

Том 64, Номер 4

ISSN 0044-4669

Апрель 2024



ФИЦ ИУ РАН

# ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ



НАУКА

— 1727 —

# СОДЕРЖАНИЕ

---

---

Том 64, номер 4, 2024 год

---

---

Памяти профессора Федора Павловича Васильева (1935–2023)  
*Л. А. Артемьева, Б. А. Будаков, А. А. Дряженков, М. М. Потапов* 565

Памяти профессора Бориса Теодоровича Поляка (1935–2023)  
*М. В. Хлебников* 571

## ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ

Стохастический градиентный спуск с предобусловленным размером шага им. Б. Т. Поляка  
*Ф. Абдухакимов, Ч. Сян, Д. Камзолов, М. Такач* 575

О некоторых работах Бориса Теодоровича Поляка по сходимости градиентных методов  
и их развитии  
*С. С. Аблаев, А. Н. Безносиков, А. В. Гасников, Д. М. Двинских, А. В. Лобанов, С. М. Пучинин,  
Ф. С. Стоякин* 587

Метод Б. Т. Поляка на основе стохастической функции Ляпунова для обоснования  
состоятельности оценок поискового алгоритма стохастической аппроксимации  
при неизвестных, но ограниченных помехах  
*О. Н. Граничин, Ю. В. Иванский, К. Д. Копылова* 627

Об избыточности невырожденности Гесса для геометрической скорости сходимости  
метода Ньютона при минимизации выпуклых функций  
*Ю. Г. Евтушенко, А. А. Третьяков* 637

Алгоритмы управления с итеративным обучением высшего порядка для линейных систем  
*П. В. Пакишин, Ю. П. Емельянова, М. А. Емельянов* 644

## МАТЕМАТИЧЕСКАЯ ФИЗИКА

Определение коэффициента теплопроводности и объемной теплоемкости вещества  
по тепловому потоку  
*А. Ю. Горчаков, В. И. Зубов* 658

## ИНФОРМАТИКА

Метод поочередных проекций для пересечения выпуклых множеств, многоагентные алгоритмы  
консенсуса и усредняющие неравенства  
*А. В. Проскурников, И. С. Забарянская* 671

---

---

# CONTENTS

---

---

**Vol. 64, issue 4, 2024**

---

---

- In memory of Prof. Fyodor Pavlovich Vasiliev (1935–2023)  
*L. A. Artemyeva, B. A. Budak, A. A. Dryazhenkov, M. M. Potapov* 565
- In memory of Prof. Boris Theodorovich Polyak (1935–2023)  
*M. V. Khlebnikov* 571

## OPTIMAL CONTROL

- Stochastic gradient descent with preconditioned Polyak step-size  
*F. Abdukhakimov, Ch.Xiang, D. Kamzolov, M. Takáč* 575
- On some works of Boris Teodorovich Polyak on the convergence of gradient methods and their development  
*S. S. Ablaev, A. N. Beznosikov, A. V. Gasnikov, D. M. Dvinskikh, A.V. Lobanov, S. M. Puchinin, F. S. Stonyakin* 587
- Polyak's method based on the stochastic Lyapunov function for justifying the consistency of estimates produced by a stochastic approximation search algorithm under an unknown-but-bounded noise  
*O. N. Granichin, Yu. V. Ivanskii, K. D. Kopylova* 627
- On the redundancy of hessian nonsingularity for linear convergence rate of the Newton method applied to the minimization of convex functions  
*Yu. G. Evtushenko, A. A. Tret'yakov* 637
- Higher-order iterative learning control algorithms for linear systems  
*P. V. Pakshin, J. P. Emelianova, M. A. Emelianov* 644

## MATHEMATICAL PHYSICS

- Determination of the thermal conductivity and volumetric heat capacity of substance from heat flux  
*A. Yu. Gorchakov, V. I. Zubov* 658

## INFORMATICS

- Alternating projection method for intersection of convex sets, multi-agent consensus algorithms, and averaging inequalities  
*A.V. Proskurnikov, I. S. Zabarianska* 671
- 
-

*Этот номер посвящен памяти члена Редколлегии журнала профессора Ф. П. Васильева и профессора Б. Т. Поляка, который был многолетним автором и рецензентом нашего журнала. Выпуск составлен из работ учеников и коллег Ф. П. Васильева и Б. Т. Поляка.*

УДК 519.653:519.658



## ПАМЯТИ ПРОФЕССОРА ФЕДОРА ПАВЛОВИЧА ВАСИЛЬЕВА (1935–2023)

© 2024 г. Л. А. Артемьева\*, Б. А. Будак, А. А. Дряженков, М. М. Потапов

*119991 Москва, Ленинские горы, 1, стр. 52, Факультет ВМК МГУ*

*\*e-mail: artemieva.luda@gmail.com*

Поступила в редакцию 03.07.2023 г.

Переработанный вариант 03.07.2023 г.

Принята к публикации 20.11.2023 г.

5 февраля 2023 г. на 88-м году жизни скончался известный российский математик, доктор физико-математических наук, профессор кафедры оптимального управления факультета вычислительной математики и кибернетики, заслуженный профессор Московского государственного университета Федор Павлович Васильев. Ушел из жизни крупный ученый, выдающийся педагог и замечательный человек.

Федор Павлович родился 24 июля 1935 г. в селе Суккулово Ермекеевского района Башкирской АССР в семье сельского учителя Васильева Павла Агаповича, 1898 года рождения, и колхозницы Анисии Ивановны, 1899 г. рождения. Отца он лишился рано, во время войны, когда учился в Суккуловской семилетней школе (1942–1949); Павел Агапович погиб на фронте в январе 1943 г. в боях под Ленинградом. Мать Анисия Ивановна умерла в 1965 г.

С 1949 по 1953 г. он учился в Белебеевском педучилище, окончил его с отличием, после чего с 1953 г. по 1955 г. работал по распределению учителем физики и математики в семилетних школах Чарджоуской области Туркменской ССР.

Любовь к математике и стремление к более глубокому ее познанию привели Федора Павловича на механико-математический факультет Московского государственного университета им. М. В. Ломоносова, куда он поступил в 1955 г. и окончил в 1960 г. с отличием. В студенческие годы ему посчастливилось учиться у многих выдающихся педагогов, из которых он с особой теплотой вспоминал Петра Лаврентьевича Ульянова (математический анализ), Анну Петровну Мишину (высшая алгебра) и Алексея Серапионовича Пархоменко (аналитическая геометрия).

С 1960 по 1963 г. Федор Павлович учился в аспирантуре механико-математического факультета, где под руководством доцента Бориса Михайловича Будака подготовил и успешно защитил в 1964 г. диссертацию «Разностный метод решения задач типа Стефана для квазилинейного параболического уравнения с разрывными коэффициентами», став кандидатом физико-математических наук.

Научная деятельность Федора Павловича Васильева развивалась непрерывно, успешно и плодотворно. В 1986 г. он защитил диссертацию «Методы решения неустойчивых экстремальных задач с неточно заданными исходными данными» на соискание ученой степени доктора физико-математических наук. Всего им было опубликовано свыше 250 научных работ, в том числе 18 монографий, учебников и учебных пособий. Основные научные результаты были получены Федором Павловичем в области методов решения краевых задач с неизвестной границей для параболических уравнений, задач быстрогодействия для управляемых систем в банаховых пространствах при наличии фазовых ограничений, по обобщению метода моментов для задач оптимального управления системами с сосредоточенными и распределенными параметрами при наличии линейных и квадратичных ограничений, методов решения двойственных задач управления и наблюдения для распределенных систем, методов регуляризации различных классов неустойчивых задач минимизации и задач равновесного программирования и регуляризации базовых итерационных методов решения таких задач, в которых неточно заданы не только целевая функция, но и допустимое множество, на котором ищется точка минимума или точка равновесия.

Ко всем исследованиям Федор Павлович Васильев активно привлекал своих учеников, 27 из которых стали кандидатами физико-математических наук. Впоследствии 9 из них успешно защитили докторские диссертации: 8 — по физико-математическим наукам и 1 — по философии. Под научным руководством Федора Павловича в разное время стажировались 4 математика из Югославии. По полученным во время этих стажировок результатам все они затем защитили в Белградском университете докторские диссертации, а один из них — М. Ячимович — был избран академиком АН Черногории.

Большое внимание Федор Павлович Васильев уделял организации научной математической деятельности. В течение длительного времени он был членом Экспертного совета по математике при ВАК СССР (1985—1990) и Экспертного совета по математике и механике ВАК РФ (1990—2014), а также входил в состав Экспертного совета по информатике при Российском фонде фундаментальных исследований (1993—1999). Федор Павлович являлся членом 3 диссертационных советов в МГУ имени М. В. Ломоносова: № 4 при ВМК МГУ (1976—1989), Д 501.001.44 (1989—2017) и К.501.001.11 (1998—2007). Он также входил в состав редколлегии математических российских и зарубежных журналов, таких как Журнал вычислительной математики и математической физики (с 1995 г.), Дифференциальные уравнения (с 2002 г.), Вычислительные методы и программирование (с 2000 г.), Вестник Московского университета. Серия 15: Вычислительная математика и кибернетика (с 1978 г.), Optimization Methods and Software (1992—2002), Mathematica Montisnigri (1993—2007). Наиболее заметный след оставил Федор Павлович именно в Журнале вычислительной математики и математической физики, где активно работал в качестве эксперта-рецензента и регулярно публиковал собственные работы.

Педагогическая деятельность Федора Павловича Васильева началась на механико-математическом факультете МГУ имени М. В. Ломоносова в 1963 г. с должности ассистента, а в 1970 г. продолжилась на вновь открывшемся факультете вычислительной математики и кибернетики (ВМК). В 1966 г. Федор Павлович становится доцентом (на мехмате), а в 1989 г. — профессором (на ВМК). В 2000 г. ему было присвоено звание заслуженного профессора МГУ. Он работал на кафедрах вычислительной математики мехмата МГУ и факультета ВМК, кафедре математической физики факультета ВМК (всеми ими заведовал академик А. Н. Тихонов), а с 1991 г. — на кафедре оптимального управления факультета ВМК (зав. каф. академик Ю. С. Осипов). За время работы Федор Павлович прочитал много основных и специальных курсов лекций, в которых затрагивались различные теоретические аспекты задач оптимизации и управления, а также методы их решения. Будучи ассистентом и доцентом, он проводил также и упражнения по математическому анализу, дифференциальным уравнениям и математическому практикуму. Много сил Федор Павлович вкладывал в воспитание специалистов по прикладной математике, руководя научной работой студентов, аспирантов, соискателей и стажеров. Фундаментальную

роль в этом играл спецсеминар по методам оптимизации, организованный в 1963 г. И. С. Березиным, Б. М. Будаком и Ф. П. Васильевым на кафедре вычислительной математики мехмата МГУ. С 1970 г. этот семинар продолжил регулярную работу на факультете ВМК МГУ, причем Федор Павлович неизменно оставался его ведущим руководителем. В настоящее время спецсеминар работает на кафедре оптимального управления факультета ВМК, теперь уже под руководством учеников Федора Павловича. Наиболее широко известным лекционным произведением Федора Павловича несомненно является годовой основной курс «Методы оптимизации» (ранее «Экстремальные задачи»), который он беспрерывно и увлеченно читал на факультете ВМК с 1970 г.. Чтение именно этого курса позволило Федору Павловичу достаточно быстро выйти на уровень блистательного лектора и высококлассного педагога. Следует отметить, что материалы читаемых им курсов лекций нередко ложились в основу соответствующих учебных пособий, некоторые из которых, посвященные методам оптимизации и задачам линейного программирования, в большинстве российских вузов являются основными учебниками уже для нескольких поколений российских студентов.

Федор Павлович Васильев занимался популяризацией научно-технических знаний, прочитал ряд циклов лекций по линии общества «Знание», дважды удостоившись диплома первой степени на Всесоюзном конкурсе на лучшие произведения научно-популярной литературы (в 1974 и 1979 гг.), неоднократно награждался грамотами Московской городской организации общества «Знание», Всесоюзного общества «Знание», а с 1975 по 1989 г. работал деканом факультета вычислительной математики в Московском городском университете технического прогресса и экономических знаний. С 1974 по 2000 г. Федор Павлович работал в библиотечном совете факультета ВМК МГУ, причем с 1987 г. являлся его председателем. Еще в студенческие годы Федор Павлович был награжден медалью «За освоение целинных земель» (1957 г.). В дальнейшем его научно-педагогические достижения были отмечены следующими наградами: медалями «За доблестный труд. В ознаменование 100-летия со дня рождения В. И. Ленина» (1970 г.), «В память 850-летия Москвы» (1997 г.), «Ветеран труда» (1998 г.), званиями «Заслуженный профессор Московского университета» (2000 г.), «Почетный работник высшего профессионального образования Российской Федерации» (2004 г.), премией имени М. В. Ломоносова за педагогическую деятельность (2010 г.), премиями по Программе развития МГУ (2016 и 2019 гг.) и Государственной премией Чувашской Республики в области естественных и технических наук (2020 г.).

Федор Павлович Васильев оставил после себя богатое научно-образовательное наследие. Светлая память о Федоре Павловиче — коллеге, ученом и учителе, мудром, добром и очень скромном человеке — навсегда останется в сердцах его учеников и коллег, родных и знакомых.

## СПИСОК ОСНОВНЫХ ПУБЛИКАЦИЙ ВАСИЛЬЕВА ФЕДОРА ПАВЛОВИЧА

### *Монографии и учебные издания*

1. Будаков Б. М., Васильев Ф. П. Приближенные методы решения задач оптимального управления. Вып. 1. М.: Изд-во МГУ, 1968. 304 с.
2. Будаков Б. М., Васильев Ф. П. Приближенные методы решения задач оптимального управления. Вып. 2. М.: Изд-во МГУ, 1969. — 299 с.
3. Васильев Ф. П. Основы численных методов решения экстремальных задач. Вып. 1. М.: Изд-во МГУ, 1972. 136 с.
4. Васильев Ф. П. Основы численных методов решения экстремальных задач. Вып. 2. М.: Изд-во МГУ, 1973. 218 с.
5. Васильев Ф. П. Лекции по методам решения экстремальных задач. М.: Изд-во МГУ, 1974. — 376 с.
6. Будаков Б. М., Васильев Ф. П. Некоторые вычислительные аспекты задач оптимального управления. М.: Изд-во МГУ, 1975. 172 с.
7. Васильев Ф. П. Численные методы решения экстремальных задач. М.: Наука, 1980. 520 с.
8. Васильев Ф. П. Методы решения экстремальных задач. М.: Наука, 1981. 400 с.
9. Васильев Ф. П. Численные методы решения экстремальных задач. М.: Наука, 1988. — 552 с.
10. Васильев Ф. П., Ишмухаметов А. З., Потапов М. М. Обобщенный метод моментов в задачах оптимального управления. М.: Изд-во МГУ, 1989. 142 с.
11. Осипов Ю. С., Васильев Ф. П., Потапов М. М. Основы метода динамической регуляризации. М.: Изд-во МГУ, 1999. 237 с.
12. F. P. Vasilyev, A. Y. Ivanitskiy. In-Depth Analysis of Linear Programming. Dordrecht: Kluwer Academic Publishers, 2001. 312 p.
13. Васильев Ф. П., Иваницкий А. Ю. Линейное программирование. М.: Факториал Пресс, 2008 (2003, 1998). 328 с.

14. *Васильев Ф. П., Куржанский М. А., Потапов М. М., Разгулин А. В.* Приближенное решение двойственных задач управления и наблюдения. М.: Макс Пресс МГУ, 2010. 384 с.
15. *Васильев Ф. П.* Методы оптимизации. Кн. 1. М.: МЦНМО, 2011. 620 с.
16. *Васильев Ф. П.* Методы оптимизации. Кн. 2. М.: МЦНМО, 2011. 433 с.
17. *Васильев Ф. П., Потапов М. М., Будаков Б. А., Артемьева Л. А.* Методы оптимизации: учебник и практикум для бакалавриата и магистратуры. М.: Юрайт, 2016. 375 с.
18. *Васильев Ф. П., Иваницкий А. Ю.* Линейное программирование. М.: МЦНМО, 2020. 416 с.

*Наиболее значимые статьи*

19. *Васильев Ф. П.* О методе конечных разностей для решения однофазной задачи Стефана для квазилинейного уравнения // Докл. АН СССР. 1963. Т. 152. № 4. С. 783—786.
20. *Васильев Ф. П., Успенский А. Б.* О методе конечных разностей для решения двухфазной задачи Стефана для квазилинейного уравнения // Докл. АН СССР. 1963. Т. 152. № 5. С. 1034—1037.
21. *Васильев Ф. П.* Разностный метод решения задач типа Стефана для квазилинейного параболического уравнения с разрывными коэффициентами // Докл. АН СССР. 1964. Т. 157. № 6. С. 1280—1283.
22. *Васильев Ф. П.* О методе прямых для решения однофазной задачи типа Стефана // Ж. вычисл. матем. и матем. физ. 1968. Т. 8. № 1. С. 64—78.
23. *Васильев Ф. П.* Условия оптимальности для некоторых классов систем, не разрешенных относительно производной // Докл. АН СССР. 1969. Т. 184. № 6. С. 1267—1270.
24. *Васильев Ф. П., Иванов Р. П.* Некоторые приближенные методы решения задач быстрогодействия в банаховых пространствах при наличии фазовых ограничений // Докл. АН СССР. 1970. Т. 196. № 3. С. 525—529.
25. *Васильев Ф. П.* Об итерационных методах решения задач быстрогодействия, связанных с параболическими уравнениями // Ж. вычисл. матем. и матем. физ. 1970. Т. 10. № 4. С. 942—957.
26. *Васильев Ф. П., Иванов Р. П.* О приближенном решении задач быстрогодействия с запаздыванием // Ж. вычисл. матем. и матем. физ. 1970. Т. 10. № 5. С. 1124—1140.
27. *Васильев Ф. П., Иванов Р. П.* О приближенном решении задачи быстрогодействия в банаховых пространствах при наличии ограничений на фазовые координаты // Ж. вычисл. матем. и матем. физ. 1971. Т. 11. № 2. С. 328—347.
28. *Васильев Ф. П.* Об условиях существования седловой точки в детерминированных играх для интегро-дифференциальных систем с запаздыванием нейтрального типа // Автоматика и телемеханика. 1972. № 2. С. 40—50.
29. *Васильев Ф. П., Прохоров Н. А.* Методы решения одной дифференциальной игры // Докл. АН СССР. 1976. Т. 227. № 2. С. 269—272.
30. *Тихонов А. Н., Васильев Ф. П.* Методы решения некорректных экстремальных задач // В кн.: Banach Center Publications. V. 3. Mathematical models and numerical methods. Warszawa, 1978. P. 297—342.
31. *Васильев Ф. П.* О регуляризации некорректных экстремальных задач // Докл. АН СССР. 1978. Т. 241. № 5. С. 1001—1004.
32. *Васильев Ф. П.* О методах поиска оптимального времени в динамических играх преследования-уклонения с программным управлением // Докл. АН СССР. 1979. Т. 246. № 4. С. 788—792.
33. *Васильев Ф. П., Воронцов М. А., Литвинова О. А.* Об оптимальном управлении процессом теплового самовоздействия // Ж. вычисл. матем. и матем. физ. 1979. Т. 19. № 4. С. 1053—1058.
34. *Васильев Ф. П.* О регуляризации некорректных задач минимизации на множествах, заданных приближенно // Ж. вычисл. матем. и матем. физ. 1980. Т. 20. № 1. С. 38—50.
35. *Васильев Ф. П., Ячимович М.* Об итеративной регуляризации метода условного градиента и метода Ньютона при неточно заданных исходных данных // Докл. АН СССР. 1980. Т. 250. № 2. С. 265—269.
36. *Васильев Ф. П., Ячимович М.* Об итеративной регуляризации метода Ньютона // Ж. вычисл. матем. и матем. физ. 1981. Т. 21. № 3. С. 775—778.
38. *Васильев Ф. П., Хромова Л. Н.* О методе высокого порядка для решения операторных уравнений // Докл. АН СССР. 1983. Т. 270. № 1. С. 28—31.
38. *Васильев Ф. П.* О регуляризации методов высокого порядка при неточно заданных исходных данных // Докл. АН СССР. 1984. Т. 279. № 2. С. 281—285.
39. *Васильев Ф. П.* О регуляризации метода Ньютона при неточном задании исходных данных // Труды МИАН СССР. 1985. Т. 167. С. 53—59.
40. *Васильев Ф. П.* Регуляризация некоторых методов минимизации высокого порядка при неточных исходных данных // Ж. вычисл. матем. и матем. физ. 1985. Т. 25. № 4. С. 492—499.

41. *Васильев Ф. П.* О методе невязки для решения неустойчивых задач минимизации // *Вестн. МГУ. Сер. 15: Вычисл. матем. и киберн.* № 4. С. 6—10.
42. *Васильев Ф. П.* Применение негладких штрафных функций в методе регуляризации неустойчивых задач минимизации // *Ж. вычисл. матем. и матем. физ.* 1987. Т. 27. № 10. С. 1443—1450.
43. *Васильев Ф. П.* О регуляризации неустойчивых задач минимизации // *Труды МИАН СССР.* 1988. Т. 185. С. 60—65.
44. *Васильев Ф. П.* Оценка скорости сходимости метода регуляризации А. Н. Тихонова для неустойчивых задач минимизации // *Докл. АН СССР.* 1988. Т. 299. № 4. С. 792—796.
45. *Васильев Ф. П., Морозов В. В., Ячимович М.* Оценка скорости сходимости метода регуляризации для задачи линейного программирования // *Ж. вычисл. матем. и матем. физ.* 1989. Т. 29. № 4. С. 631—635.
46. *Васильев Ф. П., Иваницкий А. Ю., Морозов В. А.* Оценка скорости сходимости метода невязки для задач линейного программирования с приближенными данными // *Ж. вычисл. матем. и матем. физ.* 1990. Т. 30. № 8. С. 1257—1262.
47. *Васильев Ф. П., Алиакбаров С. М.* О регуляризации неустойчивых двухуровневых задач типа стандартизации // *Ж. вычисл. матем. и матем. физ.* 1991. Т. 31. № 3. С. 363—371.
48. *Васильев Ф. П., Обрадович О.* Регуляризованный проксимальный метод для задач минимизации с неточными исходными данными // *Ж. вычисл. матем. и матем. физ.* 1993. Т. 33. № 2. С. 179—188.
49. *Васильев Ф. П., Ячимович М.* Метод стабилизации для решения лексикографических задач // *Ж. вычисл. матем. и матем. физ.* 1993. Т. 33. № 8. С. 1123—1134.
50. *Васильев Ф. П., Недич А.* Об одном варианте регуляризованного метода проекции градиента // *Ж. вычисл. матем. и матем. физ.* 1994. Т. 34. № 4. С. 511—519.
51. *Васильев Ф. П., Обрадович О.* Регуляризованный проксимальный метод для выпуклых задач минимизации // *Труды Матем. ин-та им. В. А. Стеклова РАН.* 1995. Т. 211. С. 131—139.
52. *Васильев Ф. П., Алиакбаров С. М., Мухамадиев Э. М.* Существование и устойчивость решения экстремальной задачи типа стандартизации // *Ж. вычисл. матем. и матем. физ.* 1995. Т. 35. № 3. С. 323—333.
53. *Васильев Ф. П., Недич А., Ячимович М.* Регуляризованный непрерывный метод линеаризации для задач минимизации с неточными исходными данными // *Ж. вычисл. матем. и матем. физ.* 1996. Т. 36. № 3. С. 35—43.
54. *Васильев Ф. П., Недич А., Ячимович М.* Двухшаговый регуляризованный метод линеаризации для решения задач минимизации // *Ж. вычисл. матем. и матем. физ.* 1996. Т. 36. № 5. С. 9—19.
55. *Васильев Ф. П., Иваницкий А. Ю., Морозов В. А.* Метод поточечной невязки для решения некоторых задач линейной алгебры и линейного программирования // *Ж. вычисл. матем. и матем. физ.* 1998. Т. 38. № 7. С. 1140—1152.
56. *Васильев Ф. П., Антипин А. С.* Метод стабилизации для решения задач равновесного программирования с неточно заданным множеством // *Ж. вычисл. матем. и матем. физ.* 1999. Т. 39. № 11. С. 1779—1786.
57. *Антипин А. С., Будак Б. А., Васильев Ф. П.* Регуляризованный непрерывный экстраградиентный метод первого порядка с переменной метрикой для решения задач равновесного программирования // *Дифференц. уравнения.* 2002. Т. 38. № 12. С. 1587—1595.
58. *Антипин А. С., Васильев Ф. П.* Методы регуляризации для решения задачи равновесного программирования с неточными входными данными, основанные на расширении множества // *Ж. вычисл. матем. и матем. физ.* 2002. Т. 42. № 8. С. 1158—1165.
59. *Антипин А. С., Васильев Ф. П., Шпирко С. В.* Регуляризованный экстраградиентный метод решения задач равновесного программирования // *Ж. вычисл. матем. и матем. физ.* 2003. Т. 43. № 10. С. 1451—1458.
60. *Антипин А. С., Васильев Ф. П.* Регуляризованный метод с прогнозом для решения вариационных неравенств с неточно заданным множеством // *Ж. вычисл. матем. и матем. физ.* 2004. Т. 44. № 5. С. 796—804.
61. *Васильев Ф. П., Стукалов А. С.* Условия аппроксимации равновесных задач по значению функционала // *Ж. вычисл. матем. и матем. физ.* 2004. Т. 44. № 7. С. 1196—1208.
62. *Васильев Ф. П., Стукалов А. С.* Аппроксимация равновесной задачи по аргументу // *Ж. вычисл. матем. и матем. физ.* 2004. Т. 44. № 11. С. 1972—1982.
63. *Антипин А. С., Васильев Ф. П.* Методы регуляризации для решения неустойчивых задач равновесного программирования со связанными ограничениями // *Ж. вычисл. матем. и матем. физ.* 2005. Т. 45. № 1. С. 27—40.
64. *Антипин А. С., Васильев Ф. П., Шпирко С. В.* Регуляризованный экстраградиентный метод решения задач равновесного программирования с неточно заданным множеством // *Ж. вычисл. матем. и матем. физ.* 2005. Т. 45. № 4. С. 650—660.



65. Антипин А. С., Васильев Ф. П., Делавархалафи А. Методы регуляризации со штрафными функциями для поиска точек равновесия Нэша в билинейной игре двух лиц с ненулевой суммой // Ж. вычисл. матем. и матем. физ. 2005. Т. 45. № 5. С. 813—823.
66. Антипин А. С., Будак Б. А., Васильев Ф. П. Методы решения задач равновесного программирования // Дифференц. ур-ния. 2005. Т. 41. № 1. С. 3—11.
67. F. P. Vasil'ev, A. S. Antipin. Methods for solving unstable equilibrium programming problems with coupled variables // Proc. Steklov Inst. Math. 2006. Vol. 253. Suppl. 1. P. S229—S246.
68. Антипин А. С., Васильев Ф. П., Стукалов А. С. Регуляризованный метод Ньютона для решения задач равновесного программирования с неточно заданным множеством // Ж. вычисл. матем. и матем. физ. 2007. Т. 47. № 1. С. 21—33.
69. Васильев Ф. П., Антипин А. С., Артемьева Л. А. Непрерывный экстраградиентный метод решения параметрической многокритериальной задачи равновесного программирования // Дифференц. уравнения. 2009. Т. 45. № 11. С. 1577—1585.
70. Антипин А. С., Артемьева Л. А., Васильев Ф. П. Многокритериальное равновесное программирование: экстраградиентный метод // Ж. вычисл. матем. и матем. физ. 2010. Т. 50. № 2. С. 234—241.
71. Васильев Ф. П., Антипин А. С., Артемьева Л. А. Регуляризованный непрерывный экстраградиентный метод решения параметрической многокритериальной задачи равновесного программирования // Докл. АН СССР. 2010. Т. 434. № 4. С. 439—442.
72. Васильев Ф. П., Антипин А. С., Артемьева Л. А. Регуляризованный непрерывный экстраградиентный метод решения параметрической многокритериальной задачи равновесного программирования // Дифференц. ур-ния. 2010. Т. 46. № 11. С. 1584—1600.
73. Антипин А. С., Артемьева Л. А., Васильев Ф. П. Регуляризованный экстраградиентный метод решения параметрической многокритериальной задачи равновесного программирования // Ж. вычисл. матем. и матем. физ. 2010. Т. 50. № 12. С. 2083—2098.
74. F. P. Vasil'ev, E. V. Khoroshilova, A. S. Antipin. Regularized extragradient method for finding a saddle point in an optimal control problem // Proc. Steklov Inst. Math. 2011. Vol. 275. Suppl. 1. P. S186—S196.
75. Васильев Ф. П., Антипин А. С., Артемьева Л. А. Дифференциальный экстрапроксимальный метод поиска точки равновесия в седловых играх двух лиц // Дифференц. ур-ния. 2011. Т. 47. № 11. С. 1551—1563.
76. Антипин А. С., Артемьева Л. А., Васильев Ф. П. Экстрапроксимальный метод решения седловых игр двух лиц // Ж. вычисл. матем. и матем. физ. 2011. Т. 51. № 9. С. 1576—1587.
77. Антипин А. С., Артемьева Л. А., Васильев Ф. П. Регуляризованный экстрапроксимальный метод поиска точки равновесия в седловых играх двух лиц // Ж. вычисл. матем. и матем. физ. 2012. Т. 52. № 7. С. 1231—1241.
78. Васильев Ф. П., Антипин А. С., Артемьева Л. А. Метод регуляризации для поиска точки равновесия в седловых играх двух лиц при неточно заданных входных данных // Докл. АН СССР. 2014. Т. 454. № 3. С. 257—261.
79. Васильев Ф. П., Потапов М. М., Артемьева Л. А. Регуляризованный экстраградиентный метод в многокритериальных задачах управления с неточными данными // Дифференц. ур-ния. 2016. Т. 52. № 11. С. 1555—1567.
80. F. P. Vasil'ev, A. S. Antipin, L. A. Artem'eva. Extragradient method for finding a saddle point in a multicriteria problem with dynamics // Proc. Steklov Inst. Math. 2017. Vol. 297. Suppl. 1. P. 203—210.
81. Антипин А. С., Артемьева Л. А., Васильев Ф. П. Экстраградиентный метод поиска решения задачи оптимального управления с неявно заданными граничными условиями // Ж. вычисл. матем. и матем. физ. 2017. Т. 57. № 1. С. 49—54.
82. Васильев Ф. П., Потапов М. М., Артемьева Л. А. Экстраградиентный метод коррекции противоречивых задач линейного программирования // Ж. вычисл. матем. и матем. физ. 2018. Т. 58. № 12. С. 1992—1998.
83. Васильев Ф. П., Артемьева Л. А. Регуляризованный экстраградиентный метод поиска решения задачи оптимального управления с неточно заданными входными данными // Труды Матем. ин-та им. В. А. Стеклова РАН. 2019. Т. 304. С. 137—148.

УДК 51(092)



## ПАМЯТИ ПРОФЕССОРА БОРИСА ТЕОДОРОВИЧА ПОЛЯКА (1935–2023)

© 2024 г. М.В. Хлебников<sup>1,\*</sup>

<sup>1</sup>117997 Москва, ул. Профсоюзная, 65, Институт проблем управления им. В.А. Трапезникова РАН  
\*e-mail: mkhlebnikov2008@yandex.ru

Поступила в редакцию 07.11.2023 г.  
Переработанный вариант 18.11.2023 г.  
Принята к публикации 20.11.2023 г.

3 февраля 2023 г. после тяжелой болезни ушел из жизни Борис Теодорович Поляк — прекрасный человек, выдающийся ученый, мудрый наставник, воспитавший целые поколения ученых.

Б. Т. Поляк родился 4 мая 1935 г. в Москве, где и прожил всю свою жизнь. Окончив школу с золотой медалью, он поступил на физико-химический факультет Московского института стали имени Сталина и закончил его в 1958 г., так что по образованию он — инженер-металлург. К окончанию института Борис Теодорович Поляк уже имел ряд опубликованных статей в журналах “Научные доклады высшей школы. Metallургия”, “Инженерно-физический журнал”, “Журнал технической физики” и др.

В 1960 г. Б. Т. Поляк поступил в аспирантуру мехмата МГУ, на кафедру вычислительной математики, которой тогда заведовал академик А. Н. Тихонов. Он закончил аспирантуру в 1963 г. и в следующем году защитил кандидатскую диссертацию, посвященную изучению методов типа градиентного спуска в абстрактных функциональных пространствах. Впоследствии, работая в Вычислительном центре МГУ, он начинает читать лекции, ведет семинары. Именно в это время Б. Т. Поляк находит свою тему — *оптимизацию*, которая тогда еще даже не существовала как единая наука. Одним из первых он предложил общие методы для решения оптимизационных задач при наличии ограничений и методы минимизации негладких функций.

Отметим три важнейшие статьи Б. Т. Поляка, вышедшие в 1963, 1964 и 1966 гг. в «Журнале вычислительной математики и математической физики». В [7] появляется знаменитое *условие градиентного доминирования*, оно же — *условие Поляка-Лоясевича*. Сейчас это одна из самых востребованных релаксаций понятия сильной выпуклости. В статье [8] Б. Т. Поляк предлагает метод оптимизации, впоследствии получивший название *метод тяжелого шарика*. Метод тяжелого шарика стал предтечей современных ускоренных методов выпуклой оптимизации, а сейчас его стохастические вариации являются одними из основных методов обучения нейронных сетей. В статье [9], написанной Б. Т. Поляком совместно с его учеником Е. С. Левитиным, предлагается метод условного градиента, устанавливаются неулучшаемые оценки его скорости сходимости, исследуется поведение метода проекции градиента. Эти три статьи имеют многие тысячи цитирований, они определили развитие численных методов оптимизации на десятилетия вперед.

В 1966 г. Б. Т. Поляк принимал участие в Международном математическом конгрессе, основные заседания которого проходили в Главном здании МГУ. Там он смог познакомиться с ведущими западными учеными в области оптимизации, в частности, с Ральфом Рокафелларом.

С 1971 г. и до своих последних дней Борис Теодорович Поляк работал в ИПУ РАН, в лаборатории № 7 «Адаптивных и робастных систем». После смерти Якова Залмановича Цыпкина, ближайшим соратником которого он был, Борис Теодорович Поляк возглавлял лабораторию — уже носящую имя Цыпкина — на протяжении полутора десятилетий, в свою очередь передав в 2013 г. руководство ею автору настоящей статьи.

В 1978 г. Б. Т. Поляк защитил докторскую диссертацию на тему «Методы оптимизации при наличии помех», в которой исследовал вопросы стохастической аппроксимации. Он обобщил методы оптимизации на задачи со случайными помехами, эти результаты составили основу исследования алгоритмов стохастической аппроксимации для проблем оценивания, идентификации и адаптации. Б. Т. Поляк совместно с Я. З. Цыпкиным разработал оптимальные и робастные алгоритмы для подобных задач. При этом ему удалось перенести достижения современной статистики на стохастическую оптимизацию. Большую известность получил удивительный по простоте и изяществу метод стохастической аппроксимации с усреднением, сейчас известный как метод Поляка-Рупперта-Юдицкого. По словам Бориса Теодоровича, идея метода пришла ему во сне.

В 1983 г. вышла в свет знаменитая монография «Введение в оптимизацию», написанная Б. Т. Поляком еще в 1970-х гг. Вскоре она была переведена на английский язык и издана на Западе. Монография выдержала два переиздания — в 2014 и 2019 гг., и хоть прошло более 40 лет, по ней до сих пор читаются лекции студентам по численным методам оптимизации. Не будет преувеличением сказать, что именно с трудов Бориса Теодоровича Поляка и началась современная оптимизация.

С конца 1980-х гг. интересы Б. Т. Поляка перемещаются в область робастной устойчивости. Ряд полученных им результатов вошел во многие учебники по робастному управлению. В его совместной с Я. З. Цыпкиным статье [12] 1990 г. установлен знаменитый частотный критерий робастной устойчивости, получивший название *годографа Цыпкина-Поляка*; мировую известность получила их публикация [13], вышедшая в следующем году. В 1994 г. совместно с Я. З. Цыпкиным он удостоивается Премии РАН имени А. А. Андропова за цикл работ «Робастность в задачах оценивания, оптимизации и устойчивости».

Бориса Теодоровича Поляка привлекали трудные задачи теории управления — такие как синтез регуляторов низкого порядка, построение разреженного управления, подавление ограниченных внешних возмущений, исследование эффекта всплеска. В частности, исследования, связанные с подавлением случайных ограниченных внешних возмущений в линейных системах управления, завершились написанием монографии [5], которая была удостоена Премии РАН имени Петрова 2014 г.

Сложно даже перечислить направления, которыми занимался Б. Т. Поляк, и в развитие которых он внес свой вклад: от численных методов оптимизации, математического программирования, оценивания и идентификации до рандомизированных методов, робастных версий метода главных компонент, разреженного управления, управления хаосом. Он активно занимался и преподавательской деятельностью в МФТИ, в Школе Яндекса, подготовил более 25 кандидатов и докторов наук, а в 2019 г. вышел в свет учебник [6] по теории автоматического управления, написанный при его участии.

Б. Т. Поляк — автор шести монографий, ему принадлежат более 200 статей в журналах и свыше 150 докладов на российских и международных конференциях. В последние годы он вновь возвращается к оптимизационным идеям, и появилась серия пионерских работ на стыке оптимизации и управления. Его последняя прижизненная научная статья [22], посвященная новым методам настройки ПИД-регуляторов, вышла в ноябре 2022 г.

Невозможно переоценить вклад Б. Т. Поляка в развитие оптимизации и теории управления. На его статьях и книгах воспитано не одно поколение ученых как в СССР и России, так и за рубежом. Его

научные заслуги и достижения были признаны и международной общественностью: он был почетным членом ИФАК, работал в университетах США, Франции, Италии, Израиля, Мексики, Тайваня, Финляндии и других стран, обладатель Золотой медали Европейской ассоциации по исследованию операций (Euro Gold Medal, 2012), лауреат премии Хачияна (INFORMS Optimization Society Khachiyani Prize, 2021). Трудно найти на карте мира страну, где Борису Теодоровичу не довелось побывать, он имел огромное количество научных контактов и в нашей стране, и за рубежом, которые очень часто впоследствии перерастали в дружеские.

Б. Т. Поляк был не только выдающимся ученым, он был прекрасным организатором и руководителем. В течение долгих лет он был заместителем главного редактора журнала «Автоматика и телемеханика», членом редколлегий ряда отечественных и западных журналов. Борис Теодорович четверть века руководил еженедельным научным семинаром ИПУ РАН по теории автоматического управления.

По инициативе и непосредственном активнейшем участии Б. Т. Поляка в течение десяти лет (2009—2018) проводились ежегодные Всероссийские традиционные молодежные летние школы «Управление, информация и оптимизация», на которых талантливая молодежь со всей страны получала возможность слушать лекции крупнейших отечественных и зарубежных ученых. Сейчас эта линия успешно продолжается его коллегами из НИУ ВШЭ и МФТИ.

Б. Т. Поляк вел очень активный образ жизни и до своих последних дней участвовал в различных научных мероприятиях. Он был очень скромным, доброжелательным человеком, всегда готовым поделиться научными идеями, дать добрый совет. Светлые воспоминания о ярчайшей личности Бориса Теодоровича Поляка навсегда сохранятся в сердцах всех тех, кому посчастливилось с ним общаться и работать.

ИЗБРАННЫЕ ПУБЛИКАЦИИ Б. Т. ПОЛЯКА  
(ПОЛНАЯ БИБЛИОГРАФИЯ НА <https://sites.google.com/site/lab7polyak>)

*Монографии*

1. Поляк Б. Т. Введение в оптимизацию. М.: Наука, 1983.
2. Поляк Б. Т. Методы оптимизации при наличии помех. М.: Наука, 1991.
3. Поляк Б. Т., Щербаков П. С. Робастная устойчивость и управление. М.: Наука, 2002.
4. Граничин О. Н., Поляк Б. Т. Рандомизированные алгоритмы оценивания и оптимизации при почти произвольных помехах. М.: Наука, 2003.
5. Поляк Б. Т., Хлебников М. В., Щербаков П. С. Управление линейными системами при внешних возмущениях: Техника линейных матричных неравенств. М.: ЛЕНАНД, 2014.
6. Поляк Б. Т., Хлебников М. В., Рапопорт Л. Б. Математическая теория автоматического управления. М.: ЛЕНАНД, 2019.

*Статьи*

7. Поляк Б. Т. Градиентные методы минимизации функционалов // Ж. вычисл. матем. и матем. физ. 1963. Т. 3. № 4. С. 643—653.
8. Поляк Б. Т. О некоторых способах ускорения сходимости итерационных методов // Ж. вычисл. матем. и матем. физ. 1964. Т. 4. № 5. С. 791—803.
9. Левитин Е. С., Поляк Б. Т. Методы минимизации при наличии ограничений // Ж. вычисл. матем. и матем. физ. 1966. Т. 6. № 5. С. 787—823.
10. Poljak V. T., Tsytkin J. Z. Robust Identification // Automatica. 1980. Vol. 16. No. 1. P. 53—63.
11. Поляк Б. Т. Новый метод типа стохастической аппроксимации // Автоматика и телемехан. 1990. № 7. С. 98—107.
12. Поляк Б. Т., Цыпкин Я. З. Частотные критерии робастной устойчивости и апериодичности линейных систем // Автоматика и телемехан. 1990. № 9. С. 45—54.
13. Tsytkin Ya. Z., Polyak V. T. Frequency Domain Criteria for  $l^p$ -Robust Stability of Continuous Linear Systems // IEEE Transactions on Automatic Control. 1991. Vol. 36. No. 12. P. 1464—1469.
14. Поляк Б. Т., Цыпкин Я. З. Робастный критерий Найквиста // Автоматика и телемехан. 1992. № 7. С. 25—31.
15. Поляк Б. Т. Стабилизация хаоса с помощью прогнозирующего управления // Автоматика и телемехан. 2005. № 11. С. 99—112.
16. Поляк Б. Т., Щербаков П. С. Техника D-разбиения при решении линейных матричных неравенств // Автоматика и телемехан. 2006. № 11. С. 159—174.

17. Назин С. А., Поляк Б. Т., Топунов М. В. Подавление ограниченных внешних возмущений с помощью метода инвариантных эллипсоидов // Автоматика и телемехан. 2007. № 3. С. 106—125.
18. Поляк Б. Т., Хлебников М. В., Щербаков П. С. Разреженная обратная связь в линейных системах управления // Автоматика и телемехан. 2014. № 12. С. 13—27.
19. Поляк Б. Т., Тремба А. А., Хлебников М. В., Щербаков П. С., Смирнов Г. В. Большие отклонения в линейных системах при ненулевых начальных условиях // Автоматика и телемехан. 2015. № 6. С. 18—41.
20. Поляк Б. Т., Хлебников М. В. Метод главных компонент: Робастные версии // Автоматика и телемехан. 2017. № 3. С. 130—148.
21. Fatkhullin I., Polyak B. Optimizing Static Linear Feedback: Gradient Method // SIAM Journal on Control and Optimization. 2021. Vol. 59. No. 5. P. 3887—3911.
22. Поляк Б. Т., Хлебников М. В. Новые критерии настройки ПИД-регуляторов // Автоматика и телемехан. 2022. № 11. С. 62—82.

## СТОХАСТИЧЕСКИЙ ГРАДИЕНТНЫЙ СПУСК С ПРЕДОБУСЛОВЛЕННЫМ РАЗМЕРОМ ШАГА им. Б. Т. ПОЛЯКА

© 2024 г. Ф. Абдухакимов<sup>1,\*</sup>, Ч. Сян<sup>1,\*\*</sup>, Д. Камзолов<sup>1,\*\*\*</sup>, М. Такач<sup>1,\*\*\*\*</sup>

<sup>1</sup> Университет искусственного интеллекта им. Мохамеда бин Заеда, Абу-Даби, ОАЭ

\* e-mail: farshed888@gmail.com

\*\* e-mail: chulu.xiang@mbzuai.ac.ae

\*\*\* e-mail: kamzolov.opt@gmail.com

\*\*\*\* e-mail: takac.mt@gmail.com

Поступила в редакцию 02.11.2023 г.

Переработанный вариант 16.12.2023 г.

Принята к публикации 20.12.2023 г.

Стохастический градиентный спуск (SGD) является одним из множества методов оптимизации, используемых для решения задач машинного обучения. Практичность и простота подобных методов привлекают не только исследователей, но и инженеров машинного обучения из индустрии. Однако одна из главных слабостей таких методов заключается в необходимости ручной настройки размера шага для эффективного решения каждой конкретной оптимизационной задачи, функции потерь и данных. Стохастический градиентный спуск с размером шага им. Б.Т. Поляка (SPS) — это метод, который предлагает правило обновления, не требующее точной ручной настройки размера шага для решения задачи. Цель настоящей работы — расширить SPS с помощью таких приемов предобуславливания, как методы Хатчинсона, Adam и AdaGrad, что, в свою очередь, улучшит эффективность SPS в случае с плохой обусловленностью задачи и данных. Библ. 31. Фиг. 5.

**Ключевые слова:** машинное обучение, оптимизация, адаптивный размер шага, размер шага им. Б.Т. Поляка, предобусловленность.

DOI: 10.31857/S0044466924040016, EDN: ZKWLWGL

### 1. ВВЕДЕНИЕ

В настоящей статье мы рассматриваем задачу минимизации эмпирического риска (МЭР, англ. Empirical risk minimization, ERM), имеющую вид оптимизации конечной суммы:

$$w^* \in \operatorname{argmin}_{w \in \mathbb{R}^d} \left\{ f(w) := \frac{1}{n} \sum_{i=1}^n f_i(w) \right\}, \quad (1)$$

где  $w \in \mathbb{R}^d$  является параметром весов и каждая целевая функция  $f_i : \mathbb{R}^d \rightarrow \mathbb{R}$  является гладкой и дважды дифференцируемой. Функция потерь  $f_i(w)$  вычисляет разницу между предсказанием модели с параметрами весов  $w$  и целевым значением  $y_i$ . Целью является минимизация средней потери  $f(w) = \frac{1}{n} \sum_{i=1}^n f_i(w)$  на  $n$  данных  $\{(x_i, y_i)\}_{i=1}^n$ , где  $x_i$  — входная точка данных и  $y_i$  — соответствующее целевое значение. В связи с нетривиальностью данной задачи решение в явном виде не всегда доступно, что подталкивает к применению численных методов оптимизации. Одним из таких численных методов является стохастический градиентный спуск (SGD) со следующим обновлением параметров весов:

$$w_{t+1} = w_t - \gamma_t \nabla f_i(w_t), \quad (2)$$

где  $\gamma_t \in \mathbb{R}$  — размер шага метода. Использование мини-батчей для датасетов с большой размерностью при обучении значительно уменьшает время сходимости к оптимальной точке  $w^*$ . Были проведены обширные исследования в области стохастических методов оптимизации первого порядка начиная с фундаментальных работ Г. Роббинса и С. Монро [1], Б. Т. Поляка [2], Б. Т. Поляка и А. Б. Юдицкого [3], А. С. Немировского и др. [4] и ускоренные версии от Г. Лан [5]. Стоит отметить, что каждая комбинация функции потерь и датасета требует отдельной ручной настройки размера шага  $\gamma_t$  для поиска минимума, что делает  $\gamma_t$  гиперпараметром. Эта проблема ручной настройки  $\gamma_t$  является одним из мотивирующим факторов

разработки методов с адаптивным размером шага, где  $\gamma_t$  заменена адаптивно меняющимся выражением по ходу оптимизации. В последнее время такие адаптивные методы получили широкое распространение (см. [6]–[13]) особенно в области обучения глубоких нейронных сетей.

Другое направление адаптивных стохастических методов — стохастический градиентный спуск с размером шага им. Б. Т. Поляка, который был вдохновлен размером шага для субградиентных методов, предложенным Борисом Теодоровичем Поляком в 1969 г. (см. [14], [15]). Позже был предложен стохастический вариант этого шага в [16], [17] и другие различные расширения в [18]–[24]. В следующем разделе мы детально разберем некоторые из них.

Одним из главных предметов обсуждения статьи является получение методов, предназначенных для решения задач с плохой обусловленностью с помощью техники преобуславливания градиента. Несмотря на то что достижение идеального преобуславливания практически невозможно, наше решение использует различные техники, предложенные в таких адаптивных алгоритмах, как Adam [7] и AdaGrad [6], а также метод Хатчинсона [25].

Введем обозначения. Мы наделяем прямое пространство  $w \in \mathbf{E}$  и двойственное пространство  $g \in \mathbf{E}^*$  сопряженными нормами  $\|w\|$  и  $\|g\|_*$  соответственно. Как частный случай для положительно-определенной матрицы  $B \in \mathbb{R}^{d \times d}$  мы определяем двойственные евклидовы нормы:  $\|w\|_B = \langle Bw, w \rangle^{1/2}$  и  $\|g\|_{B^{-1}} = \langle g, B^{-1}g \rangle^{1/2}$ . Отметим, что  $\nabla f(w) \in \mathbf{E}^*$  и  $\nabla^2 f(w)h \in \mathbf{E}^*$  для  $h \in \mathbf{E}$ . Оператор  $\odot$  определяется как покомпонентное умножение двух векторов, также известное как произведение Адамара. Мы обозначаем  $\text{diag}(v)$  диагональную матрицу по заданному вектору  $v$  и вектор  $\text{diagonal}(\mathbf{H}) \in \mathbb{R}^d$  как диагональ матрицы  $\mathbf{H} \in \mathbb{R}^{d \times d}$ . Для простоты мы также вводим следующее обозначение:  $(x)_+ = \max\{0, x\}$ .

## 2. ОБЗОР ЛИТЕРАТУРЫ И СВЯЗАННЫЕ РАБОТЫ

Давайте введем общее правило обновления для рассматриваемых методов как

$$w_{t+1} = w_t - \gamma_t B_t^{-1} m_t, \quad (3)$$

где  $\gamma_t$  — размер шага,  $B_t = B_t^{-1}$  — специальная преобуславливающая матрица, и  $m_t$  обозначает либо  $g_t$  (градиент или его некоторая аппроксимация), либо первый момент градиента с параметром  $\beta_1$ . Для объяснения этого обновления мы можем представить, что направление спуска  $m_t$  шкалируется и вращается преобуславливающей матрицей  $B_t$ , и делается шаг с размером шага  $\gamma_t$ . Некоторые известные адаптивные методы первого порядка пользуются слегка упрощенной формой того же правила обновления:

$$w_{t+1} = w_t - \gamma_t m_t / v_t, \quad (4)$$

где  $m_t$  и  $v_t$  — первый и второй моменты, а  $m_t/v_t$  — покоординатное деление. Упомянутые типы шагов заключают в себе одну и ту же идею преобуславливания направления спуска и могут быть для простоты использованы взаимозаменяемо на протяжении всей статьи.

Таким же образом можно описать классические методы оптимизации. Например, для получения обновления SGD требуется обозначить преобуславливающую матрицу  $B_t = I$ , первый момент  $m_t = g_t$  и размер шага  $\gamma_t$  как константу. Стоит отметить, что  $\gamma_t$  в SGD является особенно важным гиперпараметром, который требует специальной настройки в соответствии с заданными данными и функцией потерь, а методы с адаптивным размером шага, некоторые из которых используют преобуславливающую матрицу, основанную на локальной кривизне функции потерь, были представлены для устранения этой проблемы.

Классические методы с размером шага им. Б. Т. Поляка не используют такую информацию, но, тем не менее, стоит упомянуть о том, как получить классический детерминистический размер шага им. Б. Т. Поляка. Рассмотрим выпуклую функцию  $f(w)$  и ограниченное сверху расстояние от  $w_{t+1}$  до оптимальной точки  $w^*$ :

$$\|w_{t+1} - w^*\|^2 \leq Q(\gamma), \quad \text{где } Q(\gamma) = \|w_t - w^*\|^2 - 2\gamma[f(w_t) - f^*] + \gamma_t^2 \|g_t\|_*^2.$$

Здесь  $g_t$  обозначает субградиент функции  $f(w)$ , а  $f^*$  — минимум функции. Минимизируя верхнюю границу  $Q(\gamma)$ , мы получаем размер шага им. Б. Т. Поляка и можем выразить его через правило обновления (3):

$$\gamma_t = \underset{\gamma \in \mathbb{R}}{\text{argmin}} [Q(\gamma)] = \frac{f(w_t) - f^*}{\|g_t\|_*^2}, \quad B_t = I \quad \text{и} \quad m_t = g_t. \quad (5)$$

Подробный разбор доказательства приведен в [26]. Заметим, что размер шага (5) может быть применен только в том случае, когда оптимальное значение  $f^*$  уже известно. Несмотря на то что иногда это значение

известно как  $f^* = 0$  (например в задачах классификации), детерминистическая природа данного метода делает его непрактичным. Для решения этой проблемы был представлен стохастический градиентный спуск с размером шага им. Б. Т. Поляка (SPS, Stochastic Gradient Descent with Polyak Step-size) (см. [17]) вместе с более практичной версией SPS<sub>max</sub>, который ограничивает  $\gamma_t$  постоянной  $\gamma_b$ :

$$\gamma_t^{\text{SPS}} = \frac{f_i(w_t) - f_i^*}{\|\nabla f_i(w_t)\|_*^2} \quad \text{и} \quad \gamma_t^{\text{SPSmax}} = \min \left\{ \frac{f_i(w_t) - f_i^*}{\|\nabla f_i(w_t)\|_*^2}, \gamma_b \right\}. \quad (6)$$

Метод SPS все еще требует знания  $f_i^*$ , но при определенных режимах оптимизации стандартной нерегуляризованной функции потерь, таких как квадратичная задача для линейной регрессии и логистическая регрессия для классификации, оптимальное решение  $f_i^*$  равно 0. Если  $f^* = 0$ , то правило обновления SPS выражается как

$$\gamma_t = \frac{f_i(w_t)}{\|\nabla f_i(w_t)\|_*^2}, \quad H_t = I \quad \text{и} \quad m_t = \nabla f_i(w_t). \quad (7)$$

Также существует другой способ получения метода SPS. Если предположить, что выполнено условие интерполяции, то мы можем решить (1) путем выборки  $i \in \{1, 2, \dots, n\}$  н.о.р.с.в. на каждой итерации  $t$  и решением нелинейного уравнения

$$w_{t+1} = \operatorname{argmin}_{w \in \mathbb{R}^d} \|w - w^t\|^2 \quad \text{т.ч.} \quad f_i(w) = 0. \quad (8)$$

Хотя приведенная выше проекция может иметь аналитическое решение для некоторых простых функций потерь, для большинства нелинейных моделей, таких как глубокие нейронные сети, не существует решения в замкнутой форме. Поэтому вместо точного решения мы можем линейаризовать  $f_i(w)$  вокруг текущей итерации  $w^t$ , чтобы получить

$$w_{t+1} = \operatorname{argmin}_{w \in \mathbb{R}^d} \|w - w^t\|^2 \quad \text{т.ч.} \quad f_i(w^t) + \langle \nabla f_i(w^t), w - w^t \rangle = 0.$$

Правило обновления (7) и есть аналитическое решение этой задачи.

Вне режима интерполяции решение для (8) может не существовать. Поэтому вместо того, чтобы пытаться обнулить все функции потерь, мы можем попытаться приблизить их к нулю, минимизировав дополнительную переменную остатка (slack) следующим образом:

$$\begin{aligned} & \operatorname{argmin}_{w \in \mathbb{R}^d, s \geq 0} s \\ & \text{т.ч.} \quad f_i(w) \leq s \quad \text{для} \quad i = 1, 2, \dots, n; \\ & \operatorname{argmin}_{w \in \mathbb{R}^d, s \geq 0} s^2 \\ & \text{т.ч.} \quad f_i(w) \leq s \quad \text{для} \quad i = 1, 2, \dots, n; \end{aligned}$$

которые называются  $L1$ - и  $L2$ -остаточными минимизациями (см. [19]) соответственно. Отметим, что цель этого метода состоит в том, чтобы приблизить  $s$  к нулю, что позволяет решать задачи, в которых предположение интерполяции не выполняется.

### 3. РЕЗУЛЬТАТЫ

В статье мы объединяем предобусловливание и варианты остаточно-регуляризованных методов SPS. Затем мы демонстрируем, что эти новые предобусловленные методы хорошо работают на плохо масштабированных и плохо обусловленных данных.

• **Усовершенствованный SPS.** Мы расширили методы SPS и представили три новых алгоритма: PSPS, PSPSL1 и PSPSL2, которые используют метод Хатчинсона, Adam и AdaGrad для предобусловливания градиентного шага с использованием размера шага им. Б. Т. Поляка для взвешенной евклидовой нормы. Правила обновлений наших методов в явном виде описаны ниже.



- **Имплементация в PyTorch.** Мы разработали практические варианты наших методов в качестве оптимизаторов PyTorch и опубликовали программный код в нашем репозитории GitHub<sup>1</sup>.

- **Эмпирические Результаты.** Мы привели несколько экспериментов с двумя разными задачами, чтобы сравнить наши результаты с SGD, Adam, AdaGrad и с вариантами SPS, в которых не применяются какие-либо методы предобусловливания. Мы показали, что предложенные нами алгоритмы демонстрируют заметные улучшения на плохо обусловленных задачах.

#### 4. ПРЕДОБУСЛОВЛИВАНИЕ

Данные могут быть плохо масштабированы и/или плохо обусловлены, тогда предобусловливание градиента — это один из способов улучшить сходимость алгоритмов. Методы, использующие предобусловливание, имеют следующее общее правило обновления:

$$w_{t+1} = w_t - \gamma_t B_t^{-1} \nabla f_t(w_t),$$

где  $B_t \in \mathbb{R}^{d \times d}$  — обратимая и положительно-определенная матрица. Метод Ньютона — один из самых наглядных примеров метода, использующего предобусловливание. В этом случае  $B_t = \nabla^2 f(w_t)$  и  $\gamma_t = 1$ . Среди более современных и практичных методов с предобуславливанием отметим AdaHessian [27], Adagrad [6] и OASIS [28]. Эти методы включают кривизну функции потерь посредством адаптивных оценок Гессииана.

##### 4.1. Метод Хатчинсона

Метод Хатчинсона (см. [25]) используется для оценки диагонали матрицы Гессииана. Для вычисления этой оценки метод Хатчинсона использует лишь несколько произведений Гессииана на вектор, которые, в свою очередь, можно эффективно вычислить с помощью быстрого автоматического дифференцирования (см. [29]). Произведение матрицы Гессииана  $\nabla^2 f(w)$  и фиксированного вектора  $z$  можно вычислить через производную градиента по направлению. Чтобы понять, как этот метод используется для предобусловливания, сначала мы покажем, что затраты на вычисление произведения Гессииана на вектор близки к двух вычислениям градиентов, т.е.

$$\nabla^2 f(w)z = \nabla(z^T \nabla f(w)). \quad (9)$$

Затем мы можем вычислить диагональ Гессииана, используя метод Хатчинсона:

$$\text{diag}(\nabla^2 f(w)) = \mathbb{E}[z \odot (\nabla^2 f(w)z)],$$

где  $z$  — случайный вектор с распределением Радемахера<sup>2</sup> или нормальным распределением, а  $\nabla^2 f(w)z$  вычисляется с помощью произведения Гессииана на вектор, заданного в (9).

Можно доказать, что математическое ожидание  $z \odot (\nabla^2 f(w)z)$  является диагональю Гессииана (см. [30]). Используя это тождество, мы оцениваем диагональ Гессииана по заданному  $D_0$ , генерируя случайный вектор  $z$  на каждой итерации и обновляя нашу оценку с использованием средневзвешенного значения следующим образом:

$$D_t = \beta D_{t-1} + (1 - \beta) \text{diag}(z \odot \nabla^2 f(w)z),$$

где  $\beta \in (0,1)$  — параметр момента и

$$D_0 = \frac{1}{m} \sum_{i=1}^m \text{diag}(z_i \odot \nabla^2 f(w_0)z_i).$$

Наконец, чтобы гарантировать, что  $D_t$  остается положительно-определенным, несмотря на возможную невыпуклость функций потерь, мы используем усечение и сохраняем только абсолютные значения элементов следующим образом:  $(\hat{D}_t)_{j,j} = \max\{\alpha, |D_t|_{j,j}\}$ .

---

**Algorithm 1.** Аппроксимация диагонали Гессииана с использованием метода Хатчинсона

---

1: **Ввод:**  $\beta \in (0,1)$ ,  $\alpha > 0$

2: **Инициализация:**  $D_0 = \frac{1}{m} \sum_{i=1}^m \text{diag}(z_i \odot \nabla^2 f(w_0)z_i)$

3: **for**  $t = 1, 2, \dots, T$  **do**

---

<sup>1</sup> <https://github.com/fxrshed/ScaledSPS>.

<sup>2</sup>  $z_i \in \{-1, +1\}$  с равной вероятностью.

- 4: Генерируем случайный вектор  $z$  из Радемахера/нормального распределения
- 5:  $D_t = \beta D_{t-1} + (1 - \beta) \text{diag}(z \odot \nabla^2 f(w_0) z)$
- 6:  $(\hat{D}_t)_{j,j} = \max\{\alpha, |D_t|_{j,j}\}$
- 7: **Вывод:**  $\hat{D}_T$

#### 4.2. Метод AdaGrad

AdaGrad — это метод стохастической оптимизации, который аппроксимирует Гессиан функции, чтобы адаптировать размер шага в зависимости от информации о кривизне. Ключевая идея заключается в использовании информации о кумулятивном квадрате градиента для шкалирования размера шага. В форме (4) правило обновления для AdaGrad может быть задано следующим образом:

$$m_t = \mathbf{g}_t \quad \text{и} \quad v_t = \sqrt{\sum_{i=1}^t \mathbf{g}_i \odot \mathbf{g}_i}.$$

Накопление всех предыдущих градиентов в предобуславливателе  $v_t$  приводит к уменьшению размера шага  $\gamma_t$ , что повышает производительность при разреженных данных (нечастых признаках), при этом ухудшается в случае плотных данных.

#### 4.3. Метод Adam

Представленный в [7] Adam разработан для преодоления недостатков других популярных алгоритмов оптимизации, таких как AdaGrad [6] и RMSProp [31], путем включения как адаптивного размера шага, так и обновлений на основе метода “тяжелого шарика” (momentum). Правило обновления Adam предполагает вычисление скользящего среднего как для первого, так и для второго моментов градиентов. Первый момент — это среднее значение градиентов, а второй момент — нецентрированная дисперсия градиентов. Правило обновления для Adam может быть выражено в терминах (4) следующим образом:

$$m_t = \frac{(1 - \beta_1) \sum_{i=1}^t \beta_1^{t-i} \mathbf{g}_i}{1 - \beta_1^t}, \quad v_t = \sqrt{\frac{(1 - \beta_2) \sum_{i=1}^t \beta_2^{t-i} \mathbf{g}_i \odot \mathbf{g}_i}{1 - \beta_2^t}},$$

где  $0 < \beta_1, \beta_2 < 1$  — два гиперпараметра, называемых коэффициентами первого и второго моментов. Смещенные оценки корректируются путем деления их на члены коррекции смещения, которые являются степенями скоростей затухания  $\beta_1$  и  $\beta_2$  соответственно.

### 5. ПРЕДОБУСЛОВЛЕННЫЙ СТОХАСТИЧЕСКИЙ ГРАДИЕНТНЫЙ СПУСК С РАЗМЕРОМ ШАГА им. Б. Т. ПОЛЯКА

В этом разделе мы предлагаем новые методы, основанные на ранее описанных, таких как SPS. Прежде всего, чтобы описать их, мы рассмотрим задачу проекции на множество ограничений

$$w_{t+1} = \underset{w \in \mathbb{R}^d}{\text{argmin}} \|w - w_t\|^2 \quad \text{т.ч.} \quad f_i(w) = 0. \tag{10}$$

Обратите внимание, что ограничение  $f_i(w) = 0$  определено как условие интерполяции.

**Определение 1.** Мы предполагаем, что условие интерполяции выполняется для набора функций  $\{f_i(w)\}_{i=1}^n$  по заданному набору данных  $\{(x_i, y_i)\}_{i=1}^n$  с неотрицательными функциями потерь  $f_i(w) \geq 0$ , когда

$$\exists w^* \in \mathbb{R}^d \quad \text{т.ч.} \quad f_i(w^*) = 0 \quad \forall i \in \{1, 2, \dots, n\}.$$

Одним из представленных методов, используемых в настоящей работе, является использование предобуславливания для улучшения скорости сходимости в случае плохо обусловленных данных. Чтобы получить это, мы изменяем норму в проекции (10) на взвешенную норму, основанную на предобуславливателе  $B_t \succ 0$ . Другой важной частью является линейная аппроксимация условия интерполяции  $f_i(w) = 0$ . Согласно разложению Тейлора функции  $f_i(w)$ , линейное приближение (первого порядка) задается через  $f_i(w) \approx f_i(w_t) + \langle \nabla f_i(w_t), w - w_t \rangle$ . Мы используем это приближение, чтобы ослабить условие интерполяции, которое не допускает решения в явном виде для большинства нелинейных моделей. Другой способ получения аналитического решения — ввести дополнительную переменную остатка (описано позже).

**Предобусловленный SPS.** Мы рассматриваем дифференцируемую выпуклую функцию  $f_i$  и линейризацию условия интерполяции. Чтобы вывести предобусловленное правило обновления, мы используем взвешенную норму в проекции, полученный метод мы называем PSPS (Preconditioned Stochastic Gradient Descent with Polyak Step-size). В настоящей статье мы рассмотрим три варианта предобуславливания, а именно, метод Хатчинсона и предобуславливание оптимизаторов AdaGrad и Adam.

**Лемма 1 (PSPS).** Пусть  $B_t \succ 0$  для всех  $t \geq 0$ , тогда итеративный явный шаг для задачи

$$w_{t+1} = \arg \min_{w \in \mathbb{R}^d} \frac{1}{2} \|w - w_t\|_{B_t}^2, \quad \text{т.ч.} \quad f_i(w_t) + \langle \nabla f_i(w_t), w - w_t \rangle = 0$$

выражается как

$$w_{t+1} = w_t - \frac{f_i(w_t)}{\|\nabla f_i(w_t)\|_{B_t}^2} B_t^{-1} \nabla f_i(w_t).$$

Отметим, что данный шаг может быть переформулирован в виде шага (3), где

$$\gamma_t = \frac{f_i(w_t)}{\|\nabla f_i(w_t)\|_{B_t}^2} \quad \text{и} \quad m_t = \nabla f_i(w_t).$$

Аналогичным образом мы можем применить предобуславливание для методов с остатком и получить следующие два метода: PSPSL1 и PSPSL2.

**Лемма 2 (PSPSL1).** Пусть  $B_t \succ 0$  для любых  $t \geq 0$  и  $\mu, \lambda > 0$ , тогда явный вид шага для задачи

$$w_{t+1}, s_{t+1} = \arg \min_{w \in \mathbb{R}^d, s \geq 0} \frac{1}{2} \|w - w_t\|_{B_t}^2 + \mu(s - s_t)^2 + \lambda s,$$

$$\text{т.ч.} \quad f_i(w_t) + \langle \nabla f_i(w_t), w - w_t \rangle \leq s$$

выражается как

$$\gamma_t^{L1} = \frac{(f_i(w_t) - s_t + \lambda/2\mu)_+}{1/2\mu + \|\nabla f_i(w_t)\|_{B_t}^2}, \quad \gamma_t = \min \left\{ \gamma_t^{L1}, \frac{f_i(w_t)}{\|\nabla f_i(w_t)\|_{B_t}^2} \right\},$$

$$w_{t+1} = w_t - \gamma_t B_t^{-1} \nabla f_i(w_t), \quad s_{t+1} = (s_t - \frac{1}{2\mu}(\lambda + \gamma_t^{L1}))_+.$$

**Лемма 3 (PSPSL2).** Пусть  $B_t \succ 0$  для любых  $t \geq 0$  и  $\mu, \lambda > 0$ , тогда явный вид решения задачи

$$w_{t+1}, s_{t+1} = \arg \min_{w \in \mathbb{R}^d, s \in \mathbb{R}} \|w - w_t\|_{B_t}^2 + \mu(s - s_t)^2 + \lambda s^2, \quad (11)$$

$$\text{т.ч.} \quad f_i(w_t) + \langle \nabla f_i(w_t), w - w_t \rangle \leq s$$

выражается как

$$w_{t+1} = w_t - \frac{(f_i(w_t) - \mu \hat{\lambda} s_t)_+}{\hat{\lambda} + \|\nabla f_i(w_t)\|_{B_t}^2} B_t^{-1} \nabla f_i(w_t),$$

$$s_{t+1} = \hat{\lambda} \left( \mu s_t + \frac{(f_i(w_t) - \mu \hat{\lambda} s_t)_+}{\hat{\lambda} + \|\nabla f_i(w_t)\|_{B_t}^2} \right),$$

где  $\hat{\lambda} = 1 / (\mu + \lambda)$ . Здесь остаточная параметр  $\lambda$  заставляет  $s$  быть ближе к 0, пока  $\mu$  не дает  $s_{t+1}$  быть далеко от  $s_t$ .

## 6. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

В этом разделе мы представляем эксперименты, проведенные с использованием предложенных нами методов и некоторых из наиболее популярных оптимизаторов: SGD, Adam и AdaGrad. Выбор этих методов оправдан тем фактом, что все они, за исключением SGD, используют адаптивный размер шага. В наших экспериментах каждый из этих методов представлен с разными размерами шага, чтобы показать разницу в сходимости.

Мы использовали датасеты из LIBSVM<sup>3</sup>, а именно, mushrooms и colon-cancer, для иллюстрации эффективности предложенных методов, минимизирующих функцию потерь логистической регрессии и нелинейных наименьших

<sup>3</sup> <https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/>

квадратов в задачах бинарной классификации. Кроме того, каждый эксперимент дополнительно проводится на плохо обусловленной версии тех же наборов данных, где столбцы умножаются на вектор  $e = \{\exp(x_i)\}_{i=1}^d$ , где  $x_i$  генерируется из равномерного распределения с интервалом  $[-k, k]$ . На всех приведенных далее иллюстрациях термин  $k$  относится к этому коэффициенту шкалирования, где  $k = 0$  — исходные данные.

Во время обучения предложенными методами мы применяли параметры остатка  $\lambda = 0.01$  и  $\mu = 0.1$ . Для метода Хатчинсона мы применили  $\alpha = 10^{-4}$  и  $\beta = 0,999$ . Гиперпараметры (за исключением размера шага) для других методов (SGD, Adam и т.д.) были сохранены в качестве значений по умолчанию. Все эксперименты проводились с пятью различными ключами генераторов случайности (seed), используя PyTorch 1.11.0.

Оптимизируемые функции. Пусть  $\{(x_i, y_i)\}_{i=1}^n$  — это данные из выбранного датасета. Логистическая регрессия определена следующим образом:

$$f_{\text{LogReg}}(w) = \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y_i x_i^T w)),$$

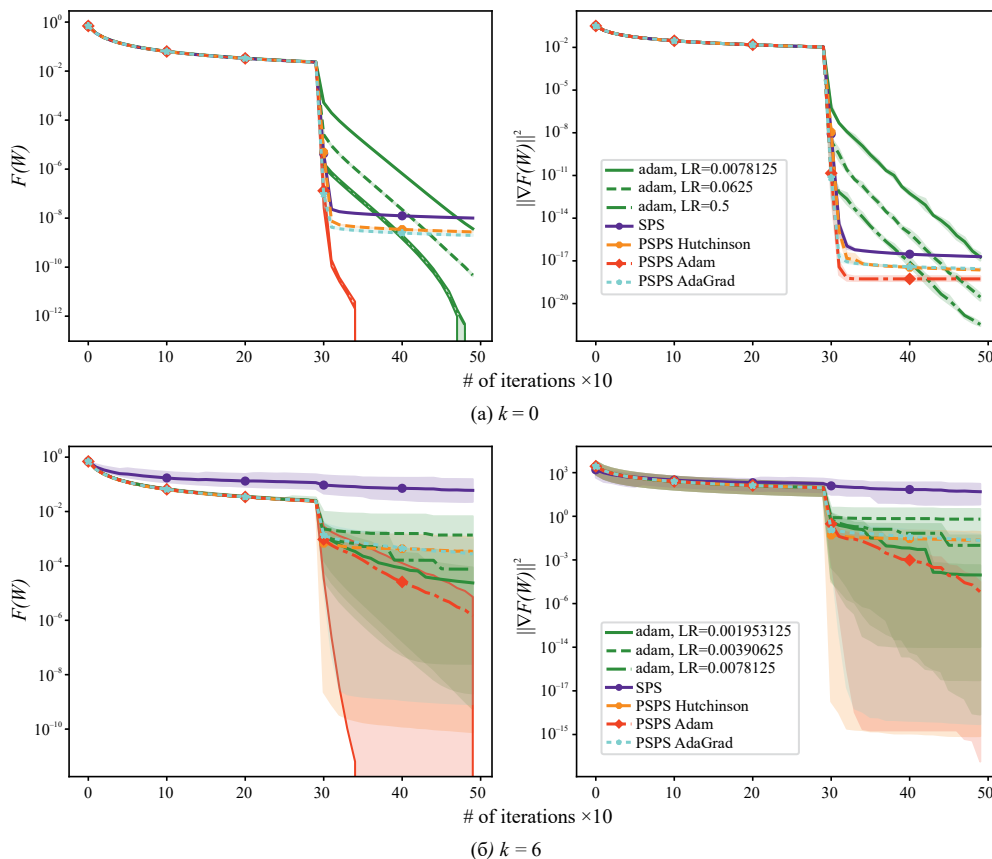
где  $x_i \in \mathbb{R}^d$  и  $y_i \in \{-1, +1\}$ . Нелинейные наименьшие квадраты заданы как

$$f_{\text{NLLSQ}}(w) = \frac{1}{n} \sum_{i=1}^n (y_i - 1 / (1 + \exp(-x_i^T w)))^2,$$

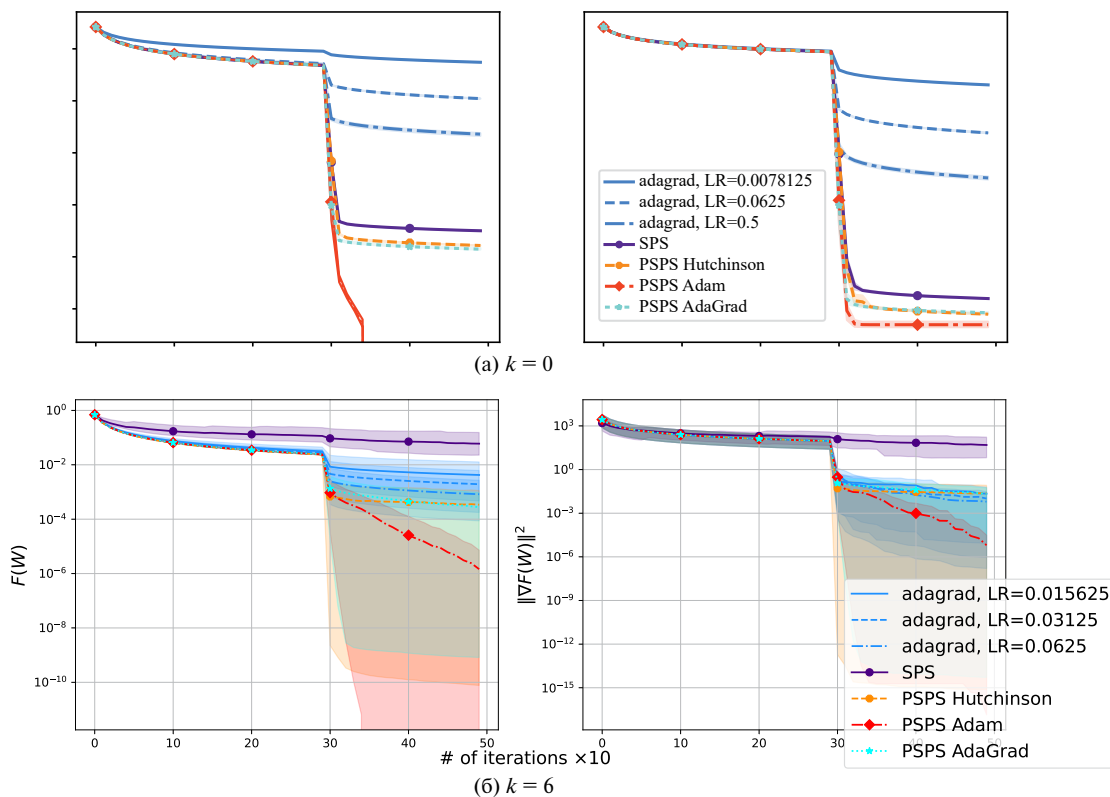
где  $y_i \in \{0, 1\}$ .

На фиг. 3 мы сравниваем скорости сходимости SPS с предобуславливателем Adam и без него. Наблюдаем, что в случае плохо обусловленных данных нам необходимо точно настроить размер шага оптимизатора Adam, чтобы избежать расхождений, поскольку выбор одинаковых размеров шага в обеих версиях данных привело к расхождению с  $k = 6$ . Кроме того, мы можем наблюдать, как различные методы предобуславливания превосходят SPS без какого-либо предобуславливания как для исходных данных, так и для плохо обусловленных. Отсутствие необходимости ручной настройки размера шага является одним из преимуществ предобусловленных методов SPS. Аналогичные результаты можно наблюдать на фиг. 6 и 9 для датасета colon-cancer. На фиг. 3б мы видим, что шкалирование данных приводит к тому, что размер шага Adam приходится уменьшать по мере увеличения коэффициента шкалирования  $k$ , чтобы метод не расхожился.

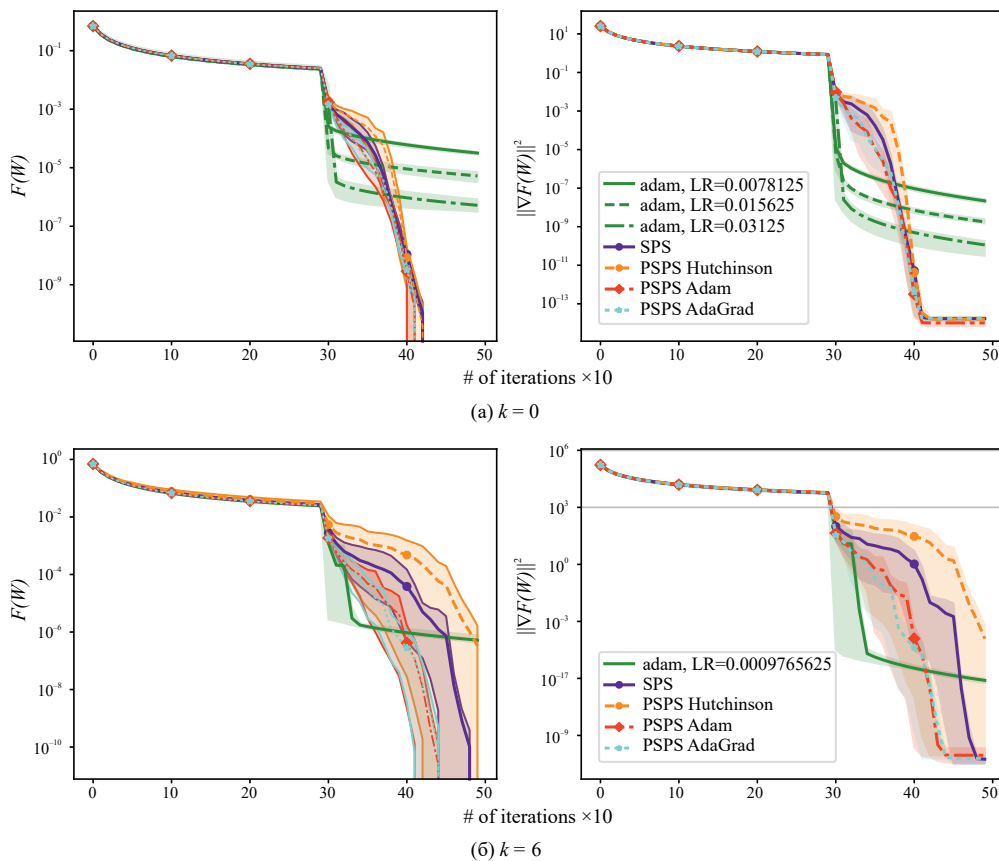
Мы также сравниваем наши методы с оригинальными SPS, SPSL1, SPSL2, SGD и Adam (фиг. 4 и 5).



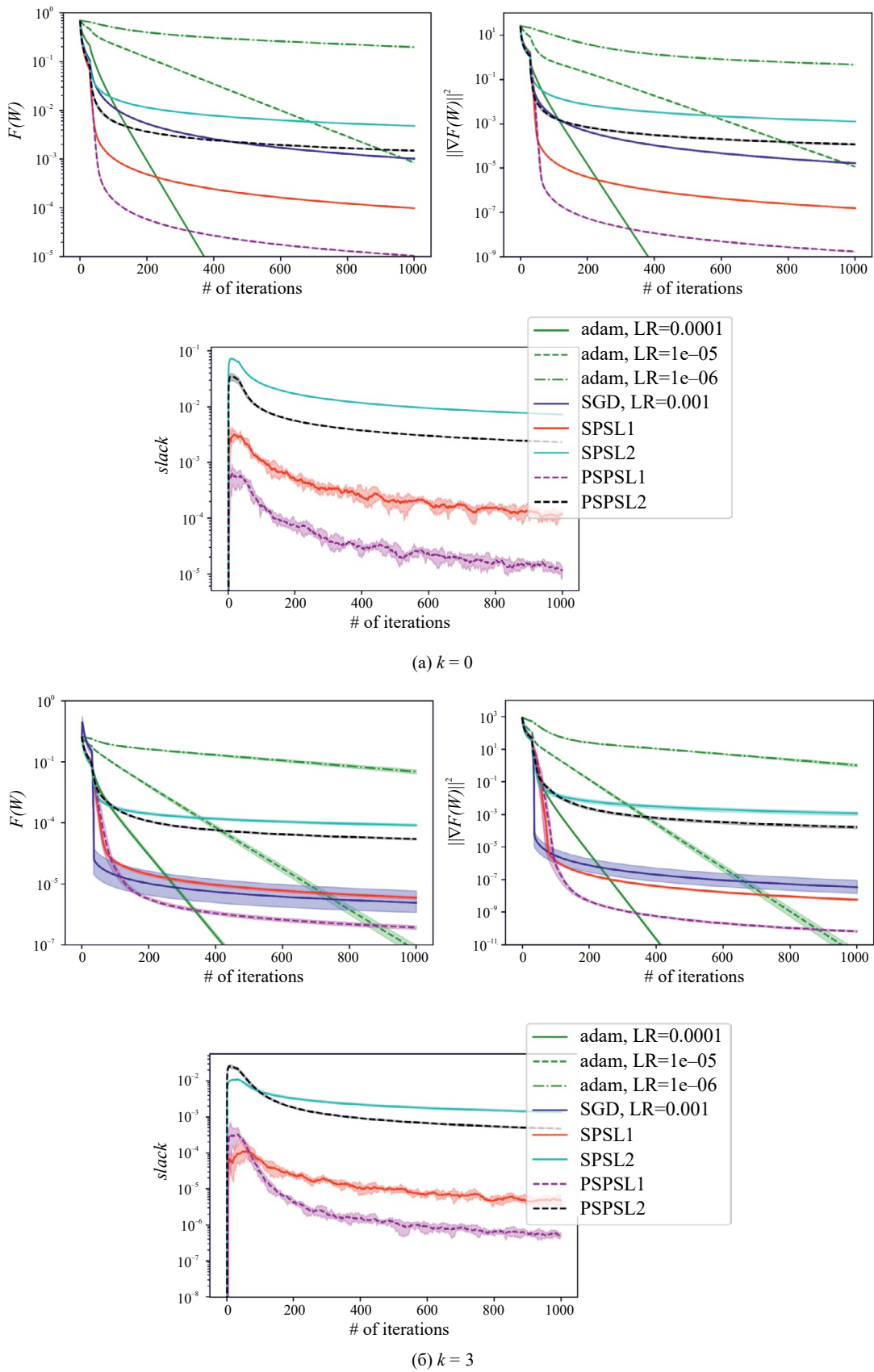
Фиг. 1. Метод Adam vs PSPPS с разным предобуславливанием для логистической регрессии на датасете mushrooms.



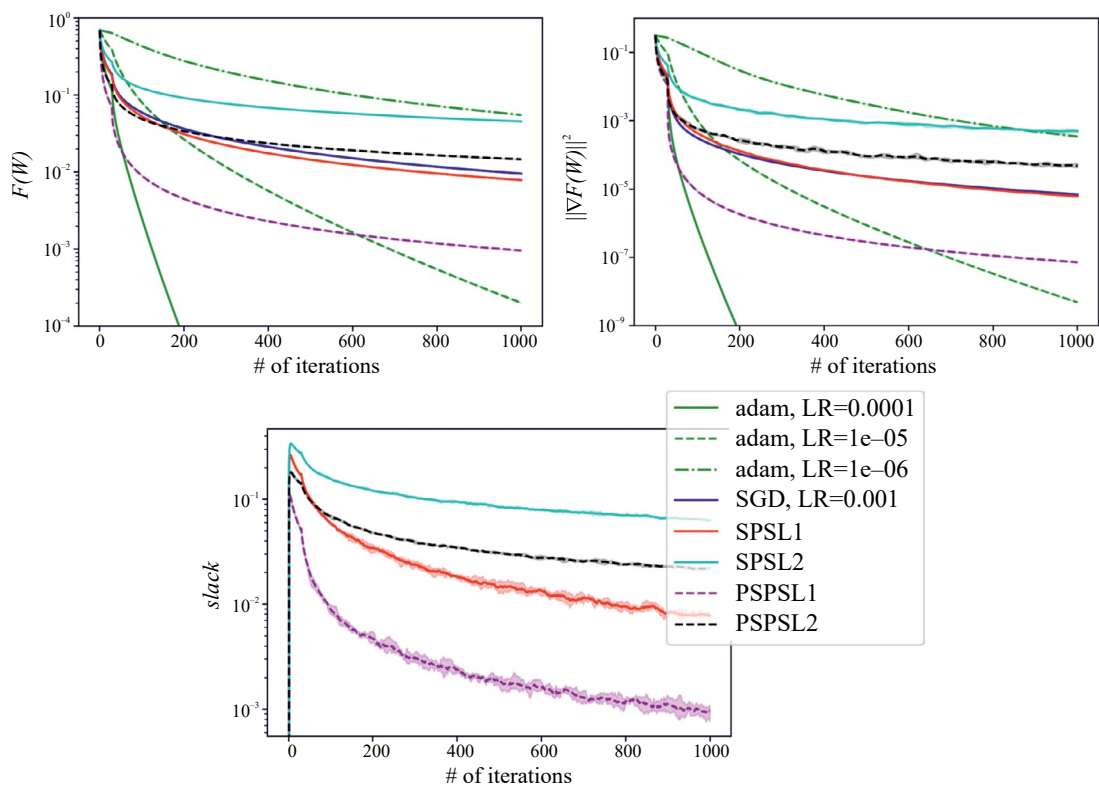
Фиг. 2. Методы AdaGrad vs PPS с разным предобуславливанием для логистической регрессии на датасете mushrooms.



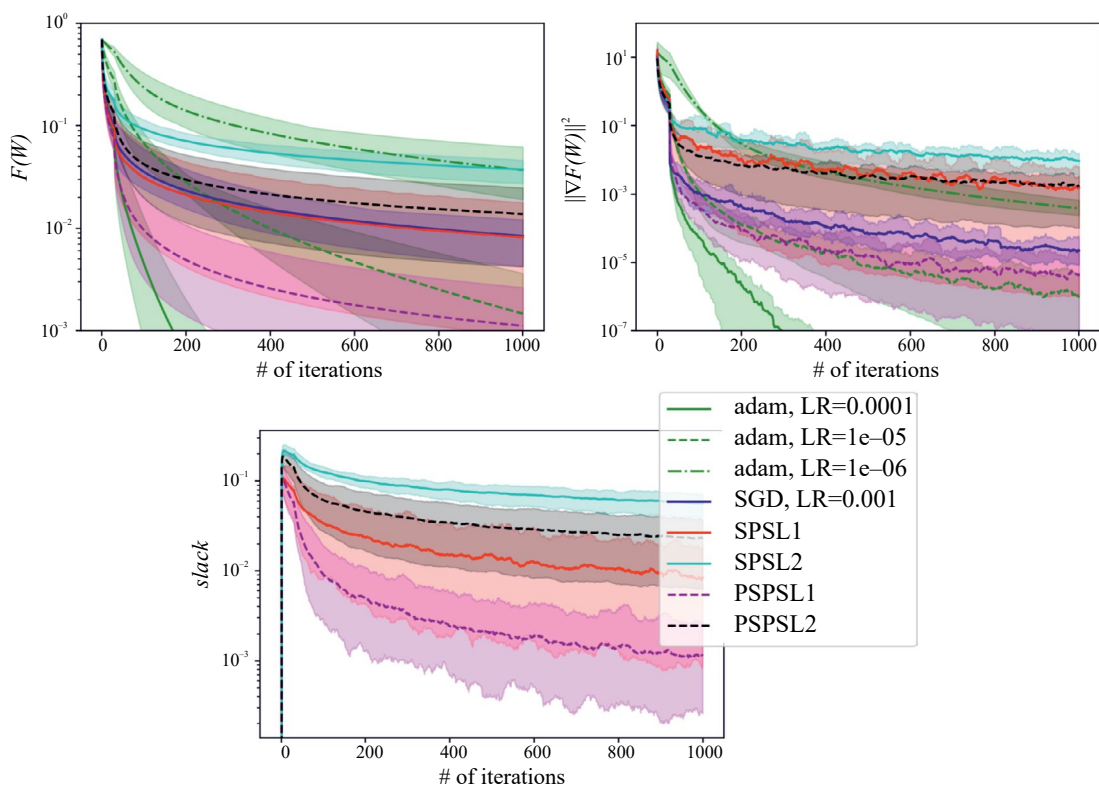
Фиг. 3. Методы Adam vs PPS с разным предобуславливанием для логистической регрессии на датасете colon-cancer.



**Фиг. 4.** Сравнение эффективности PSPSL1 и PSPSL2 с SPS, SGD и Adam для логистической регрессии на оригинальных и плохо обусловленных версиях датасета colon-cancer.



(a)  $k = 0$



(b)  $k = 3$

**Фиг. 5.** Сравнение эффективности PSPSL1 и PSPSL2 с SPS, SGD и Adam для логистической регрессии на оригинальных и плохо обусловленных версиях датасета mushrooms.

## 6. ЗАКЛЮЧЕНИЕ

В статье мы изучили влияние предобуславливания на семейство методов SPS (стохастический градиентный спуск с размером шага им. Б. Т. Поляка). Мы предложили новые методы PSPS, PSPSL1, PSPSL2 в (11)–(13). Эксперименты проводились как в выпуклых, так и в невыпуклых случаях с двумя разными датасетами. В настоящей статье отсутствует теоретический анализ предлагаемых нами методов, который может быть проведен в качестве последующей исследовательской работы. Кроме того, интересно провести эксперименты с более сложными моделями, такими как глубокие нейронные сети.

## СПИСОК ЛИТЕРАТУРЫ

1. Bekas C., Kokiopoulou E., Saad Y. An estimator for the diagonal of a matrix // *Appl. Numer. Math.* 2007. V. 57. № 11. P. 1214–1229.
2. Berrada L., Zisserman A., Kumar M. P. Training neural networks for and by interpolation. In Hal Daumé III and Aarti Singh, eds. // *Proceed. 37th Inter. Conf. Mach. Learn.* 2020. V. 119. P. 799–809.
3. Boyd S., Xiao L., Mutapcic A. Subgradient methods. lecture notes of EE392o, Stanford Univer., Autumn Quarter. 2023. V. 2004. P. 2004–2005.
4. Christianson B. Automatic Hessians by reverse accumulation // *IMA J. Numer. Anal.* 1992. V. 12. № 2. P. 135–150.
5. Duchi J., Hazan E., Singer Y. Adaptive subgradient methods for online learning and stochastic optimization // *J. Mach. Learn. Res.* 2011. V. 12. № 61. P. 2121–2159.
6. Garrigos G., Gower R. M., Schaipp F. Function value learning: Adaptive learning rates based on the polyak stepsize and function splitting in erm // *arXiv preprint arXiv:2307.14528*, 2023.
7. Gower R.M., Blondel M., Gazagnadou N., Pedregosa F. Cutting some slack for sgd with adaptive polyak step-sizes // *arXiv preprint arXiv:2202.12328*, 2022.
8. Hutchinson M.F. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines // *Comm. in Statistics-Simulation and Computat.* 1989. V. 18. № 3. P. 1059–1076.
9. Jahani M., Rusakov S., Shi Zh., Richtárik P., Mahoney M. W., Takáč M. Doubly adaptive scaled algorithm for machine learning using second-order information // *In 10th Inter. Conf. Learn. Representat. (ICLR2022)*, 2022.
10. Jiang X., Stich S. U. Adaptive sgd with polyak stepsize and line-search: Robust convergence and variance reduction // *arXiv preprint arXiv:2308.06058*, 2023.
11. Kingma D., Ba J. Adam: A method for stochastic optimization // *Inter. Conf. Learn. Representat. (ICLR)*, San Diego, CA, USA, 2015.
12. Lan G. An optimal method for stochastic composite optimization // *Math. Program.* 2012. V. 133. P. 365–397.
13. Li Sh., Swartworth W.J., Takáč M., Needell D., Gower R. M. SP2: A second order stochastic polyak method // *11th Inter. Conf. on Learn. Representat.*, 2023.
14. Li X., Orabona F. On the convergence of stochastic gradient descent with adaptive stepsizes. In Kamalika Chaudhuri and Masashi Sugiyama, eds. // *Proceed. 22nd Inter. Conf. Artific. Intelligence and Statistic.* 2019. V. 89. P. 983–992.
15. Loizou N., Vaswani Sh., Laradji I. H., Lacoste-Julien S. Stochastic polyak step-size for sgd: An adaptive learning rate for fast convergence. In Arindam Banerjee and Kenji Fukumizu, eds. // *Proceed. 24th Inter. Conf. Artific. Intelligence and Statistic.* 2021. V. 130. P. 1306–1314.
16. Loshchilov I., Hutter F. Decoupled weight decay regularization // *Inter. Conf. Learn. Representat.*, 2019.
17. Nemirovski A., Juditsky A., Lan G., Shapiro A. Robust stochastic approximation approach to stochastic programming // *SIAM J. Optimizat.* 2009. V. 19. № 4. P. 1574–1609.
18. Orvieto A., Lacoste-Julien S., Loizou N. Dynamics of sgd with stochastic polyak stepsizes: Truly adaptive variants and convergence to exact solution. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds. // *Adv. Neural Informat. Proces. System.* 2022. V. 35. P. 26943–26954.
19. Polyak B.T., Juditsky A. B. Acceleration of stochastic approximation by averaging.
20. *SIAM J. Control and Optimizat.* 1992. V. 30. № 4. P. 838–855.
21. Polyak B. T. Minimization of unsmooth functionals // *USSR Comput. Math. and Math. Phys.* 1969. V. 9. P. 14–29.
22. Polyak B. T. Introduction to optimization. Optimization Software, Inc., Publ. Division, 1987.
23. Polyak B.T. A new method of stochastic approximation type // *Avtomatika i Telemekhanika.* 1990. V. 51. P. 98–107.
24. Reddi S.J., Kale S., Kumar S. On the convergence of adam and beyond // *Inter. Conf. Learn. Representat.*, 2018.
25. Robbins H., Monro S. A stochastic approximation method // *Ann. Math. Statistic.* 1951. V. 22. P. 400–407.



26. Sadiev A., Beznosikov A., Almansoori A.J., Kamzolov D., Tappenden R., Takáč M. Stochastic gradient methods with preconditioned updates // arXiv preprint arXiv:2206.00285, 2022.
27. Schaipp F., Gower R. M., Ulbrich M. A stochastic proximal polyak step size // arXiv preprint arXiv:2301.04935, 2023.
28. Schaipp F., Ohana R., Eickenberg M., Defazio A., Gower R. M. Momo: Momentum models for adaptive learning rates // arXiv preprint arXiv:2305.07583, 2023.
29. Shi Zh., Sadiev A., Loizou N., Richtárik P., Takáč M. AI-SARAH: Adaptive and implicit stochastic recursive gradient methods // Transact. Mach. Learn. Res., 2023.
30. Tieleman T., Hinton G., et al. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude // COURSERA: Neural Networks for Machine Learn. 2012. V. 4. № 2. P. 26—31.
31. Ward R., Wu X., Bottou L. Adagrad stepsizes: Sharp convergence over nonconvex landscapes // J. Mach. Learn. Res. 2020. V. 21. № 1. P. 9047—9076.
32. Yao Zh., Gholami A., Shen Sh., Mustafa M., Keutzer K., Mahoney M. Adahessian: An adaptive second order optimizer for machine learning // Proceed. AAAI Conf. Artific. Intelligence. 2021. V. 35. P. 10665—10673.

## STOCHASTIC GRADIENT DESCENT WITH PRECONDITIONED POLYAK STEP-SIZE

**F. Abdukhakimov<sup>a,\*</sup>, Ch. Xiang<sup>a,\*\*</sup>, D. Kamzolov<sup>a,\*\*</sup>, M. Takáč<sup>a,\*\*\*\*</sup>**

*<sup>a</sup>Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE*

*\*e-mail: farshed888@gmail.com*

*\*\*e-mail: chulu.xiang@mbzuai.ac.ae*

*\*\*\*e-mail: kamzolov.opt@gmail.com*

*\*\*\*\*e-mail: takac.mt@gmail.com*

Received 02 November, 2023

Revised 16 December, 2023

Accepted 20 December, 2023

**Abstract.** Stochastic Gradient Descent (SGD) is one of the many iterative optimization methods that are widely used in solving machine learning problems. These methods display valuable properties and attract researchers and industrial machine learning engineers with their simplicity. However, one of the weaknesses of this type of methods is the necessity to tune learning rate (step-size) for every loss function and dataset combination to solve an optimization problem and get an efficient performance in a given time budget. Stochastic Gradient Descent with Polyak Step-size (SPS) is a method that offers an update rule that alleviates the need of fine-tuning the learning rate of an optimizer. In this paper, we propose an extension of SPS that employs preconditioning techniques, such as Hutchinson's method, Adam, and AdaGrad, to improve its performance on badly scaled and/or ill-conditioned datasets.

**Keywords:** machine learning, optimization, adaptive step-size, polyak step-size, preconditioning.

## О НЕКОТОРЫХ РАБОТАХ БОРИСА ТЕОДОРОВИЧА ПОЛЯКА ПО СХОДИМОСТИ ГРАДИЕНТНЫХ МЕТОДОВ И ИХ РАЗВИТИИ<sup>1)</sup>

© 2024 г. С. С. Аблаев<sup>1,4</sup>, А. Н. Безносиков<sup>1,2</sup>, А. В. Гасников<sup>1,2,3,\*</sup>, Д. М. Двинских<sup>1,2,3</sup>,  
А. В. Лобанов<sup>1,3</sup>, С. М. Пучинин<sup>1</sup>, Ф. С. Стонякин<sup>1,4</sup>

<sup>1</sup> 141700 Долгопрудный, М.о., Институтский пер., 9, НИУ МФТИ, Россия

<sup>2</sup> 127051 Москва, Б. Каретный пер., 19, стр. 1, ИППИ РАН, Россия

<sup>3</sup> 109004 Москва, ул. А. Солженицына, 25, ИСП РАН, Россия

<sup>4</sup> 295007 Симферополь, пр-т акад. Вернадского, 4,

Крымский федеральный университет им. В.И. Вернадского, Республика Крым, Россия

\* e-mail: gasnikov@yandex.ru

Поступила в редакцию: 15.09.2023 г.

Переработанный вариант 06.11.2023 г.

Принята к публикации 17.11.2023 г.

Представлен обзор современного состояния субградиентных и ускоренных методов выпуклой оптимизации, в том числе при наличии помех и доступа к различной информации о целевой функции (значении функции, градиент, стохастический градиент, старшие производные). Для невыпуклых задач рассматривается условие Поляка–Лоясиевича и приводится обзор основных результатов. Рассматривается поведение численных методов при наличии острого минимума. Цель данного обзора — показать влияние работ Б. Т. Поляка (1935–2023) по градиентным методам оптимизации и их окрестностям на современное развитие численных методов оптимизации. Библиография: 200. Табл. 1.

**Ключевые слова:** градиентный спуск, условие градиентного доминирования (Поляка–Лоясиевича), острый минимум, субградиентный метод Поляка–Шора, условие ранней остановки, метод тяжелого шарика Поляка, стохастический градиентный спуск.

DOI: 10.31857/S0044466924040028, EDN: ZKLBSS

### 1. ВВЕДЕНИЕ

Данный обзор посвящен частичному разбору нескольких работ Б. Т. Поляка о сходимости методов градиентного типа, которые на многие десятилетия определили развитие численных методов оптимизации. В частности, продолжают активно цитироваться и развиваться в настоящее время. Прежде всего, речь пойдет о работах [1]–[15].

Подчеркнем, что в данной статье не планируется описывать научный путь Бориса Теодоровича. Мы коснемся лишь пары десятков статей из более чем 250-ти. Более подробно с научным путем Б. Т. Поляка можно познакомиться, например, по статьям [16], [17].

Структура обзора следующая. В разд. 2 описываются методы негладкой оптимизации и специальный способ адаптивного выбора шага (в литературе часто называется «шаг Поляка»). В частности, приводится субградиентный метод Поляка–Шора и вариант этого метода с переключением (для задач с функциональными ограничениями). Далее обсуждается вопрос о возможной линейной скорости сходимости таких методов, если минимум острый.

В разд. 3 излагаются методы гладкой оптимизации. Начинается изложение с описания условия градиентного доминирования, которое обобщает условие сильной выпуклости целевого функционала, не предполагая при этом даже выпуклости. Показывается, что при данном условии градиентный спуск глобально линейно сходится. Рассматриваются различные обобщения данного результата. В частности, на случай, когда градиент доступен лишь с заданным уровнем относительной неточности. Затем идет изложение метода условного градиента и некоторых современных результатов вокруг этого метода. В заключении

<sup>1)</sup> Работа выполнена при финансовой поддержке Минобрнауки РФ (проект FSMG-2024-0011).

данного раздела описывается метод тяжелого шарика Поляка, который породил линейку современных ускоренных методов.

В разд. 4 приводятся результаты Поляка–Цыпкина и Поляка–Юдицкого–Рупперта, в которых возникают различные формы центральной предельной теоремы для выхода алгоритма типа стохастического градиентного спуска. Далее приводятся неасимптотические результаты, в том числе и для ускоренных вариантов стохастического градиентного спуска. В частности, рассматриваются так называемые мультипликативные помехи, которые в современных исследованиях чаще называют условием сильного роста, не ассоциируя это с пионерскими работами Б. Т. Поляка с соавторами. В заключении разд. 4 рассматриваются рандомизированные безградиентные (также называемые поисковыми) методы. В условиях повышенной гладкости целевой функции обсуждается конструкция Поляка–Цыбакова, позволяющая строить хорошую модель производной по направлению целевой функции, исходя всего из двух проб.

## 2. НЕГЛАДКАЯ ВЫПУКЛАЯ ОПТИМИЗАЦИЯ

История развития методов негладкой оптимизации начинается в 60-е годы прошлого века и достаточно подробно описана в работе [18]. Наряду с работами Н. З. Шора [19] важный вклад в развитие этой области принадлежит Б. Т. Поляку (см. [11], [15]).

### 2.1. Субградиентный метод Поляка–Шора

Если не оговорено иное, то далее в тексте рассматриваются задачи вида

$$f(x) \rightarrow \min_{x \in Q}, \quad (1)$$

где  $f(x)$  — необязательно дифференцируемая выпуклая функция,  $Q$  — выпуклое замкнутое подмножество  $\mathbb{R}^n$ ,  $f^* = f(x^*) = \min_{x \in Q} f(x)$ .

В данном пункте мы поговорим про субградиентные методы для негладкой оптимизации и вклад Бориса Теодоровича Поляка. Хорошо известно, что при минимизации недифференцируемых функций возникает ряд проблем: неприменим метод покоординатного спуска, а субградиент целевой функции не задает направление наискорейшего возрастания. В связи с этим Н. З. Шором был предложен субградиентный метод [20], являющийся прямым аналогом градиентного метода. Особенность идеи такого метода в том, что вместо градиента целевой функции в методе используется произвольный субградиент негладкой выпуклой функции. Рассмотрим случай, когда  $Q$  — замкнутое выпуклое подмножество  $\mathbb{R}^n$  с евклидовой нормой  $\|\cdot\|_2$ . Пусть  $B_2^n(x_*, R) = \{x \in \mathbb{R}^n : \|x - x_*\|_2 \leq R\}$ . Будем всюду далее обозначать субградиент  $f$  (некоторый элемент субдифференциала  $\partial f(x)$ ) в точке  $x$  как  $\nabla f(x)$ . Если функция  $f$  дифференцируема в точке  $x$ , то  $\nabla f(x)$  — ее градиент. Итерация субградиентного метода при  $h_k > 0$  имеет следующий вид:

$$x^{k+1} = \text{Pr}_Q\{x^k - h_k \nabla f(x^k)\}, \quad \nabla f(x^k) \in \partial f(x^k), \quad (2)$$

где  $\text{Pr}_Q(y) := \arg \min_{x \in Q} \{\|y - x\|_2\}$  — оператор евклидова проектирования на множество  $Q$ . Одна из главных особенностей субградиентных методов состоит в том, что значения функции в этом методе могут не убывать монотонно с ростом количества итераций. Вообще,  $f$  не обязательно убывает вдоль направления  $-\nabla f(x^k)$ , обратного направлению субградиента в текущей точке. Однако оказывается, что при этом возможно гарантировать монотонное убывание расстояния от текущей точки до точки минимума. Вторая особенность — это выбор шага субградиентного метода. Если выбирать постоянный шаг, то метод может не сходиться. Действительно, пусть для функции одной переменной  $f(x) = |x|$ ,  $x_0 = -0.01$  и выбран постоянный шаг  $h_k = 0.02$ . Тогда итеративная последовательность метода (2) будет состоять всего из двух точек  $-0.01$  и  $0.01$  и не будет сходимости к точке минимума  $0$ .

Интересный подход к выбору шага в субградиентном методе предложен Б. Т. Поляком. Он предложил при выборе шага использовать степень близости значения функции в текущей точке к минимальному. Это вполне возможно, если известно искомое минимальное значение функции  $f^*$ . Например, если систему совместных линейных уравнений

$$\langle a_i, x \rangle = b_i, \quad i = 1, 2, \dots, n, \quad x \in \mathbb{R}^n,$$

свести к минимизации функции

$$f(x) = \sum_{i=1}^n |\langle a_i, x \rangle - b_i|,$$

то  $f^* = 0$ . Также  $f^*$  бывает известно в геометрических задачах — проекция точки на множество, нахождение общей точки системы множеств. Используя  $f^*$ , можно построить адаптивный вариант шага (впервые он предложен в [11]), не содержащий таких параметров задачи, как константа Липшица целевой функции или расстояние от точки старта до множества решений:

$$h_k = \frac{f(x^k) - f^*}{\|\nabla f(x^k)\|_2^2}. \quad (3)$$

Такой шаг принято называть шагом Б. Т. Поляка. Для субградиентного метода с таким шагом известен следующий результат о сходимости (см. [2]).

**Теорема 1.** Пусть  $f(x)$  — выпуклая на  $\mathbb{R}^n$  функция, множество точек минимума  $X_*$  которой не пусто. Тогда в методе (2) с шагом (3)  $x^k \rightarrow x_* \in X_*$ . При этом  $\lim_{k \rightarrow \infty} \sqrt{k}(f(x^k) - f^*) = 0$ .

**Доказательство.** Как известно, для всякой точки минимума  $x_* \in X_*$  верны неравенства

$$\begin{aligned} 2h_k(f(x^k) - f(x_*)) &\leq 2h_k \langle \nabla f(x^k), x^k - x_* \rangle \leq \\ &\leq h_k^2 \|\nabla f(x^k)\|_2^2 + \|x^k - x_*\|_2^2 - \|x^{k+1} - x_*\|_2^2. \end{aligned}$$

Поэтому

$$\begin{aligned} \|x^{k+1} - x_*\|_2^2 &\leq h_k^2 \|\nabla f(x^k)\|_2^2 - \\ &- 2h_k(f(x^k) - f(x_*)) + \|x^k - x_*\|_2^2 = \end{aligned} \quad (4)$$

$$\begin{aligned} &= \frac{(f(x^k) - f(x_*))^2}{\|\nabla f(x^k)\|_2^2} - \frac{2(f(x^k) - f(x_*))^2}{\|\nabla f(x^k)\|_2^2} + \|x^k - x_*\|_2^2 = \\ &= -\frac{(f(x^k) - f(x_*))^2}{\|\nabla f(x^k)\|_2^2} + \|x^k - x_*\|_2^2. \end{aligned} \quad (5)$$

Таким образом, имеем

$$\frac{f(x^k) - f^*}{\|\nabla f(x^k)\|_2} \rightarrow 0.$$

Более того, неравенство  $\|x^k - x_*\|_2 \leq \|x^0 - x_*\|_2$  означает, что последовательность  $x^k$  ограничена, и тогда  $\|\nabla f(x^k)\|_2 \leq M$ . Поэтому  $f(x^k) \rightarrow f^*$ . Следовательно, найдется подпоследовательность  $x^{k_l} \rightarrow x_*$ . Итак, получаем, что  $\|x^k - x_*\|_2$  монотонно убывает, а  $\|x^{k_l} - x_*\|_2 \rightarrow 0$ . Отсюда  $x^k \rightarrow x_*$ . Ввиду (5) получаем

$$\sum_{k=0}^{\infty} \frac{(f(x^k) - f^*)^2}{\|\nabla f(x^k)\|_2^2} < \infty,$$

а из ограниченности  $\|\nabla f(x^k)\|_2$  следует  $\sum_{k=0}^{\infty} (f(x^k) - f^*)^2 < \infty$ . Если предположить, что  $\lim_{k \rightarrow \infty} \sqrt{k}(f(x^k) - f^*) > 0$ , то  $f(x^k) - f^* > a/\sqrt{k}$  для достаточно больших  $k$ , что противоречит условию  $\sum_{k=0}^{\infty} (f(x^k) - f^*)^2 < \infty$ . Итак,  $\lim_{k \rightarrow \infty} \sqrt{k}(f(x^k) - f^*) = 0$ . Теорема 1 доказана.

Предыдущий результат означает, что для достижения точности  $\varepsilon > 0$  решения задачи по функции гарантированно достаточно  $O(1/\varepsilon^2)$  итераций. Данная оценка скорости сходимости неумлучшаема на классе минимизационных задач с выпуклыми липшицевыми (как гладкими, так и негладкими) целевыми функциями. Хотя известны и другие подходы к выбору шага для субградиентного метода. Например, если при  $Q = \mathbb{R}^n$  предположить (здесь и всюду далее  $R \geq \|x^0 - x_*\|_2$ ), что

$$\|\nabla f(x)\|_2 \leq M \text{ для всякого } x \in B_2^n(x_*, R\sqrt{2}), \quad (6)$$

и выбрать шаг субградиентного метода, а также точку выхода следующим образом:

$$h = \frac{R}{M\sqrt{N}}, \quad x^{-N} = \frac{1}{N} \sum_{k=0}^{N-1} x^k, \quad (7)$$

то будет выполняться неравенство

$$f\left(x^{-N}\right) - f(x_*) \leq \frac{MR}{\sqrt{N}}. \quad (8)$$

Неравенство (8) означает, что при выборе

$$N = \frac{M^2 R^2}{\varepsilon^2}, \quad h = \frac{\varepsilon}{M^2} \quad (9)$$

будет достигаться оценка  $f\left(x^{-N}\right) - f(x_*) \leq \varepsilon$ , которая оптимальна с точностью до умножения на константу. Действительно, известно, что точная нижняя оценка на классе задач выпуклой оптимизации с условием (6) для методов первого порядка вида

$$x^{k+1} \in x^0 + \text{span}\{\nabla f(x^0), \dots, \nabla f(x^k)\}, \quad (10)$$

где для всех  $j = 0, 1, \dots, k$  верно  $\nabla f(x^j) \in \partial f(x^j)$ , имеет вид (см. [21])

$$f\left(x^N\right) - f(x_*) \geq \frac{MR}{\sqrt{N+1}}. \quad (11)$$

Представляется, что интерес выбора шага Б. Т. Поляка для субградиентного метода в том, что он в каждой точке позволяет учесть динамику значений целевой функции и не содержит параметров типа константы Липшица целевой функции или оценки расстояния от начальной точки до множества точных решений задачи. Такого типа процедуры выбора шагов в оптимизационных методах часто называют адаптивными. Разным подходам к адаптивным процедурам при выборе шагов оптимизационных методов посвящаются все новые работы (см. [22]–[24]). В этом плане можно отметить многочисленные исследования в области универсальных градиентных методов (см. [25], [26]), адаптивных методов типа AdaGrad, да и разные модификации шага Б. Т. Поляка в детерминированном и стохастическом случаях (см. [22], [27]). Важно, что, помимо процедуры выбора шага, Б. Т. Поляк еще выделил и класс задач с острым минимумом (см. [11]), для которого этот шаг позволил доказать результат о сходимости субградиентного метода со скоростью геометрической прогрессии. Далее поговорим об этом результате и его развитии в современных работах.

## 2.2. Острый минимум и линейная скорость сходимости субградиентного метода с шагом Б. Т. Поляка

Как видим, в описанных выше результатах оценки скорости сходимости субградиентного метода сублинейны. Получить линейную сходимость субградиентного метода можно лишь с помощью дополнительных предположений. Например, линейная скорость сходимости возможна для методов типа секущей гиперплоскости, применимых к задачам малой или умеренной размерности. Что касается задач большой размерности, то сходимость со скоростью геометрической прогрессии может позволить получить дополнительное предположение об остром минимуме

$$f(x) - f^* \geq \alpha \min_{x_* \in X_*} \|x - x_*\|_2, \quad (12)$$

где  $X_*$  — множество точек минимума функции  $f$  на множестве  $Q$ ,  $\alpha > 0$  — некоторое фиксированное положительное число. В частности, условие (12) верно для задачи евклидова проектирования точки  $x$  на выпуклый компакт  $X_* \subset Q$ , причем  $f^* = 0$ . Условие острого минимума было введено Б. Т. Поляком в [11]. Рассмотрим субградиентный метод (2) с шагом Б. Т. Поляка (3) для задачи минимизации  $f$ .

**Теорема 2.** Пусть  $f$  — выпуклая функция и для задачи минимизации  $f$  с острым минимумом используется алгоритм (2) с шагом Б. Т. Поляка (3). Тогда после  $k$  итераций алгоритма (2) верно неравенство

$$\min_{x_* \in X_*} \|x^{k+1} - x_*\|_2^2 \leq \prod_{i=0}^k \left( 1 - \frac{\alpha^2}{\|\nabla f(x^i)\|_2^2} \right) \min_{x_* \in X_*} \|x^0 - x_*\|_2^2.$$

**Доказательство.** Согласно (5) и условию острого минимума имеем

$$\begin{aligned} \min_{x_* \in X_*} \|x^{k+1} - x_*\|_2^2 &\leq -\frac{\alpha^2}{\|\nabla f(x^k)\|_2^2} \min_{x_* \in X_*} \|x^k - x_*\|_2^2 + \min_{x_* \in X_*} \|x^k - x_*\|_2^2 = \\ &= \left(1 - \frac{\alpha^2}{\|\nabla f(x^k)\|_2^2}\right) \min_{x_* \in X_*} \|x^k - x_*\|_2^2. \end{aligned} \quad (13)$$

Далее, получаем цепочку неравенств

$$\begin{aligned} \min_{x_* \in X_*} \|x^{k+1} - x_*\|_2^2 &\leq \left(1 - \frac{\alpha^2}{\|\nabla f(x^k)\|_2^2}\right) \min_{x_* \in X_*} \|x^k - x_*\|_2^2 \leq \\ &\leq \left(1 - \frac{\alpha^2}{\|\nabla f(x^k)\|_2^2}\right) \left(1 - \frac{\alpha^2}{\|\nabla f(x^{k-1})\|_2^2}\right) \min_{x_* \in X_*} \|x^{k-1} - x_*\|_2^2 \leq \dots \\ &\dots \leq \prod_{i=0}^k \left(1 - \frac{\alpha^2}{\|\nabla f(x^i)\|_2^2}\right) \min_{x_* \in X_*} \|x^0 - x_*\|_2^2. \end{aligned} \quad (14)$$

Теорема 2 доказана.

**Следствие 1.** Если в условиях предыдущей теоремы допустить, что  $f$  удовлетворяет условию Липшица с константой  $M > 0$  (т.е. все нормы субградиентов  $f$  равномерно сверху ограничены этой константой), то можно утверждать сходимость алгоритма (2) со скоростью геометрической прогрессии:

$$\min_{x_* \in X_*} \|x^{k+1} - x_*\|_2^2 \leq \left(1 - \frac{\alpha^2}{M^2}\right)^{k+1} \min_{x_* \in X_*} \|x^0 - x_*\|_2^2.$$

### 2.3. Субградиентные схемы с переключениями для задач с ограничениями

Б. Т. Поляк внес существенный вклад и в изучение задач математического программирования. Ему принадлежит идея так называемых схем с переключениями по продуктивным и непродуктивным шагам (см. [28]). Общая идея подхода заключается в следующем: если в текущей точке значение ограничения достаточно хорошее, то спуск выполняем по целевой функции, а в противном случае — по функции ограничения. Такого типа подходам, которые интересны ввиду малых затрат памяти на итерациях, посвящаются все новые работы как для выпуклых задач большой и сверхбольшой размерностей (см. [29]–[32]), так и для некоторых классов невыпуклых задач (см. [29]). В последние годы некоторыми из авторов статьи были исследованы адаптивные субградиентные методы с переключениями для липшицевых задач выпуклого программирования, в том числе и для некоторых нелипшицевых и/или квазивыпуклых целевых функций (см. [30], [33]–[35]). Достаточно полное исследование стохастического варианта субградиентного метода с переключениями по продуктивным и непродуктивным шагам имеется в работе [36].

В этом пункте приводится результат о сходимости субградиентных методов со скоростью геометрической прогрессии для субградиентных схем с переключениями по продуктивным и непродуктивным шагам в случае выпуклых задач с ограничениями в виде неравенств. Случай квазивыпуклых задач с квазивыпуклыми ограничениями исследован в [35]. Будем рассматривать задачу с функциональными ограничениями вида

$$\begin{aligned} \min f(x), \\ x \in Q, g(x) \leq 0, \end{aligned} \quad (16)$$

где  $f(x)$  и  $g(x)$  — липшицевы функции. Всюду далее будем считать, что задача (16) разрешима.

**Algorithm 1.** Адаптивный субградиентный метод для выпуклой целевой функции

Require:  $\delta > 0, M_g > 0, x^0, \theta_0 : \theta_0^2 \geq \frac{1}{2} \|x^* - x^0\|_2^2$ , множество  $Q$ .

1:  $I =: \emptyset$

2:  $N \leftarrow 0$

3: repeat

4: **if**  $g(x^N) \leq \delta M_g$  **then**

5:  $h_N^f = \frac{\delta}{\|\nabla f(x^N)\|_2^2}$ ,

6:  $x^{N+1} = \text{Pr}_Q(x^N - h_N^f \nabla f(x^N))$ , // “продуктивные шаги”

7:  $N \rightarrow I$ ,

8: **else**

9:  $h_N^g = \frac{\delta}{\|\nabla g(x^N)\|_2}$ ,

10:  $x^{N+1} = \text{Pr}_Q(x^N - h_N^g \nabla g(x^N))$ , // “непродуктивные шаги”

11: **end if**

12:  $N \leftarrow N + 1$ ,

13: **until**  $\frac{2\theta_0^2}{\delta^2} \leq \sum_{k \in I} \frac{1}{\|\nabla f(x^k)\|_2^2} + N - |I|$ .

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

Рассмотрим теперь алгоритм 1, где  $I$  и  $J$  — множества индексов продуктивных и непродуктивных шагов соответственно, а  $|I|$  и  $|J|$  — мощности этих множеств.

Покажем пару примеров, несколько поясняющих смысл использования таких схем для задач с ограничениями.

**Пример 1.** Рассмотрим случай задачи с несколькими ограничениями

$$g(x) = \max\{g_1(x), \dots, g_m(x)\}. \quad (17)$$

Можно сэкономить время работы алгоритма за счет рассмотрения не всех функциональных ограничений на непродуктивных шагах. То есть на непродуктивных шагах вместо субградиента ограничения  $g(x)$  можно рассматривать субградиент любого из функционалов  $g_m(x)$ , для которого верно  $g_m(x^k) > \varepsilon$ . Другие ограничения при этом можно игнорировать. Легко проверить, что результат о сходимости алгоритма 1 при этом сохранится. Аналогичное замечание можно сделать и про иные подходы указанного типа (см. [33]).

**Пример 2.** Интересным примером приложений может быть использование схем с переключениями к задачам проектирования механических конструкций, сводящихся к минимизации функций max-типа с разреженной матрицей  $A$  (см. [32], [37]):

$$\max_y \langle f, y \rangle : g(y) := \max_{1 \leq i \leq d} (\pm \langle a_i, y \rangle - 1) = \max_{1 \leq i \leq 2d} g_i(y) \leq 0. \quad (18)$$

Основное преимущество применения субградиентного метода с переключениями для задач выпуклого программирования заключается в том, что сложность одной итерации сильно понижается за счет разреженности векторов  $a_i, f$ . Из этого следует, что на каждом шаге

$$y^{k+1} = y^k - h_g \cdot \nabla g(y^k) \quad \text{или} \quad y^{k+1} = y^k + h_f \cdot f$$

в векторе  $y$  обновляется не больше чем  $r = O(1)$  элементов. Из того, что в каждом узле встречаются не более чем  $s$  стержней, следует, что обновление  $y$  влечет за собой обновление максимум  $sr$  скалярных произведений  $\langle a_i, y \rangle$ .

**Теорема 3.** После остановки алгоритма 1 для всякого  $M_g$ -липшицева квазивыпуклого ограничения  $g$  верно  $f(\hat{x}) - f(x_*) \leq \delta$  и  $g(\hat{x}) \leq \delta M_g$ . При этом в случае  $M_f$ -липшицевой функции  $f$  достаточное количество итераций для выполнения критерия остановки алгоритма 1 оценивается следующим образом:

$$N \geq \frac{2\theta_0^2 \max\{1, M_f^2\}}{\delta^2}.$$

Заметим, что на практике для реализации алгоритма 1 знание константы Липшица  $M_f$  функции  $f(x)$  необязательно, достаточно остановить алгоритм после выполнения критерия остановки.

#### 2.4. Аналог условия острого минимума для задач с ограничениями

Для выпуклых (в том числе негладких) липшицевых задач можно доказать сходимость субградиентного метода со скоростью геометрической прогрессии при дополнительном предположении о выполнении условия острого минимума (см. [35]). В частности (см. [2]), острым минимумом будет обладать задача вида

$$\begin{aligned} \min \langle c, x \rangle, \quad x \in \mathbb{R}^n, \\ Ax \leq b, \quad b \in \mathbb{R}^m, \end{aligned} \quad (19)$$

где  $c$  есть  $n$ -мерный вектор,  $A$  — матрица порядка  $m \times n$ .

Если для линейных задач (19) возможно обойтись использованием обычного условия острого минимума (12), то в общем случае для нелинейных задач это уже не так.

Для такой постановки (16) будем использовать следующую вариацию понятия острого минимума (см. [35], [38]).

**Определение 1.** Будем говорить, что для задачи вида (16) выполняется условие острого минимума (“условный” острый минимум), если при некотором  $\alpha > 0$  для всех  $x$  справедливо неравенство

$$\max\{f(x) - f^*, g(x)\} \geq \alpha \min_{x_* \in X_*} \|x - x_*\|_2. \quad (20)$$

Смысл такого подхода следующий. Поскольку задача (16) с ограничениями в виде неравенств, то в тех точках  $x$ , где эти неравенства нарушены, возможно  $f(x) \leq f^*$ , и тогда выполнение неравенства (20) возможно за счет положительного  $g$  (очевидный пример — когда в качестве ограничения выбирается функция расстояния от точки до допустимого множества). В случае же  $g(x) \leq 0$  потенциально возможна ситуация, когда неравенство  $f(x) - f^* \geq \alpha \min_{x_* \in X_*} \|x - x_*\|_2$  справедливо, а на всем допустимом множестве было бы неверным. Рассмотрим некоторые примеры таких задач с “условным” острым минимумом.

**Пример 3** (специальный вариант линейной задачи, см. [38]):

$$\min -x_1, \quad (21)$$

$$\rho \cos\left(\frac{j\pi}{10}\right)x_1 + \rho \sin\left(\frac{j\pi}{10}\right)x_2 \leq \rho, \quad j = 0, 1, \dots, 19. \quad (22)$$

Как известно (см. [38]), в этой задаче точка минимума  $x_* = (1, 0)$ , а оптимальное значение  $f^* = -1$ . Отдельно целевая функция  $-x_1$ , вообще говоря, не удовлетворяет условию острого минимума (12), поскольку зависит только от одной переменной из двух. Но при наличии ограничений (22) для этой задачи выполнен “условный” вариант острого минимума (20) при  $\alpha = \rho / 2$  (см. [38]). Выбор масштабирующего коэффициента  $\rho$  может влиять на значение параметра такого острого минимума.

**Пример 4.** Рассмотрим задачу

$$\begin{aligned} \min \{f(x_1, x_2) = |x_1| + |x_2|\}, \\ g(x_1, x_2) = \max\{\varepsilon - x_1, \varepsilon - x_2\} \leq 0, \end{aligned}$$

где  $\varepsilon > 0$ . В этом примере целевая функция удовлетворяет условию острого минимума (12). Действительно,  $f(x) = |x_1| + |x_2|$ ,  $f^* = 0$ , а  $x_* = (0, 0)$ , и неравенство  $|x_1| + |x_2| \geq \alpha \|x\|_2$  выполняется, например, при  $\alpha = 1$ . Однако функциональные ограничения смещают точку минимума  $x_* = (\varepsilon, \varepsilon)$ , и условие острого минимума нарушается. А условному острому минимуму эта задача удовлетворяет, учитывая, что  $x_* = (\varepsilon, \varepsilon)$ , а минимальное значение функции  $f^* = 2\varepsilon$ .



**Пример 5** (задача классификации с ограничениями, см. [38]). Рассмотрим множество, состоящее из  $n$  пар:  $\mathcal{D} = \{(a_i, b_i)\}_{i=1}^n$ , где  $a \in \mathbb{R}^p$  — вектор признаков, а  $b_i \in \{1, -1\}$  — множество меток при  $i = 1, 2, \dots, n$ . Пусть  $\mathcal{D}_M \in \mathbb{R}^p$ ,  $\mathcal{D}_F \in \mathbb{R}^p$  — два типа признаков. Мы хотим найти линейный классификатор  $x \in \mathbb{R}^p$ , который не только минимизирует функцию потерь, но и правильным образом обрабатывает каждый элемент из множеств  $\mathcal{D}_F$  и  $\mathcal{D}_M$ . Указанная задача сводится к следующей постановке:

$$\begin{aligned} & \min_x \frac{1}{n} \sum_{i=1}^n \max\{0, 1 - b_i a_i^T x\}, \\ & \frac{1}{n_F} \sum_{a \in \mathcal{D}_F} \sigma(a^T x) \geq \frac{\kappa}{n_M} \sum_{a \in \mathcal{D}_M} \sigma(a^T x), \\ & \frac{1}{n_M} \sum_{a \in \mathcal{D}_M} \sigma(a^T x) \geq \frac{\kappa}{n_F} \sum_{a \in \mathcal{D}_F} \sigma(a^T x), \end{aligned}$$

где  $\kappa \in (0, 1]$  — константа,  $n_M$  и  $n_F$  — количество экземпляров из  $\mathcal{D}_M$  и  $\mathcal{D}_F$  соответственно и  $\sigma(a^T x) := \max\{0, \min\{1, \{0.5 + a^T x\}\}\} \in [0, 1]$  — вероятность присваивания метки +1 предсказанию  $a$ .

Построим схему рестартов алгоритма 1 и 2 в предположении, что верно условие (20).

**Algorithm 2.** Рестарты алгоритма 1

- 
- Require:**  $\varepsilon > 0, \alpha > 0, M_g > 0, x^0, \theta_0 : \theta_0^2 \geq \frac{1}{2} \|x_* - x^0\|_2^2$ , множество  $Q$ .  
 1: Set  $p = 1$ .  
 2: **repeat**  
 3:  $\hat{x}^p$  — результат работы алгоритма 1 с параметрами  $\delta_p, \theta_p, x^0$ , где  
 4:  $x^0 = \hat{x}^p$ ,  
 5:  $\theta_p = \frac{1}{\sqrt{2^p}} \theta_0$ ,  
 6:  $\delta_p = \frac{\alpha \theta_p}{\sqrt{2} \max\{1, M_g\}}$ .  
 7: Set  $p = p + 1$ .  
 8: **until**  $p > \left\lceil 2 \log_2 \frac{\theta_0}{\varepsilon} \right\rceil$ .  
**Ensure:**  $x^p$ .
- 

Для алгоритмов 1 и 2 верна следующая

**Теорема 4.** Пусть  $f(x)$  и  $g(x)$  — липшицевы функции с константами  $M_f$  и  $M_g$  соответственно, удовлетворяющие условию (20), и известна константа  $\theta_0 > 0$  такая, что  $2\theta_0^2 \geq \|x_* - x^0\|_2^2$ . Тогда для алгоритма 1 можно подобрать параметр  $\delta > 0$  так, чтобы после

$$\left\lceil \frac{4}{\alpha^2} \max\{1, M_f^2\} \max\{1, M_g^2\} \right\rceil$$

итераций было выполнено неравенство  $\min_{x_* \in X_*} \|\hat{x} - x_*\|_2 \leq \frac{1}{\sqrt{2}} \theta_0$ . После  $p - 1$  запусков алгоритма 2.4 (рестартов алгоритма 2.3) имеем

$$\min_{x_* \in X_*} \left\| \hat{x}^{p-1} - x_* \right\|_2 \leq \frac{1}{\sqrt{2^p}} \theta_0.$$

Тогда для достижения  $\varepsilon$ -точного решения вида

$$\min_{x_* \in X_*} \left\| \hat{x}^{p-1} - x_* \right\|_2 \leq \varepsilon$$

достаточное количество обращений к субградиенту  $f$  или  $g$  можно оценить как

$$\left\lceil \frac{4}{\alpha^2} \max\{1, M_f^2\} \max\{1, M_g^2\} \left\lceil 2 \log_2 \frac{\theta_0}{\varepsilon} \right\rceil \right\rceil.$$

*2.5. Субградиентные методы для слабо выпуклых  
и относительно слабо выпуклых задач с острым минимумом*

Введенный Б. Т. Поляком класс выпуклых негладких оптимизационных задач с острым минимумом интересен тем, что на нем субградиентный метод имеет неплохие вычислительные гарантии (сходится со скоростью геометрической прогрессии). Однако представляется, что условие острого минимума достаточно существенно сужает рассматриваемый класс выпуклых задач. В этой связи интересно отметить наблюдение многих статей последних лет (см. [39]–[42]) о том, что многие возникающие в приложениях слабо выпуклые задачи имеют острый минимум. Это касается разных типов задач нелинейной регрессии, восстановления фазы, восстановления матрицы (matrix recovery) и др. Класс слабо выпуклых оптимизационных задач интересен, ему посвящаются все новые работы (см. [29], [39], [43]–[47]), но без дополнительных предположений скоростные гарантии достаточно пессимистичны. Однако в случае острого минимума недавно доказаны результаты о сходимости субградиентного метода со скоростью геометрической прогрессии при условии достаточной близости начальной точки к точному решению (см. [39]). При этом заметим, что в одном из этих результатов используется способ выбора шага, предложенный Б. Т. Поляком. Немного расскажем об этом направлении.

Как и ранее, рассматриваем задачи минимизации вида (1). Напомним, что функция  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  (аналогично и для функций  $f : Q \rightarrow \mathbb{R}$ ) называется  $\mu$ -слабо выпуклой ( $\mu \geq 0$ ), если функция  $x \rightarrow f(x) + \frac{\mu}{2} \|x\|_2^2$  выпукла. Для недифференцируемых  $\mu$ -слабо выпуклых функций  $f$  под субдифференциалом  $\partial f(x)$  в точке  $x$  можно понимать (см. [39] и цитируемую там литературу) множество всех векторов  $\nabla f(x) \in \mathbb{R}^n$ , удовлетворяющих неравенству

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + o(\|y - x\|_2) \quad \text{при } y \rightarrow x. \quad (23)$$

Известно (см. [39]), что все векторы-субградиенты  $\nabla f(x) \in \mathbb{R}^n$  из (23) автоматически удовлетворяют неравенству

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle - \frac{\mu}{2} \|y - x\|_2^2 \quad \forall x, y \in \mathbb{R}^n, \quad \nabla f(x) \in \partial f(x). \quad (24)$$

Можно проверить, что слабо выпуклые функции локально липшицевы, и поэтому в качестве субградиентов можно использовать произвольный вектор из субдифференциала Кларка (см., например, [43]).

Известно немало примеров прикладных слабо выпуклых задач, которые не являются выпуклыми. Так, слабо выпуклыми будут задачи вида (см., например, [39])

$$\min_x f(x) := h(c(x)),$$

где  $h : \mathbb{R}^m \rightarrow \mathbb{R}$  выпукло и  $M$ -липшицево, а  $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$  является  $C^1$ -гладким отображением с  $\beta$ -липшицевой матрицей Якоби. Нетрудно проверить, что при указанных допущениях  $f$  —  $M\beta$ -слабо выпукла. В частности, в указанный класс задач входят задачи нелинейной регрессии с  $h(x) = \|x\|_1$ , где  $\|\cdot\|_1$  — 1-норма.

Интересно, что слабая выпуклость позволяет существенно расширить класс задач с условием острого минимума, предложенного Б. Т. Поляком в [11]. Приведем еще пару достаточно популярных в современных приложениях примеров слабо выпуклых задач с острым минимумом.

**Пример 6** (задача восстановления фазы). Восстановление фазы — распространенная вычислительная задача, имеющая приложения в различных областях, таких как визуализация, рентгеновская кристаллография и обработка речи (см. [40], [41], [45]).

Она сводится к задаче минимизации следующей функции:

$$f(x) = \frac{1}{m} \sum_{i=1}^m \left| \langle a_i, x \rangle^2 - b_i \right|, \quad (25)$$

где  $a_i \in \mathbb{R}^n$  и  $b_i \in \mathbb{R}$  заданы для каждого  $i = 1, 2, \dots, m$ . Данная целевая функция имеет вид  $f(x) := h(c(x)) \rightarrow \min_x$ , где  $h : \mathbb{R}^m \rightarrow \mathbb{R}$  — выпуклая и  $M$ -липшицева функция, а  $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$  — это  $C^1$ -гладкое отображение с  $\beta$ -липшицевым отображением Якоби. Как отмечено выше, такие функции слабо выпуклы. В данном случае функция (25)  $\mu$ -слабо выпукла (см. [44]) при

$$\mu = 2 \max_{1 \leq i \leq m} \|a_i\|_2^2. \quad (26)$$

Более того, при соответствующей модели шума в измерениях функция имеет острый минимум (см. [40], утверждение 3).

**Пример 7** (см. [42]):

$$\text{minimize}_{U \in \mathbb{R}^{n \times r}} \left\{ \xi(U) := \frac{1}{m} \|y - A(UU^T)\|_2^2 \right\}. \quad (27)$$

Рассмотрим *робастную задачу восстановления матрицы небольшого ранга* (см. [42]), в которой измерения искажены *всплесками*. В частности, мы предполагаем, что

$$y = A(X^*) + s^*, \quad (28)$$

где  $s^* \in \mathbb{R}^m$  — вектор всплесков-возмущений, у которого небольшая часть элементов имеет произвольную величину, а остальные элементы равны нулю. Кроме того, множество ненулевых элементов предполагается неизвестным.

Хорошо известно, что  $\ell_2$ -функция потерь чувствительна к возмущениям, что приводит к недостаточной эффективности постановки (27) для восстановления базовой матрицы низкого ранга. Глобальные минимумы  $\xi$  в (27) отклоняются от базовой матрицы низкого ранга из-за всплесков, и большая доля всплесков приводит к большему возмущению. Напротив, функция потерь  $\ell_1$  более устойчива к выбросам и широко используется для обнаружения выбросов (см. [48]–[50]). Это привело к идее использовать функцию потерь  $\ell_1$  и факторизованное представление матричной переменной для решения надежной задачи восстановления малоранговой матрицы:

$$\text{minimize}_{U \in \mathbb{R}^{n \times r}} \left\{ f(U) := \frac{1}{m} \|y - A(UU^T)\|_1 \right\}. \quad (29)$$

Известно (см. [42]), что эта целевая функция слабо выпуклая и обладает острым минимумом.

Отметим, что для слабо выпуклых задач характерны многие проблемы сходимости вычислительных процедур, свойственные общей невыпуклой ситуации. Например, нет возможности гарантировать сходимость градиентного метода даже к локальному минимуму.

**Пример 8.** Действительно (см. [51]), пусть

$$f(x) \equiv f(x^{(1)}, x^{(2)}) = \frac{1}{2}(x^{(1)})^2 + \frac{1}{4}(x^{(2)})^4 - \frac{1}{2}(x^{(2)})^2.$$

Это слабо выпуклая задача, поскольку добавление к  $f$  половины квадрата нормы  $x$  приводит к выпуклой функции. Градиент целевой функции имеет вид

$$\nabla f(x) = \left( x^{(1)}, (x^{(2)})^3 - x^{(2)} \right)^T,$$

откуда следует, что существуют только три точки, которые могут претендовать на локальный минимум:  $x_1^* = (0, 0)$ ,  $x_2^* = (0, -1)$ ,  $x_3^* = (0, 1)$ . Вычисляя матрицу Гессе

$$\nabla^2 f(x) = \begin{pmatrix} 1 & 0 \\ 0 & 3(x^{(2)})^2 - 1 \end{pmatrix},$$

закключаем, что  $x_2^*$  и  $x_3^*$  являются точками изолированного локального минимума, в то время как  $x_1^*$  есть только *стационарная точка* нашей функции. Действительно,  $f(x_1^*) = 0$  и  $f(x_1^* + \varepsilon e_2) = (\varepsilon^4 / 4) - (\varepsilon^2 / 2) < 0$  при достаточно малых  $\varepsilon$ . Рассмотрим траекторию градиентного метода, начинающуюся в точке  $x_0 = (1, 0)$ . Обратим внимание на то, что вторая координата этой точки 0, поэтому вторая координата для  $\nabla f(x_0)$  также 0. Следовательно, вторая координата точки  $x_1$  равна нулю и т.д. Таким образом, вся последовательность точек, образованная градиентным методом, будет иметь нулевую вторую координату, что означает сходимость этой последовательности к  $x_1^*$ .

**Теорема 5** (см. [39]). Пусть  $f$   $\mu$ -слабо выпукла и имеет  $\alpha$ -острый минимум ( $\alpha, \mu > 0$ ), а точка  $x_0$ :  

$$\min_{x_* \in X_*} \|x^0 - x_*\|_2 \leq \alpha\gamma / \mu$$
 для некоторого фиксированного  $\gamma \in (0; 1)$ . Тогда для метода (2) с шагом (3) верно неравенство

$$\min_{x_* \in X_*} \|x^{k+1} - x_*\|_2^2 \leq \prod_{i=0}^k \left( 1 - \frac{\alpha^2(1-\gamma)}{\|\nabla f(x^i)\|_2^2} \right) \min_{x_* \in X_*} \|x^0 - x_*\|_2^2. \quad (30)$$

**Следствие 2.** Если в условиях предыдущей теоремы допустить, что  $f$  удовлетворяет условию Липшица с константой  $M > 0$ , то можно утверждать сходимость алгоритма (2) со скоростью геометрической прогрессии:

$$\min_{x_* \in X_*} \|x^{k+1} - x_*\|_2^2 \leq \left( 1 - \frac{\alpha^2(1-\gamma)}{M^2} \right)^{k+1} \min_{x_* \in X_*} \|x^0 - x_*\|_2^2.$$

Отметим, что в условиях теоремы 5 нулевой (суб)градиент возможен только в точке  $x^k \in X_*$ , поскольку справедливо следующее

**Предложение 1** (окрестность множества  $X_*$  не имеет стационарных точек). Задача минимизации  $\mu$ -слабо выпуклой функции  $f$  с  $\alpha$ -острым минимумом не имеет стационарных точек  $x$ , удовлетворяющих условию

$$0 < \min_{x_* \in X_*} \|x - x_*\|_2 < \frac{2\alpha}{\mu}. \quad (31)$$

### 3. ГЛАДКАЯ ОПТИМИЗАЦИЯ

Пожалуй, основным атрибутом гладкой оптимизации является наличие моментного ускорения, роль которого, по-видимому, впервые была отмечена в 1964 г. в работе Б. Т. Поляка (см. [9]). Отметим, что в этой работе (как и в работе по негладким методам 1969 г., см. [11]) были заложены основы разработки численных методов оптимизации на базе непрерывных аналогов (дифференциальных уравнений, порождающих при дискретизации итерационный процесс). Впоследствии аналогичным образом неоднократно разрабатывались новые численные методы оптимизации (см. [52]–[55]). Однако наличие гладкости дает возможность не только ускорения скорости сходимости. В данном разделе, следуя работам Б. Т. Поляка в основном 60-х годов прошлого века, продемонстрировано, какие дополнительные возможности приобретаются при наличии гладкости (липшицевости градиента целевой функции).

#### 3.1. Условие градиентного доминирования (Поляка–Лоясиевича)

В 1963 г. Б. Т. Поляк (см. [1]) предложил простое условие, достаточное для того, чтобы показать глобальную линейную скорость сходимости градиентного спуска для достаточно гладких задач. Это условие является частным случаем неравенства Лоясиевича, предложенного в том же году (см. [56]), и не требует сильной выпуклости (или даже выпуклости). Условие вида (32) называется условием градиентного доминирования или же условием Поляка–Лоясиевича (PL-условием) (см. [1], [56]–[58]):

$$f(x) - f(x_*) \leq \frac{1}{2\mu} \|\nabla f(x)\|_2^2 \quad \forall x \in \mathbb{R}^n. \quad (32)$$

Как правило, в литературе это условие называют условием Поляка–Лоясиевича, но стоит отметить еще работу Лежанского (см. [57]), в которой оно появилось независимо. По-видимому, правильнее называть (32) условием Лежанского–Поляка–Лоясиевича. Зачастую в теоретических приложениях данное условие используется как ослабление сильной выпуклости, так как любая  $\mu$ -сильно выпуклая функция удовлетворяет PL-условию с константой  $\mu$ .

В последние годы условие градиентного доминирования обширно исследуется в различных областях оптимизации и смежных науках. Толчком к возрождению интереса к этому условию послужила работа [58], в которой авторы показали, что PL-условие слабее тех условий, которые ранее использовались для получения линейной скорости сходимости без сильной выпуклости. Также в этой работе PL-условие использовалось для проведения нового анализа рандомизированных и жадных методов покоординатного спуска, а также стохастических градиентных методов в различных постановках. Было предложено и обобщение результатов на проксимальные градиентные методы, позволяющее несложно доказать их линейную скорость сходимости. Попутно были получены результаты сходимости для широкого круга задач машинного обучения: метода наименьших квадратов, логистической регрессии, бустинга, устойчивого метода обратного распространения ошибки, L1-регуляризации, метода опорных векторов.

Приведем несколько примеров задач, в которых условие градиентного доминирования возникает естественным образом.

**Пример 9** (PL-условие для нелинейных задач). Пусть  $f(x) = \frac{1}{2} \|\Phi(x) - y\|_2^2$ , где  $\Phi: \mathbb{R}^d \rightarrow \mathbb{R}^n$  дифференцируемо. Тогда  $f$  удовлетворяет PL-условию, если существует такое  $\mu > 0$ , что для всех  $x \in \mathbb{R}^d$  имеет место равномерная невырожденность матрицы Якоби:

$$\lambda_{\min} \left( \frac{\partial \Phi(x)}{\partial x} \cdot \left[ \frac{\partial \Phi(x)}{\partial x} \right]^T \right) \geq \mu, \quad (33)$$

где

$$\frac{\partial \Phi(x)}{\partial x} = \left( \frac{\partial \Phi_i(x)}{\partial x_j} \right)_{i,j=1}^{n,d} = D\Phi(x).$$

Действительно, достаточно написать

$$\|\nabla f(x)\|_2^2 = \|D\Phi(x)^T(\Phi(x) - y)\|_2^2 \geq \mu \|\Phi(x) - y\|_2^2 = 2\mu f(x) \geq 2\mu(f(x) - f(x^*)).$$

В качестве  $\Phi(x)$  может выступать, например, система нелинейных уравнений. В частности, из этого можно получить, что функции Розенброка и Нестерова–Скокова локально удовлетворяют условию Поляка–Лоясиевича.

Обратим внимание, что допущение (33) предполагает  $d \geq n$ . Это верно, если, например,  $\Phi$  задает перепараметризованную нейронную сеть (более точные рассуждения с опорой на использование структуры нейронной сети см. в [59]).

В некоторых задачах целевая функция обладает PL-условием не на всем пространстве, а только на каком-то множестве. Однако, если траектория вычислительного метода не выводит итеративную последовательность за рамки этого множества, то, конечно, можно применять PL-условие для анализа сходимости метода.

**Пример 10** (композиция сильно (строго) выпуклой и линейной функций). В [58] показано, что если функция  $f$  имеет вид  $f(x) = g(Ax)$ , где  $g$  — сильно выпукла, то  $f$  удовлетворяет PL-условию. Функции данного вида часто возникают в машинном обучении (например, метод наименьших квадратов). Если ослабить условие сильной выпуклости функции  $g$  до строгой выпуклости, то функция  $f$  также будет удовлетворять PL-условию, однако уже не на всем пространстве, а только на произвольном компакте. Например, из данного факта следует, что целевая функция задачи логистической регрессии

$$f(x) = \sum_{i=1}^n \log(1 + \exp(b_i a_i^T x))$$

локально (т.е. на всяком компакте) обладает PL-условием.

Приведем также пример функции с условием градиентного доминирования, связанной с теорией оптимального управления. Этот достаточно интересный класс функций с условием градиентного доминирования был получен совсем недавно И.Ф. Фатхуллиным и Б.Т. Поляком в [60].

**Пример 11** (см. [60]). Пусть задана линейная система управления

$$\dot{x}(t) = Ax(t) + Bu(t), \quad y = Cx,$$

где  $x(t) \in \mathbb{R}^n$  — состояние системы, а  $y(t) \in \mathbb{R}^l$  — управление; с начальными условиями  $x(0)$ , распределенными случайно с нулевым средним и ковариационной матрицей  $\Sigma$  ( $\mathbb{E}x(0)x(0)^T = \Sigma$ ); с критерием качества вида

$$f(K) = \mathbb{E} \int_0^{\infty} (x(t)^T Q x(t) + u(t)^T R u(t)) dt, \quad Q, R \succ 0,$$

заданным на множестве  $S$  с целью найти матрицу обратной связи по состоянию  $u(t) = -Kx(t)$ , минимизирующую  $f(K)$ . Тогда, если известен  $K_0$  — стабилизирующий регулятор, то градиент  $\nabla f(K)$  липшицев на множестве уровня  $S_0 = \{K : f(K) \leq f(K_0)\}$ , а для  $f$  выполнено условие градиентного доминирования со следующей константой  $\mu$ :

$$\mu = \frac{\lambda_1(R)\lambda_1^2(\Sigma)\lambda_1(Q)}{8f(K_*) \left( \|A\|_2 + \frac{\|B\|_2^2 f(K_0)}{\lambda_1(\Sigma)\lambda_1(R)} \right)^2},$$

где  $K_*$  — точка минимума  $f$  на множестве  $S$ . Здесь  $\lambda_i(X)$  — собственные числа матрицы  $X \in \mathbb{R}^{m \times m}$ , занумерованные в порядке возрастания их действительных частей, т.е.  $\text{Re } \lambda_1(X) \leq \text{Re } \lambda_2(X) \leq \dots \leq \text{Re } \lambda_m(X)$ .

Отметим, что функция  $f(K)$  достаточно гладкая (имеет липшицев градиент), но невыпуклая.

Таким образом, условие Поляка–Лоясиевича верно для достаточно большого класса невыпуклых задач. Тем не менее, оно позволяет обосновать сходимость градиентного метода со скоростью геометрической прогрессии. Приведем оценку скорости сходимости градиентного метода для  $L$ -гладких функций с условием Поляка–Лоясиевича. Будем рассматривать задачи минимизации функции  $f$ , удовлетворяющей PL-условию (32), а также условию Липшица градиента

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L \|x - y\|_2 \quad \forall x, y \in \mathbb{R}^n. \quad (34)$$

Справедлива следующая

**Теорема 6.** Пусть дана задача безусловной оптимизации  $\arg \min f(x)$ , где функция  $f$  является  $L$ -гладкой и удовлетворяет PL-условию с константой  $\mu$ . Тогда градиентный метод с постоянным шагом  $1/L$  вида

$$x^{k+1} = x^k - \frac{1}{L} \nabla f(x^k) \quad (35)$$

имеет линейную скорость сходимости, т.е.

$$f(x^k) - f(x_*) \leq \left(1 - \frac{\mu}{L}\right)^k (f(x^0) - f(x_*)) \leq \exp\left(-\frac{\mu}{L}k\right) (f(x^0) - f(x_*)). \quad (36)$$

**Доказательство.** В силу  $L$ -гладкости функции  $f$

$$f(x^{k+1}) \leq f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2} \|x^{k+1} - x^k\|_2^2.$$

Пользуясь правилом обновления (35), получаем

$$f(x^{k+1}) \leq f(x^k) - \frac{1}{2L} \|\nabla f(x^k)\|_2^2. \quad (37)$$

Из PL-условия имеем

$$f(x^k) - f(x_*) \leq \frac{1}{2\mu} \|\nabla f(x^k)\|_2^2. \quad (38)$$

Из (37) и (38) путем преобразований следует

$$f(x^{k+1}) - f(x_*) \leq \left(1 - \frac{\mu}{L}\right) (f(x^k) - f(x_*)).$$

Рекурсивное применение последнего неравенства и дает требуемый результат (36). Теорема 6 доказана.

Доказанная верхняя оценка скорости сходимости из теоремы 6 известна уже около 60 лет. До недавнего времени вопрос о ее оптимальности оставался открытым. В последние дни 2022 г. был выложен препринт [61], в котором обоснована оптимальность этой оценки на классе гладких задач с условием Поляка–Лоясиевича.

Отметим, что существует аналог условия Поляка–Лоясиевича для седловых задач, так называемое двухстороннее условие Поляка–Лоясиевича, которое позволяет получить ряд схожих результатов (см. [62]–[64]). Будем рассматривать седловую задачу  $\min_x \max_y f(x, y)$ . Говорят, что функция  $f(x, y)$  удовлетворяет двустороннему PL-условию, если существуют такие константы  $\mu_1$  и  $\mu_2$ , что  $\forall x, y$

$$\|\nabla_x f(x, y)\|_2^2 \geq 2\mu_1 \left( f(x, y) - \min_x f(x, y) \right),$$

$$\|\nabla_y f(x, y)\|_2^2 \geq 2\mu_2 \left( \max_y f(x, y) - f(x, y) \right).$$

Примеры функций, удовлетворяющих двустороннему PL-условию:

- $f(x, y) = F(Ax, By)$ , где  $F(\cdot, \cdot)$  — сильно-выпукло-сильно-вогнутая и  $A, B$  — произвольные матрицы.
- Невыпукло-невогнутая  $f(x, y) = x^2 + 3\sin^2 x \sin^2 y - 4y^2 - 10\sin^2 y$  ( $\mu_1 = 1/16$ ,  $\mu_2 = 1/14$ ) (см. [62]).
- Релаксация задачи Robust Least Squares (RLS):

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} F(x, y);$$

$$F(x, y) := \|Ax - y\|_M^2 - \lambda \|y - y_0\|_M^2,$$

где  $M$  — положительно полуопределена и  $\|x\|_M^2 = x^T M x$ . RLS минимизирует невязку наихудшего случая при заданном ограниченном детерминированном возмущении  $\delta$  на зашумленном векторе измерений  $y_0 \in \mathbb{R}^m$  и матрице коэффициентов  $A \in \mathbb{R}^{m \times n}$ . Саму задачу RLS можно сформулировать так ( $\rho > 0$ ) (см. [65]):

$$\min_{x \in \mathbb{R}^n} \max_{\delta: \|\delta\|_2 \leq \rho} \|Ax - y\|_2^2, \quad \text{где } \delta = y_0 - y \in \mathbb{R}^m.$$

Известно, что при  $\lambda > 1$   $F(x, y)$  удовлетворяет двустороннему PL-условию, поскольку  $F(x, y)$  можно представить как комбинацию аффинной функции и сильно-выпукло-сильно-вогнутой функции.

Для седловых задач с двусторонним условием градиентного доминирования может использоваться вариация градиентного метода — метод градиентного спуска-подъема. Этот подход сейчас довольно популярен (см. [62], [63], [66]), но был подробно проанализирован еще в малоизвестной статье Бакушинского—Поляка [67].

Далее, поговорим о поведении градиентного метода (сходимость и возможные правила ранней остановки) для класса достаточно гладких задач с условием градиентного доминирования, когда градиент целевой функции доступен с некоторой погрешностью. Б. Т. Поляк считал важными вопросы исследования влияния погрешностей доступной информации на гарантии сходимости численных методов, занимался ими и уделил им немало внимания в своей книге [2]. Так, этой тематике посвящена одна из его последних работ — [68], о которой мы немного расскажем.

Сфокусируемся на двух основных известных типах неточной информации о градиенте: абсолютной и относительной погрешностях. Типичными примерами областей, в которых возникает данная проблема неточного градиента, являются безградиентная оптимизация, в которой используется приближенно вычисляемый градиент (см. [69]–[71]), а также оптимизация в бесконечномерных пространствах, связанная с обратными задачами (см. [72], [73]).

Начнем с вопроса влияния на качество выдаваемого решения абсолютной погрешности в градиенте. Будем рассматривать задачу минимизации функции  $f$ , удовлетворяющей PL-условию (32), а также условию Липшица градиента (34). Считаем, что методу доступно не точное, а приближенное значение градиента  $\tilde{\nabla}f(x)$  в любой запрашиваемой точке  $x$ :

$$\nabla f(x) = \tilde{\nabla}f(x) + v(x), \quad \text{причем} \quad \|v(x)\|_2 \leq \Delta$$

при некотором фиксированном  $\Delta > 0$ . Тогда (32) означает, что

$$f(x) - f(x_*) \leq \frac{1}{\mu} \left( \|\tilde{\nabla}f(x)\|_2^2 + \Delta^2 \right) \quad \forall x \in \mathbb{R}^n, \quad (39)$$

откуда

$$\|\tilde{\nabla}f(x)\|_2^2 \geq \mu(f(x) - f(x_*)) - \Delta^2 \quad \forall x \in \mathbb{R}^n.$$

Заметим, что вопросы исследования влияния погрешностей градиента на оценки скорости сходимости методов первого порядка уже достаточно давно привлекают многих исследователей (см., например, [2], [74]–[77]). Однако мы сфокусируемся именно на выделенном классе, вообще говоря, невыпуклых задач. Невыпуклость целевой функции задачи, а также использование на итерациях неточно заданного градиента могут приводить к разным проблемам. В частности, при отсутствии каких-либо правил ранней остановки возможно достаточно большое удаление траектории градиентного метода от начальной точки. Это может быть проблемным, если начальное положение метода уже обладает определенными хорошими свойствами. С другой стороны, неограниченное удаление траектории градиентного метода в случае градиентного метода с помехами может привести к удалению от искомого точного решения. Опишем некоторые ситуации такого типа.

В качестве простого примера не сильно выпуклой функции, удовлетворяющей условию градиентного доминирования, можно рассмотреть

$$f(x) = \langle Ax, x \rangle, \quad (40)$$

где  $A = \text{diag}(L, \mu, 0)$  — диагональная матрица 3-го порядка с ровно двумя положительными элементами  $L > \mu > 0$ . Если для задачи минимизации функции (40) допустить наличие погрешности градиента  $v(x) = (0, 0, \Delta)$  при  $\Delta > 0$ , то при  $x^0 = (0, 0, 0)$  и  $h_k > 0$  последовательность  $x^{k+1} = x^k - h_k \tilde{\nabla}f(x^k)$  стремится к бесконечности при неограниченном увеличении  $k$ . Далее, можно рассмотреть функцию Розенброка двух переменных  $x = (x_{(1)}, x_{(2)})$

$$f(x) = 100 \left( x_{(2)} - (x_{(1)})^2 \right)^2 + (1 - x_{(1)})^2.$$

При  $x^0 = (1, 1) = x_*$  на каждом шаге градиентного метода возможна такая погрешность градиента  $v(x^k)$ , что  $x_{(2)}^k = (x_{(1)}^k)^2$  и без остановки траектория может уходить весьма далеко от точного решения  $x_*$ . Аналогично траектория градиентного метода может уходить к бесконечности для целевой функции двух переменных  $f(x) = (x_{(2)} - (x_{(1)})^2)^2$ .

Немного расскажем об одной из последних работ, подготовленных Б. Т. Поляком с соавторами (см. [68]). В этой статье поставлена цель исследовать оценку расстояния  $\|x^N - x^0\|_2$  для выдаваемых градиентным методом точек  $x^N$  и предложить правило ранней остановки, которое гарантирует некоторый компромисс между

стремлением достичь приемлемого качества выдаваемого методом решения задачи по функции и не столь существенным удалением траектории от выбранной начальной точки метода. Отметим, что правила ранней остановки в итеративных процедурах активно исследуются для самых разных типов задач. По-видимому, впервые идеология раннего прерывания итераций была предложена в статье [78], посвященной методике приближенного решения возникающих при регуляризации некорректных или плохо обусловленных задач (в упомянутой работе рассматривалась задача решения линейного уравнения). Ранняя остановка в этом случае нацелена на преодоление проблемы потенциального накопления погрешностей регуляризации исходной задачи. К данной тематике примыкают известные подходы, связанные с ранней остановкой методов первого порядка в случае использования на итерациях неточной информации о градиенте (см. [2], гл. 6, параграф 1, а также, к примеру, недавний препринт [77]). Однако известные прежде результаты для выпуклых (не сильно выпуклых) задач отличаются по сравнению с полученным в рассматриваемой работе тем, что обычно гарантируется достижение худшего уровня по функции (если сравнить с комментарием после теоремы 2 параграфа 1 гл. 6 из [2]) или оценки вида  $\|x^N - x_*\|_2 \leq \|x^0 - x_*\|_2$  без исследования  $\|x^N - x^0\|_2$ , где  $\{x^k\}_{k \in \mathbb{N}}$  — образуемая методом последовательность,  $x_*$  — ближайшее к точке старта метода  $x^0$  точное решение задачи минимизации  $f$ .

В предположении доступности значений параметров  $L > 0$  и  $\Delta > 0$  применим к задаче минимизации  $f$  градиентный метод вида

$$x^{k+1} = x^k - \frac{1}{L} \tilde{\nabla} f(x^k). \quad (41)$$

Тогда ввиду (34) для метода (41) имеем следующие соотношения:

$$\begin{aligned} f(x^{k+1}) &\leq f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2} \|x^{k+1} - x^k\|_2^2 = \\ &= f(x^k) + \frac{1}{2L} \|\nabla f(x^k) - \tilde{\nabla} f(x^k)\|_2^2 - \frac{\|\nabla f(x^k)\|_2^2}{2L} \leq \\ &\leq f(x^k) + \frac{\Delta^2}{2L} - \frac{1}{2L} \|\nabla f(x^k)\|_2^2, \end{aligned}$$

т.е.

$$f(x^{k+1}) - f(x^k) \leq \frac{\Delta}{2L} - \frac{1}{2L} \|\nabla f(x^k)\|_2. \quad (42)$$

Суммирование неравенств (42) приводит нас к оценке

$$\min_{k=0, N-1} \|\nabla f(x^k)\|_2 \leq \sqrt{\Delta^2 + \frac{2L(f(x^0) - f(x_*))}{N}} \leq \Delta + \sqrt{\frac{2L(f(x^0) - f(x_*))}{N}}, \quad (43)$$

что указывает на потенциальную возможность расходимости градиентного метода с абсолютной погрешностью задания градиента. Конкретные примеры таких ситуаций описаны выше.

С учетом (32) тогда получаем, что

$$f(x^{k+1}) - f(x^k) \leq \frac{\Delta^2}{2L} - \frac{2\mu(f(x^k) - f(x_*))}{2L} = -\frac{\mu}{L}(f(x^k) - f(x_*)) + \frac{\Delta^2}{2L},$$

откуда

$$\begin{aligned} f(x^{k+1}) - f(x_*) &\leq \left(1 - \frac{\mu}{L}\right)(f(x^k) - f(x_*)) + \frac{\Delta^2}{2L} \leq \left(1 - \frac{\mu}{L}\right)^{k+1} (f(x^0) - f(x_*)) + \\ &+ \frac{\Delta^2}{2L} \left(1 + 1 - \frac{\mu}{L} + \dots + \left(1 - \frac{\mu}{L}\right)^k\right) < \left(1 - \frac{\mu}{L}\right)^{k+1} (f(x^0) - f(x_*)) + \frac{\Delta^2}{2\mu}, \end{aligned}$$

т.е.

$$f(x^{k+1}) - f(x_*) \leq \left(1 - \frac{\mu}{L}\right)^{k+1} (f(x^0) - f(x_*)) + \frac{\Delta^2}{2\mu}. \quad (44)$$

Оценки (43) и (44), вообще говоря, неумлучшаемы. К примеру, известны нижние оценки уровня точности по функции  $O(\Delta^2 / (2\mu))$  для градиентного метода с абсолютной погрешностью задания градиента (см.,



например, разд. 2.11.1 пособия [37], а также имеющиеся там ссылки) даже на классе сильно выпуклых функций.

В [68] для градиентного метода с постоянным шагом при достаточно малом значении возмущенного градиента получена теорема, в которой указан уровень точности по функции, который возможно гарантировать после выполнения предложенного правила ранней остановки. Данный результат можно применять ко всяким  $L$ -гладким невыпуклым задачам. Далее, с использованием PL-условия получено уточнение этого результата — достаточное количество итераций градиентного метода для достижения желаемого качества точки выхода  $\hat{x}$  по функции  $f(\hat{x}) - f(x_*) = O(\Delta^2 / \mu)$ .

Теперь перейдем к ситуации относительной неточности градиента. Будем рассматривать поведения методов градиентного типа для выделенного класса достаточно гладких задач с условием Поляка—Лоясиевича в предположении доступности для метода в каждой текущей точке градиента с относительной погрешностью, т.е.

$$\|\tilde{\nabla}f(x) - \nabla f(x)\|_2 \leq \alpha \|\nabla f(x)\|_2, \quad (45)$$

где под  $\tilde{\nabla}f(x)$ , как и ранее, мы понимаем неточный градиент, а  $\alpha \in [0, 1)$  — некоторая константа, указывающая на величину относительной погрешности градиента. Данное условие на неточный градиент было введено и изучено в [2], [79].

Из (45) можно получить вариант условия градиентного доминирования (32) для относительно неточного градиента

$$f(x) - f(x_*) \leq \frac{1}{2\mu(1-\alpha)^2} \|\tilde{\nabla}f(x)\|_2^2.$$

Описанную задачу предлагается также решать методом градиентного типа вида

$$x^{k+1} = x^k - h_k \tilde{\nabla}f(x^k). \quad (46)$$

Оказывается, что можно подобрать так параметры шагов  $h_k$ , чтобы гарантировать сохранение результата о сходимости метода со скоростью геометрической прогрессии в случае относительной неточности градиента (хоть и с большим знаменателем).

Если при реализации метода (46) градиент доступен с известной относительной погрешностью  $\alpha \in [0, 1)$  и известен параметр  $L > 0$ , то выбор шага  $h_k = \frac{1}{L} \frac{(1-\alpha)}{(1+\alpha)^2}$  приводит к следующим результатам (см. параграф 1 из пособия [80] и имеющиеся там ссылки):

$$\begin{aligned} f(x^{k+1}) - f(x^k) &\leq \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2} \|x^{k+1} - x^k\|_2^2 = -h_k \langle \nabla f(x^k), \tilde{\nabla}f(x^k) \rangle + \frac{Lh_k^2}{2} \|\tilde{\nabla}f(x^k)\|_2^2 = \\ &= -h_k \|\nabla f(x^k)\|_2^2 - h_k \langle \nabla f(x^k), \tilde{\nabla}f(x^k) - \nabla f(x^k) \rangle + \frac{Lh_k^2}{2} \|\tilde{\nabla}f(x^k)\|_2^2 \leq \\ &\leq -h_k \|\nabla f(x^k)\|_2^2 + h_k \|\nabla f(x^k)\|_2 \|\tilde{\nabla}f(x^k) - \nabla f(x^k)\|_2 + \frac{Lh_k^2}{2} \|\tilde{\nabla}f(x^k)\|_2^2 \leq \\ &\leq -h_k \|\nabla f(x^k)\|_2^2 + \alpha h_k \|\nabla f(x^k)\|_2^2 + (1+\alpha)^2 \frac{Lh_k^2}{2} \|\nabla f(x^k)\|_2^2 = \left[ -(1-\alpha)h_k + (1+\alpha)^2 \frac{Lh_k^2}{2} \right] \|\nabla f(x^k)\|_2^2. \end{aligned}$$

После подстановки  $h_k$  (которое, заметим, минимизирует выражение в скобках в последнем выражении) получаем

$$f(x^{k+1}) - f(x^k) \leq -\frac{1}{2L} \frac{(1-\alpha)^2}{(1+\alpha)^2} \|\nabla f(x^k)\|_2^2. \quad (47)$$

Пользуясь условием Поляка—Лоясиевича (32), имеем следующую оценку на скорость сходимости по функции:

$$f(x^{k+1}) - f(x_*) \leq \left( 1 - \frac{\mu}{L} \frac{(1-\alpha)^2}{(1+\alpha)^2} \right) (f(x^k) - f(x_*)),$$

т.е.

$$f(x^N) - f(x_*) \leq \left( 1 - \frac{\mu}{L} \frac{(1-\alpha)^2}{(1+\alpha)^2} \right)^N (f(x^0) - f(x_*)).$$

Если вместо PL-условия (32) предполагать выполнение условия сильной выпуклости, то приведенную оценку можно уточнить (см. [80]):

$$\frac{(1-\alpha)^2}{(1+\alpha)^2} \rightarrow O\left(\frac{1-\alpha}{1+\alpha}\right).$$

В работе [82] предлагается адаптивный алгоритм для случая относительной погрешности градиента. По сути, предлагается критерий выхода из итерации, содержащий норму неточного градиента  $\|\tilde{\nabla}f(x^k)\|_2$ . Это обстоятельство затрудняет использование подхода с оценками типа (47) для нормы точного градиента. Поэтому рассматривается альтернативный вариант выбора шага для метода (46) с относительной погрешностью задания градиента, который позволит получить приемлемый аналог оценки (47) для квадрата нормы неточного градиента. Аналогично ранее рассмотренным случаям, пользуясь  $L$ -липшицевостью градиента и условием (45), можно вывести следующее неравенство (см. [82]):

$$f(x^{k+1}) \leq f(x^k) + \langle \tilde{\nabla}f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2} \|x^{k+1} - x^k\|_2^2 + \frac{\alpha}{1-\alpha} \|\tilde{\nabla}f(x^k)\|_2 \|x^{k+1} - x^k\|_2. \quad (48)$$

Таким образом, при  $\alpha \in [0; 0.5)$  для адаптивного варианта градиентного метода (46) с

$$h_k = \frac{1}{L_{k+1}} \frac{1-2\alpha}{1-\alpha} \quad (49)$$

и критерием выхода из итерации (до его выполнения  $L_{k+1}$  увеличивается в 2 раза)

$$f(x^{k+1}) \leq f(x^k) + \langle \tilde{\nabla}f(x^k), x^{k+1} - x^k \rangle + \frac{L_{k+1}}{2} \|x^{k+1} - x^k\|_2^2 + \frac{\alpha}{1-\alpha} \|\tilde{\nabla}f(x^k)\|_2 \|x^{k+1} - x^k\|_2 \quad (50)$$

неравенство (48) гарантирует выход из итерации при  $L_{k+1} \geq L$ . Тогда оценка скорости сходимости метода (46) с шагом (49) имеет вид

$$f(x^N) - f(x_*) \leq \prod_{k=0}^{N-1} \left(1 - \frac{\mu}{L_{k+1}} (1-2\alpha)^2\right) (f(x^0) - f(x_*)). \quad (51)$$

Если  $L \leq L$ , то последнее неравенство означает сходимость рассматриваемого метода со скоростью геометрической прогрессии

$$f(x^N) - f(x_*) \leq \left(1 - \frac{\mu}{2L} (1-2\alpha)^2\right)^N (f(x^0) - f(x_*)).$$

Отметим, что оценка (51) может быть применима в случае неизвестного значения  $L$ . Кроме того, даже в случае известной оценки на  $L$  потенциально возможно улучшение качества точки выхода метода при условии  $L_{k+1} < L$ .

### 3.2. Метод условного градиента (Франк–Вульфа–Левитина–Поляка)

В этом пункте мы вспомним о направлении научных исследований Б. Т. Поляка, связанном с методами условного градиента. К примеру, таким методам посвящена достаточно известная работа [10]. Несмотря на то что методы условного градиента по числу итераций проигрывают оптимальным на классе выпуклых гладких задач ускоренным методам (см. п. 3.3) для решения задач со структурой (например, на единичном симплексе или прямом произведении таких симплексов), по времени работы эти методы могут быть существенно эффективнее ускоренных из-за возможной дешевизны итерации (см. [83]–[86]). На данный момент методы условного градиента достаточно хорошо разработаны (см., например, обзорную статью [87] и монографию [88]). В частности, среди последних достижений можно отметить безградиентные варианты метода условного градиента для задач стохастической оптимизации (см. [89]), а также вариант метода условного градиента для децентрализованных оптимизационных задач на меняющихся графах (см. [90]).

Рассмотрим задачу выпуклой оптимизации ( $f(x)$  — выпуклая функция,  $Q$  — ограниченное выпуклое множество):

$$\min_{x \in Q} f(x). \quad (52)$$

Обозначим решение задачи (52) через  $x_*$  (если решение не единственно, то  $x_*$  — какое-то из решений).

**Algorithm 3.** Метод условного градиента (см. [10], [88])

**Require:**  $f$  — выпуклая, непрерывно дифференцируемая функция;  $Q$  — допустимое множество, выпуклое и компактное;  $x^1$  — начальная точка;  $N$  — количество итераций.

**Ensure:** точка  $x^N$

1: **for**  $k = 1, \dots, N - 1$  **do**

2:  $\gamma_k = \frac{2}{k+1}$ ,  $0 \leq \gamma_k \leq 1$

3:  $y^k = \arg \min_{y \in Q} \langle \nabla f(x^k), y \rangle$

4:  $x^{k+1} = (1 - \gamma_k)x^k + \gamma_k y^k$

5: **end for**

6: **return**  $x^N$

**Теорема 7** (см. [10], [88]). Пусть  $\nabla f(x)$  удовлетворяет на  $Q$  условию Липшица с константой  $L$  по отношению к некоторой норме  $\|\cdot\|$ : для всех  $x, y \in Q$  выполняется

$$\|\nabla f(y) - \nabla f(x)\|_* \leq L\|y - x\|,$$

$R = \sup_{x, y \in Q} \|x - y\|$ ,  $\gamma_k = 2 / (k + 1)$  для  $k \geq 1$ . Тогда для любого  $N \geq 2$  выполняется

$$f(x^N) - f(x_*) \leq \frac{2LR^2}{N+2}. \quad (53)$$

**Доказательство.** Из условия Липшица на градиент  $f(x)$  и выпуклости  $f(x)$  имеем оценку

$$0 \leq f(x) - f(y) - \langle \nabla f(y), x - y \rangle \leq \frac{L}{2} \|x - y\|^2$$

для всех  $x, y \in Q$ . Получаем цепочку неравенств

$$\begin{aligned} f(x^{k+1}) - f(x^k) &\leq \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2} \|x^{k+1} - x^k\|^2 = \gamma_k \langle \nabla f(x^k), y^k - x^k \rangle + \frac{L\gamma_k^2}{2} \|y^k - x^k\|^2 \leq \\ &\leq \gamma_k \langle \nabla f(x^k), x_* - x^k \rangle + \frac{L\gamma_k^2}{2} R^2 \leq \gamma_k (f(x_*) - f(x^k)) + \gamma_k^2 \frac{LR^2}{2}. \end{aligned}$$

Введем обозначение  $\delta_k = (f(x^k) - f(x_*)) / (LR^2)$ . Тогда неравенство можно переписать в виде

$$\delta_{k+1} \leq (1 - \gamma_k)\delta_k + \frac{\gamma_k^2}{2} = \left(1 - \frac{2}{k+1}\right)\delta_k + \frac{2}{(k+1)^2}.$$

Начиная с неравенства

$$\delta_2 \leq \left(1 - \frac{2}{1+1}\right)\delta_1 + \frac{2}{(1+1)^2} = \frac{1}{2},$$

применением индукции по  $k$  получаем желаемый результат. Теорема 7 доказана.

Как видим из предыдущего результата, при получении оценки скорости сходимости метода условного градиента не важен выбор конкретной нормы. Важно, чтобы параметры  $L$  и  $R$  оценивались относительно согласованных норм  $\|\cdot\|$  и  $\|\cdot\|_*$ .

Следуя [2], покажем, что оценка (53) не может быть улучшена (с точностью до числового множителя). Для этого выберем  $f(x_1, x_2) = x_1^2 + (1 + x_2)^2$ ,  $Q = \{x : |x_1| \leq 1, 0 \leq x_2 \leq 1\}$ . Тогда  $x_* = (0, 0)^T$ . При этом  $y^k = (1, 0)^T$  при  $y_1^k < 0$  и  $y^k = (-1, 0)^T$  при  $y_1^k > 0$ . Несложно показать, что для этого примера  $f(x^N) - f(x_*) \simeq 4 / N$ . Причем эта ситуация типична для случая, когда  $Q$  — многогранник, а минимум достигается в одной из его вершин, поскольку в качестве  $y^k$  могут выступать лишь вершины  $Q$ .

Отметим также, что в рассматриваемой работе Левитина–Поляка [10] впервые для метода условного градиента появилась идея использовать параметр шага, зависящий от константы Липшица градиента целевой функции  $L$ . В современных работах [87], [88], [91] такой шаг обычно применяют в виде

$$\gamma_k = \gamma_k(L) := \min \left\{ -\frac{\nabla f(x^k)^T (y^k - x^k)}{L \|y^k - x^k\|^2}, 1 \right\}, \quad (54)$$

где  $L$  — константа Липшица  $\nabla f$ . Размер соответствующего шага может рассматриваться как результат минимизации квадратичной модели  $m_k(\cdot; L)$ , переоценивающей  $f$  вдоль прямой  $x^k + \gamma_k(y^k - x^k)$ ,

$$m_k(\gamma_k; L) = f(x^k) + \gamma_k \nabla f(x^k)^T (y^k - x^k) + \frac{L\gamma_k^2}{2} \|y^k - x^k\|^2 \geq f(x^k + \gamma_k(y^k - x^k)), \quad (55)$$

где последнее неравенство следует из стандартной леммы о спуске. Интерес такого выбора шага для метода Франк—Вульфа заключается, в частности, в возможности предложить достаточные условия убывания на итерации невязки по функции не менее, чем в 2 раза. Точнее говоря, согласно лемме 2 из [87] для всякой выпуклой функции  $f$  в случае, если на  $k$ -й итерации в (55) верно  $\gamma_k = 1$ , то верно неравенство

$$f(x^{k+1}) - f^* \leq \frac{1}{2} \min \left( L \|y^k - x^k\|^2, f(x^k) - f^* \right). \quad (56)$$

При введении дополнительного условия на функцию, например, градиентного доминирования, для шага вида (54) можно получить и результаты о сходимости метода типа условного градиента со скоростью геометрической прогрессии. Например, известен такой результат для выпуклых функций с условием Поляка—Лоясиевича (см. [87]). При этом такие функции могут не быть сильно выпуклыми (например, целевая функция из задачи логистической регрессии не сильно выпукла, но удовлетворяет условию градиентного доминирования на всяком компакте, см. [58]).

**Теорема 8.** Пусть  $Q$  — компактное выпуклое множество диаметра  $R$ , а  $f$  — выпуклая функция, удовлетворяющая условию градиентного доминирования с константой  $\mu > 0$ . Пусть также существует точка минимума  $f$ , и она лежит во внутренности множества  $Q$  ( $x_* \in \text{Int}(Q)$ ), т.е. существует  $r > 0$  такой, что  $B(x_*, r) \subset Q$ . Тогда при любых  $k \geq 1$  верно

$$f(x^k) - f^* \leq (f(x^0) - f^*) \prod_{i=1}^k \varphi_i,$$

где

$$\varphi_i = \begin{cases} \frac{1}{2}, & \text{если } \alpha_i = 1, \\ 1 - \frac{r^2}{Lc^2 R^2}, & \text{если } \alpha_i < 1, \end{cases}$$

и  $c^2 = 1 / (2\mu)$ . Таким образом, имеет место сходимость по функции со скоростью геометрической прогрессии, знаменатель которой в худшем случае равен  $\max \left\{ \frac{1}{2}, 1 - \frac{r^2}{Lc^2 R^2} \right\}$ .

Отметим, что недавно в работе [91] доказан аналог теоремы 8 для варианта метода условного градиента с адаптивным подбором шага.

Что касается метода условного градиента, исследованного в самой работе Левитина—Поляка [10], то обратим внимание на две отмеченные там его особенности. Первая из них касается нижней оценки скорости сходимости такого метода. Точнее говоря, в замечании 1 на с. 805 [10] отмечено, что оценка  $O(1/k)$  невязки по функции неулучшаема с точностью до константы на классе выпуклых гладких функций  $f$ . Для того времени рассуждения на тему нижних оценок были редки, поскольку направление исследований численных методов оптимизации, основанное на нижних оценках их эффективности для классов задач, зародилось несколько позднее в монографии [52]. Вторая особенность метода условного градиента, отмеченная впервые в [10], касается повышения скоростных гарантий не путем накладывания дополнительных условий на целевую функцию, а посредством сужения класса допустимых множеств задачи  $Q$ . Например, при добавлении требования, что множество  $Q$  является сильно выпуклым. Множество  $Q$  называется сильно выпуклым, если существует функция  $\delta(\tau) = \gamma\tau^2$  ( $\gamma > 0$ ) такая, что  $(x + y) / 2 + z \in Q$  для любых  $x, y \in Q$  и любого  $z$ :  $\|z\| \leq \delta(\|x - y\|)$ . Существуют эквивалентные определения сильной выпуклости множества. Например, как было показано в [92] (теорема 1 на стр. 190), множество является сильно выпуклым тогда и только тогда, когда оно представимо в виде пересечения шаров одного радиуса. Более подробно см. на с. 789 [10], а также в [92]–[94], метод сходится по аргументу со скоростью геометрической прогрессии при условии, что  $\|\nabla f(x)\| \geq \varepsilon > 0$  на  $Q$ .

Исследование разных вариантов метода условного градиента продолжается и в настоящее время (см. [87], [88], [91], [95]). В работе [95] имеется обзор результатов последних лет по оценкам сложности для метода Франк—Вульфа, предложен универсальный (адаптивный по параметру гладкости задачи) вариант метода для задач с равномерно выпуклой структурой. В [91] адаптивный вариант метода Франк—Вульфа

с шагом (54) исследуется для выпуклых задач (вообще говоря, без равномерной структуры), причем детально проработан соответствующий аналог оценки (56).

### 3.3. Ускоренные методы выпуклой оптимизации

Для наглядности изложения рассмотрим задачу безусловной выпуклой оптимизации

$$\min_x f(x), \quad (57)$$

в которой выполняется условие  $L$ -липшицевости градиента в 2-норме (см. (34)), а  $x$  — принадлежит гильбертову пространству. Также для наглядности будем опускать числовые множители в оценках скорости сходимости.

В кандидатской диссертации Б. Т. Поляка в 1963 г. (см. также [8]) было показано, что градиентный спуск на такой задаче

$$x^{k+1} = x^k - \frac{1}{L} \nabla f(x^k)$$

будет сходиться следующим образом:

$$f(x^N) - f(x_*) \lesssim \frac{LR^2}{N},$$

где  $R = \|x^0 - x_*\|_2$  — расстояние от точки старта до ближайшего (в 2-норме) к точке старта решения  $x_*$  задачи (57) (не ограничивая общности, везде в п. 3.3 можно считать, что все константы, типа  $L$  и  $\mu$ , определены на шаре с центром в точке  $x^0$  и радиусом  $2R$ , см. [51], [80]). Оптимальная ли эта оценка? Оказывается, что, даже если выбирать шаг метода  $h = 1/L$  как-то по-другому, улучшить такую оценку на рассматриваемом классе задач можно только на числовой множитель порядка 4 (см. [96]). Однако это совсем не означает, что если рассмотреть какой-то другой метод градиентного типа, то на нем также нельзя будет получить существенно лучшую оценку скорости сходимости. Действительно, как минимум на квадратичных задачах выпуклой оптимизации метод сопряженных градиентов (первая итерация совпадает с итерацией типа градиентного спуска, аналогично и для приводимых далее моментных методов)

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k) + \beta_k (x^k - x^{k-1}), \quad (58)$$

где

$$(\alpha_k, \beta_k) \in \underset{\alpha, \beta}{\operatorname{Argmin}} f(x^k - \alpha \nabla f(x^k) + \beta (x^k - x^{k-1})),$$

дает существенно лучшую оценку скорости сходимости (при дополнительных предположениях о распределении спектра матрицы квадратичной формы оценка может быть дополнительно улучшена, см. [97]):

$$f(x^N) - f(x_*) \lesssim \frac{LR^2}{N^2}. \quad (59)$$

Наиболее тонкое исследование скорости сходимости метода сопряженных градиентов (58) (в том числе в условиях неточностей) имеется в [98]. Стоимость итерации такого метода (ввиду возможности решить задачу поиска  $\alpha_k$  и  $\beta_k$  аналитически для квадратичных задач) будет по порядку такой же, как стоимость итерации градиентного метода. Аналогичные оценки можно написать и для  $\mu$ -сильно выпуклых задач:

$$f(x^N) - f(x_*) \lesssim LR^2 \exp\left(-\frac{\mu}{L} N\right) \quad (60)$$

для градиентного спуска и

$$f(x^N) - f(x_*) \lesssim LR^2 \exp\left(-2\sqrt{\frac{\mu}{L}} N\right) \quad (61)$$

для метода сопряженных градиентов (58) на квадратичной задаче.

Заметим, что одна из форм записи метода сопряженных градиентов (58) в случае  $\mu$ -сильно выпуклой целевой функции приводит к методу Чебышёва (см. [99]):

$$x^{k+1} = x^k - \frac{4\delta_k}{L - \mu} \nabla f(x^k) + \left(\frac{2\delta_k(L + \mu)}{L - \mu} - 1\right) (x^k - x^{k-1}), \quad (62)$$

$$x^1 = x^0 - \frac{2}{L + \mu} \nabla f(x^0),$$

$$\delta_{k+1} = \frac{1}{2 \frac{L + \mu}{L - \mu} - \delta_k}, \quad \delta_1 = \frac{1}{2 \frac{L + \mu}{L - \mu} + 1}.$$

Метод тяжелого шарика Поляка (или импульсный / моментный метод Поляка) (см. [9]) при этом выглядит так:

$$x^{k+1} = x^k - \frac{4}{(\sqrt{L} + \sqrt{\mu})^2} \nabla f(x^k) + \frac{(\sqrt{L} - \sqrt{\mu})^2}{(\sqrt{L} + \sqrt{\mu})^2} (x^k - x^{k-1}). \quad (63)$$

Этот метод получается в асимптотике  $k \rightarrow \infty$  из метода Чебышёва (62), поскольку

$$\delta_\infty = \frac{1}{2 \frac{L - \mu}{L + \mu} - \delta_\infty},$$

следовательно,

$$\delta_\infty = \frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}}.$$

Приведенный вывод (с учетом оптимальности метода Чебышёва на квадратичных задачах) дает надежду, что метод тяжелого шарика (63) в некотором смысле асимптотически оптимальный. Оказывается, что “в среднем” так оно и есть (см. [100]).

Исходно метод тяжелого шарика (63) был получен с помощью первого метода Ляпунова (см. [2]), который сводит анализ скорости локальной сходимости метода на классе выпуклых задач к анализу скорости глобальной сходимости для квадратичных выпуклых задач. Б. Т. Поляк подбирал коэффициенты  $\alpha$  и  $\beta$  постоянными. Подобранные коэффициенты гарантировали локальную скорость сходимости метода (63), аналогичную скорости сходимости метода сопряженных градиентов (61). Отметим, что при этом имеется некоторая физическая аналогия в полученном таким образом методе (63) с овражным методом Гельфанда–Цетлина (см. [101]). Однако, как показал последующий анализ метода тяжелого шарика (63), в общем случае нельзя гарантировать его глобальную сходимость (см. [102]). Ее можно добиться за счет некоторых дополнительных предположений (о гладкости), но скорость глобальной сходимости получалась уже не лучше, чем у обычного градиентного спуска (см. [103]), причем то, что лучше и не получится, удалось доказать (см. [104]). Также на примере тяжелого шарика (63) хорошо демонстрируется общая особенность ускоренных методов — немонотонное убывание целевой функции по ходу итерационного процесса и наличие циклов длин  $\sim \sqrt{L/\mu}$  (после каждого цикла невязка по функции убывает в  $\sim 2$  раза). Изящное объяснение этому явлению можно найти в [105], [106].

Метод тяжелого шарика (63) сыграл очень важную роль в развитии ускоренных методов выпуклой оптимизации. Общее направление исследований тут можно описать как попытку сделать такую версию метода сопряженных градиентов, для которой бы удалось доказать глобальную оценку скорости сходимости, аналогичную приведенным выше, для обычного метода сопряженных градиентов на квадратичных задачах. Так родились, например, методы Флетчера–Ривса, Полака–Рибьера–Поляка (см. [12]). Точку здесь удалось поставить А. С. Немировскому в конце 70-х годов прошлого века (см. [52], [107]), попутно показав, что полученные оценки скорости сходимости (59) и (61) уже не могут быть дальше улучшены никакими другими методами, использующими градиент целевой функции (улучшить немного можно лишь числовой множитель). То есть получилось, что сложность класса гладких задач (сильно) выпуклой оптимизации с точностью до числовых множителей совпадает со сложностью аналогичных классов задач (сильно) выпуклой квадратичной оптимизации. Отметим, что для распределенной оптимизации (а точнее, федеративного обучения) недавно было обнаружено, что такая аналогия уже не имеет места (см. [108], [109]). Отметим также, что ускоренные методы А. С. Немировского (известно как минимум три таких метода) требовали на каждой итерации решение вспомогательной маломерной задачи оптимизации. Избавиться от этого удалось в кандидатской диссертации Ю. Е. Нестерова (научным руководителем был Б. Т. Поляк) в 1983 г. (см. [110]):

$$x^{k+1} = x^k - \frac{1}{L} \nabla f \left( x^k + \frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}} (x^k - x^{k-1}) \right) + \frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}} (x^k - x^{k-1}). \quad (64)$$

Такой ускоренный (моментный) метод Нестерова будет сходиться аналогично (61) уже глобально на классе гладких  $\mu$ -сильно выпуклых задач (аналогично можно предложить вариант метода и для выпуклых задач  $(\sqrt{L} - \sqrt{\mu}) / (\sqrt{L} + \sqrt{\mu}) \rightarrow (k - 1) / (k + 2)$  в (64)). Отметим, что недавно были предложены (см. обзор [99]) такие варианты ускоренного метода Нестерова (с аналогичными по порядку трудозатратами на каждой итерации), для которых удалось доказать, что их скорость сходимости в выпуклом и сильно выпуклых случаях оптимальны (без оговорок “с точностью до числового множителя”) на классе градиентных методов выпуклой оптимизации, использующих всевозможные линейные комбинации полученных на предыдущих итерациях градиентов (в том числе допускается вспомогательная минимизация на подпространствах, натянутых на полученные градиенты). Например, в  $\mu$ -сильно выпуклом случае оптимальный алгоритм выглядит таким образом.

---

**Algorithm 4.** Ускоренный метод Тейлора–Дрори (см. [99])

---

**Require:**  $f$  –  $\mu$ -сильно выпуклая функция с  $L$ -липшицевым градиентом, точка старта  $x^0$ .

1: **SET:**  $z^0 = x^0$ ,  $A_0 = 0$ ,  $q = \mu / L$

2: **for**  $k = 0, \dots, N - 1$  **do**

3:  $A_{k+1} = \frac{(1+q)A_k + 2(1 + \sqrt{(1+A_k)(1+qA_k)})}{(1-q)^2}$

4:  $\tau_k = 1 - \frac{A_k}{(1-q)A_{k+1}}$  и  $\delta_k = \frac{1}{2} \frac{(1-q)^2 A_{k+1} - (1+q)A_k}{1+q+qA_k}$

5:  $y^k = x^k + \tau_k (z^k - x^k)$

6:  $x^{k+1} = y^k - \frac{1}{L} \nabla f(y^k)$

7:  $z^{k+1} = (1 - q\delta_k)z^k + q\delta_k y^k - \frac{\delta_k}{L} \nabla f(y^k)$

8: **end for**

9: **return**  $z^N$

---

Как ни странно, в 1983 г. статья Ю. Е. Нестерова не вызвала какого-то большого ажиотажа. О ней вспомнили лишь спустя 20 лет уже в новом столетии, когда размеры новых задач, которые требовалось решать в анализе данных, позволяли использовать только градиентные методы (а не, скажем, методы внутренней точки) или их стохастические варианты. Во многом этому способствовало появление в 2004 г. первого издания лекций Ю. Е. Нестерова по выпуклой оптимизации (см. [51]), в которых центральную роль как раз и заняло современное (на тот момент) изложение ускоренных методов. Собственно, вот уже 20 лет идет настоящий бум ускоренных методов (см., например, [51], [111], [112]). Новые тонкие результаты о сходимости таких методов появляются и по сей день (см., например, [113]). Отметим, что ускоренные методы предложены для задач с ограничениями, с более общим понятием проектирования (неевклидова), для задач со структурой (в условиях модельной общности), в условиях неточности градиента и неточности проектирования (см. [75], [114]–[116] и цитированную тут литературу). Например, в малоизвестной работе Б. Т. Поляка [14] было показано, что при наличии малого (сколь угодно малого, но фиксированного по масштабу) аддитивного враждебного шума в градиенте (см. [2])

$$\|\tilde{\nabla}f(x) - \nabla f(x)\|_2 \leq \delta$$

нельзя гарантировать сходимость градиентного спуска. Более того, можно привести пример, когда он будет расходиться. Однако за счет ранней остановки можно решить данную проблему (см. также работу [98], в которой обсуждается ранняя остановка для метода сопряженных градиентов в условиях неточной информации). В частности, аналогичный результат имеет место и для ускоренных методов. Грубо говоря, можно показать, что по аналогии с (59) имеет место

$$f(x^N) - f(x_*) \lesssim \frac{LR^2}{N} + R\delta$$

до тех пор (для таких  $N$ ), пока первое слагаемое не станет меньше второго. В момент, когда это произойдет, нужно останавливать метод, иначе он может уже начать расходиться. При этом, если концепция шума относительная (см. [2]):

$$\|\tilde{\nabla}f(x) - \nabla f(x)\|_2 \leq \alpha \|\nabla f(x)\|_2,$$

где  $\alpha \in [0,1)$ , то, в отличие от неускоренных методов (см. п. 3.1), для ускоренных возможность сохранения порядка скорости сходимости остается только при достаточно малых  $\alpha$  для сильно выпуклых задач и должным образом убывающих на итерациях  $\alpha_k$  для выпуклых задач (см. [116], [117]).

Сложно переоценить роль ускорения в современной оптимизации. Практически все современные методы решения задач больших размерностей (в том числе для многих невыпуклых задач) используют ускорение. Ускорение можно дополнительно структурировать. Например, задачу вида

$$\min_x f(x) + g(x),$$

в случае когда  $f(x)$  — есть  $\mu_f$ -сильно выпуклая и  $L_f$ -гладкая, а  $g(x)$  — есть  $\mu_g$ -сильно выпуклая и  $L_g$ -гладкая, можно решить с относительной точностью  $\varepsilon$  за  $\mathcal{O}\left(\sqrt{L_f / (\mu_f + \mu_g)} \ln(1 / \varepsilon)\right)$  вычислений  $\nabla f$  и  $\mathcal{O}\left(\sqrt{L_g / (\mu_f + \mu_g)} \ln(1 / \varepsilon)\right)$  вычислений  $\nabla g$  (см. [111], [118]). Если использовать стандартное ускорение (не учитывая структуру задачи), то согласно (61) удалось бы получить только такой результат:  $\mathcal{O}\left(\sqrt{(L_f + L_g) / (\mu_f + \mu_g)} \ln(1 / \varepsilon)\right)$  вычислений  $\nabla f$  и  $\nabla g$ , что, очевидно, может быть хуже. Задачу вида

$$\min_{x,y} f(x, y),$$

в случае когда  $f(x, y)$  — выпуклая функция по совокупности аргументов,  $\mu_x$ -сильно выпуклая и  $L_x$ -гладкая по  $x$ , а также  $\mu_y$ -сильно выпуклая и  $L_y$ -гладкая по  $y$ , можно решить с относительной точностью  $\varepsilon$  за  $\mathcal{O}\left(\sqrt{L_x / \mu_x} \ln(1 / \varepsilon)\right)$  вычислений  $\nabla_x f$  и  $\mathcal{O}\left(\sqrt{L_y / \mu_y} \ln(1 / \varepsilon)\right)$  вычислений  $\nabla_y f$  (см. [119]). Если использовать стандартное ускорение (не учитывая структуру задачи), то (при некоторых дополнительных оговорках) согласно (61) удалось бы получить только такой результат:  $\mathcal{O}\left(\sqrt{\max\{L_x, L_y\} / \min\{\mu_x, \mu_y\}} \ln(1 / \varepsilon)\right)$  вычислений  $\nabla_x f$  и  $\nabla_y f$ , что также может быть хуже.

Более того, недавно были получены нетривиальные варианты ускоренных методов для седловых задач со структурой (см. [120]). Например, если рассматривается задача

$$\min_x \max_y p(x) + F(x, y) - q(y),$$

где  $p(x), q(y)$  — выпуклые и  $L_p, L_q$ -гладкие функции, а  $F(x, y)$  —  $L_F$ -гладкая,  $\mu_x$ -сильно выпуклая и  $\mu_y$ -сильно вогнутая, то существует такой ускоренный метод, который решит задачу с относительной точностью  $\varepsilon$  по зазору двойственности (см. [121]) за

$$\mathcal{O}\left(\left(\sqrt{\frac{L_p}{\mu_x}} + \sqrt{\frac{L_q}{\mu_y}}\right) \log \frac{1}{\varepsilon}\right)$$

вычислений  $\nabla p(x), \nabla q(y)$  и

$$\mathcal{O}\left(\max\left\{\sqrt{\frac{L_p}{\mu_x}}, \sqrt{\frac{L_q}{\mu_y}}, \frac{L_F}{\sqrt{\mu_x \mu_y}}\right\} \log \frac{1}{\varepsilon}\right)$$

вычислений  $\nabla F(x, y)$ .

Начиная с пионерской работы Ю. Е. Нестерова и Б. Т. Поляка 2006 г. [7] в нескольких ведущих мировых центрах по оптимизации стали активно разрабатываться тензорные методы (методы, использующие старшие производные). В частности, в [122] было показано, что при естественных условиях тензорные методы второго и третьего порядка (использующие производные второго и третьего порядка) могут быть реализованы практически с той же стоимостью итерации, как у метода Ньютона. Оптимальные (с точностью до числовых множителей) ускоренные тензорные методы были предложены в работах [51], [123]–[127]. Например, оптимальный тензорный метод, использующий производные порядка  $r \geq 2$ , при условии, что целевая функция  $\mu$ -сильно выпуклая и тензор  $r$ -х производных  $M_r$ -липшицев, для достижения относительной точности  $\varepsilon$  требует

$$\mathcal{O}\left(\left(\frac{M_r R^{r-1}}{\mu}\right)^{\frac{2}{3r+1}} + \ln \ln \left(\frac{1}{\varepsilon}\right)\right)$$

итераций (вычислений старших производных). При этом при  $r = 2, 3$  трудозатратность каждой итерации оптимальных методов практически такая же, как и у метода Ньютона. Отметим, что итерационная



(оракульная) сложность (необходимое число итераций) во втором слагаемом отвечает локальной итерационной сложности метода Ньютона. Проблема с методом Ньютона только в том, что нужно оказаться в окрестности квадратичной скорости сходимости. Первое слагаемое в приведенной оценке как раз отвечает за время попадания в эту окрестность. Обратим внимание, что, даже если доступны старшие производные высокого порядка ( $r$  — большое) и целевая функция достаточно гладкая, то второе слагаемое остается по порядку неизменным. Это означает, что локальная скорость сходимости метода Ньютона по порядку оптимальна в классе всевозможных численных методов оптимизации (см. [52]).

Из написанного выше может создаться впечатление, что ускорение возможно заточить под структуру любой гладкой задачи. В п. 3.1 отмечалось, что если вместо (сильной) выпуклости имеет место условие Поляка–Лоясиевича, то в общем случае ускорение невозможно. Кажется, что если ограничиться только (сильно) выпуклыми постановками, то ускорение всегда можно сделать. Как правило, так оно и есть. Но имеются такие постановки выпуклых задач, в которых доказано, что ускорение в общем случае также невозможно. К таким примерам относится задача получения для гладких выпуклых задач оптимальных (по числу коммуникаций и оракульных вызовов — вычислений градиентов) децентрализованных ускоренных методов на меняющихся графах. Децентрализованная оптимизация (см. [128], [129]) является частью распределенной оптимизации. Отметим, что одна из первых работ по распределенной оптимизации была у аспиранта Б. Т. Поляка в конце 70-х годов прошлого века (см. [130]). Бурно развивающийся подраздел распределенной оптимизации — децентрализованная оптимизация на меняющихся со временем графах (см. [131]). Оказывается (см. [132]), что в оценке числа коммуникаций невозможно в общем случае ускорение, потому как входит худшее (на итерациях) число обусловленности коммуникационного графа (если бы граф не менялся, ускорение было бы возможно за счет ускоренного консенсуса — чебышёвское ускорение). Несмотря на приведенный пример, все же во всех известных современных постановках выпуклых гладких задач (с различной структурой), как правило, удается добиться ускорения, учитывающего данную структуру. Причем основные продвижения здесь были получены буквально в последние десять–пятнадцать лет.

Также может показаться, что уже не осталось открытых вопросов, и на все вопросы для задач гладкой выпуклой оптимизации получены ответы. На самом деле это не так. Например, в упомянутой ранее седловой задаче со структурой открытым остается вопрос о возможности дополнительного разделения оракульных сложностей по числу вызовов  $\nabla p(x)$  и  $\nabla q(x)$ . Но можно не опускаться в такие частности и заметить, что даже для самой обычной задачи гладкой сильно выпуклой оптимизации до сих пор не предложен ускоренный метод, адаптивный по всем параметрам  $\mu$ ,  $L$ . При этом в классе неускоренных методов наискорейший спуск решает данную проблему (см. также [133]). К сожалению, адаптивный метод сопряженных градиентов в форме (58), как уже отмечалось, не обязан сходиться с желаемой скоростью на классе (сильно) выпуклых гладких задач. И хотя сейчас есть кандидаты (адаптивные ускоренные методы), которые, возможно, обладают желаемыми свойствами (см. [134]), однако пока это не удалось строго доказать.

#### 4. МЕТОДЫ СТОХАСТИЧЕСКОЙ ОПТИМИЗАЦИИ

Задачей стохастической оптимизации называется задача вида

$$\min_{x \in Q} f(x) := \mathbb{E}_{\xi} f(x, \xi), \quad (65)$$

в которой можно в любой точке  $x$  вызвать оракул (подпрограмму), выдающий  $\nabla f(x, \xi)$  с новой независимой реализацией  $\xi$  (допускается вызов оракула в одной точке  $x$  многократно — см. далее батчирование). При этом  $\mathbb{E}_{\xi} \nabla f(x, \xi) \equiv \nabla f(x)$ . Целью является найти  $\varepsilon$ -приближенное решение задачи (65) (по функции  $f(x)$ ) за наименьшее число вызовов оракула. Без преувеличения можно сказать, что современный анализ данных в алгоритмической своей части — это решение соответствующих задач стохастической оптимизации (см. [135]).

##### 4.1. Стохастический градиентный спуск

По аналогии с градиентным спуском для решения (65) естественно рассматривать так называемые стохастические градиентные спуски (SGD) (см. [136], [137]):

$$x^{k+1} = \pi_Q \left( x^k - \gamma_k \nabla_x f(x^k, \xi^k) \right), \quad (66)$$

где  $\pi_Q$  — обычное (евклидово) проектирование на множество  $Q$ , а  $\xi^k$  выбирается независимо от  $\xi^0, \dots, \xi^{k-1}$ .

Если  $f(x)$  — выпуклая, то, выбирая  $\gamma_k \equiv R / (M\sqrt{N})$ , можно получить

$$\mathbb{E} f(\bar{x}^N) - f(x^*) \leq \frac{MR}{\sqrt{N}},$$

где  $R = \|x^0 - x^*\|_2$  (если  $x_*$  не единственное, то в этой формуле можно выбирать ближайшее по 2-норме к  $x^0$ ),  $\mathbb{E}_\xi \|\nabla f(x, \xi)\|_2^2 \leq M^2$  при  $x \in Q$  (можно сузить на пересечение  $Q$  с некоторым шаром с центром в  $x^0$  и радиусом порядка  $2R$ , см. [138], — аналогичное замечание можно делать и относительно всех остальных констант, вводимых далее),  $\bar{x}^N = \frac{1}{N} \sum_{k=0}^{N-1} x^k$ . Если  $f(x)$  —  $\mu$ -сильно выпуклая функция, то, выбирая  $\gamma_k = 1 / (\mu(k + 1))$ , можно получить

$$\frac{\mu}{2} \mathbb{E} \|\bar{x}^N - x^*\|_2^2 \leq \mathbb{E} f(\bar{x}^N) - f(x^*) \leq \frac{M^2}{\mu N}. \quad (67)$$

Приведенные оценки в общем случае (без дополнительных предположений) не могут быть улучшены (см. [52]). То есть не существует другого способа агрегирования выборки  $\{\xi^k\}_{k=1}^N$ , который давал бы оценки лучше (с точностью до числового множителя) приведенных. Для невыпуклой  $f(x)$  гарантировать сходимость к глобальному минимуму уже нельзя. Тем не менее, на практике (66) и его вариации активно применяются и для невыпуклых задач, например, для обучения нейронных сетей.

Отметим, что для  $Q = \mathbb{R}^n$ , видоизменив сам метод (66), при некоторых дополнительных условиях результат (67) можно уточнить следующим образом (см. [139], приведенная оценка также будет неулучшаемой):

$$\mathbb{E} \|\bar{x}^N - x^*\|_2^2 \lesssim \frac{\text{Tr} \left( [\nabla^2 f(x_*)]^{-1} \Sigma [\nabla^2 f(x_*)]^{-1} \right)}{N} + \mathcal{O} \left( \frac{1}{N^{3/2}} \right), \quad (68)$$

где  $\Sigma = \mathbb{E}_\xi [\nabla f(x_*, \xi) \nabla f(x_*, \xi)^T]$ . Собственно, в этом месте остановимся, чтобы поподробнее рассказать о вкладе Б. Т. Поляка в получение такого рода результатов. В 70-е годы прошлого века Б. Т. Поляк совместно с Я. З. Цыпкиным исследовал следующие псевдоградиентные процедуры стохастического агрегирования (т.е. алгоритмы решения задачи (4)):

$$x^{k+1} = x^k - \gamma_k \phi(\nabla_x f(x^k, \xi^k)),$$

в которых за счет выбора вектор-функции  $\phi(z)$  хотелось получить как можно лучшую скорость сходимости. Базируясь на результатах [140], в работе [13] при аддитивном шуме  $\nabla f(x, \xi) = \nabla f(x) + \xi$  удалось показать, что оптимальным будет такой выбор:  $\gamma_k = k^{-1}$ ,  $\phi(z) = [\nabla^2 f(x_*)]^{-1} J^{-1} \nabla \ln p_\xi(z)$ , где  $p_\xi(z)$  — функция плотности распределения случайного вектора  $\xi$ , а информационная матрица Фишера считается по формуле  $J = \int \nabla \ln p_\xi(z) [\nabla \ln p_\xi(z)]^T p_\xi(z) dz$ . Под оптимальностью понимается следующее: при  $N \rightarrow \infty$  имеет место центральная предельная теорема (ЦПТ) в форме

$$\sqrt{N} (x^N - x_*) \in \mathcal{N} \left( 0, [\nabla^2 f(x_*)]^{-1} J [\nabla^2 f(x_*)]^{-1} \right),$$

и при указанном выше способе выбора  $\phi(z)$  ковариационная матрица наименьшая (в смысле полуопределенного отношения частичного порядка).

Однако во многих реальных приложениях плотность распределения  $p_\xi(z)$  неизвестна. Поэтому использовать ее при выборе  $\phi(z)$  нежелательно. Это приводит к корректировке оптимальной процедуры  $\phi(z) = [\nabla^2 f(x_*)]^{-1} z$  и корректировке основного результата (ЦПТ): при  $N \rightarrow \infty$

$$\sqrt{N} (x^N - x_*) \in \mathcal{N} \left( 0, [\nabla^2 f(x_*)]^{-1} \Sigma [\nabla^2 f(x_*)]^{-1} \right). \quad (69)$$

Здесь использовалось, что  $\nabla f(x_*) = 0$ . Отметим, что аддитивность шума  $\xi$  при этом не требуется. Этот результат также будет оптимальным в классе методов без доступа к  $p_\xi(z)$ . Однако даже в такой формулировке результат едва ли можно назвать практичным, поскольку для задания  $\phi(z)$  требуется знать  $\nabla^2 f(x_*)$ , что возможно, в основном, только для задач квадратичной стохастической оптимизации. Ключевое наблюдение, позволяющее решить отмеченную проблему, пришло в голову Борису Теодоровичу в конце 80-х годов во сне, и оказалось удивительным по простоте (см. [5], [6]):  $\phi(z) = z$ ,  $\gamma_k \sim k^{-\eta}$ ,  $\eta \in (1/2, 1)$ , и в качестве выхода алгоритма предлагается использовать не  $x^N$ , а  $\bar{x}^N$ . В этом случае (69) (с заменой  $x^N$  на  $\bar{x}^N$ ) останется верным. Таким образом, было показано, как избавиться от типично недоступного предобуславливателя

$[\nabla^2 f(x_*)]^{-1}$ . Аналогичное можно проделать и для стохастических вариационных неравенств (см. [141]). Асимптотический вариант (68) очевидным образом получается из (69). Получение неасимптотического варианта требует больших усилий (на появление первых таких результатов ушло еще более 20 лет, см. [142]).

Отметим, что близкая идея (однако реализованная в существенно меньшей общности) использования  $\bar{x}^N$  вместо  $x^N$  независимо была приблизительно в то же время предложена и на западе Д. Руппертом (см. [143]).

Для ряда постановок задач, например, когда множество  $Q$  является симплексом, выгоднее (с точки зрения того, как в итоговую оценку будет входить размерность  $n$  посредством  $M$  и  $R$ ) использовать неевклидово проектирование (в частности, для симплекса лучше использовать проектирование согласно дивергенции Кульбака–Ляйблера, которое приводит к экспоненциальному взвешиванию компонент стохастического градиента). Соответствующие обобщения SGD принято называть стохастическим методом зеркального спуска (stochastic mirror descent — SMD) (см. [52], [144]). Проблему неадаптивного выбора шага  $\gamma_k$  (требуется заранее знать  $N$ ) в выпуклом случае решает вариация SMD — стохастический метод двойственных усреднений (stochastic dual averaging method) (см. [145]). Однако более изящно проблема выбора шага решается в AdaGrad версии SGD (см. [146]), в которой

$$\gamma_k = \frac{R}{\sqrt{\sum_{j=1}^k \|\nabla f(x^j, \xi^j)\|_2^2}}.$$

При таком выборе шага не требуется и знание глобальной константы  $M$ . В современных работах избегаются также и от зависимости  $R$  в шаге (класс Parameter-free SGD, к которому относятся, например, DoG [147] и конструкция Mechanic [148]). К сожалению, в общем сильно выпуклом случае пока неизвестно, как можно было бы избавиться от необходимости знания  $\mu$  (продвижения имеются лишь в частных случаях, например, когда  $f(x^*)$  известно). Напомним, что аналогичная проблема была и для ускоренных детерминированных методов (см. конец п. 3.3).

В случае, если дополнительно известно, что функция  $f(x)$  — гладкая (имеет липшицев градиент), то SGD можно существенно ускорить за счет батч-параллелизации (замены стохастического градиента на выборочное среднее стохастических градиентов на независимых реализациях):

$$\nabla f(x, \xi) \rightarrow \frac{1}{b} \sum_{i=1}^b \nabla f(x, \xi^i),$$

где  $\xi^i$  — независимые одинаково распределенные, как  $\xi$ , а  $b \geq 1$  — размер батча, который можно вычислять параллельно. Действительно, рассмотрим, следуя Б. Т. Поляку [2], более точную оценку скорости сходимости SGD в гладком случае (в [2] используется глобальная оценка дисперсии  $\sigma^2$ , однако несложно показать, что достаточно использовать введенную далее дисперсию в решении  $\sigma_*^2$ , см., например, [149]):

$$\mathbb{E} \left[ \|x^N - x_*\|_2^2 \right] \leq \|x^0 - x_*\|_2^2 (1 - \gamma\mu)^N + \frac{2\gamma\sigma_*^2}{\mu}, \quad (70)$$

где  $\sigma_*^2 = \mathbb{E}_\xi \left[ \|\nabla f(x_*, \xi) - \nabla f(x_*)\|_2^2 \right] \|\nabla f(y, \xi) - \nabla f(x, \xi)\|_2 \leq L \|y - x\|_2$ ,  $\gamma_k \equiv \gamma \leq 1 / (2L)$ . Несложно также показать, что при выборе  $\gamma \equiv 1 / (2L)$  за счет батчирования ( $\sigma_*^2 \rightarrow \sigma_*^2 / b$ ) можно выравнять оба слагаемых в правой части (70) и получить такую версию (60) (тут  $N$  — число вычислений  $\nabla f(x, \xi)$ ):

$$\mathbb{E} \|x^N - x_*\|_2^2 \lesssim R^2 \exp\left(-\frac{\mu}{2L} N\right) + \frac{\sigma_*^2}{\mu^2 N}.$$

При этом для ожидаемой невязки по функции можно получить оценку

$$\mathbb{E} f(x^N) - f(x_*) \lesssim LR^2 \exp\left(-\frac{\mu}{2L} N\right) + \frac{\sigma_*^2}{\mu N}. \quad (71)$$

Отметим, что в последнее время в связи с обучением нейронных сетей огромных размеров возникает потребность в изучении роли перепараметризации, что можно сформулировать как малость дисперсии  $\sigma_*^2$ . Современное состояние развития этого направления для неускоренных стохастических градиентных методов описано, например, в [150]. Малость  $\sigma_*^2$  означает линейную скорость сходимости в небольшую окрестность решения. Такую картину, наверняка, многие наблюдали на практике, решая задачи обучения. А именно, если выбирать шаг  $\gamma$  (learning rate) достаточно большим, то в ряде случаев можно наблюдать линейную скорость сходимости SGD. Но чем больше шаг  $\gamma$ , тем больше окрестность, внутри которой метод перестает сходиться. Для дальнейшего продвижения требуется уменьшение шага или батчирование.

Многое из того, что написано выше, без каких-то существенных изменений переносится и на стохастические вариационные неравенства (седловые задачи) (см., например, обзор [121], написание которого было инициировано Б. Т. Поляком летом 2022 г.). Насколько нам известно, этот обзор, по-видимому, является последней научной работой Бориса Теодоровича.

#### 4.2. Ускоренные версии стохастического градиентного спуска

Прежде всего заметим, что (71) в форме

$$\mathbb{E}f(x^N) - f(x_*) \lesssim LR^2 \exp\left(-\frac{\mu}{2L} N\right) + \frac{\sigma^2}{\mu N},$$

где

$$\mathbb{E}_\xi \left[ \|\nabla f(x, \xi) - \nabla f(x)\|_2^2 \right] \leq \sigma^2,$$

справедливо при более слабом предположении о липшицевости градиента

$$\|\nabla f(y) - \nabla f(x)\|_2 \leq L\|y - x\|_2.$$

В таких же условиях можно улучшить (ускорить) оценку (71), если за основу брать ускоренный детерминированный метод и заменять в нем градиент на должным образом пробатченный стохастический градиент (см., например, [26], [111], [151], простое изложение имеется в [80], [152]) (следует сравнить с (61)):

$$\mathbb{E}f(x^N) - f(x_*) \lesssim LR^2 \exp\left(-\sqrt{\frac{\mu}{4L}} N\right) + \frac{\sigma^2}{\mu N}. \quad (72)$$

Аналогично для выпуклого случая (следует сравнить с (59))

$$\mathbb{E}f(x^N) - f(x_*) \lesssim \frac{LR^2}{N^2} + \frac{\sigma^2 R^2}{\sqrt{N}}. \quad (73)$$

Если дополнительно известно, что

$$\|\nabla f(y, \xi) - \nabla f(x, \xi)\|_2 \leq L\|y - x\|_2,$$

то приведенные оценки можно уточнить следующим образом (см. [153], [154]) (здесь, как и раньше,  $b$  — размер батча, только сейчас мы явно его прописываем, поскольку батчированию тут поддаются слагаемые, не только содержащие  $\sigma_*^2$ ):

$$\begin{aligned} \mathbb{E}f(x^N) - f(x_*) &\lesssim LR^2 \exp\left(-\sqrt{\frac{\mu}{4L}} N\right) + LR^2 \exp\left(-\frac{\mu}{2L} bN\right) + \frac{\sigma_*^2}{\mu bN}, \\ \mathbb{E}f(x^N) - f(x_*) &\lesssim \frac{LR^2}{N^2} + \frac{LR^2}{bN} + \frac{\sigma_*^2 R^2}{\sqrt{bN}}. \end{aligned}$$

Причем все приведенные выше оценки в п. 4.2 имеют место и для задач с ограничениями простой структуры, и для неевклидова проектирования. Более того, все эти оценки оптимальны, т.е. не могут быть в общем случае улучшены (с точностью до числовых множителей, в том числе в показателе экспоненты).

Также как и для обычного SGD на практике большую роль играет адаптивность метода. Добавление различных вариантов моментного ускорения к адаптивным методам (типа AdaGrad), упомянутым в п. 4.1, порождает популярную линейку современных методов типа Adam, AdamW, RMSProp, AdaDelta и т.д., активно использующихся для обучения нейронных сетей. Для выпуклых постановок задач имеется и теоретическое обоснование (см. [155], [156]). Однако вопрос о создании полностью адаптивного ускоренного метода для решения задач выпуклой стохастической оптимизации, насколько нам известно, пока окончательно еще не решен.

В предположении  $Q = \mathbb{R}^n$  отметим концепцию мультипликативных помех, развиваемую в работах Б. Т. Поляка в 70—80-е годы прошлого века (см. [2], [3]). На современный манер условие, которому удовлетворяют введенные помехи, можно было бы называть условием сильного роста (strong growth):

$$\mathbb{E}\|\nabla f(x, \xi)\|_2^2 \leq \rho_{sg} \|\nabla f(x)\|_2^2 + \sigma_{sg}^2, \quad \rho_{sg}, \sigma_{sg} \geq 0. \quad (74)$$

Такому условию в гладком случае, например, удовлетворяют координатные методы (см. [157]), где рандомизация в стохастическом градиенте возникает за счет случайного выбора координаты, по которой считается частная производная вместо вычисления полного градиента, при этом можно выбирать сразу несколько координат (батч) и сэмплировать не обязательно равномерно, а исходя из свойств производных по направлению (см. [158], [159]). Также под неравенство (74) подходят градиенты, к которым применяется оператор сжатия (см. [160]). Такого рода рандомизация используется в распределенной оптимизации для передачи меньшего числа информации. К примерам операторов сжатия относятся и уже упомянутый выше случайный выбор координат, различные рандомизированные квантизации и округления (см. [161]).

Для неускоренных методов, использующих стохастический градиент вида (74), начало построения теории было заложено в уже упомянутых работах [2], [3]. В связи с активным развитием машинного обучения стохастические методы оптимизации стали широко исследоваться в сообществе, в частности, было переоткрыто и предположение сильного роста (см. [162]). На данный момент для неускоренных методов, например для классического SGD вида (66), имеется хорошо разработанная теория сходимости (см., например, обзорную работу [150]). В частности, для  $L$ -гладкой выпуклой целевой функции  $f$  справедлива следующая оценка скорости сходимости после  $N$  итераций SGD:

$$\mathbb{E}f(\bar{x}^N) - f(x_*) \lesssim \frac{\rho_{sg} L \|x^0 - x_*\|_2^2}{N} + \frac{\sigma_{sg} \|x^0 - x_*\|_2}{\sqrt{N}},$$

где  $\bar{x}^N = \frac{1}{N} \sum_{k=0}^{N-1} x^k$ . Если функция является не просто выпуклой, а  $\mu$ -сильно выпуклой, то можно улучшить оценку и получить, что

$$\mathbb{E}\|x^N - x_*\|_2^2 \lesssim \exp\left(-\frac{\mu N}{\rho_{sg} L}\right) \|x^0 - x_*\|_2^2 + \frac{\sigma_{sg}^2}{\mu^2 N}.$$

Для ускоренных вариантов теория немного беднее, но основные результаты уже были получены (см. [163]). Отметим, что классический ускоренный метод (см. [110]) не подходит для такой постановки (74) и необходимо использовать дополнительный моментный член (momentum term) (см. [157], [163]). Тогда в предположении о  $L$ -гладкости и выпуклости функции  $f$  можно получить следующую оценку скорости сходимости:

$$\mathbb{E}f(\bar{x}^N) - f(x_*) \lesssim \frac{\rho_{sg}^2 L \|x^0 - x_*\|_2^2}{N^2} + \frac{\sigma_{sg} \|x^0 - x_*\|_2}{\sqrt{N}},$$

а для  $\mu$ -сильно выпуклой функции

$$\mathbb{E}\|x^N - x_*\|_2^2 \lesssim \exp\left(-\sqrt{\frac{\mu N^2}{4\rho_{sg}^2 L}}\right) \|x^0 - x_*\|_2^2 + \frac{\sigma_{sg}^2}{\mu^2 N}.$$

Отметим, что приведенные результаты удалось с некоторыми оговорками и ослаблением перенести на марковский шум (см. [164]).

Между тем легко заметить, что предположение (74) можно релаксировать до условия слабого роста (weak growth):

$$\mathbb{E}\|\nabla f(x, \xi)\|_2^2 \leq \rho_{wg}(f(x) - f(x_*)) + \sigma_{wg}^2, \quad \rho_{wg}, \sigma_{wg} \geq 0. \quad (75)$$

Если выполнено (74), то для выпуклой и  $L$ -гладкой функции  $\rho_{wg} = 2L\rho_{sg}$  и  $\sigma_{wg} = \sigma_{sg}$ . Условие (75) является не менее распространенным. В частности, одним из популярных примеров применимости (75) является гладкость в среднем, а именно, нам необходимо предположить, что для любой реализации  $\xi$  функция  $f(\cdot, \xi)$  является  $L(\xi)$ -гладкой и выпуклой, и отсюда получить

$$\begin{aligned} \mathbb{E}\|\nabla f(x, \xi)\|_2^2 &\leq 2\mathbb{E}\|\nabla f(x, \xi) - \nabla f(x_*, \xi)\|_2^2 + 2\mathbb{E}\|\nabla f(x_*, \xi)\|_2^2 \leq \\ &\leq 2\mathcal{L}^2(f(x) - f(x_*)) + 2\mathbb{E}\|\nabla f(x_*, \xi)\|_2^2, \end{aligned} \quad (76)$$

где  $\mathcal{L}^2 = \mathbb{E}[L^2(\xi)]$ . Отметим, что константа  $\mathcal{L}$  может быть значительно хуже, чем  $L$  — константа гладкости (липшицевости градиента)  $f$ . Выкладка (76) является самым популярным в литературе примером предположения (75). В частности, оно появляется в работе [165], где авторы прежде всего мотивируются классической задачей наименьших квадратов. В дальнейшем исследование предположений (75) и (76) было

обобщено на неравномерную рандомизацию, которая учитывает свойства батчей. В [166] предлагается довольно исчерпывающая теория для рандомизации вида (76) с разбором большого числа частных случаев. А именно, классический SGD (66) для выпуклой целевой функции  $f$  имеет следующие гарантии сходимости:

$$\mathbb{E}f(\bar{x}^N) - f(x_*) \lesssim \frac{\rho_{wg} \|x^0 - x_*\|_2^2}{N} + \frac{\sigma_{wg} \|x^0 - x_*\|_2}{\sqrt{N}}.$$

Если функция является дополнительно  $\mu$ -сильно выпуклой, то можно получить, что

$$\mathbb{E}\|x^N - x_*\|_2^2 \lesssim \exp\left(-\frac{2\mu N}{\rho_{wg}}\right) \|x^0 - x_*\|_2^2 + \frac{\sigma_{sg}^2}{\mu^2 N}.$$

Говоря о предположениях (74) и (75), важно заметить, что для многих частных случаев  $\sigma_{sg}$  и  $\sigma_{wg}$  равны 0, а это может значительно улучшить гарантии сходимости. Здесь можно отметить популярные и довольно часто встречающиеся примеры перепараметризации ( $\nabla f(x_*, \xi) = 0$  для всех  $\xi$ ) (см. [163]) и интерполяции ( $f(x, \xi) \geq 0$  и  $f(x_*, \xi) = 0$  для всех  $x$  и  $\xi$ ) (см. [167]). Также для упомянутых выше координатных методов справедливо, что  $\sigma_{sg} = 0$ . Но для самых простых методов со сжатием  $\sigma_{sg} \neq 0$ . Это мотивировало сообщество создать более продвинутые методы, использующие компрессию (см. [168]). Но стохастический градиент в данных подходах не получается описать с помощью (74) и (75). Можно ввести более сложное предположение (см. [169]):

$$\begin{aligned} \mathbb{E}\left[\|\nabla f(x^k, \xi^k)\|_2^2 \mid x^k\right] &\leq 2\rho_{\sigma_k,1}(f(x^k) - f(x_*)) + \rho_{\sigma_k,2}\sigma_k^2 + \sigma_{\sigma_k,1}^2, \\ \mathbb{E}\left[\|\sigma_{k+1}^2\|_2^2 \mid x^k\right] &\leq (1-p)\sigma_k^2 + 2(f(x^k) - f(x_*)) + \sigma_{\sigma_k,2}^2. \end{aligned}$$

С помощью него можно унифицировано анализировать не только многие современные методы со сжатием, но и популярные алгоритмы, использующие технику редукции дисперсии (см. [170]–[172]), а также продвинутые координатные методы (см. [173]). Важной деталью данного предположения является наличие вспомогательной последовательности  $\{\sigma_k^2\}$ , которая является уникальной для каждого метода. Эта последовательность обладает важным свойством сходимости, которое и позволяет рассмотреть функцию Ляпунова, состоящую из двух частей: классической вида  $\|x^k - x_*\|_2^2$  или  $f(x^k) - f(x_*)$  и дополнительной, завязанной на  $\sigma_k^2$ . Например, для  $\mu$ -сильно выпуклой функции метод SGD (66) с постоянным шагом  $\gamma_k = \gamma$  таким, что

$$\gamma \leq \min\left\{\frac{1}{\mu}; \frac{1}{\rho_{\sigma_k,1} + \frac{2\rho_{\sigma_k,2}\rho_{\sigma_k,1}\sigma_{\sigma_k,2}}{p}}\right\},$$

может гарантировать следующую оценку сходимости (см. [169]):

$$\mathbb{E}V_N \lesssim \exp\left(-\min\left\{\gamma\mu; \frac{p}{2}\right\}N\right)V_0 + \frac{\left(\sigma_{\sigma_k,1}^2 + \frac{\rho_{\sigma_k,2}\sigma_{\sigma_k,2}^2}{p}\right)\gamma^2}{\min\{\gamma\mu; p\}},$$

где

$$V_k = \|x^k - x_*\|_2^2 + \frac{\gamma^2\rho_{\sigma_k,2}\sigma_k^2}{2p}.$$

Похожие результаты имеются и для выпуклой целевой функции  $f$  (см. [174]), а также для стохастических вариационных неравенств и седловых задач (см. [175], [176]). Насколько нам известно, на данный момент в условиях слабого роста не известно, можно ли добиться ускорения, и если можно, то каким образом.

Наряду с предположениями (74) и (75) можно рассмотреть и похожее условие вида

$$\mathbb{E}\|\nabla f(x, \xi)\|_2^2 \leq \rho_x \|x - x_*\|_2^2 + \sigma_x^2.$$

Касательно него можно выделить работы [132], [164], [177], [178].

Все приведенные выше результаты формулировались в терминах сходимости по математическому ожиданию. Для такой сходимости было достаточно ограниченности второго момента стохастического градиента. В действительности за счет клиппирования на базе описанных методов можно строить робастные версии, которые гарантированно сходятся с такой же скоростью, но уже в терминах вероятностей больших отклонений, причем имеет место почти субгауссовская концентрация (см. [138], [179], [180]). Заметим, что идея клиппирования (нормализации градиента), как способ борьбы с тяжелыми хвостами, в скалярном случае  $n = 1$ , по-видимому, впервые появилась в 1973 г. в работе Б. Т. Поляка и Я. З. Цыпкина [181] как частный случай того, как можно выбирать функцию  $\phi(z) = \min\{1, \lambda / \|z\|_2\} z$  в псевдоградиентной процедуре из п. 4.1. Отметим, что результаты Поляка–Цыпкина, кратко описанные в п. 4.1, недавно были перенесены как раз на постановки задач, в которых аддитивный шум  $\xi$  имеет тяжелые хвосты распределения, в том числе не предполагающие наличие конечной дисперсии у шума (см. [182]).

В заключение отметим, что недавно ускоренные версии тензорных методов типа Нестерова–Поляка были распространены на достаточно гладкие задачи стохастической оптимизации. В частности, для методов второго порядка полученные результаты оптимальны по числу вызовов стохастических градиентов и числу вызовов стохастических гессианов (см. [183]).

### 4.3. Безградиентные методы

Частным случаем стохастики  $\xi$  в  $\nabla f(x, \xi)$  может быть рандомизация, которая не “дана извне”, а привнесена нами самими. Введение в метод рандомизации может иметь разные причины. Например, ярко об этом написано в статье Ю. Е. Нестерова про покомпонентные методы [157] или в фундаментальной статье А. С. Немировского и др. [144] в части рандомизации умножения матрицы на вектор из единичного симплекса. Но, пожалуй, самый известный пример рандомизации в стохастической оптимизации — это рандомизация суммы: для целевого функционала вида взвешенной суммы в качестве стохастического градиента используется случайно выбранное слагаемое (см., например, [80]). Однако в этом пункте будут описаны так называемые безградиентные (поисковые) методы или методы нулевого порядка, в которых рандомизация — это вынужденная мера, связанная с отсутствием необходимой информации. Такие методы периодически встречались в работах Бориса Теодоровича (см., например, [2], [184]) и одна из предложенных им конструкций, которая в последнее время вызывает определенный интерес, будет далее изложена.

Прежде всего рассмотрим выпуклую задачу оптимизации

$$\min_{x \in Q} f(x), \quad (77)$$

которая существенно отличается от предыдущих постановок задач, в частности от (65), тем, что оракул может выдать только значение целевой функции  $f(x)$  в запрошенной точке  $x$ . Такой оракул часто упоминается в литературе как оракул нулевого порядка или безградиентный оракул (см. [185]). Из-за невозможности получить информацию о  $l$ -й производной функции (например, градиент функции  $f$ ) для решения задачи (77) зачастую прибегают к помощи численных методов нулевого порядка, которые основываются на методах первого порядка, заменяя истинный градиент на различные модели аппроксимации градиента (см. [186]). Например, когда функция  $f(x)$  является не просто гладкой, а имеет повышенную гладкость, т.е. функция  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  имеет непрерывные частные производные до  $l$ -го порядка включительно и для всех  $x, z \in Q$  удовлетворяет условию Гёльдера:

$$\left| f(z) - \sum_{0 \leq |n| \leq l} \frac{1}{n!} D^n f(x)(z-x)^n \right| \leq L_\beta \|z-x\|_2^\beta,$$

где  $l < \beta$ ,  $L_\beta > 0$ ,  $n = (n_1, \dots, n_d)$  — мультииндекс,  $n_i > 0$  — целые,  $n! = (n_1! \dots n_d!)$ ,  $|n| = n_1 + \dots + n_d$ , и  $\forall v = (v_1, \dots, v_d) \in \mathbb{R}^d$ ,

а также  $D^n f(x)v^n = \frac{\partial^{|n|} f(x)}{\partial^{n_1} x_1 \dots \partial^{n_d} x_d} v_1^{n_1} \dots v_d^{n_d}$ , то при создании безградиентного алгоритма важно подобрать

такую аппроксимацию градиента, которая будет использовать преимущества повышенной гладкости функции ( $\beta > 2$ , где  $\beta$  — порядок гладкости функции  $f$ ). Такую оценку производной по направлению предложили в 1990 г. Б. Т. Поляк и А. Б. Цыбаков (см. [4]), которая в дальнейшем стала называться “ядерная” аппроксимация и активно использоваться (см. [142], [187]–[189]):

$$\tilde{\nabla} f(x, \mathbf{e}) = d \frac{f(x + \tau \mathbf{e}) - f(x - \tau \mathbf{e})}{2\tau} K(r) \mathbf{e}, \quad (78)$$

где  $\tau > 0$ ,  $\mathbf{e}$  — равномерно распределенный на  $S_2^d(1) := \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ ,  $r$  — равномерно распределенный на отрезке  $[-1, 1]$ ,  $\mathbf{e}$  и  $r$  независимы,  $K : [-1, 1] \rightarrow \mathbb{R}$  — фиксированная функция (ядро), которая удовлетворяет следующим условиям:

$$\mathbb{E}[K(u)] = 0, \quad \mathbb{E}[uK(u)] = 1, \quad \mathbb{E}[u^j K(u)] = 0, \quad j = 2, 3, \dots, l, \quad \mathbb{E}[|u|^\beta |K(u)|] < \infty.$$

Одним из основных достоинств этой аппроксимации градиента является тот факт, что ядерная аппроксимация (78) требует всего два вычисления значения (реализации) функции на итерации, поскольку информация о повышенной гладкости учитывается в “ядре”. Этот факт существенно улучшает оракульную сложность алгоритма, который использует конечно-разностную схему более высокого порядка в качестве оценки градиента (см. [190]), поскольку данная аппроксимация требует большего числа вызовов безградиентного оракула на каждой итерации. К 2020 г. появились интересные результаты о скорости сходимости для безградиентного алгоритма (см. [4], [142], [191], [192]): стохастический метод проекции градиента нулевого порядка, описание которого можно найти в алгоритме 5.

---

**Algorithm 5.** Стохастический метод проекции градиента нулевого порядка

---

1: **Requires:** Ядро  $K : [-1, 1] \rightarrow \mathbb{R}$ , размер шага  $\eta_k$ , сглаживающий параметр  $\tau_k$ .

2: **Initialization:** Сгенерировать скалярные числа  $r_1, \dots, r_N$ , равномерно распределенные на отрезке  $[-1, 1]$ , и вектора  $\mathbf{e}_1, \dots, \mathbf{e}_N$ , равномерно распределенные на единичной Евклидовой сфере  $S_2^d(1)$ .

3: **for**  $k = 1, \dots, N$  **do**

4:  $f_{\xi_k} := f(x_k + \tau_k r_k \mathbf{e}_k) + \xi_k, \quad f_{\xi'_k} := f(x_k - \tau_k r_k \mathbf{e}_k) + \xi'_k$

5:  $\tilde{\nabla}f(x_k, \mathbf{e}_k) := \frac{d}{2\tau_k} (f_{\xi_k} - f_{\xi'_k}) \mathbf{e}_k K(r_k)$

6:  $x_{k+1} := \text{Proj}_Q(x_k - \gamma_k \tilde{\nabla}f(x_k, \mathbf{e}_k))$

7: **end for**

8: **return**  $\{x_k\}_{k=1}^N$ .

---

Как видно из строчки 4 алгоритма 5,  $f_\xi$  выступает в роли безградиентного оракула, где  $\xi \neq \xi'$  — стохастический шум, который характеризует конкретную реализацию (т.е.  $f_\xi$  — это значение целевой функции на реализации  $\xi$ ). Именно поэтому строчку 5 называют аппроксимацией градиента с односточечной обратной связью. Данная концепция стохастического шума (см. [191]–[194]) формально определяется следующим образом:  $\mathbb{E}[\xi^2] \leq \tilde{\Delta}^2$  и  $\mathbb{E}[\xi'^2] \leq \tilde{\Delta}^2$ ,  $\tilde{\Delta} \geq 0$ , а случайные величины  $\xi$  и  $\xi'$  не зависят от  $e$  и  $r$ . Более того, эта концепция шума не требует предположения о нулевом среднем  $\xi$  и  $\xi'$ , поскольку достаточно того, что  $\mathbb{E}[\xi e] = 0$  и  $\mathbb{E}[\xi' e] = 0$ . В табл. 1 представлены результаты работ [142], [191], [192] через зависимости  $N(\epsilon)$  для различных предположений о выпуклости функции (выпуклая/ сильно выпуклая функция), где  $N$  — число последовательных итераций, совпадающее (с точностью до константы) с общим числом обращений к оракулу нулевого порядка  $T = 2N$ . Все оценки табл. 1 соответствуют случаю, когда  $\tilde{\Delta}$  не мало.

После некоторой “паузы” в 2023 г. авторам работы [189] удалось улучшить верхнюю оценку для сильно выпуклого случая за счет более качественного анализа оценки смещения ядерной аппроксимации (78), учитывая, что  $\kappa_\beta = \int |u|^\beta |K(u)| du$ :

$$\left\| \mathbb{E}[\tilde{\nabla}f(x, e)] - \nabla f(x) \right\|_2 \leq \kappa_\beta \frac{L}{(l-1)!} \cdot \frac{d}{d + \beta - 1} \tau^{\beta-1},$$

а также оценки второго момента ядерной аппроксимации (78) с  $\kappa = \int K^2(u) du$ :

$$\mathbb{E} \left[ \left\| \tilde{\nabla}f(x, e) \right\|_2^2 \right] \leq 4d \mathbb{E} \left[ \left\| \nabla f(x) \right\|_2^2 \right] + 4d\kappa L^2 \tau^2 + \frac{\kappa d^2 \tilde{\Delta}^2}{\tau^2}.$$

Основное преимущество данных оценок состоит в том, что смещение больше не зависит от размерности  $d$  асимптотически. Благодаря этому в работе [189] предоставили следующую верхнюю оценку итерационной



сложности (совпадает с оракульной сложностью) для сильно выпуклого случая, размерность в которой не зависит от порядка гладкости:

$$N = \mathcal{O} \left( \frac{d^2 L_{\beta}^{\frac{2}{\beta-1}} \tilde{\Delta}^2}{(\mu \varepsilon)^{\frac{\beta}{\beta-1}}} \right).$$

**Таблица 1.** Зависимость числа итераций  $N$  от желаемой точности задачи  $\varepsilon$ , размерности  $d$ , константы сильной выпуклости  $\mu$  и порядка гладкости функции  $\beta$

	Сильно выпуклый случай	Выпуклый случай
Нижние оценки (2020) [4], [191]	$\Omega \left( \min \left\{ \frac{d^{1+\frac{1}{\beta-1}} L_{\beta}^{\frac{2}{\beta-1}} \tilde{\Delta}^2}{(\mu \varepsilon)^{\frac{\beta}{\beta-1}}}, \frac{d^2 R^2 \tilde{\Delta}^2}{\varepsilon^2} \right\} \right)$	$\Omega \left( \min \left\{ \frac{d^{1+\frac{1}{\beta-1}} L_{\beta}^{\frac{2}{\beta-1}} R^{\frac{2\beta}{\beta-1}} \tilde{\Delta}^2}{\varepsilon^{2+\frac{2}{\beta-1}}}, \frac{d^2 R^2 \tilde{\Delta}^2}{\varepsilon^2} \right\} \right)$
Новицкий и др. (2020) [192]	$\tilde{\mathcal{O}} \left( \frac{d^{2+\frac{1}{\beta-1}} L_{\beta}^{\frac{2}{\beta-1}} \tilde{\Delta}^2}{(\mu \varepsilon)^{\frac{\beta}{\beta-1}}} \right)$	$\tilde{\mathcal{O}} \left( \frac{d^{2+\frac{1}{\beta-1}} L_{\beta}^{\frac{2}{\beta-1}} R^{\frac{2\beta}{\beta-1}} \tilde{\Delta}^2}{\varepsilon^{2+\frac{2}{\beta-1}}} \right)$
Akhavan et al. (2020) [191]	$\tilde{\mathcal{O}} \left( \frac{d^{2+\frac{2}{\beta-1}} L_{\beta}^{\frac{2}{\beta-1}} \tilde{\Delta}^2}{(\mu \varepsilon)^{\frac{\beta}{\beta-1}}} \right)$	$\tilde{\mathcal{O}} \left( \frac{d^{2+\frac{2}{\beta-1}} L_{\beta}^{\frac{2}{\beta-1}} R^{\frac{2\beta}{\beta-1}} \tilde{\Delta}^2}{\varepsilon^{2+\frac{2}{\beta-1}}} \right)$
Bach et al. (2016) [142]	$\mathcal{O} \left( \frac{d^{2+\frac{2}{\beta-1}} L_{\beta}^{\frac{2\beta}{\beta-1}} \tilde{\Delta}^{\frac{2(\beta+1)}{\beta-1}}}{(\mu \varepsilon)^{\frac{\beta+1}{\beta-1}}} \right)$	$\mathcal{O} \left( \frac{d^{2+\frac{2}{\beta-1}} (L_{\beta} R \tilde{\Delta}^2)^{\frac{2\beta}{\beta-1}}}{\varepsilon^{2+\frac{2}{\beta-1}}} \right)$

Нетрудно заметить, что в этих работах идет “борьба” за оптимальную оракульную сложность  $T = 2N$ . Однако, рассматривая безградиентный алгоритм, в последнее время авторы уделяют внимание и другим критериям оптимальности (см. [187]) — оракульной сложности  $T$ , числу последовательных итераций  $N$  и максимально допустимому уровню враждебного шума, при котором все еще удается достичь желаемой точности  $\varepsilon$ . Один из способов улучшения оценок числа последовательных итераций для безградиентного алгоритма — это взять за базу ускоренный алгоритм первого порядка (см., например, [26], [151]) и применить технику батчирования (где  $B$  — размер батча), тем самым достигнув оптимальной оценки в итерационной сложности при  $B \geq 4kd$  (см. [195]):  $N \sim \mathcal{O}(\varepsilon^{-1/2})$  и, получив (“конкурирующие” результаты с табл. 0) общее число обращений к оракулу, в выпуклом случае для любого размера батча  $B$  с  $\rho_B = \max \left\{ 1, \frac{4kd}{B} \right\}$ :

$$T = N \cdot B = \max \left\{ \mathcal{O} \left( \sqrt{\frac{\rho_B^2 L R^2}{\varepsilon}} B \right), \mathcal{O} \left( \frac{d^2 L_{\beta}^{\frac{2}{\beta-1}} \tilde{\Delta}^2}{\varepsilon^{2+\frac{2}{\beta-1}}} \right) \right\}.$$

Кроме того, исследование вопроса о максимально допустимом уровне враждебного шума, возвращаемого безградиентным оракулом со значением целевой функции, является не менее важным, поскольку в некоторых приложениях (см., например, [196]) чем больше уровень враждебного шума  $\Delta$ , тем дешевле вызов безградиентного оракула; безградиентный оракул или оракул нулевого порядка в такой концепции шума принимает следующий вид:  $f_{\delta}(x) = f(x) + \delta(x)$ ,  $|\delta(x)| \leq \Delta$ , т.е. оракул возвращает значение целевой

функции с некоторым ограниченным шумом. Например, в случае повышенной гладкости функции при выполнении условия Поляка–Лоясиевича (32) в работе [188] рассматриваются различные концепции с враждебным шумом, а также демонстрируется показательный результат преимущества рандомизированного алгоритма и эффективность использования ядерной аппроксимации (78). А в случае, когда функция не является гладкой, но гарантируется  $M$ -липшицевость функции  $f(x)$ , такой, что для всех  $x, y \in Q$ :

$$|f(y) - f(x)| \leq M \|y - x\|_2,$$

существуют работы [152], [197]–[199], авторы которых предоставили оценки на максимально допустимый уровень шума  $\Delta$  в различных настройках задачи, совпадающий с верхними границами, полученными в работе [200] для класса выпуклых  $M$ -липшицевых задач оптимизации.

Обзор современного состояния развития безградиентных методов для (сильно) выпуклых задач в условиях шума представлен, например, в [187]. Выше был описан лишь один важный, но все же частный сюжет.

Настоящая статья представляет пополненную расшифровку записи лекции 12 июля 2023 г. А. В. Гасникова на Традиционной школе им. Б. Т. Поляка по оптимизации (<https://ssopt.org/>).

### СПИСОК ЛИТЕРАТУРЫ

1. Поляк Б. Т. Градиентные методы минимизации функционалов // Ж. вычисл. матем. и матем. физ. 1963. Т. 3. № 4. С. 643–653.
2. Поляк Б. Т. Введение в оптимизацию. М.: Наука, 1983.
3. Немировский А.С., Поляк Б. Т., Цыпкин Я. З. Оптимальные алгоритмы стохастической оптимизации при мультипликативных помехах // Докл. АН СССР. 1985. Т. 284. С. 564–567.
4. Поляк Б.Т., Цыбаков А. Б. Оптимальные порядки точности поисковых алгоритмов стохастической оптимизации // Пробл. передачи информ. 1990. Т. 26. № 2. С. 45–53.
5. Поляк Б. Т. Новый метод типа стохастической аппроксимации // Автоматика и телемехан. 1990. № 7. С. 98–107.
6. Polyak B.T., Juditsky A. V. Acceleration of stochastic approximation by averaging // SIAM J. Control and Optimizat. 1992. V. 30. № 4. P. 838–855.
7. Nesterov Y., Polyak B. T. Cubic regularization of Newton method and its global performance // Math. Program. 2006. V. 108. № 1. P. 177–205.
8. Поляк Б. Т. Градиентные методы решения уравнений и неравенств // Ж. вычисл. матем. и матем. физ. 1964. Т. 4. № 6. С. 995–1005.
9. Поляк Б.Т. О некоторых способах ускорения сходимости итерационных методов // Ж. вычисл. матем. и матем. физ. 1964. Т. 4. № 5. С. 791–803.
10. Левитин Е.С., Поляк Б. Т. Методы минимизации при наличии ограничений // Ж. вычисл. матем. и матем. физ. 1966. Т. 6. № 5. С. 787–823.
11. Поляк Б. Т. Минимизация негладких функционалов // Ж. вычисл. матем. и матем. физ. 1969. Т. 9. № 3. С. 509–521.
12. Поляк Б. Т. Метод сопряженных градиентов в задачах на экстремум // Ж. вычисл. матем. и матем. физ. 1969. Т. 9. № 4. С. 807–821.
13. Поляк Б.Т., Цыпкин Я. З. Оптимальные псевдоградиентные алгоритмы адаптации // Автоматика и телемехан. 1980. № 8. С. 74–84.
14. Poljak V. Iterative algorithms for singular minimization problems // Nonlin. Program. Elsevier, 1981. P. 147–166.
15. Poljak V. T. Sharp minimum // “Generalized Lagrangians and applications”. 1982.
16. Гасников А. В. Научный путь Бориса Теодоровича Поляка. Оптимизация // Компьют. исслед. и моделирование. 2023. Т. 15. № 2. С. 235–243.
17. Fradkov A.L., Granichin O. N. Boris Teodorovich Polyak // Cybernet. and Phys. 2023. V. 12(1).
18. Polyak B. T. Subgradient methods: a survey of Soviet research // Nonsmooth Optimizat. 1978. V. 3. P. 5–29.
19. Shor N. Z. Minimization methods for non-differentiable functions. V. 3. Springer Science & Business Media, 2012.
20. Шор Н. З. Методы минимизации недифференцируемых функций и их применения. Киев: Наук. думка, 1979.
21. Drori Y., Teboulle M. An optimal variants of Kelley’s cutting-plane method // Math. Program. 2016. V. 160. № 1. P. 321–351.

22. Stochastic Polyak step-size for sgd: An adaptive learning rate for fast convergence / N. Loizou [et al.] // *Inter. Conf. on Artific. Intelligence and Statist. PMLR*. 2021. P. 1306—1314.
23. Wang X., Johansson M., Zhang T. Generalized Polyak step size for first order optimization with momentum // *arXiv preprint arXiv:2305.12939*; 2023.
24. Hazan E., Kakade S. Revisiting the Polyak step size // *arXiv preprint arXiv:1905.00313*; 2019.
25. Nesterov Y. Universal gradient methods for convex optimization problems // *Math. Program.* 2015. V. 152. № 1. P. 381—404.
26. Гасников А.В., Нестеров Ю. Е. Универсальный метод для задач стохастической композитной оптимизации // *Ж. вычисл. матем. и матем. физ.* 2018. Т. 58. № 1. С. 51—68.
27. Jiang X., Stich S. U. Adaptive SGD with Polyak stepsize and Line-search: Robust Convergence and Variance Reduction // *arXiv preprint arXiv:2308.06058v*; 2023.
28. Поляк Б.Т. Один общий метод решения экстремальных задач // *Докл. АН СССР*. 1967. Т. 174. № 1. С. 33—36.
29. Huang Y., Lin Q. Single-Loop Switching Subgradient Methods for Non-Smooth Weakly Convex Optimization with Non-Smooth Convex Constraints // *arXiv preprint*. 2023.
30. Mirror descent and convex optimization problems with non-smooth inequality constraints / A. Bayandina [et al.] // *Lect. Not. in Math.* 2018. V. 2227. P. 181—213.
31. Lagae S. New efficient techniques to solve sparse structured linear systems, with applications to truss topology optimization. 2017.
32. Nesterov Y. Subgradient methods for huge-scale optimization problems // *Math. Program.* 2014. V. 146. № 1/2. P. 275—297.
33. Адаптивные алгоритмы зеркального спуска в задачах выпуклого программирования с липшицевыми ограничениями / Ф. С. Стонякин [и др.] // *Тр. ИММ УрО РАН*. 2018. Т. 24. № 2. С. 266—279.
34. Mirror descent for constrained optimization problems with large subgradient values of functional constraints / F. S. Stonyakin [et al.] // *Comput. Res. and Model.* 2020. V. 12. № 2. P. 301—317.
35. Адаптивные субградиентные методы для задач математического программирования с квазивыпуклыми функциями / С. С. Аблаев [и др.] // *Тр. Ин-та матем. и мех. УрО РАН*. 2023. Т. 29. № 3. С. 7—25.
36. Tiapkin D., Gasnikov A. Primal-dual stochastic mirror descent for MDPs // *Inter. Conf. Artific. Intelligence and Statist. PMLR*. 2022. P. 9723—9740.
37. Выпуклая оптимизация: учебное пособие / Е. А. Воронцова [и др.]. М.: МФТИ, 2021.
38. A Parameter-free and Projection-free Restarting Level Set Method for Adaptive Constrained Convex Optimization Under the Error Bound Condition / Q. Lin [et al.] // *arXiv preprint*. 2022.
39. Subgradient methods for sharp weakly convex functions / D. Davis [и др.] // *J. of Optimizat. Theory and Appl.* 2018. V. 179. P. 962—982.
40. Duchi J.C., Ruan F. Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval // *Informat. and Inference: J. of the IMA*. 2019. V. 8. № 3. P. 471—529.
41. Eldar Y.C., Mendelson S. Phase retrieval: stability and recovery guarantees // *Appl. Comput. Harmon. Anal.* 2014. V. 36. № 3. P. 473—494.
42. Li X. Nonconvex Robust Low-rank Matrix Recovery // *arXiv preprint*. 2018.
43. Дудов С.И., Осипцев М.А. Характеризация решения задач сильно-слабо выпуклого программирования // *Матем. сб.* 2021. Т. 212. № 6. С. 43—72.
44. Incremental Methods for Weakly Convex Optimization / X. Li [et al.] // *OPT2020: 12th Ann. Workshop on Optimizat. for Mach. Learn.* 2020.
45. Davis D., Drusvyatskiy D., Paquette C. The nonsmooth landscape of phase retrieval // *IMA J. Numer. Analysis*. 2020. V. 40. № 4. P. 2652—2695.
46. Davis D., Drusvyatskiy D., Kellie M. Stochastic model-based minimization under high-order growth // *arXiv preprint*. 2018.
47. Субградиентные методы для слабо выпуклых и относительно слабо выпуклых задач с острым минимумом / Ф. С. Стонякин [и др.] // *Компьютер. исслед. и моделирование*. 2023. Т. 15. № 2. С. 393—412.
48. Li Y., Sun Y., Chi Y. Low-Rank Positive Semidefinite Matrix Recovery from Corrupted Rank-One Measurements // *IEEE Transact. Signal Proces.* 2017. V. 65. P. 397—408.
49. Robust Principal Component Analysis / E. Candes [et al.] // *J. of the ACM*. 2011.

50. A Theory on the Absence of Spurious Solutions for Nonconvex and Nonsmooth Optimization / C. Jozs [et al.] // *NeurIPS*. 2018. P. 2441—2449.
51. Lectures on convex optimization. V. 137 / Y. Nesterov [et al.]. Springer, 2018.
52. Немировский А.С., Юдин Д. Б. Сложность задач и эффективность методов оптимизации. Наука, 1979.
53. Евтушенко Ю. Г. Методы решения экстремальных задач и их применение в системах оптимизации. 1982.
54. Su W., Boyd S., Candes E. A differential equation for modeling Nesterov's accelerated gradient method: theory and insights // *Adv. Neural Informat. Proces. System*. 2014. V. 27.
55. Wilson A.C., Recht B., Jordan M. I. A Lyapunov analysis of accelerated methods in optimization // *J. Machine Learn. Res.* 2021. V. 22. № 1. P. 5040—5073.
56. Lojasiewicz S. Une propri'et'e topologique des sous-ensembles analytiques r'eels // *Les 'equations aux d'eriv'ees partielles*. 1963. V. 117. P. 87—89.
57. Lez'an'ski T. U"ber das Minimumproblem fu"r Funktionale in Banachschen r"äumen // *Mathematische Annalen*. 1963. V. 152. № 4. P. 271—274.
58. 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: Europ. Conf., ECML PKDD 2016, Riva del Garda, Italy, September 19—23, 2016, Proceed., Part I* 16. Springer, 2016. P. 795—811.
59. Liu C., Zhu L., Belkin M. Toward a theory of optimization for over-parameterized systems of non-linear equations: the lessons of deep learning // *arXiv preprint arXiv:2003.00307*; 2020. V. 7.
60. Fatkhullin I., Polyak B. Optimizing static linear feedback: Gradient method // *SIAM J. Control and Optimizat.* 2021. V. 59. № 5. P. 3887—3911.
61. Yue P., Fang C., Lin Z. On the lower bound of minimizing Polyak-Lojasiewicz functions // *The Thirty Sixth Annual Conference on Learning Theory*. PMLR. 2023. P. 2948—2968.
62. Yang J., Kiyavash N., He N. Global convergence and variance-reduced optimization for a class of nonconvex-nonconcave minimax problems // *arXiv preprint arXiv:2002.09621*; 2020.
63. Garg K., Baranwal M. Fixed-Time Convergence for a Class of Nonconvex-Nonconcave Min-Max Problems // *2022 Eighth Indian Control Conference (ICC)*. IEEE. 2022. P. 19—24.
64. Solving a class of non-convex min-max games using iterative first order methods / M. Nouiehed [и др.] // *Adv. Neural Inform. Proces. System*. 2019. V. 32.
65. El Ghaoui L., Lebret H. Robust solutions to least-squares problems with uncertain data // *SIAM J. Matrix Analys. and Appl.* 1997. V. 18. № 4. P. 1035—1064.
66. Муратида А.Я., Стонякин Ф. С. Правила остановки градиентного метода для седловых задач с двусторонним условием Поляка—Лоясиевича // *arXiv preprint arXiv:2307.09921*; 2023.
67. Бакушинский А.Б., Поляк Б. Т. О решении вариационных неравенств // *Докл. АН СССР*. 1974. Т. 219. С. 1038—1041.
68. Stonyakin F., Kuruzov I., Polyak B. Stopping rules for gradient methods for non-convex problems with additive noise in gradient // *J. Optimizat. Theory and Appl.* 2023. P. 1—21.
69. A theoretical and empirical comparison of gradient approximations in derivative-free optimization / A. S. Berahas [et al.] // *Foundat. Comput. Math.* 2022. V. 22. № 2. P. 507—560.
70. Conn A.R., Scheinberg K., Vicente L. N. Introduction to derivative-free optimization. SIAM, 2009.
71. Risteski A., Li Y. Algorithms and matching lower bounds for approximately-convex optimization // *Adv. Neural Inform. Proces. System*. 2016. V. 29.
72. Convex optimization in hilbert space with applications to inverse problems / A. Gasnikov [et al.] // *arXiv preprint arXiv:1703.00267*; 2017.
73. Kabanikhin S. I. Inverse and ill-posed problems: theory and applications. de Gruyter, 2011.
74. Devolder O., Glineur F., Nesterov Y. First-order methods of smooth convex optimization with inexact oracle // *Math. Program.* 2014. V. 146. P. 37—75.
75. Devolder O. Exactness, inexactness and stochasticity in first-order methods for large-scale convex optimization: дис... канд. / Devolder Olivier. CORE UCLouvain Louvain-la-Neuve, Belgium, 2013.
76. d'Aspremont A. Smooth optimization with approximate gradient // *SIAM J. Optimizat.* 2008. V. 19. № 3. P. 1171—1183.
77. Vasin A., Gasnikov A., Spokoiny V. Stopping rules for accelerated gradient methods with additive noise in gradient: тех. отч. / Berlin: Weierstraß-Institut fu"r Angewandte Analysis und Stochastik, 2021.

78. Емелин И.В., Красносельский М.А. Правило останова в итерационных процедурах решения некорректных задач // Автоматика и телемехан. 1978. № 12. P. 59—63.
79. Carter R. G. On the global convergence of trust region algorithms using inexact gradient information // SIAM J. on Numer. Anal. 1991. V. 28. № 1. P. 251—265.
80. Гасников А.В. Современные численные методы оптимизации. Метод универсального градиентного спуска. М.: МЦНМО, 2021.
81. De Klerk E., Glineur F., Taylor A. B. On the worst-case complexity of the gradient method with exact line search for smooth strongly convex functions // Optimizat. Lett. 2017. V. 11. P. 1185—1199.
82. Puchinin S., Stonyakin F. Gradient-Type Method for Optimization Problems with Polyak-Lojasiewicz Condition: Relative Inexactness in Gradient and Adaptive Parameters Setting // arXiv preprint arXiv:2307.14101; 2023.
83. Convex optimization: Algorithms and complexity / S. Bubeck [et al.] // Foundat. and Trend. in Mach. Learn. 2015. V. 8. № 3/4. P. 231—357.
84. Cox B., Juditsky A., Nemirovski A. Decomposition techniques for bilinear saddle point problems and variational inequalities with affine monotone operators // J. Optimizat. Theory and Appl. 2017. V. 172. P. 402—435.
85. Гасников А., Гасникова Е. Модели равновесного распределения транспортных потоков в больших сетях. 2023.
86. Efficient numerical methods to solve sparse linear equations with application to pagerank / A. Anikin [и др.] // Optimizat. Method. and Software. 2022. V. 37. № 3. P. 907—935.
87. Bomze I. M., Rinaldi F., Zeffiro D. Frank—Wolfe and friends: a journey into projection-free first-order optimization methods // 4OR. 2021. V. 19. P. 313—345.
88. Conditional gradient methods / G. Braun [и др.] // arXiv preprint arXiv:2211.14103; 2022.
89. Zero-Order Stochastic Conditional Gradient Sliding Method for Non-smooth Convex Optimization / A. Lobanov [и др.] // arXiv preprint arXiv:2303.02778; 2023.
90. Vedernikov R., Rogozin A., Gasnikov A. Decentralized conditional gradient method over time-varying graphs // arXiv preprint arXiv:2307.10978; 2023.
91. Adaptive Variant of the Frank-Wolfe Algorithm for Convex Optimization Problems / G. Aivazian [et al.] // arXiv preprint arXiv:2307.16059; 2023.
92. Vial J.-P. Strong convexity of sets and functions // J. of Math. Economic. 1982. V. 9. № 1/2. P. 187—205.
93. Vial J.-P. Strong and weak convexity of sets and functions // Math. Operat. Res. 1983. V. 8. № 2. P. 231—259.
94. Половинкин Е.С. Сильно выпуклый анализ // Матем. сб. 1996. Т. 187. № 2. С. 103—130.
95. Ito M., Lu Z., He C. A Parameter-Free Conditional Gradient Method for Composite Minimization under  $H^*$ older Condition // J. Machine Learn. Res. 2023. V. 24. P. 1—34.
96. Taylor A.B., Hendrickx J. M., Glineur F. Smooth strongly convex interpolation and exact worst-case performance of first-order methods // Math. Program. 2017. V. 161. P. 307—345.
97. Super-acceleration with cyclical step-sizes / B. Goujaud [et al.] // Inter. Conf. on Artificial Intelligence and Statist. PMLR. 2022. P. 3028—3065.
98. Немировский А.С. О регуляризирующих свойствах метода сопряженных градиентов на некорректных задачах // Ж. вычисл. матем. и матем. физ. 1986. Т. 26. № 3. С. 332—347.
99. Acceleration methods / A. d’Aspremont, D. Scieur, A. Taylor [et al.] // Foundat. and Trends in Optimizat. 2021. V. 5. № 1/2. P. 1—245.
100. Scieur D., Pedregosa F. Universal average-case optimality of Polyak momentum // Inter. Conf. on Machine Learn. PMLR. 2020. P. 8565—8572.
101. Гельфанд И.М., Цетлин М.Л. Принцип нелокального поиска в системах автоматической оптимизации // Докл. АН СССР. 1961. Т. 137. С. 295—298.
102. Lessard L., Recht B., Packard A. Analysis and design of optimization algorithms via integral quadratic constraints // SIAM J. on Optimizat. 2016. V. 26. № 1. P. 57—95.
103. Ghadimi E., Feysmahdavian H. R., Johansson M. Global convergence of the heavy-ball method for convex optimization // 2015 European control conference (ECC). IEEE. 2015. P. 310—315.
104. Goujaud B., Taylor A., Dieuleveut A. Provable non-accelerations of the heavy-ball method // arXiv preprint arXiv:2307.11291; 2023.
105. O’Donoghue B., Candes E. Adaptive restart for accelerated gradient schemes // Foundat. Comput. Math. 2015. V. 15. P. 715—732.

106. Danilova M., Kulakova A., Polyak B. Non-monotone behavior of the heavy ball method // *Difference Equations and Discrete Dynamical Systems with Applications: 24th ICDEA, Dresden, Germany, May 21–25, 2018*. Springer, 2020. P. 213–230.
107. Немировский А. Орт-метод гладкой выпуклой минимизации // *Изв. АН СССР. Техн. кибернетика*. 1982. № 2. P. 18–29.
108. Is local SGD better than minibatch SGD? / B. Woodworth [et al.] // *Inter. Conf. Machine Learn. PMLR*. 2020. P. 10334–10343.
109. The min-max complexity of distributed stochastic convex optimization with intermittent communication / B. E. Woodworth [et al.] // *Conf. Learn. Theory. PMLR*. 2021. P. 4386–4437.
110. Нестеров Ю. Е. Метод решения задачи выпуклого программирования со скоростью сходимости  $O(1/k^2)$  // *Докл. АН СССР*. 1983. Т. 269. С. 543–547.
111. Lan G. *First-order and stochastic optimization methods for machine learning*. Vol. 1. Springer, 2020.
112. Lin Z., Li H., Fang C. *Accelerated optimization for machine learning* // Nature Singapore: Springer, 2020.
113. Peng W., Wang T. The Nesterov-Spokoiny Acceleration:  $o(1/k^2)$  Convergence without Proximal Operations // *arXiv preprint arXiv:2308.14314*; 2023.
114. Inexact model: A framework for optimization and variational inequalities / F. Stonyakin [et al.] // *Optimizat. Meth. and Software*. 2021. V. 36. № 6. P. 1155–1201.
115. Zhang Z., Lan G. Solving Convex Smooth Function Constrained Optimization Is As Almost Easy As Unconstrained Optimization // *arXiv preprint arXiv:2210.05807*; 2022.
116. Accelerated gradient methods with absolute and relative noise in the gradient / A. Vasin [et al.] // *Optimizat. Method. and Software*. 2023. P. 1–50.
117. Intermediate Gradient Methods with Relative Inexactness / N. Kornilov [et al.] // *arXiv preprint arXiv:2310.00506*; 2023.
118. Optimal gradient sliding and its application to optimal distributed optimization under similarity / D. Kovalev [et al.] // *Adv. in Neural Informat. Proces. System*. 2022. V. 35. P. 33494–33507.
119. Kovalev D., Gasnikov A., Malinovsky G. An Optimal Algorithm for Strongly Convex Min-min Optimization // *arXiv preprint arXiv:2212.14439*; 2022.
120. Optimal Algorithm with Complexity Separation for Strongly Convex-Strongly Concave Composite Saddle Point Problems / E. Borodich [et al.] // *arXiv preprint arXiv:2307.12946*; 2023.
121. Smooth monotone stochastic variational inequalities and saddle point problems: A survey / A. Beznosikov [et al.] // *Europ. Math. Soc. Magazine*. 2023. № 127. P. 15–28.
122. Nesterov Y. Implementable tensor methods in unconstrained convex optimization // *Math. Program*. 2021. V. 186. P. 157–183.
123. Monteiro R.D., Svaiter B. F. An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods // *SIAM J. Optimizat.* 2013. V. 23. № 2. P. 1092–1125.
124. Near optimal methods for minimizing convex functions with Lipschitz  $p$ -th derivatives / A. Gasnikov [et al.] // *Conf. Learn. Theory. PMLR*. 2019. P. 1392–1393.
125. Kovalev D., Gasnikov A. The first optimal acceleration of high-order methods in smooth convex optimization // *Adv. Neural Inform. Proces. System*. 2022. V. 35. P. 35339–35351.
126. Optimal and Adaptive Monteiro-Svaiter Acceleration / Y. Carmon [et al.] // *Adv. n Neural Inform. Proces. System*. V. 35 / Ed. S. Koyejo [et al.]. Curran Associates, Inc., 2022. P. 20338–20350. URL: [https://proceedings.neurips.cc/paper\\_files/paper/2022/file/7ff97417474268e6b5a38bcbfae04944-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2022/file/7ff97417474268e6b5a38bcbfae04944-Paper-Conference.pdf).
127. Exploiting higher-order derivatives in convex optimization methods / D. Kamzolov [et al.] // *arXiv preprint arXiv:2208.13190*; 2022.
128. Bertsekas D., Tsitsiklis J. *Parallel and distributed computation: numerical methods*. Athena Scientific, 2015.
129. Recent theoretical advances in decentralized distributed convex optimization / E. Gorbunov [et al.] // *High-Dimensional Optimization and Probability: With a View Towards Data Science*. Springer, 2022. P. 253–325.
130. Кибардин В. М. Декомпозиция по функциям в задаче минимизации // *Автоматика и телемех.* 1979. № 9. С. 66–79.
131. Decentralized optimization over time-varying graphs: a survey / A. Rogozin [et al.] // *arXiv preprint arXiv:2210.09719*; 2022.
132. Decentralized Optimization Over Slowly Time-Varying Graphs: Algorithms and Lower Bounds / D. Metevlev [et al.] // *arXiv preprint arXiv:2307.12562*; 2023.

133. Bao C., Chen L., Li J. The Global R-linear Convergence of Nesterov's Accelerated Gradient Method with Unknown Strongly Convex Parameter // arXiv preprint arXiv:2308.14080; 2023.
134. Guminov S., Gasnikov A., Kuruzov I. Accelerated methods for weakly-quasi-convex optimization problems // Comput. Management Sci. 2023. V. 20. № 1. P. 1—19.
135. Algorithmic stochastic convex optimization / A. Beznosikov [и др.]. Springer, 2024.
136. Robbins H., Monro S. A stochastic approximation method // Ann. Math. Statistic. 1951. P. 400—407.
137. Ермолев Ю. Методы стохастического программирования. М.: Наука, 1976.
138. High-probability bounds for stochastic optimization and variational inequalities: the case of unbounded variance / A. Sadiev [et al.] // ICML. 2023.
139. Root-sgd: Sharp nonasymptotics and asymptotic efficiency in a single algorithm / C. J. Li [et al.] // Conf. Learn. Theory. PMLR. 2022. P. 909—981.
140. Невельсон М., Хасьминский Р. Стохастическая аппроксимация и рекуррентное оценивание. М.: Наука, 1972.
141. Fort G. Central limit theorems for stochastic approximation with controlled Markov chain dynamics // ESAIM: Probability and Statistic. 2015. V. 19. P. 60—80.
142. Bach F., Perchet V. Highly-smooth zero-th order online optimization // Conf. Learn. Theory. PMLR. 2016. P. 257—283.
143. Ruppert D. Efficient estimations from a slowly convergent Robbins-Monro process: тех. отч. / Cornell University Operations Research; Industrial Engineering. 1988.
144. Robust stochastic approximation approach to stochastic programming / A. Nemirovski [и др.] // SIAM J. Optimizat. 2009. V. 19. № 4. P. 1574—1609.
145. Nesterov Y. Primal-dual subgradient methods for convex problems // Math. Program. 2009. V. 120. № 1. P. 221—259.
146. Duchi J., Hazan E., Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. // J. Machine Learn. Res. 2011. V. 12. № 7.
147. Ivgi M., Hinder O., Carmon Y. DoG is SGD's Best Friend: A Parameter-Free Dynamic Step Size Schedule. 2023. arXiv: 2302.12022 [cs.LG].
148. Cutkosky A., Defazio A., Mehta H. Mechanic: A Learning Rate Tuner // arXiv preprint arXiv:2306.00144; 2023.
149. Stich S. U. Unified optimal analysis of the (stochastic) gradient method // arXiv preprint arXiv:1907.04232; 2019.
150. Gorbunov E. Unified analysis of SGD-type methods // arXiv preprint arXiv:2303.16502; 2023.
151. Lan G. An optimal method for stochastic composite optimization // Math. Program. 2012. V. 133. № 1/2. P. 365—397.
152. The power of first-order smooth optimization for black-box non-smooth problems / A. Gasnikov [et al.] // Inter. Conf. Machine Learn. PMLR. 2022. P. 7241—7265.
153. Woodworth B.E., Srebro N. An even more optimal stochastic optimization algorithm: minibatching and interpolation learning // Adv. Neural Inform. Proces. System. 2021. V. 34. P. 7333—7345.
154. Accelerated stochastic approximation with state-dependent noise / S. Pandarideva [и др.] // arXiv preprint arXiv:2307.01497; 2023.
155. Unixgrad: A universal, adaptive algorithm with optimal guarantees for constrained optimization / A. Kavis [et al.] // Adv. Neural Inform. Proces. System. 2019. V. 32.
156. Ene A., Nguyen H. L., Vladu A. Adaptive gradient methods for constrained convex optimization and variational inequalities // Proceed. of the AAAI Conf. Artificial Intelligence. 2021. V. 35. P. 7314—7321.
157. Nesterov Y. Efficiency of coordinate descent methods on huge-scale optimization problems // SIAM J. Optimizat. 2012. V. 22. № 2. P. 341—362.
158. Richtárik P., Takáč M. On optimal probabilities in stochastic coordinate descent methods // Optimizat. Lett. 2016. V. 10. P. 1233—1243.
159. Qu Z., Richtárik P. Coordinate descent with arbitrary sampling I: Algorithms and complexity // Optimizat. Meth. and Software. 2016. V. 31. № 5. P. 829—857.
160. QSGD: Communication-efficient SGD via gradient quantization and encoding / D. Alistarh [et al.] // Adv. Neural Inform. Proces. System. 2017. V. 30.
161. On biased compression for distributed learning / A. Beznosikov [et al.] // arXiv preprint arXiv:2002.12410; 2020.
162. Schmidt M., Roux N. L. Fast convergence of stochastic gradient descent under a strong growth condition // arXiv preprint arXiv:1308.6370; 2013.

163. Vaswani S., Bach F., Schmidt M. Fast and faster convergence of sgd for over-parameterized models and an accelerated perceptron // 22nd Inter. Conf. on Artificial Intelligence and Statistics. PMLR. 2019. P. 1195–1204.
164. First Order Methods with Markovian Noise: from Acceleration to Variational Inequalities / A. Beznosikov [et al.] // arXiv preprint arXiv:2305.15938; 2023.
165. Moulines E., Bach F. Non-Asymptotic Analysis of Stochastic Approximation Algorithms for Machine Learning // Adv. Neural Inform. Proces. System. V. 24 / Ed. J. Shawe-Taylor [et al.]. Curran Associates, Inc., 2011. URL: [https://proceedings.neurips.cc/paper\\_files/paper/2011/file/40008b9a5380fcacce3976bf7c08af5b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2011/file/40008b9a5380fcacce3976bf7c08af5b-Paper.pdf).
166. SGD: General analysis and improved rates / R. M. Gower [et al.] // Inter. Conf. Machine Learn. PMLR. 2019. P. 5200–5209.
167. Ma S., Bassily R., Belkin M. The power of interpolation: Understanding the effectiveness of SGD in modern over-parametrized learning // Inter. Conf. Machine Learn. PMLR. 2018. P. 3325–3334.
168. Distributed learning with compressed gradient differences / K. Mishchenko [et al.] // arXiv preprint arXiv:1901.09269; 2019.
169. Gorbunov E., Hanzely F., Richtárik P. A unified theory of SGD: Variance reduction, sampling, quantization and coordinate descent // Inter. Conf. Artificial Intelligence and Statistics. PMLR. 2020. P. 680–690.
170. Defazio A., Bach F., Lacoste-Julien S. SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives // Adv. Neural Inform. Proces. System. 2014. V. 27.
171. Johnson R., Zhang T. Accelerating stochastic gradient descent using predictive variance reduction // Adv. Neural Inform. Proces. System. 2013. V. 26.
172. Kovalev D., Horváth S., Richtárik P. Don't jump through hoops and remove those loops: SVRG and Katyusha are better without the outer loop // Algorithm. Learn. Theory. PMLR. 2020. P. 451–467.
173. Hanzely F., Mishchenko K., Richtárik P. SEGA: Variance reduction via gradient sketching // Adv. Neural Inform. Proces. System. 2018. V. 31.
174. Unified analysis of stochastic gradient methods for composite convex and smooth optimization / A. Khaled [и др.] // J. Optimizat. Theory and Appl. 2023. P. 1–42.
175. Stochastic gradient descent-ascent: Unified theory and new efficient methods / A. Beznosikov [et al.] // Inter. Conf. Artificial Intelligence and Statistics. PMLR. 2023. P. 172–235.
176. Maslovskii A.Y. A unified analysis of variational inequality methods: variance reduction, sampling, quantization, and coordinate descent // Comput. Math. and Math. Phys. 2023. V. 63. № 2. P. 147–174.
177. Explore aggressively, update conservatively: Stochastic extragradient methods with variable stepsize scaling / Y.-G. Hsieh [et al.] // Adv. Neural Inform. Proces. System. 2020. V. 33. P. 16223–16234.
178. Stochastic extragradient: General analysis and improved rates / E. Gorbunov [и др.] // Inter. Conf. Artificial Intelligence and Statistics. PMLR. 2022. P. 7865–7901.
179. Алгоритмы робастной стохастической оптимизации на основе метода зеркального спуска / А. В. Назин [и др.] // Автоматика и телемех. 2019. № 9. С. 64–90.
180. High-Probability Convergence for Composite and Distributed Stochastic Minimization and Variational Inequalities with Heavy-Tailed Noise / E. Gorbunov [и др.]. 2023. arXiv: 2310.01860 [math.OC].
181. Поляк Б. Т., Цыпкин Я. З. Псевдоградиентные алгоритмы адаптации и обучения. Автоматика и телемеханика // Автоматика и телемех. 1973. № 3. С. 45–68.
182. Nonlinear gradient mappings and stochastic optimization: A general framework with applications to heavy-tail noise / D. Jakovetić [et al.] // SIAM J. Optimizat. 2023. V. 33. № 2. P. 394–423.
183. Advancing the lower bounds: An accelerated, stochastic, second-order method with optimal adaptation to inexactness / A. Agafonov [et al.]. 2023. arXiv: 2309.01570 [math.OC].
184. Граничин О. Н., Поляк Б. Т. Рандомизированные алгоритмы оценивания и оптимизации при почти произвольных помехах. М.: Наука, 2003.
185. Rock H. An automatic method for finding the greatest or least value of a function // Comput. J. 1960. V. 3. № 3. P. 175–184.
186. Kiefer J., Wolfowitz J. Stochastic estimation of the maximum of a regression function // Ann. Math. Statistic. 1952. P. 462–466.
187. Randomized gradient-free methods in convex optimization / A. Gasnikov [et al.] // arXiv preprint arXiv:2211.13566; 2022.
188. Lobanov A., Gasnikov A., Stonyakin F. Highly Smoothness Zero-Order Methods for Solving Optimization Problems under PL Condition // Ж. вычисл. матем. и матем. физ. 2023.



189. Gradient-free optimization of highly smooth functions: improved analysis and a new algorithm / A. Akhavan [et al.] // arXiv preprint arXiv:2306.02159; 2023.
190. A theoretical and empirical comparison of gradient approximations in derivative-free optimization / A. S. Berahas [et al.] // *Foundat. Comput. Math.* 2022. V. 22. № 2. P. 507–560.
191. Akhavan A., Pontil M., Tsybakov A. Exploiting higher order smoothness in derivative-free optimization and continuous bandits // *Adv. Neural Inform. Proces. System.* 2020. V. 33. P. 9017–9027.
192. Novitskii V., Gasnikov A. Improved exploiting higher order smoothness in derivative-free optimization and continuous bandit // arXiv preprint arXiv:2101.03821; 2021.
193. Гасников А., Двуреченский П., Нестеров Ю. Стохастические градиентные методы с неточным оракулом // *Тр. Московск. физ.-тех. ин-та.* 2016. Т. 8. № 1 (29). С. 41–91.
194. Граничин О.Н., Иванский Ю.В., Копылова К.Д. Метод Б.Т. Поляка на основе стохастической функции Ляпунова для обоснования состоятельности оценок поискового алгоритма стохастической аппроксимации при неизвестных, но ограниченных помехах // *Ж. вычисл. матем. и матем. физ.* 2023.
195. Lobanov A., Bashirov N., Gasnikov A. The Black-Box Optimization Problem: Zero-Order Accelerated Stochastic Method via Kernel Approximation. 2023. arXiv: 2310.02371 [math.OC].
196. Learning supervised pagerank with gradient-based and gradient-free optimization methods / L. Bogolubsky [et al.] // *Adv. Neural Inform. Proces. System.* 2016. V. 29.
197. Noisy zeroth-order optimization for non-smooth saddle point problems / D. Dvinskikh [et al.] // *Inter. Conf. Math. Optimizat. Theory and Operat. Res.* Springer. 2022. P. 18–33.
198. Gradient-Free Federated Learning Methods with  $l_1$  and  $l_2$ -Randomization for Non-Smooth Convex Stochastic Optimization Problems / A. Lobanov [et al.] // arXiv preprint arXiv:2211.10783; 2022.
199. Accelerated Zeroth-order Method for Non-Smooth Stochastic Convex Optimization Problem with Infinite Variance / N. Kornilov [et al.]. 2022.
200. Risteski A., Li Y. Algorithms and matching lower bounds for approximately-convex optimization // *Adv. Neural Inform. Proces. System.* 2016. V. 29.

## ON SOME WORKS OF BORIS TEODOROVICH POLYAK ON THE CONVERGENCE OF GRADIENT METHODS AND THEIR DEVELOPMENT

S. S. Ablav<sup>a,d</sup>, A. N. Beznosikov<sup>a,b</sup>, A. V. Gasnikov<sup>a,b,c,\*</sup>, D. M. Dvinskikh<sup>a,b,c</sup>, A. V. Lobanov<sup>a,c</sup>, S. M. Puchinin<sup>a</sup>, F. S. Stonyakin<sup>a,d</sup>

<sup>a</sup>*Moscow Institute of Physics and Technology, Institutskiy per., 9, Dolgoprudny, Moscow oblast, 141700 Russia*

<sup>b</sup>*Institute for Information Transmission Problems, Russian Academy of Sciences, B. Karetny per., 19, build. 1, Moscow, 127051 Russia*

<sup>c</sup>*Institute for System Programming, Russian Academy of Sciences, Solzhenitsyn str., 25, Moscow, 109004 Russia*

<sup>d</sup>*Crimean Federal University, Ac. Vernadskiy Ave., Simferopol, Republic of Crimea, 295007 Russia*

\* e-mail: gasnikov@yandex.ru

Received 15 September, 2023

Revised 06 November, 2023

Accepted 17 November, 2023

**Abstract.** The paper presents a review of the current state of subgradient and accelerated convex optimization methods, including the cases with the presence of noise and access to various information about the objective function (function value, gradient, stochastic gradient, higher derivatives). For nonconvex problems, the Polyak–Lojasiewicz condition is considered and a review of the main results is given. The behavior of numerical methods in the presence of a sharp minimum is considered. The aim of this review is to show the influence of the works of B.T. Polyak (1935–2023) on gradient optimization methods and their surroundings on the modern development of numerical optimization methods.

**Keywords:** gradient descent, gradient dominance condition (Polyak–Lojasiewicz), sharp minimum, subgradient Polyak–Shor method, early stopping condition, Polyak heavy ball method, stochastic gradient descent.

# МЕТОД Б.Т. ПОЛЯКА НА ОСНОВЕ СТОХАСТИЧЕСКОЙ ФУНКЦИИ ЛЯПУНОВА ДЛЯ ОБОСНОВАНИЯ СОСТОЯТЕЛЬНОСТИ ОЦЕНОК ПОИСКОВОГО АЛГОРИТМА СТОХАСТИЧЕСКОЙ АППРОКСИМАЦИИ ПРИ НЕИЗВЕСТНЫХ, НО ОГРАНИЧЕННЫХ ПОМЕХАХ<sup>1)</sup>

© 2024 г. О. Н. Граничин<sup>1,2,\*</sup>, Ю. В. Иванский<sup>1,2</sup>, К. Д. Копылова<sup>1</sup>

<sup>1</sup> 199034 Санкт-Петербург, Университетская наб., 7–9, СПбГУ, Россия

<sup>2</sup> Санкт-Петербург, ВО, Большой пр., 61, ИПМаш РАН, Россия

\* e-mail: oleg\_granichin@mail.ru

Поступила в редакцию 15.09.2023 г.

Переработанный вариант 06.11.2023 г.

Принята к публикации 20.11.2023 г.

В 1976–1977 гг. Б.Т. Поляк опубликовал в журнале «Автоматика и телемеханика» две замечательные статьи о том, как исследовать свойства оценок итеративных псевдоградиентных алгоритмов. В первой статье 1976 г. рассматривался общий случай на основе стохастической функции Ляпунова, во второй — линейный случай. Сформулированные предположения и полученные в статьях оценки до сих пор можно считать результатами уровня «state of the art». В настоящей статье этот ставший классическим подход Б.Т. Поляка применяется к исследованию свойств оценок поискового (рандомизированного) алгоритма стохастической аппроксимации для случая неизвестных, но ограниченных помех в наблюдениях. Полученные асимптотические оценки были известны уже и ранее, точные оценки для конечного числа наблюдений публикуются впервые. Библ. 18.

**Ключевые слова:** поисковый алгоритм стохастической аппроксимации, «неизвестные, но ограниченные помехи», аппроксимация градиента, сглаживающие ядра, безградиентные методы, методы с неточным оракулом. Библ.18.

DOI: 10.31857/S0044466924040034, EDN: ZKKRRX

## 1. ВВЕДЕНИЕ

Решение задач оптимизации часто сводится к нахождению точки экстремума некоторой функции. При этом принято полагаться на использование известных итерационных процедур, использующих значения исследуемой функции в выбираемых последовательно точках. Наличие помех при измерениях оказывает существенное влияние на результат. Для практики важно предложить такую итерационную схему, которая как можно более быстрая (по количеству итераций), простая (по сложности, требуемым ресурсам и времени выполнения одной итерации) и работоспособная при почти произвольных помехах. Именно такого типа процедуры были предложены в [1]–[3], основная идея которых состоит в решении сложных многомерных (в пространстве  $\mathbb{R}^d$ ) задач оптимизации методами, похожими на процедуру Кифера–Вольфовица, но использующими на каждой итерации для формирования новой градиентной аппроксимации вместо 2d-измерений целевой функции только одно или два измерения в точках на случайно выбираемой линии, проходящей через точку предыдущей оценки. В статье 1989 г. [1] была предложена новая стохастическая рекуррентная процедура с одним измерением на итерации с пробными возмущениями на входе, для которой была обоснована состоятельность оценок при почти произвольных зависимых помехах в наблюдениях. Позже этот результат был распространен на векторный случай (см. [4]–[6]). В 1990 г. в статье Б. Т. Поляка и А. Б. Цыбакова [2] такого типа алгоритмы получили название поисковые алгоритмы стохастической аппроксимации, и для них при достаточно общих условиях была доказана асимптотическая оптимальность в том смысле, что для класса близких функций невозможно предложить итеративный алгоритм, который будет асимптотически сходиться быстрее. Похожий алгоритм в англоязычной литературе под названием метод SPSA (Simultaneously Perturbed Stochastic Approximation) в варианте с двумя измерениями был предложен в 1992 г. в статье [3] и с одним измерением — в [7]. Позднее в 1999 г. в статье [8] был обоснован новый алгоритм с двумя измерениями, в одном из которых

<sup>1)</sup> Работа выполнена в ИПМаш РАН при финансовой поддержке РФФИ (грант № 21-19-00516).

измерение псевдоградиента проводилось в возмущенной точке, а в другом — в точке предшествующей оценки без возмущения. Такая версия лучше подходит для оптимизации в реальном времени, когда данные поступают в обновляющемся потоке. Основная идея метода состоит в решении сложных многомерных (в пространстве  $\mathbb{R}^d$ ) задач методами, похожими на процедуру Кифера–Вольфовица, но используя на каждой итерации для формирования одной градиентной аппроксимации вместо 2d-измерений целевой функции только одно или два измерения в точках на случайно выбираемой линии, проходящей через точку предыдущей оценки.

В последнее время подобного рода алгоритмы привлекают внимание исследователей (см., например, [9], [10]). Их часто стали называть безградиентными методами с неточным оракулом. В 2023 г. вышла статья [11], развивающая идеи из [5]. Алгоритм, рассматриваемый в этой статье, был первоначально предложен в [12] для нахождения центров восходящих термических потоков, используемых для увеличения длительности полетов БПЛА планерного типа. При исследовании скорости сходимости оценок ключевую роль играет способ формирования размеров шагов алгоритмов. В статье [13] этот алгоритм с постоянными размерами шага использовался для решения задачи трекинга. На практике с течением времени вид исследуемой функции может несколько видоизменяться, и точки экстремумов могут дрейфовать со временем. В зависимости от скорости дрейфа и уровня помех в измерениях выбор постоянного большего размера шага алгоритма на итерации может давать более высокую скорость сходимости, но только в некоторую окрестность текущей точки экстремума, при малом постоянном размере шага оценки сходятся медленнее, но результаты точнее при отсутствии дрейфа. Важно, чтобы за время сходимости искомая точка экстремума не отдрейфовала далеко. Если нет возможности сойтись точно, есть возможность сойтись быстрее. Например, в задачах слежения за маневрирующими целями не так важно, чтобы оценки точно сошлись к искомому решению, как важно успеть отслеживать дрейф точки минимума (точность отходит на второй план, важнее становится скорость сходимости). Для задач распределенной оптимизации (в частности, распределенного отслеживания целей) в [14] была предложена модификация алгоритма, совмещенная с консенсусным протоколом локального голосования. В [15] показаны возможности ускорения сходимости по Нестерову, основанные на использовании метода тяжелого шарика Б. Т. Поляка (см. [16]).

Детальный вклад известных работ Б. Т. Поляка в развитие численных методов оптимизации анализируется в [17].

Далее в настоящей статье свойства оценок алгоритма из [13] анализируются на основе использования стохастической функции Ляпунова, опираясь на результаты статьи [18], в которой Б. Т. Поляк предложил и обосновал метод на основе стохастической функции Ляпунова для исследования свойств оценок псевдоградиентных итеративных алгоритмов стохастической оптимизации (см. также [6], гл. 5).

## 2. ПОСТАНОВКА ЗАДАЧИ, АЛГОРИТМ И ОСНОВНЫЕ ПРЕДПОЛОЖЕНИЯ

Пусть  $x_1, x_2, \dots$  — последовательность точек измерения (план наблюдения, выбираемый или контролируемый экспериментатором),  $x_n \in \mathbb{R}^d$ , в которых в каждый момент времени  $n = 1, 2, \dots$  доступно наблюдению с помехами  $v_n$  значение некоторой дифференцируемой по  $x_n$  штрафной функции (функции потерь)  $F(x_n, w_n)$ :

$$y_n = F(x_n, w_n) + v_n, \quad (1)$$

где  $F: \mathbb{R}^d \times \mathbb{R}^p \rightarrow \mathbb{R}^1$ ,  $\{w_n\}$  — неконтролируемая последовательность  $p$ -мерных случайных векторов  $w_n$  (возмущений) с одинаковым (может быть) неизвестным распределением  $P_w(\cdot)$  с компактным носителем  $\mathbb{W} = \text{supp}(P_w(\cdot)) \subset \mathbb{R}^p$ .

Рассмотрим задачу минимизации функции

$$f(x) := \mathbb{E}_w [F(x, w)] = \int_{\mathbb{R}^p} F(x, w) P_w(dw). \quad (2)$$

Здесь и далее  $\mathbb{E}$  — символ математического ожидания.

**Задача:** по доступным наблюдениям необходимо построить последовательность оценок  $\hat{\theta}_n$  неизвестного вектора  $\theta$ , минимизирующей функцию  $f(x)$ .

Для решения поставленной задачи воспользуемся итеративным алгоритмом из [13] с двумя измерениями.

**Алгоритм:** выберем произвольный вектор начальной оценки  $\hat{\theta}_0 \in \mathbb{R}^d$ , и для выбираемых или задаваемых последовательностей положительных чисел  $\{\alpha_n\}$ ,  $\{\beta_n^+\}$  и  $\{\beta_n^-\}$  с помощью пробного одновременного возмущения  $\{\Delta_n\}$  определим шаг итерации:

$$\begin{aligned} x_{2n} &= \hat{\theta}_{2n-2} + \beta_n^+ \Delta_n, \quad x_{2n-1} = \hat{\theta}_{2n-2} - \beta_n^- \Delta_n, \\ \hat{\theta}_{2n-1} &= \hat{\theta}_{2n-2}, \\ y_{2n} &= F(x_{2n}, w_{2n}) + v_{2n}, \quad y_{2n-1} = F(x_{2n-1}, w_{2n-1}) + v_{2n-1}, \\ \hat{\theta}_{2n} &= \hat{\theta}_{2n-1} - \frac{\alpha_n}{\beta_n^+ + \beta_n^-} \mathcal{K}_n(\Delta_n)(y_{2n} - y_{2n-1}), \end{aligned} \quad (3)$$

где  $\Delta_n, n = 1, 2, \dots$ , — наблюдаемая (задаваемая или контролируемая пользователем) последовательность независимых случайных векторов из  $\mathbb{R}^d$  с известными функциями распределения  $P_n(\cdot), K_n(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$  — задаваемые пользователем вектор-функции (ядра), удовлетворяющие вместе с  $P_n(\cdot)$  условиям

$$\int \mathcal{K}_n(x) P_n(dx) = 0, \quad \int \mathcal{K}_n(x) x^T P_n(dx) = I, \quad (4)$$

$$\sup_n \int \|\mathcal{K}_n(x)\|^2 P_n(dx) < \infty, \quad n = 1, 2, \dots \quad (5)$$

Следующие предположения будут использоваться для формулировки основного результата.

**Предположение 1.** Функция  $f(\cdot)$  — сильно выпуклая, т.е. имеет единственный минимум в  $\mathbb{R}^d$  в некоторой точке  $\theta = \theta(f(\cdot))$  и

$$\langle x - \theta, \nabla f(x) \rangle \geq \mu \|x - \theta\|^2 \quad \forall x \in \mathbb{R}^d$$

с некоторой постоянной  $\mu > 0$ . Здесь и далее  $\langle x, y \rangle := \sum_{i=1}^d x_i y_i$  — скалярное произведение  $x, y \in \mathbb{R}^d$ .

**Предположение 2.** Условие Липшица на градиент функции  $F(\cdot, w)$

$$\|\nabla F(x_1, w) - \nabla F(x_2, w)\| \leq A \|x_1 - x_2\|,$$

с некоторой постоянной  $A > \mu \quad \forall w \in \mathbb{R}^p, \forall x_1, x_2 \in \mathbb{R}^d$ .

**Предположение 3.**  $\forall x \in \mathbb{R}^d$  функции  $F(x, \cdot)$  и  $\nabla_x F(x, \cdot)$  равномерно на  $\mathbb{W}$  ограничены.

Пусть

$$\bar{v}_n = v_{2n} - v_{2n-1}, \quad \bar{w}_n = \begin{pmatrix} w_{2n} \\ w_{2n-1} \end{pmatrix}.$$

**Предположение 4.** О равномерной ограниченности помех наблюдения  $\bar{v}_n$  :

$\forall n \geq 1$ , если  $\bar{v}_n$  случайно, то  $\mathbb{E}\{\bar{v}_n^2\} \leq \sigma_v^2$ , иначе  $\bar{v}_n$  ограничены:  $|\bar{v}_n| \leq \sigma_v$ .

**Предположение 5.** Условия о взаимной независимости помех  $\bar{v}_n$ , возмущений  $\bar{w}_n$  и пробных возмущений  $\Delta_n$  :

$\forall n \geq 1$  случайный вектор  $\Delta_n$  не зависит от  $\bar{w}_n$ , случайные векторы  $\bar{w}_n, \Delta_n$  не зависят от  $\bar{w}_1, \dots, \bar{w}_{n-1}$ ; если  $\{\bar{v}_n\}$  — случайные величины, то  $\bar{w}_n, \Delta_n$  также не зависят от  $\bar{v}_1, \dots, \bar{v}_n$ .

### 3. АНАЛИЗ СХОДИМОСТИ ПОСЛЕДОВАТЕЛЬНОСТИ ОЦЕНОК

Для анализа сходимости оценок алгоритма (3) используется метод из [18], аналогичный второму методу Ляпунова в теории устойчивости. В качестве функции Ляпунова выберем

$$V(\hat{\theta}_{2n-2}) = \frac{1}{2} \|\hat{\theta}_{2n-2} - \theta\|^2. \quad (6)$$

Перечислим основные условия теорем 1—3 из [18], выполнение которых надо проверить.

**Условие А.** Итеративный процесс  $Y_n$ , задающий направление изменения оценки, имеет марковский характер, т.е. распределение случайного вектора  $Y_n$  зависит только от  $\hat{\theta}_{2n-2}$  и  $n$ :

$$Y_n = G_n(\hat{\theta}_{2n-2}, \bar{w}_n).$$

Для алгоритма (3) имеем

$$\mathbb{E}\{Y_n\} = \mathbb{E}\left\{\mathcal{K}_n(\Delta_n) \frac{y_{2n} - y_{2n-1}}{\beta_n^+ + \beta_n^-} \mid \mathcal{F}_{n-1}\right\}.$$

**Условие Б.**  $V(\hat{\theta}_{2n-2})$  — неотрицательна,  $\inf V(\hat{\theta}_{2n-2}) = 0$ ,  $V(\hat{\theta}_{2n-2})$  — дифференцируема, а ее градиент удовлетворяет условию Липшица:

$$\|\nabla V(x) - \nabla V(\theta)\| \leq L\|x - \theta\|.$$

**Условие В.** Условие псевдоградиентности:

$$\langle \nabla V(\hat{\theta}_{2n-2}), \mathbb{E}\{G_n(\hat{\theta}_{2n-2}, \bar{w}_n)\} \rangle \geq \delta_n V(\hat{\theta}_{2n-2}) - \gamma_n, \quad \delta_n > 0, \quad \gamma_n \geq 0.$$

Для алгоритма (3) с учетом выбранной в виде (6) функции Ляпунова получается следующее выражение:

$$\langle \nabla V(\hat{\theta}_{2n-2}), \mathbb{E}\{G_n(\hat{\theta}_{2n-2}, \bar{w}_n)\} \rangle = \left\langle \hat{\theta}_{2n-2} - \theta, \mathbb{E}\{K_n(\Delta_n) \frac{y_{2n} - y_{2n-1}}{\beta_n^+ + \beta_n^-} \mid \mathcal{F}_{n-1}\} \right\rangle.$$

Учитывая вид  $y_{2n}, y_{2n-1}$  и  $x_{2n}, x_{2n-1}$ , перепишем  $\mathbb{E}\{G_n(\hat{\theta}_{2n-2}, \bar{w}_n)\}$  в виде

$$\mathbb{E}\{G_n(\hat{\theta}_{2n-2}, \bar{w}_n)\} = \mathbb{E}\{K_n(\Delta_n) \frac{F(\hat{\theta}_{2n-2} + \beta_n^+ \Delta_n, w_{2n}) + v_{2n} - F(\hat{\theta}_{2n-2} - \beta_n^- \Delta_n, w_{2n-1}) - v_{2n-1}}{\beta_n^+ + \beta_n^-} \mid \mathcal{F}_{n-1}\}.$$

Так как в силу предположения 5 вектор  $\Delta_n$  не зависит от  $v_n$ , то

$$\mathbb{E}\{K_n(\Delta_n)(v_{2n} - v_{2n-1}) \mid \mathcal{F}_{n-1}\} = \int K_n(x) P_n(dx) \mathbb{E}\{(v_{2n} - v_{2n-1}) \mid \mathcal{F}_{n-1}\}.$$

Далее из (4) следует, что первый сомножитель в последней формуле равен нулю, а значит, и

$$\mathbb{E}\{K_n(\Delta_n)(v_{2n} - v_{2n-1}) \mid \mathcal{F}_{n-1}\} = 0.$$

Следовательно,

$$\mathbb{E}\{G_n(\hat{\theta}_{2n-2}, \bar{w}_n)\} = \frac{1}{\beta_n^+ + \beta_n^-} \mathbb{E}\{K_n(\Delta_n) F(\hat{\theta}_{2n-2} + \beta_n^+ \Delta_n, w_{2n}) - F(\hat{\theta}_{2n-2} - \beta_n^- \Delta_n, w_{2n-1}) \mid \mathcal{F}_{n-1}\}. \quad (7)$$

Воспользовавшись дважды формулой Тейлора, перепишем выражение под знаком математического ожидания:

$$\begin{aligned} & \frac{\beta_n^+}{\beta_n^+ + \beta_n^-} \nabla f(\hat{\theta}_{2n-2}) + \frac{\beta_n^+}{\beta_n^+ + \beta_n^-} \iint K_n(x) x^T \int_0^1 (\nabla_x F(\hat{\theta}_{2n-2} + t\beta_n^+ x, w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n})) dt P_n(dx) P_w(dw) + \\ & + \frac{\beta_n^-}{\beta_n^+ + \beta_n^-} \nabla f(\hat{\theta}_{2n-2}) - \frac{\beta_n^-}{\beta_n^+ + \beta_n^-} \iint K_n(x) x^T \int_0^1 \begin{pmatrix} \nabla_x F(\hat{\theta}_{2n-2} - t\beta_n^- x, w_{2n-1}) - \\ - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1}) \end{pmatrix} dt P_n(dx) P_w(dw). \end{aligned}$$

Оценим два слагаемых с интегралами по модулю:

$$\begin{aligned} & \left| \frac{\beta_n^+}{\beta_n^+ + \beta_n^-} \iint K_n(x) x^T \int_0^1 (\nabla_x F(\hat{\theta}_{2n-2} + t\beta_n^+ x, w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n})) dt P_n(dx) P_w(dw) + \right. \\ & \left. + \frac{\beta_n^-}{\beta_n^+ + \beta_n^-} \iint K_n(x) x^T \int_0^1 (\nabla_x F(\hat{\theta}_{2n-2} + t\beta_n^- x, w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})) dt P_n(dx) P_w(dw) \right| \leq \\ & \leq \left| \frac{\beta_n^+}{\beta_n^+ + \beta_n^-} \iint K_n(x) x^T \int_0^1 (\nabla_x F(\hat{\theta}_{2n-2} + t\beta_n^+ x, w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n})) dt P_n(dx) P_w(dw) \right| + \quad (8) \\ & + \left| \frac{\beta_n^-}{\beta_n^+ + \beta_n^-} \iint K_n(x) x^T \int_0^1 (\nabla_x F(\hat{\theta}_{2n-2} - t\beta_n^- x, w_{2n-1}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})) dt P_n(dx) P_w(dw) \right| \leq \\ & \leq \frac{(\beta_n^+)^2 + (\beta_n^-)^2}{\beta_n^+ + \beta_n^-} \iint \|K_n(x)\| \|x\| A \|x\| P_n(dx) P_w(dw) \leq C_1 \frac{(\beta_n^+)^2 + (\beta_n^-)^2}{\beta_n^+ + \beta_n^-}. \end{aligned}$$

Подставляем получившееся в (8) выражение в (7):

$$\left\langle \hat{\theta}_{2n-2} - \theta, \mathbb{E}\left\{\mathcal{K}_n(\Delta_n) \frac{y_{2n} - y_{2n-1}}{\beta_n^+ + \beta_n^-} \mid \mathcal{F}_{n-1}\right\}\right\rangle \geq \left\langle \hat{\theta}_{2n-2} - \theta, \nabla f(\hat{\theta}_{2n-2}) \right\rangle - C_1 \frac{(\beta_n^+)^2 + (\beta_n^-)^2}{\beta_n^+ + \beta_n^-} \|\hat{\theta}_{2n-2} - \theta\|,$$

Так как

$$2C_1 \frac{(\beta_n^+)^2 + (\beta_n^-)^2}{\beta_n^+ + \beta_n^-} \|\hat{\theta}_{2n-2} - \theta\| \leq \left( C_1 \frac{(\beta_n^+)^2 + (\beta_n^-)^2}{\beta_n^+ + \beta_n^-} \right)^2 + \|\hat{\theta}_{2n-2} - \theta\|^2,$$

то

$$\begin{aligned} \left\langle \nabla V(\hat{\theta}_{2n-2}), \mathbb{E}\left\{G_n(\hat{\theta}_{2n-2}, \bar{w}_n)\right\}\right\rangle &\geq \left\langle \hat{\theta}_{2n-2} - \theta, \nabla f(\hat{\theta}_{2n-2}) \right\rangle - \frac{1}{2} \left( C_1 \frac{(\beta_n^+)^2 + (\beta_n^-)^2}{\beta_n^+ + \beta_n^-} \right)^2 - \frac{1}{2} \|\hat{\theta}_{2n-2} - \theta\|^2 \geq \\ &\geq (2\mu - 1) \frac{1}{2} \|\hat{\theta}_{2n-2} - \theta\|^2 - \frac{1}{2} \left( C_1 \frac{(\beta_n^+)^2 + (\beta_n^-)^2}{\beta_n^+ + \beta_n^-} \right)^2. \end{aligned} \quad (9)$$

**Условие Г.** Условие на  $G_n(\cdot, \cdot)$ :

$$\mathbb{E}\left\{\|G_n(\bar{w}_n, x)\|^2\right\} \leq \sigma_n^2 + \tau_n V(x), \quad \sigma_n \geq 0, \tau_n \geq 0.$$

Учитывая ограниченность вектор-функций  $\mathcal{K}_n(\cdot)$  и компактность их носителя, последовательно выводим

$$\mathbb{E}\left\{\frac{(y_{2n} - y_{2n-1})^2}{4\beta_n^+ + \beta_n^-} \|\mathcal{K}_n(\Delta_n)\|^2 \mid \mathcal{F}_{n-1}\right\} \leq \frac{1}{2} \sup_x \mathcal{K}_n(x)^2 \mathbb{E}\{v_{2n-1}^2 + v_{2n}^2 \mid \mathcal{F}_{n-1}\} + \iiint \left( \begin{array}{l} F(\hat{\theta}_{2n-2} + \beta_n^+ x, w_{2n}) - \\ - F(\hat{\theta}_{2n-2} - \beta_n^- x, w_{2n-1}) \end{array} \right)^2 \times \quad (10)$$

$$\times \|\mathcal{K}_n(x)\|^2 P_n(dx) P_w(dw_{2n}) P_w(dw_{2n-1}) \leq C_7 \beta_n^2 \left( \|\hat{\theta}_{2n-2} - \theta\|^2 + \beta_n^2 \right) + C_8 \xi_n^2 = C_7 \beta_n^2 \|\hat{\theta}_{2n-2} - \theta\|^2 + C_8 \xi_n^2 + o(\beta_n^2),$$

использовав ограниченность вектор-функций  $\mathcal{K}_n(\cdot)$ , компактность их носителя и неравенство

$$\begin{aligned} \left| F(\hat{\theta}_{2n-2} + \beta_n^+ x, w_{2n}) - F(\hat{\theta}_{2n-2} - \beta_n^- x, w_{2n-1}) \right| &\leq \left| F(\hat{\theta}_{2n-2} + \beta_n^+ x, w_{2n}) - F(\hat{\theta}_{2n-2}, w_{2n}) \right| + \\ &+ \left| F(\hat{\theta}_{2n-2} - \beta_n^- x, w_{2n-1}) - F(\hat{\theta}_{2n-2}, w_{2n-1}) \right| + \left| F(\hat{\theta}_{2n-2}, w_{2n}) - F(\hat{\theta}_{2n-2}, w_{2n-1}) \right|, \end{aligned}$$

где  $C_i, i = 1, 2, \dots$ , — некоторые положительные константы.

Для разности применим дважды формулу Тейлора:

$$\begin{aligned} F(\hat{\theta}_{2n-2} + \beta_n^+ x, w_{2n}) - F(\hat{\theta}_{2n-2} - \beta_n^- x, w_{2n-1}) &= F(\hat{\theta}_{2n-2}, w_{2n}) + \left\langle \nabla_x F(\hat{\theta}_{2n-2} + t\beta_n^+ x, w_{2n}), \beta_n^+ x \right\rangle - \\ &- F(\hat{\theta}_{2n-2}, w_{2n-1}) - \left\langle \nabla_x F(\hat{\theta}_{2n-2} - t\beta_n^- x, w_{2n-1}), \beta_n^- x \right\rangle. \end{aligned} \quad (11)$$

Добавив и отняв  $\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})$  и  $\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})$ , получаем

$$\begin{aligned} &= F(\hat{\theta}_{2n-2}, w_{2n}) - F(\hat{\theta}_{2n-2}, w_{2n-1}) + \left\langle \nabla_x F(\hat{\theta}_{2n-2} + t\beta_n^+ x, w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n}), \beta_n^+ x \right\rangle - \\ &- \left\langle \nabla_x F(\hat{\theta}_{2n-2} - t\beta_n^- x, w_{2n-1}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1}), \beta_n^- x \right\rangle + \\ &+ \left\langle \nabla_x F(\hat{\theta}_{2n-2}, w_{2n}), \beta_n^+ x \right\rangle - \left\langle \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1}), \beta_n^- x \right\rangle. \end{aligned} \quad (12)$$

Оценим скалярные произведения

$$\begin{aligned} \left\langle \nabla_x F(\hat{\theta}_{2n-2} + t\beta_n^+ x, w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n}), \beta_n^+ x \right\rangle &\leq \left\| \nabla_x F(\hat{\theta}_{2n-2} + t\beta_n^+ x, w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n}) \right\| \|\beta_n^+ x\|, \\ \left\langle \nabla_x F(\hat{\theta}_{2n-2} - t\beta_n^- x, w_{2n-1}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1}), \beta_n^- x \right\rangle &\leq \left\| \nabla_x F(\hat{\theta}_{2n-2} - t\beta_n^- x, w_{2n-1}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1}) \right\| \|\beta_n^- x\|, \end{aligned}$$

$$\begin{aligned} \langle \nabla_x F(\hat{\theta}_{2n-2}, w_{2n}), \beta_n^+ x \rangle &\leq \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| \|\beta_n^+ x\|, \\ \langle \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1}), \beta_n^- x \rangle &\leq \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| \|\beta_n^- x\|. \end{aligned}$$

В итоге получаем оценку для (12):

$$\begin{aligned} &F(\hat{\theta}_{2n-2} + \beta_n^+ x, w_{2n}) - F(\hat{\theta}_{2n-2} - \beta_n^- x, w_{2n-1}) \leq F(\hat{\theta}_{2n-2}, w_{2n}) - F(\hat{\theta}_{2n-2}, w_{2n-1}) + \\ &+ \|\nabla_x F(\hat{\theta}_{2n-2} + \beta_n^+ x, w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| \|\beta_n^+ x\| + \|\nabla_x F(\hat{\theta}_{2n-2} - \beta_n^- x, w_{2n-1}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| \|\beta_n^- x\| + \\ &+ \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| \|\beta_n^+ x\| + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| \|\beta_n^- x\|. \end{aligned}$$

Подставив получившуюся оценку в (10), последовательно выводим

$$\begin{aligned} &\mathbb{E} \left\{ \frac{(y_{2n} - y_{2n-1})^2}{4\beta_n^+ + \beta_n^-} \|\mathcal{K}_n(\Delta_n)\|^2 \mid \mathcal{F}_{n-1} \right\} \leq \\ &\leq \frac{1}{2} \sup_x \mathcal{K}_n(x)^2 \mathbb{E}\{v_{2n-1}^2 + v_{2n}^2 \mid \mathcal{F}_{n-1}\} + \\ &+ \iiint \left( F(\hat{\theta}_{2n-2}, w_{2n}) - F(\hat{\theta}_{2n-2}, w_{2n-1}) + \|\nabla_x F(\hat{\theta}_{2n-2} + \beta_n^+ x, w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| \|\beta_n^+ x\| + \right. \\ &+ \|\nabla_x F(\hat{\theta}_{2n-2} - \beta_n^- x, w_{2n-1}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| \|\beta_n^- x\| + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| \|\beta_n^+ x\| + \\ &+ \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| \|\beta_n^- x\| \left. \right)^2 \|\mathcal{K}_n(x)\|^2 P_n(dx) P_w(dw_{2n}) P_w(dw_{2n-1}) = \\ &= \frac{1}{2} \sup_x \mathcal{K}_n(x)^2 \mathbb{E}\{v_{2n-1}^2 + v_{2n}^2 \mid \mathcal{F}_{n-1}\} + \\ &+ \iiint \left( F(\hat{\theta}_{2n-2}, w_{2n}) - F(\hat{\theta}_{2n-2}, w_{2n-1}) + \left( \|\nabla_x F(\hat{\theta}_{2n-2} + \beta_n^+ x, w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| \|\beta_n^+\| + \right. \right. \\ &+ \|\nabla_x F(\hat{\theta}_{2n-2} - \beta_n^- x, w_{2n-1}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| \|\beta_n^-\| + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| \|\beta_n^+\| + \\ &+ \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| \|\beta_n^-\| \left. \right) \|\beta_n^-\| \|\beta_n^+\| \left. \right)^2 \|\mathcal{K}_n(x)\|^2 P_n(dx) P_w(dw_{2n}) P_w(dw_{2n-1}) \leq \\ &\leq \frac{1}{2} \sup_x \mathcal{K}_n(x)^2 \mathbb{E}\{v_{2n-1}^2 + v_{2n}^2 \mid \mathcal{F}_{n-1}\} + \\ &+ \iiint \left( F(\hat{\theta}_{2n-2}, w_{2n}) - F(\hat{\theta}_{2n-2}, w_{2n-1}) + \left( A \|t\| \|\beta_n^+\| \|x\| \|\beta_n^+\| + A \|t\| \|\beta_n^-\| \|\beta_n^-\| + \right. \right. \\ &+ \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| \|\beta_n^+\| + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| \|\beta_n^-\| \left. \right) \|\beta_n^-\| \|\beta_n^+\| \left. \right)^2 \times \\ &\times \|\mathcal{K}_n(x)\|^2 P_n(dx) P_w(dw_{2n}) P_w(dw_{2n-1}) \leq \\ &\leq \frac{1}{2} \sup_x \mathcal{K}_n(x)^2 \mathbb{E}\{v_{2n-1}^2 + v_{2n}^2 \mid \mathcal{F}_{n-1}\} + \\ &+ \iiint \left( F(\hat{\theta}_{2n-2}, w_{2n}) - F(\hat{\theta}_{2n-2}, w_{2n-1}) + \left( A \|x\| (\|\beta_n^+\|^2 + \|\beta_n^-\|^2) + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| \|\beta_n^+\| + \right. \right. \\ &+ \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| \|\beta_n^-\| \left. \right) \|\beta_n^-\| \|\beta_n^+\| \left. \right)^2 \|\mathcal{K}_n(x)\|^2 P_n(dx) P_w(dw_{2n}) P_w(dw_{2n-1}) \leq \sigma_n^2 \end{aligned} \quad (13)$$

в силу ограниченности  $\mathcal{K}_n$  и компактности их носителя. Распишем разность

$$F(\hat{\theta}_{2n-2}, w_{2n}) - F(\hat{\theta}_{2n-2}, w_{2n-1}) \quad (14)$$

в виде

$$F(\theta, w_{2n}) + \langle \nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n}), (\hat{\theta}_{2n-2} - \theta) \rangle - F(\theta, w_{2n-1}) - \langle \nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n-1}), (\hat{\theta}_{2n-2} - \theta) \rangle.$$

Добавив и отняв  $\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})$  и  $\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})$ , получаем

$$\begin{aligned} &= F(\theta, w_{2n}) - F(\theta, w_{2n-1}) + \langle \nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n}), (\hat{\theta}_{2n-2} - \theta) \rangle - \\ &\quad - \langle \nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n-1}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1}), (\hat{\theta}_{2n-2} - \theta) \rangle + \\ &\quad + \langle \nabla_x F(\hat{\theta}_{2n-2}, w_{2n}), (\hat{\theta}_{2n-2} - \theta) \rangle - \langle \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1}), (\hat{\theta}_{2n-2} - \theta) \rangle. \end{aligned}$$

Оценим скалярные произведения:

$$\begin{aligned} &\langle \nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n}), (\hat{\theta}_{2n-2} - \theta) \rangle \leq \\ &\leq \|\nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| \|\hat{\theta}_{2n-2} - \theta\|, \\ &\langle \nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n-1}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1}), (\hat{\theta}_{2n-2} - \theta) \rangle \leq \\ &\leq \|\nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n-1}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| \|\hat{\theta}_{2n-2} - \theta\|, \\ &\langle \nabla_x F(\hat{\theta}_{2n-2}, w_{2n}), (\hat{\theta}_{2n-2} - \theta) \rangle - \langle \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1}), (\hat{\theta}_{2n-2} - \theta) \rangle \leq \\ &\leq \|\hat{\theta}_{2n-2} - \theta\| (\|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\|). \end{aligned}$$

Получили оценку для (14) в виде

$$\begin{aligned} &F(\hat{\theta}_{2n-2}, w_{2n}) - F(\hat{\theta}_{2n-2}, w_{2n-1}) \leq \\ &\leq F(\theta, w_{2n}) - F(\theta, w_{2n-1}) + \|\hat{\theta}_{2n-2} - \theta\| (\|\nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| + \\ &\quad + \|\nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n-1}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| + \\ &\quad + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\|). \end{aligned}$$

Подставим получившееся выражение в (13):

$$\begin{aligned} &\mathbb{E} \left\{ \frac{(y_{2n} - y_{2n-1})^2}{4\beta_n^+ + \beta_n^-} \|\mathcal{K}_n(\Delta_n)\|^2 \mid \mathcal{F}_{2n-2} \right\} \leq \\ &\leq \frac{1}{2} \sup_x \mathcal{K}_n(x)^2 \mathbb{E} \{ v_{2n-1}^2 + v_{2n}^2 \mid \mathcal{F}_{2n-2} \} + \\ &+ \left( F(\theta, w_{2n}) - F(\theta, w_{2n-1}) + \|\hat{\theta}_{2n-2} - \theta\| (\|\nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| + \right. \\ &\quad + \|\nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n-1}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| + \\ &\quad + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\|) + (A \|x\| (\beta_n^{+2} + \beta_n^{-2}) + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| \beta_n^+ + \\ &\quad \left. + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| \beta_n^-) \|x\|^2 \|\mathcal{K}_n(x)\|^2 P_n(dx) P_w(dw_{2n}) P_w(dw_{2n-1}) \right). \quad (15) \end{aligned}$$



Введем обозначения:

$$\begin{aligned}
 a &= F(\theta, w_{2n}) - F(\theta, w_{2n-1}) + \left( A\|x\|(\beta_n^{+2} + \beta_n^{-2}) + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\|\beta_n^+ + \right. \\
 &\quad \left. + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\|\beta_n^- \right) \|x\|, \\
 b &= \|\nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| + \\
 &+ \|\nabla_x F(\theta + t(\hat{\theta}_{2n-2} - \theta), w_{2n-1}) - \nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\| + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n})\| + \|\nabla_x F(\hat{\theta}_{2n-2}, w_{2n-1})\|, \\
 (a + \|\hat{\theta}_{2n-2} - \theta\|b)^2 &= a^2 + ab\|\hat{\theta}_{2n-2} - \theta\| + b^2(\|\hat{\theta}_{2n-2} - \theta\|)^2.
 \end{aligned}$$

Так как

$$\begin{aligned}
 2\|\hat{\theta}_{2n-2} - \theta\| &\leq 1 + \|\hat{\theta}_{2n-2} - \theta\|^2, \\
 \|\hat{\theta}_{2n-2} - \theta\| &\leq 0.5 + 0.5\|\hat{\theta}_{2n-2} - \theta\|^2,
 \end{aligned}$$

то

$$(a + \|\hat{\theta}_{2n-2} - \theta\|b)^2 \leq a^2 + ab(0.5 + 0.5\|\hat{\theta}_{2n-2} - \theta\|^2) + b^2(\|\hat{\theta}_{2n-2} - \theta\|)^2 = a^2 + 0.5ab + (0.5ab + b^2)\|\hat{\theta}_{2n-2} - \theta\|^2.$$

После подстановки в (15) получаем оценку

$$\begin{aligned}
 &\mathbb{E} \left\{ \frac{(y_{2n} - y_{2n-1})^2}{4\beta_n^+ + \beta_n^-} \|\mathcal{K}_n(\Delta_n)\|^2 \mid \mathcal{F}_{n-1} \right\} \leq \frac{1}{2} \sup_x \mathbf{K}_n(x)^2 \mathbb{E}\{v_{2n-1}^2 + v_{2n}^2 \mid \mathcal{F}_{n-1}\} + \\
 &+ \iiint \left( a^2 + 0.5ab + (0.5ab + b^2)\|\hat{\theta}_{2n-2} - \theta\|^2 \right) \times \|\mathcal{K}_n(x)\|^2 P_n(dx) P_w(dw_{2n}) P_w(dw_{2n-1}) \leq C_\tau V(x) + C_\sigma
 \end{aligned}$$

в силу компактности носителя  $\mathcal{K}_n$ , равномерной ограниченности функции  $F$  и  $\nabla F$ .

**Условие Д.** Условие на начальное приближение  $\mathbb{E}\{\|\hat{\theta}_0 - \theta\|^2\} < \infty$  выполнено.

Для формулировки основного результата статьи надо ввести ограничения на последовательность  $\{\alpha_n\}$ . Пусть  $\alpha_n$  удовлетворяет ограничениям

$$\begin{aligned}
 0 \leq \alpha_n \leq \frac{4\mu - 2}{LC_\tau}, \quad \alpha_n \leq \frac{2\mu - 1 - \sqrt{(2\mu - 1)^2 - 2LC_\tau}}{LC_\tau}, \\
 \alpha_n \geq \frac{2\mu - 1 + \sqrt{(2\mu - 1)^2 - 2LC_\tau}}{LC_\tau}, \quad \sum_n \alpha_n = \infty.
 \end{aligned} \tag{16}$$

Введем обозначения:

$$\begin{aligned}
 v_n &= \alpha_n \left( 2\mu - 1 - \frac{L\alpha_n C_\tau}{2} \right), \quad \phi_n = \alpha_n \gamma_n + \frac{L}{2} \alpha_n^2 C_\sigma, \quad \gamma_n = \frac{1}{2} \left( C_1 \frac{(\beta_n^+)^2 + (\beta_n^-)^2}{\beta_n^+ + \beta_n^-} \right)^2, \\
 \psi_n &= \frac{\phi_n}{v_n}, \quad \lambda_n = \left( \frac{\psi_n}{\psi_{n+1}} - 1 \right) \frac{1}{v_n}, \quad \lambda'_n = \left( 1 - \frac{\psi_{n+1}}{\psi_n} \right) \frac{1}{v_{n+1}}.
 \end{aligned}$$

**Условие Е.**  $0 \leq v_n \leq 1$ ,  $\sum_n v_n = \infty$ .

При выборе  $\alpha_n$ , удовлетворяющих ограничениям (16), условие Е выполняется.

Проведенный анализ доказывает следующую теорему, являющуюся прямым следствием теорем 1–3 из [18].

**Теорема.** Пусть выполняются предположения 1–5 и условия (4), (5) и (16).

Если  $\overline{\lim}_{n \rightarrow \infty} \psi_n \leq \psi$ ,  $\psi \geq 0$ , тогда  $\overline{\lim}_{n \rightarrow \infty} \mathbb{E}\{\|\hat{\theta}_n - \theta\|^2\} \leq \psi$ . Если при этом  $\psi_n \leq \psi$  для всех  $n$ , то

$$\mathbb{E}\{\|\hat{\theta}_n - \theta\|^2\} \leq \mathbb{E}\{\|\hat{\theta}_0 - \theta\|^2\} \prod_{i=0}^n (1 - v_i) + \psi \left( 1 - \prod_{i=0}^n (1 - v_i) \right).$$

Если  $\psi_n \rightarrow 0$ , тогда  $\mathbb{E}\{\|\hat{\theta}_n - \theta\|^2\} \rightarrow 0$ .

Если, кроме того,

а)  $\overline{\lim}_{n \rightarrow \infty} \lambda_n \leq \lambda < 1$ , то

$$\mathbb{E}\{\|\hat{\theta}_n - \theta\|^2\} \leq \frac{\psi_{n+1}}{1-\lambda} + o(\psi_{n+1});$$

б)  $\lambda_n \leq \lambda < 1$  для всех  $n$ , то

$$\mathbb{E}\{\|\hat{\theta}_n - \theta\|^2\} \leq \mu_{n+1} \left[ \frac{1}{1-\lambda} + \left( \frac{\mathbb{E}\{\|\hat{\theta}_0 - \theta\|^2\}}{\psi_0} - \frac{1}{1-\lambda} \right) \prod_{i=0}^n (1 - (1-\lambda)v_i) \right];$$

в)  $\overline{\lim}_{n \rightarrow \infty} \lambda'_n \geq \lambda > 1$ , то

$$\mathbb{E}\{\|\hat{\theta}_n - \theta\|^2\} = O\left(\prod_{i=0}^n (1 - v_i)\right);$$

г)  $\lambda'_n \geq \lambda > 1$  для всех  $n$ , то

$$\mathbb{E}\{\|\hat{\theta}_n - \theta\|^2\} \leq \left( \mathbb{E}\{\|\hat{\theta}_0 - \theta\|^2\} + \frac{\psi_0}{1-\lambda} \right) \prod_{i=0}^n (1 - v_i).$$

Если  $\sum_n \phi_n < \infty$ , тогда  $\|\hat{\theta}_n - \theta\|^2 \rightarrow 0$  с вероятностью единица.

При этом для всякого  $\varepsilon > 0$ ,  $n_0 \geq 0$  имеем

$$P\left\{\|\hat{\theta}_n - \theta\|^2 \leq \varepsilon \forall n \geq n_0\right\} \geq 1 - \frac{\mathbb{E}\{\|\hat{\theta}_0 - \theta\|^2\} + \sum_n \phi_n}{\varepsilon},$$

где  $P(\cdot)$  обозначает вероятность.

Если, кроме того,  $\overline{\lim}_{n \rightarrow \infty} \lambda'_n \geq \lambda > 1$ ,

то для всякого  $K > 0$  найдется такое  $C = C\left(K, \mathbb{E}\{\|\hat{\theta}_0 - \theta\|^2\}\right)$ , что

$$P\left\{\|\hat{\theta}_n - \theta\|^2 \leq (C + K) \prod_{i=0}^n (1 - v_i) \forall n\right\} \geq 1 - C / K,$$

а если  $\lambda'_n \geq \lambda > 1$  для всех  $n$ , то

$$P\left\{\|\hat{\theta}_n - \theta\|^2 \leq \left(K + \mathbb{E}\{\|\hat{\theta}_0 - \theta\|^2\} + \frac{\psi_0}{1-\lambda}\right) \prod_{i=0}^n (1 - v_i) \forall n\right\} \geq 1 - \frac{\mathbb{E}\{\|\hat{\theta}_0 - \theta\|^2\}}{K} - \frac{\psi_0}{K(\lambda - 1)}.$$

#### 4. ЗАКЛЮЧЕНИЕ

Борисом Теодоровичем Поляком в 70-е годы XX в. были получены результаты, которые до сих пор активно используются исследователями. В настоящей статье на основании теорем 1–3 из статьи [18] получены новые результаты о свойствах оценок алгоритма (3), расширяющие ранее установленные в [6] асимптотические свойства оценок, добавляя к ним оценки точности при конечном числе итераций.

#### СПИСОК ЛИТЕРАТУРЫ

1. Граничин О. Н. Об одной стохастической рекуррентной процедуре при зависимых помехах в наблюдении, использующей на входе пробные возмущения // Вест. Ленингр. ун-та. 1989. Т. 1. № 1. С. 19–21.
2. Поляк Б. Т., Цыбаков А. Б. О некоторых способах ускорения сходимости итерационных методов // Пробл. передачи информ. 1990. Т. 26. № 2. С. 126–133.
3. Spall J. C. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation // IEEE Trans. Autom. Control. 1992. V. 37. Iss. 6. P. 332–341.
4. Граничин О. Н. Процедура стохастической аппроксимации с возмущением на входе // Автоматика и телемехан. 1992. № 2. С. 97–104.
5. Polyak B. T., Tsybakov A. B. On stochastic approximation with arbitrary noise (the KW-case) // Adv. Sov. Math. 1992. V. 12. Iss. 8.

6. Граничин О. Н., Поляк Б. Т. Рандомизированные алгоритмы оценивания и оптимизации при почти произвольных помехах. М.: Наука, 2003.
7. Spall J. C. A one-measurement form of simultaneous perturbation stochastic approximation // *Automatica*. 1997. V. 33. Iss. 1. P. 109–112.
8. Chen H., Duncan T. E., Pasik-Duncan B. A Kiefer-Wolfowitz algorithm with randomized differences // *IEEE Trans. Autom. Control*. 1999. V. 44. Iss. 3. P. 442–453.
9. Lobanov A., Gasnikov A., Stonyakin F. Highly smoothness zero-order methods for solving optimization problems under PL condition // arXiv preprint arXiv:2305.15828; 2023.
10. Dvinskikh D., Tominin V., Tominin Y., Gasnikov A. Gradient-free optimization for non-smooth saddle point problems under adversarial noise // arXiv preprint arXiv:2202.06114; 2022.
11. Akhavan A., Chzhen E., Pontil M., Tsybakov A. B. Gradient-free optimization of highly smooth functions: improved analysis and a new algorithm // arXiv preprint arXiv:2306.02159; 2023.
12. Antal C., Granichin O. N., Levi S. Adaptive autonomous soaring of multiple UAVs using simultaneous perturbation stochastic approximation // 49th IEEE Conf. on Decision and Control (CDC), 2010. P. 3656–3661.
13. Granichin O., Amelina N. Simultaneous perturbation stochastic approximation for tracking under unknown but bounded disturbances // *IEEE Trans. Autom. Control*. 2015. V. 60. Iss. 6. P. 1653–1658.
14. Granichin O. N., Erofeeva V. A., Ivanskiy Y. V., Jiang Y. Simultaneous perturbation stochastic approximation-based consensus for tracking under unknown-but-bounded disturbances // *IEEE Trans. Autom. Control*. 2021. V. 66. Iss. 8. P. 3710–3717.
15. Erofeeva V. A., Granichin O. N., Tursunova M., Sergeenko A., Jiang Y. Accelerated simultaneous perturbation stochastic approximation for tracking under unknown-but-bounded disturbances // *Am. Control Conf. (ACC) 2022*. P. 1582–1587.
16. Поляк Б. Т. О некоторых способах ускорения сходимости итерационных методов // *Ж. вычисл. матем. и матем. физ.* 1964. Т. 4. № 5. С. 791–803.
17. Аблаев С. С., Безносиков А. Н., Гасников А. В., Двинских Д. М., Лобанов А. В., Пучинин С. М., Стонякин Ф. С. О некоторых работах Бориса Теодоровича Поляка по сходимости градиентных методов и их развитии // *Ж. вычисл. матем. и матем. физ.* 2024. Т. 64. № 4. С. 25–64.
18. Поляк Б. Т. Сходимость и скорость сходимости итеративных стохастических алгоритмов. I. Общий случай // *Автоматика и телемехан.* 1976. № 12. С. 83–94.

## POLYAK'S METHOD BASED ON THE STOCHASTIC LYAPUNOV FUNCTION FOR JUSTIFYING THE CONSISTENCY OF ESTIMATES PRODUCED BY A STOCHASTIC APPROXIMATION SEARCH ALGORITHM UNDER AN UNKNOWN-BUT-BOUNDED NOISE

O. N. Granichin<sup>a,b,\*</sup>, Yu. V. Ivanskiy<sup>a,b</sup>, K. D. Kopylova<sup>a</sup>

<sup>a</sup>*St. Petersburg State University, Universitetskaya nab., 7–9, St. Petersburg, 199034 Russia*

<sup>b</sup>*Mechanical Engineering Research Institute of the Russian Academy of Sciences, Vasilievsky Island, 61 Bolshoi Ave, St. Petersburg, Russia*

<sup>\*</sup>*e-mail: oleg\_granichin@mail.ru*

Received 15 September, 2023

Revised 06 November, 2023

Accepted 20 November, 2023

**Abstract.** In 1976–1977, Polyak published in the journal *Avtomatica i Telemekhanika* (Automation and Remote Control) two remarkable papers on how to study the properties of estimates of iterative pseudogradient algorithms. The first paper published in 1976 considered the general case based on the stochastic Lyapunov function, and the second one considered the linear case. The assumptions formulated in these papers and the estimates obtained in them can still be considered the state-of-the-art. In the current paper, Polyak's approach is applied to the study of the properties of estimates of a (randomized) stochastic approximation search algorithm for the case of unknown-but-bounded noise in observations. The obtained asymptotic estimates were already known earlier, and exact estimates for a finite number of observations are published for the first time.

**Keywords:** stochastic approximation search algorithm, unknown-but-bounded noise, gradient approximation, smoothing kernels, gradient-free methods, methods with inexact oracle.

Посвящаем эту работу памяти  
Федора Павловича Васильева

## ОБ ИЗБЫТОЧНОСТИ НЕВЫРОЖДЕННОСТИ ГЕССИАНА ДЛЯ ГЕОМЕТРИЧЕСКОЙ СКОРОСТИ СХОДИМОСТИ МЕТОДА НЬЮТОНА ПРИ МИНИМИЗАЦИИ ВЫПУКЛЫХ ФУНКЦИЙ<sup>1)</sup>

© 2024 г. Ю. Г. Евтушенко<sup>1,2,\*</sup>, А. А. Третьяков<sup>1,3,\*\*</sup>

<sup>1</sup> 119333 Москва, ул. Вавилова, 44, ФИЦ ИУ РАН, Россия,

<sup>2</sup> 141701 Долгопрудный, М.о., Институтский переулок, 9, Московский физико-технический институт  
(государственный университет), Россия

<sup>3</sup> 08-110 Siedlce, Siedlce University, Faculty of Exact and Natural Sciences, Poland

\*e-mail: yuri-evtushenko@yandex.ru

\*\*e-mail: prof.tretyakov@gmail.com

Поступила в редакцию 10.08.2023 г.

Переработанный вариант 07.11.2023 г.

Принята к публикации 07.11.2023 г.

В статье устанавливается новое свойство выпуклых функций, позволяющее добиться геометрической скорости сходимости метода Ньютона в процессе минимизации. А именно, установлено, что даже в случае вырождения гессиана в решении, ньютоновская система разрешима в окрестности точки минимума, т.е. градиент целевой функции принадлежит образу матрицы вторых производных и поэтому можно применять аналоги метода Ньютона. Библ. 10.

**Ключевые слова:** выпуклая функция, метод Ньютона, разрешимость, сходимость, скорость сходимости, регулярность.

**DOI:** 10.31857/S0044466924040045, **EDN:** ZKJRII

В задаче поиска безусловного минимума рассматриваются функции  $f(\cdot)$ , определенные и достаточно гладкие в окрестности  $U(x^*)$  точки минимума функции  $n$  вещественных переменных. Всяду далее множество точек минимума функции  $f$  обозначается как  $X^* = \text{Argmin} f$  и предполагается непустым. Необходимое условие минимума функции  $f$  в точке  $x^*$  задается равенством  $f'(x^*) = 0$ , при этом матрица вторых производных функции в точке минимума является положительно полуопределенной. Следует отметить, что исследования по методу Ньютона посвящено значительное число научных работ, среди которых укажем [1–7]. Со многими работами можно ознакомиться в обзорной статье [8]. В данной работе показывается, что несмотря на возможную вырожденность матрицы Гессе в точке  $x^*$ , в окрестности этой точки градиент целевой функции принадлежит образу ее второй производной и, следовательно, ньютоновская система относительно направления спуска разрешима в точках этой окрестности. Это топологическое свойство выпуклости и экстремальности позволяет по новому взглянуть на численные методы ньютоновского типа и обосновать скорость сходимости этих методов без обременительного предположения относительно невырожденности матрицы Гессе в точке  $x^*$ . Для задачи безусловной оптимизации получены свойства, уточняющие неравенство Лоясевича (см. [9–10]). Именно, неравенство обобщено для всего спектра производных до определенного порядка. В работе рассматриваются функции, производные которой равномерны по аргументу и в совокупности по порядку ограничены в некоторой окрестности  $U(x^*)$  точки  $x^*$ , т.е. для данной окрестности существует положительная константа  $M$  такая, что значения ее производных любого порядка не превосходят по абсолютному значению эту константу. Кроме того, объектом рассмотрения в работе будут достаточно гладкие в окрестности точки минимума функции, т.е. функции, имеющие неограниченное число порядков производных. Всяду далее без ограничения общности считаем  $f(x^*) = 0$  и обозначаем  $f^{(0)}(x) = f(x)$ . Для функции одной переменной справедлива

<sup>1)</sup> Работа выполнена при финансовой поддержке РФФИ (проект № 21-71-30005).

**Лемма 1.** Если  $x^*$ -точка изолированного локального минимума достаточно гладкой в  $U(x^*)$  функции  $f : R \rightarrow R$ , производные которой равномерно по аргументу и в совокупности по порядку ограничены в некоторой окрестности  $U(x^*)$ , то существует четная степень  $2p, p = p_f \in \mathbb{N}$ , для которой справедливы неравенства

$$f^{(2p)}(x^*) > 0, \quad f^{(k)}(x^*) = 0, \quad \frac{f^{(k)}(x)}{(x - x^*)^{2p-k}} \geq C_k, \quad k = 0, 1, 2, \dots, 2p - 1,$$

при всех  $x \in \dot{U}(x^*)$ , где положительные константы  $C_k = C_{k,f}, k = 0, 1, 2, \dots, 2p - 1$ , не зависят от  $x$ .

**Доказательство.** Поскольку  $x^*$  — точка локального минимума функции  $f$ , то  $f'(x^*) = 0$ . Сначала покажем, что в условиях леммы невозможна ситуация, когда производные  $f^{(k)}(x^*) = 0$  при любом порядке  $k, k = 1, 2, \dots$ . Действительно в этом случае для фиксированной точки  $x = x^* + t \in U(x^*)$  в силу ограниченности производных в окрестности  $x^*$  из формулы Тэйлора при нулевых производных до любого фиксированного порядка  $k$  следует неравенство  $|f(x^* + \theta t)| \leq \frac{M(|\theta t|)^k}{k!}$ , где  $M$  — обозначенная выше верхняя грань для множества значений производных функции  $f$  в указанной окрестности и  $\theta \in (0, 1)$ . Отсюда и из условия изолированности локального минимума  $x^*$  значение

$$f(x^* + t) = \int_0^1 f'(x^* + \theta t) d\theta \leq M / (k + 1)!$$

При достаточно больших  $k$  это означает противоречие с возможным предположением  $f(x) \neq 0, x \neq x^*$  что противоречит изолированности локального минимума  $x^*$ . Следовательно, существует конечное

$$k > 1 : f^{(k)}(x^*) \neq 0, \quad f^{(l)}(x^*) = 0, \quad l = 1, 2, \dots, k - 1.$$

При этом порядок  $k$  может быть только четным:  $k = 2p, p \in \mathbb{N}$  и  $f^{(2p)}(x^*) > 0$ , поскольку  $x^*$  — точка локального минимума функции  $f$ . Теперь для завершения доказательства достаточно обозначить через  $C_0$  значение  $\frac{1}{(2p + 1)!} f^{(2p)}(x^*)$ , тогда из формулы Тэйлора и ограниченности производной порядка  $2p + 1$  в окрестности  $U(x^*)$  будет вытекать первое неравенство леммы. Разложение в окрестности  $x^*$  по формуле Тэйлора производной  $f^{(k)}(x)$  дает остальные неравенства леммы при любом  $k = 1, 2, \dots, 2p - 1$ , если через  $C_k$  обозначить  $\frac{1}{(2p + 1 - k)!} f^{(2p)}(x^*)$ .

**Следствие 1.** При выполнении условий леммы 1 функция  $f$  локально выпукла.

Действительно, если  $f^{(2)}(x^*) > 0$ , то в малой окрестности  $x^*$  вторая производная будет оставаться положительной, что означает локальную выпуклость  $f(x)$ . Если же  $f^{(2)}(x^*) = 0$ , то, как показано в лемме 1,  $f^{(2p)}(x^*) > 0$  для некоторого  $p = p_f \in \mathbb{N}, p > 1$  и при этом  $f^{(k)}(x^*) = 0, k = 1, 2, \dots, 2p - 1$ . Тогда  $f^{(2)}(x) \geq C_2(x - x^*)^{2p-2}$  для любого  $x \in U(x^*)$ , положительная константа  $C_2$  определена ранее в лемме 1. Последнее неравенство также означает локальную выпуклость  $f$ .

Для функции  $n$  переменных производную  $k$ -го порядка по направлению  $h$  в точке  $x$  будем обозначать как  $f_h^{(k)}(x)$ .

**Следствие 2.** Если достаточно гладкая функция  $f : R^n \rightarrow R$ , производные которой равномерно по аргументу и в совокупности по порядку ограничены в некоторой окрестности  $U(x^*)$  изолированной точки минимума  $x^*$ , то для каждого  $h \in R^n : \|h\| = 1$ , существует четная степень  $2p_{h,f}, p_{h,f} \in \mathbb{N}$ , для которой справедливы неравенства

$$f^{(2p_{h,f})}(x^*) > 0, \quad f^{(k)}(x^*) = 0, \quad \frac{f^{(k)}(x^* + th)}{t^{2p_{h,f}-k}} \geq C_k,$$

где  $k = 0, 1, 2, \dots, 2p_{h,f} - 1$  при всех  $t \in (0, \delta_h]$ , положительные константы  $C_k = C_{k,h,f}, k = 0, 1, 2, \dots, 2p_{h,f} - 1$ , не зависят от  $t, x^* + \delta_h \in U(x^*)$ , при этом сама функция локально выпукла вдоль направления  $h$ .

Здесь элемент  $p = p_{h,f} \in \mathbb{N}$  определяется направлением  $h, \|h\| = 1$  и не зависит от малых  $t$ , для которых  $x^* + \delta_h \in U(x^*)$ , но значение  $\delta_h > 0$  зависит от  $h$  и в общем случае эта величина может быть бесконечно малой относительно  $h$ . Далее в лемме 2 будет показано, что для случая выпуклой функции  $f$  можно гарантировать существование верхней границы для величины  $p_{h,f} \in \mathbb{N}$  и положительной нижней границы для  $\delta_h$  на множестве векторов  $h : \|h\| = 1$ .

Из леммы 1 вытекает, что в точках достаточно малой выколотой окрестности решения корректно определен оператор Ньютона  $\psi(x) = x - f^{(2)}(x)^{-1}f(x)$ . Для случая произвольной размерности пространства  $R^n$  лемма 1 означает выпуклость функции  $f$  вдоль любой прямой, проходящей через точку  $x^*$ , при условии достаточной гладкости функции на пересечении прямой и окрестности  $U(x^*)$ . Кроме того, для любой прямой, проходящей через точку  $x^*$  вдоль вектора  $h$ , справедливо неравенство  $f_h^{(2m)}(x^*) \geq C_2$  для некоторой степени  $2p$ ,  $p = p_h \in \mathbb{N}$ , и некоторой константы  $C_2 = C_{2,h} > 0$ . Неравенство Лоясевича гарантирует выполнимость данного неравенства в случае аналитичности в окрестности  $U(x^*)$  функции при некоторых  $p, C_2$ , не зависящих от  $h$ . Выпуклость в указанной окрестности функции позволяет требовать лишь достаточную гладкость функции и получить свойство “устойчивости” неравенства Лоясевича, что расширяет область применения метода Ньютона. Имеет место

**Лемма 2.** Если выпуклая достаточно гладкая функция  $f : R^n \rightarrow R$ , производные которой равномерно по аргументу и в совокупности по порядку ограничены в некоторой окрестности  $U(x^*)$  изолированной точки минимума  $x^*$ , то существует четная степень  $2p$ ,  $p = p_f \in \mathbb{N}$ , константа  $C = C_f > 0$  и величина  $\delta = \delta_f > 0$ , для которых

$$f(x^* + th) \geq Ct^{2p}$$

при всех  $h : \|h\| = 1, t \in (0, \delta]$ .

**Доказательство.** Докажем существование такого  $p = p_f \in \mathbb{N}$ , обладающего свойством: если  $f_h^{(i)}(x^*) = 0, i = 0, 1, 2, \dots, 2p'(h) - 1, f_h^{(2p'(h))}(x^*) > 0$ , то  $p'(h) \leq p$ . Предположим противное, тогда для некоторых последовательностей  $h_k : \|h_k\| = 1, t_k : t_k \rightarrow +0$  имеет место неравенство  $f(x^* + t_k h_k) < C_k t_k^{m_k}$ , где  $C_k > 0$  — элементы некоторой ограниченной, а  $m_k \in \mathbb{N}$  — возрастающей последовательностей. Без ограничения общности можно считать, что  $h_k \rightarrow h, C_k \rightarrow C, k \rightarrow \infty$ . Обозначим  $\text{conv}\{h_k, h_{k+1}, h_{k+2}, \dots, h_{k+n}\}$  через  $V_k$ . В случае  $\dim V_k < n$  указанный набор векторов изменяется путем прибавления к  $n - \dim V_k$  векторам линейно независимых приращений длины порядка  $C_k t_k^{2p}, C_k \rightarrow 0, k \rightarrow \infty$ , после чего можно считать  $\dim V_k = n$ . Пусть далее  $h'_k \in \text{int} V_k, h''_k = h + \alpha_k(h'_k - h), \alpha_k \in (0, 1), \alpha_k = \inf\{\alpha > 0 : h + \alpha(h'_k - h) \in V_k\}$ . Из выпуклости и непрерывности  $f$  следует  $f(x^* + t_k h''_k) \leq 2C_k t_k^{m_k}$ . С другой стороны, из леммы 1 следует, что для некоторого  $p = p_h \in \mathbb{N}$  существует степень  $k \in \mathbb{N}, k \leq p$ , для которой производная  $f_h^{(2k)}(x^*) \geq C_2 > 0$  для некоторого  $C_2 > 0$ . Поскольку  $h_k \rightarrow h, k \rightarrow \infty$ , то при достаточно больших номерах  $k$  будут выполняться неравенства  $f(x^* + t_k h''_k) \geq C_2 t_k^{2k}$ , что противоречит предположению.

**Следствие 3.** Из доказательства леммы 2 следует существование степени  $2p, p = p_f \in \mathbb{N}$ , а также констант  $C_k = C_{k,f}, k = 1, 2, \dots, 2p' - 1$ , для которых

$$f_h^{(2p')}(x^*) > 0, f_h^{(k)}(x^*) = 0, k = 0, 1, \dots, 2p' - 1, f_h^{(k)}(x^* + th) \geq C_k t^{2p-k}, 0 < t \leq \delta,$$

при этом  $p' = p'(h) \leq p$  при всех  $h : \|h\| = 1$ .

Для точки  $x = x^* + th \in U(x^*), \|h\| = 1$ , определим базис  $G = G(h) = \{g_1, g_2, \dots, g_n\}$  пространства  $R^n$  и соответствующий индекс  $q = q(h)$  следующим образом. Рассмотрим матрицы

$$a_k = \frac{1}{(k-1)!} f^{(k)}(x^*) [h]^{k-2}, k = 2, 3, \dots$$

Тогда в точке  $x = x^* + th$  для достаточно гладкой функции  $f$  справедливы соотношения

$$f^{(2)}(x)th = \sum_{k \geq 2} (k-1) a_k h t^{k-1}, f'(x) = \sum_{k \geq 2} a_k h t^{k-1}, f(x) = \sum_{k \geq 2} \frac{a_k [h]^2 t^k}{k}. \tag{1}$$

Далее определим номер  $k_1 \geq 2$  как минимальный среди тех номеров  $k$ , для которых одномерное пространство  $L_1 = L_{k_1} = \text{Lin}\{a_k h\} \neq \{0\}$ . Прямую  $L_1$  переобозначим через  $L^1$ , вектор  $g_1$  определим как  $\frac{a_{k_1} h}{\|a_{k_1} h\|}$  и далее номер  $k_2 > k_1$  определим как минимальный номер  $k$ , для которого  $\text{Lin}\{a_k h\}$  не содержится в  $L^1$ . Тогда  $L_{k_2} = \text{Lin}\{a_{k_2} h\}, \dim(L_{k_2} \oplus L^1) = 2$ , одномерное подпространство  $L_2$  определяется как  $\text{Pr}_{(L^1)^\perp} L_{k_2}$ , вектор  $g_2$  определяется как нормированный вектор  $\text{Pr}_{L_2} a_{k_2} h$ . При этом  $\text{Pr}_{L_2} a_{k_2} h \subseteq \text{Pr}_{L_2} \text{Im} a_{k_2}$ . Далее пространство  $L^2$  определим  $L^1 \oplus L_2$ . Далее для каждого  $j = 3, 4, \dots, q = q(h)$  аналогично определяются прямые  $L_j = \text{Pr}_{(L^{j-1})^\perp} L_{k_j}, j = 3, 4, \dots, q$ , и соответствующие единичные векторы  $g_j, j = 1, 2, \dots, q$ . Номер  $q = q(h) = \dim \text{Lin}\{a_1 h, a_2 h, \dots\} \leq n$ , прямая сумма  $L^{j-1} \oplus L_j$  обозначается как  $L^j$  и является

подпространством в  $R^n$  размерности  $j$ . На конечном этапе построено подпространство  $L^q$  размерности  $q = q(h) \leq n$ . Обозначим ортогональное дополнение подпространства  $L^q$  через  $H : H = (L^q)^\perp$  в случае  $q < n$  и  $H = \{0\}$  при  $q = n$ . Определим базис  $G(h)$  пространства  $R^n$  как набор построенных единичных взаимно ортогональных векторов  $g_1, g_2, \dots, g_q$ , направленных вдоль построенных ортогональных прямых  $L_l, l = 1, 2, \dots, q$ , и произвольного ортогонального базиса  $g_{q+1}, g_{q+2}, \dots, g_n$  подпространства  $H$ .

Справедлива

**Теорема 1.** Если для достаточно гладкой функции  $f :: R^n \rightarrow R$  производные равномерно по аргументу и в совокупности по порядку ограничены в некоторой окрестности  $U(x^*)$  изолированной точки минимума  $x^*$ , то для любого фиксированного  $h : \|h\| = 1$ , система

$$f^{(2)}(x^* + th)s = f'(x^* + th) \tag{2}$$

разрешима относительно  $s$  при достаточно малых  $t$ .

**Доказательство.** Решение системы  $s$  в базисе  $G$  при любом фиксированном векторе  $h$  определим по координатно следующим образом. Положим  $\text{Pr}_H s = 0$ , далее координата  $s_q$  в соответствии с (1) определяется из условия  $\text{Pr}_{L_q} \left( (k_q - 1)a_{k_q} t^{k_q - 2} s \right) = \text{Pr}_{L_q} f'(x) = \text{Pr}_{L_q} \sum_{k \geq k_q} a_k h t^{k-1}$ . При этом  $s_q = \text{Pr}_{L_q} h t / (k_q - 1) + o(t)$ . Подстановка координаты  $s_q$  в систему (2) позволяет получить координату  $s_{q-1} = \text{Pr}_{L_{q-1}} h t / (k_{q-1} - 1) + o(t)$ . Далее последовательно определяются координаты  $s_l, l = q - 2, q - 3, \dots, 1$ . Решение системы (2) вообще говоря, не единственно, но из (1) следует, что координаты любого решения  $s$  в базисе  $G = G(h)$  имеют вид

$$s_j = \frac{\text{Pr}_{L_j} h t}{k_j - 1} + o(t), \quad j = 1, 2, \dots, q. \tag{3}$$

**Замечание 1.** Для получения единственного решения системы (2) определим задачу

$$|s|^2 \rightarrow \min_s, \quad f^{(2)}(x)s = f'(x), \tag{4}$$

решение которой существует при каждом фиксированном  $h$  при всех достаточно малых  $t$ , при условиях, указанных в теореме 1. При этом длина интервала для  $t$ , в пределах которого решение существует, зависит от  $h$ .

**Пример 1.** Для невыпуклой функции  $f(x) = x_1^4 + (x_2 - x_1^2)^2$  вывод теоремы 1 нарушается для точек параболы  $x_2 = 4x_1^2$ . Таким образом, длина интервала, для которого имеет место теорема 1, стремится к нулю, по мере приближения вектора  $h$  к  $(1, 0)$ .

Далее будет показано, что в случае выпуклой функции можно указать общий для всех векторов единичной сферы радиус окрестности переменной  $t$ , в пределах которой задача (4) разрешима. Из вида целевой функции задачи (4) следует, что ее решение  $s = f^{(2)}(x)^+ f'(x)$  где  $(\cdot)^+$  означает псевдообратный оператор на своем образе. Координаты решения этой задачи в базисе  $G$  удовлетворяют условию  $\text{Pr}_H s = 0$ .

**Теорема 2.** При выполнении условий леммы 2 при всех  $x$ , достаточно близких к  $x^*$ , задача (4) разрешима и ее решение удовлетворяет условиям

$$(f(x), s) \geq M_1 \|x - x^*\|^{2p}, \quad (f''(x)s, s) \geq M_2 \|x - x^*\|^{2p}, \tag{5}$$

где степень  $2p$  определена в лемме 2, константы  $M_1 = M_{1,f}, M_2 = M_{2,f} > 0$  не зависят от  $x$ .

**Доказательство.** Из леммы 2 и (1) следует, что при всяком  $h : \|h\| = 1$ , номера  $k_j$ , определяющие базис  $G$ , удовлетворяют условию: существует индекс  $l_1 \in \{k_1, k_1 + 1, \dots, 2p\}$ , для которого  $a_{l_1} [h]^2 \geq C > 0$ , степень  $2p$  определена в лемме 2 и от  $h$  не зависит, константа  $C > 0$  также не зависит от  $h$  и определяется в следствии к лемме 2. Последнее неравенство эквивалентно условию  $\frac{a_{l_1} [h]^2}{l_1 - 1} \geq C_1 t^{2p}$  при малых  $t$ . Обозначая  $\frac{C}{2p - 1}$  через  $M_1$ , получим неравенство в утверждении теоремы. Аналогично из следствия к лемме 2 получается и второе неравенство.

**Следствие 4.** При выполнении условий леммы 2 для выпуклой функции  $f$  при всех  $x$ , достаточно близких к  $x^*$ , имеет место разрешимость относительно  $s$  системы

$$f^{(2)}(x)s = f'(x), \tag{6}$$

что равносильно справедливости включения

$$f'(x) \in \text{Im} f^{(2)}(x) \tag{7}$$

независимо от величины ранга матрицы  $f^{(2)}(x)$ .

Далее будем обозначать множество  $\{x \in R^n : f(x) \leq f(x^*) + \varepsilon\}$  как  $X_\varepsilon^*$ , диаметр множества  $A$  как  $\text{diam}A$ . Справедлива

**Теорема 3.** При выполнении условий леммы 2 существует натуральное  $p$  и константа  $\bar{C} < \infty$ , для которых справедливо неравенство

$$\text{diam}X_\varepsilon^* < \bar{C}\varepsilon^{\frac{1}{2p}}. \tag{8}$$

**Замечание 2.** Указанные ранее свойства справедливы в предположении, что множество точек минимума функции  $f(x)$  представляет изолированную точку. Если отказаться от данного предположения, то при выполнении остальных предположений теоремы 1 справедливо представление

$$\text{Argmin}f = \{x^*\} + L, \tag{9}$$

где  $L$  — собственное подпространство  $R^n$ , которое определяется условием  $f^k(x^*)[h]^k = 0$  для всякого натурального  $k$ .

Из теорем 1, 2 не следует положительная определенность матрицы  $f^{(2)}(x)$  в окрестности точки минимума  $x^*$ . Тем не менее они гарантируют применимость метода Ньютона для поиска точки минимума гладкой функции при условии подходящей начальной точки, поскольку задача (1) позволяет получить вектор перехода в итерационной схеме без предположения выпуклости целевой функции. Кроме того, в данном случае удается получить монотонность по аргументу метода Ньютона, а при более сильных предположениях — линейную скорость сходимости. В случае же выпуклости целевой функции полученное свойство справедливо без дополнительных предположений.

Рассмотрим сначала применение результата теоремы 1 для доказательства монотонности по аргументу схемы метода Ньютона. Именно, определим оператор Ньютона  $\psi(x, s) = x - f^{(2)}(x)^+ f(x)$ , где  $f^{(2)}(x)^+ f(x) = s$  — решение задачи (4). Из процесса построения решения  $s$  в теореме 1, и замечания к теореме следует, что для любого фиксированного  $h : \|h\| = 1$ , при достаточно малых значениях  $t$  данный оператор корректно определен без предположения выпуклости функции  $f$ . Для получения оценки скорости сходимости дополнительно к предположениям теоремы 1 естественно определяется следующее свойство.

**Определение 1.** Функция  $f$  слабо регулярна в точке  $x^*$ , если существует константа  $d = d_f > 0$ , для которой  $\max_{j \leq q} |h_j| \geq d$  для всякого  $h, \|h\| = 1$ . Здесь как и ранее  $q = q(h) = \dim \text{Lin}\{a_2h, a_3h, \dots\}, h_j$  — координата вектора  $h$  с номером  $j$  базисе  $G$ .

**Теорема 4.** Если функция  $f$  слабо регулярна в точке  $x^*$ , то при выполнении условий теоремы 1 существует  $\lambda \in (0, 1)$ , для которого  $\|\psi(x^* + th, s) - x^*\| \leq \lambda t$  при всех достаточно малых  $t$ .

**Доказательство.** Для любого решения системы (1) в силу предположения слабой регулярности в точке  $x^*$  существует номер  $j$ , для которого  $\|\text{Pr}_{L_{k_j}} h\| > d$ . Отсюда

$$\begin{aligned} \|th - s\| &\leq \left\| t \sum_{l=1, l \neq k_j}^q \text{Pr}_{L_l} h - \text{Pr}_{L_l} s \right\| + t \|\text{Pr}_H h\| + |t \text{Pr}_{L_{k_j}} h - \text{Pr}_{L_{k_j}} s| \leq \\ &\leq t(1 - \alpha) + t\alpha \left( 1 - \frac{1}{k_j - 1} \right) \leq t \left( 1 - \frac{d}{k_j - 1} \right) = \lambda t, \\ \lambda = \lambda(h) &= 1 - \frac{d}{k_j - 1} \in (0, 1), \alpha = \alpha(h) \geq d. \end{aligned}$$

**Замечание 3.** Полученное неравенство не означает линейной сходимости метода Ньютона, поскольку знаменатель  $\lambda$ , который участвует в оценке, вообще говоря, зависит от  $h$ . В случае выпуклой функции полученный результат справедлив в более сильном виде:  $\|\psi(x, s) - x^*\| \leq \lambda \|x - x^*\|$  при всех  $x$  из достаточно малой окрестности  $x^*$ , поскольку для выпуклых функций справедливо следующее свойство.

**Определение 2.** Функция  $f$  равномерно регулярна в точке  $x^*$ , если существуют номер  $m = m_f \in \mathbb{N}$  и константа  $d = d_f > 0$ , для которых найдется индекс  $j \leq q : k_j \leq m$ , такой что  $\|\text{Pr}_{L_j} h\| \geq d$ .

Имеет место



**Теорема 5.** Если функция  $f$  равномерно регулярна в точке  $x^*$ , то при выполнении условий теоремы 1 существует  $\lambda \in (0,1)$ , для которого  $\|\psi(x,s) - x^*\| \leq \lambda \|x - x^*\|$  при всех  $x$  из достаточно малой окрестности  $x^*$ .

**Доказательство.** Из теоремы 4 следует, что полученный знаменатель  $\lambda = \lambda(h)$  ограничен сверху равномерно по  $h$ ,  $\|h\| = 1 : \lambda(h) = 1 - \frac{d}{k_j - 1} \leq \lambda' < 1$ , где  $j : k_j \leq m$ . Такой номер  $j$  существует, как показано в теореме 1. Кроме того, из условия равномерной регулярности следует, что  $\|h_H^\perp\| > d$  для некоторого  $d > 0$ , не зависящего от  $h$ , где  $h_A$  – проекция вектора  $h$  на подпространство  $A$ . Отсюда следует оценка для знаменателя  $\lambda$ , не зависящая от  $h$ .

**Замечание 4.** Из леммы 2 следует, выпуклая функция равномерно регулярна в точке  $x^*$ :  $k_j = k_j(h) \leq m = 2p, \|h\| = 1$ . Условие равномерной регулярности является необходимым и достаточным условием линейной скорости сходимости метода Ньютона при выполнении условий теоремы 1 без предположения о выпуклости функции  $f$ . Итерационная схема метода Ньютона имеет вид

$$x_{k+1} = x_k - f^{(2)}(x_k)^+ f(x_k),$$

а оценка скорости сходимости метода будет линейная:

$$\|x_{k+1} - x^*\| \leq \lambda \|x_k - x^*\|, \lambda \in (0,1), k = 0,1,2,\dots,$$

при начальном приближении достаточно близком к решению  $x^*$  в случае равномерно регулярной в точке минимума функции.

**Следствие 5.** При выполнении условий теоремы 2 для выпуклой функции скорость сходимости итерационной схемы метода Ньютона является линейной при любом выборе начальной точки  $x$  из достаточно малой окрестности решения. Действительно, из теоремы 2 следует, что выпуклая функция является равномерно регулярной в решении  $x^*$ , откуда вытекает линейная оценка скорости сходимости.

Для получения сходимости метода Ньютона с более высокой скоростью, чем линейная, рассмотрим модифицированный оператор  $\psi_1(x,s)$ , по координатной записи которого в базисе  $G$  представляет собой  $\psi_1(x,s)_j = x - (k_j - 1)s_j, j = 1,2,\dots,q; \psi_1(x,s)_H = x_H - g(x,h)$ , где вектор-функция  $g(x,h) : H \rightarrow H, s = f^{(2)}(x)^+ f(x)$ . Оператор  $\psi_1(x,s)_H$  позволяет определить модифицированную схему Ньютона  $x_{k+1} = \psi_1(x_k, s_k)$ . Из теоремы 1 следует

**Теорема 6.** Модифицированная схема Ньютона гарантирует сверхлинейную оценку скорости сходимости при хорошем начальном приближении в том и только том случае, когда функция  $g(x,h)$  удовлетворяет условию  $\|g(x,h) - h_H\| = o(t)$ .

## СПИСОК ЛИТЕРАТУРЫ

1. Бомаджо Б., Лебедев К. А. Метод Ньютона для нахождения экстремумов сильно выпуклых функций // Международный научно-исследовательский журнал. 2015. Выпуск 6–2 (37). С. 11–14.
2. Заботин В. И., Черняев Ю. А. Метод Ньютона для задачи минимизации выпуклой дважды гладкой функции на предвыпуклом множестве // Ж. вычисл. матем. и матем. физ. 2018. Т. 58. № 3. С. 340–345. DOI: 10.7868/S0044466918030031.
3. Budzko D., Cordero A., Torregrosa J. R. Modification of Newton's Method to extend the convergence domain // SeMA Journal. 2014. Т. 66. № 1. С. 43–53. DOI: 10.1007/s40324-014-0020-y.
4. Nesterov Y. Accelerating the cubic regularization of Newton's method on convex problems // Mathematical Programming. 2008. Т. 112. № 1. С. 159–181. DOI: 10.1007/s10107-006-0089-x.
5. Polyak B., Tremba A. New versions of Newton method: step-size choice, convergence domain and under-determined equations // Optimization Methods and Software. 2019. С. 1272–1303. DOI: 10.1080/10556788.2019.1669154.
6. Nesterov Y., Polyak B. T. Cubic regularization of Newton method and its global performance // Mathematical Programming. 2006. Т. 108. № 1. С. 177–205.
7. Поляк Б. Т. Градиентные методы минимизации функционалов // Ж. вычисл. матем. и матем. физ. 1963. Т. 3. № 4. С. 643–653.
8. Поляк Б. Т. Метод Ньютона и его роль в оптимизации и вычислительной математике // Труды Института системного анализа Российской академии наук. 2006. Т. 28. С. 44–62.
9. Colding T. H., Minicozzi W. P. Lojasiewicz inequalities and applications // arXiv:1402.5087. 2014.
10. Lojasiewicz S. Division d'une distribution par une fonction analytique de variables reelles // C. R. Acad. Sci. 1958. Т. 246. № 5. С. 683–686.

## ON THE REDUNDANCY OF HESSIAN NONSINGULARITY FOR LINEAR CONVERGENCE RATE OF THE NEWTON METHOD APPLIED TO THE MINIMIZATION OF CONVEX FUNCTIONS

Yu. G. Evtushenko<sup>a,b,\*</sup>, A. A. Tret'yakov<sup>a,c,\*\*</sup>

<sup>a</sup>*Federal Research Center "Information Science and Control," Russian Academy of Sciences,  
Vavilova St., 44, Moscow, 119333 Russia*

<sup>b</sup>*Moscow Institute of Physics and Technology,  
Institutskiy per., 9, Dolgoprudny, Moscow oblast, 141701 Russia*

<sup>c</sup>*Siedlce University, Faculty of Exact and Natural Sciences,  
Siedlce, 08-110 Poland*

\**e-mail: yuri-evtushenko@yandex.ru*

\*\**e-mail: prof.tretyakov@gmail.com*

Received 10 August, 2023

Revised 07 November, 2023

Accepted 07 November, 2023

**Abstract.** A new property of convex functions that makes it possible to achieve the linear rate of convergence of the Newton method during the minimization process is established. Namely, it is proved that, even in the case of singularity of the Hessian at the solution, the Newtonian system is solvable in the vicinity of the minimizer; i.e., the gradient of the objective function belongs to the image of the matrix of second derivatives and, therefore, analogs of the Newton method may be used.

**Keywords:** convex function, Newton method, solvability, convergence, convergence rate, regularity.

**АЛГОРИТМЫ УПРАВЛЕНИЯ С ИТЕРАТИВНЫМ ОБУЧЕНИЕМ  
ВЫСШЕГО ПОРЯДКА ДЛЯ ЛИНЕЙНЫХ СИСТЕМ<sup>1)</sup>**© 2024 г. **П. В. Пакшин<sup>1,\*</sup>, Ю. П. Емельянова<sup>1,\*\*</sup>, М. А. Емельянов<sup>1,\*\*\*</sup>**<sup>1</sup> 607227 Арзамас, ул. Калинина, 19, АПИ НГТУ им. Р.Е. Алексеева, Россия

\*e-mail: pakshinpv@gmail.com

\*\*e-mail: emelianovajulia@gmail.com

\*\*\*e-mail: mikhailemelianovarzamas@gmail.com

Поступила в редакцию 06.11.2023 г.

Переработанный вариант 18.11.2023 г.

Принята к публикации 20.11.2023 г.

Алгоритмы управления с итеративным обучением появились в связи с задачами повышения точности выполнения повторяющихся операций роботами. Они используют информацию из прошлых повторений для корректировки управляющего сигнала на текущем повторении. Чаще всего используется только информация из предыдущего повторения. Алгоритмы управления с итеративным обучением, которые используют информацию из нескольких предыдущих повторений, называются алгоритмами высшего порядка. В последнее время в литературе повысился интерес к этим алгоритмам в связи с задачами роботизированных аддитивных производств. Однако помимо того, что эти алгоритмы мало изучены, относительно их свойств имеются противоречивые оценки. В настоящей статье предлагаются новые алгоритмы управления с итеративным обучением высшего порядка для линейных дискретных и дифференциальных систем, идея построения которых основана на аналогии с многошаговыми методами в теории оптимизации, в частности, с методом тяжелого шарика. Приведен пример, подтверждающий возможность увеличения скорости сходимости ошибки обучения при использовании таких алгоритмов. Библ. 25. Фиг. 7.

**Ключевые слова:** управление с итеративным обучением, линейные системы, алгоритмы высшего порядка, повторяющиеся процессы, сходимость, устойчивость, векторная функция Ляпунова, линейные матричные неравенства.

DOI: 10.31857/S0044466924040056, EDN: ZKANBV

**1. ВВЕДЕНИЕ**

Методы и алгоритмы управления с итеративным обучением появились как инструмент повышения точности выполнения повторяющихся операций роботами (см. [1]). После появления пионерской работы [1], в силу относительной простоты и естественной структуры конечных алгоритмов, управление с итеративным обучением (УИО) быстро приобрело широкую популярность, как среди теоретиков, так и среди практиков. Для первоначального знакомства с этим активно развивающимся направлением можно рекомендовать обзоры [2], [3], а также недавнюю монографию [4]. В настоящее время алгоритмы УИО находят эффективное применение в аддитивных производствах, в частности, в установках высокоточного многослойного лазерного напыления (см. [5], [6]) в медицинских роботах для реабилитации больных, перенесших инсульт (см. [7], [8]), в устройствах поддержки желудка сердца (см. [9]) и в других многочисленных приложениях.

Одной из важнейших задач является повышение скорости сходимости алгоритмов УИО. Для решения этой задачи используются различные подходы. В настоящей работе предлагается подход, мотивированный идеей многошаговых методов в теории оптимизации (см. [10]). Хорошо известно, что такие двухшаговые методы как метод тяжелого шарика и метод сопряженных градиентов позволяют существенно ускорить сходимость градиентного метода (см. [10]–[12]). Алгоритмы УИО, аналогично алгоритму градиентного метода, для формирования управления на текущем шаге используют информацию с текущего и предыдущего шагов и, как в многошаговых методах оптимизации, можно ожидать, что учет в этих алгоритмах более глубокой “предыстории” позволит ускорить сходимость. Далее алгоритмы УИО, использующие информацию

<sup>1)</sup> Работа выполнена при финансовой поддержке РФФ (грант № 22-21-00612), <https://rscf.ru/project/22-21-00612>.

с текущего и предыдущего шагов будем называть алгоритмами первого порядка, а при использовании информации с большего числа предыдущих шагов — алгоритмами высшего порядка.

Предлагаемый подход не является абсолютно новым и в различных версиях изучался в литературе (см. [13]–[21]). Однако во всех этих работах на текущем шаге используется лишь неполная информация об управлении на прошлых шагах, что представляется нелогичным в рамках указанных аналогий с многошаговыми методами оптимизации, где роль “управления” выполняет аргумент оптимизируемой функции. Тем не менее, в работах [14], [15] показано что даже такие “неполные” структуры УИО высшего порядка обеспечивают более высокую скорость сходимости, чем структуры УИО первого порядка. Такие же результаты были получены в [16]–[19], причем в [16]–[18] утверждается, что эффект ускорения в алгоритмах высшего порядка достигается за счет эффектов обучения в течение нескольких предыдущих итераций. В [20] относительно УИО высшего и первого порядков представлены весьма неопределенные выводы. Авторы этой работы считают, что трудно обосновать достоинства той или иной стратегии УИО, хотя идея УИО высшего порядка кажется естественной. В работе [21] предложен алгоритм УИО на основе метода Нестерова (см. [22]), но скорость сходимости алгоритма по доказанному авторами составляет  $O(1/k)$ , в то время, как хорошо известно, она составляет  $O(1/k^2)$  (см. [22], [23]). Кроме того, предложенный алгоритм относится к типу причинных алгоритмов УИО и обладает известными недостатками (см. [2]), существенно ограничивающими их применение. Важно отметить, что выводы перечисленных работ сделаны на рассмотрении частных случаев и примеров, и ни в одной из них не приведен строгий сравнительный анализ в общей постановке даже для линейных систем.

В литературе изучались также другие свойства УИО высшего порядка (см. [24], [25]). В [24] утверждается, что реальной мотивацией использования УИО высшего порядка является уменьшение влияния помех и шума. В [25] рассматривается оптимальность УИО в смысле минимизации следа ковариационной матрицы ошибок управления в классе некоррелированных случайных возмущений. Показано, что УИО высшего порядка не уменьшает минимальное значение этого следа по сравнению с УИО первого порядка. Таким образом, выводы [24] и [25] противоречат друг другу. Окончательный вывод требует дополнительного исследования, поскольку в [24] не представлено достаточно полных доказательств, а в [25] в алгоритме УИО используется дискретный аналог производной необработанного случайного сигнала, и такое решение нельзя считать допустимым, поскольку уже само вычисление такой величины приводит к возрастанию дисперсии. Вопросы скорости сходимости ошибки обучения в [24] и [25] не изучались.

Наконец, отметим, что совсем недавно появился активный интерес к разработке и применению алгоритмов УИО высшего порядка в задачах аддитивных производств (см. [26]), мотивированный особенностями новых прикладных задач.

В настоящей работе, в отличие от упомянутых работ, строятся непричинные алгоритмы УИО, при этом по аналогии с многошаговыми методами на текущем повторении используется полная предыстория, т.е. полностью учитываются управляющие воздействия на определенном числе предыдущих повторений. Для построения алгоритмов используется разработанная авторами теория устойчивости повторяющихся процессов (см. [27]), в рамках которой условия сходимости ошибки обучения выражаются через свойства дивергенции векторной функции Ляпунова. В рассматриваемом линейном случае эта теория позволяет использовать эффективную технику линейных матричных неравенств. Общих результатов, доказывающих преимущество алгоритмов УИО высшего порядка в смысле увеличения скорости сходимости ошибки обучения, пока получить не удалось, но приведенный пример, использующий характеристики реального портального робота, убедительно подтверждает это преимущество и служит серьезной мотивацией для дальнейшего развития общей теории.

## 2. ДИСКРЕТНЫЕ СИСТЕМЫ

### 2.1. Постановка задачи

Рассмотрим дискретную систему в повторяющемся режиме, описываемую линейной моделью в пространстве состояний

$$\begin{aligned} x_k(p+1) &= Ax_k(p) + bu_k(p), \\ y_k(p) &= c^T x_k(p), \quad 0 \leq p \leq N-1, \quad k \geq 0, \end{aligned} \quad (1)$$

где  $k$  — номер повторения,  $x_k(p) \in \mathbb{R}^{n_x}$  — вектор состояния,  $u_k(p) \in \mathbb{R}$  — скалярная входная переменная, имеющая смысл управления, действующего на систему,  $y_k(p) \in \mathbb{R}$  — выходная переменная, которая в рассматриваемом классе задач часто называется профилем повторения,  $N$  — продолжительность повторения,  $A$ ,  $b$  и  $c$  — постоянная матрица и векторы соответствующих размеров.

Закон управления формируется в соответствии с алгоритмом

$$u_k(p) = \sum_{i=0}^d \tau_i v_{k-i}(p), \quad (2)$$

$$v_{k+1}(p) = v_k(p) + \Delta v_{k+1}(p), \quad k = 0, 1, \dots, \quad (3)$$

где  $v_k = 0$ , если  $k \in [-d, 0]$ ,  $d$  — число предыдущих повторений, информация с которых используется на текущем повторении,  $\Delta v_{k+1}$  — обновляющая поправка на повторении  $k + 1$ ,  $\tau_i$ ,  $i \in [0, d]$ , — весовые коэффициенты.

Обозначим через  $y_{\text{ref}}(p) \in \mathbb{R}$ ,  $0 \leq p \leq N - 1$ , желаемую (эталонную) траекторию, которая должна воспроизводиться на выходе системы, тогда

$$e_k(p) = y_{\text{ref}}(p) - y_k(p) \quad (4)$$

представляет собой ошибку обучения на повторении  $k$ . Поставим задачу найти такую последовательность управлений  $\{u_k\}$ , что выполняются условия

$$|e_k(p)| \leq \kappa \varrho^k, \quad \kappa > 0, \quad 0 < \varrho < 1, \quad (5)$$

$$\lim_{k \rightarrow \infty} |u_k(p)| = |u_\infty(p)| < \infty, \quad p \in [0, N - 1]. \quad (6)$$

Эти условия обеспечивают сходимость ошибки обучения к нулю при  $k \rightarrow \infty$  при сохранении ограниченности последовательности входных переменных. Далее эти условия будем называть *условиями сходимости*.

Из условий (2) и (3) следует, что эта задача сводится к нахождению обновляющей поправки и весовых коэффициентов, при которых гарантируется выполнение условий (5) и (6). В соответствии с концепцией, предложенной Я. З. Цыпкиным (см. [28]), эта задача относится к задачам обучения. При  $d = 0$  и  $\tau_0 = 1$ , когда на текущем повторении используется только информация с предыдущего повторения, эта задача известна как задача построения алгоритма управления с итеративным обучением первого порядка, которая наиболее полно изучена в литературе. При  $d > 0$  имеем задачу построения алгоритма управления с итеративным обучением высшего порядка, которая изучена значительно меньше.

Заметим, что рассмотрение систем со скалярными входными и выходными переменными связано лишь с возможностью более компактного изложения результатов, и в случае систем с векторными входными и выходными переменными принципиальных затруднений не возникает.

## 2.2. Построение модели относительно приращений переменных в форме повторяющегося процесса

Введем скалярные переменные  $\check{x}_{k,1}(p) = v_k(p)$ ,  $\check{x}_{k,2}(p) = v_{k-1}(p)$ , ...,  $\check{x}_{k,d}(p) = v_{k-d+1}(p)$ ,  $\check{x}_{k-d+1}(p) = v_{k-d}(p)$  и вектор  $\check{x}_k = [\check{x}_{k,1} \dots \check{x}_{k,d+1}]^T$ . Тогда по построению

$$\check{x}_k(p) = A_d \check{x}_{k-1}(p) + b_d v_k(p), \quad (7)$$

где

$$A_d = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}, \quad b_d = [1 \ 0 \ \dots \ 0]^T. \quad (8)$$

С использованием (7) первое уравнение в (1) можно записать как

$$x_k(p+1) = Ax_k(p) + b\theta^T \check{x}_k(p), \quad (9)$$

где  $\theta = [\tau_0 \ \tau_1 \ \dots \ \tau_d]^T$ .

Введем вспомогательные векторы

$$\eta_k(p) = x_k(p) - x_{k-1}(p), \quad \check{\eta}_k(p) = \check{x}_k(p) - \check{x}_{k-1}(p). \quad (10)$$

Тогда из (7) и (10) следует

$$\check{\eta}_k(p) = A_d \check{\eta}_{k-1}(p) + b_d \Delta v_k(p), \quad (11)$$

где  $\Delta v_k(p) = v_k(p) - v_{k-1}(p)$ . Учитывая (10), запишем уравнение в приращениях вдоль траектории повторения в виде

$$\eta_k(p+1) = A \eta_k(p) + b \theta^T A_d \check{\eta}_{k-1}(p) + b \theta^T b_d \Delta v_k(p). \quad (12)$$

Учитывая, что согласно (4)  $e_k(p) = y_{ref}(p) - c^T x_k(p)$ , и, используя (12), получим следующую систему уравнений в приращениях переменных:

$$\eta_k(p+1) = A \eta_k(p) + b \theta^T A_d \check{\eta}_{k-1}(p) + b \theta^T b_d \Delta v_k(p),$$

$$\check{\eta}_k(p) = A_d \check{\eta}_{k-1}(p) + b_d \Delta v_k(p), \quad (13)$$

$$\check{e}_k(p) = -c^T A \eta_k(p) - c^T b \theta^T A_d \check{\eta}_{k-1}(p) + \check{e}_{k-1}(p) - c^T b \theta^T b_d \Delta v_k(p),$$

где  $\check{e}_k(p) = e_k(p+1)$ .

Зададим корректирующую поправку в виде

$$\Delta v_k(p) = k_1^T \eta_k(p) + k_2 e_{k-1}(p+1), \quad (14)$$

где вектор  $k_1$  и скаляр  $k_2$  подлежат определению из условий сходимости (5), (6). Подставляя (14) в (13), запишем модель в приращениях переменных в окончательном виде:

$$\eta_k(p+1) = (A + b \theta^T b_d k_1^T) \eta_k(p) + b \theta^T A_d \check{\eta}_{k-1}(p) + b \theta^T b_d k_2 \check{e}_{k-1}(p),$$

$$\check{\eta}_k(p) = b_d k_1^T \eta_k(p) + A_d \check{\eta}_{k-1}(p) + b_d k_2 \check{e}_{k-1}(p), \quad (15)$$

$$\check{e}_k(p) = -c^T (A + b \theta^T b_d k_1^T) \eta_k(p) - c^T b \theta^T A_d \check{\eta}_{k-1}(p) + (1 - c^T b \theta^T b_d k_2) \check{e}_{k-1}(p).$$

Модель (15) записана в форме дискретного повторяющегося процесса, относящегося к классу так называемых 2D систем (см. [29]). Один из возможных подходов к получению условий сходимости ошибки обучения может быть основан на теории устойчивости повторяющихся процессов с использованием векторных функций Ляпунова (см. [27]). Эта теория далее используется в настоящей статье для построения алгоритмов управления с итеративным обучением.

### 2.3. Общие условия сходимости ошибки обучения

Введем в рассмотрение вектор  $\bar{e}_k(p) = [\check{\eta}_{k-1}^T(p) \check{e}_{k-1}(p)]^T$  и определим на траекториях системы (15) векторную функцию Ляпунова

$$V(\eta_k(p), \bar{e}_k(p)) = \begin{bmatrix} V_1(\eta_k(p)) \\ V_2(\bar{e}_k(p)) \end{bmatrix}, \quad (16)$$

где  $V_1(\eta_k(p)) > 0$ ,  $\eta_k(p) \neq 0$ ,  $V_2(\bar{e}_k(p)) > 0$ ,  $\bar{e}_k(p) \neq 0$ ,  $V_1(0) = 0$ ,  $V_2(0) = 0$ . Определим на траекториях системы (15) дискретный аналог оператора дивергенции

$$\mathcal{D}V(\eta_k(p), \bar{e}_k(p)) = V_1(\eta_k(p+1)) - V_1(\eta_k(p)) + V_2(\bar{e}_{k+1}(p)) - V_2(\bar{e}_k(p)). \quad (17)$$

Для краткости далее будем называть этот оператор просто *дивергенцией*. Дальнейший анализ основан на следующей теореме, которая является следствием теоремы 1 из [27].

**Теорема 1.** *Предположим, что существует векторная функция Ляпунова (16) и положительные скаляры  $c_1, c_2$  и  $c_3$  такие, что на траекториях системы (15) выполняются неравенства*

$$c_1 \|\eta_k(p)\|^2 \leq V_1(\eta_k(p)) \leq c_2 \|\eta_k(p)\|^2, \quad (18)$$

$$c_1 \|\bar{e}_k(p)\| \leq V_2(\bar{e}_k(p)) \leq c_2 \|\bar{e}_k(p)\|^2, \tag{19}$$

$$DV(\eta_{k+1}(p), \bar{e}_k(p)) \leq -c_3 (\|\eta_{k+1}(p)\|^2 + \|\bar{e}_k(p)\|^2). \tag{20}$$

Тогда алгоритм управления с итеративным обучением (2), (3) с обновляющей поправкой (14) гарантирует выполнение условий сходимости ошибки обучения (5), (6).

**Доказательство.** В соответствии с теоремой 1 из [27] при условиях (18)–(20) существуют  $\alpha > 0$  и  $0 < \lambda < 1$  такие, что

$$\|\eta_k(p)\|^2 + \|\bar{e}_k(p)\|^2 \leq \alpha \lambda^{k+p} \leq \alpha \lambda^k. \tag{21}$$

Следовательно,  $\|\eta_k(p)\| \leq \kappa \varrho^k$ , где  $\kappa = \sqrt{\alpha}$ ,  $\varrho = \sqrt{\lambda}$ , и, принимая во внимания определение  $\bar{e}$ ,  $|e_k(p)| \leq \kappa \varrho^k$ . Таким образом, справедливо (5). Далее, поскольку  $\Delta v_k(p)$  определено соотношением (14), из (3) с учетом (21) получим

$$|v_{k+1}(p)| \leq |v_k(p)| + \alpha_0 \lambda^{(k+p+1)/2}, \tag{22}$$

где  $\alpha_0 = \sqrt{2\alpha \max\{|k_1|, |k_2|\}}$ . Из (22) следует, что

$$|v_k(p)| \leq |v_0(p)| + \alpha_0 \lambda^{(p+1)/2} \sum_{n=0}^{k-1} \lambda^{n/2}.$$

В правой части последнего неравенства стоит сходящаяся при  $k \rightarrow \infty$  геометрическая прогрессия, следовательно, предел в левой части  $|v_\infty(p)|$  при  $k \rightarrow \infty$  существует, и при этом справедлива оценка

$$|v_\infty(p)| \leq |v_0(p)| + \frac{\alpha_0 \lambda^{(p+1)/2}}{1 - \lambda^{1/2}},$$

откуда в соответствии с (2) следует, что справедливо условие (6).

#### 2.4. Построение алгоритма

Обозначим  $\xi_k(p) = [\eta_k^T(p) \check{\eta}_{k-1}^T(p) \check{e}_{k-1}^T(p)]^T$  и введем в рассмотрение матрицы согласованных размеров

$$\bar{A} = \begin{bmatrix} A & b\theta^T A_d & 0 \\ 0 & A_d & 0 \\ -c^T A & -c^T b\theta^T A_d & 1 \end{bmatrix}, \quad \bar{b} = \begin{bmatrix} b\theta^T b_d \\ b_d \\ -c^T b\theta^T b_d \end{bmatrix}, \quad \Gamma = \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Рассмотрим дискретное неравенство Риккати

$$\bar{A}^T \bar{P} \bar{A} - (1 - \sigma) \bar{P} - \bar{A}^T \bar{P} \bar{b} [\bar{b}^T \bar{P} \bar{b} + R]^{-1} \bar{b}^T \bar{P} \bar{A} + Q \preceq 0 \tag{23}$$

относительно матрицы  $\bar{P} = \text{diag}[P_1 \ P_2] \succ 0$ , где  $P_1 \in \mathbb{R}^{n_x \times n_x}$ ,  $P_2 \in \mathbb{R}^{d+2 \times d+2}$ ,  $0 < \sigma < 1$ ,  $Q \succ 0$  и  $R \succ 0$  — соответственно весовая матрица и весовой коэффициент. Применяя лемму о дополнении Шура, получим, что если линейные матричные неравенства

$$\begin{bmatrix} (1 - \sigma) \bar{X} & \bar{X} \bar{A}^T & \bar{X} \\ \bar{A} \bar{X} & \bar{X} + \bar{b} R^{-1} \bar{b}^T & 0 \\ \bar{X} & 0 & Q^{-1} \end{bmatrix} \succeq 0, \quad \bar{X} \succ 0, \tag{24}$$

разрешимы относительно  $X = \text{diag}[X_1 \ X_2] \succ 0$ , где  $X_1$  и  $X_2$  имеют те же размерности, что и  $P_1$  и  $P_2$ , то  $\bar{P} = X^{-1}$ .

Определим

$$f^T = \underbrace{[f_1^T]}_{n_x} \quad \underbrace{0}_{d+1} \quad \underbrace{f_3}_{1} = -[\bar{b}^T \bar{P} \bar{b} + R]^{-1} \bar{b}^T \bar{P} \bar{A} \Gamma \tag{25}$$

и предположим, что выполняется следующее матричное неравенство:

$$\begin{bmatrix} M - M\Gamma - \Gamma M - Q & \Gamma \sqrt{M} \\ \sqrt{M} \Gamma & -I \end{bmatrix} \preceq 0, \tag{26}$$

где  $M = \bar{A}^T \bar{P} \bar{b} [\bar{b}^T \bar{P} \bar{b} + R]^{-1} \bar{b}^T \bar{P} \bar{A}$ . Выбирая компоненты векторной функции Ляпунова (16) в виде квадратичных форм

$$V_1(\eta_k(p)) = \eta_k^T(p) P_1 \eta_k(p), \quad V_2(\bar{e}_k(p)) = \bar{e}_k^T(p) P_2 \bar{e}_k(p)$$

и вычисляя дивергенцию функции (16) вдоль траекторий системы (15) с учетом (23) и (26), получим

$$\begin{aligned} DV(\eta_{k+1}(p), \bar{e}_k(p)) &= \xi_k^T(p) [\bar{A}^T \bar{P} \bar{A} - \Gamma M - M \Gamma + \Gamma M \Gamma - \bar{P}] \xi_k(p) = \\ &= \xi_k^T(p) [\bar{A}^T \bar{P} \bar{A} - (1 - \sigma) \bar{P} - \bar{A}^T \bar{P} \bar{b} [\bar{b}^T \bar{P} \bar{b} + R]^{-1} \bar{b}^T \bar{P} \bar{A} + Q + \\ &+ (\Gamma M \Gamma - \Gamma M - M \Gamma - Q) - \sigma \bar{P}] \xi_k(p) \leq -\xi_k^T(p) \sigma \bar{P} \xi_k(p). \end{aligned} \tag{27}$$

Поскольку  $\bar{P} = \text{diag}[P_1 \ P_2] \succ 0$ , из (27) следует, что выполнены все условия теоремы 1. Таким образом, доказан следующий результат.

**Теорема 2.** *Предположим, что для некоторого выбора весовой матрицы  $Q \succ 0$ , вектора  $\theta$  и скаляров  $R > 0$  и  $0 < \sigma < 1$  линейное матричное неравенство (24) имеет решение  $X \succ 0$ , матрица  $P = X^{-1}$  удовлетворяет неравенству (26),  $k_1 = f_1$ ,  $k_2 = f_3$ , где  $f_1$ , и  $f_3$  — компоненты вектора  $f$  из (25). Тогда алгоритм управления с итеративным обучением (2), (3) с обновляющей поправкой (14) обеспечивает выполнение условий сходимости ошибки обучения (5), (6).*

Заметим, что параметр  $\sigma$  характеризует запас устойчивости системы (15). С его увеличением, при выполнении условий теоремы 2, спектр матрицы  $(A + b\theta^T b_d k_1^T)$  и число  $(1 - c^T b \theta^T b k_2)$ , оставаясь внутри единичного круга, удаляются от его границы.

### 3. ДИФФЕРЕНЦИАЛЬНЫЕ СИСТЕМЫ

#### 3.1. Постановка задачи

Рассмотрим дифференциальную систему в повторяющемся режиме, описываемую линейной моделью в пространстве состояний

$$\begin{aligned} \dot{x}_k(t) &= Ax_k(t) + bu_k(t), \\ y_k(t) &= c^T x_k(t), \quad 0 \leq t \leq T, \quad k \geq 0, \end{aligned} \tag{28}$$

где все обозначения имеют тот же смысл, что и в (1).

Алгоритм управления формируется следующим образом:

$$u_k(t) = \sum_{i=0}^d \tau_i v_{k-i}(t), \tag{29}$$

$$v_{k+1}(t) = v_k(t) + \Delta v_{k+1}(t), \quad k = 0, 1, \dots, \tag{30}$$

где  $v_k = 0$ , если  $k \in [-d, 0]$ ,  $d$  — число предыдущих повторений, информация с которых используется на текущем повторении,  $\Delta v_{k+1}$  — обновляющая поправка,  $\tau_i, i \in [0, d]$ , — весовые коэффициенты.

Обозначим через  $y_{\text{ref}}(t) \in \mathbb{R}, 0 \leq t \leq T$ , желаемую (эталонную) траекторию, тогда

$$e_k(t) = y_{\text{ref}}(t) - y_k(t) \tag{31}$$

будет представлять собой ошибку обучения на повторении  $k$ . Для сходимости ошибки обучения последовательность входных переменных  $\{u_k(t)\}$  должна обеспечивать выполнение условий

$$|e_k(t)| \leq \kappa \varrho^k, \quad \kappa > 0, \quad 0 < \varrho < 1, \tag{32}$$

$$\lim_{k \rightarrow \infty} |u_k(t)| = |u_\infty(t)| < \infty, \quad t \in [0, T]. \tag{33}$$

Из (29) и (30) следует, что, как и в случае дискретной системы, задача сводится к нахождению обновляющей поправки, и при которой выполняются условия (32) и (33).



### 3.2. Вывод модели в приращениях и условия сходимости

Вводя переменную

$$\eta_k(t) = \int_0^t [x_k(\tau) - x_{k-1}(\tau)] d\tau,$$

перепишем (28) в виде

$$\dot{\eta}_k(t) = A\eta_k(t) + b \int_0^t \Delta u_k(\tau) d\tau.$$

Тогда с учетом (29) имеем

$$\dot{\eta}_k(t) = A\eta_k(t) + b \sum_{i=0}^d \tau_i \int_0^t \Delta v_{k-i}(\tau) d\tau,$$

и окончательно, вводя переменные  $\delta v_{k-i}(t) = \int_0^t \Delta v_{k-i}(\tau) d\tau$ ,  $i = 0, 1, \dots, d$ , имеем

$$\dot{\eta}_k(t) = A\eta_k(t) + b \sum_{i=0}^d \tau_i \delta v_{k-i}. \quad (34)$$

Введем вспомогательные переменные

$$\check{\eta}_{k1}(t) = \delta v_k(t), \quad \check{\eta}_{k2}(t) = \delta v_{k-1}(t), \quad \check{\eta}_{k3}(t) = \delta v_{k-2}(t), \quad \dots, \quad \check{\eta}_{kd+1}(t) = \delta v_{k-d}(t)$$

и запишем уравнения динамики относительно повторений:

$$\dot{\check{\eta}}_k(t) = A_d \check{\eta}_{k-1}(t) + b_d \delta v_k(t), \quad (35)$$

где  $\check{\eta}_k(t) = [\check{\eta}_{k1}(t) \check{\eta}_{k2}(t) \dots \check{\eta}_{kd+1}(t)]$ ,  $A_d$  и  $b_d$  определены в (8). Уравнение (34) с учетом (35) можно переписать в виде

$$\dot{\eta}_k(t) = A\eta_k(t) + b\theta^T A_d \check{\eta}_{k-1}(t) + b\theta^T b_d \delta v_k(t). \quad (36)$$

Из (36) и (31) следует

$$e_k(t) = -c^T A \eta_{k+1}(t) - c^T b \theta^T A_d \check{\eta}_{k-1}(t) + e_{k-1}(t) - c^T b \theta^T b_d \delta v_k(t). \quad (37)$$

Уравнения (35) и (37) задают полное описание динамики относительно повторений. Зададим обновляющую поправку в виде

$$\delta v_k(t) = k_1^T \eta_k(t) + k_2 e_{k-1}(t) \quad (38)$$

и запишем полную модель в приращениях:

$$\begin{aligned} \dot{\eta}_k(t) &= (A + b\theta^T b_d k_1^T) \eta_k(t) + b\theta^T A_d \check{\eta}_{k-1}(t) + b\theta^T b_d k_2 e_{k-1}(t), \\ \dot{\check{\eta}}_k(t) &= A_d \check{\eta}_{k-1}(t) + b_d k_1 \eta_k(t) + b_d k_2 e_{k-1}(t), \end{aligned} \quad (39)$$

$$e_k(t) = -c^T (A + b\theta^T b_d k_1^T) \eta_k(t) - c^T b \theta^T A_d \check{\eta}_{k-1}(t) + (1 - c^T b \theta^T b_d k_2) e_{k-1}(t).$$

Модель (39) записана в форме дифференциального повторяющегося процесса, относящегося к классу 2D систем. Как и в предыдущем разделе дальнейший анализ сходимости ошибки обучения будет основан на теории устойчивости повторяющихся процессов с использованием векторных функций Ляпунова из [27].

Введем в рассмотрение вектор  $\bar{e}_k(t) = [\check{\eta}_k^T(t) e_{k-1}(t)]^T$  и векторную функцию Ляпунова

$$V(\eta_k(t), \bar{e}_k(t)) = \begin{bmatrix} V_1(\eta_k(t)) \\ V_2(\bar{e}_k(t)) \end{bmatrix}, \quad (40)$$

где  $V_1(\eta) > 0$ ,  $\eta \neq 0$ ,  $V_2(\bar{e}) > 0$ ,  $\bar{e} \neq 0$ ,  $V_1(0) = 0$ ,  $V_2(0) = 0$ .

Аналог оператора дивергенции этой функции вдоль траекторий системы (39) имеет вид

$$\mathcal{D}V(\eta_k(t), \bar{e}_k(t)) = \frac{dV_1(\eta_k(t))}{dt} + \Delta_k V_2(\bar{e}_k(t)), \quad (41)$$

где  $\Delta_k V_2(\bar{e}_k(t)) = V_2(\bar{e}_{k+1}(t)) - V_2(\bar{e}_k(t))$ .

**Теорема 3.** *Предположим, что существует векторная функция Ляпунова (40) и положительные скаляры  $c_1 - c_4$  такие, что на траекториях системы (39) выполняются неравенства*

$$\begin{aligned} c_1 \|\eta_k(t)\|^2 &\leq V_1(\eta_k(t)) \leq c_2 \|\eta_k(t)\|^2, \\ c_1 \|\bar{e}_k(t)\|^2 &\leq V_2(e_k(t)) \leq c_2 \|\bar{e}_k(t)\|^2, \\ \mathcal{D}V(\eta_k(t), \bar{e}_k(t)) &\leq -c_3 (\|\eta_k(t)\|^2 + \|\bar{e}_k(t)\|^2), \\ \frac{\partial V_1(\eta)}{\partial \eta} &\leq -c_4 \|\eta\|. \end{aligned} \tag{42}$$

Тогда алгоритм управления с итеративным обучением (29), (30) с обновляющей поправкой (38) гарантирует выполнение условий сходимости ошибки обучения (32), (33).

**Доказательство.** В соответствии с теоремой 2 из [27] при условиях (42) существуют  $\alpha > 0$ ,  $\lambda > 0$  и  $0 < \zeta < 1$  такие, что

$$\|\eta_k(t)\|^2 + \|\bar{e}_k(t)\|^2 \leq \alpha \exp(-\lambda t) \zeta^k \leq \alpha \zeta^k. \tag{43}$$

Следовательно,  $\|\eta_k(t)\| \leq \kappa \varrho^k$ , где  $\kappa = \sqrt{\alpha}$ ,  $\varrho = \sqrt{\zeta}$  и, принимая во внимание определение  $\bar{e}$ ,  $|e_k(p)| \leq \kappa \varrho^k$ . Таким образом, справедливо (32). Далее, поскольку  $\delta v_k(p)$  определено соотношением (38), то, учитывая (43), по аналогии с доказательством теоремы 1 нетрудно получить оценку

$$|\delta v_{k+1}(t)| \leq \alpha_0 \exp\left(-\frac{1}{2}\lambda t\right) \zeta^{k/2}. \tag{44}$$

Из (43) следует

$$\|\eta_k(t)\| \leq \alpha \exp\left(-\frac{1}{2}\lambda t\right) \zeta^{k/2}, \quad \|\dot{\eta}_k(t)\| \leq \alpha \exp\left(-\frac{1}{2}\lambda t\right) \zeta^{k/2}, \quad |e_k(t)| \leq \alpha \exp\left(-\frac{1}{2}\lambda t\right) \zeta^{k/2}. \tag{45}$$

Из (36) с учетом (44) и (45) получим

$$\|\dot{\eta}_k(t)\| \leq \alpha_1 \exp\left(-\frac{1}{2}\lambda t\right) \zeta^{(k-1)/2},$$

где  $\alpha_1 = \alpha \max\{\|A\|, \|b\theta^T A_d\|, \|b\theta^T b_d\|\}$ , и

$$|\dot{e}_k(t)| = |c^T \dot{\eta}_k(t)| \leq \alpha_2 \exp\left(-\frac{1}{2}\lambda t\right) \zeta^{(k-1)/2},$$

где  $\alpha_2 = \alpha_1 \|c^T\|$ . С учетом двух последних неравенств

$$|\Delta v_{k+1}(t)| = \left| \frac{d}{dt} \delta v_{k+1}(t) \right| = |k_1 \dot{\eta}_k(t) + k_2 \dot{e}_k(t)| \leq \alpha_3 \exp\left(-\frac{1}{2}\lambda t\right) \zeta^{(k-1)/2},$$

где  $\alpha_3 = \alpha_0 \max\{\alpha_1, \alpha_2\}$ . Из (30) с учетом последнего неравенства следует

$$|v_{k+1}(t)| \leq |v_k(t)| + \alpha_3 \exp\left(-\frac{1}{2}\lambda t\right) \zeta^{(k-1)/2}.$$

Условие (33) теперь следует из повторения, применительно к последнему неравенству, выводов, аналогичных сделанных в доказательстве теоремы 1 после неравенства (22) с учетом (29).

### 3.3. Построение алгоритма с использованием техники линейных матричных неравенств

Введем расширенный вектор  $\xi_k(t) = [\eta_k(t)^T \ \bar{e}_k(t)^T]^T$ , аналогичный введенному в предыдущем разделе для дискретных систем. Выбирая компоненты векторной функции Ляпунова (40) в виде квадратичных форм

$$V_1(\eta_k(t)) = \eta_k^T(t) P_1 \eta_k(t), \quad V_2(\bar{e}_k(t)) = \bar{e}_k^T(t) P_2 \bar{e}_k(t)$$

и вычисляя дивергенцию этой функции вдоль траекторий системы (39), получим

$$\mathcal{D}V(\eta_k(t), \bar{e}_k(t)) = \xi_k^T(t) \left[ \bar{A}_c^T I^{(1,0)} \bar{P} + \bar{P} I^{(1,0)} \bar{A}_c + \bar{A}_c^T I^{(0,1)} \bar{P} \bar{A}_c - I^{(0,1)} \bar{P} \right] \xi_k(t), \tag{46}$$

где

$$P_1 = P_1^T \succ 0, P_2 = P_2^T \succ 0, \bar{A}_c = (\bar{A} + \bar{b}k^T \bar{H}), \bar{P} = \text{diag}[P_1 \ P_2],$$

$$P_2 = \text{diag}[P_{21} \ P_{22}], k^T = [k_1^T \ k_2], \bar{H} = \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I^{(1,0)} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, I^{(0,1)} = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}$$

и размеры блоков матриц  $\bar{P}, \bar{H}, I^{(1,0)}, I^{(0,1)}$  согласованы с размерностями переменных  $\eta, \check{\eta}, \bar{e}$ . Проводя аналогию с теорией линейно-квадратичного регулятора (ЛКР), потребуем, чтобы выполнялось неравенство

$$\mathcal{D}V(\eta_k(t), \bar{e}_k(t)) + \xi_k^T(t)[\bar{Q} + (k^T \bar{H})^T R k^T \bar{H}] \xi_k(t) \leq 0, \quad (47)$$

где  $\bar{Q} = \text{diag}[Q_1 \ Q_2] \succ 0, R \succ 0$  — весовые матрицы, тогда нетрудно видеть, что все условия теоремы 3 будут выполнены. Неравенство (47) эквивалентно матричному неравенству

$$\bar{A}_c^T I^{(1,0)} \bar{P} + \bar{P} I^{(1,0)} \bar{A}_c + \bar{A}_c^T I^{(0,1)} \bar{P} \bar{A}_c - I^{(0,1)} \bar{P} + \bar{Q} + (k^T \bar{H})^T R k^T \bar{H} \preceq 0. \quad (48)$$

Введем переменные  $\bar{X} = \bar{P}^{-1}$  и  $Z$  как решение уравнения

$$\bar{H}X = Z\bar{H}. \quad (49)$$

Тогда, применяя к (48) лемму о дополнении Шура, получим

$$\begin{bmatrix} M_{11} & M_{12} & \bar{X} & (Y^T \bar{H})^T \\ M_{12}^T & -\bar{X} & 0 & 0 \\ \bar{X} & 0 & -Q^{-1} & 0 \\ Y^T \bar{H} & 0 & 0 & -R^{-1} \end{bmatrix} \preceq 0,$$

где

$$M_{11} = (\bar{A}\bar{X} + \bar{b}Y^T \bar{H})^T I^{(1,0)} + I^{(1,0)} (\bar{A}\bar{X} + \bar{b}Y^T \bar{H}) - I^{(0,1)} \bar{X}, M_{12} = (\bar{A}\bar{X} + \bar{b}Y^T \bar{H}) I^{(0,1)}.$$

Если система матричных уравнений и неравенств (49), (50) разрешима относительно  $X \succ 0, Y$  и  $Z$ , то

$$k^T = Y^T Z^{-1}. \quad (50)$$

Сделанные выводы можно объединить в виде следующей теоремы.

**Теорема 4.** *Предположим, что для некоторого выбора весовой матрицы  $Q \succ 0$ , вектора  $\theta$  и скаляра  $R > 0$  система линейных матричных уравнений и неравенств (49), (50) разрешима относительно  $X \succ 0, Y$  и  $Z$ , а вектор  $k$  определяется по формуле (50). Тогда алгоритм управления с итеративным обучением (29), (30) с обновляющей поправкой (38) обеспечивает выполнение условий сходимости ошибки обучения (32), (33).*

#### 4. ПРИМЕР

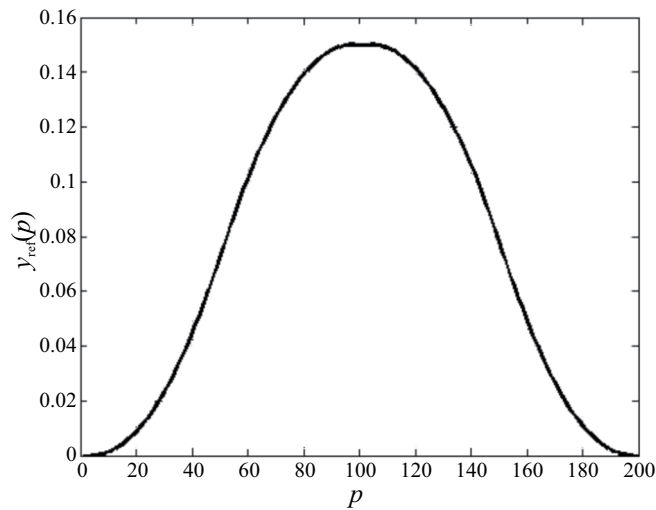
Рассмотрим управляемое движение манипулятора по горизонтальной оси, перпендикулярной направлению движения ленты конвейера многоосевого порталного робота. Модель динамики, полученная экспериментально на основе частотного анализа [31], задается передаточной функцией

$$G(s) = \frac{23.736(s + 661.2)}{s(s^2 + 426.7s + 1.744 \times 10^5)}. \quad (51)$$

Желаемая (эталонная) траектория движения продолжительностью 2 с представлена на фиг. 1. Для построения дискретной модели в пространстве состояний используем стандартные функции MATLAB, принимая период дискретности равным 0.01 с.

Рассмотрим сначала случай  $d = 1$ , при этом согласно (2), (3), (14), как в методе тяжелого шарика, на текущем шаге используется информация с двух предыдущих шагов. Введем в рассмотрение среднеквадратическую ошибку (СКО) обучения

$$E(k) = \sqrt{\frac{1}{N} \sum_{p=0}^N |e_k(p)|^2}, \quad (52)$$



Фиг. 1. Желаемая траектория движения.

и сравнение скорости сходимости алгоритмов будем проводить по числу шагов, при котором  $E(k)$  уменьшится в 10 раз относительно начального значения. Матрица  $Q$  и положительные скаляры  $R$ ,  $\sigma$  были выбраны, исходя из принципов теории ЛКР:

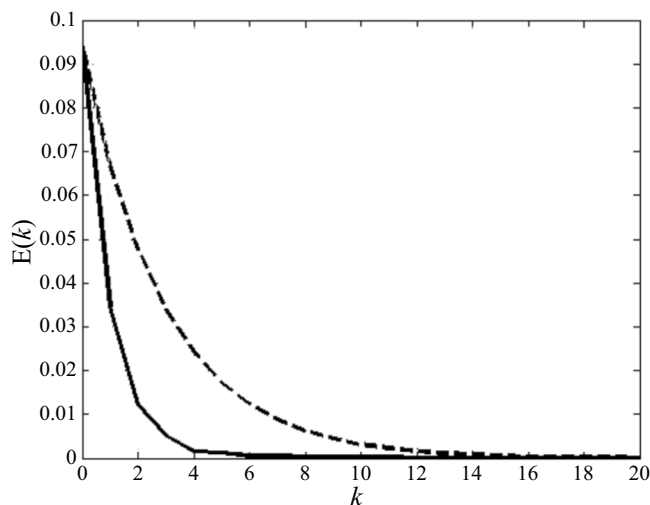
$$Q = \text{diag}[1 \ 1 \ 1 \ 10 \ 5 \times 10^5], \quad R = 10^{-3}, \quad \sigma = 0.0125.$$

Отметим, что выбор этих параметров представляет отдельную нетривиальную задачу, для более детального ознакомления с которой заинтересованный читатель может обратиться к [32]–[35]. В данном случае были использованы асимптотические свойства ЛКР (см. [35]), в соответствии с которыми малый вес  $R$  не ограничивает энергию управления, а большой вес при ошибке обучения в матрице  $Q$  придает доминирующий характер этой переменной при формировании управления.

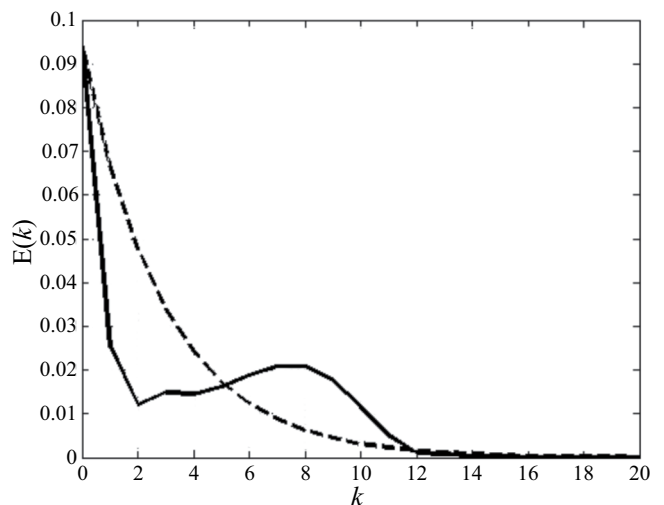
При  $\tau_0 = 1$  и  $\tau_1 = 0$  получаем стандартный алгоритм первого порядка, для которого

$$k_1^T = [-15.5 \ -12.7 \ -5180.4], \quad k_2 = 194.6.$$

Для алгоритмов второго порядка при  $\tau_0 = 1$  естественно задавать  $\tau_1$ , не превышающим 1, иначе обновленная информация будет подавляться устаревшей. При  $0 < \tau_1 < 1$  получаем ускорение сходимости относительно алгоритма первого порядка. При  $\tau_1 > 1$  в силу преобладания устаревшей информации происходит нарушение



Фиг. 2. Изменение СКО обучения в случаях алгоритма первого порядка (штриховая линия) и второго порядка (сплошная линия) при  $\tau_1 = 0.8$ .



Фиг. 3. Изменение СКО обучения в случаях алгоритма первого порядка (штриховая линия) и второго порядка (сплошная линия) при  $\tau_1 = 1.2$ .

монотонности изменения  $E(k)$  и замедление скорости сходимости относительно алгоритма первого порядка. На фиг. 2 представлено сравнение алгоритма первого порядка и алгоритма второго порядка при  $\tau_1 = 0.8$ , при этом

$$k_1^T = [-7.2 \ -5.6 \ -2296.3], \quad k_2 = 190.2.$$

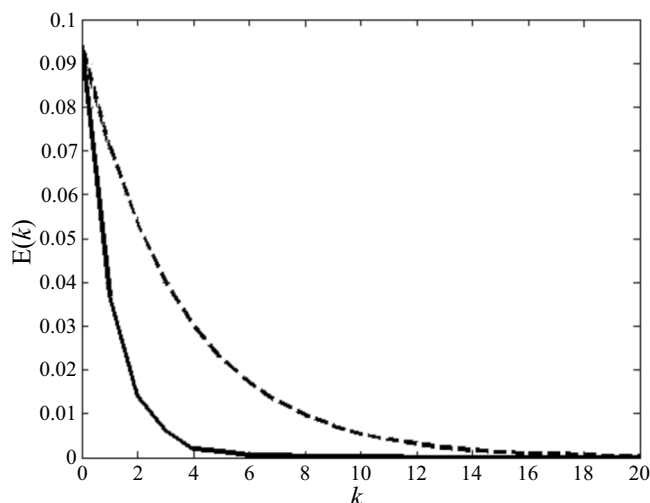
На фиг. 3 представлено то же сравнение при  $\tau_1 = 1.2$ , в этом случае

$$k_1^T = [-2.27 \ -1.74 \ -714.95], \quad k_2 = 75.72.$$

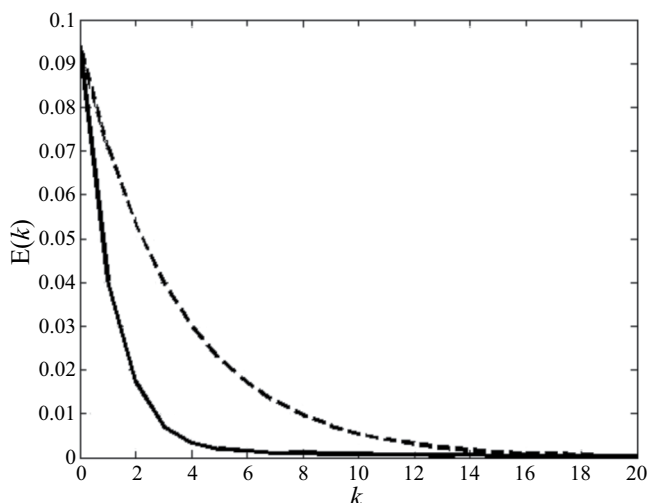
Рассмотрим случай  $d = 2$ . При этом согласно (2), (3), (14) на текущем шаге учитывается информация с трех предыдущих шагов. Сохраним все параметры теми же, что и при  $d = 1$ . Единственное отличие будет в естественном увеличении размера среднего блока матрицы  $Q$ , в результате она примет вид

$$Q = \text{diag}[1 \ 1 \ 1 \ 10 \ 10 \ 10 \ 5 \times 10^5].$$

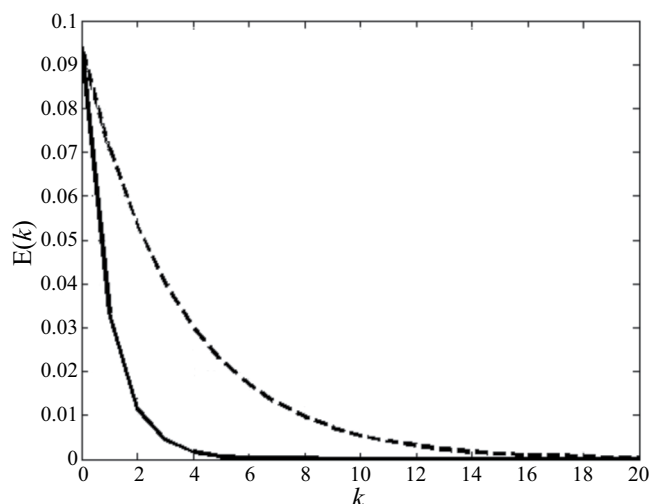
Расчеты и моделирование показали, что при  $\tau_0 = 1$  обнуление одного из коэффициентов  $\tau_1$  или  $\tau_2$  при изменении второго в интервале  $(0, 1)$  дает практически то же самое ускорение сходимости ошибки обучения, что и при  $d = 1$  (фиг. 4 и фиг. 5).



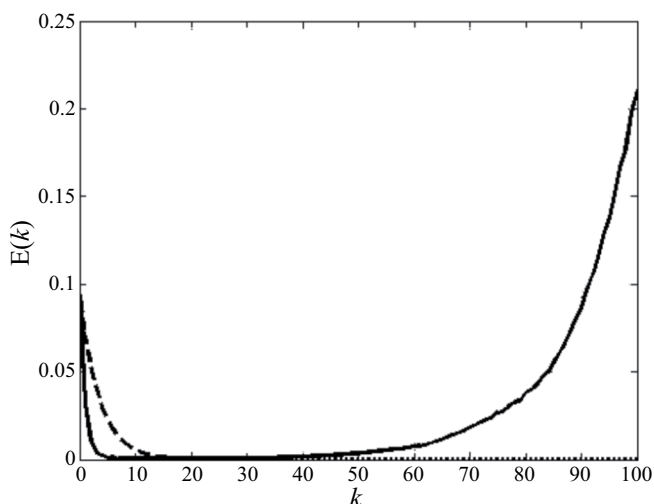
Фиг. 4. Изменение СКО обучения в случаях алгоритма первого порядка (штриховая линия) и третьего порядка (сплошная линия) при  $\tau_1 = 0.8$ ,  $\tau_2 = 0$ .



Фиг. 5. Изменение СКО обучения в случаях алгоритма первого порядка (штриховая линия) и третьего порядка (сплошная линия) при  $\tau_1 = 0$ ,  $\tau_2 = 0.8$ .



Фиг. 6. Изменение СКО обучения в случаях алгоритма первого порядка (штриховая линия) и третьего порядка (сплошная линия) при  $\tau_1 = 0.8$ ,  $\tau_2 = 0.4$ .



Фиг. 7. Изменение СКО обучения в случаях алгоритма первого порядка (штриховая линия) и третьего порядка (сплошная линия) при  $\tau_1 = 0.6$ ,  $\tau_2 = 0.7$ .

Близкая картина наблюдается, если оба коэффициента не равны нулю и  $\tau_2 < \tau_1$  (фиг. 6), но при этом, начиная с некоторого значения  $\tau_1$  ошибка обучения начинает расходиться.

При  $\tau_2 > \tau_1$  ошибка всегда расходится (фиг. 7), что объясняется противоречивым характером информации, поступающей в алгоритм.

## 5. ЗАКЛЮЧЕНИЕ

В статье предложены новые алгоритмы управления с итеративным обучением высшего порядка, отличающиеся от немногих известных тем, что они учитывают полную информацию о предыдущих шагах и по структуре аналогичны алгоритмам многошаговых методов в теории оптимизации. Примеры подтверждают ускорение сходимости ошибки обучения относительно алгоритмов первого порядка. В дальнейших исследованиях предполагается оценить робастность и эффективность предложенных алгоритмов в условиях неопределенности модели и с учетом внешних случайных возмущений и шумов измерений. Остается также открытым вопрос конструктивного выбора глубины памяти (параметр  $d$ ) и весовых коэффициентов ( $\tau_1, \dots, \tau_d$ ). Отдельным направлением является распространение полученных результатов на нелинейные системы.

Наконец, наиболее трудным является строгое общее теоретическое обоснование возможных преимуществ алгоритмов УИО высшего порядка в смысле скорости сходимости ошибки обучения. Хотя примеры вселяют определенную надежду, в общем случае эта задача пока остается открытой даже для линейных систем.

Идея этой работы появилась под влиянием докладов на традиционной молодежной летней школе им. Б. Т. Поляка (Нижний Новгород, 10—15 июля 2023 г.). Авторы выражают благодарность А. В. Гасникову за конструктивную дискуссию и указание на интересную работу [12].

## СПИСОК ЛИТЕРАТУРЫ

1. Arimoto S., Kawamura S., Miyazaki F. Bettering operation of robots by learning // *J. Robot. Syst.* 1984. V. 1. P. 123—140.
2. Bristow D. A., Tharayil M., Alleyne A. G. A survey of iterative learning control: A learning-based method for high-performance tracking control // *IEEE Control Syst. Magaz.* 2006. V. 26. № 3. P. 96—114.
3. Ahn H.-S., Chen Y. Q., Moore K. L. Iterative learning control: Survey and categorization // *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* 2007. V. 37. № 6. P. 1099—1121.
4. Rogers E., Chu B., Freeman C., Lewin P. *Iterative learning control algorithms and experimental benchmarking* Chichester: John Wiley & Sons, 2023.
5. Lim I., Hoelzle D. J., Barton K. L. A multi-objective iterative learning control approach for additive manufacturing applications // *Control Engineer. Practice.* 2017. V. 64. P. 74—87.
6. Sammons P. M., Gegel M. L., Bristow D. A., Landers R. G. Repetitive process control of additive manufacturing with application to laser metal deposition // *IEEE Transact. Control Syst. Technol.* 2019. V. 27. № 2. P. 566—575.
7. Freeman C. T., Rogers E., Hughes A.-M., Burr ridge J. H., Meadmore K. L. Iterative learning control in health care: electrical stimulation and robotic-assisted upper-limb stroke rehabilitation // *IEEE Control Syst. Magaz.* 2012. V. 47. P. 70—80.
8. Meadmore K. L., Exell T. A., Hallewell E., Hughes A.-M., Freeman C. T., Kutlu M., Benson V., Rogers E., Burr ridge J. H. The application of precisely controlled functional electrical stimulation to the shoulder, elbow and wrist for upper limb stroke rehabilitation: a feasibility study // *J. of NeuroEngineer. and Rehabilitation.* 2014. P. 11—105.
9. Ketelhut M., Stemmler S., Gesenhues J., Hein M., Abel D. Iterative learning control of ventricular assist devices with variable cycle durations // *Control Engineer. Practice.* 2019. V. 83. P. 33—44.
10. Поляк Б. Т. О некоторых способах ускорения сходимости итерационных методов // *Ж. вычисл. матем. и матем. физ.* 1964. Т. 4. № 5. С. 791—803.
11. Поляк Б. Т. *Введение в оптимизацию.* М.: Наука, 1983.
12. d'Aspremont A., Scieur D., Taylor A. *Acceleration Methods* // arXiv:2101.09545v3. 2021.
13. Ahn H.-S., Moore K. L., Cheh Y. *Iterative Learning Control: Robustness and Monotonic Convergence for Interval Systems.* London: Springer-Verlag, 2007.
14. Bien Z., Huh K. M. Higher-order iterative learning control algorithm // *IEE Proc. D-Control Theory Appl.* 1989. V. 136. P. 105—112.

15. Chen Y., Gong Z., Wen C. Analysis of a high-order iterative learning control algorithm for uncertain nonlinear systems with state delays // *Automatica*. 1998. V. 34. P. 345—353.
16. Norrlof M., Gunnarsson S. A frequency domain analysis of a second order iterative learning control algorithm // *Proc. 38th IEEE Conf. Decis. Control*. 1999. V. 2. P. 1587—1592.
17. Bu X., Yu F., Fu Z., Wang F. Stability analysis of high-order iterative learning control for a class of nonlinear switched systems // *Abstract Appl. Anal.* 2013. V. 2013. P. 1—13.
18. Wei Y.-S., Li X.-D. Robust higher-order ILC for non-linear discrete-time systems with varying trail lengths and random initial state shifts // *IET Control Theory Appl.* 2017. V. 11. P. 2440—2447.
19. Wang X., Chu B., Rogers E. Higher-order Iterative Learning Control Law Design using Linear Repetitive Process Theory: Convergence and Robustness // *IFAC PapersOnLine*. 2017. V. 50—1. P. 3123—3128.
20. Phan M. Q., Longman R. W. Higher-order iterative learning control by pole placement and noise filtering // *IFAC Proc. Volumes*. 2002. V. 35. P. 25—30.
21. Gu P., Tian S., Chen Y. Iterative learning control based on Nesterov accelerated gradient method // *IEEE Access*. 2019. V. 7. P. 115 836—115 842.
22. Нестеров Ю.Е. Метод решения задачи выпуклого программирования со скоростью сходимости  $O\left(\frac{1}{k^2}\right)$  // Докл. АН СССР. 1983. Т. 269. № 3. С. 543—547.
23. Kim D., Fessler J. A. Generalizing the optimized gradient method for smooth convex minimization // *arXiv:1607.06764*. 2018. P. 1—26.
24. Moore K. L. An iterative learning control algorithm for systems with measurement noise // *Proc. 38th IEEE Conf. Decis. Control*. 1999. V. 1. P. 270—275.
25. Saab S. S. Optimality of first-order ILC among higher order ILC // *IEEE Transact. on Automatic Control*. 2006. V. 51. P. 1332—1336.
26. Afkhami Z., Hoelzle D. J., Barton K. Robust higher-order spatial iterative learning control for additive manufacturing systems // *IEEE Transact. on Control Systems Technology*. 2023. V. 31. P. 1692—1707.
27. Pakshin P., Emelianova J., Emelianov M., Gałkowski K., Rogers E. Dissipativity and stabilization of nonlinear repetitive processes // *Systems & Control Lett.* 2016. V. 91 P. 14—20.
28. Цыпкин Я. З. Адаптация и обучение в автоматических системах. М.: Наука, 1968.
29. Rogers E., Gałkowski K., Owens D. H. Control systems theory and applications for linear repetitive processes. Berlin: Springer-Verlag, 2007.
30. Hładowski Ł., Gałkowski K., Cai Z., Rogers E., Freeman C., Lewin P. Experimentally supported 2D systems based iterative learning control law design for error convergence and performance // *Control Engineer. Practice*. 2010. V. 18. P. 339—348.
31. Hładowski Ł., Gałkowski K., Cai Z., Rogers E., Freeman C., Lewin P. Experimentally supported 2D systems based iterative learning control law design for error convergence and performance // *Control Engineer. Practice*. 2010. V. 18. P. 339—348.
32. Solheim O. A. Design of optimal control systems with prescribed eigenvalues // *Int. J. Control*. 1972. V. 15. № 1. P. 143—160.
33. Кухаренко Н. В. Определение коэффициентов квадратичных функционалов в задачах аналитического конструирования // *Технич. кибернетика*. 1977. № 4. С. 197—201.
34. Богачев А. В., Григорьев В. В., Дроздов В. Н., Коровьяков А. Н. Аналитическое регуляторов по корневым показателям // *Автоматика и телемехан.* 1979. № 8. С. 21—28.
35. Квакернаак Х., Сиван Р. Линейные оптимальные системы управления. М.: Мир, 1977.

## HIGHER-ORDER ITERATIVE LEARNING CONTROL ALGORITHMS FOR LINEAR SYSTEMS

P. V. Pakshin<sup>a,\*</sup>, J. P. Emelianova<sup>a,\*\*</sup>, M. A. Emelianov<sup>a,\*\*\*</sup>

<sup>a</sup> *Arzamas Polytechnic Institute of Nizhny Novgorod State Technical University n.a. R.E. Alekseev, Kalinina str., 19, Arzamas, 607227 Russia*

\* *e-mail: pakshinpv@gmail.com*

\*\* *e-mail: emelianovajulia@gmail.com*

\*\*\* *e-mail: mikhailemelianovarzamas@gmail.com*

Received 06 November, 2023

Revised 18 November, 2023

Accepted 20 November, 2023

**Abstract.** Iterative learning control (ILC) algorithms appeared in connection with the problems of increasing the accuracy of performing repetitive operations by robots. They use information from previous repetitions to adjust the control signal on the current repetition. Most often, information from the previous repetition only is used. ILC algorithms that use information from several previous iterations are called higher-order algorithms. Recently, interest in these algorithms has increased in the literature in connection with robotic additive manufacturing problems. However, in addition to the fact that these algorithms have been little studied, there are conflicting estimates regarding their properties. This paper proposes new higher-order ILC algorithms for linear discrete and differential systems. The idea of these algorithms is based on an analogy with multi-step methods in optimization theory, in particular, with the heavy ball method. An example is given that confirms the possibility to accelerate convergence of the learning error when using such algorithms.

**Keywords:** iterative learning control, linear systems, higher-order algorithms, repetitive, processes, convergence, stability, vector Lyapunov function, linear matrix inequalities.



*Посвящаем эту работу  
Федору Павловичу Васильеву —  
тонкому, глубокому математику  
и чуткому человеку,  
постоянно проявлявшему интерес  
к нашим исследованиям  
и поддерживавшему их*

## ОПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТА ТЕПЛОПРОВОДНОСТИ И ОБЪЕМНОЙ ТЕПЛОЕМКОСТИ ВЕЩЕСТВА ПО ТЕПЛОВОМУ ПОТОКУ<sup>1)</sup>

© 2024 г. А.Ю. Горчаков<sup>1</sup>, В.И. Зубов<sup>1,\*</sup>

<sup>1</sup>119333 Москва, ул. Вавилова, 44, ФИЦ ИУ РАН, Россия

\*e-mail: vladimir.zubov@mail.ru

Поступила в редакцию 22.09.2023 г.

Переработанный вариант 04.11.2023 г.

Принята к публикации 20.11.2023 г.

Изучение нелинейных проблем, связанных с процессом теплопередачи в веществе, очень важно для практики. Ранее Ю.А. Горчаковым и В.И. Зубовым был предложен эффективный алгоритм определения объемной теплоемкости и коэффициента теплопроводности вещества на основе результатов экспериментального наблюдения за динамикой температурного поля в объекте. В данной работе исследуется задача одновременной идентификации зависящих от температуры объемной теплоемкости и коэффициента теплопроводности исследуемого вещества по тепловому потоку на границе области. Рассмотрение осуществляется на основе первой краевой задачи для одномерного нестационарного уравнения теплопроводности. Рассматриваемая обратная коэффициентная задача сводится к вариационной задаче, которая решается градиентными методами, основанными на применении методологии быстрого автоматического дифференцирования. Исследуется вопрос единственности решения обратной задачи. Библ. 27. Фиг. 12.

**Ключевые слова:** теплопроводность, обратные коэффициентные задачи, градиент, уравнение теплопроводности, тепловой поток.

**DOI:** 10.31857/S0044466924040067, **EDN:** ZJTAJT

### ВВЕДЕНИЕ

Классическое уравнение теплопроводности часто используют для описания и математического моделирования разнообразных тепловых процессов. При этом входящие в уравнение теплопроводности плотность вещества, его удельная теплоемкость и коэффициент теплопроводности считаются известными функциями координат и температуры. Задание дополнительных краевых условий позволяет определить динамику температурного поля в изучаемом объекте.

Однако не всегда свойства вещества бывают известны. Так, при создании новых материалов приходится встречаться с ситуацией, когда объемная теплоемкость и коэффициент теплопроводности вещества зависят только от температуры, и эта зависимость неизвестна. В связи с этим возникает задача определения зависимости объемной теплоемкости и коэффициента теплопроводности вещества от температуры по результатам экспериментального наблюдения за динамикой температурного поля. К такой же задаче приходят и в том случае, когда хотят описать сложный тепловой процесс некоторой упрощенной математической моделью. Примером может служить математическое моделирование процессов распространения

<sup>1)</sup> Работа выполнялась с использованием инфраструктуры Центра коллективного пользования «Высокопроизводительные вычисления и большие данные» (ЦКП «Информатика») ФИЦ ИУ РАН (г. Москва).

тепла в сложных пористых композиционных материалах, где заметную роль играет радиационная теплопередача (см. [1], [2]).

Задача определения только коэффициента теплопроводности вещества рассматривается довольно давно. Подтверждением этому может служить большое количество публикаций, посвященных указанному вопросу (см., например, [3]—[11]). Большое внимание в этих работах уделяется не только теоретическому исследованию обратных коэффициентных задач, но и разработке численных методов их решения. Что касается одновременного определения объемной теплоемкости и коэффициента теплопроводности вещества, то работ, посвященных этому вопросу, немного (см., например, [12]—[14]).

Все указанные выше работы посвящены решению одной и той же задачи — определению теплофизических параметров исследуемого вещества. Однако используемые математические постановки этой общей задачи в разных работах разные. Разные и методы, применяемые для решения этой задачи.

В настоящей работе рассматривается одна из возможных постановок такой обратной коэффициентной задачи. Ее рассмотрение проводится на основе первой краевой задачи для одномерного нестационарного уравнения теплопроводности. Обратная коэффициентная задача сводится к вариационной задаче. В качестве целевого функционала выбрано среднеквадратичное отклонение рассчитанного теплового потока на границе образца от его экспериментального значения. Предложен алгоритм численного решения обратной коэффициентной задачи. В основе этого алгоритма лежит современная методология быстрого автоматического дифференцирования (БАД-методология, см. [15], [16]), позволившая успешно решить ряд сложных задач оптимального управления динамическими системами (см., например, [17]—[19]). Приведенные в работе примеры решения обратной коэффициентной задачи демонстрируют работоспособность предложенного алгоритма. В статье, кроме того, исследуется вопрос единственности получаемого решения сформулированной обратной задачи.

### 1. ПОСТАНОВКА ЗАДАЧИ

Рассматривается слой материала ширины  $L$ . Температура этого слоя в начальный момент времени известна. Также известно, как изменяется со временем температура на краях этого слоя. Распределение температурного поля в слое в каждый момент времени описывается решением следующей начально-краевой (смешанной) задачи:

$$C(T) \frac{\partial T(x,t)}{\partial t} - \frac{\partial}{\partial x} \left( K(T) \frac{\partial T(x,t)}{\partial x} \right) = 0, \quad (x,t) \in Q, \tag{1.1}$$

$$T(x,0) = w_0(x), \quad (0 \leq x \leq L), \tag{1.2}$$

$$T(0,t) = w_1(t), \quad T(L,t) = w_2(t) \quad (0 \leq t \leq \Theta). \tag{1.3}$$

Здесь  $x$  — декартова координата точки в слое;  $t$  — время;  $Q = \{ (0 < x < L) \times (0 < t < \Theta) \}$ ;  $T(x,t)$  — температура вещества в точке с координатой  $x$  в момент времени  $t$ ;  $C(T)$  — объемная теплоемкость вещества,  $K(T)$  — коэффициент теплопроводности;  $w_0(x)$  — заданная температура слоя в начальный момент времени;  $w_1(t)$  — заданная температура на левом краю слоя;  $w_2(t)$  — заданная температура на правом краю слоя.

Если зависимость объемной теплоемкости вещества  $C(T)$  и его коэффициента теплопроводности  $K(T)$  от температуры  $T$  известна, то, решая смешанную задачу (1.1) — (1.3), найдем распределение температуры  $T(x,t)$  в  $Q$ . Задачу (1.1)–(1.3) ниже будем называть *прямой* задачей.

Если же зависимость объемной теплоемкости вещества и его коэффициента теплопроводности от температуры неизвестна, то представляет большой интерес задача определения этих зависимостей. Одной из возможных постановок такой задачи (эта задача относится к классу задач идентификации параметров модели) является следующая: найти такие зависимости объемной теплоемкости вещества  $C(T)$  и коэффициента теплопроводности  $K(T)$  от температуры  $T$ , при которых тепловые потоки  $\Pi_0(t) = K(T(0,t)) \cdot \frac{\partial T}{\partial x}(0,t)$  на левом конце и  $\Pi_L(t) = -K(T(L,t)) \cdot \frac{\partial T}{\partial x}(L,t)$  на правом конце, полученные в результате решения сформулированной задачи (1.1)–(1.3), мало отличаются от потоков  $P_0(t)$  и  $P_L(t)$  соответственно, полученных экспериментально. Мерой отклонения этих функций может служить величина (функционал «поток»):

$$\Phi(C(T), K(T)) = \int_0^\Theta \left[ \beta_0(t) \cdot (\Pi_0(t) - P_0(t))^2 + \beta_L(t) \cdot (\Pi_L(t) - P_L(t))^2 \right] dt, \tag{1.4}$$

где  $\beta_0(t) \geq 0$ ,  $\beta_L(t) \geq 0$  — заданные весовые параметры;  $P_0(t)$  и  $P_L(t)$  — заданные тепловые потоки на концах слоя. Таким образом, задача оптимального управления состоит в определении оптимального

управления  $\{C(T), K(T)\}$  и соответствующего оптимального решения  $T(x, t)$  задачи (1.1)—(1.3), при котором функционал (1.4) достигает минимального значения.

Следует отметить, что задача идентификации объемной теплоемкости  $C(T)$  и коэффициента теплопроводности  $K(T)$ , подобная описанной выше, рассматривалась в работе [20], только там в качестве целевого функционала выбирался функционал «поле»:

$$\Phi(C(T), K(T)) = \int_0^{\Theta} \int_0^L [T(x, t) - Y(x, t)]^2 \cdot \mu(x, t) dx dt, \quad (1.5)$$

где  $\mu(x, t) \geq 0$  — заданная весовая функция,  $T(x, t)$  — температурное поле, полученное в результате решения сформулированной задачи (1.1)—(1.3),  $Y(x, t)$  — температурное поле, полученное экспериментально. В [20] отмечено, что при использовании функционала «поле» решение задачи идентификации всегда неединственно и что целесообразно определять только отношение  $\chi(T) = C(T) / K(T)$ . В случае использования функционала «поток» (1.4) это утверждение оказывается несправедливым.

Получить аналитическое решение сформулированной обратной задачи (1.1)—(1.4) удается крайне редко. Поэтому необходимо разработать алгоритм численного решения сформулированной обратной задачи.

## 2. ЗАДАЧА ОПТИМАЛЬНОГО УПРАВЛЕНИЯ В ДИСКРЕТНОМ СЛУЧАЕ

При численном решении сформулированной задачи оптимального управления искомые функции  $C(T)$  и  $K(T)$ ,  $T \in [a, b]$  аппроксимировались непрерывными кусочно-линейными функциями следующим образом. Отрезок  $[a, b]$ , на котором идентифицируются функции  $C(T)$  и  $K(T)$ , разбивался точками  $\tilde{T}_0 = a, \tilde{T}_1, \tilde{T}_2, \dots, \tilde{T}_N = b$  на  $N$  частей (они могут быть как равными, так и неравными). Каждой из точек  $\tilde{T}_n$  ( $n = 0, \dots, N$ ) ставились в соответствие числа  $c_n = C(\tilde{T}_n)$ ,  $k_n = K(\tilde{T}_n)$ . Искомые функции  $C(T)$  и  $K(T)$  аппроксимировались непрерывными кусочно-линейными функциями с узлами в точках  $\{(\tilde{T}_n, c_n)\}_{n=0}^N$  и  $\{(\tilde{T}_n, k_n)\}_{n=0}^N$  так, что

$$C(T) = c_n + \frac{c_{n+1} - c_n}{\tilde{T}_{n+1} - \tilde{T}_n} (T - \tilde{T}_n) \quad \text{при } \tilde{T}_n \leq T \leq \tilde{T}_{n+1} \quad (n = 0, \dots, N - 1),$$

$$K(T) = k_n + \frac{k_{n+1} - k_n}{\tilde{T}_{n+1} - \tilde{T}_n} (T - \tilde{T}_n) \quad \text{при } \tilde{T}_n \leq T \leq \tilde{T}_{n+1} \quad (n = 0, \dots, N - 1).$$

Область  $Q = \{(x, t) : (0 < x < L) \times (0 < t < \Theta)\}$  разбивалась сеточными линиями  $\{\tilde{x}_i\}_{i=0}^I$  и  $\{\tilde{t}^j\}_{j=0}^J$  на ряд прямоугольников с шагами  $h_i = \tilde{x}_{i+1} - \tilde{x}_i$  ( $i = 0, \dots, I - 1$ ),  $\tau^j = \tilde{t}^j - \tilde{t}^{j-1}$  ( $j = 1, \dots, J$ ). В каждом узле  $(\tilde{x}_i, \tilde{t}^j)$  расчетной области  $Q$ , характеризуемом парой индексов  $(i, j)$ , все функции задавались своими значениями в точке  $(\tilde{x}_i, \tilde{t}^j)$  (например,  $T(\tilde{x}_i, \tilde{t}^j) = T_i^j$ ). В каждом из полученных прямоугольников требовалось выполнение закона теплового баланса. В результате получалась хорошо известная конечно-разностная схема, аппроксимирующая смешанную задачу (1.1)—(1.3). Полученная система нелинейных алгебраических уравнений решалась итерационным методом с привлечением метода прогонки [21]. С помощью описанного подхода решалась смешанная задача (1.1)—(1.3) и находилась функция  $T(x, t)$  (точнее, ее аппроксимация  $T_i^j$ ).

Целевой функционал (1.4) аппроксимировался функцией  $F(c_0, c_1, \dots, c_N, k_0, k_1, \dots, k_N)$  конечного числа переменных следующим образом:

$$\Phi(C(T), K(T)) \approx F = \sum_{j=1}^J \left[ \beta_0(t^j) \cdot (\Pi_0(t^j) - P_0(t^j))^2 + \beta_L(t^j) \cdot (\Pi_L(t^j) - P_L(t^j))^2 \right] \cdot \tau^j,$$

где

$$\Pi_0(t^j) = K(T_0^j) \cdot (T_1^j - T_0^j) / h_0,$$

$$\Pi_L(t^j) = -K(T_I^j) \cdot (T_I^j - T_{I-1}^j) / h_{I-1}.$$

Для минимизации целевого функционала в работе использовался градиентный метод. Хорошо известно, что для эффективной работы градиентных методов следует использовать точное значение градиента целевой функции. Вычислить значение градиента эффективно и с машинной точностью позволяет БАД-методология (см. [15], [16], [22]). Именно применение БАД-методологии позволило успешно решить ряд сложных задач оптимального управления. В работах [17]—[19], [22] использовался «ручной» вариант БАД-методологии, при котором все необходимые формулы выводились вручную и использовались в оптимизационном коде. Однако появились стандартные пакеты программ, в автоматическом режиме выполняющие примерно ту же работу (см. [23]—[25]). В статье [26] представлены результаты сравнения градиентов, вычисленных

с помощью ручной версии БАД-методологии, с помощью стандартных пакетов программ, и сделан вывод, что стандартные пакеты программ позволяют практически с той же точностью вычислять градиенты сложных функций, что и ручная версия БАД-методологии. Поэтому в настоящей работе градиент целевой функции вычислялся с помощью пакета прикладного программного обеспечения Adept (см. [24]), при этом использовался обратный (Reverse, Adjoint) метод автоматического дифференцирования.

Для поиска минимума целевой функции применялся метод L-BFGS-B (см. [27]).

### 3. РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ РАСЧЕТОВ

Минимальное значение целевого функционала (1.4), получаемое при решении сформулированной обратной задачи, зависит от заданных экспериментальных потоков  $P_0(t)$  и  $P_L(t)$ . Если экспериментальные потоки принадлежат области достижимости функций  $\Pi_0(t)$  и  $\Pi_L(t)$ , построенных в результате решения прямой задачи (1.1)–(1.3) при допустимых  $C(T)$  и  $K(T)$ , то значение функционала равно нулю. В противном случае значение функционала больше нуля. При проведении численных экспериментов экспериментальные потоки  $P_0(t)$  и  $P_L(t)$  выбирались из области достижимости решений прямой задачи.

Для проверки работоспособности предложенного алгоритма было выполнено большое число тестовых расчетов. Эти расчеты условно можно разделить на две группы.

Все примеры первой группы базировались на том, что функция

$$T_*(x, t) = \left( \frac{n+1}{\beta} \right)^{\frac{1}{n+1}} \cdot [\alpha \cdot (t + \gamma - x)]^{\frac{1}{m-n}}, \quad \alpha = \frac{(m-n)}{(n+1)}, \quad \beta = (n+1)^{\frac{m-n}{m+1}},$$

является аналитическим решением уравнения (1.1) при  $\gamma = \text{const}$ ,  $C(T) = \beta \cdot T^n$ ,  $K(T) = T^m$ .

В соответствии с этим использовались следующие входные данные задачи:

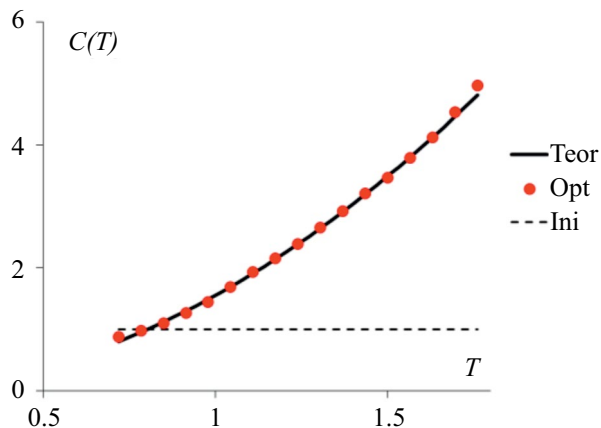
$$\begin{aligned} L &= 1, & \gamma &= 1.5, & Q &= (0, 1) \times (0, \Theta), \\ w_0(x) = T_*(x, 0) &= \left( \frac{n+1}{\beta} \right)^{\frac{1}{n+1}} \cdot [\alpha \cdot (\gamma - x)]^{\frac{1}{m-n}} & (0 \leq x \leq 1), \\ w_2(t) = T_*(1, t) &= \left( \frac{n+1}{\beta} \right)^{\frac{1}{n+1}} \cdot [\alpha \cdot (t + \gamma - 1)]^{\frac{1}{m-n}} & (0 \leq t \leq \Theta), \\ P_0(t) = \Pi_0(t) &= T_*^m(0, t) \cdot \frac{\partial T_*}{\partial x}(0, t) & (0 \leq t \leq \Theta), \\ P_L(t) = \Pi_L(t) &= -T_*^m(L, t) \cdot \frac{\partial T_*}{\partial x}(L, t) & (0 \leq t \leq \Theta), \\ a \min_{(x, t \in Q)} T_*(x, t), & & b = \max_{(x, t \in Q)} T_*(x, t). \end{aligned} \tag{3.1}$$

Что касается величины  $\Theta$ , то она в расчетах принимала различные значения.

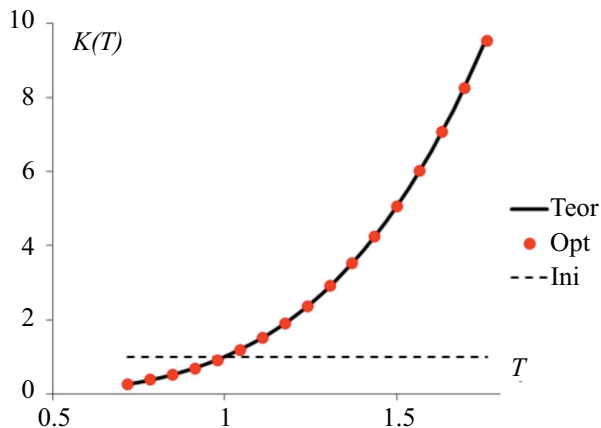
При решении прямой задачи использовалась равномерная сетка с параметрами  $I = 40$  (количество интервалов вдоль оси  $x$ ),  $J = 700$  (количество интервалов вдоль оси  $t$ ), которая обеспечивает достаточную точность вычисления поля температур и поля импульсов (сопряженных переменных). Отрезок  $[a, b]$  разбивался на 32 интервала ( $N = 32$ ).

В первом примере первой группы расчетов рассматривалась задача нахождения объемной теплоемкости и коэффициента теплопроводности с входными данными (3.1) при  $m = 4$  и  $n = 2$ . При этом полагалось, что в целевом функционале (1.4) весовые функции  $\beta_0(t)$ ,  $\beta_L(t)$  тождественно равны единице. Задача с такими параметрами исследовалась при разных значениях интервала времени  $\Theta$ .

Вначале при решении задачи оптимизации функционала (1.4) интервал времени  $\Theta = 1.5$ . Температура на левом конце слоя в этом случае изменяется от 1.246 до 1.762, а на правом — от 0.7192 до 1.438 (есть общий диапазон температур). В качестве начального приближения выбирались функции  $C(T) = 1.0$  и  $K(T) = 1.0$ . Результаты полученного решения обратной задачи представлены на фиг. 1 (функция  $C(T)$ ) и фиг. 2 (функция  $K(T)$ ). Расчеты, выполненные с использованием других (отличных от констант) начальных приближений для функций  $C(T)$  и  $K(T)$ , приводили к тому же решению обратной задачи.

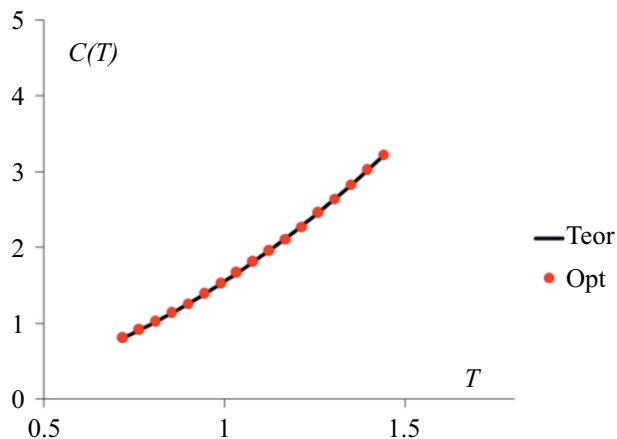


Фиг. 1. Распределение объемной теплоемкости

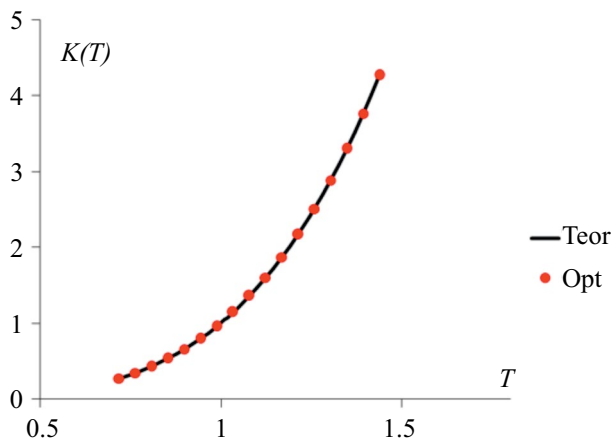


Фиг. 2. Распределение коэффициента теплопроводности

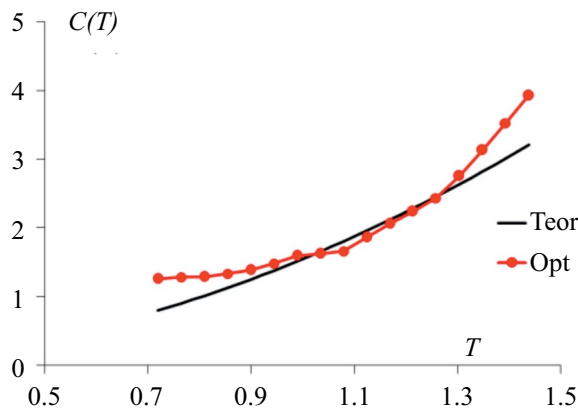
Затем при решении той же задачи полагалось  $\Theta = 0.5$ . Отметим, что здесь температура на левом конце слоя изменяется от 1.246 до 1.438, а на правом — от 0.7192 до 1.017 (есть диапазон температур, для которых потоки не задаются). В зависимости от начального приближения и параметров оптимизационного процесса получались различные решения. На фиг. 3–6 представлены два из получавшихся решений.



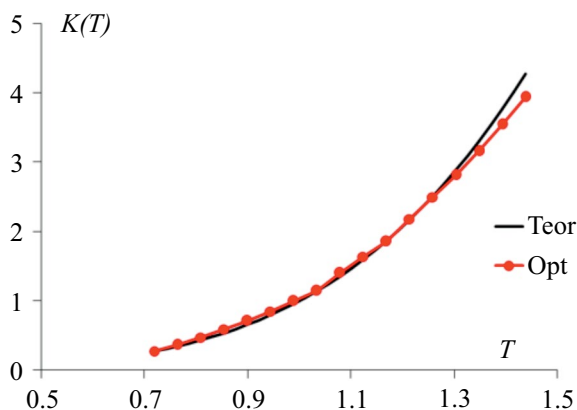
Фиг. 3. Распределение объемной теплоемкости



Фиг. 4. Распределение коэффициента теплопроводности



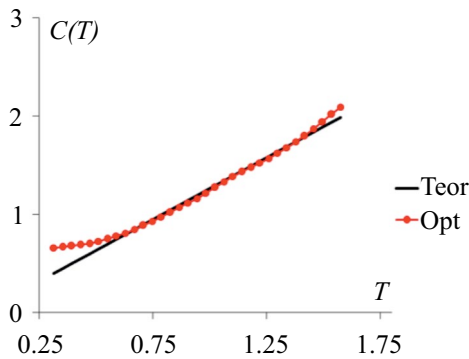
Фиг. 5. Распределение объемной теплоемкости



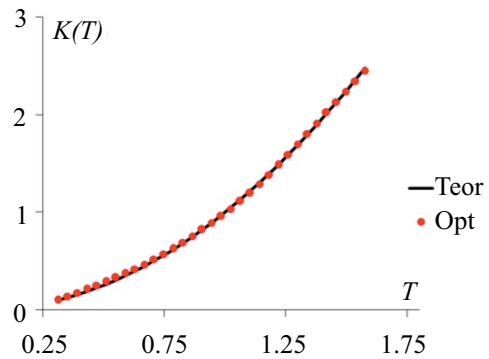
Фиг. 6. Распределение коэффициента теплопроводности

На фиг. 1—6 линии, помеченные как “Теор”, представляют собой ожидаемые решения  $C(T) = \beta \cdot T^n$ ,  $K(T) = T^m$  обратной задачи (1.1)—(1.4) с данными (3.1), а линии, помеченные как “Opt”, — решения, полученные в результате минимизации целевого функционала (1.4).

Во втором примере первой группы расчетов рассматривалась задача нахождения объемной теплоемкости и коэффициента теплопроводности с входными данными (3.1) при  $m = 2$  и  $n = 1$ . При этом, как и раньше, полагалось, что в целевом функционале (1.4) весовые функции  $\beta_0(t)$ ,  $\beta_L(t)$  тождественно равны единице. Задача с такими параметрами также исследовалась при разных значениях интервала времени  $\Theta$ . Качественно ничего нового по сравнению с предыдущим примером здесь обнаружено не было: решение обратной задачи было единственным, если интервалы температур на левом и правом концах слоя (там, где задаются тепловые потоки) пересекаются, и решение неединственно, если на отрезке  $[a, b]$  имеется диапазон температур, для которых потоки не задаются. В предельном случае, когда интервалы температур на левом и правом концах слоя не пересекаются и непрерывно следуют один за другим (на левом конце слоя температура изменяется от 0.9449 до 1.5749, а на правом — от 0.3150 до 0.9449,  $\Theta = 1.0$ ) решение также оказывается единственным, но функционал стремится к нулевому значению чрезвычайно медленно. На фиг. 7 и фиг. 8 представлены промежуточные результаты решения обратной задачи, соответствующие значению целевого функционала (1.4)  $\Phi(C(T), K(T)) \approx F \approx 10^{-9}$ .



Фиг. 7. Распределение объемной теплоемкости



Фиг. 8. Распределение коэффициента теплопроводности

Задача нахождения объемной теплоемкости и коэффициента теплопроводности с входными данными (3.1) решалась и при других значениях параметров  $m$  и  $n$ . На основе результатов, полученных при решении всех обратных задач первой группы, можно сделать вывод о том, что решение задачи определения объемной теплоемкости и коэффициента теплопроводности может быть неединственным.

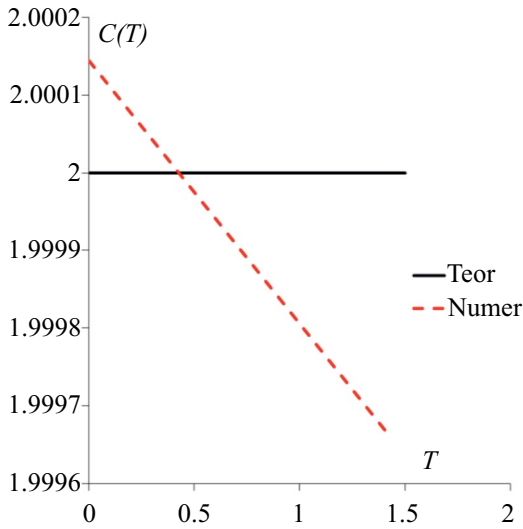
В качестве первого примера второй группы рассматривалась обратная задача (1.1)—(1.4) со следующими входными данными:

$$\begin{aligned}
 L &= 1, & Q &= (0,1) \times (0,\Theta), \\
 w_0(x) &= x^2 & & (0 \leq x \leq 1), \\
 w_1(t) &= t, & w_2(t) &= t + 1 & (0 \leq t \leq \Theta), \\
 P_0(t) &= \Pi_0(t) = 0, & P_L(t) &= \Pi_L(t) = -2 & (0 \leq t \leq \Theta), \\
 & a = 0, & & b = 2,
 \end{aligned} \tag{3.2}$$

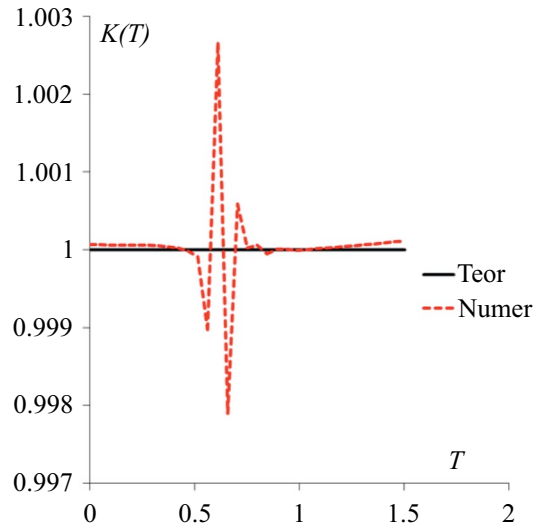
Заметим, что функция  $T_*(x,t) = x^2 + t$  является аналитическим решением уравнения (1.1) при  $C(T) = 2$  и  $K(T) = 1$ , так что пара  $\{C(T), K(T)\} = \{2,1\}$  и функция  $T_*(x,t) = x^2 + t$  являются оптимальным управлением и соответствующим оптимальным решением обратной задачи (1.1)—(1.4) с данными (3.2).

Величина  $\Theta$ , как и ранее, принимала различные значения.

При времени протекания процесса  $\Theta = 1.5$  температура на левом конце слоя изменяется от 0.0 до 1.5, а на правом — от 1.0 до 2.5 (есть общий диапазон температур). В этом случае независимо от выбора начального приближения и параметров используемого оптимизационного алгоритма получалось одно и то же ожидаемое решение обратной задачи  $C(T) = 2$ ,  $K(T) = 1$ .



Фиг. 9. Распределение объемной теплоемкости



Фиг. 10. Распределение коэффициента теплопроводности

Если же  $\Theta = 0.5$ , то в этом случае температура на левом конце слоя изменяется от 0.0 до 0.5, а на правом — от 1.0 до 1.5 (есть диапазон температур, для которых потоки не задаются). Получить решение обратной задачи оказалось чрезвычайно сложно. Целевой функционал здесь имеет большое число локальных экстремумов, и используемые методы оптимизации первого порядка часто “застревали” в локальных минимумах. Удалось так подобрать параметры оптимизационного алгоритма, который приводил к нахождению ожидаемого решения обратной задачи. При этом следует отметить, что элементы сходящейся к решению последовательности имели довольно экзотичный вид (см. фиг. 9, фиг. 10).

Во втором примере второй группы рассматривалась обратная задача (1.1)—(1.4) с такими входными данными:

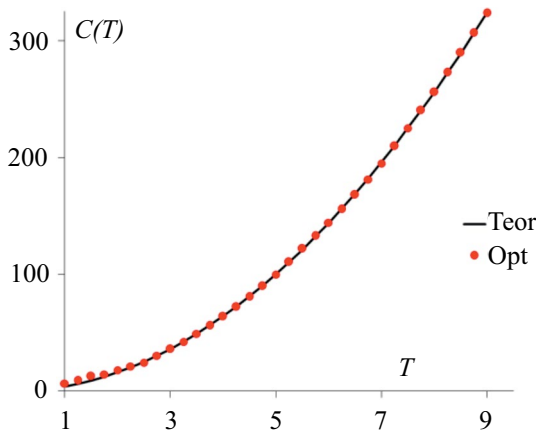
$$\begin{aligned}
 L &= 1, & \Theta &= 1, & Q &= (0,1) \times (0,1), \\
 w_0(x) &= 9 / (x + 1) & & & (0 \leq x \leq 1), \\
 w_1(t) &= 9 / (5t + 1) & & & (0 \leq t \leq 1), \\
 w_2(t) &= 9 / (7t + 2) & & & (0 \leq t \leq 1), \\
 P_0(t) = \Pi_0(t) &= K(T_{**}(0,t)) \cdot \frac{\partial T_{**}}{\partial x}(0,t) & & & (0 \leq t \leq \Theta), \\
 P_L(t) = \Pi_L(t) &= -K(T_{**}(L,t)) \cdot \frac{\partial T_{**}}{\partial x}(L,t) & & & (0 \leq t \leq \Theta), \\
 a &= 1, & b &= 9,
 \end{aligned}$$

где экспериментальное поле  $T_{**}(x,t)$  строилось численно как решение следующей смешанной задачи (1.1)—(1.3):

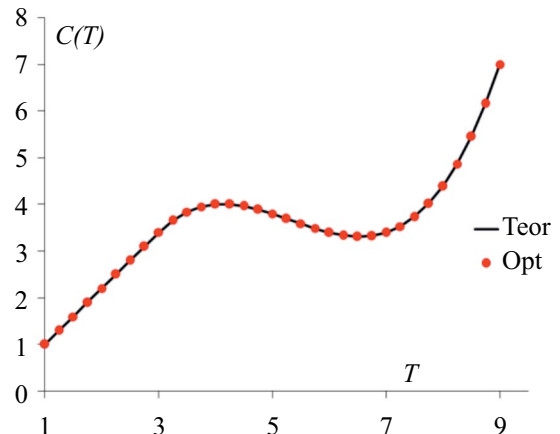
$$\begin{aligned}
 C(T) \frac{\partial T(x,t)}{\partial t} - \frac{\partial}{\partial x} \left( K(T) \frac{\partial T(x,t)}{\partial x} \right) &= 0, & (0 < x < 1), & \quad (0 < t < 1), \\
 T(x,0) &= 9 / (x + 1) & (0 \leq x \leq 1), \\
 T(0,t) &= 9 / (5t + 1) & (0 \leq t \leq 1), \\
 T(L,t) &= 9 / (7t + 2) & (0 \leq t \leq 1),
 \end{aligned}$$

при

$$\begin{aligned}
 K(T) &= \begin{cases} 0.1 \cdot (T - 3) \cdot (T - 6) \cdot (T - 7) + 3.4, & T \geq 3, \\ 1.2 \cdot (T - 3) + 3.4, & T < 3, \end{cases} \\
 C(T) &= 4 \cdot T^2, \quad T \in [a, b], \quad a = 1, \quad b = 9.
 \end{aligned}$$



Фиг. 11. Распределение объемной теплоемкости



Фиг. 12. Распределение коэффициента теплопроводности

Отметим, что в этом примере отрезки изменения температуры на левом и правом краях слоя (там, где задаются тепловые потоки) имеют непустое пересечение.

Здесь, как и в первом примере второй группы, обратная задача решалась численно при разных начальных приближениях. Независимо от начального приближения получалось одно и то же решение. На фиг. 11 и фиг. 12 представлены результаты решения задачи при начальном приближении  $C(T) = 4$ ,  $K(T) = 1$ .

Вторая группа рассмотренных примеров позволяет заключить, что при одних входных параметрах решение обратной задачи определения объемной теплоемкости и коэффициента теплопроводности является единственным, а при других параметрах той же задачи решение является неединственным.

#### 4. АНАЛИЗ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Проведенные вычислительные эксперименты показали, что единственность решения задачи одновременной идентификации объемной теплоемкости и коэффициента теплопроводности в предложенной здесь постановке существенно зависит от того, пересекаются ли отрезки изменения температуры на левом и правом краях слоя (там, где задаются тепловые потоки). Если пересечение указанных отрезков не пустое, то в проведенных расчетах получались единственные решения обратной задачи. Если же на отрезке  $[a, b]$  имеется диапазон температур, для которых потоки не задаются, то у обратной задачи решение может быть неединственным. Эти качественные выводы основаны на результатах решения большого количества конкретных задач. Но среди рассмотренных обратных задач есть задачи с такими входными параметрами, для которых неединственность решения можно обосновать теоретически. Сделаем это при следующих предположениях.

Искомые объемная теплоемкость  $C(T)$  и коэффициент теплопроводности  $K(T)$  должны принадлежать области определения, т.е. удовлетворять условиям  $C(T) \in C([a, b])$ ,  $K(T) \in C([a, b])$ ,  $C(T) > 0$ ,  $K(T) > 0$  для всех  $T \in [a, b]$ . Под областью достижимости будем понимать множество функций  $T(x, t) \in C_{x,t}^{2,1}(Q) \cap C_{x,t}^{1,0}(\bar{Q})$ ,  $P_0(t) \in C([0, \Theta])$ ,  $P_L(t) \in C([0, \Theta])$  таких, что функции  $T(x, t)$  удовлетворяют заданным условиям (1.2), (1.3), для каждой из функций  $T(x, t)$  в области определения найдутся функции  $C(T)$  и  $K(T)$ , при которых  $T(x, t)$  является решением смешанной задачи (1.1)–(1.3), а функции  $P_0(t)$  и  $P_L(t)$  определяются соотношениями

$$P_0(t) = K(T(0, t)) \cdot \frac{\partial T}{\partial x}(0, t), \quad P_L(t) = -K(T(L, t)) \cdot \frac{\partial T}{\partial x}(L, t).$$

Будем под  $\tau$ -изолинией функции  $T(x, t)$  будем понимать, как это принято, множество точек  $(x, t)$  области  $Q$ , в которых значение функции  $T(x, t)$  имеет одно и то же значение  $\tau$ . При этом отметим, что изолиния (множество) может представлять собой объединение нескольких, не связанных между собой линий.

**Теорема.** «Пусть функции  $T(x, t)$ ,  $P_0(t)$  и  $P_L(t)$  принадлежат области достижимости,  $a = \min_{(x,t \in Q)} T(x, t)$ ,  $b = \max_{(x,t \in Q)} T(x, t)$ . Пусть на отрезке  $[a, b]$  имеется интервал  $(\gamma, \delta)$  такой, что  $w_1(t) = T(0, t) \notin (\gamma, \delta)$ ,  $w_2(t) = T(L, t) \notin (\gamma, \delta)$  для всех  $0 \leq t \leq \Theta$ . Если найдется отрезок  $[\alpha, \beta] \in (\gamma, \delta) \in [a, b]$  такой, что на каждой  $\tau$ -изолинии отрезка  $\tau \in [\alpha, \beta]$  функции  $\frac{\partial T(x, t)}{\partial t}$  и  $\left(\frac{\partial T(x, t)}{\partial x}\right)^2$ , рассматриваемые как



функции некоторого параметра  $s$  вдоль этой изолинии, являются линейно зависимыми и функция  $\frac{\partial T(x,t)}{\partial x}$  не равна тождественно нулю, то решение  $\{C(T), K(T)\}$  обратной задачи (1.1)—(1.4) неединственно».

**Доказательство.** На первом этапе доказательства теоремы обратим внимание на линейную зависимость функций  $\frac{\partial T(x,t)}{\partial t}$  и  $\left(\frac{\partial T(x,t)}{\partial x}\right)^2$  вдоль каждой  $\tau$ -изолинии отрезка  $\tau \in [\alpha, \beta]$ . Это означает, что существуют функции  $A(T)$  и  $B(T)$ ,  $T \in [\alpha, \beta]$ , такие что

$$A(T) \cdot \frac{\partial T(x,t)}{\partial t} + B(T) \cdot \left(\frac{\partial T(x,t)}{\partial x}\right)^2 = 0,$$

причем  $A^2(T) + B^2(T) > 0$ . Более того,  $A(T) \neq 0$  для всех  $T \in [\alpha, \beta]$ . Действительно, предположив, что  $A(\tau) = 0$  для какого-то значения  $\tau \in [\alpha, \beta]$ , приходим к противоречию с условием отличия от тождественного нуля производной  $\frac{\partial T(x,t)}{\partial x}$ . Таким образом, справедливо следующее равенство:

$$\frac{\partial T(x,t)}{\partial t} = \tilde{B}(T) \cdot \left(\frac{\partial T(x,t)}{\partial x}\right)^2, \tag{4.1}$$

где

$$\tilde{B}(T) = -\frac{B(T)}{A(T)} \in C([\alpha, \beta]).$$

Так как функции  $T(x,t)$ ,  $P_0(t)$  и  $P_L(t)$  по предположению принадлежат области достижимости, то существуют функции  $K_*(T)$  и  $K_*(T)$  из области определения, при которых функция  $T(x,t)$  является решением смешанной задачи (1.1)—(1.3), а функции  $P_0(t)$  и  $P_L(t)$  определяются равенствами

$$P_0(t) = K_*(T(0,t)) \cdot \frac{\partial T}{\partial x}(0,t), \quad P_L(t) = -K_*(T(L,t)) \cdot \frac{\partial T}{\partial x}(L,t).$$

Покажем, что в области определения кроме них найдутся другие функции  $C(T)$  и  $K(T)$ , при которых эта же функция  $T(x,t)$  является решением смешанной задачи (1.1)—(1.3) и

$$P_0(t) = K(T(0,t)) \cdot \frac{\partial T}{\partial x}(0,t), \quad P_L(t) = -K(T(L,t)) \cdot \frac{\partial T}{\partial x}(L,t).$$

Пусть  $\psi(T) \in C([\alpha, \beta])$  — произвольная функция такая, что  $\int_{\alpha}^{\beta} \tilde{B}(\xi) \cdot \psi(\xi) d\xi = 0$ . Определим функции  $f(T) \in C([\alpha, \beta])$  и  $\sigma(T) \in C^1([\alpha, \beta])$  соотношениями

$$f(T) = \frac{K'_*(T)}{K_*(T)} + \tilde{B}(T) \cdot \psi(T), \quad \sigma(T) = \exp \left[ \int_{\alpha}^T f(\xi) d\xi \right].$$

Построим новые объемную теплоемкость  $C(T)$  и коэффициент теплопроводности  $K(T)$

$$K(T) = \begin{cases} K_*(T), & a \leq T < \alpha, \\ K_*(\alpha) \cdot \sigma(T), & \alpha \leq T \leq \beta, \\ K_*(T) \cdot K_*(\alpha) \cdot \sigma(\beta) / K_*(\beta), & \beta < T \leq b, \end{cases} \tag{4.2}$$

$$C(T) = \begin{cases} C_*(T), & a \leq T < \alpha, \\ K(T) \cdot (C_*(T) / K_*(T) + \psi(T)), & \alpha \leq T \leq \beta, \\ C_*(T) \cdot K(T) / K_*(T), & \beta < T \leq b, \end{cases}$$

и покажем, что они также являются решением обратной задачи для рассматриваемых функции-поля  $T(x,t)$  и потоков  $P_0(t)$ ,  $P_L(t)$  из области достижимости.

Во-первых, непосредственной проверкой нетрудно убедиться в том, что построенные функции  $K(T) \in C^1([a, b])$ ,  $C(T) \in C([a, b])$  и что при всех  $T \in [\alpha, \beta]$  справедливо соотношение

$$\frac{K'(T)}{K(T)} - \tilde{B}(T) \cdot \frac{C(T)}{K(T)} = \frac{K'_*(T)}{K_*(T)} - \tilde{B}(T) \cdot \frac{C_*(T)}{K_*(T)}, \tag{4.3}$$

а при всех  $T \in [a, \alpha]$  и  $T \in [\beta, b]$  — соотношения

$$\frac{K'(T)}{K(T)} = \frac{K'_*(T)}{K_*(T)}, \quad \frac{C(T)}{K(T)} = \frac{C_*(T)}{K_*(T)}. \tag{4.4}$$

Во-вторых, используя соотношения (4.3), (4.4), покажем, что при заданной в условии теоремы функции  $T(x, t)$  уравнение (1.1) выполняется не только при заданных  $(C_*(T), K_*(T))$ , но и при вновь построенных  $(C(T), K(T))$ . Действительно, при предположениях, сформулированных в условиях теоремы, уравнение (1.1) можно представить в виде

$$C(T) \cdot \frac{\partial T(x, t)}{\partial t} - K'(T) \cdot \left( \frac{\partial T(x, t)}{\partial x} \right)^2 - K(T) \cdot \frac{\partial^2 T(x, t)}{\partial x^2} = 0. \tag{4.5}$$

В тех точках области  $Q$ , в которых  $T \in [a, \alpha]$  или  $T \in [\beta, b]$ , имеем

$$\begin{aligned} \Psi(x, t) &\equiv C(T) \cdot \frac{\partial T(x, t)}{\partial t} - K'(T) \cdot \left( \frac{\partial T(x, t)}{\partial x} \right)^2 - K(T) \cdot \frac{\partial^2 T(x, t)}{\partial x^2} = \\ &= K(T) \cdot \left[ \frac{C(T)}{K(T)} \cdot \frac{\partial T(x, t)}{\partial t} - \frac{K'(T)}{K(T)} \cdot \left( \frac{\partial T(x, t)}{\partial x} \right)^2 - \frac{\partial^2 T(x, t)}{\partial x^2} \right] = \\ &= K(T) \cdot \left[ \frac{C_*(T)}{K_*(T)} \cdot \frac{\partial T(x, t)}{\partial t} - \frac{K'_*(T)}{K_*(T)} \cdot \left( \frac{\partial T(x, t)}{\partial x} \right)^2 - \frac{\partial^2 T(x, t)}{\partial x^2} \right] = \\ &= \frac{K(T)}{K_*(T)} \cdot \left[ C_*(T) \cdot \frac{\partial T(x, t)}{\partial t} - K'_*(T) \cdot \left( \frac{\partial T(x, t)}{\partial x} \right)^2 - K_*(T) \cdot \frac{\partial^2 T(x, t)}{\partial x^2} \right] = 0. \end{aligned}$$

Здесь использовались соотношения (4.4).

В тех же точках области  $Q$ , в которых  $T \in [\alpha, \beta]$ , имеем

$$\begin{aligned} \Psi(x, t) &\equiv C(T) \cdot \frac{\partial T(x, t)}{\partial t} - K'(T) \cdot \left( \frac{\partial T(x, t)}{\partial x} \right)^2 - K(T) \cdot \frac{\partial^2 T(x, t)}{\partial x^2} = \\ &= K(T) \cdot \left[ \frac{C(T)}{K(T)} \cdot \frac{\partial T(x, t)}{\partial t} - \frac{K'(T)}{K(T)} \cdot \left( \frac{\partial T(x, t)}{\partial x} \right)^2 - \frac{\partial^2 T(x, t)}{\partial x^2} \right] = \end{aligned}$$

(используем условие (4.3))

$$= K(T) \cdot \left[ \frac{C(T)}{K(T)} \cdot \frac{\partial T(x, t)}{\partial t} - \left[ \frac{K'_*(T)}{K_*(T)} - \tilde{B}(T) \cdot \left( \frac{C_*(T)}{K_*(T)} - \frac{C(T)}{K(T)} \right) \right] \cdot \left( \frac{\partial T(x, t)}{\partial x} \right)^2 - \frac{\partial^2 T(x, t)}{\partial x^2} \right] =$$

(используем условие (4.1) в первой квадратной скобке)

$$= K(T) \cdot \left[ \frac{C_*(T)}{K_*(T)} \cdot \tilde{B}(T) \cdot \left( \frac{\partial T(x, t)}{\partial x} \right)^2 - \frac{K'_*(T)}{K_*(T)} \cdot \left( \frac{\partial T(x, t)}{\partial x} \right)^2 - \frac{\partial^2 T(x, t)}{\partial x^2} \right] =$$

(используем условие (4.1) в первом слагаемом)

$$\begin{aligned} &= K(T) \cdot \left[ \frac{C_*(T)}{K_*(T)} \cdot \frac{\partial T(x, t)}{\partial t} - \frac{K'_*(T)}{K_*(T)} \cdot \left( \frac{\partial T(x, t)}{\partial x} \right)^2 - \frac{\partial^2 T(x, t)}{\partial x^2} \right] = \\ &= \frac{K(T)}{K_*(T)} \cdot \left[ C_*(T) \cdot \frac{\partial T(x, t)}{\partial t} - K'_*(T) \cdot \left( \frac{\partial T(x, t)}{\partial x} \right)^2 - K_*(T) \cdot \frac{\partial^2 T(x, t)}{\partial x^2} \right] = 0. \end{aligned}$$

Таким образом, прямая задача (1.1)–(1.3), имеет одно и то же решение  $T(x, t)$  как для функций  $C_*(T)$  и  $K_*(T)$ , так и для семейства функций  $C(T)$  и  $K(T)$ .

В-третьих, покажем, что при всех  $T \in [a, \alpha]$  и  $T \in [\beta, b]$  функции  $K(T)$  совпадают с функцией  $K_*(T)$ . Что касается отрезка  $T \in [a, \alpha]$ , то здесь совпадение следует из определения функций  $K(T)$  (см. (4.2)). Для  $T \in [\beta, b]$  имеем

$$\begin{aligned}\sigma(\beta) &= \exp \left[ \int_{\alpha}^{\beta} \left( \frac{K'_*(\xi)}{K_*(\xi)} + \tilde{B}(\xi) \cdot \psi(\xi) \right) d\xi \right] = \exp \left[ \int_{\alpha}^{\beta} \left( \frac{K'_*(\xi)}{K_*(\xi)} \right) d\xi \right] = \\ &= \exp \left[ \int_{\alpha}^{\beta} d \ln(K_*(\xi)) \right] = \exp \left[ \ln \frac{K_*(\beta)}{K_*(\alpha)} \right] = \frac{K_*(\beta)}{K_*(\alpha)}.\end{aligned}$$

Отсюда получим

$$K(T) = K_*(T) \cdot K_*(\alpha) \cdot \sigma(\beta) / K_*(\beta) = K_*(T) \cdot K_*(\alpha) \cdot \frac{K_*(\beta)}{K_*(\alpha)} / K_*(\beta) = K_*(T).$$

Как следует из соотношений (4.2), при  $T \in [a, \alpha]$  и  $T \in [\beta, b]$  функции  $C(T)$  также совпадают с функцией  $C_*(T)$ .

Итак, при выполнении условий, сформулированных в теореме, обратная задача (1.1)—(1.4) наряду с решением  $C_*(T)$  и  $K_*(T)$  имеет семейство решений  $C(T)$  и  $K(T)$ , определяемое формулами (4.2).

**Замечание.** Примерами, иллюстрирующими теорему, могут служить примеры из первой группы расчетов при  $\Theta = 0.5$ .

### ЗАКЛЮЧЕНИЕ

В работе рассмотрена одна из возможных постановок обратной коэффициентной задачи, позволяющая по результатам наблюдения за тепловыми потоками на границах слоя одновременно идентифицировать зависящие от температуры объемную теплоемкость вещества и его коэффициент теплопроводности. Предложен численный алгоритм решения этой задачи. Результатом решения обратной задачи являются как искомые параметры вещества, так и температурное поле. Свойства именно этого поля из области достижимости определяют, будет ли решение обратной задачи единственно. На основе полученных результатов сформулированы рекомендации, что для единственности решения обратной задачи следует использовать такие экспериментальные данные, при которых отрезки изменения температуры на левом и правом краях слоя (там, где задаются тепловые потоки) пересекаются. Высказано также предположение, что для получения единственного решения обратной задачи целесообразно в целевом функционале использовать как тепловой поток на границе области, так и значение температурного поля в области или в какой-то ее подобласти.

Авторы выражают искреннюю благодарность А. Ф. Албу за неоценимую помощь в проведении расчетов и обсуждение получаемых результатов.

### СПИСОК ЛИТЕРАТУРЫ

1. Зверев В. Г., Гольдин В. Д., Назаренко В. А. Радиационно-кондуктивный теплоперенос в волокнистой термостойкой изоляции при тепловом воздействии // Теплофиз. высоких температур. 2008. Т. 46. № 1. С. 119—125.
2. Алифанов О. М., Черепанов В. В. Математическое моделирование высокопористых волокнистых материалов и определение их физических свойств // Теплофиз. высоких температур. 2009. Т. 47. № 3. С. 463—472.
3. Алифанов О. М. Обратные задачи теплообмена. М.: Машиностр., 1988.
4. Yeung W. K., Lam T. T. Second-order finite difference approximation for inverse determination of thermal conductivity // Internat. Journal of Heat Mass Transfer. 1996. V. 39. P. 3685—3693.
5. Chen H. T., Lin J. Y., Wu C. H., Huang C. H. Numerical algorithm for estimating temperature-dependent thermal conductivity // Numerical Heat Transfer. 1996. V. B29. P. 509—522.
6. Самарский А. А., Вабищевич П. Н. Вычислительная теплопередача. М.: Едиториал УРСС, 2003.
7. Kim, S., Kim, M. C., Kim, K. Y. Non-iterative estimation of temperature dependent thermal conductivity without internal measurements // Internat. Journal of Heat Mass Transfer. 2003. V. 46. P. 1891—1810.
8. E. Majchrzak, K. Freus, S. Freus. Identification of temperature dependent thermal conductivity using the gradient method // J. of Applied Math. and Comput. Mechan. 2006. V. 5. Issue 1. P. 114—123.
9. Balazs Czel, Gyula Grof. Inverse identification of temperature-dependent thermal conductivity via genetic algorithm with cost function-based rearrangement of genes // Internat. Journal of Heat and Mass Transfer. 2012. V. 55. No. 15. P. 4254—4263.

10. Matsevityi Yu.M., Alekhina S.V., Borukhov V.T., Zayats G.M., Kostikov A.O. Identification of the Thermal Conductivity Coefficient for Quasi-Stationary Two-Dimensional Heat Conduction Equations // *J. of Engng Physics and Thermophysics*. 2017. V. 90. No. 6. P. 1295—1301.
11. Evtushenko Y., Zubov V., Albu A. Inverse coefficient problems and fast automatic differentiation // *J. of Inverse and Ill-Posed Problems*. 2022. V. 30(3). P. 447—460.
12. Huang C.H., Yan J.Y. An inverse problem in simultaneously measuring temperature-dependent thermal conductivity and heat capacity // *Internat. Journal of Heat Mass Transfer*. 1995. V. 38. P. 3433—3441.
13. Imani A., Ranjbar A.A., Esmkhani M. Simultaneous estimation of temperature-dependent thermal conductivity and heat capacity based on modified genetic algorithm // *Inverse Problems in Science and Engng*. 2006. V. 14. No. 7. P. 767—783.
14. Miao Cui, Kai Yang, Xiao-liang Xu, Sheng-dong Wang, Xiao-wei Gao. A modified Levenberg-Marquardt algorithm for simultaneous estimation of multi-parameters of boundary heat flux by solving transient nonlinear inverse heat conduction problems // *Internat. Journal of Heat and Mass Transfer*. 2016. V. 97. P. 908—916.
15. Евтушенко Ю. Г. Оптимизация и быстрое автоматическое дифференцирование. М.: ВЦ им. А. А. Дородницына РАН, 2013. 144 с.
16. Евтушенко Ю. Г., Zubov V. И. Об обобщенной методологии быстрого автоматического дифференцирования // *Ж. вычисл. матем. и матем. физ.* 2016. Т. 56. № 11. С. 1847—1862.
17. Евтушенко Ю. Г., Засухина Е. С., Zubov V. И. О численном подходе к оптимизации решения задачи Бюргера с помощью граничных условий // *Ж. вычисл. матем. и матем. физ.* 1997. Т. 37. № 12. С. 1449—1458.
18. Албу А. Ф., Zubov V. И. Исследование задачи оптимального управления процессом кристаллизации вещества в новой постановке для объекта сложной геометрической формы // *Ж. вычисл. матем. и матем. физ.* 2014. Т. 54. № 12. С. 1879—1893.
19. Евтушенко Ю. Г., Zubov V. И., Албу А. Ф. Оптимальное управление тепловыми процессами с фазовыми переходами. Коллективная монография. М.: МАКС Пресс, 2021. С. 248. DOI: 10.29003/m2449.978-5-317-06677-2. ISBN978-5-317-06677-2.
20. Горчаков А. Ю., Zubov V. И. Об одновременном определении коэффициента теплопроводности и объемной теплоемкости вещества // *Ж. вычисл. матем. и матем. физ.* 2023. Т. 63. № 8. С. 1279—1295.
21. Самарский А. А. Теория разностных схем. М.: Наука, 1989.
22. Албу А. Ф., Zubov V. И. Об эффективности решения задач оптимального управления с помощью методологии быстрого автоматического дифференцирования // *Труды Института матем. и механ. УрО РАН*. 2015. Т. 21. № 4. С. 20—29.
23. Hascoet L., Pascual V. The Tapenade automatic differentiation tool: principles, model, and specification // *ACM Transactions on Mathematical Software (TOMS)*. 2013. Т. 39. № 3. P. 1—43.
24. Hogan, R.J. Fast reverse-mode automatic differentiation using expression templates in C++ // *ACM Transactions on Mathematical Software (TOMS)*, 40(4), pp. 26—42 (2014).
25. Горчаков А. Ю. О программных пакетах быстрого автоматического дифференцирования // *Информационные Технологии и Вычисл. Системы*. 2018. № 1. P. 30—36.
26. Albu A., Gorchakov A., Zubov V. On the effectiveness of the fast automatic differentiation methodology // *Communications in Computer and Information Science*. 2019. V. 974. P. 264—276.
27. Yixuan Qiu, L-BFGS++, 2021. <https://github.com/yixuan/LBFGSpp/>

## DETERMINATION OF THE THERMAL CONDUCTIVITY AND VOLUMETRIC HEAT CAPACITY OF SUBSTANCE FROM HEAT FLUX

A. Yu. Gorchakov<sup>a</sup>, V. I. Zubov<sup>a,\*</sup>

<sup>a</sup>*Federal Research Center “Computer Science and Control” of the RAS,  
44 Vavilova St., Moscow, 119333 Russia*

*\*e-mail: vladimir.zubov@mail.ru*

*Received 22 September, 2023*

*Revised 04 November, 2023*

*Accepted 20 November, 2023*

**Abstract.** The study of nonlinear problems related to heat transfer in a substance is of great practical importance. Earlier, this paper's authors proposed an effective algorithm for determining the volumetric heat capacity and thermal conductivity of a substance based on experimental observations of the dynamics of the temperature field in the object. In this paper, the problem of simultaneous identification of temperature-dependent volumetric heat capacity and thermal conductivity of the substance under study from the heat flux at the boundary of the domain is investigated. The consideration is based on the first boundary value problem for a one-dimensional unsteady heat equation. The coefficient inverse problem under consideration is reduced to a variational problem, which is solved by gradient methods based on the application of fast automatic differentiation. The uniqueness of the solution of the inverse problem is investigated.

**Keywords:** thermal conductivity, coefficient inverse problems, gradient, heat equation, heat flow.

## МЕТОД ПООЧЕРЕДНЫХ ПРОЕКЦИЙ ДЛЯ ПЕРЕСЕЧЕНИЯ ВЫПУКЛЫХ МНОЖЕСТВ, МНОГОАГЕНТНЫЕ АЛГОРИТМЫ КОНСЕНСУСА И УСРЕДНЯЮЩИЕ НЕРАВЕНСТВА<sup>1)</sup>

© 2024 г. А. В. Проскурников<sup>1,\*</sup>, И. С. Забарянская<sup>2</sup>

<sup>1</sup> 10129 Turin, Italy, Corso Duca degli Abruzzi, 24, Politecnico di Torino (Политехнический университет Турина), Италия

<sup>2</sup> 199178 Санкт-Петербург, 14-я линия В.О., 29, СПбГУ, Россия

\* e-mail: avp1982@gmail.com

\*\* e-mail: akshiira@yandex.ru

Поступила в редакцию 03.11.2023 г.

Переработанный вариант 11.11.2023 г.

Принята к публикации 20.11.2023 г.

История метода поочередных проекций для нахождения общей точки нескольких выпуклых множеств в евклидовом пространстве восходит к известному алгоритму Качмаржа для решения систем линейных уравнений, который появился в 1930-х годах и впоследствии нашел широкое применения в обработке изображений и компьютерной томографии. Важную роль в исследовании данного метода сыграли работы И.И. Ерёмкина, Л.М. Брэгмана и Б.Т. Поляка, появившиеся практически одновременно и содержащие весьма общие результаты о сходимости последовательных проекций к точке на пересечении множеств, если это пересечение непусто. В настоящей статье мы рассматриваем модификацию задачи о пересечении выпуклых множеств, относящуюся к теории многоагентных систем и называемую задачей о консенсусе с ограничениями. Каждое выпуклое множество в этой задаче связано со своим агентом и, вообще говоря, недоступно другим агентам. При этом группа агентов заинтересована в нахождении общей точки этих множеств: точки, удовлетворяющей ограничениям. Распределенные аналоги метода поочередных проекций, предложенные для решения этой задачи, приводят к достаточно сложной нелинейной системе уравнений, сходимость которой, обычно, доказывается с помощью специальных функций Ляпунова. В работе дается краткий обзор данных методов и показывается их связь с теоремой о консенсусе в системе усредняющих неравенств, недавно установленной в работах первого автора и развивающей результаты о сходимости обычного метода последовательных усреднений в задаче о консенсусе. Библ. 48. Фиг. 6.

**Ключевые слова:** метод поочередных проекций, выпуклое программирование, фейеровские отображения, распределенные алгоритмы, консенсус, многоагентные системы.

DOI: 10.31857/S0044466924040078, EDN: ZJPOEZ

### 1. ВВЕДЕНИЕ

Ряд задач численного анализа и оптимизации сводятся к нахождению точки в пересечении семейства замкнутых выпуклых множеств в конечномерном либо гильбертовом векторном пространстве (или установлению того, что данное пересечение пусто). Одним из специальных случаев является проверка совместности системы линейных уравнений с нахождением одного из возможных решений, а также проверка совместности ограничений в линейном программировании; взрывной рост интереса к быстрым алгоритмам для решения таких задач был связан, в частности, с изобретением компьютерной томографии и связанных с ней алгоритмов быстрого восстановления изображений (см. [1]). К нахождению точки в пересечении полупространств аффинного пространства сводится вопрос о разделении двух множеств точек в пространстве, важный для задач обучения (см. [2], [3]) и математической диагностики (см. [4]–[6]). К задачам более общего вида приводит выпуклая оптимизация: на начальной стадии прямого (primal) метода оптимизации требуется найти некоторую (произвольную) точку, удовлетворяющую системе выпуклых ограничений. Наконец, в теории управления пересечения выпуклых множеств возникают в связи с оцениванием параметров и траекторий систем при неизвестных помехах (см. [7]); примером являются системы рекуррентных целевых неравенств (см. [8], [9]).

<sup>1)</sup> Работа выполнена при финансовой поддержке РФФ (код проекта 23-11-00229).

Большинство имеющихся в литературе алгоритмов для вычисления общей точки выпуклых множеств являются модификациями и обобщениями метода поочередных (alternating) проекций. В англоязычной литературе для данного метода также используется аббревиатура POCS (projection onto convex sets, т.е. проекции на выпуклые множества). История данного метода восходит к работам Качмаржа [10] и Чиммино [11] по релаксационным методам решения систем линейных уравнений, а впервые (в специальном случае двух линейных подпространств) он, по-видимому, упомянут в лекциях фон Неймана по геометрии гильбертова пространства, изданных в виде монографии [12]. Разработанная впоследствии теория метода поочередных проекций, во многом базируется на работах трех крупнейших специалистов в области оптимизации и численных методов: И. И. Ерёмкина, Б. Т. Поляка и Л. М. Брэгмана, чьи пионерские работы [13]–[15] появились практически одновременно. Исследование последовательных итеративных проекций привело, с одной стороны, к развитию теории фейеровских процессов (см. [16]–[19]), а с другой стороны, послужило мотивацией для изучения  $D$ -расстояния (дивергенции) Брэгмана (см. [20]), в настоящее время играющего исключительно важную роль в задачах машинного обучения, оптимизации и вычислительной геометрии.

В настоящее время существуют достаточно подробные обзоры по задаче о пересечении выпуклых множеств и ее обобщениям (в частности, сходимости фейеровских приближений), среди которых особо следует отметить монографии [21], [22] и более раннюю работу [23], а также более специальный обзор по теории оценивания [7]. С некоторыми обобщениями задачи на невыпуклый случай (проектирование на многообразиях) можно ознакомиться в работе [24], содержащей также достаточно полный обзор литературы по данной теме. Эти обзоры, однако, не покрывают *многоагентные* версии задачи о пересечении выпуклых множеств, исследуемые в последние годы под различными названиями, — например задача о консенсусе с ограничениями (см. [25]) или “оптимальном консенсусе” (см. [26]). В этих задачах каждое выпуклое множество является собственностью одного из агентов и недоступно другим агентам (это может быть вызвано как требованиями приватности, так и сложностью описания самих множеств, информацию о которых по каким-то причинам может быть затруднительно передавать через сеть). Целью группы агентов, как и в исходной задаче, является нахождение точки в пересечении всех множеств, при этом обмениваться информацией каждый из агентов может лишь с некоторым набором “соседей”. Отношения соседства описываются коммуникационным графом, который может меняться и быть неполностью известен. Распределенные децентрализованные алгоритмы для решения многоагентной задачи о пересечении множеств изучались, главным образом, в журналах по теории управления и малоизвестны специалистам из других областей, а имеющиеся обзоры по теме в основном ограничиваются специальными случаями, такими как алгоритмы решения линейных уравнений (см. [27]).

Настоящая статья восполняет указанный пробел и дает краткий обзор распределенных алгоритмов, которые были предложены для решения многоагентной задачи о пересечении выпуклых множеств. Мы покажем, что

1) в данных алгоритмах (как и в централизованном случае) операторы проектирования могут быть заменены произвольными фейеровскими (парасжимающими) отображениями,

2) сходимость к консенсусу в предложенных алгоритмах может быть доказана единым способом в рамках теории *усредняющих неравенств* на графах. Данная теория была развита в диссертации первого автора (см. [28]), а также работах [29], [30]. Для удобства читателя, мы приводим также доказательства основных результатов, прежде не публиковавшихся на русском языке.

Оставшаяся часть статьи организована следующим образом. В разд. 2 приводятся постановка задачи о пересечении выпуклых множеств, алгоритмы поочередных проекций и их обобщения (фейеровские процессы), а также краткий обзор основных результатов о сходимости этих алгоритмов, в частности, фундаментальные результаты Ерёмкина, Брэгмана и Гурина–Поляка–Райка. Основной материал статьи, относящийся к многоагентной постановке задачи, приводится в разд. 3, где мы также приводим некоторые факты о сходимости классических усредняющих алгоритмов консенсуса на переменных графах, а также ассоциированных с ними систем усредняющих неравенств. Раздел 4 завершает статью. Доказательство основных результатов о сходимости многоагентных алгоритмов приводится в приложении.

## 2. ЗАДАЧА О ПЕРЕСЕЧЕНИИ МНОЖЕСТВ И ЕЕ ОБОБЩЕНИЯ

В настоящей статье мы будем рассматривать задачу о пересечении выпуклых множеств в конечномерном евклидовом пространстве, однако ряд упоминаемых ниже результатов может быть распространен также на гильбертовы пространства (с некоторыми уточнениями, связанными с типом сходимости последовательных приближений). Проекционные методы изучались также на многообразиях, допускающих эффективное вычисление проекционного оператора (см. [24]); данное обобщение выходит за рамки настоящей статьи.

2.1. Постановка задачи. Оператор проектирования

Классическая задача о пересечении выпуклых множеств (или о совместности выпуклых ограничений), как правило формулируется следующим образом.

**Задача А.** Пусть есть набор выпуклых множеств  $\{\Xi_i\}_{i \in \mathcal{V}}$  в  $\mathbb{R}^d$ , где  $\mathcal{V}$  – некоторое конечное множество индексов, пересечение которых непусто:

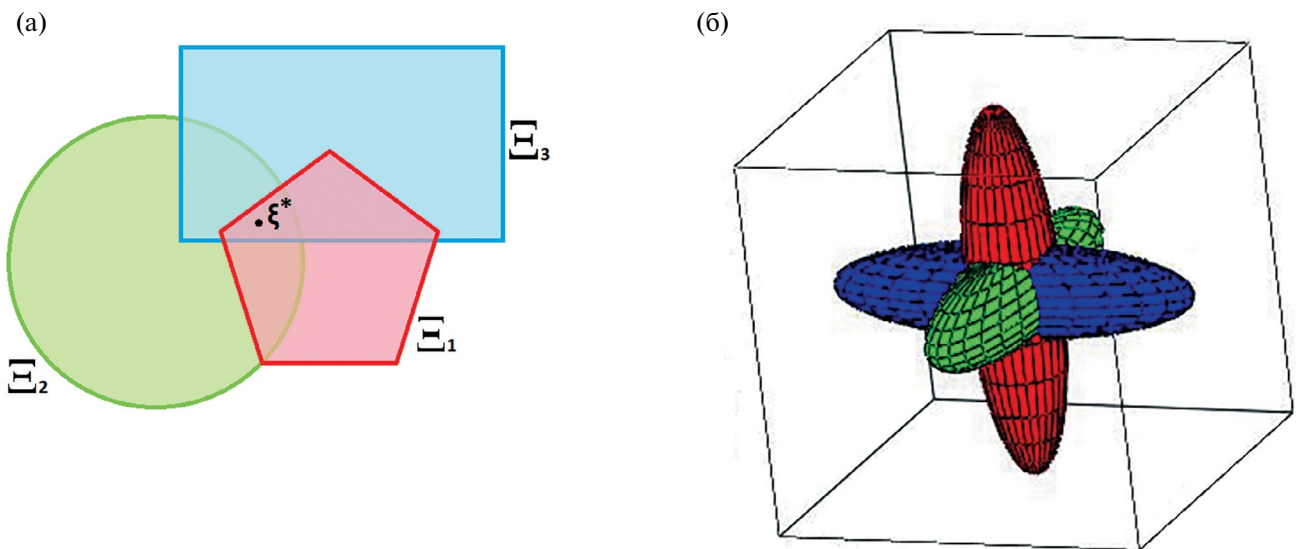
$$\Xi_* = \bigcap_{i \in \mathcal{V}} \Xi_i \neq \emptyset. \tag{1}$$

Требуется найти хотя бы одну точку  $\xi_* \in \Xi_*$ .

Более сложная задача — дополнительно распознать ситуацию, когда  $\Xi_* = \emptyset$ .

**Задача В.** Для заданного набора выпуклых множеств  $\{\Xi_i\}_{i \in \mathcal{V}}$  требуется проверить выполнение условия (1) и, если оно выполнено, найти хотя бы одну точку  $\xi_* \in \Xi_*$ .

Отметим, что в случае одномерного пространства  $d = 1$  задачи А и В решаются просто, поскольку единственно возможное выпуклое множество в  $\mathbb{R}$  — промежуток (открытый, полуоткрытый или замкнутый). Пересечение конечного семейства промежутков тривиально находится путем сортировки их левых и правых концов (фиг. 1). В этом случае мы можем определить множество  $\Xi_*$ . В случае  $d = 2$  множество  $\Xi_*$  обычно легко визуализировать с помощью компьютера, однако описать его структуру полностью или хотя бы найти одну из его точек  $\xi_*$  аналитически — достаточно трудная задача. Уже в случае трехмерного пространства  $d = 3$  даже визуализация множества  $\Xi_*$  весьма затруднительна.



Фиг. 1. Примеры пересечений выпуклых множеств: (а) — на плоскости, (б) — в пространстве.

Алгоритмы решения задач А и В, исходно предложенные в литературе, основывались на последовательных проекциях. Напомним фундаментальный результат о проектировании на выпуклое множество и определение проекционного оператора. Для любого замкнутого выпуклого множества  $\Omega \subset \mathbb{R}^d$  и  $x \in \mathbb{R}^d$  оператор проектирования  $P_\Omega : \mathbb{R}^d \rightarrow \Omega$  сопоставляет точке ближайший элемент из  $\Omega$ , т.е.  $|x - P_\Omega(x)| = \min_{y \in \Omega} |x - y|$ . Минимум расстояния всегда достигается, а ближайшая точка единственна. Это утверждение справедливо в произвольном гильбертовом пространстве (см. [31], теорема 1.4.1).

Можно показать, что  $\angle(y - P_\Omega(x), x - P_\Omega(x)) \geq \pi / 2 \forall y \in \Omega$  (фиг. 2)

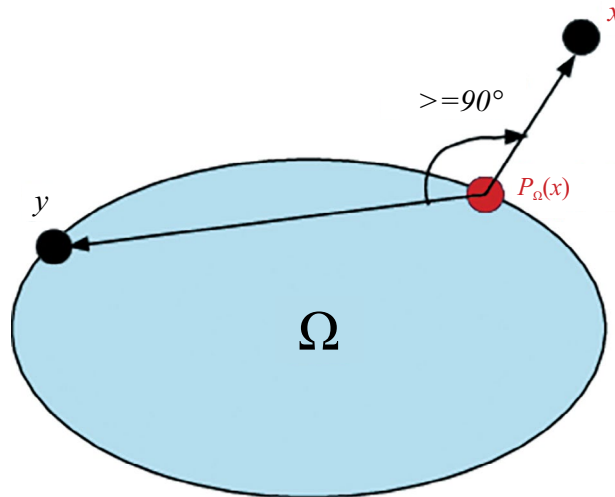
и

$$|x - y|^2 \geq |x - P_\Omega(x)|^2 + |y - P_\Omega(x)|^2 \quad \forall y \in \Omega. \tag{2}$$

<sup>2</sup> Символ  $|\cdot|$  всюду ниже обозначает обычную евклидову норму в  $\mathbb{R}^d : |x|^2 = \sum_i x_i^2 = x^T x$ .

<sup>3</sup> Символ  $\angle$  обозначает угол между векторами (в интервале  $[0, \pi]$ ).



Фиг. 2. Проекция точки  $x$  на замкнутое выпуклое множество  $\Omega$ .

В силу свойства (2), если  $y \in \Omega$  (иными словами,  $y$  — неподвижная точка  $P_\Omega$ ), то

$$|P_\Omega(x) - y| \leq |x - y|,$$

причем неравенство строгое тогда и только тогда, когда  $|x - P_\Omega(x)| > 0$ , т.е.  $x \notin \Omega$ . Непрерывное отображение с таким свойством называется *фейеровским* или (в англоязычной литературе) *парасжимающим*; ниже будет дано формальное определение.

## 2.2. Первые работы: системы линейных уравнений и неравенств

Простейшим (и первым в историческом плане) примером задачи о пересечении выпуклых множеств является решение системы линейных уравнений<sup>4</sup>

$$a_i^\top \xi = b_i \quad \forall i \in \mathcal{V},$$

где  $\xi \in \mathbb{R}^d$  — неизвестный вектор, а  $a_i \in \mathbb{R}^d, b_i \in \mathbb{R}$ , определяют коэффициенты  $i$ -го уравнения. В данном случае  $\Xi_i = \{\xi : a_i^\top \xi = b_i\}$  — аффинная гиперплоскость размерности  $d - 1$ , а поиск решения системы уравнений (более точно, одного из решений) может быть рассмотрен как пересечение этих выпуклых множеств. Подчеркнем, что система уравнений может быть как переопределенной ( $d < |\mathcal{V}|$ ), так и недоопределенной ( $d > |\mathcal{V}|$ ), уравнения могут быть линейно зависимы, а количество уравнений и переменных может быть очень велико. Такие системы уравнений, например, возникают в компьютерной томографии при восстановлении изображений по проекциям (см. [1]), леонтьевских моделях отраслевого баланса (см. [32]), вычислении вектора PageRank (см. [33], [34]) и аналогичных характеристик в задачах ранжирования и поиска, нахождении собственных векторов. Два простых проекционных алгоритма для решения линейных уравнений были практически одновременно предложены в 1930-е годы в работах Качмаржа и Чиммино.

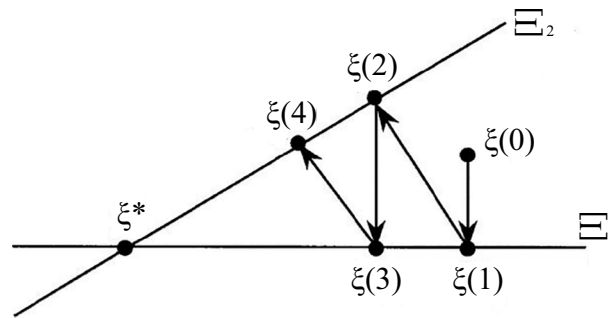
**2.2.1. Алгоритм Качмаржа.** Алгоритм Качмаржа<sup>5</sup> (см. [10], [35]) заключается в последовательном проектировании точки на гиперплоскости в циклическом порядке (для случая двух уравнений  $|\mathcal{V}| = 2$  и двух переменных  $d = 2$  процедура проиллюстрирована на фиг. 3).

Нетрудно видеть, что оператор проектирования на множество  $\Xi = \{\xi : a^\top \xi = b\}$  (где  $|a| \neq 0$ ) является аффинным отображением:

$$P_\Xi : x \mapsto x + \frac{b - a^\top x}{|a|^2} a. \quad (3)$$

<sup>4</sup> Здесь и далее элементы  $\mathbb{R}^d$  понимаются как вектор-столбцы, соответственно  $a^\top \xi$  есть скалярное произведение двух векторов  $a$  и  $\xi$ .

<sup>5</sup> Также этот алгоритм называется методом построчного действия (row action method), поскольку на каждом шаге используется одна строка матрицы ограничений.



Фиг. 3. Итеративный метод Качмаржа.

Предположим, что уравнения некоторым образом пронумерованы от 1 до  $n$ , так что  $\mathcal{V} = \{1, \dots, n\}$ , и выбрано некоторое произвольное приближение к решению  $\xi(0)$ . На итерации с номером  $t = 0, 1, \dots$  приближение  $\xi(t)$  заменяется на

$$\xi(t + 1) = P_{\Xi_i}(\xi(t)) = \xi(t) + \frac{b_i - a_i^T \xi(t)}{|a_i|^2} a_i, \text{ где } t + 1 \equiv i \pmod n.$$

В случае двух множеств данный алгоритм впервые появляется в лекциях фон Неймана (1933), изданных впоследствии в виде монографии [12], при этом рассматривались линейные замкнутые подпространства гильбертова пространства  $\Xi_1, \Xi_2$ . Фон Нейманом было доказано (см. [12], теорема 10.7), что последовательность  $\xi(t)$  сходится к проекции начальной точки  $\xi(0)$  на пространство  $\Xi_* = \Xi_1 \cap \Xi_2$  по норме. Данная сходимость экспоненциальна, а ее скорость определяется квадратом косинуса угла между подпространствами  $\Xi_1$  и  $\Xi_2$  (см. [36]). Указанный результат — экспоненциальная сходимость  $\xi(t)$  к проекции  $P_{\Xi_*}(\xi(0))$  — остается справедливым и для  $n > 2$  подпространств, однако в этом случае известны лишь верхние оценки для скорости сходимости (см. [22], [23], [36]), которая, вообще говоря, не определяется только углами между всевозможными парами подпространств  $(\Xi_i, \Xi_j)$ . Более подробно с историей алгоритма Качмаржа и его обобщениями, полученными в последние годы, можно ознакомиться в обзоре [35].

**2.2.2. Алгоритм Чиммино.** Итеративный алгоритм Чиммино, предложенный в [11], в определенном смысле является прообразом распределенных многоагентных алгоритмов, рассматриваемых в разд. 3. Его отличием от алгоритма Качмаржа является то, что на итерации  $\xi(t)$  проектируется одновременно на все гиперплоскости. Затем вычисляется центр масс полученных проекций  $\xi_b(t)$ , который может быть выбран в качестве следующего приближения  $\xi(t + 1)$ . Более общо, следующее приближение может быть получено как

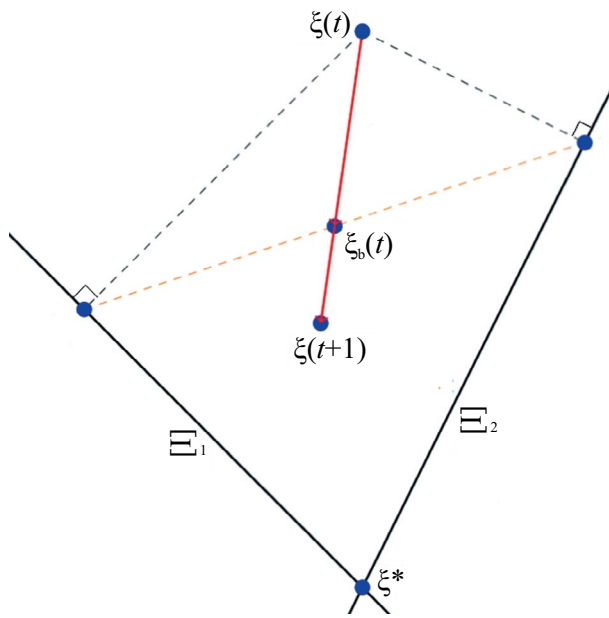
$$\xi(t + 1) = \xi(t) + \alpha(\xi_b(t) - \xi(t)),$$

где  $\alpha \in (0, 2)$  — некоторый фиксированный шаг (фиг. 4).

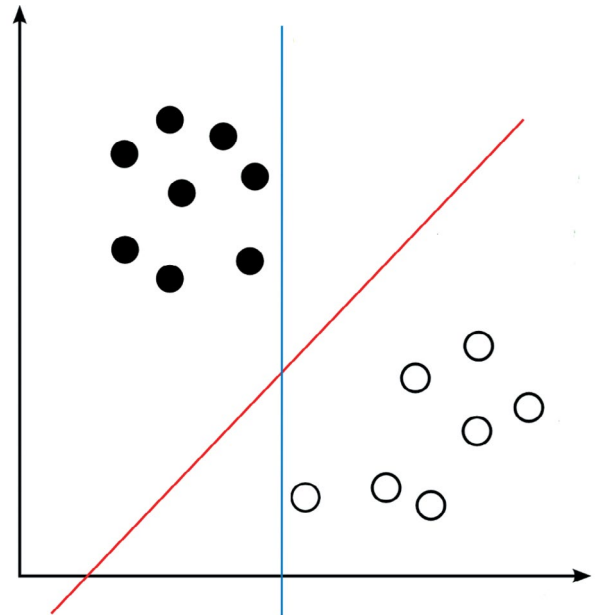
Большим преимуществом алгоритма Чиммино является возможность параллелизации операций проектирования, которые могут выполняться несколькими процессорами. Вместе с тем, число итераций данного алгоритма, требуемое для нахождения решения с заданной точностью, может быть достаточно велико. Существуют ускоренные версии алгоритма, в частности, блочный алгоритм Чиммино (см. [37]) для разреженных систем уравнений; их рассмотрение выходит за рамки настоящей статьи.

**2.2.3. Алгоритм Агмона–Мотцкина для пересечения полупространств.** Важным этапом в исследовании задачи о пересечении выпуклых множеств стали вышедшие одновременно работы Агмона [38] и Мотцкина и Шёнберга [39]; идея изучаемого в них алгоритма, как упомянуто в [38], принадлежит Мотцкину. В отличие от работ Качмаржа и Чиммино, рассматривается задача о решении системы линейных неравенств, т.е. нахождении точки в пересечении полупространств

$$\Xi_i = \left\{ \xi \in \mathbb{R}^d : a_i^T \xi \leq b_i \right\}.$$



Фиг. 4. Метод Чиммино (иллюстрация для двумерного случая).



Фиг. 5. Аффинные гиперплоскости, разделяющие два множества точек в пространстве.

Проектор на полупространство  $\Xi = \{ \xi : a^T \xi \leq b \}$  нетрудно вычислить. Отметим, что он уже не является аффинным отображением:

$$P_{\Xi} : x \mapsto x' = \begin{cases} x + \frac{b - a^T x}{|a|^2} a, & a^T x > b, \\ x, & a^T x \leq b. \end{cases} \tag{4}$$

Равенство  $a_i^T \xi = b_i$ , очевидно, может быть записано как пара неравенств, таким образом, метод из [38], [39] позволяет находить решения произвольной системы линейных уравнений и неравенств (в предположении, что такое решение существует). Отметим, что к решению неравенств сводятся многие “линейные” задачи обучения (например, алгоритмы настройки весов перцептрона), классификации и математической диагностики (см. [2], [4]–[6]). В этих задачах требуется найти гиперплоскость, разделяющую две группы точек в многомерном пространстве (фиг. 5). Задавая гиперплоскость линейным уравнением, условие разделения является системой неравенств на коэффициенты этого уравнения.

Метод из работ [38], [39] по реализации напоминает алгоритм Качмаржа (где проектор вместо (3) приобретает вид (4)), однако имеется одно существенное отличие: на каждой итерации  $t = 0, 1, \dots$  проектирование осуществляется на самое *дальнее* множество. Формально, находится индекс<sup>6</sup>

$$m(t) = \arg \max_{i \in \mathcal{V}} d(\xi(t), \Xi_i), \tag{5}$$

где  $d(\xi, \Xi) = \min_{x \in \Xi} \{ |\xi - x| \}$ . Проектирование осуществляется на полупространство с индексом  $m(t)$ , таким образом, следующее приближение имеет вид

$$\xi(t+1) = P_{\Xi_{m(t)}}(\xi(t)).$$

В работах [38], [39] была доказана (в предположении непустоты пересечения множеств) экспоненциальная сходимость этого алгоритма, а также его “релаксированного” варианта:

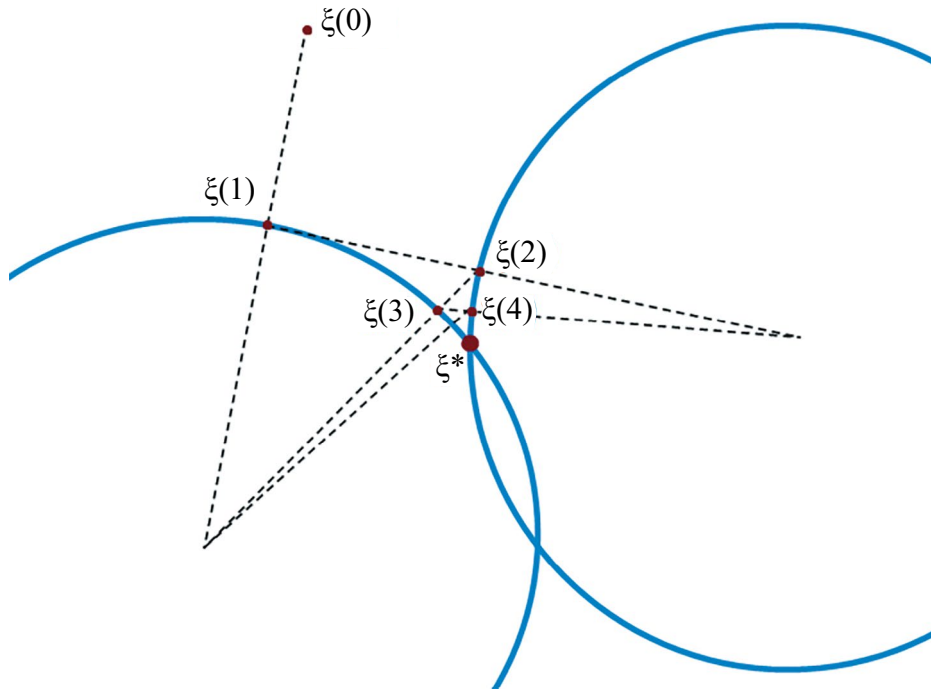
$$\xi(t+1) = (1 - \alpha)\xi(t) + \alpha P_{\Xi_{m(t)}}(\xi(t)), \quad \alpha \in (0, 2). \tag{6}$$

<sup>6</sup> В случае необщего положения, когда точка равноудалена от нескольких множеств и  $\arg \max$  определен неоднозначно, можно выбрать произвольное значение; обычно, выбирается наименьший индекс (см. [13]).

Отметим, что скорость сходимости может быть определена явно (см. [38]) и зависит от матрицы  $A$ , хотя ее вычисление нетривиально. По сложности итерация алгоритма Агмона–Мотцкина эквивалентна итерации алгоритма Чиммино, поскольку требуется вычислить расстояния до всех полупространств, прежде чем выполнить проектирование.

### 2.3. Метод попеременных проекций для произвольных множеств

Методы попеременных проекций для отыскания общей точки произвольных замкнутых выпуклых множеств чаще всего используют идею алгоритмов Качмаржа (поочередные проекции на множества в циклическом порядке, см. фиг. 6) и Агмона–Мотцкина (в этом случае на каждом шаге выбирается самое дальнее множество).



Фиг. 6. Проекции в циклическом порядке (иллюстрация для двух множеств).

Поскольку исследованию данного класса методов посвящена обширная литература, в частности, обзоры [22], [23], перечислим лишь основные работы, положившие начало целому ряду исследований по выпуклой оптимизации с ограничениями.

Исторически первой была статья И. И. Ерёмкина [40] (написанная и представленная в редакцию в 1962 г.), в которой рассматривался метод Агмона–Мотцкина (5), (6) для произвольных замкнутых выпуклых множеств  $\Xi_i$ . Было показано, что если пересечение множеств непусто, то метод сходится, и наоборот, если последовательность  $\xi(t)$  имеет предел, то ее предел автоматически принадлежит пересечению всех множеств  $\Xi_*$ .

В работе Л. М. Брэгмана [14] исследовался вопрос сходимости алгоритмов при последовательном проектировании в циклическом порядке (как в алгоритме Качмаржа) и при проекции на самое дальнее выпуклое множество, как в алгоритме Агмона–Мотцкина. Сходимость исследовалась в гильбертовом пространстве, однако понималась лишь в слабом смысле (в случае конечномерного пространства она эквивалентна сходимости по норме) и была доказана в предположении, что множества имеют непустое пересечение. Более того, была исследована сходимость модифицированного алгоритма Агмона–Мотцкина в случае полупространств  $\Xi_i = \{ \xi : a_i^T \xi \leq b_i \}$ . В этом модифицированном методе проекция осуществляется не на самое дальнее полупространство  $\Xi_i$ , а на множество, для которого ограничение  $a_i^T \xi \leq b_i$  на шаге  $t$  “наиболее нарушено”, иными словами, (5) нужно заменить условием

$$m(t) = \arg \max_{i \in V} [a_i^T \xi(t) - b_i],$$

в то время как правило проектирования (6) остается без изменения.

Дальнейшее обобщение результатов Ерёмина и Брэгмана было получено в работе Л. Г. Гурина<sup>7</sup>, Б. Т. Поляка и Э. В. Райка [15]. Как и в работе [14], исследовались два типа алгоритма попеременных проекций (в циклическом порядке и на самое дальнее множество) для замкнутых выпуклых множеств в произвольном гильбертовом пространстве. Были получены условия экспоненциальной (в норме гильбертова пространства) сходимости обоих алгоритмов в случае, если каждое из множеств  $\Xi_i$  имеет непустое пересечение с внутренностями остальных множеств:

$$\Xi_i \cap \bigcap_{j \neq i} \overset{\circ}{\Xi}_j \neq \emptyset \quad \forall i \in \mathcal{V}.$$

Последнее условие можно снять, если  $\Xi_i$  являются полупространствами (таким образом, результат [38] был обобщен на бесконечномерный случай). Приведено также некоторое условие сходимости по норме (“равномерная выпуклость” множеств  $\Xi_i$ ). Во всех указанных результатах точный оператор проектирования можно заменить его “релаксированной” версией, как в алгоритме (6) (более того, шаг  $\alpha$  может на каждой итерации алгоритма быть разным). Показано, что алгоритм циклических проекций при пустом пересечении множеств  $\Xi_i$  экспоненциально быстро сходится к периодической траектории. Был предложен также ускоренный вариант метода Агмона–Мотцкина (проекция на самое дальнее множество). В статье [15] приводится ряд интересных приложений метода последовательных проекций, в частности, к задаче аппроксимации функции полиномами на отрезке и специальной задаче оптимального управления.

**2.3.1. Дальнейшее развитие: фейеровские операторы и оптимизация.** Отметим, что алгоритмы для вычисления общей точки выпуклых множеств, рассматривавшиеся до сих пор, основаны на неявном предположении, что оператор проектирования на каждое из множеств может быть эффективно вычислен. Если в случае линейных гиперплоскостей и полупространств он допускает просто аналитическое представление, то в случае более сложных множеств (например, заданных некоторым нелинейным выпуклым неравенством  $f(\xi) \leq 0$ ) вычисление проекции возможно только численно и, вообще говоря, сводится к решению задачи выпуклой оптимизации. Вместе с тем, уже в первых работах (см. [10], [13], [38], [39]) было замечено, что оператор проекции может быть заменен “релаксированной” проекцией (6), и, более того, основное свойство, используемое в доказательстве — то, что он обладает некоторым слабым свойством сжатия и приближает любую точку за пределами выбранного множества  $\Xi$  к этому множеству. Данное наблюдение привело к развитию теории *фейеровских процессов*, основы которой были заложены Ерёминым в пионерских работах [13], [16], [42]–[44]. Наиболее полное изложение данной теории доступно в монографии [21]. Как будет показано в разд. 3, есть возможность обобщить результаты о сходимости итеративных процедур с фейеровскими операторами на многоагентный случай (см. теорему 1). В следующем пункте мы подробнее рассмотрим фейеровские отображения и приведем некоторые примеры.

Помимо того, часто интерес представляет нахождение не произвольной точки в пересечении выпуклых множеств, а точки, доставляющей оптимум некоторому функционалу (возможно, негладкому). Возможность решения задач линейного программирования с помощью метода последовательных проектирований была замечена Брэгманом (см. [14]), в последующих работах которого (см. [20], [45]) были рассмотрены итерационные методы “проектирования” в смысле некоторой “псевдометрики” на линейном пространстве, которая была названа  $D$ -расстоянием. Наиболее часто используемый тип  $D$ -расстояний называется *дивергенцией* Брэгмана и задается формулой

$$D_\varphi(x, y) = \varphi(x) - \varphi(y) - \nabla\varphi(y)(x - y)$$

для некоторой выпуклой дифференцируемой функции  $\varphi$ . Брэгманом было показано (см. [20]), что методы проекции в смысле подходящим образом построенного  $D$ -расстояния позволяют решать задачи выпуклого программирования с ограничениями в виде равенств и неравенств. Альтернативный подход, предложенный Ерёминым (см. [43]), приводит к изучению фейеровских отображений со сдвигом. Отметим, что в многоагентном случае алгоритмы оптимизации такого рода малоизучены, как правило, используются альтернативные алгоритмы, полученные децентрализацией методов проекции градиента (см. [25], [46], [47]); зачастую ограничения также заменяются барьерными функциями. Изучение алгоритмов оптимизации с ограничениями выходит за рамки настоящей статьи, достаточно полное изложение современных методов доступно, например, в монографии [48].

<sup>7</sup> К сожалению, при оцифровке переводной версии работы [41] фамилия первого автора была ошибочно написана как Gubin; такое написание до сих пор сохраняется в электронной базе ScienceDirect и попало в большинство обзоров по алгоритмам пересечения выпуклых множеств.

2.4. Фейеровские (паражсжимающие) отображения

Всюду в данном пункте мы предполагаем заданными некоторый оператор  $P : \mathbb{R}^d \rightarrow \mathbb{R}^d$  (вообще говоря, нелинейный) и некоторую норму  $\| \cdot \|$  на  $\mathbb{R}^d$  (не обязательно евклидову). Напомним, что на конечномерном пространстве все нормы эквивалентны и порождают одну и ту же топологию, соответственно понятия открытости, замкнутости и непрерывности в различных нормах также эквивалентны.

**Определение 1.** Оператор  $P$  называется *фейеровским*<sup>8</sup> относительно множества  $M \subset \mathbb{R}^d$ , или *M-фейеровским*, если выполнены три условия: [(а)]

1.  $P$  – непрерывное отображение на всем пространстве;
2. множество  $M$  неподвижно относительно  $P$ ;
3. оператор удовлетворяет условию сжатия относительно  $M$ , т.е.

$$\| P(\xi) - \xi_0 \| = \| P(\xi) - P(\xi_0) \| < \| \xi - \xi_0 \| \quad \forall \xi \notin M, \xi_0 \in M. \tag{7}$$

Очевидно, что из (7) следует, что  $P$  обладает свойством

$$\| P(\xi) - P(\xi_0) \| \leq \| \xi - \xi_0 \| \quad \forall \xi \in \mathbb{R}^d, \xi_0 \in M. \tag{8}$$

В англоязычной литературе фейеровские операторы называются часто парасжимающими (paracontracting) (см. [50]), встречаются также и иные термины, например, “притягивающий” (attracting) к множеству  $M$  оператор (см. [23]). Очевидно, что если  $\| x - y \| < \| P(x) - P(y) \|$ , то  $x \neq P(x)$ ; таким образом,  $M$  в определении 1 совпадает со множеством всех неподвижных точек  $M$ -фейеровского оператора  $\mathcal{F}(P) \triangleq \{ \zeta : P(\zeta) = \zeta \}$ ; поэтому, если не требуется явное уточнение данного множества, мы для краткости будем просто называть оператор фейеровским. Легко видеть также, что множество  $M$  автоматически является замкнутым и выпуклым. Первое свойство вытекает из непрерывности оператора  $P$ : если последовательность  $\{x_n\}$  элементов  $M$  сходится к некоторому пределу  $x \in \mathbb{R}^d$ , то, очевидно, для произвольного элемента  $a \in M$  имеем  $\| x - a \| = \lim_{n \rightarrow \infty} \| x_n - a \| = \lim_{n \rightarrow \infty} \| P(x_n) - a \| = \| P(x) - a \|$ , что возможно лишь в случае, когда  $x \in M$ . Второе свойство также легко доказать: если  $z$  – произвольный элемент отрезка, соединяющего точки  $x, y \in M$ , то  $\| x - y \| = \| x - z \| + \| z - y \| \geq \| x - P(z) \| + \| P(z) - y \| \geq \| x - y \|$ . Данное неравенство может быть выполнено только в случае, когда  $z \in M$  (в противном случае,  $\| x - P(z) \| < \| x - z \|$  и  $\| y - P(z) \| < \| y - z \|$ ). Следовательно, вместе с двумя точками  $M$  содержит и отрезок, их соединяющий.

Фейеровское отображение является естественным обобщением сжимающего (по Банаху) отображения  $P$ , для которого  $\| P(x) - P(y) \| \leq q \| x - y \| \quad \forall x, y \in \mathbb{R}^d$  при некоторой постоянной  $q \in (0, 1)$ . Принципиальным отличием сжимающих и фейеровских отображений является то, что у сжимающих отображений, в силу известной теоремы Банаха, существует ровно одна неподвижная точка, а множество неподвижных точек фейеровского отображения может быть любым замкнутым выпуклым множеством. В следующем пункте приводятся примеры фейеровских операторов.

2.4.1. Примеры фейеровских операторов.

**Пример 1.** Как мы видели, проектор  $P_\Omega$  на замкнутое выпуклое множество является фейеровским (в стандартной евклидовой норме). Более общо, можно показать, что при любом коэффициенте  $a \in (0, 2)$  оператор  $(1 - a)Id + aP_\Omega$ , где  $Id$  – тождественное отображение, также является фейеровским в евклидовой норме (см., например, лемму 2 в [40] и следствие 2.5 в [23]).

**Пример 2.** Выпуклая комбинация нескольких фейеровских отображений является фейеровским отображением. Пусть  $P_1, \dots, P_n$  – фейеровские операторы с множествами неподвижных точек  $M_1, \dots, M_n$ , и  $a_1, \dots, a_n > 0$  – коэффициенты, такие что  $\sum_{i=1}^n a_i = 1$ . Тогда легко проверить, что оператор  $P = \sum_{i=1}^n a_i P_i$  является фейеровским относительно  $M = M_1 \cap \dots \cap M_n$  (в предположении, что оно непусто). Объединяя данный пример с примером 1, можно показать, что в алгоритме Чиммино, описанном в п. 2.1, приближение  $\xi(t + 1)$  получается из  $\xi(t)$  путем применения фейеровского оператора.

**Пример 3.** Аналогично, легко проверить, что композиция фейеровских операторов  $P_n \circ \dots \circ P_1$  также является фейеровским относительно пересечения множеств неподвижных точек  $M = M_1 \cap \dots \cap M_n$  (в предположении, что оно непусто). Таким образом, последовательные  $n$  операций (где  $n$  – число множеств) в алгоритме Качмаржа эквивалентны применению фейеровского оператора.

<sup>8</sup> Термин “ $M$ -фейеровский оператор”, введенный в работах И.И. Ерёмкина, мотивирован понятием монотонной по Фейеру (относительно  $M$ ) последовательности, введенным в [39]. Так называется последовательность  $\{x_n\} \subset \mathbb{R}^d \setminus M$ , такая что  $x_{n+1} \neq x_n$  при всех  $n$  и  $\| x_n - y \| \geq \| x_{n+1} - y \|$  при всех  $n$  и  $y \in M$ . Это понятие, в свою очередь, восходит к работе Фейера [49] о расположении корней полиномов.

**Пример 4.** Пусть  $P$  в некоторой строго выпуклой норме<sup>9</sup>  $\|\cdot\|$  является нерасширяющим ( $\|P(x) - P(y)\| \leq 1$  при всех  $x, y \in \mathbb{R}^d$ ) с непустым множеством неподвижных точек  $M = \mathcal{F}(P)$ . Тогда при  $a \in (0, 1)$  оператор  $(1 - a)\text{Id} + aP$  также является нерасширяющим, а также  $M$ -фейеровским. Первое свойство проверяется непосредственно, кроме того, если  $y \in M$  и  $x \notin M$ , то  $P(x) - y \neq x - y$ , следовательно, либо  $\|P(x) - y\| < \|x - y\|$ , либо векторы  $P(x) - y$  и  $x - y$  неколлинеарны, либо  $P(x) - y = y - x$ . Во всех случаях

$$\|(1 - a)x + aP(x) - y\| = \|(1 - a)(x - y) + a(P(x) - y)\| < \|x - y\|.$$

В силу сделанного выше замечания получаем, в частности, известный факт: множество неподвижных точек нерасширяющего (относительно строго выпуклой нормы) отображения всегда является выпуклым (см. [23]).

Примеры 2—4 интересны тем, что явное вычисление проекции на множество  $M$  может быть затруднительно, тем не менее, может быть известен фейеровский оператор, приближающий точку не из  $M$  к этому множеству. На самом деле, таких примеров достаточно много. Следующий пример относится к распространенному случаю, когда выпуклое множество задается некоторым скалярным выпуклым неравенством.

**Пример 5.** Рассмотрим непрерывно дифференцируемую выпуклую функцию  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  и предположим, что множество  $M_f \triangleq \{x : f(x) \leq 0\}$  непусто (очевидно, это множество выпукло и замкнуто). Выбрав постоянную  $\alpha \in (0, 2)$ , рассмотрим отображение

$$P_{\alpha, f} : x \mapsto x' = \begin{cases} x - \alpha f(x) \frac{\nabla f(x)}{|\nabla f(x)|^2}, & f(x) > 0, \\ x, & f(x) \leq 0. \end{cases}$$

Отметим, что градиент  $\nabla f$ , в силу выпуклости, может обращаться в нуль лишь в точках (глобального) минимума  $f$ , а все точки минимума (если они существуют) должны принадлежать  $M_f$ . В силу этого, выражение  $P_{\alpha, f}(x)$  корректно определено при  $f(x) > 0$ . Нетрудно также показать, что  $P_{\alpha, f}$  — непрерывное (но, вообще говоря, недифференцируемое) отображение с множеством неподвижных точек  $M_f$ . Важное наблюдение, сделанное в [13], состоит в том, что оператор  $P_{\alpha, f}$  является фейеровским относительно евклидовой нормы. Конструкция тривиально распространяется на недифференцируемые выпуклые отображения, для которых можно выбрать непрерывную ветвь субдифференциала. Более точно, если существует непрерывное отображение  $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , такое что  $g(x) \in \partial f(x)$  при всех  $x$ , то в определении  $P_{\alpha, f}$  градиент  $\nabla f(x)$  можно заменить на  $g(x)$  (см. [13]).

Другими примерами фейеровских отображений являются проксимальные операторы и операторы градиентного спуска, соответствующие строго выпуклым функциям (см. [50]).

**2.4.2. Итерации фейеровских операторов.** Следующая простая лемма показывает, что итерации фейеровского оператора так же, как и сжимающего (в смысле Банаха) отображения, сходятся к неподвижной точке.

**Лемма 1.** Пусть оператор  $P : \mathbb{R}^d \rightarrow \mathbb{R}^d$  является фейеровским с непустым множеством неподвижных точек  $M = \mathcal{F}(P)$ . Тогда последовательность  $(\xi(t))_{t=0}^{\infty}$ , задаваемая рекуррентным соотношением

$$\xi(t + 1) = P(\xi(t)), \quad t = 0, 1, \dots,$$

при любой начальной точке  $\xi(0)$  сходится к некоторому элементу множества  $M$ .

Доказательство леммы 1 будет дано в приложении. В силу примеров 2 и 3, из леммы 1 нетрудно вывести сходимость алгоритма Качмаржа и алгоритма Чиммино (для проекторов на произвольные множества не обязательно линейной структуры). Понятие фейеровского оператора и лемму 1 можно обобщить и на многозначные отображения (возникающие, например, в алгоритме Агмона–Мотцкина, если разрешить индексу  $m(t)$  в (5) принимать несколько значений). Наиболее общие результаты о сходимости и расходимости фейеровских процессов могут быть найдены в [17], [21], [23], [51].

<sup>9</sup> Норма строго выпукла, если выполнено строгое неравенство треугольника:  $\|x + y\| < \|x\| + \|y\|$  для любых линейно независимых векторов  $x, y$ .

### 3. МНОГОАГЕНТНЫЕ АЛГОРИТМЫ ДЛЯ ПЕРЕСЕЧЕНИЯ МНОЖЕСТВ

Обратимся теперь к многоагентной<sup>10</sup> постановке задачи о пересечении выпуклых множеств. В этой задаче каждое замкнутое выпуклое множество  $\Xi_i \subseteq \mathbb{R}^d$  связано со своим агентом (это может быть программный модуль, робот или живое существо, обладающее некоторой автономностью в принятии решений). Часто это множество трактуется как ограничение, выдвигаемое агентом  $i$ . Мы считаем, что агент  $i$  располагает достаточной информацией о своем множестве  $\Xi_i \subseteq \mathbb{R}^d$ , в частности, может вычислять проекцию  $P_{\Xi_i}(\xi)$  произвольной точки  $\xi \in \mathbb{R}^d$  на множество  $\Xi_i$  либо, более общо, значение какого-либо иного фейеровского оператора  $P_i(\xi)$  с множеством неподвижных точек  $\Xi_i$ . При этом агенту  $i$  недоступна информация о множествах других агентов  $\Xi_j$ ,  $j \neq i$ . Это может быть связано как со сложной структурой множеств, информацию о которых затруднительно передать через сеть, так и с требованиями приватности (множества могут зависеть от некоторой конфиденциальной информации). Целью агентов является найти некоторую точку  $\xi_* \in \Xi_* \triangleq \bigcap_{i \in \mathcal{I}} \Xi_i$ , которая удовлетворяет всем ограничениям.

В предположении неограниченной коммуникации между агентами общая точка может, теоретически, быть найдена с помощью обычного метода попеременных проекций (обобщенного алгоритма Качмаржа). Для этого требуется некоторым образом занумеровать агентов и обеспечить после каждой операции проектирования передачу полученного приближения к точке  $\xi(t)$  следующему (в циклическом порядке) агенту. Агент 1 выполняет проектирование начальной точки  $\xi(0)$  на множество  $\Xi_1$  либо, более общо, вычисление некоторого фейеровского оператора  $P_1(\xi(0))$  и передает результат агенту 2, применяющему оператор  $P_2$  и так далее, агент  $n$  применяет оператор  $P_n$  и передает результат агенту 1. Такая процедура, однако, неудовлетворительна по ряду причин. Во-первых, будучи формально независимыми, агенты лишены возможности выполнять операции параллельно: каждый из них простаивает  $n - 1$  из последовательных  $n$  итераций. Во-вторых, с точки зрения парадигмы многоагентных систем, агенты в общем случае должны быть равнозначны и не иметь каких-либо отличительных признаков в виде уникального индекса, соответственно, для получения некоторого упорядочения от 1 до  $n$  агентам требуется осуществить специальную процедуру. Наконец, как уже было сказано, на шаге с номером  $t$  должна быть возможность коммуникации между агентами с номерами  $1 + (t \bmod n)$  и  $1 + (t + 1 \bmod n)$ . На практике это гарантировано не всегда: коммуникационный граф может быть неполным, меняться с течением времени и быть заранее неизвестным (например, ряд коммуникационных линий может отключаться вследствие отказов или необходимости передать более важную информацию).

Более удобны с точки зрения многоагентной реализации обобщенные методы Чиммино (проектирование на все множества и вычисление барицентра) и Агмона–Мотцкина (нахождение самого дальнего множества), однако они предполагают, что агенты могут обмениваться информацией с некоторым центральным узлом, который собирает у них значения  $P_i(\xi(t))$ , а затем вычисляет их барицентр (либо самую дальнюю проекцию) для нахождения  $\xi(t + 1)$ . Наличие такого центрального узла в сети также не всегда возможно, кроме того, сложность такой централизованной системы растет с ростом числа агентов.

В силу сказанного, для решения задачи о пересечении множеств, принадлежащих агентам, были предложены альтернативные алгоритмы, которые близки по идее алгоритму Чиммино, однако основываются на идее консенсуса, достигаемого путем последовательных усреднений (короткое введение в консенсусные алгоритмы дано в следующем пункте). В отличие от обычных проекционных алгоритмов, каждый агент  $i$  имеет свое собственное состояние  $\xi^i(t) \in \mathbb{R}^d$ , которое на каждом шаге становится доступным соседним (в смысле коммуникационного графа) агентам, обновляется в зависимости от состояний соседей и в конечном итоге сходится к точке из пересечения множеств, при этом данное установившееся состояние одинаково для всех агентов.

Оставшаяся часть данного раздела организована следующим образом. В п. 3.1 приведено краткое изложение консенсусных алгоритмов и условие достижения консенсуса в сети агентов с переменным направленным графом. Нам данный результат понадобится в несколько более сильном виде (робастность консенсуса к исчезающе малым возмущениям). Другой важный результат — лемма о консенсусе, гарантируемом усредняющими неравенствами, которому посвящен п. 3.2. В п. 3.3 мы приводим несколько классов алгоритмов, предложенных в литературе для решения многоагентной задачи о пересечении множеств и формулируем основной результат статьи, который будет доказан в приложении.

<sup>10</sup> Отметим, что в настоящем обзоре мы не рассматриваем общую философию теории многоагентных (мультиагентных) систем и не даем точного определения агентов, которое различается в теории управления и компьютерных науках. С историей вопроса и основными задачами теории многоагентных систем (также называемых сетевыми) можно ознакомиться, например, в статьях и монографиях [52]–[58].



### 3.1. Классические алгоритмы консенсуса. Сведения из теории графов

Алгоритмы усреднения имеют длинную историю, восходящую, с одной стороны, к динамике марковских процессов (см. [59]), а с другой стороны, к агентным (микроскопическим) моделям динамики мнений (см. [60], [61]). В данном пункте мы ограничимся лишь кратким описанием линейных алгоритмов усреднения и приведем основное (на сегодняшний день) условие достижения консенсуса.

**3.1.1. Динамика последовательного усреднения. Консенсус.** Рассмотрим множество агентов  $\mathcal{V}$ , каждый из которых описывается скалярной переменной  $x_i(t)$ ,  $i \in \mathcal{V}$ , называемой *состоянием* и меняющейся в дискретном времени  $t = 0, 1, \dots$ . На каждом шаге  $t$  каждый из агентов вычисляет *взвешенное среднее* своего собственного состояния и состояний других агентов. Агенту, однако, доступна информация не обо всех состояниях, а лишь об агентах из некоторого множества  $\mathcal{V}_i(t) \subseteq \mathcal{V}$ , которое мы будем называть *множеством соседей* агента  $i$  в момент  $t$ . Мы всегда считаем, что  $i \in \mathcal{V}_i(t)$ . В каждый момент времени  $t \in \mathbb{N}$ , агент  $i$  присваивает *веса усреднения*  $a_{ij}(t)$  всем соседям. Веса каждого агента  $i$  удовлетворяют ограничениям

$$a_{ij}(t) \geq 0 \quad \forall j \in \mathcal{V}_i(t), \quad \sum_{j \in \mathcal{V}_i(t)} a_{ij}(t) = 1. \quad (9)$$

Затем, агент обновляет свое значение следующим образом:

$$x_i(t+1) = \sum_{j \in \mathcal{V}_i(t)} a_{ij}(t) x_j(t), \quad t = 0, 1, \dots, \quad (10)$$

Доопределяя  $a_{ij}(t) \stackrel{\Delta}{=} 0$  при  $j \notin \mathcal{V}_i(t)$ , мы получаем *стохастическую* матрицу усредняющих весов  $A(t) = (a_{ij})_{i,j \in \mathcal{V}}$ . Удобно переписать (10) в компактной матричной форме:

$$x(t+1) = A(t)x(t), \quad t = 0, 1, \dots, \quad (11)$$

где  $x(t)$  обозначает вектор-столбец, составленный из состояний агентов  $x_i(t)$ .

В случае постоянной матрицы  $A$  поведение системы (11), очевидно, определяется структурой матрицы  $A$ , хорошо изученной в теории марковских цепей. В приложениях, однако, принципиально важно иметь возможность работать с переменной матрицей, поскольку каналы коммуникации между агентами могут открываться и закрываться, что приводит к изменению отношений соседства (множеств  $\mathcal{V}_i$ ) и, как следствие, необходимости перераспределять веса. Поведение систем с переменной матрицей, или бесконечными матричными произведениями, в общем случае остается мало изученным. Основные результаты можно найти в монографии [62] и обзоре [59], а также недавних работах [63]–[65]. Мы приведем ниже лишь один результат, который будет использован в дальнейшем.

Поскольку на каждом шаге агент приближает свое состояние к состояниям соседей, то можно ожидать, что в итоге все состояния окажутся одинаковыми, иными словами, будет достигнут консенсус:

$$\bar{x}_1 = \dots = \bar{x}_n, \text{ где } \bar{x}_i \stackrel{\Delta}{=} \lim_{t \rightarrow \infty} x_i(t) \quad \forall i \in \mathcal{V}. \quad (12)$$

Иногда консенсус определяется в формально более слабом смысле, например, как асимптотическая синхронизация значений агентов:

$$\lim_{t \rightarrow \infty} \max_{i,j \in \mathcal{V}} |x_j(t) - x_i(t)| = 0. \quad (13)$$

Отметим, что в нашем случае для алгоритма (11) последнее требование эквивалентно “сильному” консенсусу. Этот факт известен из литературы по теории цепей Маркова (см. [62], теорема 4.17) (синхронизацией в этом случае является слабая эргодичность обратного матричного произведения, что эквивалентно сильной эргодичности, влекущей за собой консенсус). Связано это с тем, что максимальный (соответственно, минимальный) элемент вектора  $x(t)$  в силу уравнений (10) не возрастает (соответственно, не убывает) с ростом  $t$ , сходясь к конечному пределу. Поэтому нетрудно видеть, что

$$\lim_{t \rightarrow \infty} \max_{i,j \in \mathcal{V}} |x_j(t) - x_i(t)| = \lim_{t \rightarrow \infty} \max_i x_i(t) - \lim_{t \rightarrow \infty} \min_i x_i(t),$$

а следовательно, (13) эквивалентно тому, что  $\lim_{t \rightarrow \infty} \max_i x_i(t) = \lim_{t \rightarrow \infty} \min_i x_i(t)$ , т.е. все элементы  $x(t)$  сходятся к одному и тому же конечному значению.

Всюду далее мы называем (10) (или (11)) *алгоритмом итеративного усреднения*, заданного последовательностью стохастических матриц  $A(t)$ . Подчеркнем, что хотя матрицы  $A(t)$  “скрывают” в себе структуру исходных множеств  $\mathcal{V}_i(t)$  (отношений “соседства”), в действительности именно эти множества являются первичными, поскольку они определяют информационную структуру многоагентного алгоритма (кто и кому пересылает свое состояние), в то время как веса усреднения играют существенно меньшую роль в рассматриваемой ниже задаче консенсуса, хотя и не могут быть выбраны полностью произвольно. Вместо матрицы часто говорят о (нагруженном) графе влияния между агентами. В следующем пункте мы приводим определения из теории графов.

**3.1.2. Графы влияния и связанные понятия.**

**Определение 2.** *Граф* (формально, направленный или ориентированный граф)  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  – это пара двух конечных множеств  $\mathcal{V}$  (могут быть произвольными) и  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ . Элементы  $\mathcal{V}$  называются вершинами графа, а элементы  $\mathcal{E}$  – ребрами. Обычно предполагается, что вершины индексируются от 1 до  $n$ , так что  $\mathcal{V} = [1 : n]$ . Ребро  $(i, j)$  соединяет вершину  $i$  с вершиной  $j$  (и также обозначается как  $i \rightarrow j$ ). Во всех графах, рассматриваемых ниже, в каждой вершине имеется петля, т.е. ребро  $(i, i)$ .

Следующие два понятия относятся к типам связности (направленных) графов.

**Определение 3.** Последовательность ребер  $v_0 \rightarrow v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_s$  называется маршрутом, соединяющим вершину  $v_0$  с вершиной  $v_s$ , а число ребер в маршруте  $s$  называется его *длиной*. Граф называется *сильно связным*, если каждые две разные вершины соединены маршрутом. Граф называется *квазисильно связным*, если существуют маршруты из некоторой вершины (называемой корнем) во все остальные вершины.

Отметим, что квазисильно связный граф имеет исходящее *остовное дерево* (spanning tree) (см. [66]), и наоборот, наличие такового в графе влечет квазисильную связность. Поэтому часто в критериях консенсуса квазисильная связность формулируется как существование остовного дерева в графе (см. [67], [68]).

Следуя традиции теории многоагентных систем, мы будем рассматривать графы, вершины которых ассоциируются с агентами, а ребро  $j \rightarrow i$  ассоциируется с влиянием агента  $j$  на агента  $i$  в заданный момент времени. Применительно к алгоритму (10), влияние означает, что состояние агента  $j$  участвует в формировании следующего состояния агента  $i$ , т.е.  $a_{ij}(t) \neq 0$ . Соответственно, вес  $a_{ij}(t)$  можно трактовать как силу влияния или вес соответствующего ребра в графе. Это мотивирует следующее определение.

**Определение 4.** Для заданной квадратной матрицы  $A = (a_{ij})_{i,j \in \mathcal{V}}$  определим ассоциированный граф  $(\mathcal{V}, \mathcal{E}[A])$ , вершины которого находятся в взаимно однозначном соответствии с индексным множеством  $\mathcal{V}$  и  $\mathcal{E}[A] = (j, i) : a_{ij} \neq 0$ . Тройка  $\mathcal{G}[A] \triangleq (\mathcal{V}, \mathcal{E}[A], A)$  в дальнейшем называется *нагруженным графом*, определяемым матрицей  $A$ .

**3.1.3. Предположения о весах влияния.** Несмотря на то что выбор матрицы весов, на самом деле, имеет существенно меньшую роль, чем структура графа влияния, консенсус не может быть достигнут при произвольном выборе весов. Например, циклический алгоритм

$$x_1(t+1) = x_2(t), \quad x_2(t+1) = x_3(t), \quad \dots, \quad x_n(t+1) = x_1(t)$$

не только не приводит к консенсусу, но и вообще не сходится в случае попарно различных начальных состояний  $x_i(0)$ , поскольку вектор  $x(t+1)$  просто получается из  $x(t)$  циклической перестановкой. Для исключения подобного поведения требуются некоторые условия, аналогичные апериодичности марковских цепей. Наиболее распространенное условие, накладываемое на матрицы в большинстве работ – это свойство равномерной положительности диагональных элементов стохастической матрицы:

$$a_{ii}(t) \geq \eta > 0 \quad \forall i \in \mathcal{V} \quad \forall t. \tag{14}$$

Помимо этого ясно, что для достижения консенсуса связь между агентами не должна полностью прерываться, более того, веса влияния не могут становиться слишком малыми с течением времени. Это легко проиллюстрировать на примере двух агентов и матрицы  $a_{11}(t) = a_{22}(t) = 1 - a(t)$ ,  $a_{12}(t) = a_{21}(t) = a(t)$ . Нетрудно проверить, что

$$x_1(t+1) - x_2(t+1) = (1 - 2a(t))(x_1(t) - x_2(t)) = \dots = (x_1(0) - x_2(0)) \prod_{s=0}^t (1 - 2a(s)).$$

Если  $|a(t)| \neq 1/2$  при всех  $t$  и  $\sum_{t \geq 0} a(t) < \infty$ , то произведение в правой части сходится к ненулевому значению, соответственно, консенсус не достигается.

В связи с этим обычно вводятся некоторые требования “усиленной” связности графа влияний, которая должна сохраняться при отбрасывании слишком слабых связей между агентами. Мы будем рассматривать лишь наиболее распространенное свойство равномерной (или повторяющейся) связности.

**Определение 5.** Если дана неотрицательная матрица  $A \in \mathbb{R}^{\mathcal{V} \times \mathcal{V}}$ , то ее  $\varepsilon$ -скелет — это матрица  $A^\varepsilon$ , полученная обнулением всех элементов, которые меньше  $\varepsilon$ , т.е.

$$a_{ij}^\varepsilon \triangleq \begin{cases} a_{ij}, & a_{ij} \geq \varepsilon, \\ 0 & \text{в противном случае.} \end{cases}$$

Нагруженный граф  $\mathcal{G}[A^\varepsilon]$  называется  $\varepsilon$ -скелетом нагруженного графа  $\mathcal{G}[A]$ . Мы говорим, что нагруженный граф (квази)сильно  $\varepsilon$ -связен, если его  $\varepsilon$ -скелет является (квази)сильно связным<sup>11</sup>.

**Определение 6.** Имея последовательность матриц  $A(t) \in \mathbb{R}^{\mathcal{V} \times \mathcal{V}}$ , где  $t \in [t_0 : t_1]$ , мы определяем объединение соответствующих графов над  $[t_0 : t_1]$  как граф

$$\bigcup_{t=t_0}^{t_1} \mathcal{G}[A(t)] \triangleq \mathcal{G}\left[\sum_{t=t_0}^{t_1} A(t)\right].$$

Графы  $\mathcal{G}[A(t)]$ ,  $t \in \mathcal{T}$ , где  $\mathcal{T} = [t_0 : t_1]$ , называются совместно (квази)сильно  $\varepsilon$ -связными на интервале  $\mathcal{T}$ , если их объединение на  $\mathcal{T}$  является (квази)сильно  $\varepsilon$ -связным.

**Определение 7.** Меняющийся во времени граф  $\mathcal{G}[A(\cdot)]$  является равномерно (квази)сильно связным, если существуют такие константы  $\varepsilon > 0$  и  $T > 0$ , что графы  $\mathcal{G}[A(t)]$  совместно (квази)сильно  $\varepsilon$ -связны в каждый момент времени  $[t_0 : t_0 + T]$ ,  $t_0 \geq 0$ .

Отметим, что часто предполагается дополнительное ограничение на веса усреднения, которое на практике обычно выполняется. Именно, считается, что ненулевые элементы матрицы весов не могут быть слишком малы:

$$a_{ij}(t) \in \{0\} \cup [\varepsilon, 1] \quad \forall t = 0, 1, \dots \quad (15)$$

При этом понятие равномерной связности просто означает, что объединение графов за достаточно длинный период  $T$  является связным (в сильном или квазисильном смысле).

**3.1.4. Достаточное условие робастного консенсуса.** При изучении многоагентных алгоритмов пересечения множеств мы будем использовать следующее условие консенсуса, которое на самом деле гарантирует робастность консенсуса к исчезающим на бесконечности возмущениям. Наряду с (11), рассмотрим возмущенную систему

$$x(t+1) = A(t)x(t) + f(t), \quad t = 0, 1, \dots \quad (16)$$

Справедлива следующая

**Лемма 2.** Пусть  $A(t)$  — стохастические матрицы с равномерно положительными диагональными элементами (14). Тогда алгоритм (11) устанавливает консенсус (12), если переменный граф  $\mathcal{G}[A(\cdot)]$  обладает равномерной квазисильной связностью. Более того, если возмущение  $f(\cdot)$  исчезает на бесконечности  $f(t) \xrightarrow{t \rightarrow \infty} 0$ , то

$$\max x(t) - \min x(t) \xrightarrow{t \rightarrow \infty} 0 \quad (17)$$

для любого решения (16).

Лемма 2 обычно доказывается при дополнительном предположении (15), в таком виде она доступна в [46]. Как показано в диссертации [28], это условие можно снять, однако доказательство становится существенно более сложным. Можно также распространить рассуждения из работы [69], доказывающей непрерывный аналог леммы 2 (наряду с более общими критериями робастного консенсуса), на случай систем с дискретным временем. Все доказательства из указанных источников, однако, достаточно сложны технически, и мы их здесь не приводим.

<sup>11</sup> Иными словами, граф является (квази)сильно  $\varepsilon$ -связным, если он остается (квази)сильно связным после удаления всех дуг, которые имеют вес меньше, чем  $\varepsilon$ .

### 3.2. Консенсус в усредняющих неравенствах

В этом пункте мы введем важный инструмент для доказательства главного результата настоящей статьи. Именно, наряду с уравнениями (11), мы будем рассматривать *рекуррентные усредняющие неравенства* следующего вида:

$$x(t+1) \leq A(t)x(t), \quad t \in \mathbb{N}. \quad (18)$$

Неравенство между векторами здесь понимается покомпонентно.

Следует отметить, что рекуррентное неравенство не может рассматриваться как алгоритм управления для группы агентов, поскольку его решение не является однозначно определенным. Однако такие неравенства представляются удобным инструментом в анализе многоагентных алгоритмов (см. [30], [61], [70]). Может показаться удивительным, что столь слабое ограничение на вектор состояния системы агентов в принципе позволяет установить какие-то свойства, однако, как будет показано, в предположении равномерной сильной связности графа может, как и в случае уравнений, быть доказан консенсус (12). В отличие от уравнений (11), решения которых ограничены, решения неравенств только полуограничены сверху. В частности, некоторые компоненты могут сходиться к  $-\infty$  при  $t \rightarrow \infty$ . Так консенсус в случае неограниченных решений означает, что  $\bar{x}_i = -\infty \forall i$ . В практических приложениях, однако, мы часто знаем априорную нижнюю границу для решения, таким образом, ситуация консенсуса в бесконечности не реализуется.

Как мы увидим, при изучении решений неравенств часто важно знать поведение вектора невязок между правой и левой частью:

$$\Delta_i(t) \triangleq \sum_j a_{ij}(t)x_j(t) - x_i(t+1) \geq 0, \quad (19)$$

в частности, сходимость этих невязок к нулю:

$$\Delta_i(t) \xrightarrow{t \rightarrow \infty} 0. \quad (20)$$

Следующая ключевая лемма гарантирует как достижение консенсуса, так и сходимость к нулю невязок для любого ограниченного решения (18).

**Лемма 3.** Пусть последовательность стохастических матриц  $A(t)$  имеет равномерно положительные диагональные элементы  $a_{ii}(t) \geq \eta > 0$ , а переменный граф  $\mathcal{G}[A(\cdot)]$  равномерно сильно связан. Тогда для любого решения неравенства (18) имеет место консенсус (12), а если решение ограничено снизу, то выполнено (20) при любом  $i \in \mathcal{V}$ .

Доказательство леммы 3 дано в приложении; специальные случаи этой леммы (предполагающие равномерную положительность всех ненулевых элементов матрицы, а не только диагональных) были доказаны в работах первого автора [30], [70]. Отметим, что требование равномерной сильной связности может быть существенно ослаблено, наиболее общий результат приведен в диссертации [28].

### 3.3. Консенсусные алгоритмы для пересечения выпуклых множеств

Рассмотрим несколько классов алгоритмов, предложенных в литературе для решения многоагентной задачи о пересечении выпуклых множеств. Как будет показано, техника усредняющих неравенств позволяет достаточно просто доказать сходимость состояния всех агентов в каждом из указанных алгоритмов к некоторой консенсусной точке, лежащей в пересечении множеств.

Именно, предположим, что выпуклое замкнутое множество  $\Xi_i \subset \mathbb{R}^d$ , связанное с агентом  $i \in \mathcal{V}$ , является множеством неподвижных точек фейеровского относительно некоторой нормы  $\|\cdot\|$  отображения  $P_i$  (в наиболее распространенном случае,  $P_i$  – проектор на множество  $\Xi_i$ , а норма евклидова). Предположим, что агент  $i$  способен вычислять знач.  $P_i(\xi)$  в произвольной заданной точке  $\xi$ .

**3.3.1. Классы консенсусных алгоритмов.** Мы покажем, что при выполнении некоторых ограничений на граф, точка, принадлежащая  $\Xi_*$ , может быть вычислена одним из следующих алгоритмов, основанных на последовательном усреднении с одновременным применением фейеровских операторов:

$$\xi^i(t+1) = P_i \left[ \sum_{j \in \mathcal{V}} a_{ij}(t) \xi^j(t) \right], \quad i \in \mathcal{V}, \quad (21)$$

$$\xi^i(t+1) = P_i \left[ \sum_{j \in \mathcal{V}} a_{ij}(t) P_j(\xi^j(t)) \right], \quad i \in \mathcal{V}, \quad (22)$$

$$\xi^i(t+1) = \sum_{j \in \mathcal{V}} a_{ij}(t) P_j(\xi^j(t)), \quad i \in \mathcal{V}, \quad (23)$$

$$\xi^i(t+1) = a_{ii}(t) P_i(\xi^i(t)) + \sum_{j \neq i} a_{ij}(t) \xi^j(t), \quad i \in \mathcal{V}. \quad (24)$$

Во всех указанных алгоритмах, как и прежде,  $A(t) = (a_{ij}(t))$  – стохастические матрицы. Отметим, что в вырожденном случае, когда у агентов нет ограничений ( $\Xi_i = \mathbb{R}^d$  и  $P_i = \text{Id}$ ) алгоритмы (21)–(24) превращаются в стандартную процедуру последовательного усреднения, применяемую для достижения консенсуса (см. [54], [59], [71]) (с единственной разницей, что состояние агента является многомерным вектором, а не скаляром – это, очевидно, никоим образом не меняет условий сходимости алгоритма). Вместе с тем, в общем случае условие сходимости существенно более ограничительно: требуется равномерная сильная связность, а не квазисильная, как в лемме 2. Это ограничение на граф накладывается во всех известных авторам работах. Метод доказательства сходимости, изложенный в настоящей статье, сводит каждый из алгоритмов к системе усредняющих неравенств, сильная связность требуется для применения леммы 3.

Происхождение и особенности алгоритмов (21)–(24). Алгоритм (21) был предложен в основополагающей статье [25] для случая проекционных операторов  $P_i$ , посвященной проблемам распределенной оптимизации, а затем развит в [46], где были ослаблены некоторые предположения (в частности, требование симметричности матрицы  $a_{ij}$ ) и исследована робастность алгоритма [25] к запаздываниям<sup>12</sup>. В более поздней работе [50] алгоритм (21) был предложен для поиска общей точки фейеровских<sup>13</sup> отображений  $P_i$ . Существенной особенностью данного алгоритма является то, что агент сначала получает состояния соседей, а затем, вычисляет  $P_i$  от усредненного значения. Отметим, что если  $P_i$  – проектирование на  $\Xi_i$ , то состояние агента  $i$  не покидает множества  $\Xi_i$  при  $t \geq 0$ . В этом случае агент стремится сблизить свое состояние с состояниями соседей, не нарушая собственное ограничение.

Алгоритм (24) был изначально предложен [72] для решения систем линейных уравнений, однако работает и для более общего случая. В этом случае  $\Xi_i$  – линейные гиперплоскости, а операторы проектирования  $P_i$  можно вычислить аналитически. Как и в алгоритме (21), агенты обмениваются своими состояниями. Однако имеются два принципиальных отличия: во-первых, агент может вычислять значение  $P_i$ , не дожидаясь получения соседних измерений (на практике, это может привести к ускорению вычислений), а во-вторых, состояние агента  $i$ , как правило, не принадлежит  $\Xi_i$  ни на одном шаге.

Алгоритм (23) аналогичен по структуре алгоритму Чиммино, с той разницей, что каждый агент, во-первых, имеет свое приближение к точке из пересечения множеств  $\xi^i$ , а во-вторых, вместо барицентра всех проекций вычисляется некоторая выпуклая комбинация проекций соседних агентов. Близкие по структуре распределенные алгоритмы изучались в [17], однако в этой работе не вводились явно агенты, и не рассматривался случай переменного графа. Отметим, что в случае этого алгоритма агенты не только не раскрывают соседям информацию о своем множестве, но даже не обмениваются состояниями (это может быть существенно для защиты информации).

Наконец, алгоритм (22) был предложен в работе [73], также посвященной алгоритмам решения алгебраических уравнений. Как и в алгоритме (23), агенты обмениваются значениями  $P_i$ , однако затем агент  $i$  повторно вычисляет значение  $P_i$ . В случае проекций, как уже обсуждалось, это гарантирует, что состояние агента  $i$  не покидает желаемого множества. Как будет показано ниже, это также позволяет избежать некоторых патологических точек равновесия, возникающих в системе (23) (при пустом пересечении множеств  $\Xi_* \neq \emptyset$ ).

**3.3.2. Теорема о консенсусе с ограничениями.** Предположим, что  $\Xi_* = \bigcap \Xi_i = \bigcap \mathcal{F}(P_i)$  непусто. Будем говорить, что установлен консенсус с ограничениями (см. [25]), если при произвольных начальных состояниях агентов существуют и равны между собой следующие пределы:

$$\lim_{t \rightarrow \infty} \xi^1(t) = \dots = \lim_{t \rightarrow \infty} \xi^n(t) \in \Xi_*. \quad (25)$$

<sup>12</sup> Как вытекает из результатов [28], [30], на самом деле к коммуникационным запаздываниям робастны все алгоритмы (21)–(24), поскольку такие запаздывания не нарушают свойств усредняющих алгоритмов и неравенств. В настоящем обзоре мы не рассматриваем вопросы робастности, которые в основном изучаются в рамках теории управления многоагентными системами и представляют интерес, если агенты ограничены в возможности быстро передавать информацию (например, связаны через Интернет).

<sup>13</sup> Отметим, что [50] накладывает ряд ограничений, в частности, свойство Фейера должно выполняться относительно  $\ell_p$  нормы на  $\mathbb{R}^d$ ; как мы увидим ниже, это ограничение можно снять.

Анализ алгоритмов (21)–(24) опирается на свойства сходимости усредняющих неравенств (18) и лемму 2 о робастности консенсуса. Сформулируем основной результат, который будет доказан в приложении.

**Теорема 1.** *Предположим, что отображения  $P_i : \mathbb{R}^d \rightarrow \mathbb{R}^d$  являются фейеровскими относительно некоторой общей нормы  $\|\cdot\|$  и имеют по крайней мере одну общую неподвижную точку, т.е.  $\Xi_* = \bigcap_{i \in \mathcal{V}} \mathcal{F}(P_i) \neq \emptyset$ . Пусть последовательность стохастических матриц  $A(t)$  имеет равномерно положительные диагональные элементы  $a_{ii}(t) \geq \eta > 0$ , и переменный граф  $G[A(\cdot)]$ , соответствующий  $A(t)$ , равномерно сильно связан. Тогда каждый из алгоритмов (21)–(24) находит общую неподвижную точку  $\{P_i\}$ , т.е. состояния агентов  $\xi^i$  сходятся к консенсусу (25).*

Отметим что теорема 1 в ее полной общности публикуется впервые. Сходимость алгоритмов (21), (22) и (24) с помощью рекуррентных усредняющих неравенств впервые была доказана в [30] при дополнительном предположении (15), которое впоследствии было снято в [28]. Исследование алгоритма (23), однако, проведено не было.

**3.3.3. Замечания к теореме 1.** Отметим, что в условиях отсутствия общих точек у множеств  $\Xi_i$  сходимость векторов  $\xi^i(t)$  к общему значению  $\xi_*$  при  $t \rightarrow \infty$  для алгоритмов (21), (22) и (24), очевидно, невозможна. Для алгоритмов (21) и (22) это так, поскольку  $\xi^i(t) \in \Xi_i$  при любом  $i \in \mathcal{V}$ , таким образом, общий предел автоматический является общей точкой всех  $\Xi_i$ . Для (24) это следует из непрерывности  $P_i$ : предполагая, что  $\xi^i(t) \xrightarrow{t \rightarrow \infty} \xi_*$ , мы имели бы

$$a_{ii}(t) \left[ P_i(\xi^i(t)) - \xi^i(t) \right] = \xi^i(t+1) - \sum_{j \in \mathcal{V}} a_{ij}(t) \xi^j(t) \xrightarrow{t \rightarrow \infty} 0,$$

что в силу равномерной положительности  $a_{ii}$  вновь означает, что  $\xi_* = P_i(\xi_*)$  при всех  $i$ , т.е.  $\xi_*$  – общая неподвижная точка семейства операторов. Алгоритм (23), однако, является исключением: консенсус при этом алгоритме возможен, теоретически, и при пустом пересечении множеств  $\Xi_i$ . Рассмотрим, к примеру, систему из двух агентов с множествами ограничений  $\Xi_1 = \{-1\}$ ,  $\Xi_2 = \{1\}$  и весовыми коэффициентами  $a_{11} = a_{12} = a_{21} = a_{22} = 1/2$ . Очевидно, что  $\xi^1(t) \equiv \xi^2(t) \equiv 0$  – точка равновесия системы (23), не принадлежащая ни одному из множеств  $\Xi_i$ .

Вместе с тем, в случае, когда консенсус не достигается, вопрос о поведении траекторий системы является, насколько известно авторам, открытым. Неизвестно в частности, даже при постоянной матрице весов, всегда ли в этом случае имеет место сходимость к периодической или квазипериодической траектории. Более того, даже ограниченность решений, как видно из доказательства в приложении, опирается на непустоту  $\Xi_* = \bigcap_i \Xi_i$  и, за исключением специальных случаев (например,  $P_i$  – проектор на ограниченное множество  $\Xi_i$ ), вообще говоря, не является очевидным свойством.

Следует отметить также, что скорость сходимости алгоритмов (21)–(24) достаточно трудно оценить явно даже в случае постоянного графа. В случае же переменного графа, даже точная скорость сходимости обычных консенсусных алгоритмов ( $P_i = \text{Id}$ ) до конца не исследована; ряд оценок на скорость сходимости может быть найден в обзоре [74]. Как показано в недавней работе [75], оценка скорости сходимости алгоритмов консенсуса сводится к вычислению наибольшего коэффициента эргодичности для некоторого компактного множества стохастических матриц. Существуют “ускоренные” модификации алгоритма (21) для решения систем линейных уравнений (см. [27], [73]), для которых доказана экспоненциальная сходимость к консенсусу. Данные алгоритмы не были исследованы в случае произвольных фейеровских операторов, а их анализ требует техники, выходящей за рамки данного обзора.

В настоящей статье мы не рассматриваем стохастические варианты алгоритма пересечения выпуклых множеств, в частности, рандомизированную версию алгоритма (24) из работы [76], а также рандомизированные модификации алгоритма Качмаржа (см., например, обзор литературы в [77]). Следует отметить, что для некоторых из рассмотренных выше алгоритмов существуют аналоги с непрерывным временем, например, в работе [26] приведен следующий аналог алгоритма (24):

$$\dot{\xi}^i(t) = \sum_{j \in \mathcal{V}} a_{ij}(t) (\xi^j(t) - \xi^i(t)) + P_i(\xi^i(t)) - \xi^i(t). \tag{26}$$

Исследование данного алгоритма возможно с помощью теории усредняющих неравенств в непрерывном времени (см. [28], [29]); мы не приводим соответствующие результаты, поскольку они полностью аналогичны критериям сходимости в дискретном времени.

## 4. ЗАКЛЮЧЕНИЕ

Авторы благодарны редакторам специального выпуска, посвященного Б. Т. Поляку, за возможность представить данный обзор, который представляет расширенный материал лекции, прочитанной первым автором на Традиционной летней школе по управлению и оптимизации им. Б. Т. Поляка летом 2023 г. В обзоре представлена история задачи о пересечении выпуклых множеств в пространстве, а также недавние результаты первого автора, относящиеся к многоагентной постановке этой задачи и децентрализованным алгоритмам для ее решения. Борис Теодорович Поляк считал статью [15], в которой был получен ряд важных результатов об экспоненциальной сходимости проекционных алгоритмов, одной из важнейших своих работ. Об этой статье он, в частности, рассказывал на семинаре ИПУ РАН, организованном в честь присуждения ему премии им. Хачияна<sup>14</sup>. Согласно базе данных Google Scholar, английская версия данной статьи к настоящему моменту процитирована более 1000 раз.

Следует отметить, что хотя с доказательства первой теоремы фон Неймана, устанавливающей сходимость метода последовательных проекций для двух линейных подпространств, прошло уже 90 лет, теория проекционных алгоритмов далека от своего завершения. Это относится как к классической (централизованной) версии задачи, так и многоагентной (децентрализованной) постановке. Как уже было сказано, достаточно много вопросов связано со скоростью сходимости алгоритмов, в частности, точная оценка скорости неизвестна даже для алгоритма Качмаржа. При замене проекционных операторов фейеровскими отображениями, ассоциированными с множествами, оценки скорости сходимости удается получить лишь в специальных случаях (см. [44]). В случае децентрализованного алгоритма, даже для консенсуса без ограничений (когда все множества совпадают с пространством) оценка скорости сходимости состояний агента к консенсусному значению является нетривиальной задачей (см. [74]).

Другим открытым вопросом является нахождения минимальных (необходимых и достаточных) условий связности графа. Как показано в диссертации [28], равномерная связность в леммах 2 и 3 может быть существенно ослаблена, однако приведенное условие все равно является только достаточным, а не необходимым, кроме того, его проверка для произвольного переменного графа затруднительна.

Перспективным направлением исследований представляется также развитие проекционных методов для многообразий (см. [24]) на многоагентный случай. Данная задача связана с вопросом о сходимости консенсусных алгоритмов на многообразиях (см. [78]).

## ПРИЛОЖЕНИЕ

**1. Доказательство леммы 3.** Предположим, что все условия леммы 3 выполнены, в частности,  $a_{ii}(t) \geq \eta$ . Рассмотрим сначала ситуацию, когда условие равномерной связности выполнено с  $T = 1$ , т.е. все графы  $\mathcal{G}[A(t)]$  являются сильно  $\varepsilon$ -связными. Обозначим число агентов  $n = |\mathcal{V}|$  и, при каждом  $t = 0, 1, \dots$  занумеруем элементы  $\mathcal{V}$  в порядке возрастания компонент  $x(t)$ :

$$\mathcal{V} = \{\sigma_1(t), \dots, \sigma_N(t)\},$$

$$y_1(t) \triangleq x_{\sigma_1(t)}(t) \leq y_2(t) \triangleq x_{\sigma_2(t)}(t) \leq \dots \leq y_n(t) \triangleq x_{\sigma_n(t)}(t).$$

Мы докажем следующее вспомогательное неравенство:

$$y_{k+1}(t+1) \leq (1 - \varepsilon_0)y_n(t) + \varepsilon_0 y_k(t) \quad \forall k = 0, 1, \dots, n-1, \quad (27)$$

где  $\varepsilon_0 = \min(\varepsilon, \eta)$ . Действительно, обозначим  $I \triangleq \{\sigma_{k+1}(t), \dots, \sigma_n(t)\}$ . Так как  $\mathcal{G}[A(t)]$  – сильно  $\varepsilon$ -связный граф, то существует  $i \in I$  и  $m \in J \triangleq \mathcal{V} \setminus I$  такие, что  $a_{im}(t) \geq \varepsilon \geq \varepsilon_0$ . Следовательно,

$$x_i(t+1) \leq a_{im}(t) \underbrace{x_m(t)}_{\leq y_k(t)} + \sum_{j \in \mathcal{V} \setminus \{m\}} a_{ij}(t) \underbrace{x_j(t)}_{\leq y_n(t)} \leq (1 - \varepsilon_0)y_n(t) + \varepsilon_0 y_k(t).$$

С другой стороны, используя неравенство  $a_{ii}(t) \geq \eta \geq \varepsilon_0$ , для каждого  $j \in J$  мы имеем

$$x_j(t+1) \leq a_{jj}(t) \underbrace{x_j(t)}_{\leq y_k(t)} + \sum_{\ell \in \mathcal{V} \setminus \{j\}} a_{j\ell}(t) \underbrace{x_\ell(t)}_{\leq y_m(t)} \leq (1 - \varepsilon_0)y_n(t) + \varepsilon_0 y_k(t).$$

Следовательно, все компоненты  $x$  с индексами из множества  $J \cup \{i\}$  не больше правой части из (27). Для доказательства (27) остается заметить, что мощность последнего множества равна  $k + 1$ .

<sup>14</sup> Видеозапись семинара доступна по ссылке [https://www.youtube.com/watch?v=\\_6p5qQ15fpw](https://www.youtube.com/watch?v=_6p5qQ15fpw)

Поскольку  $y_n(t) = \max x(t)$  не возрастает, существует предел  $y_* = \lim_{t \rightarrow \infty} y_n(t) \geq -\infty$ . Если  $y_* = -\infty$ , то, очевидно,  $x_i(t) \xrightarrow{t \rightarrow \infty} -\infty$ . Иначе, мы можем использовать (27) для доказательства того, что  $y_k(t) \xrightarrow{t \rightarrow \infty} y_* > -\infty$ , используя обратную индукцию по  $k = n, n-1, \dots, 1$ , иными словами, достигается консенсус (12).

Случай произвольного периода  $T > 1$  сводится к уже рассмотренному случаю: достаточно заметить, что в силу равномерной положительности диагональных элементов графы матриц

$$\tilde{A}(k) = A(kT)A(kT+1)\dots A(kT+T-1)$$

являются сильно  $\tilde{\varepsilon}$ -связными при некотором  $\tilde{\varepsilon} = \tilde{\varepsilon}(\varepsilon, \eta) > 0$ . Поэтому сходимость к консенсусу имеет место для подпоследовательности  $\tilde{x}(k) = x(kT)$ , удовлетворяющей рекуррентному неравенству

$$\tilde{x}(k+1) \leq \tilde{A}(k)\tilde{x}(k).$$

Поскольку для любого  $\ell = 1, 2, \dots, T-1$  справедливы неравенства

$$\begin{aligned} \tilde{x}(k+1) &= x(kT+T) \leq A(kT+T-1)\dots A(kT+\ell)x(kT+\ell), \\ x(kT+\ell) &= A(kT+\ell-1)\dots A(kT)\tilde{x}(k), \end{aligned}$$

то из сходимости  $\tilde{x}(k) \xrightarrow{k \rightarrow \infty} \tilde{c}1_n$  (где  $\tilde{c} \in \mathbb{R}$  и  $1_n$  обозначает вектор-столбец из единиц размерности  $n$ ) и стохастичности матриц  $A(t)$  вытекает сходимость  $x(kT+\ell) \xrightarrow{k \rightarrow \infty} \tilde{c}1_n$  при любом  $\ell = 1, 2, \dots, T-1$ , следовательно, достигается консенсус между агентами.

Наконец, свойство (20) для ограниченных решений также очевидно из того, что матрицы  $A(t)$  стохастические (и, в частности, равномерно ограничены): если  $x(t) \xrightarrow{t \rightarrow \infty} c1_n$  (при некотором  $c \in \mathbb{R}$ ), то  $A(t)1_n = 1_n$ , а следовательно,

$$A(t)x(t) = A(t)(x(t) - c1_n) + c1_n \xrightarrow{t \rightarrow \infty} c1_n.$$

## 2. Доказательство леммы 1 и теоремы 1.

### Лемма о фейеровском отображении.

Доказательства теоремы и леммы используют следующий технический результат.

**Лемма 4.** Пусть  $P$  – фейеровский оператор в некоторой норме  $\|\cdot\|$ , и  $\xi^0$  – его неподвижная точка. Обозначим  $d(\xi) \triangleq \|\xi - \xi^0\| - \|P\xi - \xi^0\| \geq 0$  и рассмотрим ограниченную последовательность векторов  $\xi(t)$  такую, что  $d(\xi(t)) \xrightarrow{t \rightarrow \infty} 0$ . Тогда

$$\|P\xi(t) - \xi(t)\| \xrightarrow{t \rightarrow \infty} 0.$$

**Доказательство.** Предположим противное:  $\|P\xi(t_r) - \xi(t_r)\| \geq \varepsilon$  для любого  $r \in \mathbb{N}$ ,  $\varepsilon > 0$ , и последовательности  $t_r \rightarrow \infty$ . Переходя к меньшей подпоследовательности, можно предположить, не умаляя общности, что векторы  $\xi(t_r)$  сходятся к пределу  $\xi_* \in \mathbb{R}^d$ . Поскольку  $P$  – непрерывное отображение, имеем  $\|P\xi_* - \xi_*\| \geq \varepsilon$  при  $d(\xi_*) = 0$ . Пришли к противоречию с (7), так как  $\xi_* \notin \mathcal{F}(P)$ , в то время как  $\|P\xi_* - \xi_*\| = \|\xi_* - \xi^0\|$ . Лемма 4 доказана.

**Доказательство леммы 1.** Очевидно, что последовательность  $\xi(t)$  из леммы 1 ограничена в силу того, что расстояния от  $\xi(t)$  до любой точки из  $M$  не возрастают. Невозрастающая последовательность

$$a(t) \triangleq \|\xi(t) - \xi^0\| \geq 0$$

для произвольного элемента  $\xi^0 \in M$  имеет конечный предел при  $t \rightarrow \infty$ . В частности, поскольку  $P\xi(t) = \xi(t+1)$ , то

$$a(t) - a(t+1) = \|\xi(t) - \xi^0\| - \|P\xi(t) - \xi^0\| = d(\xi(t)) \xrightarrow{t \rightarrow \infty} 0.$$

В силу леммы 4, мы доказали, что  $\|P\xi(t) - \xi(t)\| \xrightarrow{t \rightarrow \infty} 0$ .

Отметим теперь, что в силу ограниченности, существует подпоследовательность  $\xi(t_k)$ , сходящаяся к некоторому элементу  $\xi_* \in \mathbb{R}^d$ . Очевидно, что



$$\|P(\xi_*) - \xi_*\| = \lim_{k \rightarrow \infty} \|P(\xi(t_k)) - \xi(t_k)\| = 0,$$

т.е.  $\xi_* \in M$ . Поскольку  $\xi^0$  в рассуждении выше произвольно, мы можем выбрать  $\xi^0 = \xi^*$ . Поскольку при  $t \geq t_k$ , мы имеем

$$0 \leq \|\xi(t) - \xi_*\| \leq \|\xi(t_k) - \xi_*\|,$$

где правая часть может быть сделана сколь угодно малой при большом  $k$ , то  $\|\xi(t) - \xi_*\| \xrightarrow{t \rightarrow \infty} 0$ . Лемма 1 доказана.

**Доказательство теоремы 1.** Сначала введем некоторые вспомогательные обозначения. Выберем произвольную точку  $\xi^0 \in \Xi_*$  и положим, по определению,  $\delta_i(\xi) \triangleq \|\xi - \xi^0\| - \|P_i \xi - \xi^0\| \geq 0$ . Пусть  $\zeta^i(t) = \sum_{j \in \mathcal{V}} a_{ij}(t) \xi^j(t)$ . Центральная идея доказательства состоит в исследовании свойств векторов  $x(t) = (x_i(t))_{i \in \mathcal{V}}$ , компоненты которых  $x_i(t) \triangleq \|\zeta^i(t) - \xi^0\|$  обозначают расстояния от состояний агентов  $\xi^i(t)$  до выбранной неподвижной точки.

**Шаг 1.** Покажем сначала, что для каждого из указанных в теореме алгоритмов векторы  $x(t)$  удовлетворяют усредняющему рекуррентному неравенству (18).

В случае алгоритма (21) имеем

$$\begin{aligned} x_i(t+1) &= \|P_i(\zeta^i(t)) - \xi^0\| \stackrel{(8)}{\leq} \|\zeta^i(t) - \xi^0\| = \\ &= \sum_j a_{ij}(t) (\xi^j(t) - \xi^0) \leq \sum_j a_{ij}(t) \|\xi^j(t) - \xi^0\| = \sum_j a_{ij}(t) x_j(t). \end{aligned} \tag{28}$$

Аналогично рассматривается случай (22). Обозначив  $\bar{\zeta}^i(t) \triangleq \sum_{j \in \mathcal{V}} a_{ij}(t) P_j(\xi^j(t))$ , имеем

$$\begin{aligned} x_i(t+1) &= \|P_i(\bar{\zeta}^i(t)) - \xi^0\| \stackrel{(8)}{\leq} \|\bar{\zeta}^i(t) - \xi^0\| = \\ &= \sum_j a_{ij}(t) (P_j(\xi^j(t)) - \xi^0) \leq \sum_j a_{ij}(t) \|P_j(\xi^j(t)) - \xi^0\| \stackrel{(8)}{\leq} \\ &\leq \sum_j a_{ij}(t) \|\xi^j(t) - \xi^0\| = \sum_j a_{ij}(t) x_j(t). \end{aligned} \tag{29}$$

В случае алгоритма (23) нетрудно видеть, что

$$\begin{aligned} x_i(t+1) &= \sum_j a_{ij}(t) (P_j(\xi^j(t)) - \xi^0) \leq \sum_j a_{ij}(t) \|P_j(\xi^j(t)) - \xi^0\| \stackrel{(8)}{\leq} \\ &\leq \sum_j a_{ij}(t) \|\xi^j(t) - \xi^0\| = \sum_j a_{ij}(t) x_j(t). \end{aligned} \tag{30}$$

В случае алгоритма (24) имеем

$$\begin{aligned} x_i(t+1) &= \left\| a_{ii}(t) (P_i(\xi^i(t)) - \xi^0) + \sum_{j \neq i} a_{ij}(t) (\xi^j(t) - \xi^0) \right\| \leq \\ &\leq a_{ii}(t) \|P_i(\xi^i(t)) - \xi^0\| + \sum_{j \neq i} a_{ij}(t) \|\xi^j(t) - \xi^0\| \stackrel{(8)}{\leq} \\ &\leq a_{ii}(t) \|\xi^i(t) - \xi^0\| + \sum_{j \neq i} a_{ij}(t) x_j(t) = \sum_{j \in \mathcal{V}} a_{ij}(t) x_j(t). \end{aligned} \tag{31}$$

Заметим, что  $x_i(t)$  неотрицательны в силу определения. В силу теоремы 3 усредняющее неравенство (18) устанавливает консенсус, т.е.  $x_i(t) \xrightarrow{t \rightarrow \infty} c \geq 0 \forall i$ , где  $c$  зависит от конкретного решения. Из теоремы 3 также следует, что

$$\Delta_i(t) \xrightarrow[t \rightarrow \infty]{} 0 \quad \forall i \in \mathcal{V}, \tag{32}$$

где невязки  $\Delta_i(t)$  определены формулой (19).

**Шаг 2.** Исследуя структуру невязок  $\Delta_i$ , покажем, что

$$\|\xi^i(t+1) - \zeta^i(t)\| \xrightarrow[t \rightarrow \infty]{} 0 \quad \forall i, \tag{33}$$

$$\|P_i(\xi^i(t)) - \xi^i(t)\| \xrightarrow[t \rightarrow \infty]{} 0 \quad \forall i. \tag{34}$$

Векторы  $\xi^i(t)$  (а значит, и  $\zeta^i(t)$ ) ограничены, так как расстояния  $x_i(t)$  от них до  $\xi^0$  ограничены сверху в силу (18). Для алгоритма (21) из неравенства (28) следует, что

$$\Delta_i(t) \geq \|\zeta^i(t) - \xi^0\| - \|P_i(\zeta^i(t)) - \xi^0\| = \delta_i(\zeta^i(t)). \tag{35}$$

Применяя лемму 4 к  $P = P_i$  из (32) получаем  $\|P_i(\zeta^i(t)) - \zeta^i(t)\| \xrightarrow[t \rightarrow \infty]{} 0$ , что эквивалентно (33), поскольку  $\xi^i(t+1) = P_i(\zeta^i(t))$ . Чтобы вывести (34), заметим, что

$$0 \leq \|P_i(\xi^i(t+1)) - \xi^i(t+1)\| \leq \|P_i(\zeta^i(t)) - \zeta^i(t)\| + \|\zeta^i(t) - \xi^i(t+1)\| \xrightarrow[t \rightarrow \infty]{} 0.$$

В случае алгоритма (24), заметим, что (31) дает

$$\Delta_i(t) \geq a_{ii}(t) \left( \|\xi^i(t) - \xi^0\| - \|P_i(\xi^i(t)) - \xi^0\| \right) \geq \eta \delta_i(\xi^i(t)), \tag{36}$$

где  $\eta > 0$  – константа из предположения теоремы. Применяя лемму 4, можно доказать (34), что влечет за собой также (33), поскольку

$$\xi^i(t+1) = \zeta^i(t) + a_{ii}(t) (P_i(\xi^i(t)) - \xi^i(t)).$$

Доказательство в случае алгоритма (22) объединяет две вышеупомянутые оценки. Во-первых, (29) приводит к неравенству

$$\Delta_i(t) \geq \|\bar{\zeta}^i(t) - \xi^0\| - \|P_i(\bar{\zeta}^i(t)) - \xi^0\| = \delta_i(\bar{\zeta}^i(t)), \tag{37}$$

которое аналогично (35) и, поскольку  $P_i(\bar{\zeta}^i(t)) = \xi^i(t+1)$ , влечет

$$\|\xi^i(t+1) - \bar{\zeta}^i(t)\| \xrightarrow[t \rightarrow \infty]{} 0. \tag{38}$$

Во-вторых, (29) также влечет (36), что, в свою очередь, влечет (34). Таким образом,

$$\|\bar{\zeta}^i(t) - \zeta^i(t)\| \xrightarrow[t \rightarrow \infty]{} 0. \tag{39}$$

Это доказывает (33), так как  $\|\xi^i(t+1) - \zeta^i(t)\| \leq \|\xi^i(t+1) - \bar{\zeta}^i(t)\| + \|\bar{\zeta}^i(t) - \zeta^i(t)\|$ .

Доказательство для алгоритма (23) аналогично доказательству для алгоритма (22). В этом случае (38) выполнено по определению  $\bar{\zeta}^i$  (левая часть равна 0), а неравенство (36) легко следует из (30), поскольку

$$a_{ij}(t) \left( \|\xi^j(t) - \xi^0\| - \|P_j(\xi^j(t)) - \xi^0\| \right) \geq 0 \quad \forall j \neq i.$$

По аналогии с алгоритмом (22) можно проверить выполнение уравнений (34) и (33).

Таким образом, мы доказали, что при всех алгоритмах, изучаемых в теореме, решения обладают тем свойством, что

$$f^i(t) \stackrel{\Delta}{=} \xi^i(t+1) - \zeta^i(t) = \xi^i(t+1) - \sum_{j \in \mathcal{V}} a_{ij}(t) \xi^j(t) \xrightarrow[t \rightarrow \infty]{} 0.$$

Отметим теперь, что состояния агентов  $\xi^i$ , в силу определения  $f^i(t)$ , могут рассматриваться как траектории усредняющей системы с возмущением:

$$\xi^i(t+1) = \sum_j a_{ij}(t) \xi^j(t) + f^i(t),$$

где  $f^i(t)$  — нечто, сходящееся к нулю. В силу леммы 2 наличие исчезающих на бесконечности возмущений<sup>15</sup> не нарушает синхронизации между агентами:  $\left\| \xi^i(t) - \xi^j(t) \right\|_{t \rightarrow \infty} \rightarrow 0$ .

**Шаг 3.** Рассмотрим теперь состояние агента с индексом  $i \in \mathcal{V}$ . Поскольку векторы  $\xi^i(t)$  ограничены, то существует последовательность  $t_r \rightarrow \infty$  такая, что  $\xi^i(t_r) \xrightarrow{r \rightarrow \infty} \xi_* \in \mathbb{R}^d$ . Ввиду синхронизации мы имеем  $\xi^j(t_r) \rightarrow \xi_*$  для каждого  $j \in \mathcal{V}$ . Из свойства (34) следует, что  $P_i(\xi_*) = \xi_* \forall i$ , откуда  $\xi_* \in \Xi_*$ . Остается показать, что  $\xi^i(t) \xrightarrow{t \rightarrow \infty} \xi_*$ . Заметим, что на шаге 1 мы не уточняли выбор  $\xi^0$ , которая может быть произвольной точкой в  $\Xi_*$ . Мы доказали, что для любой такой точки расстояния  $x_i(t) = \left\| \xi^i(t) - \xi^0 \right\|$  сходятся к некоторому значению  $c$ , зависящему от  $\xi^0$  и начальных условий. В частности, подставляя  $\xi^0 = \xi_*$ , получаем, что существуют пределы

$$x_i(t) = \lim_{t \rightarrow \infty} \left\| \xi^i(t) - \xi_* \right\| = c_* \quad \forall i.$$

Вспоминая, что  $x_i(t_r) \xrightarrow{r \rightarrow \infty} 0$ , имеем  $c_* = 0$ , что доказывает достижение консенсуса (25), причем общий предел состояний агентов в (25) равен  $\xi_* \in \Xi_*$ .

### СПИСОК ЛИТЕРАТУРЫ

1. Gordon R., Bender R., Herman G. T. Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography // Journal of Theoretical Biology. 1970. Т. 29. No 3. С. 471—481.
2. Гелиг А. Х., Матвеев А. С. Введение в математическую теорию обучаемых распознающих систем и нейронных сетей. СПб.: Изд-во СПбГУ, 2014.
3. Якубович В. А. Некоторые общие теоретические принципы построения обучаемых опознающих систем // Вычислительная техника и вопросы программирования. Ленинград: Изд-во ЛГУ. 1965. С. 3—71.
4. Demyanov V. F. Mathematical diagnostics via nonsmooth analysis // Optimization Methods and Software. — 2005. Т. 20. No 2/3. С. 197—218.
5. Оптимизационные методы в задачах диагностики / К. И. Ананьев [и др.] // Вестник СПбГУ. Прикл. матем. Информ. Прогр. упр. 2011. Т. 10. No 3. С. 3—12.
6. Малоземов В. Н., Плоткин А. В. Строгое полиномиальное отделение двух множеств // Вестник СПбГУ. Математика. Механика. Астрономия. 2019. Т. 6, No 2. С. 232—240.
7. Combettes P. L. The Foundations of Set Theoretic Estimation // Proceedings of IEEE. 1993. Т. 81. No 2. С. 182—208.
8. Петров И. П., Тимофеев А. В. Конечно-сходящиеся рекуррентные алгоритмы решения целевых неравенств при наличии ограничений // Ж. вычисл. матем. и матем. физ. 1975. Т. 15. No 6. С. 1582—1588.
9. Фомин В. Н., Фрадков А. Л., Якубович В. А. Адаптивное управление динамическими объектами. — М.: Наука, 1981.
10. Kaczmarz S. Angenäherte Auflösung von Systemen linearer Gleichungen // Bulletin International de l'Académie Polonaise des Sciences et des Lettres. 1937. Т. 35. С. 355—357.
11. Cimmino G. Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari // La Ricerca Scientifica. 1938. Т. 2. No 9. С. 326—333.
12. Von Neumann J. Functional operators, Vol. II. The geometry of orthogonal spaces. — Princeton, NJ: Princeton University Press, 1950. — (Annals of Mathematical Studies). — Reprint of mimeographed lecture notes first distributed in 1933.
13. Еремин И. И. Релаксационный метод решения систем неравенств с выпуклыми функциями в левых частях // Докл. АН СССР. 1965. Т. 160. No 5. С. 994—996.
14. Брэгман Л. М. Нахождение общей точки выпуклых множеств методом последовательного проектирования // Докл. АН СССР. 1965. Т. 162. No 3. С. 487—490.

<sup>15</sup> Формально, лемма 2 была сформулирована для случая агентов со скалярными состояниями  $x_i \in \mathbb{R}$ , однако она, очевидно, сохраняет свою силу и для многомерного случая (достаточно рассмотреть проекции состояний агентов на каждую координатную ось).

15. Гурин Л. Г., Поляк Б. Т., Райк Э. В. Методы проекций для отыскания общей точки выпуклых множеств // Ж. вычисл. матем. и матем. физ. 1967. Т. 7. No 6. С. 1211—1228.
16. Еремин И. И., Мазуров В. Д. Итерационный метод решения задачи выпуклого программирования // Докл. АН СССР. 1966. Т. 170. No 1. С. 57—60.
17. Бердникова Е. А., Ерёмин И. И., Попов Л. Д. Распределенные фейеровские процессы для систем линейных неравенств и задач линейного программирования // Автомат. и телемех. 2004. No 2. С. 16—32.
18. Ерёмин И. И., Попов Л. Д. Замкнутые фейеровские циклы для несовместных систем выпуклых неравенств // Изв. вузов. Матем. 2008. No 1. С. 11—19.
19. Ерёмин И. И., Попов Л. Д. Фейеровские процессы в теории и практике: обзор последних результатов // Изв. вузов. Матем. 2009. No 1. — С. 44—65.
20. Брэгман Л. М. Релаксационный метод нахождения общей точки выпуклых множеств и его применение для решения задач выпуклого программирования // Ж. вычисл. матем. и матем. физ. 1967. Т. 7, No 3. С. 620—631.
21. Васин И. И., Еремин И. И. Операторы и итерационные процессы фейеровского типа: теория и приложения. — Ижевск: Регулярная и хаотическая динамика, 2005.
22. Escalante R., Raydan M. Alternating Projection Methods. — Society for Industrial, Applied Mathematics Philadelphia, 2011.
23. Bauschke H. H., Borwein J. M. On Projection Algorithms For Solving Convex Feasibility Problems // SIAM Review. 1996. Т. 38. No 3. С. 367—426.
24. Lewis A. S., Malick J. Alternating Projections on Manifolds // Mathematics of Operations Research. 2008. Т. 33. No 1. С. 216—234.
25. Nedic A., Ozdaglar A., Parrilo P. Constrained consensus and optimization in multi-agent networks // IEEE Trans. Autom. Control. 2010. Т. 55. No 4. С. 922—938.
26. Shi G., Johansson K., Hong Y. Reaching an optimal consensus: dynamical systems that compute intersections of convex sets // IEEE Trans. Autom. Control. 2013. Т. 58. No 3. С. 610—622.
27. Solving a system of linear equations: from centralized to distributed algorithms / P. Wang [и др.] // Ann. Rev. Control. 2019. Т. 47. С. 306—322.
28. Проскурников А. В. Усредняющие алгоритмы и неравенства в задачах многоагентного управления и моделирования. Диссертация д.ф.-м.н. — Санкт-Петербург, 2021.
29. Proskurnikov A., Cao M. Differential inequalities in multi-agent coordination and opinion dynamics modeling // Automatica. 2017. Т. 85. С. 202—210.
30. Proskurnikov A., Calafiore G., Cao M. Recurrent averaging inequalities in multi-agent control and social dynamics modeling // Ann. Rev. Control. 2020. Т. 49. С. 95—112.
31. Балакришнан А. В. Прикладной функциональный анализ. — М.: Наука, 1980.
32. Peterson B., Olinick M. Leontief models, Markov chains, substochastic matrices, and positive solutions of matrix equations // Mathematical Modelling. 1982. Т. 3. No 3. С. 221—239.
33. Polyak B., Tremba A. Regularization-based solution of the PageRank problem for large matrices // Autom. Remote Control. 2012. Т. 73. No 11. С. 1877—1894.
34. Гасников А. В., Дмитриев Д. Ю. Об эффективных рандомизированных алгоритмах поиска вектора PageRank // Ж. вычисл. матем. и матем. физ. 2015. Т. 55. No 3. С. 355—371.
35. Sznajder R. Kaczmarz Algorithm Revisited // Technical Transactions Fundamental Sciences. 2015. No 2. С. 248—254.
36. Deutsch F. The Angle Between Subspaces of a Hilbert Space // Approximation Theory, Wavelets and Applications / под ред. S. P. Singh. — Dordrecht: Springer, 1995. С. 107—130.
37. A block projection method for sparse matrices / M. Arioli [и др.] // SIAM Journal on Scientific and Statistical Computing. 1988. Т. 13. С. 326—333.
38. Agmon S. The Relaxation Method For Linear Inequalities // Canadian Journal of Mathematics. 1954. Т. 6. С. 382—392.
39. Motzkin T. S., Schoenberg I. J. The Relaxation Method For Linear Inequalities // Canadian Journal of Mathematics. 1954. Т. 6. С. 393—404.

40. Ерёмин И. И. Обобщение релаксационного метода Моцкина—Агмона // Успехи мат. наук. 1965. Т. 20, No 2. С. 183—187.
41. Gurin L., Polyak B., Raik E. The method of projections for finding the common point of convex sets // USSR Comp. Math. Math. Phys. 1967. Т. 7. No 6. С. 1—24.
42. Ерёмин И. И. О скорости сходимости в методе фейеровских приближений // Матем. заметки. 1968. Т. 4. No 1. С. 53—61.
43. Ерёмин И. И. Фейеровские отображения и задача выпуклого программирования // Сиб. матем. журн. 1969. Т. 10. No 5. С. 1034—1047.
44. Ерёмин И. И. Применение метода фейеровских приближений к решению задач выпуклого программирования с негладкими ограничениями // Ж. вычисл. матем. и матем. физ. 1969. Т. 9. No 5. С. 1153—1160.
45. Брэгман Л. М. Релаксационный метод нахождения общей точки выпуклых множеств и его применение для задач оптимизации // Докл. АН СССР. 1966. Т. 171. No 5. С. 1019—1022.
46. Lin P., Ren W. Constrained consensus in unbalanced networks with communication delays // IEEE Trans. Autom. Control. 2014. Т. 59. No 3. С. 775—781.
47. Notarstefano G., Notarnicola I., Camisa A. Distributed Optimization for Smart Cyber-Physical Networks // Foundations and Trends in Systems and Control. 2019. Т. 7. No 3. С. 253—383.
48. Гасников А. В. Современные численные методы оптимизации. Универсальный градиентный спуск. — М.: МФТИ, 2018.
49. Fejér L. Über die Lage der Nullstellen von Polynomen, die aus Minimumforderungen gewisser Art entspringen // Math. Ann. 1922. No 85. С. 41—48.
50. Fullmer D., Morse A. S. A distributed algorithm for computing a common fixed point of a finite family of parac-  
ontractions // IEEE Trans. Autom. Control. 2018. Т. 63. No 9. С. 2833—2843.
51. Vasin V. V., Eremin I. I. Operators and Iterative Processes of Fejér Type: Theory and Applications. — De Gruyter, 2009.
52. Rzevski G., Skobelev P. Managing complexity. — Southampton, UK: WIT Press, 2014.
53. Wooldridge M. An Introduction To Multiagent Systems. — Chichester, England: John Wiley & Sons Ltd., 2002.
54. Проскурников А. В., Фрадков А. Л. Задачи и методы сетевого управления // Автоматика и телемеханика. 2016. No 10. С. 3—39.
55. Граничин О. Н. Как действительно устроены сложные информационно-управляющие системы? // Сто-  
хастическая оптимизация в информатике. 2016. Т. 12. No 1. С. 3—19.
56. Proskurnikov A. V., Granichin O. N. Evolution of clusters in large-scale dynamics networks // Cybern. Phys. 2018. Т. 7. No 3. С. 102—129.
57. Проблемы сетевого управления / А. Л. Фрадков [и др.]. — Ижевск: ИКИ, 2015.
58. Marik V., Gorodetsky V., Skobelev P. Multi-agent technology for industrial applications: barriers and trends // Proc. of IEEE Int. Conf. Syst., Man, Cybern. 2020. С. 1980—1987.
59. Агаев Р. П., Чеботарев П. Ю. Сходимость и устойчивость в задачах согласования характеристик // Управ-  
ление большими системами. 2010. Т. 30.1. С. 470—505.
60. Proskurnikov A., Tempo R. A tutorial on modeling and analysis of dynamic social networks. Part I // Ann. Rev. Control. 2017. Т. 43. С. 65—79.
61. Proskurnikov A., Tempo R. A tutorial on modeling and analysis of dynamic social networks. Part II // Ann. Rev. Control. 2018. Т. 45. С. 166—190.
62. Seneta E. Non-negative matrices and Markov chains. — New York: Springer-Verlag, 1981.
63. Leizarowitz A. On infinite products of stochastic matrices // Linear Alg. Appl. 1992. Т. 168. С. 189—219.
64. Bolouki S., Malhame R. P. Consensus algorithms and the decomposition-separation theorem // IEEE Trans. Au-  
tom. Control. 2016. Т. 61. No 9. С. 2357—2369.
65. Touri B., Nedic A. On backward product of stochastic matrices // Automatica. 2012. Т. 48. No 8. С. 1477—1488.
66. Chebotarev P., Agaev R. Forest matrices around the Laplacian matrix // Linear Alg. Appl. 2002. Т. 356. С. 253—274.
67. Ren W., Beard R. Distributed consensus in multi-vehicle cooperative control: theory and applications. — London: Springer-Verlag, 2008.
68. Ren W., Cao Y. Distributed coordination of multi-agent networks. — Springer, 2011.

69. Shi G., Johansson K. Robust consensus for continuous-time multi-agent dynamics // *SIAM J. Control Optim.* 2013. Т. 51. No 5. С. 3673—3691.
70. Proskurnikov A., Cao M. Modulus consensus in discrete-time signed networks and properties of special recurrent inequalities // *Proc. of IEEE Conf. Decision and Control.* 2017. С. 2003—2008.
71. Proskurnikov A., Matveev A. Popov-type criterion for consensus in nonlinearly coupled networks // *IEEE Trans. Cybernetics.* 2015. Т. 45. No 8. С. 1537—1548.
72. You K., Song S., Tempo R. A networked parallel algorithm for solving linear algebraic equations // *Proc. IEEE Conf. Decision and Control.* 2016. С. 1727—1732.
73. Mou S., Liu J., Morse A. A distributed algorithm for solving a linear algebraic equation // *IEEE Trans. Autom. Control.* 2015. Т. 60. No 11. С. 2863—2878.
74. Olshevsky A., Tsitsiklis J. Convergence speed in distributed consensus and averaging // *SIAM Rev.* 2011. Т. 53. No 4. С. 747—772.
75. Proskurnikov A. V., Calafiore G. C. Delay Robustness of Consensus Algorithms: Continuous-Time Theory // *IEEE Trans. Autom. Control.* 2023. Т. 68. No 9. С. 5301—5316.
76. Shi G., Johansson K. Randomized optimal consensus of multi-agent systems // *Automatica.* 2012. Т. 48. Вып. 12. С. 3018—3030.
77. Steinerberger S. Randomized Kaczmarz Converges Along Small Singular Vectors // *SIAM Journal on Matrix Analysis and Applications.* 2021. Т. 42. No 2. С. 608— 615.
78. Sarlette A., Sepulchre R. Consensus Optimization on Manifolds // *SIAM Journal on Control and Optimization.* — 2009. Т. 48, No 1. С. 56—76.

# ALTERNATING PROJECTION METHOD FOR INTERSECTION OF CONVEX SETS, MULTI-AGENT CONSENSUS ALGORITHMS, AND AVERAGING INEQUALITIES

A. V. Proskurnikov<sup>a,\*</sup>, I. S. Zabarianska<sup>b,\*\*</sup>

<sup>a</sup>*Politecnico di Torino,*

*Corso Duca degli Abruzzi, 24, Turin, 10129 Italy*

<sup>b</sup>*St. Petersburg State University,*

*14th line Vasilyevskiy Island, 29, St. Petersburg, 199178 Russia*

<sup>\*</sup>*e-mail: avp1982@gmail.com*

<sup>\*\*</sup>*e-mail: akshiira@yandex.ru*

Received 03 November, 2023

Revised 11 November, 2023

Accepted 20 November, 2023

**Abstract.** The history of the alternating projection method for finding a common point of several convex sets in Euclidean space goes back to the well-known Kaczmarz algorithm for solving systems of linear equations, which was devised in the 1930s and later found wide applications in image processing and computed tomography. An important role in the study of this method was played by I.I. Eremin's, L.M. Bregman's, and B.T. Polyak's works, which appeared nearly simultaneously and contained general results concerning the convergence of alternating projections to a point in the intersection of sets, assuming that this intersection is nonempty. In this paper, we consider a modification of the convex set intersection problem that is related to the theory of multi-agent systems and is called the constrained consensus problem. Each convex set in this problem is associated with a certain agent and, generally speaking, is inaccessible to the other agents. A group of agents is interested in finding a common point of these sets, that is, a point satisfying all the constraints. Distributed analogues of the alternating projection method proposed for solving this problem lead to a rather complicated nonlinear system of equations, the convergence of which is usually proved using special Lyapunov functions. A brief survey of these methods is given, and their relation to the theorem ensuring consensus in a system of averaging inequalities recently proved by the first author is shown (this theorem develops convergence results for the usual method of iterative averaging as applied to the consensus problem).

**Keywords:** alternating projection method, convex programming, Fejér mappings, distributed algorithms, consensus, multi-agent systems.