

НА ПРАВАХ РУКОПИСИ

НЕСТЕРОВ ЮРИЙ ЕВГЕНЬЕВИЧ

Алгоритмическая выпуклая оптимизация

СПЕЦИАЛЬНОСТЬ 01.01.07 - ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА

АВТОРЕФЕРАТ ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ
ДОКТОРА ФИЗИКО-МАТЕМАТИЧЕСКИХ НАУК

МОСКВА - 2013

Работа выполнена в Лаборатории структурных методов
анализа данных в предсказательном моделировании (ПреМоЛаб)
Московского физико-технического института
(государственного университета)

Официальные оппоненты: доктор физико-математических наук,
профессор, академик РАН
Евтушенко Юрий Гаврилович,
Вычислительный центр
им. А.А. Дородницына РАН, директор

доктор физико-математических наук,
профессор
Тихомиров Владимир Михайлович,
кафедра общих проблем управления
механико-математического факультета
МГУ им. М.В.Ломоносова, профессор

доктор технических наук
Горнов Александр Юрьевич,
лаборатория оптимального управления
Института динамики систем и теории
управления Сибирского отделения РАН,
главный научный сотрудник

Ведущая организация: Институт проблем управления
им. В.А.Трапезникова РАН

Защита состоится _____ в _____
час. на заседании диссертационного совета Д 212.156.05 при Московском
физико-техническом институте (ГУ) по адресу: 141700, МО, г. Долгопруд-
ный, Институтский пер. д. 9, ауд. 903 КПМ.

С диссертацией можно ознакомиться в библиотеке МФТИ (ГУ).

Автореферат разослан " _____ " _____ 2013 г.

Ученый секретарь диссертационного совета

Федько О.С.

Общая характеристика работы

Актуальность темы и степень ее разработанности

Настоящая диссертация посвящена разработке новых методов решения нелинейных выпуклых задач оптимизации. Теория методов оптимизации является одной из самых востребованных областей численного анализа. Наиболее продвинутой ее частью посвящена решению выпуклых задач. Эти постановки базируются на солидном математическом фундаменте, разработанном в основном в первой половине 20го столетия математиками Г. Минковским, К. Каратеодори, Э. Хелли, В. Фенхелем, А. Александровым и другими. Поначалу выпуклый анализ развивался независимо от теории экстремальных задач. Однако после основополагающей монографии Р.Т. Рокафеллара¹, центр развития этой науки окончательно сместился в сторону теории оптимизации². В настоящее время существует большое количество прекрасных книг как по выпуклому анализу, так и по его оптимизационным приложениям. Очень важно, что выпуклые задачи представляют собой практически единственный класс оптимизационных задач, допускающих построение методов с глобальными скоростными характеристиками, приемлемыми для большинства практических приложений. Это обстоятельство привело к появлению большого количества интересных методов и подходов. Основные приоритетные результаты в этой области принадлежат отечественным ученым А.С. Антипину, Ф.П. Васильеву, Е.Г. Гольштейну, В.Ф. Демьянову, Ю.Г. Евтушенко, Ю.М. Ермольеву, А.Д. Иоффе, В.Г. Карманову, Б.С. Мордуховичу, А.С. Немировскому, Б.Т. Поляку, Р.А. Поляку, Б.Н. Пшеничному, А.М. Рубинову, В.М. Тихомирову, Л.Г. Хачияну, Н.З. Шору, Д.Б. Юдину, и другим. Результаты по теории сложности экстремальных задач могут рассматриваться как естественное развитие традиций, заложенных еще Л.В. Канторовичем³.

Однако, начиная с выдающейся работы Н. Кармаркара, опубликованной в 1985 г., и примерно до 2000 г. развитие теории и методов оптимизации было в основном связано с прогрессом в теории полиномиальных методов внут-

¹Рокафеллар Р.Т. Выпуклый анализ. - М.: Мир, 1973 - 420с.

²Иоффе А.Д., Тихомиров В.М. Теория экстремальных задач. - М.: Наука, 1974 - 460с.

³Канторович Л.В. Функциональный анализ и прикладная математика // Успехи мат. наук - 1948. - Т.3, вып.1 - С.89-185.

ренной точки. Были получены новые и очень эффективные методы решения задач линейного программирования, которые существенно подняли планку соревнования с традиционным симплекс-методом. Была разработана общая теория самосогласованных функций⁴, которая позволяла строить полиномиальные методы внутренней точки для всех выпуклых задач с явной структурой. Появилась теория конических нелинейных задач, в рамках которой удалось построить длинношаговые прямо-двойственные методы внутренней точки, способные ускоряться и стартовать из недопустимых точек. Новые эффективные методы появились в квадратичном и полуопределенном программировании. Очень важно, что прогресс в этом направлении сопровождался и иницировался детальным теоретическим анализом эффективности предлагаемых алгоритмов. Оценки скорости сходимости стали важным аргументом в пользу новых методов. Может быть, даже более важным чем результаты предварительных численных экспериментов.

К концу этого периода выяснилось, что все это время запросы пользователей росли быстрее, чем возможности новых методов. Громадный прогресс в скорости компьютеров, оперативной памяти и возможностях вспомогательных программ, небывалая доступность различной информации через интернет, дешевизна и доступность вычислительных средств - все это вместе на порядки упростило процесс создания математических моделей и, как неизбежное следствие, существенно увеличило их размеры. Худшей ситуации для методов внутренней точки придумать было нельзя, так как трудоемкость этих методов растет (по крайней мере) пропорционально кубу размерности.

Цели и задачи

В связи с этим в начале 2000х годов встал вопрос о частичной реабилитации почти забытых методов градиентного типа. Однако полностью вернуться на позиции середины 80х годов оказалось невозможным. Дело в том, что к 1985г. теория методов выпуклой оптимизации представляла собой цельный монолит⁵, заверченный в основном усилиями А.С. Немировского и Д.Б. Юдина.

⁴Nesterov Yu., Nemirovskii A. Interior point polynomial methods in convex programming: Theory and Applications // Philadelphia: SIAM, 1994 - 520p.

⁵Поляк Б.Т. Введение в оптимизацию. - М.: Наука, 1983.-433.

Разработанная ими теория сложности⁶, дающая верхние оценки на возможную эффективность методов минимизации для различных классов задач, была подкреплена оптимальными методами, которые эти оценки как раз и реализовывали. Предположения их вычислительной модели (оракул типа черный ящик) полностью соответствовали существовавшему тогда стилю написания оптимизационных пакетов программ, где пользователю предлагалось написать подпрограмму вычисления значения функции и градиента, закрытую для самого пакета⁷. Таким образом, в то время казалось что все важные вопросы в этой области были уже заданы и на (почти) все из них получены ответы.

Однако к 2000му году этот монолит дал трещину. Дело в том, что область реальных приложений для полиномиальных методов внутренней точки практически совпадает с областью применения чернойящичных методов. И там, где новые методы были применимы (т.е. были в состоянии сделать хотя бы одну итерацию), они как правило оказывались и в теории, и на практике лучше неулучшаемых оптимальных методов. Кажущееся противоречие легко объяснялось тем, что методы внутренней точки для построения хорошего (самоогласованного) барьера требовали доступа к внутренней структуре оракула, т.е. они не были чернойящичными.

Таким образом, к началу 2000х годов стало ясно, что некоторые из особо пессимистических границ теории сложности можно скорее всего отодвинуть - для этого был, по крайней мере, один пример (самосогласованные функции). И было понятно как - надо использовать структуру задачи. Этими соображениям и объясняется основное направление исследований автора, приведшее к появлению излагаемых ниже результатов.

Основной целью настоящей диссертации является разработка более мощных и быстрых методов оптимизации, которые значительно расширяют возможности, обрисованные в стандартной теории сложности.

⁶Немировский А.С., Юдин Д.Б. Сложность задач и эффективность методов оптимизации. - М.: Наука, 1977. - 540с.

⁷Евтушенко Ю.Г. Методы решения экстремальных задач и их применение в системах оптимизации. - М.: Наука, 1982. - 412с.

Новизна, методология и методы исследования

Все приводимые ниже результаты являлись на момент их публикации новыми. Большинство из них было получено в 2002-2007 годах и опубликовано ближе к концу десятилетия в ведущих оптимизационных журналах. Все они связаны с разработкой новых методов оптимизации для различных классов выпуклых задач.

Новые методы обладают улучшенными глобальными скоростными характеристиками и/или возможностями аппроксимации двойственных решений. Их разработка стала возможной благодаря уточнению понятия модели целевой функции. Полезно различать три типа моделей.

1. Геометрические модели. В них обычно обосновывается факт включения некоторого простого выпуклого множества, содержащего текущую точку в более сложный объект (допустимое множество, надграфик целевой функции). В результате появляется возможность сдвинуться в нужном направлении, сохраняя при этом допустимость решения. В качестве примеров можно привести следующие ситуации.

- *Дикинский эллипсоид.* Этот эллипсоид, определяемый гессианом самосогласованной функции, лежит в допустимой области.
- *Аппроксимация 1го порядка.* Для функции $f(x)$, чей градиент удовлетворяет условию Липшица с константой L , можно написать следующую аппроксимацию:

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

Минимизируя эту аппроксимацию на допустимом множестве, можно улучшить значение целевой функции.

- *Аппроксимация 2го порядка.* Для функции $f(x)$, у которой гессиан является липшицевым с константой M , верна верхняя оценка

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}\langle \nabla^2 f(x)(y - x), y - x \rangle + \frac{1}{6}M\|y - x\|^3.$$

Минимизируя эту аппроксимацию, получаем модифицированный вариант метода Ньютона, обладающий глобальными гарантиями эффективности.

- *Верхняя аппроксимация для системы уравнений.* Для уравнения $F(x) = 0$, где отображение $F : R^n \rightarrow R^m$ имеет липшицев якобиан ∇F с константой M , можно написать верхнюю оценку

$$\|F(y)\| \leq \|F(x) + \nabla F(x)(y - x)\| + \frac{1}{2}M\|y - x\|^2.$$

Минимизация этой оценки дает модифицированный метод Гаусса-Ньютона.

2. Алгебраические модели. В них описывается конкретный способ представления целевой функции. Для выпуклых функций естественной формой представления является максимум (дискретный или непрерывный) выпуклых функций. Например, может случиться, что нам известно следующее представление:

$$f(x) = \max_{u \in Q_2} \{\langle Ax, u \rangle - \phi(u)\}.$$

Эта модель оказывается полезной в прямо-двойственных субградиентных методах, где она позволяет строить приближенные двойственные решения. При благоприятных обстоятельствах эта модель может использоваться в технике сглаживания, что на порядок увеличивает скорость сходимости соответствующих методов.

3. Структурные модели. В них описывается структура целевой функции. Обычно в ней выделяются два взаимодействующих объекта, один простой и один сложный. Простой объект может иметь плохие функциональные характеристики. Однако он должен допускать решение некоторых вспомогательных задач с его участием в явном виде. Примером может являться составная целевая функция, которая формируется как сумма гладкой выпуклой функции и простой выпуклой функции общего вида. Простая функция может быть сколь угодно плохой, даже разрывной (например, индикаторной функцией выпуклого множества).

В этой работе мы покажем, как и насколько увеличиваются возможности методов оптимизации, если они построены с учетом вышеупомянутых моделей целевой функции (или их комбинаций). Мы приведем новые алгоритмы следующих типов.

- Прямо-двойственные субградиентные методы для решения негладких выпуклых задач (параграфы 3.1 и 3.2).

- Методы минимизации составных функций (параграф 3.3).
- Методы решения вариационных неравенств (Глава 4).
- Методы второго порядка (Глава 5).
- Алгоритмы техники сглаживания (Глава 6).
- Методы для нахождения решений выпуклых задач с относительной точностью (Глава 7).

Теоретическая и практическая значимость работы

В работе предлагаются новые методы для нахождения приближенных решений для нескольких важных классов выпуклых задач. Все методы являются новыми. При этом они превосходят известные алгоритмы по крайней мере в одном из следующих аспектов:

- вычисление дополнительных характеристик решения (например, наряду с прямым решением аппроксимируется также и решение двойственной задачи или производится контроль точности решения);
- обладают гарантиями глобальной эффективности (например, новые методы второго порядка);
- сходятся быстрее, чем разрешает стандартная теория сложности (например, алгоритмы техники сглаживания).

Подобная эффективность многих из разработанных методов ранее считалась недостижимой. Поэтому можно ожидать существенное увеличение размеров решаемых задач, обладающих выпуклой структурой.

Положения, выносимые на защиту

- Прямо-двойственные методы для минимизации негладких функций. Обладают неулучшаемой оценкой скорости сходимости. Позволяют формировать приближенное решение двойственной задачи и осуществлять контроль точности.

- Барьерный субградиентный метод. Позволяет минимизировать негладкие функции на множествах заданных самосогласованным барьером. Обладает оптимальной скоростью сходимости. Позволяет максимизировать положительные вогнутые функции с относительной точностью.
- Градиентные методы минимизации составных функций. Применимы к функциям, представимым в виде суммы гладкой функции, заданной череноящичным оракулом и простой функцией общего вида. Позволяют сохранить скорость сходимости, характерную для гладкого случая.
- Метод решения вариационных неравенств с гладким оператором. Обладает неулучшаемой оценкой скорости сходимости. Является двойственным аналогом экстраградиентного метода.
- Методы решения квазивариационных неравенств. Существенно расширен класс неравенств, обладающих единственным решением, и построен быстрый метод их решения.
- Кубическая регуляризация метода Ньютона. С помощью кубической верхней оценки целевой функции построен новый метод второго порядка. Метод работает и на невыпуклых функциях. На выпуклых задачах он сходится со скоростью $O(1/k^2)$, где k - это счетчик итераций. Это первый метод второго порядка, обладающий такой скоростью сходимости на вырожденных выпуклых задачах.
- Ускоренный метод Ньютона. За счет использования небольшой памяти кубический метод Ньютона удается ускорить. Теперь он сходится как $O(1/k^3)$.
- Модифицированный метод Гаусса-Ньютона. Предлагается новый метод для решения систем нелинейных уравнений. Он основан на использовании верхней квадратичной аппроксимации для нормы невязки. На невырожденных задачах допускает глобальную оценку эффективности.
- Сглаживание явной модели целевой функции. Для подходящей модели эта техника позволяет решать негладкие задачи с трудоемко-

стью $O(1/\epsilon)$ итераций метода градиентного типа, где ϵ - это требуемая точность решения. Тем самым преодолевается неулучшаемая оценка $O(1/\epsilon^2)$, справедливая для чернойщичных методов.

- Условие обратного зазора. Методы, основанные на этой интерпретации, позволяют применять технику сглаживания к более широкому классу задач, в частности включающему сильно выпуклые функции. Это улучшает оценку трудоемкости до $O(1/\epsilon^{1/2})$.
- Техника сглаживания в полуопределенной оптимизации. Показывается, как применять технику сглаживания к симметричным функциям собственных значений симметрических матриц.
- Оптимизация с относительной точностью. Показывается, как получать приближенные решения с относительной точностью для специальных классов однородных задач. Получаемые оценки трудоемкости практически не зависят от исходных данных.
- Эллипсоидальная аппроксимация выпуклых тел используется для нахождения правильной евклидовой нормы, позволяющей существенно ускорить скорость сходимости методов оптимизации.

Публикации и апробация результатов

Результаты, включенные в данную работу, опубликованы в 19 статьях в ведущих отечественных и международных журналах (см. [1-6, 9-18, 20-22]). В работе также использовались фрагменты из монографий автора [7], [8] и русской версии [19]. В двух работах [11] и [20], написанных в соавторстве, соискателю принадлежат результаты, указанные в положениях, выносимых на защиту данной диссертации.

Результаты, включенные в диссертацию, неоднократно докладывались на научных семинарах в ведущих университетах России и за рубежом: семинар кафедры общих проблем управления, мех-мат МГУ, руководители: проф. М.И.Зеликин, проф. В.Ю.Протасов, проф. В.М.Тихомиров и проф. А.В.Фурсиков (апрель 2012); семинар лаборатории Адаптивных и робастных

систем им. Я.З.Цыпкина, ИПУ РАН, руководитель – профессор Б.Т.Поляк (март 2011); семинар лаборатории ПреМоЛаб, МФТИ, руководитель – профессор В.Г.Спокойный (декабрь 2011, апрель 2012); семинар по оптимизации, Технологический Университет Джорджии (США), руководитель – профессор А.С.Немировский (апрели 2008 - 2012гг.); семинар по исследованию операций института IFOR (ETHZ, Zurich), руководитель – профессор Н.-J.Luethi (март 2008, сентябрь 2011).

В качестве **пленарных и приглашенных докладов**, результаты автора докладывались на многочисленных международных конференциях:

International Conference on Operations Research, Цюрих, 2011;

International conference "Foundations of Computational Mathematics", Будапешт, 2011;

International Conference on Machine Learning (NIPS), Ванкувер, 2010;

International Conference on Optimization and Applications (CIRM), Барселона, 2010;

International Workshop on Quadratic Optimization, Левен, 2010;

IPAM Workshop on Continuous Optimization, Лос Анжелес, 2010;

International Congress of Mathematicians, Хидерабад, 2010;

International conference ОПТИМА, Будва, 2009;

20th International Symposium on Mathematical Programming, Чикаго, 2009;

7th EUROPT Workshop on Operations Research, Реманген, 2009;

7th Brazilian Conference on Continuous Optimization, Кампинас, 2008;

International Conference on Optimization and Operations Research, Северобайкальск, 2008;

School "New Paradigms in Optimization", Аскона, 2008;

International conference "High Performance Optimization", Амстердам, 2008;

INFORMS meeting on Continuous Optimization, Атланта, 2008;

School on Continuous Optimization, Эйн Геди, 2007;

International Conference on Continuous and Discrete optimization, Ватерлоо, 2007;

International Conference on Nonlinear Optimization, Люмини, 2007;

School on Operations Research and Optimization, Зиналь, 2007;

International Workshop on Continuous Optimization VOCAL, Будапешт,

2006;

International Symposium on Mathematical Programming, Рио-де-Жанейро, 2006;

International Conference on Semidefinite Programming, Сингапур, 2006;

Annual INFORMS meeting. Сан-Франциско, 2005;

International conference on positive polynomials, Люмини, 2005;

International Workshop on Nonlinear Optimization, Обервольфах, 2005;

French-German-Spain conference on Operations Research, Авиньон, 2004;

International conference "High Performance Optimization", Амстердам, 2004;

International workshop on semidefinite programming, Ватерлоо, 2004.

На основе результатов диссертации подготовлены курсы лекций, которые были прочитаны в различных европейских университетах:

- Университет г.Льеж (Бельгия), 2012;
- Независимый московский университет, Москва, 2012;
- Ecole nationale de la statistique et de l'administration économique (Paris Tech), Париж (Франция), 2012;
- Университет г.Вены (Австрия), 2013;
- V Традиционная Школа "Управление, информация и оптимизация", Сенеж, Моск. обл., 2013.

Структура и объем работы

Диссертация состоит из введения, шести глав, заключения, и списка литературы. Полный объем составляет 367 страниц, список литературы содержит 125 наименований.

Основное содержание работы

Во **Введении** обосновывается актуальность темы исследования и степень ее разработанности. Указывается, что основной целью диссертации является

разработка более мощных и быстрых методов оптимизации, которые значительно расширяют возможности, обрисованные в стандартной теории сложности. Обосновывается научная новизна и значимость работы. Основные аргументы из Введения приведены в первом параграфе автореферата.

Глава 1 является вводной и содержит все необходимые вспомогательные результаты. В параграфе 1.1 вводятся различные классы выпуклых функций, ранжированные по степени гладкости и выпуклости. Доказываются полезные неравенства.

Параграф 1.2 начинается с рассмотрения базовых градиентных методов, *прямого* и *двойственного*, которые применяются к задаче

$$\text{Найти } f^* = \min_{x \in Q} f(x), \quad (1.2.1)$$

где Q - это выпуклое замкнутое множество, и $f \in \mathcal{C}_L^{1,1}(Q)$. Обозначим через $T_Q(x) \in Q$ оптимальное решение следующей задачи минимизации:

$$\min_y \{ \langle \nabla f(x), y - x \rangle + \frac{1}{2}L\|y - x\|^2 : y \in Q \}. \quad (1.2.2)$$

В этой вспомогательной задаче норма может и не быть евклидовой.

Для описания простейших методов нам потребуется понятие *брегмановской проекции*. Обозначим через $d(x)$ некоторую *прокс-функцию* множества Q . Предполагается что функция $d(x)$ является непрерывной и сильно выпуклой на Q с параметром выпуклости единица. Для простоты предположим что $d(x)$ дифференцируема. Обозначим через x_0 ее *центр*:

$$x_0 = \arg \min_x \{d(x) : x \in Q\}.$$

Не ограничивая общности, будем считать что $d(x_0) = 0$. Обозначим через

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

брегмановское расстояние между точками x и y . Теперь мы можем определить *брегмановскую проекцию*:

$$\mathcal{B}_Q(x, g) = \arg \min_y \{ \langle g, y - x \rangle + \rho(x, y) : y \in Q \}.$$

В терминах этого отображения удобно записывать методы минимизации первого порядка.

Прямой градиентный метод	(1.2.7)
$x_{k+1} = \mathcal{B}_Q(x_k, \frac{1}{L}\nabla f(x_k)), k = 0, 1, \dots$	

Теорема 1.2.1 Пусть $f \in \mathcal{C}_L^{1,1}(Q)$. Тогда для всех $k \geq 1$ выполнено неравенство

$$f(\bar{x}_k) - f^* \leq \frac{1}{k} \sum_{i=1}^k [f(x_i) - f^*] \leq \frac{1}{k} L d(x^*), \quad (1.2.8)$$

где $\bar{x}_k = \frac{1}{k} \sum_{i=1}^k x_i$.

Двойственный метод пересчитывает простейшую модель целевой функции.

Двойственный градиентный метод
<p>Инициализация: Положим $\psi_0(x) = \frac{1}{2}\rho(x_0, x)$.</p> <p>При $k \geq 0$ повторяем: Вычисляем $x_k = \arg \min_{x \in Q} \psi_k(x)$.</p> <p>Пересчитываем $\psi_{k+1}(x) = \psi_k(x) + \frac{1}{L}[f(x_k) + \langle \nabla f(x_k), x - x_k \rangle]$.</p>
(1.2.10)

Теорема 1.2.2 Пусть последовательность $\{x_k\}_{k \geq 0}$ сформирована методом (1.2.10). Обозначим $y_k = T_Q(x_k)$, и $\bar{y}_k = \frac{1}{k+1} \sum_{i=0}^k y_i, k \geq 0$. Тогда при всех $k \geq 0$ выполнено неравенство

$$f(\bar{y}_k) - f^* \leq \frac{1}{k+1} \sum_{i=0}^k [f(y_i) - f^*] \leq \frac{L}{k+1} d(x^*). \quad (1.2.11)$$

И, наконец, быстрый градиентный метод может рассматриваться как комбинация этих двух алгоритмов.

Быстрый градиентный метод
<p>При $k \geq 0$ повторяем следующие операции:</p> <p>Вычисляем $f(x_k)$ и $\nabla f(x_k)$. Находим $y_k = T_Q(x_k)$.</p> <p>Находим $z_k = \arg \min_x \left\{ Ld(x) + \sum_{i=0}^k \frac{i+1}{2} [f(x_i) + \langle \nabla f(x_i), x - x_i \rangle] : x \in Q \right\}$.</p> <p>Полагаем $x_{k+1} = \frac{2}{k+3} z_k + \frac{k+1}{k+3} y_k$.</p>
(1.2.18)

Теорема 1.2.3 Пусть последовательности $\{x_k\}_{k=0}^\infty$ и $\{y_k\}_{k=0}^\infty$ сформированы методом (1.2.18). Тогда при всех $k \geq 0$ справедлива оценка

$$\frac{(k+1)(k+2)}{4}f(y_k) \leq \min_x \left\{ Ld(x) + \sum_{i=0}^k \frac{i+1}{2}[f(x_i) + \langle \nabla f(x_i), x - x_i \rangle] : x \in Q \right\}. \quad (1.2.19)$$

Таким образом,

$$f(y_k) - f(x^*) \leq \frac{4Ld(x^*)}{(k+1)(k+2)}, \quad (1.2.20)$$

где x^* - это оптимальное решение задачи (1.2.1).

Этот параграф заканчивается описанием модификаций этих методов, предназначенных для минимизации сильно выпуклых функций. Сначала показывается что понятие сильной выпуклости у гладких функций естественным образом связывается только с евклидовой нормой (для этого доказыва-ется, что число обусловленности дважды дифференцируемой функции в ℓ_1 -норме не может быть меньше размерности пространства). Затем приводятся оценки скорости сходимости простейших градиентных методов, которые являются *непрерывными* по параметру сильной выпуклости.

В последнем параграфе 1.3 этой главы приводятся необходимые сведения из теории самосогласованных барьеров (см. [19]). Эти результаты необходи-мы для обоснования сложности решения вспомогательных задач в методах градиентного типа.

Глава 2 посвящена современным методам субградиентного типа. Основ-ные результаты этой главы опубликованы в работах [18, 21, 22]. В параграфе 2.1 описываются новые прямо-двойственные субградиентные ме-тоды для минимизации негладких функций. Первые (прямые) субградиент-ные методы были предложены в работах Н.З.Шора⁸ и Б.Т.Поляка⁹. Двой-ственный субградиентный метод был известен под именем *зеркального спус-ка*¹⁰. Однако в обоих случаях методы имели только одну управляющую по-следовательность шаговых множителей. В этом параграфе мы показываем что требуются две такие последовательности: одна для движения в прямом

⁸Shor N.Z. Minimization Methods for Nondifferentiable Functions. - Berlin: Springer-Verlag, 1985. - 312p.

⁹Поляк Б.Т. Общий метод решения экстремальных задач // ДАН СССР - 1967 - Т.8 - С.593-597.

¹⁰Немировский А.С., Юдин Д.Б. Сложность задач и эффективность методов оптимизации. - М.: Наука, 1977, - 540с.

пространстве, и одна - для движения в двойственном. В результате получаемые методы освобождаются от излишних логарифмических множителей в оценках трудоемкости. Остановимся подробнее на структуре предлагаемых методов.

Пусть выпуклое замкнутое множество Q снабжено прокс-функцией $d(x)$. Это множество может быть неограниченным. Нам потребуются две вспомогательные функции для множества Q опорного типа:

$$\begin{aligned}\xi_D(s) &= \max_{x \in Q} \{ \langle s, x - x_0 \rangle : d(x) \leq D \}, \\ V_\beta(s) &= \max_{x \in Q} \{ \langle s, x - x_0 \rangle - \beta d(x) \},\end{aligned}\tag{2.1.11}$$

где $D \geq 0$ и $\beta > 0$ являются параметрами. Первая функция является обычной опорной функцией множества $\mathcal{F}_D = \{x \in Q : d(x) \leq D\}$. Вторая функция представляет собой сглаженный вариант первой. Заметим что обе функции положительны. Нам потребуется следующее соотношение между функциями (2.1.11).

Лемма 2.1.15 *При всех $s \in E^*$ и $\beta \geq 0$ имеем: $\xi_D(s) \leq \beta D + V_\beta(s)$.*

Рассмотрим следующие последовательности:

$$X_k = \{x_i\}_{i=0}^k \subset Q, \quad G_k = \{g_i\}_{i=0}^k \subset E^*, \quad \Lambda_k = \{\lambda_i\}_{i=0}^k \subset R_+.$$

Обычно, тестовые точки x_i и веса λ_i генерируются некоторым методом, а вектора (субградиенты) g_i вычисляются оракулом $\mathcal{G}(\cdot)$ типа черный ящик,

$$g_i = \mathcal{G}(x_i), \quad i \geq 0,$$

связанным с конкретной выпуклой задачей.

Мы собираемся приближать прямые и двойственные решения нашей задачи с помощью следующих усредненных объектов:

$$\begin{aligned}S_k &= \sum_{i=0}^k \lambda_i, & \hat{x}_{k+1} &= \frac{1}{S_k} \sum_{i=0}^k \lambda_i x_i, \\ s_{k+1} &= \sum_{i=0}^k \lambda_i g_i, & \hat{s}_{k+1} &= \frac{1}{S_k} s_{k+1},\end{aligned}\tag{2.1.19}$$

где $\hat{x}_0 = x_0$ и $s_0 = 0$. Качество последовательности X_k естественным образом описывается следующей *функцией зазора*:

$$\delta_k(D) = \max_x \left\{ \sum_{i=0}^k \lambda_i \langle g_i, x_i - x \rangle : x \in \mathcal{F}_D, \right\}, \quad D \geq 0. \quad (2.1.20)$$

Полезной является и *сглаженная функция зазора*

$$\Delta_k(\beta, D) = \beta D + \sum_{i=0}^k \lambda_i \langle g_i, x_i - x_0 \rangle + V_\beta(-s_{k+1}).$$

При этом, для всех неотрицательных D и β имеем: $\delta_k(D) \leq \Delta_k(\beta, D)$.

Поскольку множество Q простое, значения функций зазора могут быть легко вычислены. При некоторых D эти значения могут быть отрицательными. Однако, если решение x^* нашей задачи существует, то для $D \geq d(x^*)$ значение $\delta_k(D)$ неотрицательно независимо от последовательностей X_k , Λ_k и G_k , используемых для его определения.

Рассмотрим теперь общую схему *двойственных усреднений* (ДУ).

Инициализация: Set $s_0 = 0 \in E^*$. Choose $\beta_0 > 0$.	
Итерация ($k \geq 0$):	
1. Вычисляем $g_k = \mathcal{G}(x_k)$.	(2.1.24)
2. Полагаем $\lambda_k > 0$. Set $s_{k+1} = s_k + \lambda_k g_k$.	
3. Выбираем $\beta_{k+1} \geq \beta_k$. Полагаем $x_{k+1} = \pi_{\beta_{k+1}}(-s_{k+1})$.	

Теорема 2.1.30 Пусть последовательности X_k , G_k и Λ_k построены схемой (2.1.24). Тогда при всех $k \geq 0$ и $D \geq 0$ имеем:

$$\delta_k(D) \leq \Delta_k(\beta_{k+1}, D) \leq \beta_{k+1} D + \frac{1}{2\sigma} \sum_{i=0}^k \frac{\lambda_i^2}{\beta_i} \|g_i\|_*^2. \quad (2.1.25)$$

Форма неравенств (2.1.25) приводит к естественной стратегии выбора параметров β_i . Определим последовательность:

$$\hat{\beta}_0 = \hat{\beta}_1 = 1, \quad \hat{\beta}_{i+1} = \hat{\beta}_i + \frac{1}{\hat{\beta}_i}, \quad i \geq 1.$$

Разумность ее определения подтверждается соотношением $\hat{\beta}_{k+1} = \sum_{i=0}^k \frac{1}{\hat{\beta}_i}$, $k \geq 0$. Таким образом, эта последовательность может использоваться для балансировки двух членов, возникающих в правой части неравенства (2.1.25).

Скорость роста этой последовательности оценивается так: $\sqrt{2k-1} \leq \hat{\beta}_k \leq \frac{1}{1+\sqrt{3}} + \sqrt{2k-1}$, $k \geq 1$.

Имеются две стратегии для выбора параметров λ_i : а) *простые усреднения*: $\lambda_k = 1$; б) *взвешенные усреднения*: $\lambda_k = \frac{1}{\|g_k\|_*}$. Приведем схему ДУ-метода для первого способа.

Метод Простых Усреднений

Инициализация: Set $s_0 = 0 \in E^*$. Выбираем $\gamma > 0$.	(2.1.31)
Итерация ($k \geq 0$):	
1. Вычисляем $g_k = \mathcal{G}(x_k)$. Set $s_{k+1} = s_k + g_k$.	
2. Выбираем $\beta_{k+1} = \gamma \hat{\beta}_{k+1}$. Полагаем $x_{k+1} = \pi_{\beta_{k+1}}(-s_{k+1})$.	

Теорема 2.1.31 *Предположим что $\|g_k\|_* \leq L$, $k \geq 0$. Для метода (2.1.31) имеем $S_k = k + 1$ и $\delta_k(D) \leq \hat{\beta}_{k+1} \left(\gamma D + \frac{L^2}{2\sigma\gamma} \right)$.*

Покажем как эта техника работает на примере задачи

$$\min_x \{f(x) : x \in Q\}, \quad (2.1.32)$$

где f - выпуклая на E функция и множество Q является выпуклым и замкнутым. В этом случае имеется следующая интерпретация функции зазора $\delta_k(D)$. Обозначим

$$\begin{aligned} l_k(x) &= \frac{1}{S_k} \sum_{i=0}^k \lambda_i [f(x_i) + \langle g_i, x - x_i \rangle], \quad (g_i \in \partial f(x_i)) \\ \hat{f}_k(D) &= \min_x \{l_k(x) : x \in \mathcal{F}_D\}, \\ f_D^* &= \min_x \{f(x) : x \in \mathcal{F}_D\}. \end{aligned}$$

Поскольку функция $f(\cdot)$ выпукла, в силу определения функции зазора (2.1.20) имеем:

$$\frac{1}{S_k} \delta_k(D) = \frac{1}{S_k} \sum_{i=0}^k \lambda_i f(x_i) - \hat{f}_k(D) \geq f(\hat{x}_{k+1}) - f_D^*.$$

Таким образом, в силу Теоремы 2.1.31, имеем

$$f(\hat{x}_{k+1}) - f_D^* \leq \frac{0.5 + \sqrt{2k+1}}{k+1} \left(\gamma D + \frac{L^2}{2\sigma\gamma} \right).$$

Соответствующая алгоритмическая схема выглядит так:

$$x_{k+1} = \arg \min_{x \in Q} \left\{ \frac{1}{k+1} \sum_{i=0}^k [f(x_i) + \langle g_i, x - x_i \rangle] + \frac{\beta_{k+1}}{k+1} d(x) \right\}, \quad k \geq 0,$$

Для этого метода не требуется ограниченность множества Q . В то же время, вычисляемые оценки $\hat{f}_k(D)$ обеспечивают надежный контроль точности. В оставшейся части параграфа рассмотрены применения ДУ-методов к минимаксным задачам, задачам с функциональными ограничениями, седловым задачам, вариационным неравенствам и задачам стохастической оптимизации.

В параграфе 2.2 предлагается *барьерный субградиентный метод* (БСМ). В нем сглаживание функции зазора производится с помощью самосогласованного барьера. Пусть выпуклое замкнутое множество $Q \subset E$ не содержит прямых и снабжено ν -самосогласованным барьером F . Для произвольного $x \in \text{int } Q$ введем следующую локальную норму:

$$\|s\|_x^* = \langle s, [\nabla^2 F(x)]^{-1} s \rangle^{1/2}, \quad s \in E^*.$$

Рассмотрим другое выпуклое множество $\hat{P} \subseteq E$. Нас будет интересовать их пересечение $P = \hat{P} \cap Q$, которое предполагается ограниченным. Обозначим через x_0 его *условный аналитический центр*:

$$x_0 = \arg \min_{x \in P_0} F(x) \in P_0 \stackrel{\text{def}}{=} \hat{P} \cap \text{int } Q \subseteq P.$$

Для множества P , введем гладкую аппроксимацию ее опорной функции:

$$U_\beta(s) = \max_{x \in \hat{P}} \{ \langle s, x - x_0 \rangle - \beta [F(x) - F(x_0)] \}, \quad s \in E^*, \quad (2.2.3)$$

где $\beta > 0$ - это параметр сглаживания. Обозначим через $u_\beta^*(s)$ единственное решение задачи максимизации (2.2.3).

Рассмотрим задачу выпуклой оптимизации в следующей форме:

$$\text{Найти } f_\star \stackrel{\text{def}}{=} \max_x \{ f(x) : x \in P \}, \quad (2.2.13)$$

где f - это вогнутая функция. Мы предполагаем что f супердифференцируема на P_0 и что множество P является простым. Приведем

общую схему БСМ.

Инициализация: Полагаем $s_0 = 0 \in E^*$.	(2.2.14)
Итерация ($k \geq 0$):	
<ol style="list-style-type: none"> 1. Выбираем $\beta_k > 0$ и вычисляем $x_k = u_{\beta_k}^*(s_k)$. 2. Выбираем $\lambda_k > 0$ и полагаем $s_{k+1} = s_k + \lambda_k \nabla f(x_k)$. 	

Для анализа скорости БСМ нам потребуется следующие *функции зазора*:

$$l_k(y) = \sum_{i=0}^k \lambda_i \langle \nabla f(x_i), y - x_i \rangle, \quad l_k^* \stackrel{\text{def}}{=} \max_{y \in P} l_k(y), \quad k \geq 0.$$

Теорема 2.2.1 Пусть параметры метода (2.2.14) удовлетворяют соотношению

$$\lambda_k \|\nabla f(x_k)\|_{x_k}^* \leq \beta_k \leq \beta_{k+1}, \quad k \geq 0. \quad (2.2.15)$$

Обозначим $S_k = \sum_{i=0}^k \lambda_i$, и $A_k = \sum_{i=0}^k \beta_i \omega_* \left(\frac{\lambda_i}{\beta_i} \|\nabla f(x_i)\|_{x_i}^* \right)$, где $\omega_*(\tau) = -\tau - \ln(1 - \tau)$. Тогда

$$l_k^* \leq A_k + \beta_{k+1} \nu \left[1 + 2 \ln \left(1 + \sqrt{\frac{A_k}{\beta_{k+1} \nu} + c(Q) \frac{S_k}{\beta_{k+1}} \|\nabla f(x_0)\|_{x_0}^*} \right) \right], \quad k \geq 0, \quad (2.2.16)$$

где $c(Q) = 1$ если Q является конусом и барьер $F(\cdot)$ логарифмически однороден. В противном случае, $c(Q) = 3$.

Оценим теперь скорость сходимости БСМ при решении некоторых специальных задач.

Определение 2.2.1 Мы говорим что $f \in \mathcal{B}_M(P)$ если $\|\nabla f(x)\|_x^* \leq M$ для любого $x \in P_0$.

Для функции $f \in \mathcal{B}_M(P)$, в БСМ можно выбрать следующие значения параметров:

$$\lambda_k = 1, \quad k \geq 0, \quad \beta_0 = \beta_1, \quad \beta_k = M \cdot \left(1 + \sqrt{\frac{k}{\nu}} \right), \quad k \geq 1. \quad (2.2.19)$$

Теорема 2.2.2 Пусть задача (2.2.13) с $f \in \mathcal{B}_M(P)$ решается методом (2.2.14) с параметрами заданными в (2.2.19). Тогда при любом $k \geq 0$ имеем:

$$\frac{1}{S_k} l_k^* \leq 2M \cdot \left(\sqrt{\frac{\nu}{k+1}} + \frac{\nu}{k+1} \right) \cdot \left(1 + \ln \left(2 + \frac{1}{2} c(Q) \sqrt{\nu(k+1)} \right) \right). \quad (2.2.20)$$

Интересным приложением для БСМ является решение задачи максимизации положительной вогнутой функции. В этом случае он обеспечивает нахождение решения с относительной точностью с трудоемкостью, которая никак не зависит от свойств конкретной задачи (например, размера области). Другими примерами приложений являются вычисление дробных покрытий, вычисление максимальных многопродуктовых потоков, минимаксные задачи с неотрицательными коэффициентами, полуопределенная релаксация задач булевой квадратичной оптимизации и оптимизация в реальном времени как альтернатива стохастической оптимизации.

В параграфе 2.3 рассматриваются градиентные методы для минимизации составных функций. Предполагается что целевая функция задачи имеет вид

$$\phi(x) = f(x) + \Psi(x), \quad (2.3.1)$$

где $f(x)$ - дифференцируемая функция заданная чернойщичным оракулом, а $\Psi(x)$ - выпуклая замкнутая *простая* функция общего вида. Для любого $y \in Q$ определим *составное градиентное отображение*:

$$\begin{aligned} m_L(y; x) &= f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2} \|x - y\|^2 + \Psi(x), \\ T_L(y) &= \arg \min_{x \in Q} m_L(y; x), \end{aligned} \quad (2.3.10)$$

где норма $\| \cdot \|$ евклидова ($\|h\| = \langle Bh, h \rangle^{1/2}$, $B = B^T \succ 0$), а L - положительный параметр (в обычном градиентном отображении $\Psi(\cdot) \equiv 0$). Теперь можно определить аналог градиентного направления для гладкой функции:

$$g_L(y) = L \cdot B(y - T_L(y)) \in E^*. \quad (2.3.11)$$

Если $Q \equiv E$ и $\Psi \equiv 0$, то $g_L(y) = \nabla \phi(y) \equiv \nabla f(x)$ при любых $L > 0$. Наше предположение о простоте функции Ψ означает в точности реализуемость

операции (2.3.10). Теперь можно определить градиентную итерацию.

Градиентная итерация $\mathcal{G}(x, M)$	
ПОЛАГАЕМ:	$L := M.$
ПОВТОРЯЕМ:	$T := T_L(x),$ Если $\phi(T) > m_L(x; T)$ то $L := L \cdot \gamma_u,$
ПОКА НЕ ВЫПОЛНИТСЯ:	$\phi(T) \leq m_L(x; T).$
ВЫХОД:	$\mathcal{G}(x, M).T = T, \quad \mathcal{G}(x, M).L = L,$ $\mathcal{G}(x, M).S = S_L(x).$

(2.3.26)

Обычно выбирают $\gamma_u = \gamma_d = 2$. Теперь можно рассмотреть прямой градиентный метод.

Прямой градиентный метод $\mathcal{GM}(y_0, L_0)$	
$y_{k+1} = \mathcal{G}(y_k, L_k).T,$	(2.3.28)
$M_k = \mathcal{G}(y_k, L_k).L,$	
$L_{k+1} = \max\{L_0, M_k/\gamma_d\}.$	

Можно показать что N_k , общее число вызовов оракула после k итераций метода (2.3.28) не может быть большим: $N_k \leq \left[1 + \frac{\ln \gamma_d}{\ln \gamma_u}\right] \cdot (k + 1) + \frac{1}{\ln \gamma_u} \cdot \left(\ln \frac{\gamma_u L_f}{\gamma_d L_0}\right)_+.$

Теорема 2.3.4 Пусть функция f выпукла на Q . Предположим что она достигает своего минимума на Q в точке x^* и что ее множества уровня ограничены:

$$\|y - x^*\| \leq R \quad \forall y \in Q : \phi(y) \leq \phi(y_0). \quad (2.3.35)$$

Если $\phi(y_0) - \phi(x^*) \geq \gamma_u L_f R^2$, то $\phi(y_1) - \phi(x^*) \leq \frac{\gamma_u L_f R^2}{2}$. В противном случае,

$$\phi(y_k) - \phi(x^*) \leq \frac{2\gamma_u L_f R^2}{k+2}, \quad k \geq 0. \quad (2.3.36)$$

К составным функциям можно применять и быстрый градиентный метод.

Быстрый градиентный метод $\mathcal{A}(x_0, L_0, \mu)$	
Инициализация: $\psi_0(x) = \frac{1}{2}\ x - x_0\ ^2, A_0 = 0.$	
Итерация $k \geq 0$	
ПОЛАГАЕМ:	$L := L_k.$
ПОВТОР:	Найти a из квадратного уравнения $\frac{a^2}{A_k+a} = 2\frac{1+\mu A_k}{L}.$ Полагаем $y = \frac{A_k x_k + a v_k}{A_k + a}$, и вычисляем $T_L(y).$
Если:	$\langle \phi'(T_L(y)), y - T_L(y) \rangle \geq \frac{1}{L} \ \phi'(T_L(y))\ _*^2.$ То СТОП Повтор Иначе $L := L \cdot \gamma_u.$
ПОЛАГАЕМ:	$y_k := y, M_k := L, a_{k+1} := a,$ $L_{k+1} := M_k / \gamma_d, x_{k+1} := T_{M_k}(y_k),$ $\psi_{k+1}(x) := \psi_k(x) + a_{k+1}[f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1} \rangle + \Psi(x)].$

(2.3.52)

Теорема 2.3.6 *Предположим что градиент функции f удовлетворяет условию Липшица с константой L_f . И пусть параметр $L_0 \in (0, L_f)$. Тогда скорость сходимости метода $\mathcal{A}(x_0, L_0, 0)$ оценивается следующим образом:*

$$\phi(x_k) - \phi(x^*) \leq \frac{\gamma_u L_f \|x^* - x_0\|^2}{k^2}, \quad k \geq 1. \quad (2.3.53)$$

Мы видим что составное градиентное отображение полностью устраняет негативное влияние простой добавки на скорость сходимости соответствующих градиентных методов.

В **Главе 3** рассматриваются методы решения вариационных неравенств. Основные результаты этой главы опубликованы в работах [13, 20].

Вариационные неравенства (ВН) с монотонными операторами представляют собой максимально общий класс задач, обладающих выпуклой структурой¹¹. В этом виде могут быть сформулированы выпуклые задачи оптимизации, задачи нахождения седловых точек у выпукло-вогнутых функций, различные задачи нахождения равновесия и т.д. В параграфе 3.1 мы рассматриваем методы для гладких монотонных операторов. Они основываются на простых правилах пересчета некоторой линейной функции, которая пред-

¹¹Вайнберг М. М. Вариационный метод и метод монотонных операторов в теории нелинейных уравнений. - М.: Наука, 1972. - 385с.

ставляет собой модель нашей задачи. Это правило называется шагом *двойственной экстраполяции*. Применяя это правило последовательно, мы получим простую итеративную схему с оценкой сложности $O(\frac{1}{\epsilon})$, справедливой для ВН с липшицевым монотонным оператором. Та же схема с правильной стратегией одномерного поиска может быть применена к монотонным оператором с *ограниченным изменением* и, в частности, к негладким задачам. В этом случае оценка сложности превращается в $O(\frac{1}{\epsilon^2})$. Опишем подробнее нашу конструкцию.

Нас будет интересовать следующая задача решения *вариационного неравенства*:

$$\text{Найти } x^* \equiv x^*(Q) \in Q : \quad \langle g(x), x - x^* \rangle \geq 0 \quad \forall x \in Q, \quad (3.1.1)$$

где оператор g монотонен. Пусть функция $d(x)$ является сильно выпуклой на Q с параметром выпуклости $\sigma > 0$. Определим *брегмановское расстояние*

$$\rho(x, y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle, \quad x \in \tilde{Q}, y \in Q.$$

Выберем произвольную точку $\bar{x} \in \tilde{Q}$ в качестве *центра* множества Q . Заметим что множество Q может быть неограниченным. Мы будем описывать качество произвольной точки $x \in Q$ как приближенного решения задачи (3.1.1) с помощью следующей *условной функции близости*:

$$f_D(x) = \max_{y \in Q} \{ \langle g(y), x - y \rangle : \rho(\bar{x}, y) \leq D \},$$

где D - фиксированный положительный параметр.

Теперь мы можем определить основной шаг нашего метода (*двойственную экстраполяцию*). Обозначим

$$T_\beta(z, s) = \arg \max_x \{ \langle s, x - z \rangle - \beta \rho(z, x) : x \in Q \}.$$

Зафиксируем некоторые положительные значения β и λ . Рассмотрим отображение $\mathcal{E}_{\beta, \lambda}(s)$, которое переводит произвольную точку $s \in E^*$ в новую точку s_+ :

$$(x, y, s_+) = \mathcal{E}_{\beta, \lambda}(s) \quad \Leftrightarrow \quad \begin{cases} x &= T_\beta(\bar{x}, s), \\ y &= T_\beta(x, -\lambda g(x)), \\ s_+ &= s - \lambda g(y). \end{cases}$$

Этот шаг может применяться для решения ВН различной степени гладкости.

Метод решения гладких вариационных неравенств	
Инициализация:	Выбираем $\bar{x} \in \tilde{Q}$. Фиксируем $\beta = \frac{L}{\sigma}$ и выбираем $D > 0$. Задаем точность $\epsilon > 0$. Полагаем $s_{-1} = 0 \in E^*$.
Итерация:	$(x_k, y_k, s_k) = \mathcal{E}_{\beta,1}(s_{k-1}), k \geq 0$.
:	Если $\sum_{i=0}^k \langle g(y_i), y_i - \bar{x} \rangle + \xi_D(s_k) \leq \epsilon \cdot (k + 1)$, то Стоп.
Результат:	$\tilde{y}_k = \frac{1}{k+1} \sum_{i=0}^k y_i$.

(3.1.20)

Теорема 3.1.2 Пусть оператор $g(x)$ будет липшицевым на Q . Тогда

$$f_D(\tilde{y}_k) \leq \frac{LD}{\sigma \cdot (k+1)}, \quad k \geq 0. \quad (3.1.21)$$

Критерий прерывания в методе (3.1.20) обеспечивает неравенство $f_D(\tilde{y}_k) \leq \epsilon$. Это критерий будет выполнен не более чем через $1 + \lfloor \frac{LD}{\sigma\epsilon} \rfloor$ итераций.

Метод для ВН с ограниченным изменением оператора	
Инициализация:	Выбираем $\bar{x} \in \tilde{Q}$ и $D > 0$. Для $k \geq 0$ положим $\beta_k = M \sqrt{\frac{k+1}{\sigma D}}.$ Выберем точность $\epsilon > 0$. Положим $s_{-1} = 0 \in E^*$.
Итерация:	$(x_k, y_k, s_k) = \mathcal{E}_{\beta_k,1}(s_{k-1}), k \geq 0$.
Остановка:	Если $\sum_{i=0}^k \langle g(y_i), y_i - \bar{x} \rangle + \xi_D(s_k) \leq \epsilon \cdot (k + 1)$, то Стоп.
Результат:	$\tilde{y}_k = \frac{1}{k+1} \sum_{i=0}^k y_i$.

(3.1.23)

Теорема 3.1.3 Предположим что оператор $g(x)$ имеет ограниченное изменение на множестве Q :

$$\|g(x_1) - g(x_2)\|_* \leq M \quad \forall x_1, x_2 \in Q.$$

Тогда для любого $k \geq 0$ имеем: $f_D(\tilde{y}_k) \leq 2M \sqrt{\frac{D}{\sigma \cdot (k+1)}}$. Критерий прерывания в методе (3.1.23) обеспечивает выполнение неравенства $f_D(\tilde{y}_k) \leq \epsilon$. Этот критерий сработает не более чем за $1 + \lfloor \frac{4M^2 D}{\sigma \epsilon^2} \rfloor$ итераций.

В заключение параграфа мы показываем как применять полученные методы для вычисления седловых точек и для решения билинейных матричных игр,

В параграфе 3.2 мы показываем как решать сильно монотонные ВН и квазивариационные неравенства. Рассмотрим непрерывный оператор $g(x) : Q \rightarrow E^*$, который является *сильно монотонным*:

$$\langle g(x) - g(y), x - y \rangle \geq \mu \|x - y\|^2, \quad \forall x, y \in Q. \quad (3.2.2)$$

Мы будем искать решение следующего ВН:

$$\text{Найти } x^*(Q) \in Q : \langle g(x^*(Q)), y - x^*(Q) \rangle \geq 0 \quad \forall y \in Q. \quad (3.2.3)$$

Для описания качества приближенного решения задачи (3.2.3), нам потребуется следующая *мера близости*:

$$f(x) = \sup_{y \in Q} \{ \langle g(y), x - y \rangle + \frac{1}{2} \mu \|y - x\|^2 \}. \quad (3.2.4)$$

Для $\beta > 0$, обозначим

$$\begin{aligned} \psi_y^\beta(x) &= \langle g(y), y - x \rangle - \frac{1}{2} \beta \|x - y\|^2, \\ \Psi_k(x) &= \sum_{i=0}^k \lambda_i \psi_{y_i}^\mu(x), \quad S_k = \sum_{i=0}^k \lambda_i. \end{aligned}$$

Пусть оператор g является липшицевым с константой L . Для простоты предположим что константы μ и L нам известны. Обозначим через $\gamma = \frac{L}{\mu} \geq 1$ *число обусловленности* оператора g .

Метод для сильно монотонных ВН	
Инициализация:	Выбираем $\bar{x} \in Q$. Полагаем $\lambda_0 = 1$, и $y_0 = \bar{x}$.
Итерация	$(k \geq 0)$: $x_k = \arg \max_{x \in Q} \Psi_k(x),$ $y_{k+1} = \arg \max_{x \in Q} \psi_{x_k}^L(x),$ $\lambda_{k+1} = \frac{1}{\gamma} \cdot S_k.$
Результат:	$\tilde{y}_k = \frac{1}{S_k} \sum_{i=0}^k \lambda_i y_i.$

(3.2.11)

Теорема 3.2.3 *Если оператор g сильно монотонен и липшицев, то*

$$\frac{\mu}{2} \cdot \|\tilde{y}_k - x^*(Q)\|^2 \leq f(\tilde{y}_k) \leq f(\bar{x}) \cdot \gamma^2 \cdot \exp\left\{-\frac{k}{\gamma+1}\right\}, \quad k \geq 0. \quad (3.2.12)$$

Во второй части параграфа рассматриваются *квазивариационные неравенства* (КВН):

$$\text{Найти } x_* \in Q(x_*) : \quad \langle g(x_*), y - x_* \rangle \geq 0, \quad \forall y \in Q(x_*), \quad (3.2.14)$$

где точечно-множественное отображение $Q : E \rightarrow 2^E$ принимает выпуклые и замкнутые значения. Существование решения этой задачи гарантируется только при малой скорости изменения множеств $Q(x)$ ¹²: существует $\alpha < \frac{1}{\gamma(\gamma + \sqrt{\gamma^2 - 1})}$ такая что

$$\|\pi_{Q(x)}(z) - \pi_{Q(y)}(z)\| \leq \alpha \|x - y\|, \quad \forall x, y, z \in E, \quad (3.2.15)$$

где $\pi_Q(x)$ - евклидова проекция точки x на множество Q . Нам удастся существенно ослабить это условие с помощью новой алгоритмической схемы.

Обозначим через $\tilde{y}_N(Q, \hat{x})$ результат работы N -шагового метода (3.2.11), применяемого к решению ВН с оператором g на множестве Q и центральной точкой $\bar{x} = \arg \max_{x \in Q} \psi_{\hat{x}}^L(x)$. Обозначим $\delta \stackrel{\text{def}}{=} 1 - \alpha\gamma$ и предположим что оно положительно. Зададим

$$\hat{N} = \lfloor 2(\gamma + 1) \ln \frac{12\gamma}{1 - \alpha\gamma} \rfloor + 1,$$

¹²Noor M. A., Oettli W. On general nonlinear complementarity problems and quasi-equilibria // *Le Matematiche* - 1994 - V.XLIX - 313-331p.

и рассмотрим следующую двухуровневую схему.

$$\begin{aligned} & \text{Выбираем } u_0 \in E. \text{ Для } k \geq 0 \text{ повторяем:} \\ & \hat{x}_k = \pi_{\mathcal{Q}(u_k)}(u_k), \quad u_{k+1} = \tilde{y}_{\hat{N}}(\mathcal{Q}(u_k), \hat{x}_k). \end{aligned} \quad (3.2.30)$$

Теорема 3.2.8 Пусть $\delta > 0$. Тогда существует единственное решение x_* КВН (3.2.14), и метод (3.2.30) сходится со следующей скоростью:

$$\|u_k - x_*\| \leq \frac{1}{\delta} \cdot \exp\left\{-\frac{\delta}{2}k\right\} \cdot \|u_0 - T(u_0)\|. \quad (3.2.31)$$

В **Главе 4** рассматриваются методы второго порядка. Представленные результаты опубликованы в работах [14, 16, 11].

В параграфе 4.1 описывается кубическая регуляризация метода Ньютона. Метод с таким названием появился впервые в статье А.Беннета¹³. Первый результат о локальной квадратичной сходимости метода принадлежит Л.В.Канторовичу¹⁴. Однако оценки глобальной скорости сходимости этого метода на вырожденных задачах так и не были получены¹⁵.

Мы рассматриваем задачу безусловной минимизации функции f с липшицевым Гессианом:

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\|, \quad \forall x, y \in R^n.$$

Тогда нетрудно видеть что вспомогательная функция

$$\xi_{2,x}(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle \nabla^2 f(x)(y - x), y - x \rangle + \frac{L}{6} \|y - x\|^3$$

является *верхней оценкой второго порядка* для целевой функции:

$$f(y) \leq \xi_{2,x}(y) \quad \forall y \in R^n.$$

Таким образом, можно попытаться формировать следующую тестовую точку в методе второго порядка с помощью следующей задачи минимизации:

$$x_{k+1} \in \text{Arg min}_y \xi_{2,x_k}(y) \quad (4.1.1)$$

¹³Bennet A.A. Newton's method in general analysis. // Proc. Nat. Ac. Sci. USA - 1916 - V.2, N10. - P.592-598.

¹⁴Канторович Л.В. Функциональный анализ и прикладная математика // Успехи мат. наук - 1948 - Т.3, вып.1 - С.89-185.

¹⁵Conn A.B., Gould N.I.M., Toint Ph.L. Trust Region Methods - Philadelphia: SIAM, 2000. - 1005p.

(здесь Argmin обозначает глобальный минимум). Этот подход называется *кубической регуляризацией* метода Ньютона. Заметим что задача (4.1.1) может быть невыпуклой и может иметь много локальных минимумов. Однако, наш подход реализуем за полиномиальное время так как эта вспомогательная задача оказывается эквивалентной задаче минимизации специальной *выпуклой функции* одной переменной.

Положим

$$\bar{f}_M(x) = \min_y \left[f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle \nabla^2 f(x)(y - x), y - x \rangle + \frac{M}{6} \|y - x\|^3 \right].$$

Введем отображение

$$T_M(x) \in \text{Arg min}_y \left[\langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle \nabla^2 f(x)(y - x), y - x \rangle + \frac{M}{6} \|y - x\|^3 \right], \quad (4.1.5)$$

где "Arg" означает что $T_M(x)$ выбирается из множества глобальных минимумов соответствующей задачи минимизации. Пусть $L_0 \in (0, L]$ - положительный параметр. Рассмотрим следующий метод.

Кубическая регуляризация метода Ньютона	
Инициализация: Выбираем $x_0 \in R^n$.	
Итерация k, ($k \geq 0$):	
1. Находим $M_k \in [L_0, 2L]$ такое что	
$f(T_{M_k}(x_k)) \leq \bar{f}_{M_k}(x_k)$.	
2. Полагаем $x_{k+1} = T_{M_k}(x_k)$.	(4.1.16)

Поскольку $\bar{f}_M(x) \leq f(x)$, этот процесс является монотонным: $f(x_{k+1}) \leq f(x_k)$. Про эту схему удастся доказать много интересных результатов.

- Для невыпуклых функций, для нормы градиента справедлива глобальная оценка $O(1/k^{2/3})$, где k - это счетчик итераций. В невырожденной ситуации, метод обладает локальной квадратичной скоростью сходимости. Невырожденные стационарные точки и точки максимума являются точками отталкивания для минимизирующей последовательности.
- В выпуклой ситуации, глобальная скорость сходимости по невязке по функции составляет $O(1/k^2)$.

- Получены глобальные оценки скорости сходимости для звездно-выпуклых функций, градиентно доминируемых функций и функций, оценивающих невязку в системе нелинейных невырожденных уравнений.

В параграфе 4.2 описывается ускоренная версия кубической регуляризации метода Ньютона, применяемой для минимизации выпуклой функции. Ускорение стало возможным за счет применения аппарата оценочных функций (см. §§ 2.2 и 3.3).

Ускоренная версия кубической регуляризации
Инициализация: Выбираем $x_0 \in E$. Полагаем $M = 2L$ и $N = 12L$. Вычисляем $x_1 = T_{L_3}(x_0)$ и полагаем $\psi_1(x) = f(x_1) + \frac{N}{6}\ x - x_0\ ^3$.
Итерация k, ($k \geq 1$): <ol style="list-style-type: none"> 1. Вычисляем $v_k = \arg \min_{x \in E} \psi_k(x)$ и полагаем $y_k = \frac{k}{k+3}x_k + \frac{3}{k+3}v_k$. 2. Вычисляем $x_{k+1} = T_M(y_k)$ и пересчитываем $\psi_{k+1}(x) = \psi_k(x) + \frac{(k+1)(k+2)}{2} \cdot [f(x_{k+1}) + \langle \nabla f(x_{k+1}), x - x_{k+1} \rangle]$.

(4.2.28)

Теорема 4.2.2 Если последовательность точек $\{x_k\}_{k=1}^{\infty}$ сформирована методом (4.2.28), то для любого $k \geq 1$ выполняется неравенство:

$$f(x_k) - f(x^*) \leq \frac{14 L_3 \|x_0 - x^*\|^3}{k(k+1)(k+2)}, \quad (4.2.29)$$

где x^* - оптимальное решение задачи, а L_3 - константа Липшица для Гесса целевой функции.

В параграфе 4.3 описывается модифицированный метод Гаусса-Ньютона. Он применяется для нахождения приближенного решения уравнения

$$F(x) = 0, \quad x \in E_1. \quad (4.3.4)$$

Для измерения качества такого решения, вводится функция близости $\phi(u)$, $u \in E_2$, которая выпукла, неотрицательна и обнуляется только в начале координат. Она должна удовлетворять условию Липшица с единичной константой и иметь острый минимум в нуле:

$$\phi(u) \geq \gamma_{\phi} \|u\|, \quad \forall u \in E_2, \quad (4.3.5)$$

при некотором $\gamma_\phi \in (0, 1]$. Например, можно взять $\phi(u) = \|u\|$. Тогда $\gamma_\phi = 1$.

Таким образом, задача (4.3.4) преобразуется в следующую задачу *безусловной минимизации*:

$$\min_{x \in E_1} \{ f(x) \equiv \phi(F(x)) \} \stackrel{\text{def}}{=} f^*. \quad (4.3.6)$$

Рассматривая *локальную модель* нашей целевой функции:

$$\psi(x; y) = \phi(F(x) + \nabla F(x)(y - x)), \quad y \in E_1,$$

приходим к следующему методу *Гаусса-Ньютона*

$$x_{k+1} \in \text{Arg} \min_{y \in E_1} \psi(x_k; y).$$

Такие схемы очень хорошо изучены^{16,17}. Однако глобальная эффективность таких методов до сих пор неизвестна.

В этом параграфе, для функций с липшицевым якобианом

$$\|\nabla F(x) - \nabla F(y)\|_* \leq L\|x - y\|, \quad \forall x, y \in \mathcal{F}, \quad (4.3.7)$$

предлагается *модифицированный метод Гаусса-Ньютона*:

$$x_{k+1} \in \text{Arg} \min_{y \in E_1} [\psi(x; y) + \frac{1}{2}M\|y - x\|^2],$$

где "Arg" означает что точка x_{k+1} выбрана из множества глобальных минимумов соответствующей задачи минимизации. Этот процесс обеспечивает монотонное убывание целевой функции. Более того, при условии глобальной невырожденности системы уравнений удастся доказать глобальную оценку скорости сходимости процесса. При естественных предположениях, локальная сходимость будет квадратичной.

В **Главе 5** описывается техника сглаживания, применяемая для построения ускоренных методов минимизации негладких функций. Основные результаты главы опубликованы в работах [9, 10, 12].

¹⁶Ortega J., Rheinboldt W. Iterative Solution of Nonlinear Equations in Several Variables. - NY: Academic Press, 1970. - 630p.

¹⁷Conn A.R., Gould N.I.M., Toint Ph.L. Trust Region Methods. - Philadelphia: SIAM, 2000. - 1005p.

В параграфе 5.1 рассматривается задача минимизации

$$\text{Find } f^* = \min_x \{f(x) : x \in Q_1\}, \quad (5.1.3)$$

где выпуклое замкнутое множество Q_1 в конечномерном вещественном пространстве E_1 является ограниченным, а выпуклая функция $f(x)$ непрерывна на Q_1 . Очень часто известна *структура* целевой функции в задаче (5.1.3):

$$f(x) = \hat{f}(x) + \max_u \{\langle Ax, u \rangle_2 - \hat{\phi}(u) : u \in Q_2\}, \quad (5.1.4)$$

где $\langle \cdot, \cdot \rangle_2$ - скалярное произведение в пространстве E_2 , выпуклая функция $\hat{f}(x)$ непрерывна на Q_1 , выпуклое замкнутое множество Q_2 из конечномерного вещественного пространства E_2 ограничено, выпуклая функция $\hat{\phi}(u)$ непрерывна на Q_2 , а линейный оператор A отображает E_1 в E_2^* . В этом случае задача (5.1.3) может быть переписана в *сопряженной форме*:

$$\max_u \{\phi(u) : u \in Q_2\}, \quad (5.1.5)$$

$$\phi(u) = -\hat{\phi}(u) + \min_x \{\langle Ax, u \rangle_2 + \hat{f}(x) : x \in Q_1\}.$$

Рассмотрим *прокс-функцию* $d_2(u)$ множества Q_2 с параметром выпуклости $\sigma_2 = 1$. Обозначим через

$$u_0 = \arg \min_u \{d_2(u) : u \in Q_2\}$$

ее *прокс-центр*. И пусть $d_2(u_0) = 0$. Выберем положительное μ в качестве *параметра сглаживания*. Рассмотрим следующую функцию (пусть $\hat{f}(x) \equiv 0$):

$$f_\mu(x) = \max_u \{\langle Ax, u \rangle_2 - \hat{\phi}(u) - \mu d_2(u) : u \in Q_2\}. \quad (5.1.6)$$

Обозначим через $u_\mu(x)$ единственное оптимальное решение этой задачи.

Можно показать что функция f_μ является $O(\mu)$ -аппроксимацией исходной функции. С другой стороны, ее градиент является липшицевым с константой $O(1/\mu)$. Это обстоятельство позволяет находить ϵ -решение задачи (5.1.3) за $O(1/\epsilon)$ итераций быстрого градиентного метода (см. §1.2), применяемого к задаче минимизации сглаженной функции f_μ . Получаемые оценки трудоемкости на порядок улучшают нижние оценки трудоемкости черноточечных методов, полученных в теории сложности.

В конце параграфа рассматриваются применения техники сглаживания для нахождения решений антагонистических матричных игр, непрерывной задачи размещения, вариационных неравенств с линейным оператором и минимизации кусочно-линейных функций. Обсуждаются вопросы вычислительной устойчивости и приводятся обнадеживающие результаты численных экспериментов.

В параграфе 5.2 рассматривается ситуация когда прямая и сопряженная форма задачи (5.1.3) имеют одинаковую сложность. В этом случае можно применять симметричные методы, которые чередуют шаги в прямом и сопряженном пространстве.

Пользуясь обозначениями §5.1, заметим что при любых $x \in Q_1$ и $u \in Q_2$ выполняется неравенство

$$\phi(u) \leq f(x). \quad (5.2.12)$$

Более того, при естественных предположениях в задачах (5.1.3), (5.1.5) будет нулевой зазор двойственности. Однако, $f_{\mu_2}(x) \leq f(x)$ и $\phi(u) \leq \phi_{\mu_1}(u)$, где последняя функция определена аналогично (5.1.6). Такая ситуация открывает принципиальную возможность обеспечить выполнение условия *обратного зазора*:

$$f_{\mu_2}(\bar{x}) \leq \phi_{\mu_1}(\bar{u}) \quad (5.2.13)$$

при некоторых $\bar{x} \in Q_1$ и $\bar{u} \in Q_2$. Ясно что из условия (5.2.13) нетрудно получить оценку качества соответствующей прямо-двойственной пары (\bar{x}, \bar{u}) :

$$0 \leq \max\{f(\bar{x}) - f^*, f^* - \phi(\bar{u})\} \leq f(\bar{x}) - \phi(\bar{u}) \leq \mu_1 D_1 + \mu_2 D_2, \quad (5.2.14)$$

где $D_1 = \max_{x \in Q_1} d_1(x)$, и $D_2 = \max_{u \in Q_2} d_2(u)$.

Оказывается условие обратного зазора (5.2.13) можно поддерживать при параметрах сглаживания стремящихся к нулю. Покажем как это можно сделать, уменьшив μ_1 и не меняя μ_2 .

Для $x \in Q_1$ определим *прямое градиентное отображение*:

$$T_{\mu_2}(x) = \arg \min_{y \in Q_1} \left\{ \langle \nabla f_{\mu_2}(x), y - x \rangle_1 + \frac{1}{2} L_1(f_{\mu_2}) \|y - x\|_1^2 \right\}. \quad (5.2.16)$$

Теорема 5.2.1 Пусть точки $\bar{x} \in Q_1$ и $\bar{u} \in Q_2$ удовлетворяют условию обратного зазора (5.2.13) при некоторых положительных μ_1 и μ_2 . Зафиксируем $\tau \in (0, 1)$ и выберем $\mu_1^+ = (1 - \tau)\mu_1$,

$$\begin{aligned}\hat{x} &= (1 - \tau)\bar{x} + \tau x_{\mu_1}(\bar{u}), \\ \bar{u}_+ &= (1 - \tau)\bar{u} + \tau u_{\mu_2}(\hat{x}), \\ \bar{x}_+ &= T_{\mu_2}(\hat{x}).\end{aligned}\tag{5.2.19}$$

Тогда пара (\bar{x}_+, \bar{u}_+) удовлетворяет условию (5.2.13) с параметрами сглаживания μ_1^+ и μ_2 при всех не слишком больших τ :

$$\frac{\tau^2}{1 - \tau} \leq \frac{\mu_1}{L_1(f_{\mu_2})}.\tag{5.2.20}$$

Это правило пересчета, применяемое поочередно в прямом и сопряженном пространстве, позволяет решать задачи минимизации со структурой (5.1.4) за $O(1/\epsilon)$ итераций. В конце параграфа мы также рассматриваем сильно выпуклый случай и показываем как построить алгоритм его решения с трудоемкостью $O(1/\epsilon^{1/2})$ итераций.

В параграфе 5.3 мы рассматриваем применение техники сглаживания в полуопределенной оптимизации. А именно, мы показываем как можно сглаживать симметричные функции собственных значений симметрических матриц.

Прежде всего нам потребуются условия гладкости симметричных функций собственных значений.

Пусть функция одной вещественной переменной $f(\tau)$ задана рядом

$$f(\tau) = a_0 + \sum_{k=1}^{\infty} a_k \tau^k$$

с $a_k \geq 0$ при $k \geq 2$. Предположим что ее область определения $\text{dom } f = \{\tau : |\tau| < R\}$ не пуста. Для симметрической матрицы X рассмотрим следующую симметричную функцию собственных значений:

$$F(X) = \sum_{i=1}^n f(\lambda^{(i)}(X)).$$

Ясно что $\text{dom } F = \{X \in \mathcal{S}_n : \lambda_{\max}(X) < R, \lambda_{\min}(X) > -R\}$.

Теорема 5.3.1 Для любого $X \in \text{dom } F$ и $H \in \mathcal{S}_n$ имеем

$$\langle \nabla^2 F(X)H, H \rangle \leq \sum_{i=1}^n \nabla^2 f(\lambda^{(i)}(|X|))(\lambda^{(i)}(|H|))^2.$$

С помощью этого факта можно оценивать степень гладкости различных симметричных функций собственных значений. Приведем два примера.

1. Квадрат матричной l_p -нормы. Для целого $p \geq 1$ рассмотрим следующую функцию:

$$F_p(X) = \frac{1}{2} \|\lambda(X)\|_{(2p)}^2 = \frac{1}{2} \langle X^{2p}, I_n \rangle_M^{1/p}.$$

Тогда для любого матричного направления H имеем:

$$\langle \nabla^2 F_p(X)H, H \rangle_M \leq (2p - 1) \|\lambda(H)\|_{(2p)}^2 = (2p - 1) \|H\|_{(2p)}^2.$$

2. Энтропийное сглаживание максимального собственного значения. Рассмотрим функцию $E(X) = \ln \sum_{i=1}^n e^{\lambda^{(i)}(X)}$. Тогда при любом H

$$\langle \nabla^2 E(X)H, H \rangle_M \leq \left[\sum_{i=1}^n e^{\lambda^{(i)}(X)} \right]^{-1} \sum_{i=1}^n e^{\lambda^{(i)}(X)} (\lambda^{(i)}(|H|))^2 \leq \|H\|_{(\infty)}^2.$$

В заключение параграфа показывается как можно использовать сглаженные функции для минимизации максимального собственного значения и спектрального радиуса семейства симметрических матриц, линейно зависящих от параметров. Оценка трудоемкости соответствующих методов имеет порядок $O(1/\epsilon)$.

В **Главе 6** приводятся алгоритмы, позволяющие получать приближенные решения экстремальных задач с *относительной точностью*. Соответствующие результаты опубликованы в статьях [15, 17].

В параграфе 6.1 рассматривается коническая задача безусловной минимизации

$$\text{Найти } f^* = \min_x \{f(x) : x \in \mathcal{L}\}, \tag{6.1.1}$$

$$\mathcal{L} = \{x \in R^n : Cx = b\},$$

где выпуклая функция $f(x)$ является однородной степени единица, C - это $p \times n$ -матрица, и $b \neq 0$. Основное предположение о задаче (6.1.1) выглядит так:

$$\text{dom } f \equiv R^n, \quad 0 \in \text{int } \partial f(0). \quad (6.1.2)$$

Отсюда следует что $f^* > 0$ и задача отыскания решения задачи (6.1.1) с некоторой относительной точностью поставлена корректно.

Зафиксируем некоторую норму $\|\cdot\|$ в R^n . Пусть положительные параметры $\gamma_0 \leq \gamma_1$ удовлетворяют следующему условию:

$$B_{\|\cdot\|}(\gamma_0) \subseteq \partial f(0) \subseteq B_{\|\cdot\|}(\gamma_1). \quad (6.1.3)$$

Обозначим $\alpha = \frac{\gamma_0}{\gamma_1} < 1$. Этот параметр определяет сложность нахождения решения задачи (6.1.1) с некоторой относительной точностью.

Обозначим через x_0 проекцию начала координат на аффинное подпространство \mathcal{L} относительно нормы $\|\cdot\|$:

$$\|x_0\| = \min_x \{\|x\| : Cx = b\}.$$

Эта точка снабжает нас важной информацией о размере оптимального решения.

Теорема 6.1.1

1. Для любого $x \in R^n$ имеем

$$\gamma_0 \cdot \|x\| \leq f(x) \leq \gamma_1 \cdot \|x\|. \quad (6.1.6)$$

Таким образом, функция $f(x)$ липшицева на R^n относительно нормы $\|\cdot\|$ с константой γ_1 . Более того,

$$\alpha f(x_0) \leq \gamma_0 \cdot \|x_0\| \leq f^* \leq f(x_0) \leq \gamma_1 \cdot \|x_0\|. \quad (6.1.7)$$

2. Для любого оптимального решения x^* задачи (6.1.1) имеем:

$$\|x_0 - x^*\| \leq \frac{2}{\gamma_0} f^* \leq \frac{2}{\gamma_0} f(x_0). \quad (6.1.8)$$

Простейший метод решения задачи (6.1.1) выглядит так.

Субградиентный метод $G_N(R)$	
<p>Для $k := 0 \dots N$ повторяем Вычисляем $f(x_k)$ и $g(x_k)$. $x_{k+1} := \pi_{\mathcal{L}} \left(x_k - \frac{R}{\sqrt{N+1}} \cdot \frac{g(x_k)}{\ g(x_k)\ ^*} \right).$ Результат: $G_N(R) = \arg \min \{ f(x) : x = x_0, \dots, x_N \}.$</p>	(6.1.10)

Теорема 6.1.2 Для фиксированного δ из $(0, 1)$ выберем

$$\hat{\rho} \stackrel{\text{def}}{=} \frac{1}{\gamma_0} f(x_0), \quad N = \left\lfloor \frac{1}{\alpha^4 \delta^2} \right\rfloor. \quad (6.1.13)$$

Тогда $f(G_N(\hat{\rho})) \leq (1 + \delta) \cdot f^*$.

Этот метод можно ускорить за счет правильной стратегии обновления. Положим

$$\hat{N} = \left\lfloor \frac{e}{\alpha^2} \cdot \left(1 + \frac{1}{\delta}\right)^2 \right\rfloor,$$

где e - число Эйлера (2.718 ...). Рассмотрим следующий процесс. Полагаем $\hat{x}_0 = x_0$, и для $t \geq 1$ повторяем следующие операции:

$\hat{x}_t := G_{\hat{N}} \left(\frac{1}{\gamma_0} f(\hat{x}_{t-1}) \right);$ <p>если $f(\hat{x}_t) \geq \frac{1}{\sqrt{e}} f(\hat{x}_{t-1})$ то $T := t$ и СТОП.</p>	(6.1.14)
---	----------

Теорема 6.1.3 Число точек в процессе (6.1.14) ограничено:

$$T \leq 1 + 2 \ln \frac{1}{\alpha}. \quad (6.1.15)$$

Последняя из этих точек удовлетворяет неравенству $f(\hat{x}_T) \leq (1 + \delta) f^*$. Общее число градиентных шагов в процессах нижнего уровня (6.1.14) не превосходит

$$\frac{e}{\alpha^2} \cdot \left(1 + \frac{1}{\delta}\right)^2 \cdot \left(1 + 2 \ln \frac{1}{\alpha}\right). \quad (6.1.16)$$

В заключительной части этого параграфа мы показываем как можно еще снизить трудоемкость получения δ -решения в случае когда целевая функция является составной. А именно, мы предполагаем что $f(x) = F(A(x))$,

где $A(x)$ - линейный оператор, а $F(v)$ - *простая* нелинейная выпуклая однородная функции. Тогда, комбинируя операцию сглаживания с применением быстрого градиентного метода, можно добиться трудоемкости $O(1/(\alpha\delta))$ (см. теорему 6.1.5).

В параграфе 6.2 мы показываем как можно еще снизить трудоемкость нахождения δ -решения за счет предварительного вычисления подходящей евклидовой нормы. Мы начинаем с описания эффективных алгоритмов построения эллипсоидов Джона для различных типов выпуклых множеств. Эллипсоиды $W_r(v, G) \subset E^*$ представляются в следующей форме:

$$W_r(v, G) = \{s \in E^* : \|s - v\|_G^* \equiv \langle s - v, G^{-1}(s - v) \rangle^{1/2} \leq r\},$$

где $G \succ 0$ - линейный оператор из E в E^* . Если $v = 0$, то мы переходим к обозначению $W_r(G)$. Эллипсоид $W_1(v, G)$ называется β -*аппроксимацией* выпуклого множества $C \subset E^*$, ($\beta \geq 1$) если

$$W_1(v, G) \subseteq C \subseteq W_\beta(v, G).$$

Параметр β называется *радиусом* эллипсоидальной аппроксимации.

Сначала приводятся алгоритмы, работающие с центрально-симметричными телами и с выпуклыми телами общего вида, близкие к алгоритму Хачияна¹⁸. Доказывается что они получают приемлемую аппроксимацию за $O(n)$ шагов, где n - размерность пространства. Далее мы рассматриваем знако-инвариантные тела. Показывается, что они допускают $O(n^{1/2})$ -аппроксимацию, которая обеспечивается *диагональной* матрицей. Нахождение такой матрицы также требует $O(n)$ итераций.

Во второй части параграфа мы показываем, как найденные метрики могут использоваться для нахождения решений с относительной точностью у конкретных типов задач минимизации. В частности, мы показываем, как можно эффективно минимизировать максимум модулей линейных форм и решать билинейную матричную игру. Для последней задачи общее число шагов градиентного метода оценивается как

$$4e(1 + \ln(2\sqrt{n}))\sqrt{2n \ln m} \left(1 + \frac{1}{\delta}\right) \quad (6.2.34)$$

¹⁸Khachiyan L.G. Rounding of polytopes in the real number model of computation. // Mathematics of Operations Research - 1996 - V.21, N2. - P.307-320.

(см. теорему 6.2.5).

В **заключении** приводятся основные результаты и выводы диссертации. В частности, мы показали, что правильное использование модели экстремальной задачи существенно увеличивает возможности методов минимизации по генерации полезной информации. Мы можем формировать приближенные решения сопряженных задач, пользоваться точными критериями останковки, решать задачи с относительной точностью и т.д.

Более того, мы показали, что информация, полученная из "черного ящика", позволяет ускорять эти методы, превышая даже максимально возможные скорости, полученные в теории сложности. По-видимому, в ближайшем будущем мы увидим дальнейшее уточнение моделей целевых функций, которое позволит эффективно решать задачи сверхбольших размеров, возникающих в больших социальных сетях (телекоммуникации, интернет).

В настоящий момент теория выпуклой оптимизации впервые за долгое время оказалась не защищенной никакой теорией сложности. Для ее разработки и обновления, по-видимому, потребуются совсем другие принципы классификации задач оптимизации, разработка которых является одним из самых интересных направлений для будущих исследований.

Основные публикации автора по теме диссертации

1. Нестеров Ю.Е. Метод минимизации выпуклых функций со скоростью сходимости $O(1/k^2)$ // Докл. АН СССР - 1983 - Т.269, вып.3 - С. 543-547.
2. Нестеров Ю.Е. О классе методов безусловной минимизации с высокой скоростью сходимости // Журн. выч.мат. и мат.физ. - 1984 - Т.24, вып.7 - С. 1090-1093.
3. Нестеров Ю.Е. Метод линейного программирования с трудоемкостью $O(n^3L)$ операций // Экономика и мат. методы - 1988 - Т.24, вып.1 - С. 174-176.
4. Нестеров Ю.Е. Полиномиальные методы для задач линейного и квад-

- ратичного программирования // Известия АН СССР - 1988 - вып.3 - С. 119 - 128.
5. Нестеров Ю.Е. Об одном подходе к разработке оптимальных методов минимизации гладких выпуклых функций // Экономика и мат.методы - 1988 - Т.24, вып.3 - С. 509-517.
 6. Нестеров Ю.Е. Двойственные полиномиальные алгоритмы для линейного программирования // Кибернетика - 1989 - вып.1 - С. 34-54.
 7. Нестеров Ю.Е. Эффективные методы нелинейного программирования. - М.: Радио и Связь, 1989. - 280с.
 8. Yu. Nesterov. Introductory lectures on convex optimization. A basic course. - Boston: Kluwer, 2004. - 236p.
 9. Nesterov Yu. Smooth minimization of non-smooth functions // Mathematical Programming - 2005 - V.103, N1.-P.127-152.
 10. Nesterov Yu. Excessive gap technique in non-smooth convex minimization // SIAM J. Optim. - 2005 - V.16, N1. - P.235-249.
 11. Nesterov Yu., Polyak B. Cubic regularization of Newton's method and its global performance // Mathematical Programming - 2006 - V.108, N1. - P.177-205.
 12. Nesterov Yu. Smoothing technique and its applications in semidefinite optimization // Mathematical Programming - 2007 - V.110, N2. - P.245-259.
 13. Nesterov Yu. Dual extrapolation and its application for solving variational inequalities and related problems // Mathematical Programming - 2007 - V.109, N2-3. - P.319-344.
 14. Nesterov Yu. Modified Gauss-Newton scheme with worst-case guarantees for its global performance // Optimization Methods and Software - 2007 - V.22, N3. - P.469-483.

15. Nesterov Yu. Rounding of convex sets and efficient gradient methods for linear programming problems // Optimization Methods and Software - 2008 - V.23, N1. - P.109-128.
16. Nesterov Yu. Accelerating the cubic regularization of Newton's method on convex problems // Mathematical Programming - 2008 - V.112, N1. - P.159-181.
17. Nesterov Yu. Unconstrained Convex Minimization in Relative Scale // Mathematics of Operations Research - 2009 - V.34, N1. - P.180-193.
18. Nesterov Yu. Primal-dual subgradient methods for convex problems // Mathematical Programming - 2009 - V.120, N1. - P.261-283.
19. Нестеров Ю.Е. Введение в выпуклую оптимизацию. - М.: Изд. МЦНМО, 2010. - 279 с.
20. Nesterov Yu., Scramali L. Solving strongly monotone variational and quasi-variational inequalities // Discrete and Continuous Dynamical Systems - 2011 - V.31, N4. - P.1383-1296.
21. Nesterov Yu. Barrier subgradient method // Mathematical Programming - 2011 - V.127, N1. - P.31-56.
22. Nesterov Yu. Gradient methods for minimizing composite функцияс // Mathematical Programming - 2013 - V.140, N1. - P.125-161.

НЕСТЕРОВ ЮРИЙ ЕВГЕНЬЕВИЧ

Алгоритмическая выпуклая оптимизация

АВТОРЕФЕРАТ

Подписано в печать 19.09.2013. Формат 60 × 84 1/16. Усл. печ. л. 1.9

Тираж 150 экз. Заказ номер 906

Федеральное государственное автономное образовательное
учреждение высшего профессионального образования

"Московский физико-технический институт (государственный университет)"

Отдел оперативной полиграфии "Физтех-полиграф"

141700, Московская обл., г. Долгопрудный, Институтский пер., 9