также комплементарных задач (задач дополнительности).

Основные положения, выносимые на защиту:

1. Был разработан аналог алгоритма зеркального спуска для задач оптимизации нелипшицевых квазивыпуклых функций, при наличии квазивыпуклых липшицевых ограничений типа неравенств.
2. Был предложен вариант метода зеркального спуска для задач выпуклого программирования на классе относительно липшицевых задач, включая задачи онлайн оптимизации, а также задачи в стохастической постановке; были получены теоретические оценки скорости сходимости алгоритмов зеркального спуска для решения задачи оптимизации с функциями, допускающими представление в абстрактной модельной общности.
3. Была предложена модификация алгоритма зеркального спуска для вариационных неравенств с монотонным относительно ограниченным оператором, а также доказана оценка скорости сходимости, которую можно считать оптимальной.
4. Была предложена модификация техники рестартов адаптивного проксимального зеркального метода для вариационных неравенств с сильно монотонным гельдеровым оператором и доказана оценка скорости сходимости, являющаяся оптимальной при $\nu = 0$ и $\nu = 1$.
5. Была описана техника ускорения алгоритма решения сильно выпукло-вогнутых седловых задач с пониженным уровнем гладкости.

Научная новизна:

1. Впервые был предложен аналог метода зеркального спуска с переключениями для задач минимизации квазивыпуклого целевого функционала с квазивыпуклыми липшицевыми ограничениями типа неравенств.
2. Впервые была предложена техника рестартов адаптивного проксимального зеркального метода для сильно монотонных вариационных неравенств с гельдеровыми операторами.

3. Впервые был предложен ускоренный метод для седловой задачи с пониженным уровнем гладкости.

<u>**Достоверность**</u> полученных результатов обусловлена публикацией 12 статей, индексируемых базой Scopus и Web of Science. Ниже приведен список публикаций по материалам диссертации.

Публикации повышенного уровня:

1. Bayandina, A., Dvurechensky, P., Gasnikov, A., Stonyakin, F., Titov, A. Mirror Descent and Convex Optimization Problems with Non-smooth Inequality Constraints. In: Giselsson, P., Rantzer, A. (eds) Large-Scale and Distributed Optimization. Lecture Notes in Mathematics. – 2018. – Т. 2227. – С. 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. – Т. 59. – №. 5. – С. 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. – С. 1-26.
4. Ablaev, S. S., Titov, A. A., Stonyakin, F. S., Alkousa, M. S., Gasnikov, A. Some Adaptive First-Order Methods for Variational Inequalities with Relatively Strongly Monotone Operators and Generalized Smoothness. In: Olenev, N., Evtushenko, Y., Jaćimović, M., Khachay, M., Malkova, V., Pospelov, I. (eds) Optimization and Applications. OPTIMA 2022. Lecture Notes in Computer Science. – 2022. – Т. 13781. – С. 135-150.

Публикации стандартного уровня:

1. Titov, A.A., Stonyakin, F.S., Gasnikov, A.V., Alkousa, M.S. Mirror Descent and Constrained Online Optimization Problems. In: Evtushenko, Y., Jaćimović, M., Khachay, M., Kochetov, Y., Malkova, V., Posypkin, M. (eds) Optimization and Applications. OPTIMA 2018. Communications in Computer and Information Science. – 2019. – Т. 974, – С. 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. In: Khachay, M., Kochetov, Y., Pardalos, P. (eds) Mathematical Optimization Theory and Operations Research. MOTOR 2019. Lecture Notes in Computer Science. – 2019. – T. 11548, – 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 In: Kochetov, Y., Bykadorov, I., Gruzdeva, T. (eds) Mathematical Optimization Theory and Operations Research. MOTOR 2020. Communications in Computer and Information Science. – 2020. – T. 1275. – 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.

Прочие публикации:

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.

Апробация работы. Основные результаты работы докладывались на: следующих конференциях, воркшопах и симпозиумах:

1. Всероссийская научная конференция МФТИ, 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. Международный симпозиум по применению численных методов оптимизации для решения обратных задач, 2021, Москва, Россия.
7. Moscow Conference on Combinatorics and Applications, 2021, Долгопрудный, Россия.
8. International Conference "Optimization without Borders", 2021, Sochi, Russia.

Содержание работы

В первой главе диссертации приведена формальная постановка задачи и предложены некоторые модификации метода зеркального спуска задачи минимизации функций с нестандартными условиями роста. В п. 1.1 представлена общая постановка задачи оптимизации, а также базовые определения, используемые в диссертации. Важным фокусом диссертации, как было отмечено ранее, является отказ от работы с классическим расстоянием в пользу расстояния в более общем смысле. Введем некоторые базовые понятия, касающиеся так называемого расстояния Брэгмана.

Пусть $(E, ||\cdot||)$ — некоторое нормированное конечномерное векторное пространство, а E^* — пространство непрерывных линейных функционалов, определенных на E — его сопряженное. Зададим норму сопряженного пространства следующим образом:

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

где под $\langle y,x \rangle$ будем обозначать значение непрерывного линейного функционала y в точке $x \in E$. Рассмотрим выпуклое компактное подмножество $X \subset E$, а также две выпуклые субдифференцируемые функции $f(x) : X \to \mathbb{R}$ и $g(x) : X \to \mathbb{R}$.

Определение 1. *Пусть $d(x) : X \to \mathbb{R}^+$ — некоторая непрерывно дифференцируемая и 1-сильно выпуклая функция относительно нормы $\|\cdot\|$, то есть*

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

Будем называть функцию $d(x)$ прокс-функцией или функцией, порождающей расстояние (distance generating function).

Определение 2. *Будем говорить, что $V_d(y, x) = V(y, x)$ — расстояние Брэгмана, порожденное прокс-функцией $d(\cdot)$, если выполнено следующее равенство*

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

В п. 1.2 рассматривается адаптивная модификация метода зеркального спуска [18] для решения задачи оптимизации с функциональным ограничением:

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

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

со скоростью сходимости (количество итераций, достаточное для получения ε–точности рассматриваемой задачи) $O\left(\frac{1}{\varepsilon^2}\right)$, в предположении, что целевая функция и функциональное ограничение удовлетворяют условию Липшица, то есть для всех $x, y \in X$ выполнены следующие неравенства:

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

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

Здесь и далее под ε будем понимать точность решения рассматриваемой задачи.

Определение 3. *Будем говорить, что точка z является ε-точным решением задачи (7-8), если выполнены следующие неравенства:*

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

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

В случае, если целевая функция $f(x)$ не удовлетворяет условию Липшица, но имеет липшицев градиент, рассматривается соответствующая модификация метода зеркального спуска [18] с аналогичной скоростью сходимости. Естественным примером возникновения постановки задачи с таким классом гладкости являются квадратичные функции.

В п. 1.3 рассматривается случай невыпуклых функций, в том числе, как одновременная квазивыпуклость целевой функции и функционального ограничения, так и случай квазивыпуклости лишь целевого функционала. Скорость сходимости предложенных методов также составляет $O\left(\frac{1}{\varepsilon^2}\right)$. Методы минимизации квазивыпуклых функций нашли массу применений во многих прикладных задачах, среди которых в рамках диссертации была рассмотрена задача поиска внутренней нормы доходности.

Определение 4 ([10]). *Функция $f : X \to \mathbb{R}$ называется квазивыпуклой, если выполнено следующее неравенство:*

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

При работе с квазивыпуклыми функциями вместо классического (суб)градиента нередко рассматривается следующее множество [25]

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

Здесь и далее будем понимать под $Df(x)$ произвольный вектор из $\hat{D}f(x)$:

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

Для заданной функции $f(x)$ и каждого субградиента $\nabla f(x)$ в точке $y \in X$ определим следующую функцию, которая будет использоваться для характеристики сложности Алгоритма 1:

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

Введем следующее понятие проксимального оператора.

Определение 5. *Для всякого $x \in X$ и $p \in E^*$ определим проксимальный оператор* $\mathrm{Mirr}_x(p)$ *следующим образом:*

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

Алгоритм 1 Модификация адаптивного зеркального спуска для квазивыпуклых функций

Require: $\varepsilon > 0$; Θ_0, такая что $d(x_*) \leq \Theta_0^2$, C_f,C_g

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

Определим $I = \emptyset$, $k = 0$

repeat

  if $g(x_k) \leq \varepsilon M_g$ **then**

    $h_k^f = \frac{C_f}{\|Df(x_k)\|_*}$

    $x_{k+1} = \mathrm{Mirr}_{x_k}\Big(h_k^f Df(x_k)\Big)$ "продуктивный шаг"

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

  else

    $h_k^g = \frac{C_g}{\|Dg(x_k)\|_*}$

    $x_{k+1} = \mathrm{Mirr}_{x_k}\Big(h_k^g Dg(x_k)\Big)$ "непродуктивный шаг"

  end if

  $k = k + 1$

until $\frac{2\Theta_0^2}{\varepsilon^2} \leq N$

Ensure: $\widetilde{x} := \arg\min\limits_{i \in I} f(x_i)$

Теорема 1. *Пусть $f(x)$ — квазивыпуклая функция, $g(x)$ — квазивыпуклая функция, удовлетворяющая условию Липшица с константой M_g. Тогда после $N = \left\lceil \frac{2\Theta_0^2}{\varepsilon^2} \right\rceil$ шагов работы Алгоритма 1 выполнены следующие неравенства:*

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

<u>Вторая глава</u> прежде всего посвящена обобщению условия Липшица (2) в случае замены нормы разности на расстояние в некотором обобщенном смысле, точнее — на дивергенцию Брэгмана.

Пункт 2.1 посвящен мотивации в рассмотрении классов относительно гладких и относительно липшицевых функций. Важной особенностью данных концепция является ослабление требований, предъявляемых к прокс-функции (1), а именно замена условия 1–сильной выпуклости на обычную выпуклость. В диссертации описывается, как следующее определение относительной гладкости функции нашло применение в решении задачи оптимального проектирования.

Определение 6 ([16]). *Будем говорить, что функция $f(x)$ удовлетворяет условию относительной гладкости с константой L (или является L-относительно гладкой), если для всех $x, y \in X$ выполнено следующее неравенство:*

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

Основным же фокусом главы 2 являются относительно липшицевые функции, которые позволили по новому посмотреть на многие известные прикладные задачи, среди которых были рассмотрены метод опорных векторов для задачи бинарной классификации и задача поиска общей точки эллипсоидов.

Определение 7 ([15]). *Будем говорить, что функция $f(x)$ удовлетворяет условию относительной липшицевости с константой M_f (или является M_f-относительно липшицевой), если для всех $x, y \in X$ выполнено следующее неравенство:*

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

Пример 1 (Метод опорных векторов). *Рассмотрим оптимизационную постановку задачи бинарной классификации, решаемой методом опорных векторов с l_2- регуляризацией [27; 29]:*

$$f(x) := \frac{1}{n} \sum_{i=1}^{n} f_j(x) \to \min_x, \tag{21}$$

$$f_j(x) := \max\left\{0{,}1 - y_i x^T w_i\right\} + \frac{\lambda}{2}\|x\|_2^2,$$

где w_i — вектор признаков элемента выборки, а $y_i \in \{-1{,}1\}$ — метка класса. Очевидно, что рассматриваемая функция $f(x)$ не является ни диффе-

ренцируемой, ни липшицевой (в силу регуляризации), таким образом, использование классических (суб)градиентных методов для решения задачи (21) *затруднительно. В [15] показано, что рассматриваемая функция $f(x)$ является 1-относительно липшицевой в случае рассмотрения следующей прокс-функции:*

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

Таким образом, используя подход со стохастической аппроксимацией (суб)градиента целевой функции $f(x)$ и прокс-функцию вида (22), *классический алгоритм зеркального спуска в стохастической постановке:*

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

гарантирует точку $\tilde{x}$, являющейся стохастическим ε-точным решением задачи (21)*:*

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

$$g(z) \le \varepsilon. \tag{25}$$

Более того, число шагов работы алгоритма составляет $O(\frac{1}{\varepsilon^2})$. Заметим, что $\nabla f(x)$ в (23) *является стохастическим (суб)градиентом, удовлетворяющим*

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

и

$$\|\nabla f(x,\xi)\|_* \le M_f, \ \ \|\nabla g(x,\zeta)\|_* \le M_g \quad \textit{почти наверное}. \tag{27}$$

Пример 2 (Задача о поиске общей точки n заданных эллипсоидов)**.** *Рассмотрим n эллипсоидов, каждый из которых задается следующим образом:*

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

где $A_i \in \mathbb{S}^m_{++}$, $i = 1, \ldots, n$. Задача заключается в поиске такой точки $x \in \mathbb{R}^m$, что

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

Стоит отметить, что распространенные для решения подобных задач методы внутренней точки применимы только в случае относительно небольшой размерности m, n. Рассмотрим задачу о поиске пересечения эллипсоидов в следующем виде минимизации функции:

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

Рассматриваемая функция $f(x)$ не является ни дифференцируемой, ни липшицевой. Пусть $\sigma := \max\limits_{0 \le i \le n} \|A_i\|_2^2$, где $\|A_i\|_2$ — спектральный радиус A_i; $\rho := 2 \max\limits_{0 \le i \le n} \|A_i b_i\|_2$, $\gamma := \max\limits_{0 \le i \le n} \|b_i\|_2^2$.

В [15] показано, что рассматриваемая функция $f(x)$ является 1-относительно липшицевой в случае рассмотрения следующей прокс-функции:

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

Более того, классический алгоритм зеркального спуска (23) (с использованием обычного (суб)градиента $\nabla f(x)$ вместо стохастического) гарантирует ε-точное решение задачи (30) за $O(\frac{1}{\varepsilon^2})$ итераций работы алгоритма.

Таким образом, уже в рамках концепции относительной липшицевости функций в пункте 2.2 предлагаются различные адаптивные модификации метода зеркального спуска. Важной особенностью предложенных методов является допущение представления функции в некоторой абстрактной общности, являющейся естественным обобщением предложенной концепции (δ, L)–модели функции [4]. Также рассматриваются обобщения предложенных методов в случае нескольких функциональных ограничений. Более точно, предполагается, что функции допускают следующее представление в модельной общности.

Определение 8. *Пусть $\delta > 0$. Будем говорить, что функции $f(x)$ и $g(x)$ допускают относительно липшицевую (δ, ϕ, V)–модель в точке $y \in X$, если*

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

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

где $\psi_f(y,x)$ и $\psi_g(y,x)$ являются выпуклыми функциями по первой переменной и $\psi_f(x,x) = \psi_g(x,x) = 0$ для всех $x \in X$. При этом функции $\phi_f(\cdot)$, $\phi_g(\cdot)$ могут пониматься как некоторые характеристики вертикальной структуры функций $f(x)$ и $g(x)$, соответственно.

Мотивируем введенное Определение 8 следующими примерами [26], опуская при этом постановку задачи в модельной общности, то есть положим

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

Пример 3. *Пусть прокс-функция $d(x)$ вновь является 1-сильно выпуклой, а (суб)градиент $f(x)$ ограничен ($f(x)$ является M_f-липшицевой). При таких предположениях известно, что дивергенция Брэгмана удовлетворяет следующему неравенству:*

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

Тогда:

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

таким образом, можно рассмотреть

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

Пример 4. *Рассмотрим задачу максимизации положительной вогнутой функции $q(x) : X \to \mathbb{R}_+$:*

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

Тогда функция $f(x) := -\log q(x)$ *будет удовлетворять неравенству* (36) *с* $M_f = 1$. *При этом задачу максимизации* (38) *можно решить путем стандартной минимизации функции* $f(x)$.

Пример 5 (Задача композитной оптимизации [2; 16; 23])**.** *Предложенный далее метод, а также его модификации, рассмотренные в диссертации, применимы для задачи композитной оптимизации следующего вида:*

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

где $r(x), \eta(x) : Q \to \mathbb{R}$ *простые выпуклые функции. В таком случае для всех* $x, y \in X$ *верно следующее:*

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

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

Определим проксимальный оператор для шага $h > 0$ следующим образом:

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

Теорема 2. *Пусть* $f(x)$ *и* $g(x)$ *— выпуклые функции, допускающие представление в модельной общности* (32), (33), *соответственно,* $\varepsilon > 0, \delta > 0$. *Как и ранее, предположим, что существует такая константа* $\Theta_0 > 0$, *что* $d(x_*) \leq \Theta_0^2$. *Предположим, что на определенном шаге работы Алгоритма 2 был выполнен критерий остановки, тогда верно следующее неравенство:*

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

Далее, рассматриваются различные варианты уточнения вида модели функции и предлагаются оптимальные алгоритмы решения задач соответствующих классов гладкости.

Алгоритм 2 Метод зеркального спуска в модельной общности.

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

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

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

$k = 0$

repeat

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

    $x_{k+1} = \mathrm{Mirr}_{h^f}\left(x_k, \psi_f\right)$ "продуктивный шаг"

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

  else

    $x_{k+1} = \mathrm{Mirr}_{h^g}\left(x_k, \psi_g\right)$ "непродуктивный шаг"

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

  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: $\widetilde{x} := \frac{1}{|I|} \sum\limits_{k \in I} x_k$

П. 2.3 посвящен задаче онлайн-оптимизации с функциональным ограничением следующего вида:

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

При этом были предложены алгоритмы решения задачи как в классической постановке, так и в случае, если функции допускают представление в модельной общности. Предложенные методы также являются оптимальными [9] с точки зрения того, что количество непродуктивных шагов соизмеримо с общим числом итераций алгоритмов N. Также в диссертации был проанализирован случай отрицательного регрета и получены соответствующие теоретические оценки количества непродуктивных шагов в данном случае.

П. 2.4 посвящен стохастической постановке задачи оптимизации с сохранением предположений о гладкости целевой функции и функционального ограничения, более того, предложенные методы также имеют оптимальные оценки скорости сходимости $O\left(\frac{1}{\varepsilon^2}\right)$.

В п. 2.5 рассматривается задача минимизации относительно сильно выпуклой

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

целевой функции с функциональным ограничением и предлагается процедура рестартов введенного ранее алгоритма зеркального спуска со скоростью сходимости $O\left(\frac{M^2}{\mu\varepsilon}\right)$, где M — максимальная константа относительной липшицевости целевой функции M_f и функционального ограничения M_g:

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

<u>В главе 3</u> рассматривается задача поиска решения вариационного неравенства для некоторого оператора $g : Q \longrightarrow \mathbb{R}^n$, где Q — некоторое выпуклое замкнутое подмножество $\mathbb{R}^n$:

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

а также седловой задачи

$$f^* = \min_x \max_y f(x,y), \tag{48}$$

где $f(x,y) : Q_x \times Q_y \to \mathbb{R}$ — выпуклая по x и вогнутая по y функция, $Q_x \subset E_1$, $Q_y \subset E_2$ — выпуклые компактные подмножества некоторых нормированных конечномерных векторных пространств с заданными нормами $\| \cdot \|_1$, $\| \cdot \|_2$, соответственно.

В п. 3.1 рассматривается адаптивный проксимальный метод решения вариационных неравенств с липшицевым оператором, гарантирующий ε–точное решение за не более, чем $O\left(\frac{1}{\varepsilon}\right)$ итераций, что является оптимальной оценкой.

В пункте 3.2 предлагается аналог метода зеркального спуска для вариационных неравенств с относительно ограниченным

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

и монотонным оператором:

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

гарантирующий ε–точное решение задачи (47) за $O\left(\frac{1}{\varepsilon^2}\right)$ итераций.

Далее, в п. 3.3 предлагается ускоренный метод решения седловой задачи (48) в предположении, что градиент целевой функции частично удовлетворяет условию Гельдера, $\nu \in [0,1]$, при этом по одной переменной является гладким (градиент удовлетворяет условию Липшица):

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

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

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

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

При этом количество итераций для достижения ε–точного решения составляет

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

где

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

где D — диаметр множества определения $f(x,\cdot)$.

В п. 3.4 впервые предлагается техника рестартов для вариационных неравенств с гельдеровым сильно монотонным оператором:

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

при этом допускается, что вычисление оператора, удовлетворяющего условию Гельдера с константой L_ν, допустимо с некоторыми погрешностями. Скорость сходимости алгоритма при этом составляет

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

где R_0, Ω являются некоторыми характеристиками рассматриваемого пространства. Стоит отметить, что для $\nu = 0$ скорости сходимости предложенного алгоритма и ускоренного метода п.3.3 совпадают, в то время как для $\nu > 0$ асимптотическая скорость сходимость ускоренного метода п. 3.3 является лучшей.

Определение 9. *Предположим, что для некоторого $\delta_u > 0$ (неконтролируемая ошибка) и для любого $\delta_c > 0$ (контролируемая ошибка) существует такая константа $L(\delta_c) \in (0, +\infty)$, что $\forall x,y \in Q$ мы можем посчитать такие $\tilde{g}(x,\delta_c,\delta_u)$ и $\tilde{g}(y,\delta_c,\delta_u) \in E^*$, что выполнены неравенства:*

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

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

Тогда будем называть оператор $\tilde{g}(\cdot,\delta_c,\delta_u)$ неточным оракулом g.

Теорема 3 ([7]). *Предположим, что $g(\cdot)$ и $\tilde{g}(\cdot,\delta_c,\delta_u)$ удовлетворяют (59) и (60). Тогда, $\forall k \geq 1$ и любого $u \in Q$:*

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

Более того, общее количество вызовов оракула не превосходит

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

Алгоритм 3 Адаптивный проксимальный метод для вариационных неравенств в концепции неточного оракула

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

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

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

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

4: **repeat**

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

6: Вычислить

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

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

7: $i_k = i_k + 1$

8: **until**

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

9: $k = k + 1$

10: **end for**

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

Теорема 4. *Предположим, что оператор $g(x)$ является $\mu > 0$–сильно монотонным. Также, предположим, что прокс-функция $d(x)$ удовлетворяет $d(x) \leq \frac{\Omega}{2} \quad \forall x \in Q : \ \|x\| \leq 1$, а начальная точка $x_0 \in Q$ и $R_0 > 0$ такие, что $\|x_0 - x_*\|^2 \leq R_0^2$. Тогда для $p \geq 0$, последовательность x_p, генерируемая Алгоритмом 4, удовлетворяет*

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

а точка x_p, являющаяся результатом работы Алгоритма 4, удовлетворяет

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

В **заключении** приведены основные результаты работы, которые заключаются в следующем.

Алгоритм 4 Техника рестартов адаптивного проксимального метода для вариационных неравенств в концепции неточного оракула

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} — результат работы Алгоритма 3 с точностью $\frac{\mu\varepsilon}{2}$, δ_u, δ_{pu}, прокс функцией $d_p(\cdot)$ и критерием остановки $\sum_{i=0}^{k-1} M_i^{-1} \geq \frac{\Omega}{\mu}$
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.

В рамках диссертации впервые были предложены:

1. аналог метода зеркального спуска с переключениями для задач минимизации квазивыпуклого целевого функционала с квазивыпуклыми липшицевыми ограничениями типа неравенств;
2. техника рестартов адаптивного проксимального зеркального метода для сильно монотонных вариационных неравенств с гельдеровыми операторами;
3. ускоренный метод для (гельдеровой) седловой задачи с пониженным уровнем гладкости.

В рамках поставленных задач, был разработан аналог метода зеркального спуска с переключениями для решения задачи минимизации квазивыпуклых нелипшицевых функций с квазивыпуклыми липшицевыми ограничениями типа неравенств со скоростью сходимости $O\left(\frac{1}{\varepsilon^2}\right)$. Был предложен алгоритм решения задачи минимизации квазивыпуклой целевой функции, не удовлетворяющей условию Липшица, но при этом имеющей липшицев градиент, с квазивыпуклым ограничением. Были предложены оптимальные методы зеркального спуска для относительно липшицевой задачи оптимизации с функциональным ограничением в случае онлайн и стохастической постановки задачи. При этом была рассмотрена концепция модельной общности функции соответствующего класса гладкости. Были предложены соответствующие модификации метода зеркального спуска для относительно липшицевых задач в случае стохастической постановки задачи оптимизации.

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. — Т. 15, № 1. — С. 229—251.

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

24. *Nesterov Y.* Universal gradient methods for convex optimization problems // Mathematical Programming. — 2015. — Т. 152, № 1. — С. 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. Т. 4. — 2019.

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

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

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

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

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

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

33. *Поляк Б. Т.* Введение в оптимизацию // Наука. — 1983.