Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»
Физтех-школа прикладной математики и информатики
Кафедра дискретной математики

На правах рукописи

УДК 519.85

СТОНЯКИН ФЕДОР СЕРГЕЕВИЧ

АДАПТИВНЫЕ МЕТОДЫ ДЛЯ ВАРИАЦИОННЫХ НЕРАВЕНСТВ И ЗАДАЧ МИНИМИЗАЦИИ ФУНКЦИОНАЛОВ С ОБОБЩЁННЫМИ УСЛОВИЯМИ ГЛАДКОСТИ

01.01.07 — вычислительная математика

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
доктора физико-математических наук

Москва — 2020

Работа прошла апробацию на кафедре дискретной математики Федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)».

Научный консультант: доктор физико-математических наук, профессор Орлов Игорь Владимирович

Ведущая организация: Федеральное государственное бюджетное образовательное учреждение высшего образования «Омский государственный технический университет»

Защита состоится «28» декабря 2020 года в 12:00 на заседании диссертационного совета ФПМИ.01.01.07.003 по адресу 141701, Московская область, г. Долгопрудный, Институтский переулок, д. 9.

С диссертацией можно ознакомиться в библиотеке и на сайте Московского физико-технического института (национального исследовательского университета):
https://mipt.ru/education/post-graduate/soiskateli-fiziko-matematicheskie-nauki.php

Работа представлена «18» сентября 2020 г. в Аттестационную комиссию Федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)» для рассмотрения советом по защите диссертаций на соискание учёной степени кандидата наук, доктора наук в соответствии с п. 3.1 ст. 4 Федерального закона «О науке и государственной научно-технической политике».

Актуальность темы. Задачи математического программирования встречаются в самых разных приложениях и весьма актуальна проблема разработки эффективных алгоритмических процедур для их численного решения. Среди этих задач особо можно выделить различные классы задач выпуклой оптимизации. Выпуклый анализ и выпуклая оптимизация основаны на солидном математическом фундаменте, разработанном в основном в первой половине 20-го столетия математиками Г. Минковским, К. Каратеодори, Э. Хелли, В. Фенхелем, А. Александровым и другими.

Многие выпуклые задачи являются негладкими, что привело к необходимости развития методов недифференцируемой оптимизации. Разработки в этой области начались в СССР в 60-е годы прошлого века. Такие методы предназначались, прежде всего, для решения задач линейного программирования большой размерности, возникавших при моделировании важнейших экономических, экологических и других реальных явлений. Для этого в работах отечественных математиков Б. Т. Поляка, В. Ф. Демьянова, Л. В. Васильева и В. Н. Малоземова, Н. З. Шора, Б. Н. Пшеничного были исследованы необходимые условия экстремума некоторых классов недифференцируемых функций, а также предложены итеративные процедуры, обобщающие классические методы градиентного спуска. В США и странах Западной Европы методы недифференцируемой оптимизации начали активно изучаться в 70-е годы прошлого века. При этом отдельные работы в данном направлении, обусловленные преимущественно потребностями развития теории игр (которая используется при моделировании конкуренции в рыночной экономике), встречались и раньше. Наряду с численными алгоритмами, исследовались свойства обобщённых градиентов (или субдифференциалов) и полученные на их основе условий оптимальности в том числе и для невыпуклых функций. Позднее круг приложений негладкого анализа существенно расширился. Широко известны разработки в данной области Р. Рокафеллара, Ф. Кларка, Дж. Варги, Ф. Мишеля и Ж.-П. Пено, Ж.-П. Обена и И. Экланда, Д. Ауссела, А. Д. Иоффе, Дж. Борвейна и К. Жу, а также многих других авторов.

После выхода монографии Р. Т. Рокафеллара[1)] основное направление развития выпуклого анализа сместилось в сторону теории методов оптимизации [2)]. По выпуклому анализу и его оптимизационным приложениям известно довольно много работ и замечательных книг. Основные приоритетные результаты в этой области принадлежат отечественным ученым А. С. Антипину, Ф. П. Васильеву, Е. Г. Гольштейну, В. Ф. Демьянову, Ю. Г. Евтушенко, Ю. М. Ермольеву, А. Д. Иоффе, Б. С. Мордуховичу, А. С. Немировскому, Ю. Е.

[1)] Рокафеллар Р. Т. Выпуклый анализ. — М.: Мир, 1973 — 420 с.

[2)] Иоффе А. Д., Тихомиров В. М. Теория экстремальных задач. — М.: Наука, 1974 — 460 с.

Нестерову, Б. Т. Поляку, Р. А. Поляку, Б. Н. Пшеничному, А. М. Рубинову, В. М. Тихомирову, Л. Г. Хачияну, Н. З. Шору, Д. Б. Юдину и многим другим.

Проблема разработки эффективных алгоритмических процедур для оптимизационных задач с функционалами различного уровня гладкости (в том числе и с нелипшицевыми) довольно актуальна в приложениях. Отметим лишь некоторые примеры негладких задач, которые исследуются в последние годы. Так, негладкие задачи возникают при решении проблем проектирования механических конструкций (Truss Topology Design)[3)4)5)]. Также, вообще говоря, негладкими являются задачи распределения ресурсов[6)], которые сводятся к максимизации совокупной полезности производителей при совместном использовании имеющихся ресурсов. При этом возможны и целевые функционалы, которые имеют более слабый уровень гладкости по сравнению с условием Липшица[7)]. Как частный случай, выделим задачу распределения ресурсов в компьютерных сетях [8)] [9)]. Можно отметить также и некоторые геометрические негладкие оптимизационные задачи типа Ферма-Торричелли-Штейнера [10)], задачи о наименьшем покрывающем шаре, Convex Feasibility Problem [11)].

Наиболее известный подход к пониманию эффективности алгоритмических процедур для задач оптимизации основан на теории сложности, восходящей к известной монографии[12)]. Разработанная А.С. Немировским и Д.Б. Юдиным теория верхних оценок возможной эффективности методов выпуклой минимизации для различных классов задач была подкреплена оптимальными методами, которые реализовывали эти оценки. Для класса методов, у которых на каждой итерации разрешается не более чем $\mathrm{O}(1)$ раз обращаться к оракулу (подпрограмме) для расчёта градиента $\nabla f(x)$, оценка числа итераций N, необходимых для достижения точности решения задачи ε (по функции)

[3)] Shpirko S., Nesterov Y. Primal-dual Subgradient Methods for Huge-scale Linear Conic Problem // SIAM Journal on Optimization, 2014. Vol. 24, no. 3. P. 1444–1457.

[4)] Nesterov Y. Subgradient methods for Huge-Scale Optimization Problems // Math. Prog., 2015. Vol. 146, no. 1–2. P. 275 – 297.

[5)] Lagae S. New efficient techniques to solve sparse structured linear systems, with applications to truss topology optimization. Ecole polytechnique de Louvain, Université catholique de Louvain, 2017.

[6)] Kelly F. P., Maulloo A. K. and Tan D. K. H. Rate control for communication networks: shadow prices, proportional fairness and stability. // J. Oper Res Soc. 1998. Vol. 49 (3). P. 237 – 252.

[7)] Nedic A. and Ozdaglar A. Subgradient Methods in Network Resource Allocation: Rate Analysis. Proc. of CISS, 2008.

[8)] Рохлин Д. Б. Распределение ресурсов в сетях связи с большим числом пользователей: стохастический метод градиентного спуска. // Теория вероятностей и её применения, 2020 (в печати).

[9)] Ivanova A., Stonyakin F., Pasechnyuk D, Vorontsova E., Gasnikov A. Adaptive Mirror Descent for the Network Utility Maximization Problem. `https://arxiv.org/abs/1911.07354v1`, 2020 IFAC Congress.

[10)] Mordukhovich B. S., Nam N. M. Applications of variational analysis to a generalized Fermat–Torricelli problem // J. Optim. Theory Appl. 2011. Vol. 148, no 3. P. 431–454.

[11)] Beck A. First-Order Methods in Optimization // MOS-SIAM Series on Optimization. SIAM. 2017. — 467 p.

[12)] Немировский А. С., Юдин Д. Б. Сложность задач и эффективность методов оптимизации. — М.: Наука, 1977. — 540 с.

Таблица 1: Оптимальные оценки количества обращений к (суб)градиенту.

	$\lvert f(y)-f(x)\rvert \leqslant$ $\leqslant M\lVert y-x\rVert$	$\lVert \nabla f(y)-\nabla f(x)\rVert_* \leqslant$ $\leqslant L\lVert y-x\rVert$
$f(x)$ выпукла	$O\left(\frac{M^2R^2}{\varepsilon^2}\right)$	$O\left(\sqrt{\frac{LR^2}{\varepsilon}}\right)$
$f(x)$ μ-сильно выпукла в $\lVert\cdot\rVert$-норме	$O\left(\frac{M^2}{\mu\varepsilon}\right)$	$O\left(\sqrt{\frac{L}{\mu}}\left\lceil \ln\left(\frac{\mu R^2}{\varepsilon}\right)\right\rceil\right)$ ($\forall\ N$)

$f\left(x^N\right)-\min\limits_{x\in \mathrm{R}^n} f\left(x\right)\leq\varepsilon$, указана в таблице 1 в зависимости от класса рассматриваемых задач (x^N — выход работы метода). Важно, что эти оценки никак не связны с размерностью решаемой задачи.

Отметим, что в таблице 1 $R=\left\lVert x^0-x_*\right\rVert_2$ — расстояние от точки старта до точки-решения задачи x_* ($\lVert\cdot\rVert_2$ — евклидова норма). Если решение не единственно, то под x_* при этом понимается (евклидова) проекция точки x^0 на множество решений задачи. При этом оптимальные оценки достигаются для хорошо известного быстрого градиентного метода Ю.Е. Нестерова [13)].

После этого на некоторый период градиентные методы были практически забыты. Начиная с выдающейся работы Н. Кармаркара[14)] и примерно до 2000 г. развитие теории и методов оптимизации было в основном связано с прогрессом в теории полиномиальных методов внутренней точки. К примеру, была разработана общая теория самосогласованных функций [15)], которая позволяла строить полиномиальные методы внутренней точки для выпуклых задач с явной структурой. Однако трудоёмкость методов внутренней точки существенно возрастает при увеличении размерности решаемой задачи. В связи с указанным обстоятельством в 2000-е годы вновь возник интерес к теории сложности [16)], а также к подкрепляющим её оптимизационным методам градиентного типа. Итерации таких методов требуют меньших затрат памяти, что интересно ввиду приложений к задачам оптимизации с большими данными. В то же время оценки эффективности для задач негладкой оптимизации представляются довольно пессимистичными. Например, сложность задачи минимизации выпуклого липшицева (возможно, и негладкого) функционала достигает $O\left(\varepsilon^{-2}\right)$ итераций градиентного типа, а для задачи минимизации выпуклого функционала с лип-

[13)] Нестеров Ю.Е. Метод минимизации выпуклых функций со скоростью сходимости $O(1/k^2)$ // Доклады АН СССР. 1983. Т. 269. № 3. С. 543–547.

[14)] Karmakar N. A new polynomial time algorithm for linear propramming // Combinatorica. 1984. V. 4, No. 4. P. 373 – 395.

[15)] Nesterov Yu., Nemirovskii A. Interior point polynomial methods in convex programming: Theory and Applications // Philadelphia: SIAM, 1994 — 520 p.

[16)] Немировский А. С., Юдин Д. Б. Сложность задач и эффективность методов оптимизации. — М.: Наука, 1977. — 540 с.

шицевым градиентом — $O\left(\sqrt{\varepsilon^{-1}}\right)$. В этой связи естественно возникает вопрос о том, насколько возможно приблизить скорость сходимости численных методов для задач негладкой оптимизации (и вообще, задач с липшицевыми целевыми функционалами) к гладкому случаю. В этом плане хорошо известен подход, основанный на технике сглаживания [17)]. Однако этот подход существенно связан со специально подобранной структурой рассматриваемых задач, которая позволяет подобрать удачную аппроксимацию негладкой целевой функции некоторым гладким сильно выпуклым аналогом. С другой стороны, даже в случае применимости техника сглаживания может не приводить к удовлетворительным результатам [18)].

Поэтому вполне естественны дальнейшие усилия по нахождению новых подходов к проблеме повышению эффективности работы алгоритмических методов для негладких оптимизационных задач. Среди них можно выделить так называемые *универсальные методы*, исследованию которых было положено начало в работе [19)]. Указанные подходы основаны на построении для задач выпуклой оптимизации с гёльдеровым градиентом (субградиентом при $\nu = 0$) целевого функционала (параметр $\nu \in [0; 1]$ фиксирован)

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L_\nu \|x - y\|^\nu \quad \forall x, y \in Q$$

аналога стандартной квадратичной интерполяции с искусственно введённой неточностью $\delta > 0$

$$f(x) + \langle \nabla f(x), y - x \rangle \leqslant f(y) \leqslant f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|^2 + \delta, \text{ где}$$

$$L = L_\nu \left[\frac{L_\nu}{2\delta} \cdot \frac{1-\nu}{1+\mu}\right]^{\frac{1-\nu}{1+\nu}}.$$

При этом $\nu = 1$ соответствует гладкому случаю (градиент f удовлетворяет свойству Липшица), а $\nu = 0$ — в общем случае негладкому (f удовлетворяет свойству Липшица). Универсальность при этом понимается как возможность адаптивной настройки при работе метода на оптимальный в некотором смысле уровень гладкости задачи и величину соответствующей константы Гёльдера L_ν градиента (субградиента при $\nu = 0$) целевого функционала. Оказывается, что возможность такой настройки может позволить для некоторых задач улучшить скорость сходимости по сравнению с оптимальными теоретическими оценками

---

17) Nesterov Yu. E., Smooth minimization of non-smooth functions // Mathematical Programming. 2005. Vol. 103. P. 127 – 152.

18) Lagae S. New efficient techniques to solve sparse structured linear systems, with applications to truss topology optimization. Ecole polytechnique de Louvain, Université catholique de Louvain, 2017.

19) Nesterov Yu. Universal gradient methods for convex optimization problems. // Math. Program. Vol. 152. No. 1 – 2. P. 381 -– 404.

на классе задач [20]. Как известно, погрешности при решении задач оптимизации возникают в силу разных причин [21]. Они могут быть естественно связанными с неточностью доступных данных, заменой бесконечномерной задачи конечномерным аналогом и т.д. Возможны и искусственные неточности, возникающие в ходе математического исследования рассматриваемых задач. Помимо указанной выше идеологии универсальных методов в этом плане можно отметить и неточности, связанные с регуляризацией задачи, а также со сглаживанием [22].

Поэтому естественно возникает проблема описания влияния погрешностей задания целевого функционала и градиента на теоретические оценки скорости сходимости методов. Для градиентных методов выпуклой оптимизации известен подход к этой проблеме, основанный на недавно введённом понятии неточного оракула [23] [24]. Известно, что для обычного (неускоренного) градиентного метода в оценке скорости сходимости не происходит накопления величин, связанных с погрешностями. Однако для оптимальных при отсутствии погрешностей на классе гладких задач ускоренных методов (например, для быстрого градиентного метода) в итоговой оценке скорости сходимости величины погрешностей могут накапливаться. Отметим подходы к этой проблеме в случае погрешностей неслучайной природы (аппроксимативный градиент) [25], а также — для случайного аддитивного шума (погрешности) [26] [27][28] при задании градиента.

Также весьма важной является задача построения методов для вариационных неравенств и седловых задач с адаптацией как к уровню гладкости оператора, так и к величинам погрешностей. Отметим, что вариационные неравенства имеют широкий спектр приложений в физике, технике, экономике, а в последнее время и в машинном обучении. К примеру, отмечена целесообразность их

[20] Nesterov Yu., Universal gradient methods for convex optimization problems // Math. Program. Series A and B. 2015. Vol. 152, No. 1–2. P. 381-–404.

[21] Поляк Б. Т. Введение в оптимизацию. — М.: Наука, 1983. — 433 с.

[22] Nesterov Yu. E., Smooth minimization of non-smooth functions // Math. Program. 2005. Vol.103, No. 1. P. 127 – 152.

[23] Devolder, O., Glineur, F., Nesterov, Yu. First-order methods of smooth convex optimization with inexact oracle. // Math. Program. 2014. Vol. 146, № 1–2. P. 37 – 75.

[24] Devolder O. Exactness, Inexactness and Stochasticity in First-Order Methods for Large-Scale Convex Optimization. // PhD thesis (2013). https://dial.uclouvain.be/pr/boreal/object/boreal:128257.

[25] D'Aspremont A. Smooth optimization with approximate gradient. // SIAM Journal of Optimization. 2008. Vol. 19, no. 3. P. 1171–1183.

[26] Gorbunov E., Dvinskikh D., Gasnikov A. Optimal Decentralized Distributed Algorithms for Stochastic Convex Optimization. https://arxiv.org/abs/1911.07363v2, 2019.

[27] А. В. Гасников. Современные численные методы оптимизации. Метод универсального градиентного спуска. — М: МФТИ, 2018. https://arxiv.org/ftp/arxiv/papers/1711/1711.00394.pdf

[28] Cohen M. B., Diakonikolas J., Orecchia L. On Acceleration with Noise-Corrupted Gradients. Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, Proceedings of Machine Learning Research 80, 2018.

применения в задачах описания генеративно-состязательных сетей [29]. В качестве аналога гладкой оптимизационной задачи можно рассматривать задачу нахождения решения вариационного неравенства с монотонным липшицевым оператором, а в качестве аналога негладкой оптимизационной задачи — вариационное неравенство с монотонным ограниченным оператором. Оптимальным с точки зрения нижних оценок количества обращений к оракулу оператора поля для указанных классов задач будет экстраградиентный метод [30], а также его более современный вариант — проксимальный зеркальный метод [31]. В работе [32] предложен адаптивный метод для вариационных неравенств со случайной погрешностью задания оператора и поставлена задача разработки метода с адаптивной настройкой на уровень гладкости оператора с исследованием вопросов накопления в итоговой оценке качества найденного решения величины абсолютных помех неслучайной природы при задании оператора вариационного неравенства.

Настоящая диссертационная работа посвящена проблеме построения адаптивных методов для задач выпуклой оптимизации, вариационных неравенств и седловых задач с адаптивной настройкой как на уровень гладкости целевого функционала (оператора), так и на величину абсолютной погрешности. При этом ставится задача обосновать применимость разрабатываемых алгоритмических методов на классе задач со структурой, которая допускает их описание с точки зрения некоторой концепции абстрактной неточной модели оптимизируемой функции подобно [33] [34].

Цель диссертационного исследования — разработка оптимальных алгоритмических методов для задач выпуклой негладкой оптимизации, а также подходов к использованию неточных моделей для вариационных неравенств с монотонными операторами, выпукло-вогнутых седловых задач и задач минимизации

[29] Antonakopoulos, K., Belmega, V., Mertikopoulos, P. An adaptive Mirror-Prox method for variational inequalities with singular operators. — Advances in Neural Information Processing Systems 32, 2019 — pp. 8453–8463.

[30] Г.М. Корпелевич. Экстраградиентный метод для отыскания седловых точек и других задач, // Экономика и матем. методы. 1976. Т. 12. № 4. С. 747–756.

[31] Nemirovski A. Prox-method with rate of convergence O(1/t) for variational inequalities with Lipschitz-continuous monotone operators and smooth convex-concave saddle point problems. SIAM Journal on Optimization. 2004. Vol. 15(1). P. 229 – 251.

[32] Bach F. and Levy K. Y. A universal algorithm for variational inequalities adaptive to smoothness and noise. In COLT'19: Proceedings of the 32nd Annual Conference on Learning Theory, 2019.

[33] Tyurin A. I., Gasnikov A. V. Fast gradient descent method for convex optimization problems with an oracle that generates a (δ, L)-model of a function in a requested point. // Computational Mathematics and Mathematical Physics. 2019. Vol.59, No. 7. P. 1137 – 1150.

[34] Stonyakin, F. S., Dvinskikh, D., Dvurechensky, P., Kroshnin, A., Kuznetsova, O., Agafonov, A., Gasnikov, A., Tyurin, A., Uribe, C. A., Pasechnyuk, D., Artamonov, S. Gradient Methods for Problems with Inexact Model of the Objective. // In: Khachay M., Kochetov Y., Pardalos P. (eds) Mathematical Optimization Theory and Operations Research. MOTOR 2019. Lecture Notes in Computer Science. Springer, Cham. 2019. Vol. 11548. P. 97 – 114.

функционалов, которые позволили бы предложить алгоритмические методы с приемлемыми вычислительными гарантиями и адаптацией оценок качества решения к уровню гладкости задачи, а также величинам погрешностей.

Можно выделить такие **задачи** диссертационного исследования:

1. Распространение идеологии универсальных градиентных методов на существенно более широкие классы задач, а именно — на вариационные неравенства и седловые задачи. Разработка концепций абстрактных неточных оптимизационных моделей (аналоги (δ, L)-оракула и (δ, L)-модели функции) для вариационных неравенств и седловых задач, которые позволяют получить приемлемые оценки скорости сходимости с учётом естественных и некоторых типов искусственно введённых погрешностей для таких задач.

2. Исследование новых подходов к построению неточной оптимизационной модели целевого функционала для минимизационных задач с раздельным учётом абсолютных погрешностей задания целевой функции и градиента (субградиента). Разработка для соответствующих классов задач алгоритмических методов, которые могли бы позволить более гибко учитывать степень негладкости задачи, а также возникающие естественные и искусственно введённые неточности с сохранением приемлемых вычислительных гарантий.

3. Описание условий на целевой функционал оптимизационной задачи, которые позволяют гарантировать линейную скорость сходимости метода с точностью до величин, соответствующих погрешностям.

4. Исследование возможности распространения указанных подходов с сохранением оптимальных вычислительных гарантий на некоторые классы невыпуклых оптимизационных задач.

5. Разработка адаптивных алгоритмических схем зеркального спуска для задач негладкого выпуклого программирования с возможностью обоснования оптимальных вычислительных гарантий на классах задач с нелипшицевыми целевыми функционалами, возникающих в приложениях.

Теоретические исследования по обоснованию необходимых результатов и оценок скорости сходимости разработанных методов были выполнены с использованием **методов** математического анализа, выпуклого и функционального анализа. Тестирование предложенных алгоритмов выполнено с использованием компьютерных программ в среде CPython 3.7.

Сформулируем основные положения диссертации, которые выносятся на защиту:

1. Предложен единообразный подход к построению оптимизационной модели для вариационных неравенств с ν-гёльдеровыми монотонными операторами (при всех $\nu \in [0; 1]$), который позволил разработать универсальный метод для вариационных неравенств и выпукло-вогнутых седловых задач соответствую-

щего уровня гладкости. Получены оценки скорости сходимости этого метода, указывающие на оптимальность предложенного подхода как на классе вариационных неравенств с липшицевыми операторами ($\nu = 1$), так и на классе вариационных неравенств с ограниченными операторами ($\nu = 0$).

2. Введено новое понятие абстрактной неточной модели (аналог неточного оракула) для вариационных неравенств и седловых задач. Соответственно, выделен новый класс вариационных неравенств и седловых задач, которые допускают в произвольной запрашиваемой точке существование такой модели, причём с условиями типа относительной гладкости. Для этого класса задач предложен адаптивный вариант проксимального зеркального метода, получена оценка его скорости сходимости и доказано, что в ней не накапливаются определяемые погрешностями величины. Тем самым, существенно расширены границы применимости методов экстраградиентного типа, для которых можно с уверенностью утверждать сохранение оптимальных оценок скорости сходимости или близких к оптимальным в случае погрешностей. Обоснована применимость данного метода к популярным для вопросов обработки изображений композитным седловым задачам.

3. С использованием усреднения адаптивно подбираемых констант Липшица оператора на итерациях предложен адаптивный метод для вариационных неравенств с липшицевыми сильно монотонными операторами с гарантией линейной скорости сходимости.

4. Предложен новый подход к понятию абстрактной неточной оптимизационной модели для градиентных методов оптимизации, показана его применимость к достаточно широкому выделенному классу задач негладкой оптимизации. Особенность предложенного подхода — раздельный учёт влияния погрешностей разного типа путём введения несколькими параметров погрешностей. Предложен метод градиентного типа с адаптивной настройкой в оценке скорости сходимости некоторых из этих параметров и показано, что такая адаптивная настройка может повышать качество найденного решения. Доказано, что величины, связанные со всеми типами погрешностей, не накапливаются в итоговых оценках для неускоренных методов. Предложена дополнительная процедура для ускоренного метода, гарантирующая отсутствие в определённому смысле накопления одного из типов погрешностей. Обоснована применимость разработанной методики на некотором классе негладких задач оптимизации с вычислительными гарантиями, близкими к оптимальным (с точностью до умножения на логарифмический множитель).

5. Введено новое понятие неточной модели целевой функции с несколькими параметрами, соответствующими свойствам сильной выпуклости, гладкости, а также с раздельным учётом погрешностей задания целевого функционала и

градиента. Обоснована близкая к линейной скорость сходимости предложенных градиентных методов с адаптивным выбором шага для задач минимизации функционалов, которые допускают существование указанного типа оптимизационной модели в произвольной запрошенной точке. Разработан адаптивный градиентный метод и обоснована близкая к линейной скорость его сходимости в случае, если вместо сильной выпуклости целевого функционала относительно евклидовой нормы выполняется условие градиентного доминирования.

6. Комбинированием метода дихотомии с неточным решением одномерной двойственной задачи и методов градиентного типа для вспомогательных многомерных задач предложен подход к задачам выпуклого программирования с одним функциональными ограничением. Методика основана на введённом в работе адаптивном критерии остановки, который учитывает неточность решения вспомогательных задач. На базе полученной оценки скорости сходимости для аналога метода дихотомии на квадрате с учётом погрешности нахождении градиента предложен подход к задаче выпуклого программирования с двумя функциональными ограничениями. Показана оптимальность предложенной методики для достаточно гладких задач с сильно выпуклым целевым функционалом как в случае одного, так и двух функционалов ограничений.

7. Для зеркальных спусков с переключениями по продуктивным и непродуктивным шагам введены новые адаптивные критерии остановки, выполнение которых гарантирует достижение приемлемого качества решения задачи выпуклого программирования вне зависимости от уровня гладкости целевого функционалам. Обоснована оптимальность оценок скорости сходимости для некоторых типов целевых функционалов, не удовлетворяющих условию Липшица. В частности доказано, что на классе задач с гёльдеровыми выпуклыми целевыми функционалами сохранится оценка сложности $O(\varepsilon^{-2})$, оптимальная даже на более узком классе липшицевых выпуклых целевых функционалов. Рассмотрено приложение к задаче оптимизации высоконагруженной компьютерной сети. Также на базе разработанных методов зеркального спуска предложены методы для минимизации с относительной точностью выпуклого однородного функционала и выпуклыми функционалами ограничений. Эти методы гарантируют достижение заданной относительной точности приближённого решения задачи за оптимальное число итераций при существенно более общих предположениях в сравнении с известными аналогичными результатами (0 — не обязательно внутренняя точка субдифференциала целевой функции в нулевой точке).

8. С помощью техники рестартов (перезапусков) указанных в п. 7 адаптивных зеркальных спусков впервые разработаны методы для задач сильно выпуклой оптимизации с сильно выпуклым липшицевым функционалом ограничения. Впервые получены оптимальные оценки скорости сходимости на некото-

рых классах задач с негладкими целевыми функционалами max-типа, которые не удовлетворяю условию Липшица.

9. Обоснована применимость некоторых из разработанных адаптивных зеркальных спусков с переключениями к задачам минимизации произвольного квазивыпуклого субдифференцируемого по Кларку целевого функционала с ненулевым субградиентом в любой точке и выпуклым функциональным ограничением. При этом показано, что сохраняются оценки скорости сходимости методов, которые оптимальны для более узкого класса задач выпуклого программирования с целевыми функционалами соответствующего уровня гладкости.

10. Предложен новый адаптивный метод зеркального спуска для задач онлайн-оптимизации в случае выпуклых (возможно, негладких) липшицевых целевых функционалов и нескольких выпуклых липшицевых функциональных ограничений. Обоснована оптимальность метода в терминах нижних оракульных оценок на рассмотренном классе задач.

К наиболее важным результатам, по нашему мнению, стоит отнести результаты пп. 1, 2, 4, 5, 7 и 8. В частности, результаты пп. 2, 4 и 5 о модельной общности для вариационных неравенств и минимизационных задач приводят к существенному расширению границ применимости методов градиентного типа с сохранением приемлемых вычислительных гарантий. Результаты пп. 7 и 8 указывают на существенное расширение класса задач (преимущественно, негладкого) выпуклого программирования с не обязательно ограниченными субградиентами, для которых сохраняются заведомо оптимальные вычислительные гарантии.

Научная новизна, степень достоверности и апробация результатов диссертации

Все основные результаты работы являются новыми и строго математически обоснованными. По тематике диссертации опубликована 22 основные печатные работы, из которых 8 вышли без соавторов. Из указанных работ 21 опубликованы в изданиях, индексируемых в базе Scopus или Web of Science и 1 статья опубликованы в журнале из перечня ВАК РФ, который индексируется в базе RSCI. Из совместных работ в диссертацию включены только результаты, которые принадлежат соискателю.

По материалам диссертационной работы автор выступал с 90-минутными докладами на следующих научных семинарах и конференциях:

- Общемосковский постоянный научный семинар «Теория автоматического управления и оптимизации». Институт проблем управления им. В. А. Трапезникова РАН под руководством профессора Б.Т. Поляка.

- Научный семинар кафедры общих проблем управления механико-математического факультета МГУ имени М.В. Ломоносова «Задачи дифференциальных уравнений, анализа и управления: теория и приложения» под руководством член-корреспондента РАН М.И. Зеликина, член-корреспондента РАН В.Ю. Протасова, профессора В.М. Тихомирова и профессора А.В. Фурсикова.

- Научный семинар кафедры дискретной математики МФТИ под руководством профессора А.М. Райгородского.

- Научный семинар кафедры алгебры и функционального анализа Крымского федерального университета им. В.И. Вернадского под руководством профессора И.В. Орлова.

- Научный семинар кафедры математического анализа Российского университета дружбы народов под руководством профессора В.И. Буренкова.

- Научный семинар "Конструктивный негладкий анализ и недифференцируемая оптимизация"под руководством профессора В.Н. Малозёмова.

- Традиционные школы «Информация. Управление. Оптимизация» 2018 и 2019 гг.

- IX Международная молодежная научно-практическая конференция с элементами научной школы "Прикладная математика и фундаментальная информатика, посвящённая 80-летию со дня рождения академика РАН Ю.Г. Евтушенко, Омск, апрель 2019 г.

По материалам диссертационной работы были секционные доклады, в частности, на следующих научных конференциях:

- 23 Симпозиум по математическому программированию, Бордо, Франция, июль 2018 г.

- International IFAC Congress, Германия, Берлин, июль 2020 г.

- Conference on graphs, networks, and their applications. МФТИ, г. Долгопрудный, май 2019 г.

- International conferences «Mathematical Optimization Theory and Operations Research», июль 2019 и 2020 гг.

- IX и X Международные конференции «Optimization and applications», Петровац, Черногория, 2018 и 2019 гг.

- IV и V Международные конференции «Квазилинейные уравнения и обратные задачи», Долгопрудный, МФТИ, декабрь 2018 и 2019 гг.

- International Workshop «Optimization at work», Долгопрудный, МФТИ, октябрь 2017 г.

Структура и основное содержание диссертации

Диссертационная работа состоит из введения, 5 глав, заключения и списка литературы. Объём работы составляет 417 страниц. Список цитируемых источников состоит из 178 наименований. Нумерация упоминаемых в автореферате утверждений и алгоритмов совпадает с нумерацией их в диссертации.

Во **введении к диссертации** указана актуальность темы исследования и степень ее разработанности, а также сформулированы основные результаты, которые выносятся на защиту.

Глава 1 диссертации является вводной и содержит обзор литературных источников по теме, а также набор необходимых вспомогательных результатов. В частности, для оценки качества решения оптимизационных задач вводится широко используемый в оптимизации аналог расстояния между точками x и y, который называют *расхождением* или *дивергенцией Брэгмана*. Обычно *дивергенция Брэгмана* вводится с использованием вспомогательной 1-сильно выпуклой функции d (порождает расстояния), которая дифференцируема во всех точках $x \in Q$:

$$V(y,x) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \quad \forall x, y \in Q,$$

где $\langle \cdot, \cdot \rangle$ — скалярное произведение в $\mathbb{R}^n$. В частности, для стандартной евклидовой нормы $\| \cdot \|_2$ и расстояния в $\mathbb{R}^n$ можно считать, что $V(y,x) = d(y-x) = \frac{1}{2}\|y-x\|_2^2$ для произвольных $x, y \in Q$. Приводятся сведения о недавно введённом условии относительной гладкости[35)] для оптимизационных задач. Такой класс задач связан с заменой стандартного условия Липщица градиента f на ослабленный вариант

$$f(x) + \langle \nabla f(x), y - x \rangle \leqslant f(y) \leqslant f(x) + \langle \nabla f(x), y - x \rangle + LV(y,x)$$

и предполагает лишь выпуклость (но не сильную выпуклость) порождающей функции d. Условие относительной гладкости позволяет применить вариант градиентного метода для некоторых задач, которые ранее решались лишь с помощью методов внутренней точки. В частности, речь идет об известной задаче

[35)] H. H. Bauschke, J. Bolte, and M. Teboulle. A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications. // Math. Oper. Res. 2017. Vol. 42. P. 330 -– 348.

построения оптимального эллипсоида, покрывающего заданный набор точек, которая имеет приложения в статистике и анализе данных[36]. Далее, описаны аналогичные понятие относительной сильной выпуклости и относительной липшицевости (ограниченности субградиента) для минимизационных задач. Подробно обсуждается понятия неточного оракула и неточной модели [37] целевой функции. Приведено немало примеров прикладных задач, для которых целесообразно использовать понятия неточного оракула и неточной модели, относительной гладкости, относительной сильной выпуклости и относительной липшицевости целевой функции. Отметим, что настоящая диссертационная работа посвящена, в частности, разработке методов с приемлемыми вычислительными гарантиями для задач минимизации целевых функционалов указанных выше классов при наличии функциональных ограничений (что в общем случае усложняет структуру задачи), а также для вариационных неравенств и седловых задач.

Глава 2 диссертации посвящена новой методике решения вариационных неравенств и седловых задач, которая позволяет учитывать погрешности задания оператора, а также применима к вариационным неравенствам с гёльдеровыми монотонными операторами и выпукло-вогнутым седловым задачам соответствующего уровня гладкости.

Во **введении** к главе 2, в частности, напоминается постановка задачи решения вариационного неравенства (далее сокращенно — ВН), а также необходимые понятия и результаты. Для оператора $G : Q \to \mathbb{R}^n$, заданного на выпуклом компакте $Q \subset \mathbb{R}^n$ под *вариационным неравенством* понимаем неравенство вида

$$\langle G(x_*),\ x_* - x\rangle \leqslant 0. \tag{1}$$

Отметим, что в (1) требуется найти $x_* \in Q$ (это x_* и называется (строгим) решением ВН), для которого

$$\max_{x\in Q}\langle G(x_*),\ x_* - x\rangle \leqslant 0.$$

Для монотонного оператора поля G можно рассматривать также задачу отыскания *слабого решения вариационного неравенства*

$$\langle G(x),\ x_* - x\rangle \leqslant 0, \tag{2}$$

т.е. нахождения $x_* \in Q$, для которого (2) верно при всех $x \in Q$.

[36] Lu H., Freund R. M., and Nesterov Y. Relatively-smooth convex optimization by first-order methods and applications. // SIAM Journal on Optimization. 2018. Vol. 28, No. 1. P. 333 – 354.

[37] Tyurin A. I., Gasnikov A. V. Fast gradient descent method for convex optimization problems with an oracle that generates a (δ, L)-model of a function in a requested point. // Computational Mathematics and Mathematical Physics. 2019. Vol.59, No. 7. P. 1137 – 1150.

Один из основных подходов главы 2 — использование введённых автором [38] аналогов (δ, L)-оракула (и, в более общем случае, (δ, L)-модели функции) для вариационных неравенств и седловых задач. Для удобства будем рассматривать даже более общую задачу нахождения решения $x_* \in Q$ абстрактной задачи равновесного программирования

$$\psi(x, x_*) \geqslant 0 \quad \forall x \in Q \tag{3}$$

для некоторого выпуклого компакта $Q \subset \mathbb{R}^n$, а также функционала $\psi : Q \times Q \to \mathbb{R}$. Если предположить абстрактную монотонность функционала ψ:

$$\psi(x, y) + \psi(y, x) \leqslant 0 \;\; \forall x, y \in Q,$$

то всякое решение (3) будет также и решением двойственной задачи равновесного программирования

$$\psi(x_*, x) \leqslant 0 \quad \forall x \in Q. \tag{4}$$

В общем случае делается предположение о существовании решения x_* задачи (3).

В разделе 2.1 рассмотрено следующее понятие (δ, L)-модели для указанного выше класса задач и предлагается вариант проксимального зеркального метода с адаптивным выбором шага и адаптивным критерием остановки.

Определение 2.1.1. *Будем говорить, что функционал ψ допускает (δ, L)-модель $\psi_\delta(x, y)$ при некоторых фиксированных $L > 0$ и $\delta > 0$ на множестве Q относительно дивергенции Брэгмана $V(y, x)$, если для всяких $x, y, z \in Q$ верны:*

(i) $\psi(x, y) \leqslant \psi_\delta(x, y) + \delta$*;*

(ii) $\psi_\delta(x, y)$ *— выпуклый функционал по первой переменной;*

(iii) $\psi_\delta(x, x) = 0$*;*

(iv) (абстрактная δ-монотонность)

$$\psi_\delta(x, y) + \psi_\delta(y, x) \leqslant \delta;$$

(v) (обобщённая относительная гладкость)

$$\psi_\delta(x, y) \leqslant \psi_\delta(x, z) + \psi_\delta(z, y) + LV(x, z) + LV(z, y) + \delta. \tag{5}$$

[38] F. S. Stonyakin. On the Adaptive Proximal Method for a Class of Variational Inequalities and Related Problems. // Proceedings of the Steklov Institute of Mathematics. 2020. Vol. 309(1). P.S139 – S150.

Естественно возникает идея обобщить этот метод на абстрактные задачи (3) и (4) в предположениях их разрешимости, а также существования неточной модели, удовлетворяющей определению 2.1.1. При этом будем учитывать погрешность δ в (5), а также погрешность $\tilde{\delta}$ решения вспомогательных задач на итерациях согласно одному из достаточно известных в алгоритмической оптимизации подходов:

$$x := \arg\min_{y\in Q}^{\tilde{\delta}} \varphi(y), \text{ если } \langle \nabla\varphi(x), x-y\rangle \leqslant \tilde{\delta} \quad \forall y \in Q.$$

Ниже описана $(N+1)$-ая итерация предложенного метода $(N = 0, 1, 2, \ldots)$ с выбором начального приближения $x^0 = \arg\min\limits_{x\in Q} d(x)$ для фиксированной точности $\varepsilon > 0$, а также некоторой константы $L_0 \leqslant 2L$.

Алгоритм 1 Адаптивный метод для концепции (δ, L)-модели для ВН.

---

1. $N := N+1$; $L_{N+1} := \frac{L_N}{2}$.

2. Вычисляем:
   $y^{N+1} := \arg\min\limits_{x\in Q}^{\tilde{\delta}}\big\{\psi_\delta(x, x^N) + L_{N+1}V(x, x^N)\big\}$,
   $x^{N+1} := \arg\min\limits_{x\in Q}^{\tilde{\delta}}\big\{\psi_\delta(x, y^{N+1}) + L_{N+1}V(x, x^N)\big\}$
   до тех пор, пока не будет выполнено:

$$\begin{aligned}\psi_\delta(x^{N+1}, x^N) \leqslant \psi_\delta(y^{N+1}, x^N) + \psi_\delta(x^{N+1}, y^{N+1}) + \\ + L_{N+1}V(y^{N+1}, x^N) + L_{N+1}V(x^{N+1}, y^{N+1}) + \delta.\end{aligned} \tag{6}$$

3. **Если** (6) не выполнено, **то** $L_{N+1} := 2L_{N+1}$ и повторяем п. 2.

4. **Иначе** переход к п. 1.

5. Критерий остановки метода:

$$S_N := \sum_{k=0}^{N-1} \frac{1}{L_{k+1}} \geqslant \frac{\max\limits_{x\in Q} V(x, x^0)}{\varepsilon}.$$

Получена следующая

Теорема 2.1.6. *Пусть функционал ψ допускает (δ, L)-модель $\psi_\delta(x, y)$ при некоторых фиксированных $L > 0$ и $\delta > 0$ на множестве Q относительно* *дивергенции Брэгмана $V(y,x)$. Тогда после остановки алгоритма 1 для всякого $x \in Q$ будет выполнено неравенство:*

$$-\frac{1}{S_N}\sum_{k=0}^{N-1}\frac{\psi_\delta(x,y^{k+1})}{L_{k+1}} \leqslant \frac{V(x,x^0)}{S_N} + 2\tilde{\delta} + \delta \leqslant \varepsilon + 2\tilde{\delta} + \delta, \tag{7}$$

а также

$$\psi(\tilde{y},x) \leqslant \frac{V(x,x^0)}{S_N} + 2\tilde{\delta} + 3\delta \leqslant \varepsilon + 2\tilde{\delta} + 3\delta \tag{8}$$

при

$$\tilde{y} := \frac{1}{S_N}\sum_{k=0}^{N-1}\frac{y^{k+1}}{L_{k+1}}. \tag{9}$$

Замечание 2.1.7. *Ввиду* (5) *и выбора $L_0 \leqslant 2L$ гарантированно будет верно $L_{k+1} \leqslant 2L \ \ \forall k = \overline{0, N-1}$. Поэтому $S_N \geqslant \frac{N}{2L}$ и оценки* (7)–(8) *означают, что для всякого $x \in Q$ будут верны неравенства:*

$$\psi(\tilde{y},x) \leqslant \varepsilon + 2\tilde{\delta} + 3\delta \tag{10}$$

после выполнения не более, чем $O\left(\varepsilon^{-1}\right)$ итераций предлагаемого метода. При этом нетрудно проверить, количество решений вспомогательных задач в п.2 алгоритма на N итерациях метода не превышает $2N + \log_2 \frac{L}{L_0}$, т.е. стоимость итерации в среднем будет сопоставимой со стоимостью итерации классического экстраградиентного метода, предполагающей решение двух вспомогательных задач на каждой итерации. Отметим, что оценка с точностью до числового множителя оптимальна для вариационных неравенств с монотонным липшицевым оператором[39]*.*

Замечание 2.1.8. *Для обычных слабых вариационных неравенств* (2) *неравенство* (10) *можно заменить на*

$$\max_{x\in Q}\langle G(x), \tilde{y} - x\rangle \leqslant \varepsilon + 2\widetilde{\delta} + 3\delta. \tag{11}$$

Отметим, что неравенство вида (11) *довольно часто используют как критерий качества решения вариационного неравенства.*

В разделе 2.2 рассмотрен случай гёльдерова оператора поля вариационного неравенства G

$$||G(x) - G(y)||_* \leqslant L_\nu ||x-y||^\nu \quad \forall x,y \in Q \tag{12}$$

[39] Nemirovsky A. S. Information-based complexity of linear operator equations. Journal of Complexity, 8(2):153–175, 1992.

для произвольного $\nu \in [0; 1]$, причем $L_0 < +\infty$ (другие константы L_ν ($\nu \neq 0$) могут быть бесконечными) Можно показать, что при условии (12)

$$\langle G(z) - G(y), z - x\rangle \leqslant \frac{L}{2}||z - x||^2 + \frac{L}{2}||z - y||^2 + \frac{\varepsilon}{2} \tag{13}$$

для константы

$$L = L_\nu^{\frac{2}{1+\nu}} (2\varepsilon)^{\frac{\nu-1}{1+\nu}},$$

которая зависит от введённой искусственной неточности ε. На базе интерполяции (13) алгоритм 1 с заменой $V(y, x)$ на $\frac{1}{2}\|y - x\|^2$ сводится к *универсальному* методу для вариационных неравенств, который предполагает адаптивную настройку на уровень гладкости оператора G. Доказано, что в таком случае разработанный метод позволяет получить ε -приближённое решение задач (1) – (2) с оценкой достаточного количества итераций для достижения приемлемого качества решения

$$O\left(\varepsilon^{-\frac{2}{1+\nu}}\right), \tag{14}$$

которая оптимальна с точностью до постоянного множителя при $\nu = 0$ и $\nu = 1$. При этом адаптивность метода на практике может приводить к ускорению работы метода по сравнению с (14), что продемонстрировано на примере экспериментов.

В разделе 2.3 введён аналог понятия (δ, L)-модели для более узкого по сравнению с предыдущим разделом класса седловых задач и выписана оценка скорости сходимости алгоритма 1. Обоснована применимость этой концепции к композитным седловым задачам, которые возникают в задачах обработки изображений[40)].

В разделе 2.4 приведены некоторые численные эксперименты, демонстрирующие преимущества использования разработанных адаптивных (универсальных) методов для рассмотренных в предыдущих разделах главы 2 классов задач. В частности, рассматриваются расчёты лагранжевы седловые задачи для негладкой геометрической задачи Ферма-Торричелли-Штейнера и задачи о наименьшем покрывающем шаре с функциональными ограничениями, а также задачи решения билинейных матричных игр.

В **разделе 2.5** с использованием методики усреднений адаптивно подбираемых констант Липшица оператора поля получена оценка скорости сходимости предложенного автором [41)] адаптивного аналога метода [42)] [43)] для вариацион-

[40)] Chambolle A. and Pock T. A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision, 40(1):120–145, May 2011.

[41)] Ф.С. Стонякин. Адаптивный аналог метода Ю. Е. Нестерова для вариационных неравенств с сильно монотонным оператором. // Сибирский журнал вычислит. матем. 2019. Т. 22, № 2. С. 201 -– 211.

[42)] Nesterov Yu., Scrimali L. Solving strongly monotone variational and quasi-variational inequalities. // Discrete and Continuous Dynamical Systems, 2007

[43)] Нестеров Ю. Е. Алгоритмическая выпуклая оптимизация. // Дисс. ...докт. физ.-мат. наук: 01.01.07 /

ных неравенств с сильно монотонным липшицевым оператором на выпуклом замкнутом множестве Q. Полученная оценка скорости адаптивного метода может оказаться эффективнее по сравнению с известной оценкой для неадаптивного варианта метода, что проиллюстрировано некоторыми численными экспериментами.

Глава 3 диссертационной работы посвящена новым результатам в области методов градиентного типа с адаптивной настройкой на величины погрешностей данных, которые могут быть как естественными (приближённые значения целевой функции или градиента), так и искусственными (связаны с более низким уровнем гладкости целевого функционала). Отличительная особенность подхода — наличие не одного, а двух соответствующих возможным погрешностям параметров (в частности, для гладких задач погрешности задания целевого функционала и его градиента). При этом такой подход даёт возможность избежать накопления значений одного из этих параметров погрешностей (задания градиента в гладком случае) в теоретических оценках скорости сходимости метода.

Во **введении к главе 3** приводится мотивировка для разрабатываемого подхода к аналогу неточного оракула (точнее, (δ, L)-модели) целевой функции с несколькими параметрами, которые описывают разные типы неточностей. Например, для задач задач минимизации с помощью предлагаемого подхода можно раздельно учитывать погрешности задания целевого функционала f и градиента ∇f. Если положить, что для всякого $x \in Q$ при некотором фиксированном $\Delta > 0$ верно $\left\|\nabla f(x) - \widetilde{\nabla} f(x)\right\|_* \leqslant \Delta,$ для некоторого доступного приближенного значения $\widetilde{\nabla} f(x)$ градиента ∇f, то будет верно неравенство $\left|\left\langle \nabla f(x) - \widetilde{\nabla} f(x), y - x \right\rangle\right| \leqslant \Delta\|y - x\|$. Это означает, что для всякого функционала f с L-липшицевым градиентом и при условии доступности в точке x лишь приближённого значения $f_\delta(x)$ такого, что $f_\delta(x) \leqslant f(x) \leqslant f_\delta(x) + \delta$ при некотором $\delta > 0$, будет верно неравенство

$$\begin{gathered} f_\delta(x) + \left\langle \widetilde{\nabla} f(x), y - x \right\rangle - \Delta\|y - x\| \leqslant f(y) \leqslant \\ \leqslant f_\delta(x) + \left\langle \widetilde{\nabla} f(x), y - x \right\rangle + \frac{L}{2}\|y - x\|^2 + \Delta\|y - x\| + \delta \ \ \forall x, y \in Q. \end{gathered} \tag{15}$$

Можно ввести (причём в абстрактной модельной общности) следующий ана-

Нестеров Юрий Евгеньевич; [Место защиты: Моск. физ.-техн. ин-т (гос. ун-т)]. — Москва, 2013. — 367 с.: ил. РГБ ОД, 71 15-1/114.

лог неравенства (15) с параметрами $\delta, \gamma, \Delta \geqslant 0$:

$$\begin{gathered} f(x) + \left\langle \widetilde{\nabla} f(x), y - x \right\rangle - \delta - \gamma \|y - x\| \leqslant f(y) \leqslant \\ \leqslant f(x) + \left\langle \widetilde{\nabla} f(x), y - x \right\rangle + \frac{L}{2}\|y - x\|^2 + \Delta\|y - x\| + \delta \ \ \forall x, y \in Q. \end{gathered} \tag{16}$$

Смысл такого обобщения заключается в том, что возможны различные значения параметров γ и Δ в (16). Как показано в разделе 3.3 далее, влияние величины Δ на оценку итогового качества решения может быть уменьшено. В главе 3, в частности, рассмотрены численные эксперименты для некоторых примеров негладких задач в случае $\delta = \gamma = 0$ при $\Delta > 0$. Если положить $\gamma = 0$, то $\widetilde{\nabla} f(x)$ — δ-субградиент f в точке x и параметр $\Delta > 0$ может указывать в этом случае на скачки $\widetilde{\nabla} f(x)$ в точках негладкости f. Вполне естественно рассматривать Δ как искусственную неточность для негладкой задачи. В таком случае Δ в неравенстве

$$f(x) + \left\langle \widetilde{\nabla} f(x), y - x \right\rangle \leqslant f(y) \leqslant f(x) + \left\langle \widetilde{\nabla} f(x), y - x \right\rangle + \frac{L}{2}\|y - x\|^2 + \Delta\|y - x\|. \tag{17}$$

вполне можно рассматривать как оценку скачков субдифференциалов f вдоль всевозможных векторных отрезков $[x; y]$. Несложно понять, что если f в (17) удовлетворяет условию Липшица с константой $M > 0$, то $\Delta \leqslant 2M$. Отметим, что возможна ситуация, когда Δ существенно меньше M. Похожие на (17) условия рассматривались в [44) 45)] для случая, когда f представим в виде суммы гладкого и негладкого целевого функционала. Рассматриваемые в главе 3 подходы к введению оптимизационной модели позволяют, в частности, рассматривать более широкий класс целевых функционалов, которые не обязательно представимы в виде суммы гладкого и негладкого слагаемых.

В **разделе 3.1** предложено следующее понятие неточной модели целевой функции.

Определение 3.1.1. *Будем говорить, что f допускает $(\delta, \gamma, \Delta, L)$-модель в точке $x \in Q$, если для некоторой выпуклой по первой переменной функции $\psi(y, x)$ такой, что $\psi(x, x) = 0$, будет верно неравенство*

$$\begin{gathered} f(x) + \psi(y, x) - \gamma\|y - x\| + \delta \leqslant f_\delta(x) + \psi(y, x) - \gamma\|y - x\| \\ \leqslant f(y) \leqslant f_\delta(x) + \psi(y, x) + \delta + \Delta\|y - x\| + LV(y, x) \leqslant \\ \leqslant f_\delta(x) + \psi(y, x) + \delta + \Delta\|y - x\| + LV(y, x) \end{gathered} \tag{18}$$

для произвольных $x, y \in Q$.

44) Devolder O. Exactness, Inexactness and Stochasticity in First-Order Methods for Large-Scale Convex Optimization. // PhD thesis (2013).https://dial.uclouvain.be/pr/boreal/object/boreal:128257.

45) Lan G. Gradient sliding for composite optimization. // Math. Program. 2016. Vol. 159, no. 1–2, P. 201–235.

Покажем пример, поясняющий смысл использования модельной общности в предыдущем определении. В этом контексте можно отметить задачу выпуклой композитной оптимизации $f(x) = g(x) + h(x) \to \min$, где g — гладкая выпуклая функция, а h — выпуклая не обязательно гладкая функция простой структуры (операция проектирования на любое множество уровня h не сильно затратна). Если при этом для градиента ∇g задано его приближение $\widetilde{\nabla} g$: $\left\|\widetilde{\nabla} g(x) - \nabla g(x)\right\| \leqslant \Delta$, причем

$$g(y) \geqslant g(x) + \langle \widetilde{\nabla} g(x), y - x \rangle - \Delta\|y - x\| - \delta,$$

то можно положить $\psi(y, x) = \langle \widetilde{\nabla} g(x), y - x \rangle + h(y) - h(x)$ и будет верно неравенство (18) при $\gamma = \Delta$.

Предложен следующий алгоритм 6, допускающий адаптивную оценку качества найденного решения к величинам параметров (δ, Δ, L)-модели.

Алгоритм 6 Адаптивный градиентный метод для функций, допускающих $(\delta, \gamma, \Delta, L)$-модель в запрошенной точке.

Require: $x^0 \in Q$ — начальная точка, $V(x_*, x^0) \leqslant R^2$, параметры:
$L_0 > 0,\ \Delta_0 > 0,\ \delta_0 > 0$: $L_0 \leqslant 2L,\ \Delta_0 \leqslant 2\Delta,\ \delta_0 \leqslant 2\delta$.

1: $L_{k+1} := {}^{L_k}\!/_2,\ \Delta_{k+1} := {}^{\Delta_k}\!/_2,\ \delta_{k+1} := {}^{\delta_k}\!/_2$.
2: $x^{k+1} := \arg\min\limits_{x \in Q}\{\psi(x, x^k) + LV(x, x^k)\}$.
3: **if** $f_\delta(x^{k+1}) \leqslant f_\delta(x^k) + \psi(x^{k+1}, x^k) + L_{k+1}V(x^{k+1}, x^k) + \Delta_{k+1}\left\|x^{k+1} - x^k\right\| + \delta_{k+1}$ **then**
4: $k := k + 1$ и выполнение п. 1.
5: **else**
6: $L_{k+1} := 2 \cdot L_{k+1};\ \Delta_{k+1} := 2 \cdot \Delta_{k+1};\ \delta_{k+1} := 2 \cdot \delta_{k+1}$ и выполнение п. 2.
7: **end if**

Ensure: $\widehat{x} := \frac{1}{S_N} \sum\limits_{k=0}^{N-1} \frac{x^{k+1}}{L_{k+1}},\ S_N := \sum\limits_{k=0}^{N-1} \frac{1}{L_{k+1}}$.

Имеет место следующая

Теорема 3.1.4. *Пусть $f : Q \to \mathbb{R}$ — выпуклая функция, допускающая $(\delta, \gamma, \Delta, L)$-модель $\psi(y, x)$ в любой точке $x \in Q$, а также $V(x_*, x^0) \leqslant R^2$, где x^0 — начальное приближение и x_* — точное решение, ближайшее к x^0 с точки зрения дивергенции Брэгмана. Тогда после N итераций для выхода $\hat{x}$ алгоритма 6 будет верно неравенство*

$$f(\hat{x}) - f(x_*) \leqslant \frac{R^2}{S_N} + \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{\delta_{k+1} + \Delta_{k+1}\left\|x^{k+1} - x^k\right\| + \gamma\|x^k - x_*\|}{L_{k+1}} + \delta. \quad (19)$$

Отметим, что вспомогательная задача п. 2 листинга алгоритма 6 решается не более

$$2N + \max\left\{\log_2 \frac{2L}{L_0}, \log_2 \frac{2\delta}{\delta_0}, \log_2 \frac{2\Delta}{\Delta_0}\right\} \qquad (20)$$

раз.

Как видим, параметр δ рассмотренной не допускает полной адаптивной настройки. Поэтому по сути предложено лишь частичное решение проблемы разработки алгоритмического метода с адаптивной настройкой в оценке скорости сходимости параметров, соответствующих разным типам неточностей.

Замечание 3.1.5. *Оценка (20) показывает, что в среднем трудоемкость итерации предложенного адаптивного алгоритма превышает трудоемкость неадаптивного метода не более, чем в постоянное число раз. Отметим также, что при $k = 0, 1, 2, \ldots$ $L_{k+1} \leqslant 2CL$, $C = \max\left\{1, \frac{2\delta}{\delta_0}, \frac{2\Delta}{\Delta_0}\right\}$. Поэтому $S_N \leqslant \frac{N}{2CL}$, что указывает на скорость сходимости метода $O(\varepsilon^{-1})$, но при наличии в оценке (19) слагаемых, определяемой параметрами δ и Δ (при этом ввиду адаптивности метода δ_k и Δ_k могут быть меньше δ и Δ соответственно).*

Далее, в **разделе 3.2** предложен вариант быстрого градиентного метода для задач выпуклой минимизации с адаптивным выбором шага и адаптивной настройкой на величины параметров (δ, Δ, L)-модели (аналога $(\delta, \gamma, \Delta, L)$-модели при $\gamma = 0$ и замене $V(y, x)$ на $\frac{1}{2}\|y - x\|^2$) и получена оценка качества найденного решения. При этом возможна адаптивная настройка некоторых из параметров модели, соответствующих погрешностям задания целевой функции и градиента. Известные подходы к проблеме накопления ошибки градиента для ускоренного метода основаны на нетривиальной концепции аппроксимативного градиента [46] или на понимании погрешности как случайной величины [47] [48] [49]. Подход настоящей работы позволяет получить результат о скорости сходимости варианта ускоренного метода в модельной общности.

В **разделе 3.3** рассмотрен специальный класс задач выпуклой негладкой оптимизации, к которым применим подход определения 3.1.1 ($\delta = \gamma = 0$,

[46] D'Aspremont A. Smooth optimization with approximate gradient. // SIAM Journal of Optimization. 2008. Vol. 19, № 3. P. 1171–1183.

[47] Gorbunov E., Dvinskikh D., Gasnikov A. Optimal Decentralized Distributed Algorithms for Stochastic Convex Optimization. `https://arxiv.org/abs/1911.07363v2`, 2019.

[48] А. В. Гасников. Современные численные методы оптимизации. Метод универсального градиентного спуска. — М.: МФТИ, 2018. urlhttps://arxiv.org/ftp/arxiv/papers/1711/1711.00394.pdf.

[49] Cohen M. B., Diakonikolas J., Orecchia L. On Acceleration with Noise-Corrupted Gradients. Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

$\Delta > 0$). Показано, что для этой ситуации возможно модифицировать неускоренный метод так, чтобы гарантированно достигалось ε-точное решение задачи минимизации f за

$$O\left(\frac{L}{\varepsilon}\right) + O\left(\frac{\Delta^2}{\varepsilon^2}\right) \tag{21}$$

обращений к (суб)градиенту целевого функционала. Для ускоренного метода аналогичной модификацией можно добиться того, чтобы гарантированно достигалось ε-точное решение задачи минимизации f за

$$O\left(\sqrt{\frac{L}{\varepsilon}}\right) + O\left(\frac{\Delta^2}{\varepsilon^2}\right) \tag{22}$$

обращений к (суб)градиенту целевого функционала. По сути получен некоторый аналог результата диссертации О. Деволдера (п. 4.7.2) [50]. Однако важным и новым является то, уже необязательно оптимизируемый функционал имеет вид суммы гладкого и негладкого слагаемого и рассмотрена модельная общность. Отметим, что в случае целевого функционала, равного сумме гладкого и M-липшицева функционалов, параметр Δ может быть существенно меньше M. Также предлагается метод с адаптивным выбором шага с гарантированными оценками скорости сходимости, которые близки к (21) и (22). Отметим, что при этом для адаптивного ускоренного метода оценка может увеличиться на некоторый логарифмический множитель вида $O(log_2(\varepsilon^{-3}))$, для адаптивного неускоренного — на множитель вида $O(log_2(\varepsilon^{-1}))$.

Опишем указанный результат для неускоренного метода в предположении, что целевой функционал допускает (δ, Δ, L)-модель ψ при $\delta = 0$. Заметим, что при этом уже требуется 1-сильная выпуклость прокс-функции в определении дивергенции Брэгмана.

Пусть на (k+1)-й итерации алгоритма 6 ($k = 0, 1, \dots, N-1$) верно неравенство $L \leqslant L_{k+1} \leqslant 2L$ (как показано в п. 2 доказательства теоремы 3.1.4, этого можно всегда добиться выполнением не более чем постоянного числа операций п. 2 листинга алгоритма 6). Для каждой итерации алгоритма 6 ($k = 0, 1, \dots, N-1$) предложим такую процедуру:

Повторяем операции п. 2 p раз, увеличивая L_{k+1} в два раза при неизменной $\Delta_{k+1} \leqslant 2\Delta$.

(23)

Процедуру (23) остановим в случае выполнения одного из неравенств:

$$\Delta_{k+1}\left\|x^{k+1} - x^k\right\| \leqslant \frac{\varepsilon}{2},$$

[50] Devolder O. Exactness, Inexactness and Stochasticity in First-Order Methods for Large-Scale Convex Optimization. // PhD thesis (2013). https://dial.uclouvain.be/pr/boreal/object/boreal:128257.

или
$$f(x^{k+1}) \leqslant f(x^k) + \psi(x^{k+1}, x^k) + 2^{p-1}L\left\|x^{k+1} - x^k\right\|^2.$$

Отметим, что здесь мы полагаем f точно заданной, то есть $f_\delta = f$ $(\delta = 0)$ и $\psi(y, x) = \langle \nabla f(x), y - x \rangle$, где ∇f — некоторый субградиент f. Получена верхняя оценка на p и доказан следующий результат (см. (21)).

Теорема 3.3.9. *Для выхода $\hat{x}$ модифицированного алгоритма 6 с учетом дополнительной процедуры* (23) *неравенство $f(\hat{x}) - f^* \leqslant \varepsilon$ будет гарантированно выполнено не более, чем после*

$$\left\lceil \frac{4LR^2}{\varepsilon} + \frac{64\Delta^2R^2}{\varepsilon^2} \right\rceil \cdot \left\lceil \log_2\left(1 + \frac{16\Delta^2}{\varepsilon L}\right) \right\rceil \tag{24}$$

вычислений субградиента f.

Замечание 3.3.12. *Если не предполагать, что на (k+1)-й итерации ($k = 0, 1, \ldots, N-1$) модифицированного алгоритма 6 выполнено неравенство $L \leqslant L_{k+1} \leqslant 2L$ и предусмотреть полностью адаптивную настройку параметров L и Δ в методе, то оценка* (24) *может увеличиться не более, чем в*

$$\left\lceil \max\left\{\frac{2L}{L_0}, \frac{2\Delta}{\Delta_0}\right\} \right\rceil$$

раз. Отметим также, что логарифмический множитель в (24) *можно опустить, если рассматривать неадаптивный вариант алгоритма 6 с фиксированным параметром $L_{k+1} = 2^pL$ при подходящем натуральном p.*

В **разделе 3.5** предложен метод для вариационных неравенств с адаптивной настройкой величины, которая соответствует аддитивному шуму при задании оператора[51)]. В частности, получен аналог теоремы 3.3.9 для вариационных неравенств.

В **главе 4** рассмотрены некоторые адаптивные методы для сильно выпуклых задач оптимизации, которые гарантируют линейную скорость сходимости в случае отсутствия погрешностей. Хорошо известно, что в случае сильной выпуклости целевого функционала оценки скорости сходимости градиентного метода существенно улучшаются. Например, для сильно выпуклого целевого функционала с липшицевым градиентом известно, что градиентный метод сходится с линейной скоростью. Весьма популярен вопрос о том, насколько можно

[51)] F. Stonyakin, E. Vorontsova and M. Alkousa. New Version of Mirror Prox for Variational Inequalities with Adaptation to Inexactness. // 10th International Conference on Optimization and Applications, OPTIMA-2019. Communications in Computer and Information Sciences. 2020. Vol. 1145. P. 427 – 442.

условие сильной выпуклости ослабить. В этом случае известен подход, основанный на использовании вместо сильной выпуклости условия градиентного доминирования Поляка-Лоясиевича ((PL)-условие) [52) 53)]

$$f(x) - f(x_*) \leqslant \frac{1}{2\mu}\|\nabla f(x)\|_2^2 \quad \forall x \in Q, \tag{25}$$

где x_* — точное решение задачи минимизации f. Известно, неравенство (25) в предположении липшицевости градиента f позволяет получить оценку скорости сходимости

$$f(x^N) - f(x_*) \leqslant \left(1 - \frac{\mu}{L}\right)^N \left(f(x^0) - f(x_*)\right) \leqslant \exp\left(-\frac{\mu}{L}N\right)\left(f(x^0) - f(x_*)\right). \tag{26}$$

В **разделе 4.1** введено следующее ослабление условия липшицевости градиента [54)]

$$f(y) \leqslant f(x) + \langle \widetilde{\nabla} f(x), y - x\rangle + \frac{L}{2}\|y - x\|_2^2 + \Delta\|y - x\|_2 \quad \forall\, x, y \in Q$$

для некоторых $\delta, \Delta > 0$. Например, это предположение естественно в случае, если f удовлетворяет условию Липшица градиента и $\widetilde{\nabla} f(x)$ — некоторое возмущенное с точностью Δ значение градиента $\nabla f(x)$. Предложен следующий алгоритм 10 с адаптивным подбором шага с настройкой на величины L и Δ и показана оценка скорости сходимости, аналогичная (26).

Для данного метода доказано, что либо невязка $\min\limits_k f(x^k) - f(x_*)$ убывает со скоростью геометрической прогрессии при увеличении k (см. (27)), либо эта невязка ограничена сопоставимой с Δ величиной (в т.ч. и в случае неположительного шага). Справедлива следующая

Теорема 4.1.2. *Пусть для некоторого натурального числа k верно $\Delta_{k+1} \leqslant \Delta$ и $h_i > 0$ при всех $i = \overline{1,k}$. Тогда после k итераций алгоритма 10 для всякого $C > 1$ будет выполняться одно из двух неравенств*

$$f(x^{k+1}) - f(x_*) \leqslant \left(1 - \frac{\mu}{\hat{L}}\left(\frac{C-1}{C+1}\right)^2\right)^{k+1} (f(x^0) - f(x_*)), \tag{27}$$

где $\hat{L} \leqslant 2L\max\left\{1, \frac{2\Delta}{\Delta_0}\right\}$ или $\min\limits_{i=\overline{1,k+1}} f(x^i) - f^* < \frac{(C+1)^2\Delta^2}{2\mu}$.

---

52) Поляк Б. Т. Градиентные методы минимизации функционалов // Журн. вычислит. математики и мат. физики. 1963. Том 3, № 4. С. 643–653.

53) Karimi H., Nutini J., Schmidt M. Linear Convergence of Gradient and Proximal-Gradient Methods Under the Polyak-Lojasiewicz Condition. Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2016. Lecture Notes in Computer Science, vol 9851. Springer, Cham.

54) Ф. С. Стонякин. Адаптация к величинам погрешностей для некоторых методов оптимизации градиентного типа. // Труды ИММ УрО РАН. 2019. Т. 25, № 4. С. 210 – 225.

Алгоритм 10 Адаптивный градиентный метод для функций, удовлетворяющих (PL)-условию.

Require: x^0 — начальная точка, параметры $\Delta_0,\ L_0$
$(2\mu \leqslant L_0 < 2L,\ \Delta_0 \leqslant 2\Delta)$.

1: $L_{k+1} := {}^{L_k}\!/_2,\ \Delta_{k+1} := {}^{\Delta_k}\!/_2$.
2: $x^{k+1} = x^k - h_k \widetilde{\nabla} f(x^k)$, $h_k = \frac{1}{L_{k+1}} - \frac{\Delta_{k+1}}{L_{k+1}\widetilde{g}_{x^k}}$, $\widetilde{g}_{x^k} = \left\|\widetilde{\nabla} f(x^k)\right\|_2$.
3: **repeat**
4:     **if** $f(x^{k+1}) \leqslant f(x^k) + \left\langle \widetilde{\nabla} f(x^k), x^{k+1} - x^k \right\rangle + \dfrac{L_{k+1}}{2}\left\|x^{k+1} - x^k\right\|_2^2 + \Delta_{k+1}\left\|x^{k+1} - x^k\right\|_2$ **then**
5:         $k := k + 1$ и выполнение п. 1.
6:     **else**
7:         $L_{k+1} := 2 \cdot L_{k+1};\ \Delta_{k+1} := 2 \cdot \Delta_{k+1}$ и выполнение п. 2.
8:     **end if**
9: **until** $k \geqslant N$

Ensure: x^{k+1}.

В **разделе 4.2** вводится аналог понятия (δ, L, μ)-оракула [55] для задач оптимизации, который применим к классу задач, близкому к гладким сильно выпуклым задачам оптимизации. Предложен метод с адаптивной настройкой параметра гладкости L, обоснована близкая к линейной скорость сходимости (с точностью до величины, определяемой погрешностями).

Определение 4.2.1. *Будем говорить, что функция f допускает (δ, L, μ)-модель в точке x для некоторой константы $\mu > 0$, если для любого $y \in Q$ верно:*

$$\mu V(y, x) \leqslant f(y) - (f_\delta(x) + \psi(y, x)) \leqslant LV(y, x) + \delta,$$

где $\psi(y, x)$ — выпуклая по y функция, $\psi(x, x) = 0$, $\delta > 0$.

Отметим, что обычное свойство μ-сильно выпуклости функции

$$\frac{\mu}{2}||y - x||^2 \leq f(y) - f(x) - \langle \nabla f(x), y - x \rangle$$

здесь заменяется неравенством $\mu V(y, x) + f_\delta(x) + \psi(y, x) \leqslant f(x)$. При достаточно стандартном [56] условии на выбор прокс-функции обычная сильная выпуклость может гарантировать выполнение последнего неравенства.

[55] Devolder O. Exactness, Inexactness and Stochasticity in First-Order Methods for Large-Scale Convex Optimization. // PhD thesis (2013).

[56] Nemirovski A. Lectures on modern convex optimization analysis, algorithms, and engineering applications. Philadelphia: SIAM, 2013.

Замечание 4.2.3. *Пусть для всяких* $x, y \in Q$ *верно* $d(y-x) \leqslant C_n\|y-x\|^2$, *где* n *— размерность пространства и* $C_n = O(\log n)$. *Тогда* $V(y,x) \leqslant C_n\|y-x\|^2$ *и* μC_n*-сильная выпуклость* $\frac{\mu}{2}||y-x||^2 \leqslant f(y) - f(x) - \psi(y,x)$ *влечёт относительную* μ*-сильную выпуклость:* $\mu V(y,x) + f_\delta(x) + \psi(y,x) \leqslant f(y)$.

В **разделе 4.3** по аналогии с определением 3.1.1 введена концепция (δ, Δ, L, μ)-модели целевой функции и предложен метод градиентного типа с адаптивной настройкой параметров неточности этой модели. По сути в этой части работы предложен метод с адаптивной настройкой параметров неточной модели на класс задач с условием, близким к сильной выпуклости целевого функционала. Обоснована скорость сходимости этого метода, близкая к линейной (с точностью до величин, определяемых погрешностями).

В **разделе 4.4** предложен адаптивный метод[57)] для задачи сильно выпуклой оптимизации с одним функциональным ограничением

$$f(x) \to \min, \quad x \in Q \subset \mathbb{R}^n,\ g(x) \leqslant 0, \tag{28}$$

где Q — выпуклый компакт в конечномерном нормированном пространстве $\mathbb{R}^n$, f и g — выпуклые функционалы, g удовлетворяет условию Липшица

$$|g(y) - g(x)| \leqslant M_g\ \|x-y\|_2\ \ \forall x, y \in Q$$

при некоторой постоянной $M_g > 0$ ($\|\cdot\|_2$ — евклидова норма). При этом если функционалов ограничений несколько $\{g_p(x)\}_{p=1}^m$, то вполне можно рассмотреть задачу с одним ограничением $g(x) = \max\limits_{p=\overline{1,m}} g_p(x)$, которое будет заведомо удовлетворять условию Липшица, если все g_p удовлетворяют условию Липшица. Вполне естественно рассматривать подход, основанный на замене (28) двойственной к ней задачей

$$\varphi(\lambda) = \min_{x\in Q}\{f(x) + \lambda g(x)\} \to \max_{\lambda \geqslant 0}. \tag{29}$$

В этом случае двойственная функция зависит от одной двойственной переменной $\lambda \geqslant 0$. Если выполнены условия Слейтера для задачи (28), то возможные значения λ ограничены отрезком. Это позволяет применять метод дихотомии для нахождения значения двойственной переменной λ, которое близко к соответствующему оптимальному значению λ_*, для которого

$$\lambda_* \cdot g(x(\lambda_*)) = 0.$$

[57)] Stonyakin F. S., Alkousa M. S., Titov A. A., Piskunova V. V. On Some Methods for Strongly Convex Optimization Problems with One Functional Constraint. Mathematical Optimization Theory and Operations Research. MOTOR 2019. Lecture Notes in Computer Science, vol 11548. Springer, Cham, 2019.

Однако для эффективного решения (29) необходимо эффективно решать вспомогательную задачу многомерной минимизации по x функционала $f(x)+\lambda g(x)$ при фиксированном λ. Вообще говоря, такая задача может быть решена лишь с некоторой точностью методами оптимизации. Это приводит к погрешностям при нахождении $\varphi(\lambda)$ и ее производной $\varphi'(\lambda)$. Также если $f(x)+\lambda g(x)$ не сильно выпукла, то φ может быть негладкой в точке λ. Поэтому при рассмотрении указанного подхода вполне естественно потребовать сильную выпуклость целевого функционала. Исследовано влияние указанных погрешностей при решении двойственной задачи на качество решения прямой задачи с учётом одномерности двойственной переменной, а также условий на гладкость и (сильную) выпуклость f или g. Предложен алгоритм с адаптивным критерием остановки вида

$$\lambda \cdot |g(x_\varepsilon(\lambda)| \leqslant \varepsilon \tag{30}$$

для задач вида (28), где $x_\varepsilon(\lambda)$ — ε-точное решение по функции вспомогательной задачи минимизации (29) при текущем двойственном множителе λ. Рассмотрен класс задач с сильно выпуклым целевым функционалом f при следующих типах предположений для f и g:

$$|f(x)-f(y)| \leqslant M_f\|x-y\|_2, \quad |g(x)-g(y)| \leqslant M_g\|x-y\|_2 \tag{31}$$

или

$$\|\nabla f(x)-\nabla f(y)\| \leqslant L_f\|x-y\|_2, \quad \|\nabla g(x)-\nabla g(y)\| \leqslant L_g\|x-y\|_2 \tag{32}$$

для всех $x, y \in Q$ и для некоторых действительных положительных чисел M_f, M_g, L_f, L_g. Доказано, что при условиях (32) предлагаемый метод позволяет получить приемлемое качество решения задачи (28) после не более чем $O\left(\log_2^2 \frac{1}{\varepsilon}\right)$ обращений к подпрограмме, выдающей градиент f или g. В предположениях (31) (т.е. для негладких задач) оптимальной оценки сложности предложенного подхода обосновать не удалось. Однако адаптивность предложенного критерия остановки (30) может позволить существенно повысить скорость его работы по сравнению с теоретическими оценками и для некоторых примеров негладких задач, что проиллюстрировано экспериментами.

В **разделе 4.5** предложен подход для задач выпуклого программирования с двумя функционалами ограничений, который аналогичен рассмотренному в разделе 4.4. В этом случае двойственная задача уже двумерна и вполне естественно применить к ней вариант метода минимизации выпуклой липшицевой функции двух переменных на квадрате с фиксированной стороной [58]. Идея

[58] Пасечнюк Д. А., Стонякин Ф. С. Об одном методе минимизации выпуклой липшицевой функции двух переменных на квадрате. // Компьютерные исследования и моделирование. 2019. Т. 11, № 3. С. 379 – 395.

метода — деление квадрата на меньшие части и постепенное их удаление так, чтобы в оставшейся достаточно малой части все значения целевой функции были достаточно близки к оптимальному. Показано, что метод может работать для задач выпуклой гладкой оптимизации при наличии погрешностей решения вспомогательных одномерных задач, а также при вычислении направлений градиентов. Исследована возможная корреляция между указанными погрешностями. Также описана ситуация, когда возможно уменьшить временные затраты на решение вспомогательных одномерных задач. Данная методика приводит к подходу для решения двойственных задач к задаче многомерной сильно выпуклой минимизации с двумя выпуклыми функционалами ограничений. Этот подход основан на использовании для 2-мерной двойственной задачи предложенного метода Ю.Е. Нестерова с аналогичным (30) критерием остановки, который соответствует подходящему значению возмущенного градиента двойственной задачи.

Глава 5 диссертационной работы посвящена новым вариантам схем зеркального спуска с переключениями для следующего класса задач оптимизации

$$f(x) \to \min \tag{33}$$

с M_g-липшицевым выпуклым функциональным ограничением $g(x) \leqslant 0$. Далее будем полагать, что x_* — одно из решений поставленной задачи (33). Предлагаются алгоритмические методы зеркального спуска с новыми адаптивными критериями остановки, выполнение которых гарантирует достижение приемлемого качества решения, например (с точностью умножения на некоторые постоянные): $f(\hat{x}) - f(x_*) \leqslant \varepsilon$ и $g(\hat{x}) \leqslant \varepsilon$. Использование адаптивных критериев позволяет ускорить работу методов, а также в некоторых случаях применять их и для задач, для которых не удаётся установить приемлемые оптимальные оценки сложности. В случае негладкости целевого функционала или функциональных ограничений естественно использовать субградиентные методы с переключениями по продуктивным и непродуктивным шагам, восходящие к хорошо известным работам [59) 60)]. Зеркальные спуски для задач выпуклой оптимизации с ограничениями были предложены в [61)]. В работе [62)], предложены некоторые методы зеркального спуска для задач вида (33) с адаптивным выбором шагов и, самое важное, — с адаптивными критериями остановки. Глава 5 диссертации посвящена развитию некоторых идей указанной работы.

[59)] Polyak B. T. A general method of solving extremum problems, Sov. Math. Dokl., 8:3 (1967).

[60)] Шор Н. З. Применение обобщённого градиентного спуска в блочном программировании, Кибернетика, 1967, № 3, 53–55.

[61)] Немировский А. С., Юдин Д. Б. Сложность задач и эффективность методов оптимизации. М.: Наука, 1979. 384 с.

[62)] A. Bayandina, P. Dvurechensky, A. Gasnikov, F. Stonyakin, A. Titov. Mirror descent and convex optimization problems with non-smooth inequality constraint// Lecture Notes in Mathematics. 2018 Vol. 2227, P. 181-213.

Через $(E, \|\cdot\|)$ обозначим конечномерное нормированное векторное пространство и E^* — сопряженное пространство к E со стандартной нормой $\|y\|_* = \max\limits_x\{\langle y, x\rangle, \|x\| \leq 1\}$, где $\langle y, x\rangle$ — значение линейного непрерывного функционала y в точке $x \in E$. Пусть $Q \subset E$ — некоторое замкнутое выпуклое множество. Стандартно введём оператор проектирования

$$\mathrm{Mirr}_x(p) = \arg\min_{u \in Q} \left\{\langle p, u\rangle + V(u, x)\right\} \text{ для всяких } x \in Q \text{ и } p \in E^*$$

и сделаем предположение о том, что оператор $\mathrm{Mirr}_x(p)$ легко вычислим.

В **разделе 5.1** выполнено теоретическое исследование метода

Алгоритм 17 Адаптивный зеркальный спуск, разные условия гладкости целевого функционала.

if $g(x^N) \leqslant \varepsilon$ **then**
  $x^{N+1} = Mirr_{x^N}\left(\frac{\varepsilon \nabla f(x^N)}{\|\nabla f(x^N)\|_*}\right)$ // *"продуктивные шаги"* $(N \in I)$
else
  $x^{N+1} = Mirr_{x^N}\left(\frac{\varepsilon \nabla g(x^N)}{\|\nabla g(x^N)\|_*^2}\right)$ // *"непродуктивные шаги"* $(N \notin I)$
end if

Отметим, что алгоритм 17 применим к условным задачам выпуклой минимизации и позволяет получить оптимальные оценки скорости сходимости для достаточно широкого класса целевых функционалов в случае липшицева функционала ограничения. Например, в задачах с квадратичными целевыми функционалами мы сталкиваемся с ситуацией, когда такой функционал не удовлетворяет обычному свойству Липшица (или константа Липшица достаточно большая), но градиент удовлетворяет условию Липшица. Можно рассматривать и более широкий класс уже негладких целевых функционалов

$$f(x) = \max_{1 \leqslant i \leqslant m} f_i(x), \text{ где} \tag{34}$$

$$f_i(x) = \frac{1}{2}\langle A_i x, x\rangle - \langle b_i, x\rangle + \alpha_i,\ i = 1, \ldots, m, \tag{35}$$

в случае, когда A_i $(i = 1, \ldots, m)$ — положительно определённые матрицы: $x^T A_i x \geqslant 0$ $\forall x \in Q$. Отметим, что функционалы вида (34)–(35) вообще говоря, не удовлетворяют условию Липшица.

Показано, как можно оценить скорость сходимости предлагаемого метода. Для всякого ненулевого конечного субградиента $\nabla f(x)$ целевого функционала f рассматривается следующая величина

$$v_f(x, y) = \left\langle \frac{\nabla f(x)}{\|\nabla f(x)\|_*}, x - y \right\rangle, \quad x \in Q,$$

где x_* — искомое решение задачи (33). Случай $\nabla f(x) = 0$ здесь опускается, поскольку тогда x автоматически будет искомой точкой x_*. Пусть x^0 — начальное приближение и постоянная Θ_0 такова, что $V(x_*, x^0) \leqslant \Theta_0^2$. В таком случае для алгоритма 17 справедлива следующая

Теорема 5.1.1. *Пусть $\varepsilon > 0$ — фиксированное число и выполнен критерий остановки*
$$\frac{2\Theta_0^2}{\varepsilon^2} \leqslant \sum_{k \notin I} \frac{1}{\|\nabla g(x^k)\|_*^2} + |I|$$
алгоритма 17. Тогда $\min\limits_{k \in I} v_f(x^k, x_*) < \varepsilon$. Отметим, что алгоритм 17 работает не более
$$N = \left\lceil \frac{2\max\{1, M_g^2\}\Theta_0^2}{\varepsilon^2} \right\rceil \tag{36}$$
итераций.

Хорошо известно, что оценка (36) оптимальна с точностью умножения на константу на классе задач выпуклого программирования с липшицевым целевым функционалом или ограничением. С использованием теоремы 5.1.1 можно, в частности, оценить скорость сходимости алгоритма 17 для следующего класса, вообще говоря, негладких целевых функционалов.

Следствие 5.1.4. *Пусть $f(x) = \max\limits_{i=\overline{1,m}} f_i(x)$, где f_i дифференцируема для всякого $x \in Q$ и $\|\nabla f_i(x) - \nabla f_i(y)\|_* \leqslant L_i\|x - y\| \quad \forall x, y \in Q$. Тогда после остановки алгоритма 17 верны оценки:*
$$\min_{k \in I} f(x^k) - f(x_*) \leqslant \varepsilon \cdot \|\nabla f(x_*)\|_* + \frac{L\varepsilon^2}{2} \ \text{ и } \ \max_{k \in I} g(x^k) \leqslant \varepsilon, \ \text{ где } L = \max_{i=\overline{1,m}} L_i.$$

Доказано, что если задача (33) разрешима и её целевой функционал f удовлетворяет условию Гёльдера $|f(x) - f(y)| \leqslant M_\nu \|x - y\|^\nu \ \ \forall x, y \in Q$ для некоторого $\nu \in [0; 1)$, то для алгоритма 17 сохраняется асимптотически оптимальная оценка скорости сходимости $O(\varepsilon^{-2})$.

В **разделе 5.2** исследованы следующие две алгоритмические схемы [63] [64] с критериями проверки продуктивности шага, который связан с нормой субградиента ограничения в текущей точке и с соответствующими критериями остановки. Для проведённых экспериментов установлено, что такие методы могут

[63] F. Stonyakin, A. Stepanov, A. Gasnikov and A. Titov. Mirror Descent for Constrained Optimization Problems with Large Subgradient Values. // Компьютерные исследования и моделирование. 2020, № 2. С. 301 – 317.

[64] Ф. С. Стонякин, И. В. Баран. О некоторых алгоритмах для условных задач оптимизации с относительной точностью по целевому функционалу. // Труды ИММ УрО РАН. 2020. Т. 26, № 3. С. 198 – 210.

Таблица 2: Алгоритмы 19 – 20.

	Критерий остановки
Алгоритм 19	$2\frac{\Theta_0^2}{\varepsilon^2} \leqslant \sum\limits_{j\in I} \frac{1}{\|\nabla f(x^N)\|_*^2} + N - \lvert I \rvert$
Алгоритм 20	$2\frac{\Theta_0^2}{\varepsilon^2} \leqslant N$
	Оценки
Алгоритм 19	$f(\bar{x}^N) - f(x_*) \leqslant \varepsilon,\ \max\limits_{k\in I} g(x^k) \leqslant \varepsilon M_g$
Алгоритм 20	$\min\limits_{k\in I} v_f(x^k, x_*) \leqslant \varepsilon,\ \max\limits_{k\in I} g(x^k) \leqslant \varepsilon M_g$

работать быстрее по сравнению с предложенными в [65], если нормы субградиентов функционала ограничения достаточно велики.

Алгоритм 19 Адаптивный зеркальный спуск.

1: **if** $g(x^N) \leqslant \varepsilon\|\nabla g(x^N)\|_*$ **then**

2: $\quad x^{N+1} = Mirr_{x^N}\left(\frac{\varepsilon\nabla f(x^N)}{\|\nabla f(x^N)\|_*^2}\right)$ // *"продуктивные шаги"* $(N \in I)$

3: **else**

4: $\quad x^{N+1} = Mirr_{x^N}\left(\frac{\varepsilon\nabla g(x^N)}{\|\nabla g(x^N)\|_*}\right)$ // *"непродуктивные шаги"* $(N \notin I)$

5: **end if**

Ensure:

$$\hat{x} = \sum_{k\in I} h_k x^k / \sum_{k\in I} h_k.$$

Алгоритм 20 Адаптивный зеркальный спуск с фиксированным числом шагов.

1: **if** $g(x^N) \leqslant \varepsilon\|\nabla g(x^N)\|_*$ **then**

2: $\quad x^{N+1} = Mirr_{x^N}\left(\frac{\varepsilon\nabla f(x^N)}{\|\nabla f(x^N)\|_*}\right)$ // *"продуктивные шаги"* $(N \in I)$

3: **else**

4: $\quad x^{N+1} = Mirr_{x^N}\left(\frac{\varepsilon\nabla g(x^N)}{\|\nabla g(x^N)\|_*}\right)$ // *"непродуктивные шаги"* $(N \notin I)$

5: **end if**

Ensure: $\hat{x} = argmin_{x^k,\, k\in I}\, f(x^k)$.

В виде таблицы 2 ниже приведём сравнительный анализ полученных результатов для рассмотренных в разделе 5.2 алгоритмических схем.

[65] A. Bayandina, P. Dvurechensky, A. Gasnikov, F. Stonyakin, A. Titov. Mirror descent and convex optimization problems with non-smooth inequality constraint// Lecture Notes in Mathematics. 2018 Vol. 2227, P. 181-213.

В **разделе 5.3** показано, что оценки скорости сходимости для алгоритмов 17 и 20 сохраняются в случае, если f — квазивыпуклый функционал. При этом в качестве аналога субградиента мы рассматриваем элементы $\nabla_{Cl} f(x)$ субдифференциала Кларка в предположении локальной липшицевости целевого функционала, а также $\nabla_{Cl} f(x) \neq 0$ при $x \neq x_*$.

В **разделе 5.4** предложены схемы зеркального спуска с переключениями, которые оптимальны для некоторых классов сильно выпуклых задач негладкой оптимизации. Рассмотрим задачу (33) для μ-сильно выпуклых функционалов f и g с одинаковым параметром $\mu > 0$. Несколько модифицируем предположения на прокс-функцию и допустим, что $d(x)$ ограничена на единичном шаре относительно выбранной нормы $\|\cdot\|$:

$$d(x) \leqslant \Theta_0^2, \quad \forall x \in Q : \|x\| \leqslant 1.$$

Также предположим, что для начальной точки $x^0 \in Q$ существует такое $R_0 > 0$, что $\|x_0 - x_*\|^2 \leqslant R_0^2$. Для построения методов решения сильно выпуклой задачи (33) использована идея рестартов (перезапусков) алгоритмов 17 – 20.

В качестве примера приведём оценку для следующего класса негладких сильно выпуклых целевых функционалов. Пусть

$$f(x) = \max_{i=\overline{1,m}} f_i(x), \tag{37}$$

где f_i дифференцируемы во всякой точке $x \in Q$ и имеют липшицев градиент, т.е. существуют $L_i > 0$ такие, что $\|\nabla f_i(x) - \nabla f_i(y)\|_* \leqslant L_i \|x - y\|$ при всяких $x, y \in Q$. Рассмотрим функцию $\tau : \mathbb{R}^+ \to \mathbb{R}^+$:

$$\tau(\delta) = \max \left\{ \delta \|\nabla f(x_*)\|_* + \frac{\delta^2 L}{2},\ \delta \right\}, \quad \text{где } L = \max_{i=\overline{1,m}} \{L_i\}.$$

Ясно, что функция τ возрастает, $\tau(0) = 0$ и поэтому для всякого $\varepsilon > 0$ существует $\hat{\varphi}(\varepsilon) > 0 :\ \tau(\hat{\varphi}(\varepsilon)) = \varepsilon$. Рассмотрим[66)] следующий метод и результаты об оценке скорости сходимости для задачи (33) при сделанных допущениях.

[66)] Ф. С. Стонякин, М. Алкуса, А. Н. Степанов, А. А. Титов. Адаптивные алгоритмы зеркального спуска для задач выпуклой и сильно выпуклой оптимизации с функциональными ограничениями. // Дискретн. анализ и исслед. опер. 2019. Т. 26, № 3. С. 88–114.

Алгоритм 23 Адаптивный алгоритм зеркального спуска для сильно выпуклых функционалов.

Require: $\varepsilon > 0$; начальная точка x^0; Θ_0 удовл. $d(x) \leqslant \Theta_0^2 \quad \forall x \in Q : \|x\| \leqslant 1$; X; $d(\cdot)$; параметр сильной вып. μ; R_0 удовл. $\|x^0 - x_*\|^2 \leqslant R_0^2$.

1: Set $d_0(x) = d\left(\frac{x-x^0}{R_0}\right)$.
2: Set $p = 1$.
3: **repeat**
4: Set $R_p^2 = R_0^2 \cdot 2^{-p}$.
5: Set $\varepsilon_p = \frac{\mu R_p^2}{2}$.
6: Set x^p — выход алгоритма 17 с точностью $\hat{\varphi}(\varepsilon_p)$, прокс-функцией $d_{p-1}(\cdot)$ и Θ_0^2.
7: $d_p(x) \leftarrow d\left(\frac{x-x^p}{R_p}\right)$.
8: Set $p = p + 1$.
9: **until** $p > \log_2 \frac{\mu R_0^2}{2\varepsilon}$.

Теорема 5.4.4. *Пусть f удовлетворяет* (37)*. Если f и g — μ-сильно выпуклые функционалы на $Q \subset \mathbb{R}^n$ и $d(x) \leqslant \Theta_0^2$ для всех $x \in Q$ таких, что $\|x\| \leqslant 1$. Пусть начальное приближение $x^0 \in Q$ и число $R_0 > 0$ заданы так, что $\|x^0 - x_*\|^2 \leqslant R_0^2$. Тогда для $\widehat{p} = \left\lceil \log_2 \dfrac{\mu R_0^2}{2\varepsilon} \right\rceil$ выход $x_{\widehat{p}}$ есть ε-решение задачи* (33)*, а также верно $\|x^{\widehat{p}} - x_*\|^2 \leqslant \frac{2\varepsilon}{\mu}$. При этом количество итераций алгоритма 17 при работе алгоритма 23 согласно пункту 6 листинга не превышает*

$$\widehat{p} + \sum_{p=1}^{\widehat{p}} \frac{2\Theta_0^2 \max\{1, M_g^2\}}{\hat{\varphi}^2(\varepsilon_p)}, \quad \text{где } \varepsilon_p = \frac{\mu R_0^2}{2^{p+1}}.$$

Замечание 5.4.5. *Предыдущую оценку количества итераций работы алгоритма 17 при работе алгоритма 23 (рестарты) можно несколько конкретизировать при условии $\varepsilon < 1$. В этом случае при всяком $\delta < 1$ имеем $\tau(\delta) \leqslant C\delta$ для некоторой константы C. Поэтому можно считать, что $\hat{\varphi}(\varepsilon) = \widehat{C} \cdot \varepsilon$ для соответствующей константы $\widehat{C} > 0$ и с точностью до умножения на константу имеем:*

$$N \leqslant \hat{p} + \frac{64\Theta_0^2 \max\{1, M_g^2\}}{\mu\varepsilon},$$

что с точностью до постоянного множителя соответствует оптимальной нижней оценке на указанном классе задач.

В **разделе 5.5** описан адаптивный метод зеркального спуска для задач онлайн-оптимизации для выпуклых липшицевых целевых функционалов с липшицевым функциональным ограничением [67], обоснована оценка скорости сходимости и оптимальность метода с точки зрения известных нижних оценок на соответствующем классе задач.

В **разделе 5.6** обсуждаются приложения построенной теории к некоторым прикладным задачам. Например, показано, как предложенные методы зеркального спуска с переключениями могут быть применены к задаче распределения ресурсов, в частности к задаче оптимизации высоконагруженной компьютерной сети. Уточним постановку задачи[68] [69] [70]. Допустим, что имеется компьютерная сеть с n пользователями (узлами), которые обмениваются пакетами через фиксированный набор m соединений. Структура сети задана матрицей маршрутизации $C = (C_i^j) \in \mathbb{R}^{m\times n}$, столбцы которой $\mathbf{C}_i \neq 0$, $i = 1, \ldots, n$ есть булевы m-мерные векторы такие, что $C_i^j = 1$ в случае использования узлом i соединения j, в противном случае $C_i^j = 0$. Ограничения на пропускную способность соединений задаются вектором $\mathbf{b} \in \mathbb{R}_+^m$ со строго положительными компонентами. Пользователи оценивают качество работы сети с помощью функций полезности $u_p(x_p)$, $p = 1, \ldots, n$, где $x_p \in \mathbb{R}_+$ — скорость передачи данных p–го пользователя. Согласно [71] в качестве критерия оптимальности системы можно принять сумму функций полезностей для всех пользователей и тогда задача максимизации суммарной полезности сети при заданных ограничениях на пропускную способность соединений формулируется следующим образом:

$$\max_{\left\{C\mathbf{x}=\sum\limits_{p=1}^{n}\mathbf{C}_p x_p\right\}\leq \mathbf{b}} \left\{U(\mathbf{x}) = \sum_{k=1}^{n} u_p(x_p)\right\},$$

где $\mathbf{x} = (x_1, \ldots, x_n)$. Решением данной задачи будет оптимальное распределение ресурсов x_*. Отметим, что целевые функционалы тут могут быть самого разного уровня гладкости, в том числе даже формально не удовлетворяющие условию Липшица: $u_p(x_p) = \ln x_p$ или $u_p(x_p) = \sqrt{x_p}$. Использование для такой задачи разработанных субградиентных схем с переключениями по продуктивным и непродуктивным шагам представляется более выгодными в случае большого количества соединений m, так как оценки скорости сходимости не

[67] A. A. Titov, F. S. Stonyakin, A. V. Gasnikov, M. S. Alkousa. Mirror descent and constrained online optimization problems. // 9th International Conference on Optimization and Applications, OPTIMA-2018. Communications in Computer and Information Sciences. 2019. Vol. 974. P. 64 – 78.

[68] Рохлин Д. Б. Распределение ресурсов в сетях связи с большим числом пользователей: стохастический метод градиентного спуска. // Теория вероятностей и её применения, 2020 (В печати).

[69] Ivanova A., Stonyakin F., Pasechnyuk D, Vorontsova E., Gasnikov A. Adaptive Mirror Descent for the Network Utility Maximization Problem. https://arxiv.org/abs/1911.07354v1. IFAC Congress 2020.

[70] Ф.С. Стонякин. Адаптивные алгоритмические методы в негладкой оптимизации. — Симферополь: «ПОЛИПРИНТ», 2020.

[71] Kelly F. P., Maulloo A. K. and Tan D. K. H. Rate control for communication networks: shadow prices, proportional fairness and stability. // J. Oper. Res Soc., 1998. Vol. 49, № 3. P. 237 – 252.

зависят от размерности задачи. При этом показано, что алгоритмы 17 и 20 приводят к оптимальным оценкам сложности для гельдеровых функций полезности. Для логарифмической функции полезности $u_p(x_p) = \ln x_p$ можно модифицировать допустимое множество задачи отступами от нуля по значениям переменных ($x_p \geqslant n\varepsilon$) и показать оценку сложности $O(\varepsilon^{-4})$ для алгоритмов 17, 19 и 20. Если же для некоторого $R > 0$ добавить условие ограниченности $\|x\|_2 \leqslant R$ и оценку сложности $O(\varepsilon^{-4})$ для алгоритмов 17 и 20 можно заменить на $O(n^4\varepsilon^{-2}\ln^4 n\varepsilon)$. Однако при этом оценка сложности зависит от величины, соответствующей количеству пользователей. Отметим при этом, что для проведённых экспериментов получилось, что алгоритм 19 за счёт адаптивного критерия остановки работает существенно быстрее по сравнению с алгоритмами 17 и 20.

Обсуждаются результаты численных экспериментов для других задач (в частности, аналоги задачи Ферма-Торричелли-Штейнера и задачи о наименьшем покрывающем шаре) с ограничениями по сравнению скоростей работы алгоритмов 17, 19 и 20 между собой, а также с некоторыми аналогами [72)73)]. По результатам проведённых экспериментов оказалось, что для алгоритма 19 за счёт предложенного адаптивного критерия остановки приемлемое качество решения достигается в несколько раз быстрее субградиентного метода[74)].

В **разделе 5.7** рассмотрены приложения полученных результатов о зеркальных спусках к двум специальным классам задач. Во-первых, вместо условия Липшица для целевого функционала f и ограничения g рассмотрены неравенства $\langle \nabla f(x), x-y\rangle \leqslant M_f\sqrt{2V(y,x)}$ и $\langle \nabla g(x), x-y\rangle \leqslant M_g\sqrt{2V(y,x)}$ для не обязательно 1-сильно выпуклой прокс-функции и дивергенции Брэгмана. Эти условия покрывают, в частности, класс так называемых относительно непрерывных (точнее, относительно липшицевых) функционалов [75)] Рассмотрен частично адаптивный аналог алгоритма 19 при таких общих предположениях [76)]. Отметим, что использование таких общих предположений (относительной липшицевости) для функционала ограничения g позволяет расширить границы применимости по сравнению с упомянутыми ранее алгоритмами 17 – 20, оптимальность оценок для которых обоснована в предыдущих разделах диссертации

[72)] Nesterov Y. Subgradient methods for Huge-Scale Optimization Problems // Math. Prog., 2015. — Vol. 146, no. 1–2. — P. 275–297.

[73)] Lagae S. New efficient techniques to solve sparse structured linear systems, with applications to truss topology optimization. Ecole polytechnique de Louvain, Université catholique de Louvain, 2017.

[74)] Nesterov Y. Subgradient methods for Huge-Scale Optimization Problems // Math. Prog., 2015. — Vol. 146, no. 1–2. — P. 275–297.

[75)] Lu H. "Relative-Continuity"for Non-Lipschitz Non-Smooth Convex Optimization using Stochastic (or Deterministic) Mirror Descent. // arXiv:1710.04718 (2018). `https://arxiv.org/abs/1710.04718v3`.

[76)] Ф.С. Стонякин. Адаптивные алгоритмические методы в негладкой оптимизации. — Симферополь: «ПОЛИПРИНТ», 2020.

лишь в предположении липшицевости g.

Важная часть **раздела 5.7** посвящена приложениям полученных результатов о зеркальных спусков с переключениями к задаче минимизации (вообще говоря, негладкого) выпуклого однородного функционала с относительной точностью при наличии функциональных ограничений. Такая постановка восходит к работам [77] [78]. Как показано в упомянутых выше работах, подход к оценке качества решения задачи в с точки зрения именно относительной точности вполне оправдан для разных прикладных задач (линейное программирование, проектирование механических конструкций и др.), если нет необходимости слишком точно решать поставленную задачу. Известно, что достаточно широкий класс задач оптимизации с относительной точностью можно сводить к минимизации выпуклой однородной функции. Были рассмотрены на выпуклом замкнутом множестве $Q \subset \mathbb{R}^n$ задача минимизации выпуклой однородной функции вида

$$f(x) \to \min_{x \in Q} \tag{38}$$

с выпуклым функциональным ограничением $g(x) \leqslant 0$. В разделе 5.7 показано, что к такому классу задач можно применять разработанные адаптивные субградиентные схемы с переключениями по продуктивным и непродуктивным шагам. Это позволяет рассматривать более широкий класс функционалов ограничений, а также избегать дополнительной задачи проектирования при выборе начальной точки x^0. Сначала рассмотрено предположение упомянутых выше работ Ю.Е. Нестерова о том, что 0 есть внутренняя точка субдифференциала $\partial f(0)$. Как известно (см. теорему 6.1.1 из докторской диссертации Ю.Е. Нестерова), при таком предположении для некоторого $\gamma_0 > 0$ верно неравенство $\|x^0 - x_*\| \leqslant \frac{2}{\gamma_0} f^*$. Получена следующая [79]

Теорема 5.7.3. *Пусть выпуклый однородный функционал f M_f-липшицев на Q при некотором $M_f > 0$, 0 — внутренняя точка субдифференциала $\partial f(0)$, а также для некоторого $C > 0$ верно $2V(x_*, x^0) \leqslant C^2 \|x_* - x^0\|^2$ и начальная точка x^0 выбрана так, что $\|x^0\| := \min_{x \in Q}\{\|x\|\}$. Тогда для всякого $\delta > 0$ можно подобрать входные параметры $\varepsilon > 0$ и $\Theta_0 > 0$ алгоритма 19 так, что после*

$$N \geqslant \frac{4C^2 \max\{1, M_f^2\}}{\gamma_0^2 \delta^2}$$

[77] Nesterov Yu. Rounding of convex sets and efficient gradient methods for linear programming problems // Optimization Methods and Software. 2008. Vol.23, N1. P.109-128.

[78] Nesterov Yu. Unconstrained Convex Minimization in Relative Scale // Mathematics of Operations Research. 2009.Vol.34, N 1. P.180-193.

[79] Ф.С. Стонякин, И.В. Баран. О некоторых алгоритмах для условных задач оптимизации с относительной точностью по целевому функционалу. // Труды ИММ УрО РАН. 2020. Т. 26, № 3. С. 198 – 210.

итераций этого метода гарантированно будут выполнены следующие неравенства: $f(\widehat{x}) \leqslant f^*(1+\delta)$ *и* $g(\widehat{x}) \leqslant \delta f^* \|\nabla g(\widehat{x})\|_*$.

Далее, получены аналоги результаты теоремы 5.7.3 в существенно более общей ситуации, когда 0 не обязательно есть внутренняя точка субдифференциала целевого функционала f в нуле. Точнее говоря, рассмотрен следующий ослабленный вариант этого условия

$$B_{\gamma_0}^{K^*}(0) \subseteq \partial f(0) \subseteq B_{\gamma_1}^{K^*}(0), \tag{39}$$

где K^* — сопряженный конус к некоторому полунормированному конусу $K \subset \mathbb{R}^n$ с конус-полунормой $\|\cdot\|_K$ (отличие от обычной полунормы в том, что $\|\alpha x\|_K = \alpha \|x\|_K$ лишь для $\alpha \geqslant 0$), где под *сопряженным конусом* K^* понимается набор функционалов вида $\psi_\ell = \max\{0, \ell(x)\}$ для линейных функционалов $\ell : K \to \mathbb{R} : \ell(x) \leqslant C_\ell \|x\|_K$ при некотором $C_\ell > 0$ $\forall\, x \in K$. Ясно, что K^* будет выпуклым конусом с операциями сложения $\psi_{\ell_1} \oplus \psi_{\ell_2} := \psi : \quad \psi(x) = \max\{0, \ell_1(x) + \ell_2(x)\}$ и умножения на скаляр $\lambda \geqslant 0$ $\psi_{\lambda\ell}(x) = \lambda\psi_\ell(x) = \lambda \max\{0, \ell(x)\}$ $\forall\, x \in K$. На K^* можно ввести норму $\|\psi_\ell\|_{K^*} = \sup_{\|x\|_K \leqslant 1} \max\{0, \ell(x)\} = \sup_{\|x\|_K \leqslant 1} \ell(x)$ и шар $B_r^{K^*}(0) = \{\psi_\ell \in K^* \mid \|\psi_\ell\|_{K^*} \leqslant r\}$. Из аналога теоремы об опорном функционале в нормированных конусах [80] получаем, что $\|x\|_K = \max_{\psi_\ell \in B_1^{K^*}(0)} \ell(x)$.

Не уменьшая общности рассуждений, будем полагать $K = \bigcup\limits_{r \geqslant 0} B_r^K(0)$, а также $x_* \in K$ для точного решения x_* рассматриваемой задачи минимизации f на Q. Для вывода оценок скорости сходимости методов с относительной точностью необходимо знать оценку R величины расстояния от точки старта x^0 до искомого решения x_*. Однако в выпуклых конусах, вообще говоря, не задана операция вычитания и поэтому в качестве аналога нормы разности можно использовать метрику $d^K(x^0, x_*)$, где

$$d^K(x,y) = \sup_{\|\psi_\ell\|_{K^*} \leqslant 1} |\psi_\ell(x) - \psi_\ell(y)|.$$

Некоторые условия, при которых нормированный конус допускает существование метрики такого типа, исследованы в [81] [82]. Получен аналог теоремы 6.1.1 из докторской диссертации Ю.Е. Нестерова для указанного выше предположения (39) ($x^0, x_* \in K$, причём $\|x^0\|_K = \min\{\|x\|_K,\ x \in Q\}$).

[80] Stonyakin F.S. An analogue of the Hahn–Banach theorem for functionals on abstract convex cones.// Eurasian Math. J. 2016. Vol. 7, no. 3. P. 89–99.

[81] Ф. С. Стонякин. Сублинейный аналог теоремы Банаха–Мазура в отделимых выпуклых конусах с нормой. // Матем. заметки. 2018. Т. 104, № 1. С. 118 — 130.

[82] F.S. Stonyakin. Hahn–Banach type theorems on functional separation for convex ordered normed cones. // Eurasian Math. J. 2019. Vol. 10. No. 1. P. 59 -– 79.

Теорема 5.7.8. 1) $\forall\, x \in K \quad \gamma_0\|x\|_K \leqslant f(x) \leqslant \gamma_1\|x\|_K$. *Более того,*

$$\frac{\gamma_0}{\gamma_1} f(x^0) \leqslant \gamma_0\|x^0\|_K \leqslant f^* \leqslant f(x^0) \leqslant \gamma_1\|x^0\|_K.$$

2) *Для всякого точного решения* $x_* \in K$ *справедливо неравенство:*

$$d^K(x^0, x_*) \leqslant \|x^0\|_K + \|x_*\|_K \leqslant \frac{2}{\gamma_0} f^* \leqslant \frac{2}{\gamma_0} f(x^0).$$

Для применимости к поставленной задаче выпуклой однородной минимизации приведенного выше адаптивного алгоритма 19 достаточно выбрать прокс-структуру так, чтобы для некоторой константы $\hat{\omega} > 0$ было верно

$$V(x_*, x^0) \leqslant \hat{\omega} d^K(x_*, x^0). \tag{40}$$

Если верно (40) и $\hat{\omega} d^K(x_*, x^0) = \Theta_0^2$, то для задачи (38) справедлива следующая

Теорема 5.7.11. *Пусть выпуклый однородный функционал f M_f-липшицев для некоторого $M_f > 0$. Тогда после $N \geqslant \frac{8\max\{1, M_f^2\}}{\gamma_0^2\delta^2}$ итераций алгоритма 19 гарантированно будут верны неравенства:*

$$f(\widehat{x}) \leqslant f(x_*)(1+\delta) \text{ и } g(\widehat{x}) \leqslant \frac{M_g\Theta_0^2\gamma_0\delta}{2}.$$

Ясно, что в случае $0 \in Q$ можно просто выбрать $x^0 = 0$, что позволяет избегать необходимости дополнительно решать вспомогательную задачу минимизации нормы для выбора начальной точки x^0. Степень соблюдения условий на функциональные ограничения задачи при этом предлагается проверять с помощью неравенств-проверки продуктивности итераций зеркального спуска. Заметим также, что для предыдущего результата существенен выбор начальной точки x^0. Иными словами, в случае $0 \notin Q$ важно быть уверенным в существовании точки, реализующей минимально возможное значение конус-полунормы $\|\cdot\|_K$ на множестве Q. Некоторые условия разрешимости задач такого типа в абстрактных нормированных конусах для множества Q получены в [83)] и обсуждаются в завершающих трёх подпунктах раздела 5.7 диссертации.

В **заключительных замечаниях к главе 5** подведены итоги описанных ней исследований и рассмотрены модификации алгоритмов 17 – 20 в случае, если вместо субградиентов целевого функционала и функционала ограничения использовать δ-субградиенты.

В **заключении** к диссертаци сформулированы основные выводы диссертации и сформулированы дальнейшие возможные направления развития проведённых исследований.

[83)] Ф. С. Стонякин. О сублинейных аналогах слабых топологий в нормированных конусах. // Матем. заметки. 2018. Т. 103, № 5. С. 794 — 800.

Основные публикации автора по материалам диссертации

1. А. В. Гасников, П. Е. Двуреченский, Ф. С. Стонякин, А. А. Титов. Адаптивный проксимальный метод для вариационных неравенств. // Журнал вычислит. мат. и мат. физики. 2019. Т. 59, № 5. С.889 -– 894.

2. F. S. Stonyakin. On the Adaptive Proximal Method for a Class of Variational Inequalities and Related Problems. // Proceedings of the Steklov Institute of Mathematics. 2020. Vol. 309(1). P.S139 – S150.

3. Ф.С. Стонякин. Адаптивный аналог метода Ю. Е. Нестерова для вариационных неравенств с сильно монотонным оператором. // Сибирский журнал вычислит. матем. 2019. Т. 22, № 2. С. 201 -– 211.

4. F. Stonyakin, E. Vorontsova and M. Alkousa. New Version of Mirror Prox for Variational Inequalities with Adaptation to Inexactness. // 10th International Conference on Optimization and Applications, OPTIMA-2019. Communications in Computer and Information Sciences. 2020. Vol. 1145. P. 427 – 442.

5. Ф. С. Стонякин. Адаптация к величинам погрешностей для некоторых методов оптимизации градиентного типа. // Труды ИММ УрО РАН. 2019. Т. 25, № 4. С. 210 — 225.

6. Ф. С. Стонякин, М. Алкуса, А. Н. Степанов, А. А. Титов. Адаптивные алгоритмы зеркального спуска для задач выпуклой и сильно выпуклой оптимизации с функциональными ограничениями// Дискретн. анализ и исслед. опер. 2019. Т. 26, № 3. С. 88 — 114.

7. Д. А. Пасечнюк, Ф. С. Стонякин. Об одном методе минимизации выпуклой липшицевой функции двух переменных на квадрате. // Компьютерные исследования и моделирование. 2019. Т.11, № 3. С.379 — 395.

8. F. S. Stonyakin, M. S. Alkousa, A. A. Titov and V. V. Piskunova. On Some Methods for Strongly Convex Optimization Problems with One Functional Constraint. //In: M. Khachay et al. (Eds.): MOTOR 2019. Lecture Notes in Computer Science. 2019. Vol. 11548. P. 82 – 96.

9. F. S. Stonyakin, D. Dvinskikh, P. Dvurechensky, A. Kroshnin, O. Kuznetsova, A. Agafonov, A. Gasnikov, A. Tyurin, C. A. Uribe, D. Pasechnyuk, S. Artamonov. Gradient Methods for Problems with Inexact Model of the Objective. // In: M. Khachay et al. (Eds.): MOTOR 2019. Lecture Notes in Computer Science. 2019. Vol. 11548. P. 97 – 114.

10. A. A. Titov, F. S. Stonyakin, A. V. Gasnikov, M. S. Alkousa. Mirror descent and constrained online optimization problems. // 9th International Conference on Optimization and Applications, OPTIMA-2018. Communications in Computer and Information Sciences. 2019. Vol. 974. P. 64 – 78.

11. F. S. Stonyakin, A. A. Titov. One Mirror Descent algorithm for convex constrained optimization problems with non-standard growth properties.// School-

Seminar on Optimization Problems and their Applications, OPTA-SCL 2018. CEUR-WS 2018, Vol. 2098. P. 372 – 384.

12. A. Bayandina, P. Dvurechensky, A. Gasnikov, F. Stonyakin, A. Titov. Mirror descent and convex optimization problems with non-smooth inequality constraint. // Lecture Notes in Mathematics. 2018 Vol. 2227, P. 181 – 213.

13. Ф. С. Стонякин, М. С. Алкуса, А. Н. Степанов, М. А. Баринов. Адаптивные алгоритмы зеркального спуска в задачах выпуклого программирования с липшицевыми ограничениями. // Труды ИММ УрО РАН. 2018. Т. 24, № 2. С. 266 -– 279.

14. А. Д. Агафонов, Ф. С. Стонякин. Градиентные методы для задач оптимизации, допускающие существование неточной сильно выпуклой модели целевой функции. // Труды МФТИ. 2019. Том 11 (44), № 3. С. 4 – 19.

15. F. Stonyakin, A. Stepanov, A. Gasnikov and A. Titov. Mirror Descent for Constrained Optimization Problems with Large Subgradient Values. // Компьютерные исследования и моделирование. 2020, Т. 12, № 2. — С. 301 – 317.

16. Ф. С. Стонякин, И. В. Баран. О некоторых алгоритмах для условных задач оптимизации с относительной точностью по целевому функционалу. // Труды ИММ УрО РАН. 2020. Т. 26, № 3. С. 198 – 210.

17. F. S. Stonyakin. Hahn–Banach type theorems on functional separation for convex ordered normed cones. // Eurasian Math. J. Vol. 10, no. 1, 2019. P. 59 -– 79.

18. Ф. С. Стонякин. Сублинейный аналог теоремы Банаха–Мазура в отделимых выпуклых конусах с нормой. // Матем. заметки. 2018. Т.104, № 1. С. 118 -– 130.

19. Ф. С. Стонякин. О сублинейных аналогах слабых топологий в нормированных конусах. // Матем. заметки. 2018. Т. 103, № 5. С. 794 -– 800.

20. F. S. Stonyakin. Subdifferential calculus in abstract convex cones. // Constructive Nonsmooth Analysis and Related Topics (Dedicated to the Memory of V.F. Demyanov). Proceedings of CNSA – 2017. — P. 316 – 319.

21. F. S. Stonyakin. An analogue of the Hahn–Banach theorem for functionals on abstract convex cones. // Eurasian Math. J. Vol. 7, no. 3, 2016. P. 89 -– 99.

22. P. Dvurechensky, A. Gasnikov, E. Nurminsky and F. Stonyakin. Advances in Low-Memory Subgradient Optimization. // In: A. M. Bagirov et al.(eds.). Numerical Nonsmooth Optimization. State of the Art Algorithms. Springer Nature Switzerland AG 2020. P. 19 – 59.