Н. З. ШОР

МЕТОДЫ МИНИМИЗАЦИИ НЕГЛАДКИХ ФУНКЦИЙ И МАТРИЧНЫЕ ЗАДАЧИ ОПТИМИЗАЦИИ

КИШИНЭУ • ЭВРИКА • 2009

НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК УКРАИНЫ
ИНСТИТУТ КИБЕРНЕТИКИ ИМ. В. М. ГЛУШКОВА

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ И МОЛОДЕЖИ
РЕСПУБЛИКИ МОЛДОВА

АКАДЕМИЯ ТРАНСПОРТА,
ИНФОРМАТИКИ И КОММУНИКАЦИЙ

Н. З. ШОР

МЕТОДЫ МИНИМИЗАЦИИ НЕГЛАДКИХ ФУНКЦИЙ И МАТРИЧНЫЕ ЗАДАЧИ ОПТИМИЗАЦИИ

Сборник избранных трудов

КИШИНЭУ * ЭВРИКА * 2009

CZU 519.8
Ш 78

Descrierea CIP a Camerei Nationale a Cartii

Шор, Н. З.

Методы минимизации негладких функций и матричные задачи оптимизации: Сб. избранных трудов / Н. З. Шор; редкол.: И. В. Сергиенко, Д. И. Соломон, П. И. Стецюк [и др.]; Нац. акад. наук Украины, Ин-т кибернетики им. В. М. Глушкова, Акад. Транспорта, Информатики и Коммуникаций. – К.: Б. и., 2009 (Tipogr. ASM). – 240 p.

Texte: lb. engl., rusa. – Bibliogr. la sfarsitul art. – 150 ex.

ISBN 978-9975-62-267-7.

519.8
Ш 78

Сборник содержит работы академика Н.З.Шора, основателя направления недифференцируемой оптимизации в теории математического программирования. В сборник включены его работы по методам минимизации негладких функций и их применению в задачах большой размерности, квадратичного, булевого и матричного программирования.

Книга рассчитана на специалистов в области математического программирования и его приложений, а также студентов и аспирантов соответствующих специальностей.

ISBN 978-9975-62-267-7

АКАДЕМИК Н. З. ШОР (1937 – 2006)

"Теория всего класса алгоритмов с растяжением
пространства ещё далека от совершенства.
Нам кажется достаточно реалистической целью
построение такого алгоритма, который
по своей практической эффективности
не уступал бы r-алгоритму и был столь же
хорошо обоснован, как метод эллипсоидов".

Шор

Предисловие

В 2008 году был издан сборник [1] избранных научных трудов академика Наума Зуселевича Шора, основателя направления негладкой (недифференцируемой) оптимизации в математическом программировании. В сборник вошли двенадцать научных работ, которые оказали большое влияние на развитие теории и практики численных методов минимизации недифференцируемых функций и сыграли важную роль при развитии ряда направлений в математическом программировании. Первая часть сборника содержит восемь статей [2]–[9] по численным методам оптимизации недифференцируемых функций и их приложениям в задачах блочного программирования большой размерности. Вторая часть содержит четыре статьи [10]–[13], связанные с использованием методов недифференцируемой оптимизации при решении матричных, полиномиальных и дискретных задач оптимизации. В сборник также включен самый полный список научных работ Наума Зуселевича, который имелся на момент его подготовки.

Структурное разделение статей сборника на две части было обусловлено тем, что среди многочисленных приложений методов недифференцируемой оптимизации можно выделить два главных направления, которым Наум Зуселевич придавал очень большое значение. Первая часть сборника отражает период в научной деятельности Наума Зуселевича (с 1960 по 1985 годы), когда главным двигателем при разработке численных методов была необходимость решать негладкие задачи, которые возникают при декомпозиции блочных задач большой размерности. Вторая часть сборника отражает период (с 1985 по 2006 годы), который, главным образом, был связан с исследованием «задач матричной оптимизации»[1] и адаптацией субградиентных методов для их решения. Такое пристальное внимание к матричной оптимизации было продиктовано тем, что к ней сводятся многие важные классы многоэкстремальных задач. Начиная с 1985 года, Н. З. Шором уделял большое внимание совершенствованию методов минимизации недифференцируемых функций для решения различного рода задач с негладкими матричными функциями.

---

[1]Этот термин использовался Н. З. Шором и его учениками, начиная с середины 80-х годов. Под «задачами матричной оптимизации» подразумевались все те задачи оптимизации, где коэффициенты симметричных матриц заданы линейными функциями, а целевые функции либо ограничения включают функции от матрицы (детерминант, след, максимальное собственное число, сумма взвешенных максимальных собственных чисел, и т.д.)

Каждый из условно выделенных периодов характеризуется достижениями, которые признаны ведущими специалистами в области оптимизации. Первый период дал метод обобщенного градиентного (субградиентного) спуска (предложен Н. З. Шором в 1962 году), который положил начало численным методам недифференцируемой оптимизации. На его основе были разработаны схемы декомпозиции, позволяющие решать задачи производственно-транспортного планирования большой размерности. В 1969–1971 годы субградиентные методы были значительно усовершенствованы благодаря применению операции растяжения пространства, что позволило эффективно решать задачи минимизации негладких функций с овражными особенностями. Широкую известность получили частные случаи такого рода процессов: метод эллипсоидов, на основе которого были получены полиномиальные алгоритмы для ряда задач выпуклого и дискретного программирования; r-алгоритмы, которые до сих пор являются одним из наиболее эффективных средств решения задач недифференцируемой оптимизации. При минимизации гладких функций r-алгоритмы оказались конкурентоспособными с наиболее удачными реализациями методов сопряженных направлений и методов квазиньютоновского типа.

Второй период дал новые методы для решения задач полиномиальной и матричной оптимизации, новые результаты в области сложности комбинаторных задач, новые методы анализа многоэкстремальных и NP-трудных задач математического программирования. Одним из главных достижений второго периода является техника лагранжевых (двойственных) оценок в квадратичных невыпуклых задачах. С помощью квадратичных невыпуклых моделей описываются задачи полиномиальной оптимизации, экстремальные задачи на графах, задачи дискретного и булевого программирования и др. Эту технику можно считать альтернативой использованию методов внутренних точек для решения задач полуопределенного программирования (semidefinite programming)[2]. Ее использование позволило установить тесную связь между задачами полиномиальной оптимизации и 17-й проблемой Гильберта, каса-

[2]Многие задачи полуопределенного программирования целесообразно рассматривать как частный случай задач недифференцируемой оптимизации и применять для их решения эффективные методы минимизации негладких выпуклых функций. Условие неотрицательной определенности симметричной $(n{\times}n)$-матрицы X (принято обозначать $X\succeq 0$) эквивалентно тому, что минимальное собственное число этой матрицы $\lambda_n(X)\geq 0$. Но $\lambda_n(X)$ – вогнутая недифференцируемая функция элементов матрицы, т.е. если элементы матрицы $X(u) = \{x_{ij}\}_{i,j=1}^{n}$ являются линейными функциями от вектора варьируемых параметров $u \in R^m$, то условие $X(u)\succeq 0$ эквивалентно выпуклому негладкому ограничению $\varphi(u) = -\lambda_n(X(u)) \leq 0$.

ющейся представления неотрицательных рациональных форм в виде суммы квадратов рациональных форм. Очень интересными оказались результаты, связанные с NP-трудной комбинаторной задачей нахождения взвешенного максимального независимого множества вершин графа. Для нее лагранжевы оценки оказались тесно связаны с известными числами Ловаса $\vartheta(G, w)$ и $\vartheta'(G, w)$.

Обьем сборника [1] не позволил отразить ряд важных идей, которые проходили через работы как первого, так и второго периодов научной деятельности Н. З. Шора. Для первого периода недостаточно отражены работы по разработке субградиентных алгоритмов с преобразованием пространства и процедур, которые применялись при решении негладких задач большой размерности. Для второго периода в тени остались субградиентные алгоритмы для решения ряда матричных задач, связанных с построением описанных и вписанных экстремальных эллипсоидов. Не нашли отражения также и первые работы, с которых начиналась техника лагранжевых (двойственных) оценок в квадратичных невыпуклых задачах. Настоящий сборник избранных трудов академика Н. З. Шора восполняет этот пробел. Принцип его построения такой же, как и для первого сборника: два периода научной деятельности – две части книги.

В первую часть сборника включены девять работ. Сюда вошли две обзорные статьи [14] и [15]. Работа [14] содержит достаточно полный обзор результатов, полученных по субградиентным методам минимизации негладких функций и их приложениям до 1976 года[3]. Вторая обзорная работа [15] дополняет первую. Здесь дается сравнительная характеристика трех основных направлений в методах недифференцируемой оптимизации: обобщенных градиентных методов, схемы отсечения и ε-субградиентных методов. В частности, в работе дается интерпретация метода эллипсоидов Д. Б. Юдина и А. С. Немировского с позиций субградиентых методов с растяжением пространства.

Остальные семь статей в первой части сборника помещены в хронологическом порядке. Первые три связаны с обоснованием ряда вариантов обобщенного градиентного спуска с растяжением пространства в направлении субградиента. В работе [16] впервые использовалось растяжение пространства для ускорения сходимости субградиентных методов

[3] В 70-80 годы эта работа была настольной для многих, кто был связан с приложением методов недифференцируемой оптимизации в блочных задачах математического программирования. В статью включено детальное описание схемы декомпозиции по переменным (декомпозиция Бендерса). Своей актуальности эта работа не потеряла и сейчас.

при минимизации функций овражного типа. В работе [17] для минимизации выпуклых функций предложен и обоснован алгоритм обобщенного градиентного спуска с растяжением пространства и регулировкой шага, использующей априорное знание значения функции в точке минимума. В работе [18] этот алгоритм модифицируется для случая, когда значение функции в точке минимума неизвестно.

Работа [19] обобщает результаты о сходимости обобщенного градиентного спуска со скоростью геометрической прогрессии. В статье [20] описываются две модификации метода обобщенного градиентного спуска с растяжением пространства в направлении разности двух последовательных градиентов применительно к решению минимаксных задач. В статье [21] определяется класс почти-дифференцируемых функций. Для нахождения локальных минимумов таких функций предлагается метод с растяжением пространства, который может быть использован при решении систем нелинейных уравнений. В работе [22] разработана процедура построения «существенной» части линейных операторов растяжения пространства с помощью собственных значений и собственных векторов. Эта работа продолжает исследования, начатые в [20] для решения негладких задач большой размерности.

Во вторую часть сборника в хронологическом порядке включены одиннадцать работ. Первые четыре – это пионерские работы по технике лагранжевых (двойственных) оценок для невыпуклых квадратичных задач. В статье [23] впервые предложен метод получения оценок в квадратичных экстремальных задачах с булевыми переменными, который впоследствии использовался при исследовании ряда NP-трудных комбинаторных задач и экстремальных задач на графах. Работа [24] посвящена численным методам решения оценочных задач в экстремальных задачах с квадратичными целевым функционалом и ограничениями. Показано, что в многоэкстремальном случае оценки снизу минимального значения целевой функции можно получить методами двойственности, решая выпуклую оптимизационную задачу с ограничениями типа неотрицательной определенности матриц. В работе [25] методика оптимальных лагранжевых оценок была применена к исследованию оценок снизу целевой функции в полиномиальных задачах математического программирования путем их сведения к квадратичным экстремальным задачам. Она была дополнена идеей использовать функционально избыточные ограничения для улучшения точности этих оценок. Для случая минимизации полиномов от одной переменной это позволило доказать, что соответствующие оценки являются точными. Эта идея нашла более полное изложение в работе [26] для задачи минимизации полиномов от многих переменных.

Следующие пять работ связаны с различными вариантами алгоритмов для построения оптимальных по объему вписанных в многогранник и описанных вокруг многогранника эллипсоидов. В работе [27] задача нахождения эллипсоида минимального объема, включающего заданный набор точек, сведена к специальной задаче матричной оптимизации, которая заключается в максимизации вогнутой негладкой функции от элементов симметричной матрицы при условии положительной определенности этой матрицы. Для её решения разработана модификация r-алгоритма. Идея этого метода легла в основу разработки приближенного алгоритма для построения эллипсоида минимального объема, включающего пересечение заданных эллипсоидов, который изложен в работе [28]. В работе [29] алгоритм субградиентного типа с растяжением пространства использован для построения эллипсоида максимального объема, вписанного в многогранник. Метод базируется на анализе соответствующей задачи матричной оптимизации. В работе [30] для построения описанного эллипсоида минимального объема разработан алгоритм, в основу которого положены последовательные сжатия пространства. В работе [31] подобный алгоритм, но уже с последовательным растяжением пространства, предложен и применен для построения эллипсоида максимального объема, вписанного в многраннник.

В статье [32] подытожены исследования по задаче построения эллипсоида минимального обьема, описанного вокруг многогранника, заданного конечным набором точек, и задаче построения эллипсоида максимального обьема, вписанного в многогранник, заданный конечной системой линейных неравенств. Здесь наиболее полно отражена картина «сложности» всех пяти построенных алгоритмов для решения этих задач (три алгоритма для первой задачи и два алгоритма для второй задачи), и даются рекомендации по использованию алгоритмов в зависимости от количества точек и линейных неравенств. Завершает вторую часть сборника работа [33], связанная с алгоритмами построения инвариантного эллипсоида минимального обьема для устойчивой динамической системы с заданным конечным множеством начальных положений. Алгоритмы разработаны на базе субградиентных методов с применением последовательных преобразований фазового пространства.

Всего в оба сборника включены тридцать две работы, характеризующие исследования академика Н. З. Шора и его школы. Из них семнадцать работ относятся к первому периоду (восемь включены в первый сборник и девять – во второй) и пятнадцать работ – ко второму периоду (четыре включены в первый сборник и одиннадцать во второй).

Они достаточно полно характеризуют исследования по численным методам минимизации недифференцируемых функций и их приложениям в сложных экстремальных задачах, которые проводились в Институте кибернетики имени В. М. Глушкова НАН Украины за последние 50 лет. Хочется надеяться, что эти два сборника избранных трудов Наума Зуселевича станут помощниками в работе для специалистов в области математического программирования и его приложений, а также студентов и аспирантов соответствующих специальностей.

Объем двух книг не позволил поместить ряд результатов, которые были получены под руководством Н. З. Шора. В частности, не нашли достаточно полного отражения результаты по развитию методов эллипсоидов для задач выпуклого программирования, стохастическим многоэтапным задачам, специальным задачам дробно-линейного программирования, использованию субградиентных методов в пакетах прикладных программ и др. Чтобы восполнить этот пробел, планируется издать серию тематических сборников научных трудов. В настоящее время готовятся три таких сборника: Соломон Д. И. «Дробно-линейное программирование и недифференцируемая оптимизация»; Журбенко Н. Г. «Субградиентные методы минимизации негладких функций»; Стецюк П. И. «Методы эллипсоидов и их обобщения». Их издание планируется в 2009–2010 гг.

Редакционная коллегия
Киев, Кишинэу, май 2009 г.

Литература

1. Шор Н. З. Методы недифференцируемой оптимизации и сложные экстремальные задачи: сборник избранных трудов. – Кишинеу, Эврика, 2008. – 270 с.

2. Шор Н. З.,Журбенко Н. Г., Лиховид А. П., Стецюк П. И. Развитие алгоритмов недифференцируемой оптимизации и их приложения // Кибернетика и системный анализ. – 2003. – № 4. – С. 80–94.

3. Шор Н. З. Применение обобщенного градиентного спуска в блочном программировании // Кибернетика. – 1967. – № 3. – С. 53–55.

4. Ермольев Ю. М., Шор Н. З. Метод случайного поиска для задач двухэтапного стохастического программирования и его обобщение // Кибернетика. – 1968. – № 1. – С. 90–92.

5. Шор Н. З., Щепакин М. Б. Алгоритм решения двухэтапной задачи стохастического программирования // Кибернетика. – 1968. – № 3. – С. 56–58.

6. Шор Н. З., Журбенко Н. Г. Метод минимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов // Кибернетика. – 1971. – № 3. – С. 51–59.

7. Шор Н. З. Метод отсечения с растяжением пространства для решения задач выпуклого программирования // Кибернетика. – 1977. – № 1. – С. 94–95.

8. Беляева Л. В., Журбенко Н. Г., Шор Н. З. О методе решения одного класса динамических распределительных задач // Экономика и мат. методы. – 1978. – Т. 14. – Вып. 1. – С. 137–146.

9. Шор Н. З., Бардадым Т. А., Журбенко Н. Г., Лиховид А. П., Стецюк П. И. Использование методов негладкой оптимизации в задачах стохастического программирования // Кибернетика и системный анализ. – 1999. – № 5. – С. 33–47.

10. Шор Н. З. Минимизация матричных функций и недифференцируемая оптимизация // Обозрение прикладной и промышленной математики. – М.: научн. изд.-во "ТВП", 1995. – Т. 2. – С. 113–138.

11. Шор Н. З., Стецюк П. И. Использование модификации r-алгоритма для нахождения глобального минимума полиномиальных функций // Кибернетика и системный анализ. – 1997. – № 4. – С. 28–49.

12. Шор Н. З. Роль избыточных ограничений в улучшении двойственных оценок для полиномиальных оптимизационных задач // Кибернетика и системный анализ. – 1998. – № 4. – С. 106–121.

13. Shor N. Z. and Stetsyuk P. I. Lagrangian Bounds in Multiextremal Polynomial and Discrete Optimization Problems // Journal of Global Optimization. – 2002. – **23**. – P. 1–41.

14. Шор Н. З. Обобщенные градиентные методы минимизации негладких функций и их применение к задачам математического программирования // Экономика и мат. методы. – 1976. – № 12. – Вып.2. – С. 337–356.

15. Шор Н. З. Новые направления в развитии методов негладкой оптимизации // Кибернетика. – 1977. – № 6. – С. 87–91.

16. Шор Н. З., Билецкий В. И. Метод растяжения пространства для ускорения сходимости в задачах овражного типа // Тр. семинара Науч. совета АН УССР по кибернетике "Теория оптимальных решений", Киев, 1969. – № 2. –С. 3–18,

17. Шор Н. З. Использование операций растяжения пространства в задачах минимизации выпуклых функций // Кибернетика. – 1970. – № 1. – С. 6–12.

18. Шор Н. З. О скорости сходимости метода обобщенного градиентного спуска с растяжением пространства // Кибернетика. – 1970. – № 2. – С. 80–85.

19. Шор Н. З., Гамбурд П. Р. Некоторые вопросы сходимости обобщенного градиентного спуска // Кибернетика. – 1971. – № 6. – С. 82–84.

20. Шор Н. З., Шабашова Л. П. О решении минимаксных задач методом обобщенного градиентного спуска с растяжением пространства // Кибернетика. – 1972. – № 1. – С. 82–88.

21. Шор Н. З. О классе почти-дифференцируемых функций и одном методе минимизации функций этого класса. // Кибернетика. – 1972. – № 4. – С. 65–70.

22. Шор Н. З., Гершович В. И. Об одной модификации алгоритмов минимизации градиентного типа с растяжением пространства для решения задач большой размерности // Кибернетика. – 1981. – № 5. – С. 67–70.

23. Шор Н. З., Давыдов А. С. О методе получения оценок в квадратичных экстремальных задачах с булевыми переменными // Кибернетика. – 1985. – № 2. – С. 48–50.

24. Шор Н. З. Квадратичные оптимизационные задачи. – Известия АН СССР. "Техническая кибернетика". – 1987. – № 1. –С. 128–139.

25. Шор Н. З. Об одном подходе к получению глобальных экстремумов в полиномиальных задачах математического программирования // Кибернетика. – 1987. – № 5. – С. 102–106.

26. Шор Н. З. Об одном классе оценок глобального минимума полиномиальных функций // Кибернетика. – 1987. – № 6. – С. 9–11.

27. Шор Н. З., Стеценко С. И. Использование точных штрафов при построении описанных эллипсоидов минимального объема // Кибернетика. – 1989. – № 2. – С. 117–119.

28. Шор Н. З., Давыдов А. С., Стеценко С. И. Алгоритмы построения оптимальных описанных эллипсоидов на основе методов негладкой оптимизации // Сб. научных трудов "Методы решения экстремальных задач и смежные вопросы", ИК АН УССР, Киев, 1989.

29. Шор Н. З., Березовский О. А. Использование алгоритма субградиентного типа с растяжением пространства для построения эллипсоида максимального объема, вписанного в многогранник // Кибернетика. – 1989. – № 6. – С. 119–120.

30. Шор Н. З., Стеценко С. И. Алгоритм последовательного сжатия пространства для построения описанного эллипсоида минимального объема // Исследование методов решения экстремальных задач, ИК АН УССР, Киев, 1990. – С. 25–29.

31. Шор Н. З., Березовский О. А. Построение эллипсоида максимального объема, вписанного в многранник, с использованием последовательного растяжения пространства // Сб."Кибернетика и вычислительная техника", Киев, 1992. – Вып.93. – С. 1–6.

32. Shor N. Z. and Berezovski O. A. *New algorithms for constructing optimal circumscribed and inscribed ellipsoids.* Optimization Methods and Software, Vol. 1, P. 283–299, 1992.

33. Шор Н. З., Березовский О. А. Алгоритмы построения инвариантного эллипсоида минимального обьема для устойчивой динамической системы // Кибернетика и системный анализ. – 1995. – № 3. – С. 130–137.

Ч А С Т Ь I

МЕТОДЫ МИНИМИЗАЦИИ НЕГЛАДКИХ ФУНКЦИЙ

Обобщенные градиентные методы минимизации негладких функций и их применение к задачам математического программирования

Н. З. Шор

Экономика и математические методы.
– 1976. – Том XII. – № 2. – С. 337–356.

1. Введение

В последние годы резко возрос интерес к проблемам безусловной оптимизации. Широко распространилось понимание того, что сведение задачи математического программирования к задаче минимизации функции без ограничений является во многих случаях достаточно эффективным путем к ее решению, что проблемы разработки эффективных методов безусловной оптимизации и их обоснования далеко не тривиальны и что в этой, казалось бы, распаханной вдоль и поперек области исследований имеется возможность значительного прогресса. С точки зрения приложений особенно важно иметь эффективные алгоритмы для минимизации «плохих» функций: овражных, негладких. Если алгоритм рассчитан на решение широкого круга задач, то у исследователя появляется возможность выбора математической модели оптимизируемого процесса, наиболее адекватно и экономно отображающей реальный процесс, реализации более гибких средств эквивалентного преобразования первоначальной задачи в некоторую стандартную и достаточно простую форму.

Проблемы минимизации функций, не всюду дифференцируемых, естественным образом появляются при решении самых разнообразных задач математического программирования, когда используются схемы декомпозиции, минимизируются функции максимума, применяются методы штрафных функций с негладкими «штрафами», а также при решении задач оптимального планирования и проектирования, в которых технико-экономические характеристики задаются в виде кусочно-гладких функций, и т. д.

В нашей стране уже накоплен серьезный опыт использования обобщенных градиентных методов минимизации негладких функций для решения задач математического программирования большой размерности. Исходя из него, многие сложные задачи оптимального планирования и проектирования можно успешно решать следующим образом:

1) сводя к задачам безусловной минимизации или минимизации функций при простейших ограничениях (для этого применяются схемы декомпозиции, метод штрафных функций и т. д.);

2) минимизируя полученные, как правило, негладкие функции, с помощью той или иной модификации метода обобщенного градиентного спуска (ОГС).

Метод ОГС был разработан в 1961 г. [1]. Сначала он применялся для минимизации кусочно-гладких выпуклых функций, появляющихся при решении транспортных и транспортно-производственных задач [1]–[3], затем для класса произвольных выпуклых функций [4]–[5] и задач выпуклого программирования в гильбертовом пространстве [6]. Широкое развитие получили стохастические аналоги ОГС [7]. Алгоритмы решения задач математического программирования, построенные на основе ОГС, отличаются простотой и, что особенно важно при большой размерности, экономным использованием оперативной памяти ЭВМ. К недостаткам метода ОГС относится его довольно медленная сходимость и ограниченность применения – только к классу выпуклых функций.

В связи с этим в последние годы в Институте кибернетики АН УССР разработаны ускоренные варианты обобщенных градиентных методов и их обоснование для достаточно широкого класса негладких функций. Основное внимание было уделено исследованию методов градиентного типа с растяжением пространства, их экспериментальной и практической проверке. В данном обзоре кратко излагаются основные результаты работ по градиентным методам минимизации негладких функций, описывается разнообразное практическое применение этих методов, особенно при решении задач оптимального планирования и проектирования большой размерности.

2. Некоторые классы негладких функций. Обобщения понятия градиента

Начнем с класса почти-дифференцируемых функций.

Определение. *Функция $f(x)$, заданная на n-мерном евклидовом пространстве E_n, называется почти-дифференцируемой, если: а) в любой ограниченной области она удовлетворяет условию Липшица; б) почти везде дифференцируема; в) ее градиент непрерывен на том множестве M, где он определен.*

Этот класс достаточно широк, чтобы охватить выпуклые и вогнутые функции, кусочно-гладкие и функции, встречающиеся в минимаксных задачах. В то же время для функций этого класса легко ввести обобщенное понятие градиента, благодаря чему удается построить аналоги градиентных процессов для нахождения локального минимума.

Определение. *Почти-градиентом функции $f(x)$ в точке $\overline{x}$ называется вектор, являющийся предельной точкой некоторой последовательности градиентов $g(x_1)$, $g(x_2)$, $\ldots, g(x_k)$, $\ldots$, где $\{x_k\}_{k=1}^{\infty}$ – последовательность точек, сходящихся к $\overline{x}$, и такая, что во всех ее точках $f(x)$ дифференцируема.*

В [8] показано, что множество $G_f(\overline{x})$ почти-градиентов почти-дифференцируемой функции $f(x)$ в точке $\overline{x}$ является непустым, ограниченным и замкнутым. В точке, где $f(x)$ дифференцируема, почти-градиент совпадает с градиентом, для выпуклой функции он является субградиентом, для кусочно-гладкой – совпадает с градиентом к одному из примыкающих к данной точке кусков, для функции максимума от конечного числа непрерывно дифференцируемых функций – с градиентом к одной из функций, на которой достигается максимум. Таким образом, в большинстве случаев при решении практических задач вычисление почти-градиента не вызывает особых трудностей.

Выпуклые на E_n функции составляют подкласс почти-дифференцируемых функций. Напомним понятие обобщенного градиента.

Вектор $g(x_0)$ называется обобщенным градиентом (субградиентом) выпуклой функции $f(x)$ в точке x_0, если он удовлетворяет следующему соотношению для любых $x \in E_n : f(x) - f(x_0) \geq (g(x_0), x - x_0)$.

Пусть $\widehat{G}_f(\overline{x})$ – множество обобщенных градиентов выпуклой функции $f(x)$ в точке $\overline{x}$. Это множество является выпуклым, замкнутым и ограниченным. Если $\widehat{G}_f(\overline{x})$ состоит из единственной точки, то в $\overline{x}$ функция $f(x)$ дифференцируема и обобщенный градиент совпадает с градиентом.

Пусть $\{f_i(x)\}_{i=1}^{m}$ – семейство выпуклых функций, $f(x) = \sum\limits_{i=1}^{m} c_i f_i(x)$, $c_i \geq 0$. Тогда $f(x)$ – выпуклая функция и $\widehat{G}_f(x)$ состоит из всевозможных векторов вида

$$g_f(x) = \sum_{i=1}^{m} c_i\, g_{f_i}(x), \quad g_{f_i}(x) \in \widehat{G}_{f_i}(x), \quad i = 1, 2, \ldots, m.$$

Если $f(x) = \max\limits_{1 \leq i \leq m} f_i(x)$, то $f(x)$ – выпуклая функция и

$\widehat{G}_f(x) = \operatorname{conv} \bigvee\limits_{i \in I(x)} \widehat{G}_{f_i}(x)$, где $I(x)$ – множество индексов i таких, что $f(x) = f_i(x)$, conv – символ выпуклого замыкания.

Таким образом, в большинстве практически важных случаев сложность вычисления обобщенного градиента не отличается от сложности вычисления градиента. Как показано в [9], выпуклая функция почти везде дифференцируема (т. е. за исключением точек множества меры 0, обобщенный градиент совпадает с градиентом). При этом градиент выпуклой функции непрерывен на множестве точек, где он определен.

Пусть $\eta \in E_n, \ \eta \neq 0$. Производная по направлению от выпуклой функции $f(x)$ существует и вычисляется по формуле

$$f'_\eta(x) = \max_{g \in \widehat{G}_f(\overline{x})} (g, \eta),$$

Рассмотрим луч: $y = x + t\,\eta, \ t \geq 0$. Если $0 \leq t_1 \leq t_2$, то $f'_\eta\,(x + t_1\,\eta) \leq$ $\leq f'_\eta\,(x + t_2\,\eta)$. Это свойство монотонности производной по направлению позволяет легко построить алгоритм, вычисляющий производную по направлению $f'_\eta\,(x_0)$ с заданной точностью, если предположить, что имеется возможность вычисления $f(x)$ со сколь угодно высокой точностью и в точке x_0 $f(x)$ дифференцируема.

В самом деле, справедливо неравенство: $f\,(x + h\,\eta) - f(x) \geq h\,f'_\eta(x) \geq$ $\geq f(x) - f\,(x - h\,\eta), \ h > 0$, которое дает двухсторонние оценки на производную, и легко подобрать h такое, чтобы погрешность не превышала заданную. Так как вычисление градиента сводится к определению частных производных, то отсюда вытекает алгоритм расчета градиента с заданной точностью в той точке, где он существует.

Легко привести примеры, показывающие, что если обобщенный градиент в данной точке определен неоднозначно, то не существует алгоритма, который бы гарантировал для любого $\varepsilon > 0$ построение вектора g_ε такого, что

$$\min_{g \in \widehat{G}_f(x)} \|g_\varepsilon - g\| < \varepsilon.$$

Однако показано, что можно построить алгоритм, вычисляющий вектор g^*, для которого при заданных $\varepsilon, \ \delta$ имеется такая точка $\overline{x}$, лежащая в δ-окрестности точки x_0, что

$$\min_{g \in \widehat{G}_f(\overline{x})} \|g^* - g\| < \varepsilon.$$

Такой алгоритм дает возможность получить вектор, который близок к обобщенному градиенту $f(x)$, взятому в точке, лежащей в сколь угодно малой окрестности заданной точки. Хотя этот алгоритм носит универсальный характер, на практике он еще не применялся, и мы его здесь описывать не будем, так как в большинстве случаев удается обойтись более простыми средствами вычисления обобщенного градиента.

Замечание. Можно показать, что крайние точки множества обобщенных градиентов $\widehat{G}_f(x)$ являются почти-градиентами, так что $\widehat{G}_f(x)$ является выпуклым замыканием множества почти-градиентов $f(x)$ в x. В некоторых случаях удобно использовать и для невыпуклой функции понятие обобщенного почти-градиента в точке x_0, как произвольного элемента выпуклого замыкания множества почти-градиентов в x_0.

3. Обобщенный градиентный спуск

Пусть задана выпуклая функция $f(x)$. Под методом ОГС будем понимать процедуру построения последовательности $\{x_k\}_{k=1}^{\infty}$ вида

$$x_{k+1} = x_k - h_{k+1}\left(x_k\right) g_f\left(x_k\right), \quad g_f\left(x_k\right) \in \widehat{G}_f\left(x_k\right), \quad k = 0, 1, \ldots, \tag{1}$$

где x_0 – нулевое приближение – задается до начала процесса спуска.

Выражение (1) внешне ничем не отличается от формулы, задающей спуск в обычных градиентных методах, тем более, что обобщенный градиент $g_f(x)$ почти всюду совпадает с градиентом. Однако существенное различие состоит в выборе метода регулировки шага. Дело в том, что обычно применяемые для дифференцируемых функций приемы, например $h_{k+1}\left(x_k\right) = h$ в методе с постоянным шаговым множителем или выбор h_{k+1} из условия $\min\limits_{h_{k+1}} f(x_k - h_{k+1} g_f\left(x_k\right))$, как в методе наискорейшего спуска, в случае произвольной выпуклой функции, вообще говоря, неприменимы. Проблема регулировки шага при использовании обобщенного градиентного спуска должна решаться по-другому.

Рассмотрим процесс обобщенного градиентного спуска с постоянным шагом, когда

$$h_{k+1}\left(x_k\right) = \frac{h}{\left\|g_f\left(x_k\right)\right\|}, \quad h > 0.$$

Пусть Ω – множество минимумов функции $f(x)$, $\Omega \neq \emptyset$.

В дальнейшем для простоты изложения будем предполагать, что для $k = 0, 1, 2, \ldots,$ $g_f\left(x_k\right) \neq 0$ (если $g_f\left(x_k\right) = 0$, то $x_k \in \Omega$).

Лемма 1. *Пусть при использовании метода обобщенного градиентного спуска* $h_{k+1}\left(x_k\right) = \dfrac{h}{\left\|g_f\left(x_k\right)\right\|}$, $h > 0$. *Тогда для любого* $\varepsilon > 0$

и $x^* \in \Omega$ найдутся такие $k = k^*$ и $\overline{x} \in E_n$, что будет выполняться свойство

$$f\left(\overline{x}\right) = f\left(x_{k^*}\right), \text{ причем } \|\overline{x} - x^*\| < \frac{h}{2}\,(1+\varepsilon).$$

Доказательство. В соответствии с условиями леммы

$$x_{k+1} = x_k - \frac{h\,g_f\left(x_k\right)}{\|g_f\left(x_k\right)\|}, \quad k = 0, 1, 2, \ldots.$$

Для любого $x_k \in E_n$

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &= \left\| x_k - x^* - \frac{h\,g_f\left(x_k\right)}{\|g_f\left(x_k\right)\|} \right\|^2 = \\ &= \|x_k - x^*\|^2 + h^2 - 2h\left(x_k - x^*, \frac{g_f\left(x_k\right)}{\|g_f\left(x_k\right)\|}\right) \leq \\ &\leq \|x_k - x^*\|^2 + h^2 - 2h\,b\left(x^*, x_k\right), \end{aligned}$$

где $b\left(x^*, x_k\right) = \min\limits_{y \in \{x\,:\, f(x) = f(x_k)\}} \|y - x^*\|$.

Если для всех $k = 0, 1, 2, \ldots,\quad b\left(x^*, x_k\right) \geq \dfrac{h}{2}\,(1+\varepsilon)$, то

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \varepsilon h^2 \leq \|x_0 - x^*\|^2 - \varepsilon(k+1)\,h^2$$

при $k = 0, 1, 2, \ldots.$ Но $\|x_{k+1} - x^*\|^2 \geq 0$, что противоречит предыдущему выражению. Значит, найдется такое $k^* = k^*_\varepsilon\left(x^*\right)$, что $b\left(x^*, x^*_k\right) < \dfrac{h}{2}\,(1+\varepsilon)$.

Лемма доказана. Она позволяет непосредственно получить следствия, формулируемые здесь в виде теорем.

Теорема 1. *Для произвольного $\delta > 0$ можно найти такое h_δ, что при применении ОГС с шагом $h_{k+1}\left(x_k\right) = h_\delta / \|g_f\left(x_k\right)\|$ при любом $x_0 \in E_n$ либо найдется k^*, что $x_{k^*} \in \Omega$, либо такая подпоследовательность $k_1 < k_2 < k_3 < \ldots$, что $\lim\limits_{i \to \infty} f\left(x_{k_i}\right) - \min\limits_{x \in E_n} f(x) < \delta$.*

Теорема 2. *Если функция $f(x)$ имеет область минимумов Ω, содержащую сферу радиуса $r > h/2$, то при использовании метода ОГС с $h_{k+1}\left(x_k\right) = h / \|g_f\left(x_k\right)\|$ найдется такое k^*, что $x_{k^*} \in \Omega$.*

Лемма 1 гарантирует уменьшение расстояния до точек области минимума на данном шаге только в случае достаточно большого расстояния от нее. С уменьшением шага h сокращается и это расстояние. Поэтому, чтобы получить обычные теоремы сходимости, нужно, чтобы $h_{k+1}(x_k) \to 0$, но это стремление $h_{k+1}(x_k)$ к 0 не должно осуществляться слишком быстро, чтобы гарантировать сходимость $\{x_k\}_{k=0}^{\infty}$ к области Ω.

Так мы приходим к ставшим уже классическими условиям

$$h_{k+1}(x_k) = \frac{h_k}{\|g_f(x_k)\|}, \quad \lim_{k\to\infty} h_k = 0, \quad \sum_{k=1}^{\infty} h_k = +\infty.$$

Имеется несколько вариантов доказательства теоремы о сходимости метода ОГС [5, 6, 10]. Все они основаны на изучении поведения последовательности $\{\rho_k\}_{k=0}^{\infty}$, где $\rho_k = \min\limits_{k\in\Omega} \|x_k - x\|$. Наиболее общий результат (для случая выпуклых функций, определенных в гильбертовом пространстве, когда минимизация производится при наличии ограничений) получен Б. Т. Поляком [6]. В [5] приводится доказательство теоремы о сходимости для конечномерного случая. Остановимся на некоторых из этих результатов.

Теорема 3. [*10*]. *Пусть $f(x)$ – выпуклая функция, определенная в E_n, область минимумов которой Ω ограничена, и задана последовательность положительных чисел $\{h_k\}$, $k = 1, 2, \ldots$, обладающая следующими свойствами: $\lim\limits_{k\to\infty} h_k = 0$, $\sum\limits_{k=1}^{\infty} h_k = \infty$. Тогда последовательность $\{x_k\}$, $k = 0, 1, \ldots$, образованная по формуле*

$$x_{k+1} = \begin{cases} x_k - h_{k+1}\dfrac{g_f(x_k)}{\|g_f(x_k)\|}, & g_f(x_k) \neq 0, \\ x_k, & g_f(x_k) = 0, \end{cases}$$

при произвольном $x_0 \in E_n$ характеризуется: $\lim\limits_{k\to\infty} \min\limits_{x\in\Omega} \|x_k - x\| = 0$ и $\lim\limits_{k\to\infty} f(x_k) = \min\limits_{x\in E_n} f(x) = f^*$.

Теорема 4. *Если при условиях теоремы 3 мы образуем последовательность $\{x_k\}$ по формуле $x_{k+1} = x_k - h_{k+1}\, g_f(x_k)$, $k = 0, 1, \ldots$, где $x_0 \in E_n$ – произвольная точка, $\lim\limits_{k\to\infty} h_k = 0$, $\sum\limits_{k=1}^{\infty} h_k = \infty$ и последовательность $\{g_f(x_k)\}$ ограничена, то*

$$\lim_{k\to\infty} \min_{x\in\Omega} \|x_k - x\| = 0 \quad \textit{и} \quad \lim_{k\to\infty} f(x_k) = f^*.$$

В [6] доказана аналогичная теорема, соответствующая методу проекции обобщенного градиента, для случая минимизации выпуклых функционалов, определенных в гильбертовом пространстве при ограничениях довольно общего вида. Приведем формулировку этой теоремы с небольшими изменениями.

Пусть $f(x)$ – выпуклый функционал, определенный на выпуклом множестве Q гильбертова пространства E, $Q = Q_1 \bigcap Q_2$, где $Q_1 = \{x : \ g(x) \leq 0\}$, $g(x)$ – выпуклый на E функционал; Q_1 имеет внутренние точки (множество внутренних точек – Q_1^0); Q_2 – выпуклое замкнутое множество, причем $Q_1^0 \bigcap Q_2 \neq \emptyset$. Обозначим через $P_{Q_2}(x)$ проекцию точки x на Q_2, т. е.

$$\|x - P_{Q_2}(x)\| = \inf_{y \in Q_2} \|x - y\|.$$

Метод обобщенного градиентного спуска заключается в построении $\{x_k\}_{k=0}^{\infty}$

$$x_{k+1} = P_{Q_2}\left(x_k + \Delta x_k\right), \tag{2}$$

где Δx_k – произвольный опорный функционал к множеству $\{x : f(x) \leq$ $\leq f\left(x_k\right)\}$, если $x_k \in Q_1$, и к множеству $\{x : g(x) \leq g\left(x_k\right)\}$, если $x_k \overline{\in} Q_1$. (Линейный функционал c называется опорным к $R \subset E$ в точке $x \in R$, если $(c, x) \leq (c, y)$ для всех $y \in R$.)

Теорема 5. *Если $f(x)$ можно непрерывно продолжить на E,*

$$\lim_{k \to \infty} \|\Delta x_k\| = 0, \quad \sum_{k=0}^{\infty} \|\Delta x_k\| = \infty,$$

то при любом x_0 в методе (2) найдется подпоследовательность x_{k_n} такая, что $x_{k_n} \in Q$, $\lim\limits_{n \to \infty} f\left(x_{k_n}\right) = f^* = \inf\limits_{x \in Q} f(x)$. Если, кроме этого, $S = \big\{x : f(x) = f^*; \quad x \in Q\big\}$ непусто и из условия $f(x) \to f^*$, $x \in Q$ следует, что

$$\rho(x, S) = \inf_{y \in S} \|x - y\| \to 0, \text{ то } \lim_{k \to \infty} \rho\left(x_k, S\right) = 0.$$

Для ускорения сходимости при некоторых дополнительных предположениях можно применить другую регулировку шагового множителя [11]–[13].

Теорема 6. *Пусть $f(x)$– выпуклая функция, определенная на E_n, и для всех $x \in E_n$ выполняется неравенство при некотором φ, $0 \leq \varphi < \pi/2$,*

$$\Big(g_f(x), x - x^*(x)\Big) \geq \cos\varphi \, \|g_f(x)\| \cdot \|x - x^*(x)\|, \tag{3}$$

где $x^*(x)$ – *точка, принадлежащая множеству минимумов функции* $f(x)$ *и лежащая на кратчайшем расстоянии от* x. *Тогда при заданном* x_0, *если мы выберем величину* h_1, *удовлетворяющую неравенствам*

$$h_1 \geq \|x^*(x_0) - x_0\| \cos\varphi, \qquad \frac{\pi}{4} \leq \varphi < \frac{\pi}{2}, \tag{4}$$

$$h_1 \geq \frac{\|x^*(x_0) - x_0\|}{2\cos\varphi}, \qquad 0 \leq \varphi < \frac{\pi}{4}, \tag{5}$$

определим $\{h_k\}_{k=1}^{\infty}$ *в соответствии с рекуррентной формулой*

$$h_{k+1} = h_k r(\varphi), \quad k = 1, 2, \ldots, \tag{6}$$

где

$$r(\varphi) = \sin\varphi, \qquad \frac{\pi}{4} \leq \varphi < \frac{\pi}{2}, \tag{7}$$

$$r(\varphi) = \frac{1}{2\cos\varphi}, \qquad 0 \leq \varphi < \frac{\pi}{4}, \tag{8}$$

и будем вычислять $\{x_k\}_{k=1}^{\infty}$

$$x_{k+1} = x_k - h_{k+1} \frac{g_f(x_k)}{\|g_f(x_k)\|},$$

то либо при некотором $\overline{k}$ $g_f\left(x_{\overline{k}}\right) = 0$ *и* $x_{\overline{k}}$ *принадлежит области минимумов, либо при всех* $k = 0, 1, 2, \ldots$ *будут выполняться неравенства*

$$\|x_k - x^*(x_k)\| \leq \frac{h_{k+1}}{\cos\varphi}, \qquad \frac{\pi}{4} \leq \varphi < \frac{\pi}{2}, \tag{9}$$

$$\|x_k - x^*(x_k)\| \leq 2\cos\varphi\, h_{k+1}, \qquad 0 \leq \varphi < \frac{\pi}{4}, \tag{10}$$

Доказательство. Рассмотрим сначала случай $\pi/4 \leq \varphi \leq \pi/2$. Для $k = 0$ равенство (9) выполняется. Пусть оно справедливо для $k = r$ и $g_f(x_r) \neq 0$. Докажем его выполнение для $k = r + 1$

$$\|x_{r+1} - x^*(x_{r+1})\|^2 \leq \|x_{r+1} - x^*(x_r)\|^2 =$$

$$= \|x_r - x^*(x_r)\|^2 - 2h_{r+1}\left(x_r - x^*(x_r), \frac{g_f(x_r)}{\|g_f(x_r)\|}\right) + h_{r+1}^2.$$

Воспользовавшись последовательно неравенствами (3), (9) и учитывая, что при $\varphi,\ \pi/4 \le \varphi < \pi/2,\ 1 - 2\cos^2\varphi \ge 0$.

$$\begin{aligned}\left\|x_{r+1} - x^*\left(x_{r+1}\right)\right\|^2 &\le \left\|x_r - x^*\left(x_r\right)\right\|^2\left(1 - 2\cos^2\varphi\right) + h_{r+1}^2 \le \\ &\le \frac{h_{r+1}^2}{\cos^2\varphi}\left(1 - 2\cos^2\varphi\right) + h_{r+1}^2 = h_{r+1}^2\left(\frac{\sin\varphi}{\cos\varphi}\right)^2.\end{aligned}$$

Принимая во внимание (6), (7), получаем

$$\left\|x_{r+1} - x^*\left(x_{r+1}\right)\right\|^2 \le \frac{h_{r+2}^2}{\cos^2\varphi}.$$

Таким образом, для $\pi/4 \le \varphi \le \pi/2$ теорема справедлива.

Рассмотрим случай $0 \le \varphi < \dfrac{\pi}{4}$. Для $k = 0$ (10) выполняется. Пусть оно справедливо и для $k = r,\ g_f\left(x_r\right) \ne 0$. Докажем его выполнение для $k = r + 1$.

Воспользовавшись (3), (10), (6), (8), имеем

$$\begin{aligned}&\left\|x_{r+1} - x^*\left(x_{r+1}\right)\right\|^2 \le \left\|x_r - x^*\left(x_r\right)\right\|^2 - 2h_{r+1}\cos\varphi\left\|x_r - x^*\left(x_r\right)\right\| + h_{r+1}^2 \le \\ &\le \left\|x_r - x^*\left(x_r\right)\right\|^2 - \left\|x_r - x^*\left(x_r\right)\right\|^2 + h_{r+1}^2 = h_{r+1}^2 = \left(2\cos\varphi\, h_{r+2}\right)^2.\end{aligned}$$

Теорема доказана.

Замечание. Из доказательства очевидно, что для справедливости теоремы достаточно выполнения условия (3) лишь для точек последовательности $\{x_k\}_{k=0}^{\infty}$.

Таким образом, если угол φ заранее известен, то, регулируя шаг по формулам (6) и (7), (8), мы можем получить сходимость к минимуму со скоростью геометрической прогрессии и знаменателем $q = r(\varphi)$.

В (3) $\cos\varphi$ показывает степень вытянутости поверхностей уровня функции $f(x)$. Если в некоторой окрестности минимума $f(x)$ не существует $\varphi < \pi/2$ такого, что для любого x из этой окрестности выполняется (3), то будем называть эту функцию существенно овражной. При минимизации существенно овражных функций приведенный в теореме способ регулировки шаговых множителей неприменим. В этом случае нужно использовать универсальный метод выбора шаговых множителей, как в теореме 4.

Сформулируем теорему, аналогичную предыдущей, непосредственно в терминах, характеризующих степень «вытянутости» поверхностей уровня.

Теорема 7. [*11*]. *Пусть выпуклая функция* $f(x)$ *определена на* E_n, x^* *– единственная точка минимума. Задано начальное приближение* x_0 *и числа* σ *и* h_1, *причем* $\sigma \geq \sqrt{2}$, $h_1 \geq \|x_0 - x^*\| / \sigma$.

Рассмотрим множество $Y = \{y : \|y - x^*\| \leq \sigma h_1\}$. *Если для любой пары точек* $x, z \in Y$ *и такой, что* $f(x) = f(z) \neq f(x^*)$, *выполняется условие* $\|x - x^*\| / \|z - x^*\| \leq \sigma$, *то последовательность* $\{x_k\}_{k=0}^{\infty}$, *образованная с помощью рекуррентных формул*

$$x_{k+1} = x_k - h_{k+1} \frac{g_f(x_k)}{\|g_f(x_k)\|},$$

причем $h_{k+1} = h_k \sqrt{\sigma^2 - 1} / \sigma$, *сходится к* x^* *со скоростью геометрической прогрессии,* $\|x_k - x^*\| \leq h_{k+1} \sigma$, $k = 0, 1, \ldots$, *за исключением случая, когда для некоторого* $k = \overline{k}$ $g_f(x_k) = 0$, *т. е.* $x_{\overline{k}} = x^*$.

В [14] предложен еще один способ регулировки шаговых множителей в методе ОГС, основанный на априорном знании значения функции в точке минимума.

Пусть $f(x)$ – выпуклая функция, определенная в E_n, x^* – единственная точка минимума, $f(x^*) = f^*$. Построим итеративный процесс вида

$$x_{k+1} = x_k - \frac{\gamma [f(x_k) - f^*]}{\|g_f(x_k)\|^2} g_f(x_k). \tag{11}$$

Теорема 8. *При* $0 < \gamma < 2$ *процесс* (11) *сходится к* x^* *из любого начального приближения.*

В [14] получены также условия, при которых метод (11) сходится со скоростью геометрической прогрессии.

Теорема 9. *Пусть* $f(x)$ *– сильно выпуклый функционал, причем* $f(x) - f(x^*) \geq m \|x - x^*\|^2$ *удовлетворяет условию Липшица на области* $\{x : \|x - x^*\| \leq \|x_0 - x^*\|\}$ *с константой М. Тогда*

$$\|x_k - x^*\| \leq q^k \|x_0 - x^*\|,$$

$$q = \left(1 - \frac{\gamma (2 - \gamma) m^2}{M^2}\right)^{1/2} < 1.$$

В [15] метод ОГС обобщается на класс почти-дифференцируемых функций. Доказана следующая теорема.

Теорема 10. *Пусть $f(x)$ – почти-дифференцируемая функция, x^* – точка локального минимума этой функции такая, что*

$$f(x^*) = \min_{x \in S_r} f(x), \quad S_r = \{x : \|x - x^*\| \le r\}, \quad \|x^* - x_0\| \le r$$

и для любого ε, $0 < \varepsilon < r$, $\inf\limits_{x \in S_r \setminus S_\varepsilon} (g_f(x), x - x^*) > 0$. Тогда последовательность $\{x_k\}_{k=0}^{\infty}$, образованная по формулам

$$x_{k+1} = \begin{cases} \overline{x}_{k+1}, & \text{если } \overline{x}_{k+1} \in S_r, \\ x_0, & \text{если } \overline{x}_{k+1} \overline{\in} S_r, \end{cases}$$

где $\overline{x}_{k+1} = x_k - h_k \dfrac{g_f(x_k)}{\|g_f(x_k)\|}$, при условиях $h_k > 0$, $\sum\limits_{k=0}^{\infty} h_k = \infty$; $\lim\limits_{k \to \infty} h_k = 0$ сходится к x^*.

В [7] введено понятие обобщенного стохастического градиента, предложен и обоснован метод случайного поиска, являющийся стохастическим аналогом ОГС. В дальнейшем этот метод получил свое развитие в работах Ю. М. Ермольева и его учеников и является довольно эффективным средством решения задач стохастического программирования.

Из других исследований метода ОГС отметим работу М. А. Шепилова [16], в которой сформулирован результат об его устойчивости к малым ошибкам в определении x_k и $g_f(x_k)$.

Теорема 11. *Если функция $f(x)$ выпукла в E_n, множество*

$$\Omega = \{x : f(x) = f^*\}$$

не пусто, то при любой начальной x_0 в $x_{k+1} = x_k - \lambda_k \dfrac{g_f(\overline{x}_k)}{\|g_f(\overline{x}_k)\|}$, где

$$\|\overline{x}_k - x_k\| \le \delta_k, \quad \lim_{k \to \infty} \delta_k = 0, \quad \lambda_k > 0, \quad \lim_{k \to \infty} \lambda_k = 0,$$

$$\sum_{k=0}^{\infty} \lambda_k \delta_k < \infty, \qquad \sum_{k=0}^{\infty} \lambda_k^2 < \infty, \quad \sum_{k=0}^{\infty} \lambda_k = \infty,$$

справедливо равенство $\lim\limits_{k \to \infty} x_k = x^* \in \Omega$.

Эта теорема вместе со способом приближенного вычисления обобщенного градиента, упомянутым нами на стр. 15, обосновывает возможность построения универсального алгоритма нахождения минимума выпуклой функции $f(x)$ (при условии, что мы имеем возможность вычислять значения функции $f(x)$ в произвольной точке с любой точностью).

В [16] получены также некоторые новые достаточные условия сходимости ОГС, когда множество минимумов не является ограниченным. Е. Г. Гольштейн в [17] предложил и обосновал алгоритм, переносящий метод ОГС на задачи нахождения седловой точки выпукло-вогнутой функции. Аналогичный алгоритм без обоснования сходимости описан в [18].

4. Метод градиентного типа с растяжением пространства в направлении почти-градиента

Анализ метода обобщенного градиентного спуска показывает, что с помощью одних лишь попыток регулировки шагового множителя трудно добиться в достаточно общем случае значительного ускорения сходимости. В самом деле, медленная сходимость связана с тем, что антиградиент образует угол, близкий к $\pi/2$, с направлением на точку минимума. В такой ситуации расстояние до точки минимума убывает на величину, намного меньшую, чем длина шага, а значит, и скорость уменьшения шага не может быть слишком большой, если мы хотим гарантировать сходимость к минимуму.

Имеется простая возможность изменения углов между градиентом и направлением на точку минимума – использование линейных неортогональных преобразований пространства. Возникает идея построения в процессе последовательных приближений линейных операторов, изменяющих метрику пространства, и выбора направления спуска, соответствующего антиградиенту в пространстве с новой метрикой. Это направление может значительно отличаться от направления антиградиента.

Как же строить операторы преобразования? Когда рассматривается класс дважды непрерывно дифференцируемых функций, то основная идея, лежащая в построении алгоритмов с изменяемой метрикой, состоит в получении тем или иным способом матрицы, близкой к обратной гессиану в точке минимума, т. е. фактически используется идея квадратической аппроксимации минимизируемой функции и имитации метода Ньютона-Рафсона без вычисления вторых производных [19]. Для минимизации негладких функций такой подход в принципе невозможен.

Например, для любых кусочно-линейных функций гессиан почти везде равен 0, и матрицы, обратной гессиану, просто не существует. Поэтому для построения алгоритмов градиентного типа с преобразованием пространства, рассчитанных на достаточно широкий класс негладких функций, нужно использовать другие идеи.

Автором были предложены два типа алгоритма с изменяемой метрикой. Оба они связаны с использованием операции растяжения пространства в определенных направлениях.

Рассмотрим алгоритмы с растяжением пространства в направлениях: сначала градиента, а затем разности двух последовательных градиентов.

Пусть задан вектор $\xi \in E_n$, $\|\xi\| = 1$ и число $\alpha \geq 0$. Каждый вектор $x \in E_n$ однозначно представим в виде

$$x = \gamma_\xi(x)\,\xi + d_\xi(x), \tag{12}$$

$$(\xi, d_\xi(x)) = 0, \tag{13}$$

При этом

$$\gamma_\xi(x) = (x, \xi)\,, \quad d_\xi(x) = x - (x, \xi)\ \xi. \tag{14}$$

Определение. *Оператором растяжения пространства E_n в направлении ξ с коэффициентом α назовем $R_\alpha(\xi)$, действующий следующим образом на вектор x, представленный в форме* (12)–(14)

$$R_\alpha(\xi)\,x = \alpha\gamma_\xi(x)\,\xi + d_\xi(x) = x + (\alpha - 1)(x, \xi)\,\xi, \tag{15}$$

где $R_\alpha(\xi)$ – линейный симметричный оператор. Легко видеть, что

$$R_{\alpha\beta}(\xi) = R_\alpha(\xi)\,R_\beta(\xi)\,, \quad R_\alpha(\xi)\,R_{1/\alpha}(\xi) = R_1(\xi) = I, \quad \alpha > 0,$$

где $R_0(\xi)$ – оператор проектирования на подпространство, ортогональное ξ. Оператор $R_\alpha(\xi)$ при $n \geq 2$ имеет два собственных числа: $\lambda_1 = \alpha$, $\lambda_2 = 1$. Первому из них соответствует подпространство собственных векторов, порожденное вектором ξ, второму – подпространство собственных векторов, состоящее из векторов, ортогональных ξ.

Рассмотрим класс алгоритмов минимизации почти-дифференцируемых функций, на каждом шаге которых движение в направлении, противоположном почти-градиенту, будет сочетаться с операцией растяжения пространства в том же направлении.

Пусть $f(x)$ – почти-дифференцируемая функция, $x_0 \in E_n$ – заданное начальное приближение, $B_0 = A_0^{-1}$ – неособая матрица (в частности, можно брать $B_0 = I$).

Определим бесконечно шаговый процесс, $(k+1)$-й шаг которого описывается следующим образом, $k = 0, 1, 2, \ldots$. Вычисляем: $g_f(x_k)$ (если $g_f(x_k) = 0$, процесс останавливается)

$$g_{\varphi_k}(y_k) = B_k^* g_f(x_k) = \widetilde{g}_k, \tag{16}$$

где $\varphi_k(y) = f(B_k y)$; $y_k = A_k x_k$; $A_k = B_k^{-1}$; B_k^* – оператор, сопряженный оператору B_k.

Формула (16) дает возможность найти почти-градиент от функции $\varphi_k(y) = f\left(A_k^{-1} y\right)$, которая получается из $f(x)$ при использовании линейного преобразования пространства $y = A_k x$

$$\xi_{k+1} = \frac{g_{\varphi_k}(y_k)}{\|g_{\varphi_k}(y_k)\|} = \frac{\widetilde{g}_k}{\|\widetilde{g}_k\|},$$

$$x_{k+1} = x_k - B_k\, h_{k+1}\, \xi_{k+1}. \tag{17}$$

Выражение (17) получается из формулы

$$A_k\, x_{k+1} = y_k - h_{k+1}\, \xi_{k+1}, \tag{18}$$

определяющей шаг обобщенного градиентного процесса для функции $\varphi_k(y)$, с последующим применением к обеим частям (18) оператора B_k для отображения в основное пространство E_n. Здесь h_{k+1} – шаговый множитель,

$$B_{k+1} = A_{k+1}^{-1} = B_k R_{\beta_{k+1}}(\xi_{k+1}), \tag{19}$$

где $\beta_{k+1} = 1/\alpha_{k+1}$, $\alpha_{k+1} > 1$ – коэффициент растяжения пространства.

Формула (19) дает возможность вычислить оператор B_{k+1}, обратный результирующему оператору A_{k+1} преобразования пространства, который получается в результате последовательного применения операторов растяжения пространства в направлении нормированных почти-градиентов $\xi_1, \ldots, \xi_{k+1}$ с коэффициентами $\alpha_1, \ldots, \alpha_{k+1}$:

$$A_{k+1} = R_{\alpha_{k+1}}(\xi_{k+1})\, A_k,$$

откуда

$$B_{k+1} = A_{k+1}^{-1} = B_k R_{1/\alpha_{k+1}}(\xi_{k+1}) = B_k R_{\beta_{k+1}}(\xi_{k+1}).$$

Переходим к $(k+2)$-му шагу. Только что описанное семейство процедур минимизации будем называть РП-алгоритмами (алгоритмами с растяжением пространства). Конкретная реализация РП-алгоритма зависит от выбора последовательностей $\{\alpha_k\}$ и $\{h_k\}$. Ниже мы остановимся на некоторых способах выбора этих последовательностей, обеспечивающих при определенных условиях сходимость к точке минимума.

Справедливы следующие теоремы.

Теорема 12. *Пусть x^* – точка локального минимума функции $f(x)$ и в процессе реализации алгоритма (16)–(18) выполняются условия:*

1) $\|x_k - x^*\| \leq d, \quad d > 0, \qquad k = 1, 2, \ldots,$

2) $1 + \delta \leq \alpha_k \leq \alpha^*, \qquad \alpha_k = 1/\beta_k, \quad \delta > 0.$

Тогда существует такая подпоследовательность $\{x_{k_p}\}_{p=1}^{\infty}$ и $c > 0$, что

$$\|\widetilde{g}_{k_p}\| \leq c \left(\prod_{j=1}^{k_p} \alpha_j \right)^{-1/n}, \quad p = 1, 2, \ldots .$$

Теорема 13. *Пусть $f(x)$ – почти-дифференцируемая функция, определенная в некоторой сферической окрестности S_d точки x^* локального минимума, и в тех точках, где функция дифференцируема, ее производная по направлению $\mu(x) = x - x^*$ удовлетворяет неравенству*

$$N \leq \frac{f'_{\mu(x)}(x)}{f(x) - f(x^*)} \leq M, \quad M > N > 0, \quad x \neq x^*.$$

Тогда, если в алгоритме (16), (17), (19) примем

1) $x_0 \in S_d,$

2) $h_{k+1} = \dfrac{2MN}{M+N} \cdot \dfrac{f(x_k) - f(x^*)}{\|\widetilde{g}_k\|},$

3) $\alpha_{k+1} = \alpha = \dfrac{M+N}{M-N},$

то найдутся константа $\overline{c}$ и подпоследовательность индексов $\{k_p\}_{p=1}^{\infty}$, $k_p < k_{p+1}$, такая, что $f(x_{k_p}) - f(x^*) \leq \overline{c}\, \alpha^{-k_p/n}$.

Описанный в теореме 13 способ выбора коэффициента растяжения и шагового множителя находит непосредственное применение при решении систем нелинейных уравнений $f_i(x) = 0,\ i = 1, \ldots, n$, путем сведения к задаче $\min \max\limits_{1 \le i \le n} |f_i(x)|$. При этом, как показано в [8], в регулярном случае при достаточно хорошем начальном приближении константы M и N можно выбирать близкими к единице, что обеспечивает быструю сходимость.

Если $f(x)$ – выпуклая функция, то N можно всегда выбирать больше или равным единице. При решении выпуклых задач, минимаксных и выпуклого программирования обычно $f(x^*)$ неизвестно, поэтому представляет интерес вопрос о подборе $f(x^*)$ в процессе счета. Справедлива следующая теорема.

Теорема 14. *Пусть выпуклая функция* $f(x)$ *обладает следующими свойствами:* а) *существует постоянная* $M > 1$ *такая, что если* $\varphi(\alpha) = f[(1-\alpha)x_1 + \alpha x_2], \quad 0 \le \alpha < 1$ *строго убывает по* α, *то выполняется неравенство* $f'_{x_1 - x_2}(x_1) \le M[f(x_1) - f(x_2)]$;
б) $\lim\limits_{\|x\| \to \infty} f(x) = +\infty$. *Тогда, если при применении алгоритма* (16), (17), (19)

$$\alpha_{k+1} = \frac{M+1}{M-1}$$

$$h_{k+1} = \frac{2M}{M+1} \cdot \frac{[f(x_k) - \overline{f}]}{\|\widetilde{g}_k\|}$$

$\overline{f}$ *выбрано большим или равным* $f^* = \min\limits_{x \in E_n} f(x)$, *то последовательность* $\{h_k\}$ *является ограниченной и для произвольного* $\varepsilon > 0$ *найдется* $\overline{k}$ *такое, что* $f(x_{\overline{k}}) \le \overline{f} + \varepsilon$ *(счет прекращается, если на некотором шаге* $f(x_{\overline{k}}) \le \overline{f}$*); если* $\overline{f}$ *выбрано меньшим* f^*, *то последовательность* h_k *является неограниченной.*

Эта теорема позволяет построить алгоритм минимизации функции $f(x)$, удовлетворяющей условию теоремы, при неизвестном f^* [20]. Этот алгоритм связан с подбором f^* с использованием следующих признаков: если при некотором $\overline{f}$ h_k превосходит некоторое число, то $\overline{f}$ увеличивается; при приближении $f(x_k)$ к $\overline{f}$ – уменьшается [1].

[1]Подробное описание и обоснование алгоритма см. в [21]

5. Метод градиентного типа с растяжением пространства в направлении разности двух последовательных градиентов

Этот метод был предложен в [20]. В [22] отдельные модификации его экспериментально исследованы применительно к задачам минимизации гладких функций, а в [23]–[24] кусочно-гладких функций. В настоящее время с его помощью решено большое число сложных практических задач и можно с уверенностью говорить о высокой эффективности этого метода, особенно для задач минимаксного типа, нелинейного программирования (с использованием негладких штрафных функций), для минимизации функций «овражного» типа.

Рассмотрим основную его модификацию – так называемый $r(\alpha)$-алгоритм применительно к минимизации почти-дифференцируемых функций, в частности кусочно-гладких.

Пусть $f(x)$ – почти-дифференцируемая функция, $\alpha > 1$ $\beta = 1/\alpha$. $r(\alpha)$-алгоритмом минимизации назовем итеративную процедуру следующего типа: на каждой итерации вычисляем n-мерные векторы x_k, $\widetilde{g}_k$ и матрицу размерности $n \times n$ $B_k = \{b_{ij}\}$; $x_0 \in E_n$, $\widetilde{g}_0 = 0$, B_0 – единичная матрица. Пусть проделаны k итераций, вычислены x_k, $\widetilde{g}_k$, B_k. На $(k+1)$-й итерации находим:

1) $g_f\,(x_k)$ – почти-градиент функции $f(x)$ в точке x_k; если он определяется неоднозначно, то берем такой почти-градиент, что $(B_k\,\widetilde{g}_k, g_f\,(x_k)) \leq 0$;

2) $g_k^* = B_k^*\,g_f\,(x_k)$, g_k^* – почти-градиент функции $\varphi_k(y) = f\,(B_k\,y)$ в точке $y = A_k\,x_k$, где $A_k = B_k^{-1}$;

3) $r_k = g_k^* - \widetilde{g}_k$, r_k – разность двух почти-градиентов от функции $\varphi_k(y)$, вычисленных в точках $y_k = A_k\,x_k$, $\widetilde{y}_k = A_k\,x_{k-1}$;

4) $\xi_{k+1} = \dfrac{r_k}{\|r_k\|}$;

5) $B_{k+1} = B_k\,R_\beta\,(\xi_{k+1})$, B_{k+1} – матрица, обратная A_{k+1}, A_{k+1} – матрица преобразования пространства после $(k+1)$-го шага, $A_{k+1} = R_\alpha\,(\xi_{k+1})\,A_k$, $R_\alpha(\xi)$ – оператор растяжения пространства в направлении ξ; $\|\xi\| = 1$, α – коэффициент растяжения пространства, β – коэффициент «сжатия» пространства градиентов;

6) $\widetilde{g}_{k+1} = R_\beta(\xi_{k+1})\ g_k^* = B_{k+1}^*\, g_f(x_k)$, $\widetilde{g}_{k+1}$ – значение почти-градиента функции $\varphi_{k+1}(y) = f(B_{k+1}\ y)$ в точке $y = A_{k+1}\, x_k$;

7) $x_{k+1} = x_k - h_{k+1}\, B_{k+1}\, \widetilde{g}_{k+1}$, где h_{k+1} выбирается из условия

$$h_{k+1} = \arg \overline{\min_{h \geq 0}} f(x_k - h\, B_{k+1}\, \widetilde{g}_{k+1}) \tag{20}$$

(операция $\overline{\min}$ означает определение ближайшего к нулю локального минимума); применим к обеим частям 7) оператор A_{k+1} : $y_{k+1} = A_{k+1}\, x_{k+1} = A_{k+1}\, x_k - h_{k+1}\, \widetilde{g}_{k+1} = \widetilde{y}_{k+1} - h_{k+1}\, \widetilde{g}_{k+1}$ и получим, что 7) фактически реализует шаг наискорейшего спуска для функции $\varphi_{k+1}(y)$;

8) переход к $(k+2)$-й итерации с запоминанием x_{k+1}, $\widetilde{g}_{k+1}$, B_{k+1} или окончание работы алгоритма при выполнении некоторого критерия остановки.

В отличие от других градиентных методов минимизации негладких функций: обобщенного градиентного спуска, ОГС РП [10], [25], $r(\alpha)$-алгоритм обеспечивает монотонность спуска благодаря определенному по (20) способу выбора шага спуска: $f(x_0) \geq f(x_1) \geq \ldots \geq f(x_k)$. Однако он существенно отличается от обычных релаксационных методов следующим: если $h_k = 0$, то это не значит, что процесс спуска прекращается. При выполнении серии итераций с нулевым шагом x_k стоит на месте, но изменяются B_k и $\widetilde{g}_k$. В это время как бы в скрытой форме осуществляется поиск подходящего направления спуска.

Чтобы избежать в некотором смысле «патологических» случаев, предположим:

а) $\lim\limits_{\|x\| \to \infty} f(x) = +\infty$, что гарантирует существование минимума $f(x)$ и минимума в (20);

б) $f(x)$ обладает следующим свойством: в любой ограниченной области D для любого $\delta > 0$ найдется такое $\varepsilon > 0$, что если $x,\ y \in D$, $f(x) - f(y) < \varepsilon$ и на отрезке между x и y $f(x)$ убывает, то $\|x - y\| < \delta$ (для выпуклых функций это свойство является аналогом сильной выпуклости).

Наиболее общий результат о сходимости $r(\alpha)$-алгоритмов, доказанный к настоящему времени, сформулирован в следующей теореме.

Теорема 15. *Если $f(x)$ – непрерывная кусочно-гладкая функция, для которой выполняются свойства а) и б), то последовательность $\{x_k\}$, генерируемая $r(\alpha)$-алгоритмом при любом $\alpha > 1$, имеет в качестве своей предельной некоторую точку $\overline{x}$, множество почти-градиентов которой образует линейно-зависимое семейство векторов. При этом в силу монотонности процесса спуска последовательность $f(x_k)$ сходится к $f(\overline{x})$.*

Доказательство содержится в [26]. Так как оно довольно громоздко, то мы его приводить не будем; заметим только, что это доказательство основывается на изучении поведения ширины выпуклого замыкания множества почти-градиентов от функций $\varphi_k(y) = f\left(A_k\, y\right)$ в точках $\{y_k\}$. Из теоремы 15 вытекает следующая теорема.

Теорема 16. *Если выполняются условия теоремы 15, x^* – изолированная точка локального минимума, x_0 – такая точка, что связная компонента множества $\{x : \ f(x^*) \leq f(x) \leq f(x_0)\}$, содержащая x^*, x_0, не имеет кроме x^* других точек z, у которых семейство $G_f(z)$ линейно-зависимо, то последовательность $\{x_k\}_{k=0}^{\infty}$, генерируемая $r(\alpha)$-алгоритмом, сходится к x^*.*

Скорость сходимости слабо изучена (так, в [22] описан результат о скорости сходимости предельного варианта $r(\alpha)$-алгоритма). Многочисленные вычислительные эксперименты (см. [22], [23]) показали высокую эффективность $r(\alpha)$-алгоритма и его модификаций, которые в основном связаны с различными способами приближенного поиска минимума по направлению.

В [27] предложены видоизмененные схемы вычислений в методах с растяжением пространства. Суть их состоит в том, чтобы вместо матриц B_k, запоминать и преобразовывать $H_k = B_k\, B_k^*$. Так как H_k – симметричная матрица, то для ее запоминания требуется $n^2/2$ ячеек памяти. Уменьшается при этом также число арифметических операций, связанных с преобразованием матриц и векторов. Так, $r(\alpha)$-алгоритм по новой схеме вычислений сводится к следующему.

Принимаем $H_0 = I$. На k-м шаге имеем $H_k,\ x_k,\ g_f\left(x_{k-1}\right)$.

Вычисляем:

$$g_f\left(x_k\right),\ e_k = g_f\left(x_k\right) - g_f\left(x_{k-1}\right);$$

$$H_{k+1} = H_k - \left(1 - \beta_k^2\right) \frac{\left(H_k\, e_k\right)\left(H_k\, e_k\right)^T}{\left(e_k, H_k\, e_k\right)},$$

$$x_{k+1} = x_k - t_k\, p_k; \qquad p_k = H_{k+1}\, g_f\left(x_k\right); \quad t_k = \arg\,\overline{\min} f\left(x_k - t_{p_k}\right).$$

В [28, 29] опубликованы два похожих по идее алгоритма минимизации произвольных выпуклых функций. По форме они близки, с одной стороны, к методу наискорейшего спуска решения минимаксных задач В. Ф. Демьянова, а с другой, – к одной из модификаций метода сопряженных градиентов. Насколько можно судить по первым кратким публикациям, эти алгоритмы при удачном выборе параметров обладают определенными преимуществами в смысле быстроты сходимости по сравнению с методом обобщенного градиентного спуска и требуют дополнительно запоминания одного или нескольких векторов размерности n, если минимизируемая функция определена в E_n.

6. Применение обобщенных градиентных методов в задачах математического программирования

Схемы декомпозиции. Рассмотрим схему декомпозиции для задач выпуклого программирования с разбиением множества переменных на две части. Найти

$$\min_{x,y} f_0(x,y) \tag{21}$$

при ограничениях

$$f_i(x,y) \le 0, \quad i = 1,\dots,n, \tag{22}$$

где x, y – векторные переменные, $x = (x^{(1)},\dots,x^{(l)})$, $y = (y^{(1)},\dots,y^{(m)})$; f_0, f_i – выпуклые функции по совокупности переменных x, y, $i = 1,\dots,n$.

Зафиксируем $x = \overline{x}$ и рассмотрим задачу: найти

$$\min_{y} f_0(\overline{x},y) \tag{23}$$

$$f_i(\overline{x},y) \le 0, \quad i = 1,\dots,n, \tag{24}$$

Для тех значений $\overline{x}$, для которых существует оптимальное решение задачи (23)–(24), определим функцию $\Phi(x)$

$$\Phi(\overline{x}) = \min_{y\in D(\overline{x})} f_0(\overline{x},y), \tag{25}$$

где $D\,(\overline{x})$ – множество значений, удовлетворяющих (24).

Теорема 17. *Если задача (21)–(22) имеет решение, то функция $\Phi(x)$ в соответствии с (25) является выпуклой, определенной на выпуклом подмножестве $W \subseteq E_l$. Если для некоторого $\overline{x} \in W$ выполняется условие Слейтера в задаче (23)–(24), то обобщенный градиент функции $\Phi(x)$ в точке $x = \overline{x}$ вычисляется по формуле*

$$g_\Phi(\overline{x}) = g^x_{L_u}\Big(\overline{x}, y(\overline{x})\Big), \tag{26}$$

где $L_u(x,y) = f_0(x,y) + \sum_{i=1}^n u_i f_i(x,y)$; $y(\overline{x})$ – одно из оптимальных значений y в (23)–(24); $u = \{u_i\}_{i=1}^n$ – множители Лагранжа задачи (23)–(24), полученные в соответствии с теоремой Куна–Таккера; $g^x_{L_u}(\overline{x}, y(\overline{x}))$ – проекция такого обобщенного градиента функции $L_u(x) = L_u(x,y)$ на подпространство E^x_l, у которого проекция на подпространство E^y_m равна нулю (обобщенный градиент берется в точке $z = (\overline{x}, y(\overline{x}))$).

Следствие. *Если в условиях предыдущей теоремы дополнительно потребовать, чтобы $f_\alpha(x,y)$, $\alpha = 0,1,\ldots,n$, были непрерывно дифференцируемыми по y, то формула (26) уточняется следующим образом*

$$g_\Phi(\overline{x}) = g^x_{f_0}(x, y(\overline{x})) + \sum_{i=1}^n u_i(\overline{x})\, g^x_{f_i}(\overline{x}, y(\overline{x})), \tag{27}$$

где $g^x_{f_\alpha}(\overline{x}, y(\overline{x}))$ – проекции произвольных обобщенных градиентов функции f_α в точке $\overline{z} = (\overline{x}, y(\overline{x}))$ на E^x_l.

Выражение (27) позволяет легко строить алгоритмы решения задач выпуклого программирования (в частности, линейного) с использованием схемы разложения по координатам. Для этого мы должны предположить возможность получения за конечное число шагов решения задачи (23)–(24), множителей Лагранжа $u = \{u_i\}_{i=1}^n$, а также «частных» обобщенных градиентов $g^x_{f_\alpha}$, $\alpha = 0,1,\ldots,n$. Это реализуемо, например, когда (23)–(24) является задачей линейного или квадратичного программирования. Осложнения могут быть связаны с тем, что (23)–(24) не при всех $\overline{x}$ будет иметь решения. Стандартный выход из такой ситуации заключается в использовании метода «штрафных функций»; задача (21)–(22) заменяется следующей: найти

$$\min_{x,y,v}\left[f_0(x,y) + M\sum_{i=1}^n v_i\right] \tag{28}$$

при
$$f_i(x,y) - v_i \le 0, \quad v_i \ge 0, \quad i = 1,\dots,n, \tag{29}$$
где M – достаточно большое положительное число.

Задача (23)–(24) заменяется такой: найти
$$\min_{y,v}\left[f_0(\overline{x},y) + M\sum_{i=1}^{n} v_i\right] \tag{30}$$
при
$$f_i(\overline{x},y) - v_i \le 0, \quad v_i \ge 0. \tag{31}$$

Легко видеть, что (23)–(24) имеет допустимое решение при любом $\overline{x}$ с выполнением условия Слейтера.

Итак, мы можем предположить, что задача (21)–(22) сведена к такому виду, что функция $\Phi(\overline{x})$ определена для любого $\overline{x}$ и имеется возможность вычисления обобщенного градиента $g_\Phi(x)$ в произвольной точке x. Короче говоря, можно применить метод обобщенного градиентного спуска или методы с растяжением пространства к задаче минимизации функции $\Phi(x)$.

Исследуем теперь схему разложения по ограничениям. Она является в некотором смысле двойственной к схемам разложения по переменным. Пусть имеется задача выпуклого программирования, множество ограничений которой некоторым образом разбито на две части: найти
$$\min f_0(x), \quad x \in E_l, \tag{32}$$
при
$$f_i(x) \le 0, \quad i = 1,\dots,m, \tag{33}$$
$$\varphi_j(x) \le 0, \quad j = 1,\dots,n. \tag{34}$$

Рассмотрим задачу следующего вида: найти
$$L(u) = \min_{x\in D}\left[f_0(x) + \sum_{i=1}^{m} u_i\, f_i(x)\right], \tag{35}$$
где $u = \{u_i\}_{i=1}^{m}$ – m-мерный вектор с неотрицательными компонентами; D – множество, определяемое ограничениями (34).

В качестве непосредственно следствия теоремы Куна–Таккера получаем утверждение: если решение задачи (32)–(34) существует и ограничения ее удовлетворяют условию Слейтера, D – ограниченное множество, то $L(u)$ – вогнутая функция, определенная для всех $u \ge 0$, максимум функции $L(u)$ существует и достигается в точке u^*, при этом u^* является u-компонентой седловой точки функции Лагранжа задачи (32)–(34)

$$\Phi(x,u,v) = f_0(x) + \sum_{i=1}^{m} u_i\, f_i(x) + \sum_{j=1}^{n} v_j\, \varphi_j(x).$$

Определения максимума вогнутой функции $L(u)$ и минимума выпуклой функции $L_1(u) = -L(u)$ – эквивалентны.

Минимум выпуклой функции $L_1(u)$ при ограничении $u \geq 0$ можно находить, используя метод обобщенного градиентного спуска с проектированием на множество

$$\{u: \; u \geq 0\} \; : \; u_{k+1}^{(i)} = \max\{0, u_k^{(i)} + h_{k+1}\, f_i(x^*\,(u_k))\}, \quad i = 1, \ldots, m.$$

Операция проектирования вектора на это множество очень проста и сводится к тому, что положительные и нулевые координаты проектируемого вектора остаются неизменными, а отрицательные заменяются нулем; $x^*\,(u_k)$ – оптимальный план (35).

Для частного случая – задачи линейного программирования – приведенная схема разложения подробно рассмотрена в [30]. Конкретные приложения к решению задач большого объема описаны в [1]–[4], [12], [31], [32].

Задачи выпуклого программирования и общие задачи нелинейного программирования. Пусть $f_\nu(x)$, $\nu = 0, \ldots, m$ – выпуклые функции, определенные на E_n. Найдем

$$\min f_0(x) \tag{36}$$

при ограничениях

$$f_i(x) \leq 0, \quad i = 1, \ldots, n. \tag{37}$$

Если существует седловая точка (x^*, λ^*), $\lambda^* \geq 0$ функции Лагранжа $L(x,\lambda) = f_0(x) + \sum\limits_{i=1}^{m} \lambda_i\, f_i(x)$, $\lambda = (\lambda_1, \ldots, \lambda_m)$, то, как показано в [33], задача (36)–(37) эквивалентна задаче минимизации по x $L^+\left(x, \overline{\lambda}\right)$ при любом $\overline{\lambda} \geq \lambda^*$, где

$$L^+\,(x,\lambda) = f_0(x) + \sum_{i=1}^{m} \lambda_i^+(x)\, f_i(x);$$

$$\lambda_i^+(x) = \begin{cases} \lambda_i, & f_i(x) > 0, \\ 0, & f_i(x) \leq 0. \end{cases}$$

Таким образом, имея оценки сверху для множителей Лагранжа в задаче выпуклого программирования (во многих практических задачах такого рода оценки довольно легко получить), мы сводим ее к задаче минимизации негладкой выпуклой функции, для решения которой можно применять любой из изложенных выше обобщенных градиентных методов. При этом очевидное преимущество по сравнению с «гладким» вариантом метода штрафных функций (минимизации $\widetilde{L}\left(x,\overline{\lambda}\right) = f_0(x) + \sum\limits_{i=1}^{m} \lambda_i^+(x)\, f_i^2(x))$ состоит в том, что для получения решения с высокой точностью не нужно выбирать слишком большие штрафные коэффициенты или увеличивать их в процессе счета, что отрицательно влияет на скорость сходимости.

Метод негладких штрафных функций можно с успехом использовать для получения локальных экстремумов в общих задачах нелинейного программирования. Например, такой: найти

$$\min f_0(x)$$

при ограничениях вида

$$\varphi_i(x) \leq 0, \quad i = 1, \ldots, m, \quad \psi_j(x) = 0, \quad j = 1, \ldots, l.$$

Заменим ее задачей минимизации функции

$$L_{\overline{\lambda},\overline{\mu}}(x) = f_0(x) + \sum_{i=1}^{m} \overline{\lambda}_i^+(x)\, \varphi_i(x) + \sum_{j=1}^{l} \overline{\mu}_j\, |\psi_j(x)|$$

при достаточно больших $\{\overline{\lambda}_i\}_{i=1}^m$ и $\{\overline{\mu}_j\}_{j=1}^l$. Для минимизации $L_{\overline{\lambda},\overline{\mu}}(x)$ лучше всего использовать обобщенный градиентный метод с растяжением пространства в направлении разности двух последовательных градиентов.

Другой способ решения состоит в сведении задач выпуклого программирования к системе выпуклых неравенств с неизвестным параметром – значением функционала в точке минимума. В самом деле, если f^* – минимальная величина $f_0(x)$ в (36)–(37), то эта задача эквивалентна системе неравенств

$$\begin{cases} f_0(x) - f^* \leq 0, \\ f_i(x) \leq 0, \quad i = 1, \ldots, m. \end{cases}$$

Для решения этой задачи наиболее подходит вариант обобщенного градиентного метода с растяжением пространства в направлении градиента с автоматическим подбором параметра f^*, аналогичный алгоритму минимизации с неизвестным значением функции в точке минимума [21].

При использовании схем декомпозиции в линейном и выпуклом программировании часто возникают задачи минимизации выпуклых функций $\Phi(x)$ при простейших ограничениях вида $x \geq 0$. Если к решению такого рода задач попытаться применить градиентный метод с растяжением пространства, то нужно применить либо метод штрафных функций, либо некоторый аналог метода проекции градиента, что затрудняется в связи с применением неортогональных преобразований пространства.

В [22] исследован более простой способ сведения данной задачи к безусловной оптимизации, который заключается в том, что рассматривается минимизация функции $F(y) = \Phi(y^+)$, где компоненты вектора y^+ равны модулям компонент вектора y (схема «зеркального отражения»). Хотя $F(y)$ многоэкстремальна, но все локальные минимумы этой функции являются глобальными. Для минимизации $F(y)$ можно применить обобщенный градиентный метод с растяжением пространства в направлении разности двух последовательных градиентов. Функция $\widetilde{F}(y) = \Phi(\{y_i^2\})$ имеет «ложные» стационарные точки (например, $y = 0$) и поэтому она менее удобна.

Схема «зеркального отражения» с успехом применялась при решении задач: линейного и выпуклого программирования с использованием схем декомпозиции; квадратичного программирования, которые, следуя [34], сводились к задачам минимизации квадратичных функций при простейших ограничениях; задач нелинейного программирования (в сочетании с методом штрафных функций) для «изъятия» односторонних и двусторонних ограничений на отдельные переменные.

Из специальных задач математического программирования отметим нелинейную транспортную, которая играет большую роль при расчете и выборе оптимальных параметров гидравлических сетей. Речь идет об оптимальном распределении потока на сети при заданных мощностях источников и стоков в соответствии с определенной целевой функцией, зависящей от параметров потока (обычно сепарабельная функция, аддитивные компоненты которой являются функциями от модуля потока вдоль определенного ребра сети). Если сделать на сети разрезы таким образом, чтобы сеть превратилась в дерево (число этих разрезов равно числу независимых контуров в сети), то, задавая потоки вдоль разрезов, можно однозначно определить, используя первый закон Кирхгофа, потоки вдоль остальных ребер. Таким образом, нелинейная транспортная задача сводится к задаче минимизации некоторой функции от потоков вдоль разрезов сети. Как правило, эта функция оказывается негладкой, а при решении задач оптимального проектирования – невыпуклой. Для получения локальных минимумов эффективным средством оказался градиентный метод с растяжением пространства в направлении разности двух последовательных градиентов [22, 23].

Обобщенные градиентные методы особенно эффективны при решении задач минимизации функций вида: $\varphi(x) = \max\limits_{1 \le i \le m} f_i(x)$.

В [23, 24] описан опыт решения минимаксных задач с использованием $r(\alpha)$-алгоритмов. При решении минимаксных задач большой размерности (1200 переменных) в [23] успешно был использован метод, промежуточный между ОГС и $r(\alpha)$-алгоритмом, в котором вместо матрицы B_k запоминалась последовательность векторов $\xi_1, \ldots, \xi_k$. И. Г. Овруцким проведена серия численных экспериментов по применению $r(\alpha)$-алгоритма для решения задач интерпретации результатов геофизических измерений по критерию минимизации максимального отклонения результатов, полученных с помощью теоретической модели, от результатов измерений. Эта задача сводится к нахождению минимума $\varphi_i(x)$, где $\varphi_1(x) = \max\limits_{1 \le i \le m} |f_i(x) - a_i|$; x – вектор неизвестных параметров теоретической модели; $f_i(x)$ – теоретическое значение величины в i-й точке; a_i – ее измеренное значение. Оказалось, что решение этой задачи требует, как правило, меньше машинного времени, чем решение аналогичной задачи по методу наименьших квадратов, т. е. нахождение минимума $\varphi_2(x) = \sum\limits_{i=1}^{m} [f_i(x) - a_i]^2$. Это объясняется тем, что вычисление почти-градиента функции $\varphi_1(x)$ требует гораздо меньше арифметических операций, чем градиент $\varphi_2(x)$.

В заключение отметим, что в Институте кибернетики АН УССР разработаны программы по обобщенным градиентным методам с растяжением пространства на языках «АЛГОЛ» и «ФОРТРАН» применительно к ЭВМ БЭСМ-6 и серии ЕС. С их помощью решаются практические задачи отраслевого планирования (в основном транспортно-производственные задачи большой размерности), оптимального проектирования сложных технических объектов (например, Единой газоснабжающей системы СССР), определения параметров в нелинейных моделях регрессии и др. Для оценки эффективности таких программ проведены многочисленные модельные эксперименты, результаты которых опубликованы в [22, 23].

Литература

1. Шор Н. З. Применение метода градиентного спуска для решения сетевой транспортной задачи // В сб. Материалы научного семинара по теоретическим и прикладным вопросам кибернетики и исследования операций. – К.: Научн. совет АН УССР по кибернетике. – Вып. 1. – С. 9–17.

2. Бакаев Л. А., Михалевич В. С., Брановицкая С. В., Шор Н. З. Методика и опыт решения сетевых транспортных задач большого объема на ЭЦВМ // Математические методы и проблемы производства. – М.: Экономиздат, 1963. – С. 247–257.

3. Шор Н. З., Горбач Г. И. Решение задач распределительного типа методом обобщенного градиентного спуска // Теория оптимальных решений. – К., 1968. – № 3. – С. 59–71.

4. Шор Н. З. О структуре алгоритмов численного решения задач оптимального планирования и проектирования // Диссертация на соискание степени канд. физ.-мат. наук. – К., 1964. – 135 с.

5. Ермольев Ю. М. Методы решения нелинейных экстремальных задач // Кибернетика. – 1966. – № 4.

6. Поляк Б. Т. Один общий метод решения экстремальных задач // Докл. АН СССР. – 1967. – **174**. – № 1.

7. Ермольев Ю. М., Шор Н. З. Метод случайного поиска для задач двухэтапного стохастического программирования // Кибернетика. – 1968. – № 1. – С. 90–92.

8. Шор Н. З. О классе почти-дифференцируемых функций и одном методе минимизации функций этого класса // Кибернетика. – 1972. – № 4. – С. 65–70.

9. Буземан Н. З. Выпуклые поверхности. – М.: Наука, 1969.

10. Шор Н. З. Обобщенный градиентный спуск // Труды I Зимней школы по математическому программированию (Дрогобыч). – М.: ЦЭМИ АН СССР, 1969. – Вып. 3. – С. 578–585.

11. Шор Н. З. О скорости сходимости обобщенного градиентного спуска // Кибернетика. – 1968. – № 3. – С. 98–99.

12. Шор Н. З., Щепакин М. Б. Алгоритм решения двухэтапной задачи стохастического программирования // Кибернетика. – 1968. – № 3. – С. 56–58.

13. Шор Н. З., Гамбурд П. Р. Некоторые вопросы сходимости обобщенного градиентного спуска // Кибернетика. – 1971. – № 6. – С. 82–84.

14. Поляк Б. Т. Минимизация негладких функционалов // Журн. вычислительной математики и мат. физики. – 1969. – № 3.

15. Баженов Л. Г. Об условиях сходимости метода минимизации почти-дифференцируемых функций // Кибернетика. – 1972. – № 4.

16. Шепилов М. А. О градиентных и штрафных методах в задачах математического программирования // Автореферат дисс. – М., 1974.

17. Гольштейн Е. Г. Обобщенный градиентный метод отыскания седловых точек // Экономика и мат. методы. – 1972. – **VIII**. – Вып. 4.

18. Михалевич В. С., Ермольев Ю. М., Шкурба В. В., Шор Н. З. Сложные системы и решение экстремальных задач // Кибернетика. – 1967. – № 5. – С. 29–39.

19. Huang H. V. Unified Approach to Quadratically Convergent Algorithms for Function Minimization // Optimization Theory and Applications. – 1970. – **5**. – № 6.

20. Шор Н. З. Методы минимизации недифференцируемых функций и их приложения // Автореферат диссертации на соискание степени доктора физ.-мат. наук. – К., 1970. – 43 с.

21. Шор Н. З. О скорости сходимости метода обобщенного градиентного спуска с растяжением пространства // Кибернетика. – 1970. – № 2. – С. 80–85.

22. Шор Н. З., Журбенко Н. Г. Метод минимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов // Кибернетика. – 1971. – № 3. – С. 51–59.

23. Шор Н. З., Шабашова Л . П. О решении минимаксных задач методом обобщенного градиентного спуска с растяжением пространства // Кибернетика. – 1972. – № 1. – С. 82–88.

24. Шабашова Л. П. Градиентные методы решения нелинейных минимаксных задач // Диссертация на соискание степени канд. физ.-мат. наук. – Днепропетровск, 1973. – 135 с.

25. Шор Н. З. Использование операций растяжения пространства в задачах минимизации выпуклых функций // Кибернетика. – 1970. – № 1. – С. 6–12.

26. Шор Н. З. Исследование сходимости метода градиентного типа с растяжением пространства в направлении разности двух последовательных субградиентов // Кибернетика. – 1975. – № 4. – С. 48–53.

27. Скоков В. А. Замечание к методам минимизации, использующим операцию растяжения пространства // Кибернетика. – 1974. – № 4.

28. Wolfe P. On Conjugate Subgradient Method for Minimization of Nondifferentiable Functionals // Math. Progr. – 1974. – **7**. – № 3.

29. Lemarechal Cl. Note on an Extension of «Davidon» Methods to Nondifferentiable Functions // Math. Progr. – 1974. – **7**. – № 3.

30. Шор Н. З. Применение обобщенного градиентного спуска в блочном программировании // Кибернетика. – 1967. – № 3. – С. 53–55.

31. Шор Н. З., Иванова Л. В. Об одном итеративном методе решения задач линейного программирования и матричных игр // Теория оптимальных решений. – К., 1969. – № 3. – С. 22–30.

32. Шор Н. З., Росина Н. И. Схема разложения задач линейного и выпуклого программирования и ее применение для решения задач планирования перевозок // Доклады I Всесоюзной конференции по оптимизации и моделированию транспортных сетей. – К., 1967. – С. 225–237.

33. Еремин И. И. О методе «штрафов» в выпуклом программировании // Кибернетика. – 1966. – № 4.

34. Кюнци Г. П., Крелле В. Нелинейное программирование. – М.: Сов. радио, 1965.

Новые направления в развитии методов негладкой оптимизации

Н. З. Шор

Кибернетика. – 1977. – № 6. – С. 87–91.

За последние два – три года в области негладкой оптимизации произошли довольно заметные изменения, вызванные повышенным интересом к ней со стороны широкого круга «оптимизаторов» – как прикладников, так и теоретиков.

Появились новые направления – это, в первую очередь, ε-субградиентные методы, различным модификациям которых посвящено уже довольно много работ (см., например, [2]). На новой идейной основе возродился интерес к схемам с последовательным отсечением. В то же время накопился большой экспериментальный материал по использованию семейства обобщенных градиентных методов как простейших (метод обобщенного градиентного спуска), так и методов с растяжением пространства. Все это требует определенного осмысливания, чтобы понять, что же на самом деле достигнуто и в каком направлении надо двигаться дальше.

Несмотря на то, что при построении и обосновании процессов релаксационного типа [3, 4] при определенных предположениях гладкости и регулярности получены значительные результаты, имеется еще большая группа важных в практическом отношении задач, которые с трудом поддаются решению алгоритмами этого типа. Главным образом, это задачи линейного и нелинейного программирования высокой размерности. Как показывает вычислительная практика, применение такого испытанного средства, как симплекс-метод, при решении задач линейного программирования высокой размерности редко приводит к успеху, что связано с резким возрастанием числа итераций и накоплением ошибок округления. Использование классических методов декомпозиции (например, метода Данцига – Вулфа) оказалось на практике малоэффективным. В то же время использование схем декомпозиции для задач блочной структуры приводит к задачам минимизации, как правило, функций с разрывным градиентом от связывающих переменных или от множителей Лагранжа, соответствующих связывающим ограничениям.

К задачам минимизации функций с разрывным градиентом сводятся многие модели игрового характера, «многокритериальные» модели оптимального планирования и проектирования, когда в качестве целевой функции выступает функция максимума (как правило, негладкая).

Большой интерес представляет метод негладких штрафных функций. Дело в том, что негладкие штрафные функции определенного вида обладают несомненным преимуществом по сравнению с обычно применяемыми гладкими функциями штрафа: при использовании негладких штрафных функций, как правило, нет необходимости устремлять штрафные коэффициенты к ∞ и можно получить задачу, в точности эквивалентную первоначальной задаче нелинейного программирования, при определенных конечных значениях штрафных коэффициентов [5, 6].

Необходимость в минимизации негладких функций, состоящих из астрономического числа гладких «кусков», возникает и при решении задач дискретного программирования, когда мы пытаемся в методе ветвей и границ получить оценки границ путем перехода к двойственной задаче [2]. Функции с разрывным градиентом могут также непосредственно входить в модель задачи оптимального планирования, проектирования или исследования операций как результат кусочно-гладкой аппроксимации технико-экономических характеристик реальных объектов.

Таким образом, сфера применения методов негладкой оптимизации огромна. Создание эффективных методов негладкой оптимизации является ключом к решению многих вычислительных проблем математического программирования, в особенности для задач высокой размерности.

Ниже приводится краткий сравнительный анализ алгоритмов решения достаточно общих классов задач минимизации функций с разрывным градиентом, не требующих для своей реализации априорного задания конкретной структуры минимизируемой функции. Предоставляется лишь возможность вычисления значений функции и ее градиентов (или их аналогов в случае недифференцируемости – определенных представителей обобщенных градиентных множеств) в произвольной точке.

Если исключить такие универсальные средства, как методы систематического перебора и случайного поиска, то можно выделить три основных класса специфических методов негладкой оптимизации, для реализации которых используется вычисление обобщенных градиентов. Наиболее развитым из них является класс немонотонных обобщенных градиентных методов, которые начиная с 1961 года активно разрабатываются и исследуются в Институте кибернетики АН УССР [1]. Наиболее простым и общим является метод обобщенного градиентного спуска

(ОГС). Его можно применять как для минимизации выпуклых функций, так и для получения локальных экстремумов в случае более общего класса локально липшицевых (почти-дифференцируемых) функций [7]. Он не требует дополнительной памяти, однако недостатком его является медленная сходимость и не всегда хорошо контролируемая точность решения. Из алгоритмов с растяжением пространства наиболее хорошо экспериментально исследованы методы с растяжением пространства в направлении разности двух последовательных градиентов (r-алгоритмы) [9, 10]. Если размерность пространства, в котором происходит минимизация, не превышает 200–300, целесообразность использования r-алгоритмов вместо ОГС в большинстве случаев не вызывает сомнений. При бóльших размерностях солидная доля вычислительной работы уходит на преобразование матрицы растяжения пространства. Трудоемкость одной итерации в r-алгоритме может превысить в несколько раз трудоемкость одной итерации в методе ОГС, и здесь целесообразность применения r-алгоритма тем больше, чем меньше относительная доля времени, идущего на преобразование матрицы, и чем большая точность требуется от результата. В каждом конкретном случае нужно оценить время решения. При этом можно пользоваться следующим эмпирическим правилом для r-алгоритмов: относительная точность решения по функционалу увеличивается на порядок через каждые $n - 1,5n$ итераций (n – размерность минимизируемой функции). Что касается метода ОГСРП [1], то его, по-видимому, стоит применять, только когда известно значение функции в точке минимума. В противном случае r-алгоритм является более эффективным.

Для решения задач стохастического программирования с успехом применяются стохастические аналоги метода ОГС [11].

В США и Западной Европе градиентными методами минимизации негладких функций серьезно начали заниматься примерно с 1973 г. сначала в связи с приложениями в области дискретного программирования [12], а затем – в целом для решения задач большой размерности [13]. Об исследовании работ по этой проблеме на Западе достаточно полное представление дает сборник статей [2]. С тех пор интерес и количество печатных работ в этом направлении растут лавинообразно. Особенно интенсивно развивается направление так называемой ε-субградиентной оптимизации, по идее близкое, с одной стороны, к алгоритмам В. Ф. Демьянова решения минимаксных задач [14], т. е. к методам ε-наискорейшего спуска, а с другой (особенно в формальном отношении), – к алгоритмам метода сопряженных градиентов (или «давидоновского» типа) [15, 16].

К настоящему времени разработаны лишь монотонные варианты ε-субградиентных методов, которые по количеству вычислений на одной итерации в какой-то степени являются промежуточными между методом ОГС и методами с растяжением пространства. К сожалению, эти варианты еще очень недостаточно исследованы экспериментально, а теоретически обоснованная скорость сходимости (точность порядка $1/k$ за k^3 итераций [16] не дает оснований к оптимизму. Проведенные нами численные эксперименты показали значительное преимущество r-алгоритмов по сравнению с ε-субградиентными методами.

И, наконец, в связи с исследованием оптимальных алгоритмов минимизации выпуклых функций заново возродился интерес к схемам последовательных отсечений опорными гиперплоскостями. Среди алгоритмов этого типа два были наиболее известны.

1. Метод секущих Келли [17]. Основан на кусочно-линейной аппроксимации графика выпуклой функции с использованием опорных гиперплоскостей. На каждом шаге этого метода приходится решать задачу линейного программирования, причем от шага к шагу число ограничений задачи может возрастать.

2. Метод центрированных сечений А. Ю. Левина. Основан на проведении опорных гиперплоскостей к поверхности уровня выпуклой функции, проходящих через центр тяжести области локализации оптимума [18]. Последовательные процедуры отсечения в методе центрированных сечений оказываются весьма эффективными, однако сложность вычисления центра тяжести в многомерном случае сводит на нет практическую ценность метода (см. [19]).

В последнее время обнаружились очень интересные связи между алгоритмами последовательных отсечений и алгоритмами с растяжением пространства [19, 20]. Приведем некоторые дополнительные результаты в этом направлении. В работе [20] показано, что одна из модификаций метода обобщенного градиентного спуска с растяжением пространства в направлении градиента (ОГСРП) фактически реализует модифицированный метод центрированных сечений (ММ ЦС), рассмотренный в [19] применительно к задаче выпуклого программирования, и позволяет последовательно строить n-мерные эллипсоиды Φ_k, $k = 0, 1, 2, \ldots$, внутри которых локализован оптимум, причем объемы $v_0, v_1, \ldots, v_k, \ldots$ этих эллипсоидов представляют собой убывающую геометрическую прогрессию со знаменателем

$$q_n = \sqrt{\frac{n-1}{n+1}} \left(\frac{n}{\sqrt{n^2-1}} \right)^n .$$

Этот результат может обобщаться и уточняться по нескольким направлениям.

1. Из анализа доказательства основной леммы [20] видно, что в нем, по существу, используется только следующее свойство субградиентного поля направлений $g(x)$:

$$(g(x), x - x^*) \geq 0, \tag{1}$$

где x^* – искомая оптимальная точка.

Поэтому все, что было сказано в [20] по поводу решения задачи выпуклого программирования, можно перенести на более общий случай векторного поля, обладающего свойством (1).

Итак, пусть имеются векторное поле $g(x)$, определенное на E_n, и точка $x^* \in E_n$, лежащая на расстоянии, не превышающем R от заданной точки x_0, для которой справедливо соотношение (1) при любом $x \in E_n$ (причем $g(x) \neq 0$ при $x \neq x^*$).

При $n > 1$ рассмотрим следующий итеративный алгоритм.

Перед первым шагом имеем $x_0 \in E_n$; $B_0 = I_n$ – единичная матрица, $h_0 = R/(n+1)$. Пусть проделано k шагов и получено $x_k \in E_n$, B_k – матрицу размера $n \times n$ и h_k.

Переходим к $(k+1)$-му шагу, на котором вычисляем:

1) $g(x_k)$;

2) $\xi_k = B_k^* g(x_k) / \|B_k^* g(x_k)\|$;

3) $x_{k+1} = x_k - h_k B_k \xi_k$;

4) $B_{k+1} = B_k \cdot R_\beta(\xi_k)$, $\quad \beta = \sqrt{(n-1)/(n+1)}$; $\quad R_\beta(\xi_k)$ – оператор растяжения пространства, $R_\beta(\xi_k)x = x + (\beta - 1)(x, \xi)\xi$ [1];

5) $h_{k+1} = h_k \cdot r$, $\quad r = n/\sqrt{n^2 - 1}$.

Справедливо утверждение следующей теоремы.

Теорема 1. *Последовательность $\{x_k\}_{k=0}^{\infty}$, генерируемая вышеприведенным алгоритмом, удовлетворяет неравенствам:*

$$\|A_k(x_k - x^*)\| \leq h_k(n+1); \quad A_k = B_k^{-1}.$$

Доказательство теоремы 1 практически не отличается от доказательства основной леммы в [20].

Применим теорему 1 к определению седловой точки выпукло-вогнутой функции. Пусть задана выпукло-вогнутая функция двух векторных переменных

$$f(x, y), \quad x \in E_n, \quad y \in E_m, \quad z = \{x, y\} \in E_n \times E_m \equiv E_{n+m}.$$

Пусть z^* – седловая точка этой функции, z_0 – заданное начальное приближение и априори известно, что $\|z_0 - z^*\| \le R$.

Рассмотрим псевдоградиентное множество

$$G(z) = G_f^x(x, y) \times (-G_f^y(x, y)),$$

где $G_f^x(x, y)$ – множество частных субградиентов функции $f(x, y)$, которая рассматривается как функция от x при фиксированном y; $-G_f^y(x, y)$ – множество субградиентов от функции $-f(x, y)$ по y при фиксированном x. Сформируем векторное поле $g(z)$ следующим образом:

$$g(z) = \{g_f^x(z), -g_f^y(z)\}; \quad g_f^x(z) \in G_f^x(z); \quad g_f^y(z) \in G_f^y(z).$$

Покажем, что $(g(z), z - z^*) \ge 0$. Из определения седловой точки следует

$$f(x, y^*) \ge f(x^*, y^*) \ge f(x^*, y).$$

Далее,

$$0 \le f(x, y^*) - f(x^*, y) = f(x, y^*) - f(x, y) + f(x, y) - f(x^*, y) \le$$

$$\le (g_f^x(z), x - x^*) - (g_f^y(z), y - y^*) = (g(z), z - z^*).$$

Таким образом, для локализации седловой точки z^* можно применить алгоритм 1)–5).

2. Для задачи безусловной минимизации выпуклой функции $f(x)$, определенной на E_n, хотелось бы перейти от оценки скорости сходимости алгоритма, выраженной в терминах уменьшения объема области локализации минимума, к оценке скорости сходимости по функционалу. Для этого докажем следующий факт, относящийся к классу алгоритмов с растяжением пространства в направлении градиента (ОГСРП) [1].

Теорема 2. *Пусть в процессе применения алгоритма ОГСРП выполняются следующие условия:*

1) $\|g(x_k)\| \le d, \quad k = 0, 1, \ldots;$

2) $1 < \alpha = \alpha_k, \quad \beta_k = 1/\alpha_k$, *и пусть* $\widetilde{g}(x_k) = B_k^* g(x_k), \quad A_k = B_k^{-1}$.

Тогда $\min\limits_{1 \le r \le k} \|g_r(x_r)\| \le \dfrac{d\sqrt{k(\alpha^2 - 1)}}{\sqrt{\alpha^{2k/n} - 1}}$, $k = 1, 2, \ldots$

Доказательство. Запишем матрицу A_k в виде произведения ортогональной матрицы O_k и симметричной положительно определенной матрицы $S_k : A_k = O_k \cdot S_k$. Так как $A_k = R_\alpha(\xi_k) \ldots R_\alpha(\xi_1)$, то произведение собственных чисел матрицы S_k равно

$$\det S_k = \det A_k = \alpha^k.$$

Пусть $\{\lambda_i^{(k)}\}_{i=1}^n$ – последовательность собственных чисел оператора S_k, записанных в порядке неубывания, $\{e_i^{(k)}\}_{i=1}^n$ – соответствующая ортонормированная система собственных векторов

$$S_k e_i^{(k)} = \lambda_i^{(k)} e_i^{(k)}; \quad \prod_{i=1}^n \lambda_i^{(k)} = \alpha^k; \quad \lambda_n^{(k)} \geq \alpha^{\frac{k}{n}}. \tag{2}$$

Пусть $O_k e_i^{(k)} = \overline{e}_i^{(k)}$. Система векторов $\{\overline{e}_i^{(k)}\}_{i=1}^n$ также является ортонормированной ввиду ортогональности O_k. Запишем $g_f(x_k)$ в следующем виде:

$$g_f(x_k) = \sum_{i=1}^n g_i^{(k)} e_i^{(k)}; \quad \sum_{i=1}^n (g_i^{(k)})^2 \leq d^2.$$

Далее,

$$\xi_{k+1} = \frac{\widetilde{g}_k}{\|\widetilde{g}_k\|} = \sum_{i=1}^n \mu_i^{(k)} \overline{e}_i^{(k)}, \text{ где } \mu_i^{(k)} = \frac{g_i^{(k)}}{\lambda_i^{(k)} \|\widetilde{g}_k\|}.$$

Доказательство проведем методом от противного.
Пусть для некоторого k

$$v_k = \min \|\widetilde{g}_r\| > d \cdot \frac{\sqrt{k(\alpha^2 - 1)}}{\sqrt{\alpha^{2k/n} - 1}}. \tag{3}$$

Тогда для $r \leq k$, $1 \leq i \leq n$ из (3) получаем:

$$|\mu_i^{(r)}| \leq \frac{|g_i^{(r)}| \sqrt{\alpha^{2k/n} - 1}}{\lambda_i^{(r)} d \sqrt{k(\alpha^2 - 1)}}. \tag{4}$$

Используя это соотношение, оценим при r, $1 \leq r \leq k-1$, разность $(\lambda_n^{(r+1)})^2 - (\lambda_n^{(r)})^2$ $\quad(\lambda_n^{(r)}$ – максимальное собственное число оператора $S_r)$. Рассмотрим произвольный вектор a, $\|a\| = 1$:

$$a = \sum_{i=1}^n a_i e_i^{(r)}; \quad \sum_{i=1}^n a_i^2 = 1; \quad A_r a = \sum_{i=1}^n a_i \lambda_i^{(r)} \overline{e}_i^{(r)}; \tag{5}$$

$$\lambda_n^{(r+1)} = \max_{\|a\|=1} \|A_{r+1} a\| = \max_{\|a\|=1} \|R_\alpha(\xi_{r+1}) A_r a\|. \tag{6}$$

Воспользуемся формулой (см. [1]):

$$\|R_\alpha(\xi) x\| = \sqrt{\|x\|^2 + (\alpha^2 - 1)(x, \xi)^2}.$$

Приняв $x = A_r a$, из (4)–(6) получаем:

$$\left[\lambda_n^{(r+1)}\right]^2 = \max_{\|a\|=1}\left[\sum_{i=1}^n a_i^2(\lambda_i^{(r)})^2 + (\alpha^2-1)\left(\sum_{i=1}^n a_i\lambda_i^{(r)}\mu_i^{(r)}\right)^2\right] \leq$$

$$\leq \max_{\|a\|=1}\left[(\lambda_n^{(r)})^2+(\alpha^2-1)\left(\sum_{i=1}^n \left|a_i g_i^{(r)}\right|\right)^2 \frac{(\alpha^{\frac{2k}{n}}-1)}{kd^2(\alpha^2-1)}\right] \leq \left[\lambda_n^{(r)}\right]^2+\frac{\alpha^{\frac{2k}{n}}-1}{k};$$

$$\left[\lambda_n^{(r+1)}\right]^2 < \left[\lambda_n^{(r)}\right]^2 + \frac{\alpha^{\frac{2k}{n}}-1}{k}; \quad r<k.$$

Отсюда $\left[\lambda_n^{(k)}\right]^2 < 1 + (\alpha^{\frac{2k}{n}} - 1) = \alpha^{\frac{2k}{n}}$. Это противоречит $\lambda_n^{(k)} \geq \alpha^{\frac{k}{n}}$. Теорема доказана. Как следствие получаем следующий результат.

Теорема 3. *В предположениях теоремы 2 при использовании описанного нами алгоритма* 1)–5) *и* $\|x_0 - x^*\| \leq R$

$$\rho_k = \min_{0\leq r\leq k}\left[f(x_r)-f(x^*)\right] \leq \frac{Rd_k\sqrt{k(\alpha^2-1)}}{\sqrt{1-\beta^{2k/n}}}\cdot(q_n)^{k/n}, \quad k=1,2,\ldots,$$

где $d_k = \max_{0\leq r\leq k}\|g(x_r)\|$.

3. Так как для достаточно больших n $q_n \approx 1 - 1/(2n)$, то теорема 3 гарантирует скорость сходимости «рекордов» отклонений функционалов от оптимального значения, соответствующую скорости сходимости геометрической прогрессии со знаменателем $s \approx 1 - 1/(2n^2)$. В то же время, как показано в [19], оптимальные алгоритмы должны обеспечивать такую же скорость сходимости по функционалу, как геометрическая прогрессия со знаменателем $1 - c/n$, однако известные алгоритмы, которые обеспечивают соответствующую скорость сходимости по отношению к числу вычислений субградиентов, требуют слишком большого числа дополнительных операций на одной итерации [19]. Таким образом, перед нами стоит весьма интересная и актуальная задача: построить эффективные с вычислительной точки зрения алгоритмы, которые для достаточно широкого класса выпуклых функций сходятся по функционалу со скоростью геометрической прогрессии, знаменатель которой имеет зависимость от размерности в форме $q = 1 - c/n$, $c > 0$. Можем надеяться, что подобные алгоритмы удастся получить, комбинируя схемы отсечения с преобразованием пространства.

Большой интерес представляет проблема построения немонотонных вариантов ε-субградиентных процессов, обладающих существенно ускоренной сходимостью по сравнению с методами ОГС и не требующих значительной дополнительной памяти. Решение этих проблем даст дополнительный стимул для использования методов минимизации негладких функций при решении сложных задач математического программирования.

Литература

1. Шор Н. З. Обобщенные градиентные методы минимизации негладких функций и их применение к задачам математического программирования // Экономика и матем. методы. – 1976. – Вып. 2. – № 12. – С. 337–356.

2. Mathematical Programming. Study 3, Nondifferentiable Optimization. (Edited by M. L. Balinski and P. Wolfe). – North–Holland Publishing Company. – Amsterdam, 1975.

3. Зойтендейк Г. Методы возможных направлений. – М.: ИЛ, 1963.

4. Пшеничный Б. Н., Данилин Ю. М. Численные методы в экстремальных задачах. – М.: Наука, 1975.

5. Еремин И. И. О методе штрафов в выпуклом программировании // Кибернетика. – 1967. – № 4.

6. Bertsekas D. P. Necessary and Sufficient Conditions for a Penalty Method to be Exact // Math. Program. – 1975. – V. 9. – № 1.

7. Баженов Л. И. Об условиях сходимости методов минимизации почти-дифференцируемых функций // Кибернетика. – 1972. – № 4.

8. Шор Н. З., Журбенко Н. Г. Метод минимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов // Кибернетика. – 1971. – № 3. – С. 51–59.

9. Шор Н. З., Шабашова Л. П. О решении минимаксных задач методом обобщенного градиентного спуска с растяжением пространства // Кибернетика. – 1972. – № 1. – С. 82–88.

10. Овруцкий И. Г., Шор Н. З. Применение методов минимизации негладких функций для решения задач интерпретации гравиметрических наблюдений // Кибернетика. – 1976. – № 2. – С. 57–64.

11. Ермольев Ю. М. Методы стохастического программирования. – М.: Наука, 1976.

12. Held M., Karp R. M. The Traveling Salesman Problem and Minimum Spanning Trees: Part II // Math. Program. – 1973. – V. 4. – № 2.

13. Held M., Wolfe P., Crowder H. P. Validation of Subgradient Optimization. // Math. Program. – 1974. – V. 6. – № 1.

14. Демьянов В. Ф., Малоземов В. Н. Введение в минимакс. – М.: Наука, 1972.

15. Lemaréchal C. An Etension of Davidon Methods to Nondifferentiable Problems // Math. Progr. Study 3, 1975.

16. Wolfe P. A Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions // Math. Progr. Study 3, 1975.

17. Kelley G. E. The «Cutting Plane» Method for Solving Convex Programs // SIAM Journal, 1960. – V. 8. – № 4.

18. Левин А. Ю. Об одном алгоритме минимизации выпуклых функций // Докл. АН СССР. – М., 1965. – Т. 160. – № 6.

19. Юдин Д. Б., Немировский А. С. Информационная сложность и эффективные методы решения выпуклых экстремальных задач // Экономика и мат. методы. – 1976. – Вып. 2. – № 12. – С. 337–356.

20. Шор Н. З. Методы отсечения с растяжением пространства для решения задач выпуклого программирования // Кибернетика. – 1977. – № 1.

Метод растяжения пространства для ускорения сходимости в задачах овражного типа

Н. З. Шор, В. И. Билецкий
Теория оптимальных решений. – 1969. – № 2. – С. 3–16.

1. Введение

Мы будем рассматривать задачу минимизации выпуклой функции $f(x)$ в n-мерном евклидовом пространстве E_n. От функции $f(x)$ мы не будем требовать непрерывности производных по направлению, поэтому вместо поля градиентов нам придется рассматривать его аналог – поле обобщенных градиентов [2], т. е. совокупность векторов $\widehat{g}\,(x_0)$, обладающих следующим свойством:

$$f(x) - f\,(x_0) \geq (\widehat{g}\,(x_0)\,, x - x_0).$$

Пусть единственный минимум функции $f(x)$ достигается в точке $x^* = 0$ и $f(x^*) = 0$. Если $f(x) \not\equiv 0$, то поверхности $f(x) = c > 0$ являются замкнутыми и выпуклыми. Для любого $c > 0$ и α, $|\alpha| = 1$ найдутся положительные $\overline{t}_c^{\,\alpha}$ и $\underline{t}_c^{\,\alpha}$ такие, что

$$f\left(\overline{t}_c^{\,\alpha}\alpha\right) = f\left(-\underline{t}_c^{\,\alpha}\alpha\right) = c.$$

Обозначим $\overline{t}_c^{\,\alpha} + \underline{t}_c^{\,\alpha}$ через t_c^α. $\displaystyle\max_{|\alpha_1|=|\alpha_2|=1} \frac{t_c^{\alpha_1}}{t_c^{\alpha_2}} = k(c)$ назовем локальным коэффициентом овражности, $k = \displaystyle\lim_{\varepsilon\to+0}\sup_{c\leq\varepsilon} k(c)$ – коэффициент овражности. Если $k = +\infty$, то функция $f(x)$ называется существенно овражной.

Из работы [2] ясно, что скорость сходимости обобщенного градиентного спуска, если вся траектория спуска проходит в области $f(x) \leq c$ может быть линейной с показателем $\sqrt{1 - [\,\sup_{r\leq c} k(r)\,]^2}/\sup_{r\leq c} k(r)$, если $\sup_{r\leq c} k(r) < \infty$. Если функция $f(x)$ является существенно овражной, то обычный обобщенный градиентный спуск не гарантирует даже линейной скорости сходимости с показателем, произвольно близким к единице.

В то же время существенно овражные функции не являются в каком-то смысле патологическими, практически редко встречающимися. Особенно естественно они появляются в задачах минимаксного типа. Как показывает пример $f(x) = \max\left[x^2 + y^2, (x-1)^2 + y^2\right]$, в этих задачах существенная овражность является скорее не исключением, а правилом.

В связи с этим получение высокой точности минимизации в задачах минимаксного типа представляет значительные трудности.

Методы решения задач овражного типа, описанные в литературе [1], обычно основываются на следующей идее: рассматривая некоторый отрезок минимизируемой последовательности, нужно определить направление «вдоль оврага» и в этом направлении двигаться с большим шагом.

Предлагаемый ниже метод основан на другой идее, суть которой сводится к тому, чтобы сделать преобразования пространства E_n, направленные на уменьшение локальных коэффициентов овражности.

Плохая сходимость градиентных методов для овражного типа обусловлена тем, что направление градиента (обобщенного градиента) обычно образует угол, близкий к $\frac{\pi}{2}$ с направлением «русла» оврага. Поэтому мы приняли в качестве подходящей процедуры преобразование пространства E_n, растяжение этого пространства в направлении обобщенного градиента функции $f(x)$, взятого в рассматриваемой точке.

2. Оператор растяжения пространства

Пусть задан вектор $\xi \in E_n, \|\xi\| = 1$ и число $\alpha > 0$. Запишем разложение произвольного вектора $X \in E_n$ в следующем виде:

$$X = \gamma_\xi(x) \cdot \xi + d_\xi(x),$$

причем $(\xi, d_\xi(x)) = 0$. Тогда оператором растяжения в направлении ξ с коэффициентом α назовем оператор $R_\alpha(\xi)$, действующий следующим образом:

$$R_\alpha(\xi)X = \alpha\gamma_\xi(x)\xi + d_\xi(x).$$

Легко видеть, что $R_{\alpha\beta}(\xi) = R_\alpha(\xi)R_\beta(\xi)$. В частности,

$$R_\alpha(\xi)R_{1/\alpha}(\xi) = R_1(\xi) = E.$$

Пусть координаты вектора ξ в некоторой ортонормированной системе координат, определяемой ортами $e_1, \ldots, e_n$, равны $\xi_1, \ldots, \xi_n$. Тогда в этой системе координат преобразованию $R_\alpha(\xi)$ соответствует матрица $B_\alpha(\xi)$ с элементами b_{ij}, определяемыми по следующим формулам:

$$b_{ij} = \begin{cases} (\alpha-1)\xi_i\,\xi_j & \text{для} \quad i \neq j; \\ 1+(\alpha-1)\xi_i^2 & \text{для} \quad i = j. \end{cases}$$

Заметим, что применение матрицы $B_\alpha(\xi)$ к произвольному вектору $x = (x_1, \ldots, x_n)$ требует всего $3n$ операций умножения и $\approx 3n$ операций сложения и вычитания. Умножение матрицы на матрицу $B_\alpha(\xi)$ требует порядка cn^2 операций умножения, сложения и вычитания.

Пусть задана функция $f(x)$ и задано линейное преобразование $y = Ax$. Рассмотрим f как функцию от y: $x = A^{-1}y$; $f(x) = f(A^{-1}y)$. Пусть $\widehat{g}(x_0)$ – обобщенный градиент функции $f(x)$ в точке $x = x_0$, т. е.

$$f(x) - f(x_0) \geq (\widehat{g}(x_0), x - x_0),$$

или

$$f(A^{-1}y) - f(A^{-1}y_0) \geq \left(\widehat{g}(x_0), A^{-1}(y - y_0)\right) = \left((A^*)^{-1}\widehat{g}(x_0), y - y_0\right).$$

Отсюда, для $\varphi(y) = f(A^{-1}y)$,

$$\varphi(y) \cdot \varphi(y_0) \geq \left((A^*)^{-1}\widehat{g}(x_0), y - y_0\right). \qquad (1)$$

Таким образом, обобщенный градиент к функции $\varphi(y)$ в точке $y = y_0$ равен $(A^*)^{-1}\,\widehat{g}(x_0)$, где $x_0 = A^{-1}y_0$. Формула (1) дает нам возможность вычислить обобщенный градиент функции, полученной в результате линейного преобразования аргументов, не преобразуя вида исходной функции. Заметим, что если $A = B_k B_{k-1} \ldots B_1$, то $A^{-1} = B_1^{-1} \ldots B_{k-1}^{-1} B_k^{-1}$.

Таким образом, мы будем рассматривать класс алгоритмов следующего вида:

0-й шаг. Задаемся начальным приближением x_0 и матрицей $A_0 = A_0^{-1} = E$; $f(x) = f_0(x)$.

k-й шаг. Вычислены x_{k-1}, A_{k-1}^{-1}. Находим $\widehat{g}_{f_0}(x_{k-1})$, $\widehat{g}_{f_k}(A_{k-1}x) = (A^*)_{k-1}^{-1}\,\widehat{g}_f(x_{k-1}) = \eta_k$ (здесь $f_k(z) = f(A_{k-1}^{-1}z)$). Затем вычисляем:

$$\begin{aligned} \xi_k &= \eta_k/||\eta_k||; \\ x_k &= x_{k-1} - A_{k-1}^{-1} h_k \xi_k; \\ B_k^{-1} &= B_{\alpha_k}^{-1}(\xi_k) \;=\; B_{1/\alpha_k}(\xi_k); \\ A_k^{-1} &= A_{k-1}^{-1} \cdot B_{\alpha_k}^{-1}(\xi_k) \;=\; A_{k-1}^{-1} \cdot B_{1/\alpha_k}(\xi_k); \end{aligned}$$

h_k и α_k – положительные константы.

Различные модификации алгоритма будут отличаться друг от друга способом выбора последовательности h_k и α_k. Вообще говоря, можно рассматривать более общий класс алгоритмов, в котором оператор последовательности B_k может отличаться от оператора растяжения, но в данной статье мы ограничимся только рассмотрением операторов растяжения.

3. Некоторые условия сходимости метода растяжения пространства

Для вывода условий сходимости метода растяжения пространства введем некоторые определения.

Определение 1. *Назовем R-оператором оператор, являющийся произведением операторов растяжения пространства, т. е. оператор следующего вида:*

$$A = R_{\alpha_m}(\xi_m) \cdot R_{\alpha_{m-1}}(\xi_{m-1}) \cdots R_{\alpha_1}(\xi_1).$$

R^+-оператор – это R-оператор, у которого $\alpha_i \geq 1,\ i = 1, \ldots, m$.
R^--оператор – это R-оператор, у которого $\alpha_i \leq 1,\ i = 1, \ldots, m$.

Определение 2. *Коэффициентом A – расширения вектора $x \neq 0$ будем называть следующее число:*

$$K_A(x) = \frac{\|Ax\|}{\|x\|}.$$

Лемма 1. *Если A – R^+-оператор (R^--оператор), то для всех $x \neq 0$ $K_A(x) \geq 1$ $(K_A(x) \leq 1)$.*

Доказательство. Пусть $\alpha \geq 1$.

$$\begin{aligned} \|R_\alpha(\xi)\, x\| &= \sqrt{\|\alpha\gamma_\xi(x)x\|^2 + \|d_\xi(x)\|^2} \geq \\ &\geq \sqrt{\|\gamma_\xi(x)x\|^2 + \|d_\xi(x)\|^2} = \|x\|. \end{aligned}$$

Отсюда, $\dfrac{\|R_\alpha(\xi)\, x\|}{\|x\|} \geq 1$.

Если A – R^+-оператор, то $A = R_{\alpha_m}(\xi_m)\dots R_{\alpha_1}(\xi_1)$, причем $\alpha_i \geq 1$, $i = 1,\dots,m$.

$$\frac{\|Ax\|}{\|x\|} = \prod_{k=1}^{m} \frac{\|A_k x\|}{\|A_{k-1}x\|},$$

где $A_k = R_{\alpha_k}(\xi_k)\cdots R_{\alpha_1}(\xi_1)\,, \quad k = 1,\dots,m; \quad A_0 = E.$

Так как $\dfrac{\|A_k x\|}{\|A_{k-1}x\|} = \dfrac{\|R_{\alpha_k}(\xi_k)\cdot A_{k-1}x\|}{\|A_{k-1}x\|} \geq 1$, то $\dfrac{\|Ax\|}{\|x\|} \geq 1.$

Для R^--операторов доказательство леммы аналогичное.

Лемма 2. *Пусть A – произвольный линейный оператор, $z = \alpha x + \beta y$, причем x и y – линейно независимы, $\|x\| = \|y\| = 1$, $|\alpha| + |\beta| = 1$. Тогда*

$$\frac{\|Az\|}{\|z\|} \leq \frac{\sqrt{2}\max(K_A(x), K_A(y))}{\sqrt{1 - |(x,y)|}}.$$

Доказательство.

$$\begin{aligned} \|Az\| &= \|\alpha Ax + \beta Ay\| \leq |\alpha|\, K_A(x)\|x\| + |\beta| K_A(y)\,\|y\| \leq \\ &\leq (|\alpha| + |\beta|)\max(K_A(x), K_A(y)) = \max(K_A(x), K_A(y)), \\ \|z\| &= \|\alpha x + \beta y\| = \sqrt{\alpha^2 + \beta^2 + 2\,\alpha\,\beta\,(x,y)}. \end{aligned}$$

Найдем $\min(\alpha^2 + \beta^2 + 2\,\alpha\,\beta\,(x,y))$ при условии $|\alpha| + |\beta| = 1$. Очевидно, что минимум будет достигнут при таких значениях α и β, при которых $\alpha\,\beta\,(x,y) \leq 0$. Поэтому нужно минимизировать $|\alpha|^2 + |\beta|^2 - 2\,|\alpha|\,|\beta|\;|(x,y)|$ при условии $|\alpha| + |\beta| = 1$. Несложный подсчет показывает, что минимум достигается при $|\alpha| = |\beta| = \dfrac{1}{2}$ и равен $\dfrac{1 - |(x,y)|}{2}$. Отсюда, $\|z\| \geq \dfrac{\sqrt{1 - |(x,y)|}}{\sqrt{2}}$ и $K_A(z) = \dfrac{\|Az\|}{\|z\|} \leq \dfrac{\sqrt{2}\max(K_A(x), K_A(y))}{\sqrt{1 - |(x,y)|}}$.

Определение 3. *Пусть задано множество $\Sigma = \{A_\gamma\}$ линейных операторов. Вектор $x \neq 0$ назовем Σ-особым, если*

$$\sup_{A_\gamma \in \Sigma} K_{A_\gamma}(x) < \infty.$$

Лемма 3. *Совокупность Σ-особых векторов и вектор 0 образуют линейное подпространство пространства E_n, которое мы будем обозначать E_Σ.*

Доказательство. Обозначим множество Σ-особых векторов через M_Σ. Если $x \in M_\Sigma$, то и $a\,x \in M_\Sigma$, $(a \neq 0)$. Если $x \in M_\Sigma$ и $y \in M_\Sigma$ и $x = by$ при любом b, то $x + y \in M_\Sigma$, что видно из леммы 2. Отсюда следует утверждение леммы 3.

Определение 4. *Назовем коэффициентом объемного расширения подпространства E_k под действием оператора A отношение неориентированного объема множества $\{y : Ax = y;\ x \in S_k\}$ к объему S_k, где S_k – сфера единичного радиуса, соответствующего подпространству. Этот коэффициент обозначим $V_A\,(E_k)$.*

Лемма 4. *Коэффициент объемного расширения подпространства E_k под действием оператора $R_\alpha\,(\xi)$, $\|\xi\| = 1$ равен $\sqrt{1 + (\alpha^2 - 1)\,\|\psi\|^2}$, где ψ – проекция вектора ξ на подпространство E_k.*

Доказательство леммы следует из очевидных геометрических соображений.

Теорема 1. *Пусть выпуклая функция $f(x)$ достигает минимума в точке* 0 *и при применении алгоритма растяжения пространства (см. стр.* 52*)* $\alpha_k \geq 1+\varepsilon$, $k = 1, 2, \ldots$ *и последовательность $\{x_k\}$ обладает следующим свойством: $\|A_k x_k\| \leq c$, $k = 0, 1, \ldots$, где c – некоторая константа. Тогда $\lim\limits_{k\to\infty} f\,(x_k) = f(0)$.*

Доказательство. Из того, что $\|A_k x_k\| \leq c$ следует, что $\|x_k\| \leq c$. Следовательно, эта последовательность имеет точки сгущения. Пусть Σ – множество операторов $\{A_k\}$. E_Σ – подпространство особых векторов множества Σ. Коэффициент объемного расширения подпространства E_Σ ограничен. Отсюда вытекает в силу леммы 4 и свойства $\alpha_k \geq 1 + \varepsilon$, что проекция вектора ξ_k на подпространство $A_k E_\Sigma$ стремится к нулю, т. е. производные по направлению, лежащему в подпространстве $A_k E_\Sigma$ от функции $\varphi\,(y) = f\left(A^{-1} x\right)$ стремятся к нулю.

С другой стороны, все точки сгущения последовательности $\{x_k\}$ должны лежать в подпространстве E_Σ, т.к. в противном случае не выполнялось бы $\|A_k x_k\| \leq c$ для всех x_k. Пусть $\overline{x}$ – точка сгущения, и $f(\overline{x}) \neq f(0)$. $\{x'_i\}$ – некоторая подпоследовательность последовательности $\{x_k\}$, сходящаяся к $\overline{x}$.

Далее, $f\,(x'_i) - f(x'_i - \overline{x}) \leq \varphi'_i \cdot \|\overline{x}\|$, где φ'_i – производная в направлении, определяемом вектором $\overline{x}$ в точке x'_i. Но из сказанного выше, эта производная стремится к 0. Отсюда,

$$f(x'_i) \leq f(x'_i - \overline{x}) + \delta_i, \quad \text{где } \delta_i \to 0.$$

Перейдем к пределу при $i \to \infty$. Получим

$$\lim_{i\to\infty} f(x_i') = f(\overline{x}) \le f(0), \quad \text{т. е. } f(\overline{x}) = f(0).$$

Доказательство теоремы 1 завершено.

Ниже мы рассмотрим способы регулировки шага и выбора последовательности коэффициента расширения $\{\alpha_k\}$, обеспечивающие условия теоремы 1 для широкого класса выпуклых функций $f(x)$. При этом мы будем предполагать, что нам заранее известно значение функции в точке минимума $x = x^*$ и без потери общности будем считать, что $x^* = 0$ и $f(x^*) = 0$.

Теорема 2. *Пусть выпуклая функция $f(x)$ достигает своего минимума в точке $x = 0$ и $f(x) = 0$. Выберем начальную точку x_0 и пусть в окрестности $\|x\| \le \|x_0\|$ $f(x)$ обладает следующим свойством:*

$$Nf(x) \le |f_x'| \cdot \|x\| \le Mf(x), \tag{2}$$

где M, N – некоторые положительные константы, f_x' – обобщенная производная по направлению x в точке x. Тогда, если мы в методе растяжения пространства будем выбирать шаг

$$h_k = \gamma f(x_{k-1}) / \|\eta_k(x_{k-1})\|, \qquad \gamma = 2MN/(M+N)$$

(η_k – значение обобщенного градиента в растянутом пространстве), а коэффициент растяжения $\alpha_k = (M+N)/(M-N)$, то последовательность $\{r_k\} = \{\|A_k x_k\|\}$ будет возрастающей.

Примечание. *Для любой выпуклой функции N можно выбирать не меньше 1.*

Доказательство.

$$\|A_k x_k\|^2 = \|R_\alpha(\xi_k) A_{k-1} x_k\|^2 = \|R_{\alpha_k}(\xi_k) A_{k-1} x_{k-1} - R_\alpha(\xi_k) h_k \xi_k\|^2 =$$

$$= \left[1 + \left(\alpha_k^2 - 1\right) \frac{(A_{k-1}x_{k-1}, \xi_k)^2}{\|A_{k-1}x_{k-1}\|^2}\right] \|A_{k-1}x_{k-1}\|^2 -$$

$$-2(R_{\alpha_k}(\xi_k) A_{k-1}x_{k-1}, \alpha_k h_k \xi_k) + \alpha_k^2 h_k^2 =$$

$$= \|A_{k-1}x_{k-1}\|^2 + \left(\alpha_k^2 - 1\right)(A_{k-1}x_{k-1}, \xi_k)^2 - 2\alpha_k^2 h_k (A_{k-1}x_{k-1}, \xi_k) + \alpha_k^2 h_k^2 =$$

$$= \|A_{k-1}x_{k-1}\|^2 + \alpha_k^2((A_{k-1}x_{k-1}, \xi_k) - h_k)^2 - (A_{k-1}x_{k-1}, \xi_k)^2. \tag{3}$$

Далее, $(A_{k-1}x_{k-1}, \xi_k) = \dfrac{\left\| f'_{k-1}\left(A_{k-1}x_k\right)\right\| \, \|A_{k-1}x_k\|}{\|\eta_k\|}$.

Из неравенства (2) имеем :

$$\frac{Mh_k}{\gamma} \geq (A_{k-1}x_{k-1}, \xi_k) \geq \frac{Nh_k}{\gamma}. \tag{4}$$

Рассмотрим 2 случая :

1) $(A_{k-1}x_{k-1}, \xi_k) \geq h_k$. Тогда: из $h_k \geq \dfrac{\gamma\,(A_{k-1}x_{k-1}, \xi_k)}{M}$,

$$\alpha_k^2\left[(A_{k-1}x_{k-1}, \xi_k) - h_k\right]^2 \leq \alpha_k^2\left[(A_{k-1}x_{k-1}, \xi_k) - \frac{\gamma\,(A_{k-1}x_{k-1}, \xi_k)}{M}\right]^2 =$$

$$= \frac{(M+N)^2}{(M-N)^2}\cdot\left[1 - \frac{2MN}{(M+N)\,M}\right]^2 (A_{k-1}x_{k-1}, \xi_k)^2 = (A_{k-1}x_{k-1}, \xi_k)^2 .$$

Из (3) имеем: $\|A_kx_k\|^2 \leq \|A_{k-1}x_{k-1}\|^2$.

2) $(A_{k-1}x_{k-1}, \xi_k) \leq h_k$.

Из неравенства (4) имеем : $h_k \leq \dfrac{\gamma\,(A_{k-1}x_{k-1}, \xi_k)}{N}$,

$$\alpha_k^2\left[(A_{k-1}x_{k-1}, \xi_{k-1}) - h_k\right]^2 \leq \alpha_k^2\left[(A_{k-1}x_{k-1}, \xi_k) - \frac{\gamma\,(A_{k-1}x_{k-1}, \xi_k)}{N}\right]^2 =$$

$$= \frac{(M+N)^2}{(M-N)^2}\cdot\left[1 - \frac{2MN}{(M+N)\,N}\right]^2 (A_{k-1}x_{k-1}, \xi_k)^2 = (A_{k-1}x_{k-1}, \xi_k)^2 .$$

Отсюда $\|A_kx_k\|^2 \leq \|A_{k-1}x_{k-1}\|^2$.

Анализ этих случаев показывает, что всегда $\|A_kx_k\|^2 \leq$ $\leq \|A_{k-1}x_{k-1}\|^2$, что и требовалось доказать.

4. Численные эксперименты

Для ЭВМ «М-220» была разработана исследовательская программа, с помощью которой отрабатывались различные модификации метода растяжения пространства.

1 модификация. $h_k = h = const$; $\alpha_k = \alpha = const$.

Пример 1. $f(x) = x_1^2 + 40x_2^2$. Начальное приближение: $x_1^{(0)} = 1$; $x_2^{(0)} = 1$; $h = 1$.

а) $\alpha = \dfrac{5}{3}$, 20-я итерация: $x_1^{(20)} = 0,0023$; $\quad x_2^{(20)} = 0,00026$.

б) $\alpha = \dfrac{100}{49}$, 25-я итерация: $x_1^{(25)} = 0,000018$; $\quad x_2^{(25)} = 0,000026$.

Пример 2. $f(x) = x_1^2 + |x_2|^3$; $x_1^{(0)} = 1$; $x_2^{(0)} = 1$; $h = 1$, $\alpha = \dfrac{5}{3}$. После 40 итераций: $x_1^{(40)} = 0,0000044$; $x_2^{(40)} = 0,000127$.

Пример 3. $f(x) = \max\{x_1^2+x_2^2;\ 10[(x_1-1)^2+x_2^2]\}$; $x_1^{(0)} = 0$; $x_2^{(0)} = 1$; $h = 1$, $\alpha = 2$. После 23 итераций: $x_1^{(23)} = 0,75986$; $x_2^{(23)} = 0,0051986$. Истинный минимум: $x_1^* = 0,7598$; $\quad x_2^* = 0$.

Пример 4. $f(x) = x_1^2 + 10x_2^2 + 30x_3^2 + 50x_4^2 + 90x_5^2 + 100x_6^2$; $x^0 = (1,1,1,1,1,1)$; $h = 1$, $\alpha = 2$. После 100 итераций:

$x_1^{(100)} = -1,42 \cdot 10^{-5}$; $\quad x_2^{(100)} = -2,48 \cdot 10^{-6}$; $\quad x_3^{(100)} = 3,23 \cdot 10^{-7}$;
$x_4^{(100)} = 2,34 \cdot 10^{-6}$; $\quad x_5^{(100)} = -5,8 \cdot 10^{-8}$; $\quad x_6^{(100)} = 1,08 \cdot 10^{-6}$.

Эта же функция оптимизировалась при том же начальном приближении, $\alpha = 2$, но h_k выбиралась по формуле:

$$h_k = \frac{2f(x_k)}{\|\eta_k\|}.$$

После 50 итераций получили:

$x_1^{(50)} = 1,84 \cdot 10^{-9}$; $\quad x_2^{(50)} = 3,21 \cdot 10^{-10}$; $\quad x_3^{(50)} = 1,6 \cdot 10^{-10}$;
$x_4^{(50)} = 3,81 \cdot 10^{-10}$; $\quad x_5^{(50)} = 8,52 \cdot 10^{-11}$; $\quad x_6^{(50)} = 7,87 \cdot 10^{-11}$.

Приблизительно такие же результаты получились при $\alpha = \dfrac{5}{3}$ за 60 итераций.

Наиболее интересным примером для апробирования данного метода была следующая задача.

Решить систему 3-х уравнений с 3-мя неизвестными:

$$\begin{cases} S_1 = x_1 + x_2 + x_3 - 1 = 0 \\ S_2 = 0,5x_1^2 + 0,2x_2^3 + 0,5x_3^2 - x_3 + 0,5 = 0 \\ S_3 = x_1 + x_2 + 0,5x_3^2 - 0,5 = 0. \end{cases}$$

Как известно, эта задача эквивалентна такой задаче: минимизировать функцию $\varphi = \sum_{i=1}^3 S_i^2$ или аналогично: минимизировать функцию $\varphi = \max\limits_i S_i^2 \quad (i = 1,2,3)$. Минимум этой функции существует и равняется 0. Координаты точки, в которой достигается минимум функции φ, являются решением данной системы.

Функция φ минимизировалась при переменном h, переменном и постоянном α и начальных условиях $x_1^{(0)} = 0,2$; $x_2^{(0)} = 0,2$; $x_3^{(0)} = 0,5$. Точность результатов достаточно высока.

Предлагается вниманию один из вариантов счета на ЭВМ.

Минимизировать φ при $\gamma = 3,5$ и $\alpha = 2$; $\varphi = \max_i S_i^2$.

Результаты даны в приложении.

Как видно из приведенных примеров, предлагаемый метод является эффективным для минимизации функции «овражного» типа.

ПРИЛОЖЕНИЕ

Функция $\varphi = \max\limits_{i=1,2,3} S_i^2$ минимизировалась при начальных условиях

$$x_1^{(0)} = 0,2, \quad x_2^{(0)} = 0,2, \quad x_3^{(0)} = 0,5.$$

Время минимизации: $t \approx 1$ сек. (без печати).

Истинный минимум: $x_1^* = 0, \quad x_2^* = 0, \quad x_3^* = 1.$

Результаты счета $\left(\alpha = 2, \ \ h_k = 3,5 \cdot \dfrac{f(x_k)}{||\eta_k||}\right)$:

№ итерации	x_1	x_2	x_3
1	$-2,34196905 \cdot 10^{-2}$	$1,78810362 \cdot 10^{-1}$	$9,41450773 \cdot 10^{-1}$
2	$-8,16830448 \cdot 10^{-2}$	$8,41567071 \cdot 10^{-2}$	$8,83365350 \cdot 10^{-1}$
3	$-3,19177314 \cdot 10^{-4}$	$1,54071584 \cdot 10^{-1}$	$9,31868333 \cdot 10^{-1}$
4	$-6,39540071 \cdot 10^{-2}$	$8,71700629 \cdot 10^{-2}$	$9,06798667 \cdot 10^{-1}$
5	$-1,46799376 \cdot 10^{-2}$	$1,01292623 \cdot 10^{-1}$	$9,65876271 \cdot 10^{-1}$
6	$-5,30606773 \cdot 10^{-2}$	$3,33666809 \cdot 10^{-2}$	$9,79782951 \cdot 10^{-1}$
7	$-2,74751833 \cdot 10^{-2}$	$-7,79663206 \cdot 10^{-2}$	$1,13537478 \cdot 10^{0}$
8	$-2,48891073 \cdot 10^{-2}$	$3,08417906 \cdot 10^{-1}$	$7,32521830 \cdot 10^{-1}$
9	$-3,74401296 \cdot 10^{-2}$	$2,73236657 \cdot 10^{-2}$	$1,00958546 \cdot 10^{0}$
10	$-2,86302821 \cdot 10^{-3}$	$-6,80732447 \cdot 10^{-3}$	$1,00914536 \cdot 10^{0}$
			(продолжение следует)

(продолжение)			
№ итерации	x_2	x_2	x_3
11	$-2,62364389 \cdot 10^{-3}$	$-6,90444696 \cdot 10^{-3}$	$1,00992183 \cdot 10^{0}$
12	$-3,08415124 \cdot 10^{-3}$	$-5,05653411 \cdot 10^{-3}$	$1,00778048 \cdot 10^{0}$
13	$-3,53400917 \cdot 10^{-3}$	$-1,01101322 \cdot 10^{-3}$	$1,00481517 \cdot 10^{0}$
14	$-4,37964526 \cdot 10^{-3}$	$-3,87906298 \cdot 10^{-3}$	$1,00029942 \cdot 10^{0}$
15	$-6,69755703 \cdot 10^{-3}$	$1,99772981 \cdot 10^{-2}$	$9,86871125 \cdot 10^{-1}$
16	$-2,22250175 \cdot 10^{-3}$	$1,01189804 \cdot 10^{-2}$	$1,01241284 \cdot 10^{0}$
17	$-4,61722055 \cdot 10^{-3}$	$5,76726461 \cdot 10^{-3}$	$9,98830214 \cdot 10^{-1}$
18	$-4,80155521 \cdot 10^{-3}$	$7,05823834 \cdot 10^{-3}$	$9,97758124 \cdot 10^{-1}$
19	$-4,34916597 \cdot 10^{-3}$	$3,98878782 \cdot 10^{-3}$	$1,00035068 \cdot 10^{0}$
20	$-4,02906569 \cdot 10^{-3}$	$1,87312521 \cdot 10^{-3}$	$1,00216320 \cdot 10^{0}$
21	$4,97682155 \cdot 10^{-4}$	$-2,62913445 \cdot 10^{-3}$	$1,00213871 \cdot 10^{0}$
22	$-1,47189105 \cdot 10^{-3}$	$3,75047645 \cdot 10^{-3}$	$9,97721415 \cdot 10^{-1}$
23	$2,52797550 \cdot 10^{-3}$	$-4,83942860 \cdot 10^{-4}$	$9,97955756 \cdot 10^{-1}$
24	$-7,20102717 \cdot 10^{-4}$	$2,25606586 \cdot 10^{-3}$	$9,98463367 \cdot 10^{-1}$
25	$8,83980114 \cdot 10^{-4}$	$-2,34598877 \cdot 10^{-4}$	$9,99349147 \cdot 10^{-1}$
26	$1,16860382 \cdot 10^{-3}$	$-2,43455684 \cdot 10^{-4}$	$9,99075955 \cdot 10^{-1}$
27	$2,63516281 \cdot 10^{-4}$	$-2,09003892 \cdot 10^{-4}$	$9,99944715 \cdot 10^{-1}$
28	$-2,71840906 \cdot 10^{-4}$	$-1,86175041 \cdot 10^{-4}$	$1,00045859 \cdot 10^{0}$
29	$7,40435296 \cdot 10^{-4}$	$-2,27783026 \cdot 10^{-4}$	$9,99481177 \cdot 10^{-1}$
30	$-1,22790770 \cdot 10^{-4}$	$4,98292871 \cdot 10^{-4}$	$9,99624297 \cdot 10^{-1}$
31	$-1,57573533 \cdot 10^{-4}$	$7,54403286 \cdot 10^{-4}$	$9,99403321 \cdot 10^{-1}$
32	$-4,38755065 \cdot 10^{-5}$	$-8,17559547 \cdot 10^{-5}$	$1,00012563 \cdot 10^{0}$
33	$-4,86606774 \cdot 10^{-5}$	$-4,65926937 \cdot 10^{-5}$	$1,00009523 \cdot 10^{0}$
			(продолжение следует)

(продолжение)			
№ итерации	x_2	x_2	x_3
34	$-5,01445367 \cdot 10^{-5}$	$-3,56555189 \cdot 10^{-5}$	$1,00008581 \cdot 10^{0}$
35	$-6,32631989 \cdot 10^{-5}$	$6,07874988 \cdot 10^{-5}$	$1,00000246 \cdot 10^{0}$
36	$-7,51153687 \cdot 10^{-5}$	$1,47949611 \cdot 10^{-4}$	$9,99927172 \cdot 10^{-1}$
37	$-5,65799717 \cdot 10^{-5}$	$1,16541053 \cdot 10^{-5}$	$1,00004492 \cdot 10^{0}$
38	$2,31719515 \cdot 10^{-5}$	$-6,70734832 \cdot 10^{-5}$	$1,00004390 \cdot 10^{0}$
39	$-4,58273267 \cdot 10^{-5}$	$1,42374285 \cdot 10^{-5}$	$1,00003159 \cdot 10^{0}$
40	$4,91778943 \cdot 10^{-6}$	$-2,46727375 \cdot 10^{-5}$	$1,00001975 \cdot 10^{0}$
41	$5,37726374 \cdot 10^{-6}$	$-2,96939818 \cdot 10^{-5}$	$1,00002431 \cdot 10^{0}$
42	$3,45020334 \cdot 10^{-6}$	$-8,62901495 \cdot 10^{-6}$	$1,00000517 \cdot 10^{0}$

Литература

1. Гельфанд И. М., Цетлин М. Л. О некоторых способах управления сложными системами // УМН. – 1962. – Т. XYII. – Вып. 1.

2. Шор Н. З. О скорости сходимости обобщенного градиентного спуска // Кибернетика. – 1968. – № 3. – С. 98–99.

Использование операции растяжения пространства в задачах минимизации выпуклых функций

Н. З. Шор

Кибернетика. – 1970. – № 1. – С. 87–91.

1. Введение

Рассмотрим задачу минимизации выпуклой функции $f(x)$, определенной на всем n-мерном евклидовом пространстве E_n. При этом от функции $f(x)$ не будем требовать непрерывности производных по направлению. В этих условиях, как известно, обычные градиентные методы или методы покоординатного спуска не работают, и приходится использовать метод обобщенного градиентного спуска [2], который сходится при определенной регулировке шага для произвольной выпуклой функции $f(x)$, имеющей минимум. Однако, если использовать универсальный способ регулировки шага, т. е. требовать, чтобы $\sum_{k=1}^{\infty} h_k = \infty$ и $h_k \to 0$, где h_k – длина приращения вектора-аргумента на k-ом шаге, то скорость сходимости получается весьма медленной. В работе [3] показано, что при определенных ограничениях, накладываемых на $f(x)$, при помощи специального способа регулировки шага можно добиться сходимости обобщенного градиентного спуска со скоростью геометрической прогрессии. Точнее, если выпуклая функция $f(x)$ достигает своего минимума в точке x^*, и мы выбрали начальное приближение x_0 и начальный шаг h_0 таким образом, что существует $\sigma \geq \sqrt{2}$ такое, что:

1) $\|x_0 - x^*\| \leq h_0\sigma$;

2) для произвольной пары x, y такой, что

$$f(x) = f(y) \neq f(x^*); \quad \|x - x^*\| \leq h_0\sigma; \quad \|y - y^*\| \leq h_0\sigma,$$

выполняется неравенство $\dfrac{\|x - x^*\|}{\|y - y^*\|} \leq \sigma$, то, выбирая шаг в методе обобщенного градиентного спуска по формуле $h_{k+1} = h_k \dfrac{\sqrt{\sigma^2 - 1}}{\sigma}$, получим сходящуюся последовательность $\{x_k\}$, $k = 0, 1, 2, \ldots$; при этом

$\|x_k - x^*\| \leq h_k\sigma$. На практике σ может оказаться довольно большим числом, при этом знаменатель геометрической прогрессии оказывается близким к единице и скорость сходимости медленной. Более того, для некоторых $f(x)$ вообще невозможно подобрать $\sigma < \infty$, удовлетворяющее выше приведенным условиям. Такие функции будем называть существенно овражными.

Следует отметить, что существенно овражные функции встречаются на практике весьма часто, особенно в нелинейных задачах минимаксного типа. Например, функция

$$f(x_1, x_2) = \max\left[x_1^2 + x_2^2, (x_1 - 1)^2 + x_2^2\right]$$

является существенно овражной. Из выше сказанного видно, какое большое значение имеет ускорение сходимости метода обобщенного градиентного спуска.

Ниже будет изложен метод минимизации выпуклых функций, представляющий собой комбинацию метода обобщенного градиентного спуска и преобразования пространства аргументов. Как показали теоретические исследования и численные эксперименты, этот метод, лишь незначительно отличаясь по сложности от обобщенного градиентного спуска, обладает большими преимуществами по сравнению с ним в скорости сходимости.

2. Вычисление обобщенного градиента функции при линейном преобразовании пространства аргументов

Обобщенным градиентом выпуклой функции $f(x)$ в точке x_0 называется вектор $\widehat{g}_f(x_0)$, удовлетворяющий следующему неравевству для произвольного $x \in E_n$:

$$f(x) - f(x_0) \geq (\widehat{g}_f(x_0), x - x_0)\,. \tag{1}$$

Пусть произведено неособенное линейное преобразование пространства E_n, определяемое оператором A:

$$y = Ax; \quad x = A^{-1}y.$$

Рассмотрим функцию

$$\varphi(y) = f(A^{-1}y).$$

Функция $\varphi(y)$ является выпуклой. В самом деле, для произвольных $\alpha, \beta \geq 0, \quad \alpha + \beta = 1$

$$\varphi(\alpha y_1 + \beta y_2) = f\left(\alpha A^{-1} y_1 + \beta A^{-1} y_2\right)^2 \leq$$
$$\leq \alpha f\left(A^{-1} y_1\right) + \beta f\left(A^{-1} y_2\right) = \alpha\varphi(y_1) + \beta\varphi(y_2).$$

Вычислим обобщенный градиент функции $\varphi(y)$ в точке $y_0 = Ax_0$:

$$\varphi(y) - \varphi(y_0) = f\left(A^{-1} y\right) - f\left(A^{-1} y_0\right) \geq$$
$$\geq \left(\widehat{g}_f(x_0), A^{-1}(y - y_0)\right) = \left((A^*)^{-1}\widehat{g}_f(x_0), y - y_0\right).$$

A^* – оператор, сопряженный A. Таким образом,

$$\varphi(y) - \varphi(y_0) \geq \left((A^*)^{-1}\widehat{g}_f(x_0), y - y_0\right). \tag{2}$$

Сравнивая формулы (2) и (1), получаем, что $(A^*)^{-1}\widehat{g}_f(x_0)$ является обобщенным градиентом функции $\varphi(y)$ в точке $y_0 = Ax_0$.

$$\widehat{g}_\varphi(y_0) = (A^*)^{-1}\,\widehat{g}_f(x_0); \qquad x_0 = A^{-1} y_0. \tag{3}$$

Так как обобщенный градиент в заданной точке определяется неоднозначно, то равенство (3) следует понимать следующим образом: между множествами обобщенных градиентов $\{\widehat{g}_\varphi(y_0)\}$ и $\{\widehat{g}_f(x_0)\}$ можно установить такое взаимно однозначное соответствие, что будут выполняться равенства (3). Формулы (3) позволяют вычислять обобщенный градиент функции $\varphi(x)$.

3. Операторы растяжения пространства

Пусть задан вектор $\xi \in E_n$, $\|\xi\| = 1$, и число $\alpha > 0$. Каждый вектор $x \in E_n$ однозначно представим в следующем виде:

$$x = \gamma_\xi(x) \cdot \xi + d_\xi(x), \qquad (\xi, d_\xi(x)) = 0. \tag{4}$$

Определение 1. Оператором растяжения в направлении ξ с коэффициентом α назовем оператор $R_\alpha(\xi)$, действующий следующим образом на вектор x, представленный в форме (4):

$$R_\alpha(\xi)x = \alpha\gamma_\xi(x) \cdot \xi + d_\xi(x). \tag{5}$$

Из этого определения сразу следует:

1) $R_{\alpha\beta}(\xi) = R_\alpha(\xi) \cdot R_\beta(\xi)$;

2) $R_\alpha(\xi) \cdot R_{1/\alpha}(\xi) = R_1(\xi) = E$.

Замечание. Оператор растяжения с $\alpha < 1$ мы будем иногда называть оператором сжатия.

Пусть координаты вектора ξ в некоторой ортонормированой системе координат равны $\xi^{(1)}, \ldots, \xi^{(n)}$. Тогда в этой системе координат преобразованию $R_\alpha(\xi)$ соответствует матрица $R_\alpha(\xi)$ с элементами $\{r_{ij}\}$, вычисляемыми по следующим формулам:

$$r_{ij} = \begin{cases} (\alpha - 1)\xi^{(i)} \cdot \xi^{(j)} & \text{для} \quad i \neq j; \\ (\alpha - 1)(\xi^{(i)})^2 + 1 & \text{для} \quad i = j. \end{cases} \tag{6}$$

Заметим, что умножение матрицы $R_\alpha(\xi)$ на вектор требует порядка cn операций умножения, сложения и вычитания, а умножение $R_\alpha(\xi)$ на матрицу требует порядка cn^2 арифметических операций.

Определение 2. Коэффициентом объемного расширения подпространства $R_m \subseteq E_n$ под воздействием оператора A назовем $\sqrt{G}$, где G – определитель Грама, построенный на векторах $\sigma_1, \ldots, \sigma_m$, $\sigma_i = Ae_i$, $i = 1, \ldots, m$, $\{e_i\}$ – ортонормированный базис подпространства R_m. Этот коэффициент будем обозначать $v_{R_m} \cdot (A)$.

Из свойств определителя Грама следуют утверждения:

1) $v_{R_m}(A_1 A_2) = v_{A_2 R_m}(A_1) \cdot v_{R_m}(A_2)$; (7)

2) $v_{E_n}(R_\alpha(\xi)) = \alpha$; (8)

3) $v_{R_m}(R_\alpha(\xi)) = \sqrt{1 + (\alpha^2 - 1)\|\psi\|^2}$, (9)

где ψ – проекция вектора ξ на подпространство R_m;

4) пусть R_{n-m} – ортогональное дополнение пространства R_m в E_n,

тогда

$$v_{R_m}(R_\alpha(\xi)) \cdot v_{R_{n-m}}(R_\alpha(\xi)) \geq \alpha; \tag{10}$$

из (9) вытекает важное следствие –

5) если $\alpha \geq 1$, то $v_{R_m}(R_\alpha(\xi)) \geq 1$ для произвольного R_m.

4. Обобщенный градиентный спуск с растяжением пространства

Ниже мы рассмотрим класс алгоритмов минимизации выпуклых функций, на каждом шаге которых движение в направлении обобщенного антиградиента будет сочетаться с операцией растяжения пространства аргументов в направлении обобщенного градиента. Алгоритмы этого класса мы будем называть алгоритмами обобщенного градиентного спуска с растяжением пространства (сокращенно ОГСРП-алгоритмы).

I. Для заданной выпуклой функции $f(x)$ имеется алгоритм вычисления $\widehat{g}_f(x)$ в произвольной точке $x \in E_n$; заданы операторы вычисления последовательностей положительных чисел $\{h_k\}$ и $\{\alpha_k\}$, $\quad k = 1, 2, \ldots$

II. 0-й шаг алгоритма.
Выбираем начальное приближение $x = x_0$ и матрицу $B_0 = A_0^{-1} = E$.

III. $(k+1)$-й шаг алгоритма. $k = 0, 1, \ldots$

Вычисляем:

1) $\widehat{g}_f(x_k)$;
2) $\widehat{g}_{\varphi_k}(y_k) = B_k^* \widehat{g}_f(x_k)$. $\qquad (11)$

Примечание. $\quad \varphi_k(y) = f(B_k y); \quad y_k = A_k x_k$.

3) $\xi_{k+1} = \dfrac{\widehat{g}_{\varphi_k}(y_k)}{\|\widehat{g}_{\varphi_k}(y_k)\|}$; $\qquad (12)$
4) h_{k+1};
5) α_{k+1};
6) $x_{k+1} = x_k - B_k h_{k+1} \xi_{k+1}$; $\qquad (13)$
7) $R^{-1}_{\alpha_{k+1}}(\xi_{k+1}) = R_{1/\alpha_{k+1}}(\xi_{k+1})$;
8) $B_{k+1} = A^{-1}_{k+1} = B_k \cdot R_{1/\alpha_{k+1}}(\xi_{k+1})$. $\qquad (14)$

Различные варианты алгоритмов ОГСРП будут отличаться друг от друга способами вычисления последовательностей $\{\alpha_k\}$ и $\{h_k\}$.

Теорема 1. *Пусть в процессе выполнения ОГСРП выполняются следующие условия для* $k = 0, 1, 2, \ldots$*:*

$$\inf_{x^* \in M^*} \|A_k(x_k - x^*)\| \le r; \tag{15}$$

$$\alpha_k \geq 1 + \varepsilon, \tag{16}$$

где r и ε – некоторые положительные константы, M^* – множество точек минимума $f(x)$, которое предполагается ограниченным. Тогда

$$\lim_{k\to\infty} f(x_k) = f(x^*), \quad x^* \in M^*.$$

Доказательство. Для доказательства введем некоторые определения. Пусть задано множество операторов $\mathfrak{M} = \{A_\gamma\}$. Подпространство $R_m \subseteq E_n$ назовем особым по отношению к множеству $\mathfrak{M}$, если $\sup v_{R_m}(A_\gamma) < \infty$.

В дальнейшем под $v_x(A)$ мы будем понимать $v_{R_1}(A)$, где R_1 порождено вектором $x \neq 0$. Докажем лемму.

Лемма. *Пусть заданы 2 линейно независимых вектора x, y, оператор A и коэффициент растяжения $v_x(A) = \dfrac{\|A_x\|}{\|x\|}$ и $v_y(A) = \dfrac{\|A_y\|}{\|y\|}$. Тогда справедливо следующее неравенство для произвольного вектора $z = ax + by, \quad z \neq 0$:*

$$v_z(A) \leq \frac{\sqrt{2}\max\left(v_x(A), v_y(A)\right)}{\sqrt{1 - |\cos(\widehat{x,y})|}} \tag{17}$$

Доказательство.

$$\begin{aligned} v_z(A) &= \frac{\|A(ax+by)\|}{\|ax+by\|} \leq \frac{|a|\cdot\|Ax\| + |b|\cdot\|Ay\|}{\sqrt{a^2\|x\|^2 + 2ab(x,y) + b^2\|y\|^2}} \leq \\ &\leq \frac{\Big[\|x\|\cdot|a| + |b|\cdot\|y\|\Big]\cdot\max\Big(v_x(A), v_y(A)\Big)}{\sqrt{a^2\|x\|^2 + 2ab\|x\|\|y\|\cos(\widehat{x,y}) + b^2\|y\|^2}}. \end{aligned} \tag{18}$$

Найдем минимум выражения $a^2\|x\|^2 + 2ab\|x\|\cdot\|y\|\cos\varphi + b^2\|y\|^2$ по $a, b, \|x\|, \|y\|$ при ограничении $|a|\cdot\|x\| + |b|\cdot\|y\| = c$, где $c > 0$. Легко видеть, что этот минимум достигается при $|a|\cdot\|x\| = |b|\cdot\|y\| = \dfrac{c}{2}$ и при $ab\cos\varphi \leq 0$ равен $\dfrac{c^2}{2}(1 - |\cos\varphi|)$. Отсюда и из (18) получаем:

$$v_z(A) \leq \frac{c\max(v_x(A), v_y(A))}{\sqrt{\dfrac{c^2}{2}(1 - |\cos(\widehat{x,y})|)}} = \frac{\sqrt{2}\max(v_x(A), v_y(A))}{\sqrt{1 - |\cos(\widehat{x,y})|}}.$$

Доказательство леммы завершено.

Следствие. Совокупность особых векторов по отношению к множеству операторов $\mathfrak{M}$ образует подпространство.

Доказательство. Выберем максимальное число линейно независимых особых векторов. Тогда из предыдущей леммы следует, что и произвольная линейная комбинация этих векторов будет особым вектором, а это значит, что совокупность особых векторов образует подпространство.

Теперь перейдем к доказательству теоремы. Так как A_k является произведением операторов растяжения (см.(14)), то из (16), (7), (9) следует, что последовательность $v_{R_m}(A_k)$ для произвольного подпространства $R_m \subseteq E_n$ будет неубывающей по k, причем для всех $k = 0, 1, 2, \ldots$ $v_{R_m}(A_k) \geq 1$.

Множество минимумов M^* функции $f(x)$ является замкнутым, поэтому в формуле (15) инфимум достигается в некоторой точке, которую мы будем обозначать $x^*(x_k)$.

Так как $\|A_k(x_k - x^*(x_k))\| = v_{x_k - x^*(x_k)}(A_k) \times \|x_k - x^*(x_k)\|$ и $v_{x_k - x^*(x_k)}(A_k) \geq 1$, то из (15) следует:

$$\|x_k - x^*(x_k)\| \leq r. \tag{19}$$

Пусть $\overline{R}_m$ – пространство особых векторов по отношению к множеству $\{A_k\}$, $k = 0, 1, \ldots$; $\overline{R}_{n-m}$ – ортогональное дополнение $\overline{R}_m$ в E_n. Из (19) и ограниченности M^* следует, что последовательность $\{x_k\}$ имеет предельные точки. Пусть $\overline{x} \notin M^*$ – некоторая предельная точка, $\{x_{k_i}\}$ – некоторая последовательность, сходящаяся к этой точке. Покажем, что $\overline{x} - x^*(\overline{x}) \in \overline{R}_m$.

Рассмотрим разложения

$$\overline{x} - x^*(\overline{x}) = \overline{y} + \overline{z}; \qquad x_{k_i} - x^*(x_{k_i}) = y_{k_i} + z_{k_i};$$

где $\overline{y}, y_{k_i} \in \overline{R}_m$; $\quad \overline{z}, z_{k_i} \in \overline{R}_{n-m}$.

Легко показать, что $\lim\limits_{k\to\infty} \min\limits_{z \in \overline{R}_{n-m}} v_z(A_k) = \infty$. В самом деле, рассмотрим пересечение множеств: $\mathfrak{M}_d = \bigcap\limits_{k=0}^{\infty} \{z : z \in \overline{R}_{n-m}, \quad v_z(A_k) \leq d\}$, где $d > 0$. Если это пересечение не пусто, то найдется $z \neq 0, \quad z \in \overline{R}_{n-m}$, что $v_z(A_k) \leq d$ для $k = 0, 1, 2, \ldots$ Но тогда $z \in \overline{R}_m$. Получили противоречие. Следовательно,

$$\lim_{k\to\infty} \min_{z \in \overline{R}_{n-m}} v_z(A_k) = \infty. \tag{20}$$

Если $\overline{z} \neq 0$, то, начиная с некоторого i, будет выполняться неравенство $\|z_{k_i}\| \geq \delta > 0$, так как $\lim\limits_{i\to\infty} z_{k_i} = \overline{z}$. Но тогда $\|A_{k_i}(x_{k_i} - x^*(x_{k_i}))\| \geq$

$\geq v_{z_{k_i}}(A_{k_i}) \cdot \delta - \|A_{k_i} y_{k_i}\|$. Так как $\{\|A_{k_i} y_{k_i}\|\}$ ограничена, то из (20) получаем $\|A_{k_i}(x_{k_i} - x^*(x_{k_i}))\| \to \infty$ при $i \to \infty$, но это противоречит (15). Таким образом, $\overline{z} = 0$, т. е.

$$\overline{x} - x^*(\overline{x}) \in \overline{R}_m.$$

По определению $\overline{R}_m$ существует конечный предел последовательности

$$v_{\overline{x}-x^*(\overline{x})}(A_k), \quad k = 0, 1, \ldots$$

Но тогда

$$\lim_{k\to\infty} \frac{v_{\overline{x}-x^*(\overline{x})}(A_{k+1})}{v_{\overline{x}-x^*(\overline{x})}(A_k)} = 1. \tag{21}$$

В силу (9), (16), (21) проекция вектора ξ_k на направление $A_k(\overline{x} - x^*(\overline{x}))$ должна стремится к 0, откуда обобщенная производная от функции $f(x)$ по направлению $\overline{x} - x^*(\overline{x})$ в точке x_k должна стремится к 0 при $k \to \infty$.

Замечание 1. Под обобщенной производной от функции $f(x)$ по направлению η в заданной точке x_0 понимается проекция обобщенного вектора градиента в этой точке на указанное направление. Мы будем для обобщенной производной по направлению применять обозначение $f'_\eta(x_0)$. Для производной по направлению от выпуклой функции $f(x)$ справедливо неравенство

$$\left|f'_{x-x_0}(x)\right| \cdot \|x - x_0\| \geq f(x) - f(x_0). \tag{22}$$

Продолжим доказательство теоремы.

Из неравенства (22) получаем:

$$|f'_{\overline{x}-x^*(\overline{x})}(x_{k_i})| \cdot \|\overline{x} - x^*(\overline{x})\| \geq f(x_{k_i}) - f(x_{k_i} - \overline{x} + x^*(\overline{x})). \tag{23}$$

Перейдем к пределу при $i \to \infty$.

Получим из (23):

$$0 \geq f(\overline{x}) - f(x^*(\overline{x})).$$

Значит, $f(\overline{x}) = f(x^*(\overline{x}))$, т. е. $\overline{x} \in M^*$.

Таким образом, все предельные точки последовательности $\{x_k\}$ принадлежат области минимумов функции $f(x)$, что и доказывает теорему 1.

В теореме 1 мы постулировали определенные свойства последовательности $\{A_k x_k\}$. В теореме 2 при некоторых предположениях, накладываемых на $f(x)$, мы сформулируем способ выбора последовательностей $\{h_k\}$ и $\{\alpha_k\}$, при котором требуемые в теореме 1 свойства $\{A_k x_k\}$ выполняются.

Теорема 2. *Пусть множество M^* точек минимума выпуклой функции $f(x)$ ограничено, и $f(x)$ принимает на M^* значение m^*.*

Введем обозначения:

$$\rho(x) = \min_{x^* \in M^*} \|x - x^*\| \, ;$$

$x^*(x)$ – точка, для которой $\|x - x^*(x)\| = \rho(x)$. Пусть для всех x, удовлетворяющих условию $\rho(x) \leq c$, $x \notin M^*$ и $x^* \in M^*$ выполняются следующие неравенства:

$$N\left[f(x) - m^*\right] \leq \left|f'_{x-x^*}(x)\right| \cdot \|x - x^{**}\| \leq M\left[f(x) - m^*\right], \qquad (24)$$

где x^{**} – ближайшая к x точка минимума, лежащая на луче $y = x + t(x - x^*)$, $t \geq 0$.

Тогда, если при применении ОГСРП

1) $\rho(x_0) \leq c$; (25)

2) $h_{k+1} = \dfrac{2MN}{M+N} \cdot \dfrac{[f(x_k) - m^*]}{\|\widehat{g}_{\varphi_k}(y_k)\|}$; (26)

3) $1 + \varepsilon \leq \alpha_{k+1} \leq \dfrac{M+N}{M-N}$, (27)

то будут выполнены условия теоремы 1. (Здесь M, N, c, ε – положительные константы.)

Доказательство. Для доказательства достаточно показать, что последовательность $\min\limits_{x^* \in M^*} \|A_k(x_k - x^*)\|$ является невозрастающей.

Используя (14), (13), получаем :

$$\begin{aligned}
&\min_{x^* \in M^*} \|A_k(x_k - x^*)\|^2 = \\
&= \min_{x^* \in M^*} \|R_{\alpha_k}(\xi_k) \cdot A_{k-1}(x_k - x^*)\|^2 = \\
&= \min_{x^* \in M^*} \|R_{\alpha_k}(\xi_k) \cdot A_{k-1}(x_{k-1} - x^*) - R_{\alpha_k}(\xi_k) h_k \xi_k\|^2 \leq \\
&\leq \left\|R_{\alpha_k}(\xi_k) \cdot A_{k-1}(x_{k-1} - x^*_{k-1}) - R_{\alpha_k}(\xi_k) h_k \xi_k\right\|^2,
\end{aligned} \qquad (28)$$

где x^*_{k-1} – значение x^*, при котором достигается минимум в выражении:

$$\min_{x^* \in M^*} \|A_{k-1}(x_{k-1} - x^*)\| \, . \qquad (29)$$

Введем обозначение:

$$A_{k-1}(x_{k-1} - x_{k-1}^*) = z_{k-1}. \tag{30}$$

Из (28), (9) получаем :

$$\begin{aligned}
&\min_{x^* \in M^*} \|A_k(x_k - x^*)\|^2 \le \|R_{\alpha_k}(\xi_k) z_{k-1} - R_\alpha(\xi_k) \cdot h_k \xi_k\|^2 = \\
&= \|R_{\alpha_k}(\xi_k) \cdot z_{k-1}\|^2 - 2h(R_{\alpha_k}(\xi_k) z_{k-1}, R_{\alpha_k} h_k \xi_k) + \alpha_k^2 h_k^2 = \\
&= \left[1 + (\alpha_k^2 - 1)\frac{(z_{k-1}, \xi_k)^2}{\|z_{k-1}\|^2}\right] \cdot \|z_{k-1}\|^2 - 2\alpha_k^2 h_k (z_{k-1}, \xi_k) + \alpha_k^2 h_k^2 = \\
&= \|z_{k-1}\|^2 + \alpha_k^2 \left[(z_{k-1}, \xi_k) - h_k\right]^2 - (z_{k-1}, \xi_k)^2.
\end{aligned} \tag{31}$$

Оценим (z_{k-1}, ξ_k), используя (11), (12) :

$$\begin{aligned}
(z_{k-1}, \xi_k) &= \left(A_{k-1}(x_{k-1} - x_{k-1}^*), \xi_k\right) = \left(x_{k-1} - x_{k-1}^*, A_{k-1}^* \xi_k\right) = \\
&= \left(x_{k-1} - x_{k-1}^*, \frac{\widehat{g}_f(x_{k-1})}{\|\widehat{g}_{\varphi_{k-1}}(y_{k-1})\|}\right) = \\
&= \frac{f'_{x_{k-1}} - x_{k-1}^*(x_{k-1}) \cdot \|x_{k-1} - x_{k-1}^*\|}{\|\widehat{g}_{\varphi_{k-1}}(y_{k-1})\|}
\end{aligned} \tag{32}$$

(здесь мы воспользовались определением обобщенной производной по направлению (см. замечание 1)).

Рассмотрим два случая.

I. $(z_{k-1}, \xi_k) \le h_k$; из (24), (26), (32) получаем :

$$\begin{aligned}
h_k &= \frac{[f(x_{k-1}) - m^*]}{\|\widehat{g}_{\varphi_{k-1}}(y_{k-1})\|} \cdot \frac{2MN}{M+N} \le \\
&\le \frac{f'_{x_{k-1} - x_{k-1}^*}(x_{k-1}) \cdot \|x_{k-1} - x_{k-1}^*\| \cdot 2M}{\|\widehat{g}_{\varphi_{k-1}}(y_{k-1})\| \cdot (M+N)} = (z_{k-1}, \xi_k) \cdot \frac{2M}{M+N}.
\end{aligned}$$

Получаем

$$\begin{aligned}
&\alpha_k^2 [(z_{k-1}, \xi_k) - h_k]^2 - (z_{k-1}, \xi_k)^2 \le \\
&\le \alpha_k^2 \left[(z_{k-1}, \xi_k) - (z_{k-1}, \xi_k) \cdot \frac{2M}{M+N}\right]^2 - (z_{k-1}, \xi_k)^2 \le \\
&\le \left[\left(\frac{M+N}{M-N}\right)^2 \cdot \left(1 - \frac{2M}{M+N}\right)^2 - 1\right] \cdot (z_{k-1}, \xi_k)^2 = 0.
\end{aligned}$$

II. $(z_{k-1}, \xi_k) \geq h_k$; из (24), (26), (32) получаем :

$$h_k \geq f'_{x_{k-1}-x^*_{k-1}}(x_{k-1}) \cdot \left\|x_{k-1} - x^*_{k-1}\right\| \cdot \frac{2N}{\left\|\widehat{g}_{\varphi_{k-1}}(y_{k-1})\right\| \cdot (M+N)} =$$

$$= (z_{k-1}, \xi_k) \cdot \frac{2N}{M+N}.$$

Отсюда

$$\alpha_k^2 \left[(z_{k-1}, \xi_k) - h_k\right]^2 - (z_{k-1}, \xi_k)^2 \leq$$

$$\leq \alpha_k^2 \left[(z_{k-1}, \xi_k) - (z_{k-1}, \xi_k) \cdot \frac{2N}{M+N}\right]^2 - (z_{k-1}, \xi_k)^2 \leq$$

$$\leq \left[\left(\frac{M+N}{M-N}\right)^2 \cdot \left(1 - \frac{2N}{M+N}\right)^2 - 1\right] \cdot (z_{k-1}, \xi_k)^2 = 0.$$

Таким образом, в обоих случаях:

$$\alpha_k^2 \cdot \left[(z_{k-1}, \xi_k) - h_k\right]^2 - (z_{k-1}, \xi_k)^2 \leq 0. \tag{33}$$

Подставляя (33) в (31) получаем :

$$\min_{x^* \in M^*} \|A_k(x_k - x_k^*)\|^2 \leq \|z_{k-1}\|^2 = \min_{x^* \in M^*} \left\|A_{k-1}(x_{k-1} - x^*_{k-1})\right\|^2.$$

Отсюда сразу следует теорема 2.

Если функция $f(x)$ имеет единственный минимум в точке x^*, то в качестве следствия из теоремы 2 получаем теорему 3.

Теорема 3. *Пусть минимум выпуклой функции $f(x)$ достигается в единственной точке x^* и $f(x^*) = m^*$, а для всех x, удовлетворяющих условию $\rho(x) \leq c$, выполняются следующие неравенства:*

$$N \cdot [f(x) - m^*] \leq \left|f'_{x-x^*}(x)\right| \cdot \|x - x^*\| \leq M\,[f(x) - m^*].$$

Тогда, если при применении ОГСРП

(1) $\rho(x_0) \leq c$,

(2) $h_{k+1} = \dfrac{2MN}{M+N} \cdot \dfrac{[f(x) - m^*]}{\|\widehat{g}_{\varphi_k}(y_k)\|}$,

(3) $1 + \varepsilon \leq \alpha_{k+1} \leq \dfrac{M+N}{M-N}$,

то будут выполнены условия теоремы 1 *(M, N, c, ε – положительные константы).*

5. Обсуждение результатов

Теоремы 2 и 3 в сочетании с теоремой 1 дают нам возможность строить сходящиеся процедуры ОГСРП. Что можно сказать о скорости сходимости? Если выполняются условия теоремы 3, то $\|A_k(x_k - x^*)\| \le c$ для $k = 0, 1, \ldots$ В силу (7), (8) коэффициент объемного расширения

$$v_{E_n}(A_k) = \prod_{i=1}^{k} \alpha_i.$$

Если обозначить объем n-мерного шара $\|y\| \le c$ через v_c, то вектор $x_k - x^*$ будет локализован в объеме $v_c/v_{E_n}(A_k) = v_c/\prod_{i=1}^{k} \alpha_i$. Если $\alpha_i = \dfrac{M+N}{M-N}$, $i = 1, 2, \ldots$, то объем области локализации $x_k - x^*$ уменьшается не медленнее, чем геометрическая прогрессия со знаменателем $\dfrac{M-N}{M+N}$. Следует отметить, что в силу (22) $N \ge 1$, т. е. при небольших M этот знаменатель существенно отличен от 1. Заметим также, что константы M и N не связаны непосредственно с «коэффициентом овражности» σ (см. § 1) и инвариантны по отношению к линейным преобразованиям пространства E_n.

Если $f(x) = (Kx, x)$ – положительно определённая квадратичная форма, то $M{=}N{=}2$. В этом случае α_k может быть сколько угодно большим положительным числом. В пределе при $\alpha_k \to \infty$ ОГСРП переходит в один из методов ортогонализации для решения систем линейных уравнений [1], который сходится за n шагов.

Теоремы 2 и 3 можно применять непосредственно, когда заранее известно значение $f(x)$ в точках минимума (m^*).

Это имеет место, если нам нужно решать систему из равенств и выпуклых неравенств:

$$\begin{cases} F_i(x) = 0; & i = 1, \ldots, r; \\ H_j(x) \le 0; & j = 1, \ldots, p. \end{cases} \tag{34}$$

Решение этой системы можно свести к задаче минимизации

$$f(x) = \max\left[0, \max_i F_i^2, \max_j H_j(x)\right].$$

Если решение системы (34) существует, то $m^* = \min f(x) = 0$. В общем случае m^* неизвестно. В настоящее время нами исследуются алгоритмы, в которых m^* подбирается в процессе расчетов. Кроме того, исследуются алгоритмы ОГСРП, в которых $\{\alpha_k\}$ и $\{h_k\}$ строятся по схемам, отличным от приведенных в данной статье. В настоящее время в ИК АН УССР В.И.Билецким разработана стандартная программа для ЭВМ М-220, с помощью которой проведено большое число экспериментальных расчетов по различным модификациям ОГСРП. Результаты расчетов показывают, что скорость сходимости ОГСРП практически намного выше, чем при использовании обобщенного градиентного спуска без растяжения пространства. Подробные данные о результатах экспериментальных расчетов содержатся в статье [4]. В заключение следует отметить, что операцию растяжения пространства удобно сочетать с различными методами градиентного типа. Можно надеяться, что для многих задач применение операции растяжения пространства повлечет существенное ускорение сходимости процесса минимизации.

Литература

1. Фаддеев Д. К., Фаддеева В. Н. Вычислительые методы линейной алгебры. – М.: Физматгиз, 1960.

2. Шор Н. З. Обобщенный градиентный спуск // Труды первой зимней школы по математическому программированию (г. Дрогобыч). – М., 1969. – Вып. 3. – С. 578–585.

3. Шор Н. З. О скорости сходимости обобщенного градиентного спуска // Кибернетика. – 1968. – № 3. – С. 80–85.

4. Шор Н. З., Билецкий В. И. Метод растяжения пространства для ускорения сходимости в задачах овражного типа // Теория оптимальных решений. – 1969. – Вып. 2. – С. 3–18.

О скорости сходимости метода обобщенного градиентного спуска с растяжением пространства

Н. З. Шор

Кибернетика. – 1970. – № 2. – С. 80–85.

1. Введение

Мы будем рассматривать выпуклую функцию $f(x)$, определенную на всем евклидовом пространстве E_n и обладающую следующим свойством:

$$\lim_{\|x\| \to \infty} f(x) = +\infty. \tag{1}$$

Из свойства (1) следует, что область минимумов функции является замкнутым ограниченным выпуклым множеством. Пусть функция $f(x)$ принимает на $\mathfrak{M}^*$ значение m^*.

В [1] для минимизации выпуклых функций был предложен метод обобщенного градиентного спуска с растяжением пространства (сокращенно ОГСРП). Обоснование этого метода было проведено для специального случая регулировки шага и коэффициента растяжения пространства, причем значение m^* предполагалось известным.

В настоящей статье, напомнив основные определения и результаты [1], перейдем к подробному исследованию алгоритма, когда m^* заранее неизвестно.

Затем мы получим некоторые оценки скорости сходимости ОГСРП и остановимся на вопросах практического использования этого метода.

2. Основные определения и результаты

Обобщенным градиентом выпуклой функции $f(x)$ в точке x_0 будем называть вектор $\widehat{g}_f\,(x_0)$, удовлетворяющий следующему неравенству для произвольного $x \in E_n$:

$$f(x) - f(x_0) \geq (\widehat{g}_f(x_0), x - x_0). \tag{2}$$

Пусть задан вектор $\eta \neq 0$. Обобщенной производной функции $f(x)$ в направлении η в точке x_0 мы будем называть число $f'_\eta\,(x_0)$, равное проекции вектора $\widehat{g}_f\left(x_0\right)$ на направление η:

$$f'_\eta\left(x_0\right)=\frac{\left(\widehat{g}_f\left(x_0\right),\eta\right)}{\|\eta\|}. \tag{3}$$

Взяв η равным $x-x_0$, получим следующее неравенство:

$$f'_{x-x_0}\left(x_0\right)\cdot\|x-x_0\|\leq f(x)-f\left(x_0\right) \tag{4}$$

или

$$f'_{x-x_0}(x)\cdot\|x-x_0\|\geq f(x)-f\left(x_0\right). \tag{5}$$

Пусть B – некоторый линейный оператор, определенный на E_n. Рассмотрим функцию

$$\varphi(y)=f\left(By\right); \tag{6}$$

$\varphi(y)$ является выпуклой функцией, и обобщенный градиент $\varphi(y)$ можно вычислить по следующей формуле:

$$\widehat{g}_\varphi\left(y_0\right)=B^*\widehat{g}_f\left(x_0\right), \tag{7}$$

где $x_0=By_0$, B^* – оператор, сопряженный оператору B.

Пусть задан некоторый вектор ξ, $\|\xi\|=1$. Оператором растяжения $R_\alpha(\xi)$ в направлении ξ с коэффициентом α будем называть линейный оператор, действующий следующим образом на произвольный вектор $x\in E_n$:

$$R_\alpha(\xi)x=\alpha x_\xi+\left(x-x_\xi\right), \tag{8}$$

где $x_\xi=(x,\xi)\xi$.

Коэффициентом растяжения вектора $x\neq 0$ под воздействием оператора A будем называть величину $k_x\left(A\right)$, определяемую по формуле:

$$k_x\left(A\right)=\frac{\|Ax\|}{\|x\|}. \tag{9}$$

Если A, B – два оператора, то

$$k_x\left(AB\right)=k_{Bx}\left(A\right)k_x\left(B\right). \tag{10}$$

Для оператора растяжения $R_\alpha(\xi)$ справедлива формула

$$k_x\Big(R_\alpha(\xi)\Big)=\sqrt{1+\left(\alpha^2-1\right)\cdot\frac{\left(x,\xi\right)^2}{\|x\|^2}}. \tag{11}$$

Алгоритмом обобщенного градиентного спуска с растяжением пространства (сокращенно ОГСРП-алгоритмом) мы будем называть алгоритм минимизации выпуклых функций,описание которого следует ниже.

I. Предпосылки применения алгоритма ОГСРП.

Для заданной выпуклой функции $f(x)$ имеется алгоритм вычисления $\widehat{g}_f(x)$ в произвольной точке $x \in E_n$; заданы операторы вычисления последовательностей положительных чисел $\{h_k\}$ и $\{\alpha_k\}$, $k = 1, 2, \ldots$.

II. 0-й шаг алгоритма Выбираем начальное приближение $x = x_0$ и матрицу $B_0 = A_0^{-1} = E$.

III. $(\mathbf{k+1})$-й шаг алгоритма; $\mathbf{k = 0, 1}, \ldots$ Вычисляем:

1) $\widehat{g}_f\left(x_k\right);$

2) $\widehat{g}_{\varphi_k}\left(y_k\right) = B_k^* \widehat{g}_f\left(x_k\right).$ (12)

Примечание. $\varphi_k(y) = f\left(B_k y\right); \quad y_k = A_k x_k;$

3) $\xi_{k+1} = \widehat{g}_{\varphi_k}\left(y_k\right) / \left\|g_{\varphi_k}\left(y_k\right)\right\|;$ (13)

4) $h_{k+1};$

5) $\alpha_{k+1};$

6) $x_{k+1} = x_k - B_k h_{k+1} \xi_{k+1};$ (14)

7) $R_{\alpha_{k+1}}^{-1}\left(\xi_{k+1}\right) = R_{1/\alpha_{k+1}}\left(\xi_{k+1}\right);$

8) $B_{k+1} = A_{k+1}^{-1} = B_k R_{1/\alpha_{k+1}}\left(\xi_{k+1}\right).$ (15)

Различные варианты алгоритмов ОГСРП отличаются друг от друга способами вычисления последовательностей $\{h_k\}$ и $\{\alpha_k\}$. В [1] мы подробно рассмотрели случай выбора $\{h_k\}$ и $\{\alpha_k\}$ при известном значении функции в точке минимума и доказали при некоторых предположениях сходимость этого метода. При этом h_{k+1} выбиралось по следующей формуле:

$$h_{k+1} = \frac{\gamma\left[f\left(x_k\right) - m\right]}{\left\|\widehat{g}_{\varphi_k}\left(y_k\right)\right\|}, \tag{16}$$

где γ – некоторая константа, не меньшая 1, $m = m^*$, равной значению функции $f(x)$ в точке минимума. Ниже мы рассмотрим поведение последовательности $\{h_k\}$ в том случае, если m в (16) выбрано неверно и на основе этого получим алгоритм минимизации $f(x)$, когда m^* неизвестно.

3. Алгоритм минимизации с неизвестным m^*

Сначала мы покажем, что если в формуле (16) $m < m^*$, то $\sup\limits_k h_k = \infty$. Для этого докажем лемму.

Лемма 1. *Пусть в процессе выполнения алгоритма* ОГСРП *соблюдаются для всех* $k = 1, 2, \ldots$ *следующие условия:*

$$1) \quad \alpha_k \geq 1 + \varepsilon; \tag{17}$$

$$2) \quad f(x_k) - m^* \leq c, \tag{18}$$

где ε *и* c *– заданные положительные константы. Тогда*

$$\lim_{k\to\infty} \widehat{g}_{\varphi_k}(y_k) = 0.$$

Доказательство. Рассмотрим последовательность операторов $\{B_k^*\}$. Из (15) получаем:

$$B_{k+1}^* = R_{1/\alpha_{k+1}}(\xi_{k+1}) \cdot B_k^*. \tag{19}$$

Из (17) и (11) вытекает, что для произвольного $x \neq 0$

$$k_x\left(R_{1/\alpha_{k+1}}(\xi_{k+1})\right) \leq 1.$$

Применяя (10), (19), получим, что последовательность $\{k_x(B_k^*)\}$ является невозрастающей по k. Значит, для любого x существует $\lim\limits_{k\to\infty} k_x(B_k^*)$.

Из свойства (1) функций $f(x)$ следует, что область

$$R_c = \{x : f(x) - m^* \leq c\}$$

является ограниченной; следовательно, ограниченной является $\|\widehat{g}_f(x)\|$ при $x \in R_c$.

Введем обозначение:

$$\varrho = \sup_{x\in R_c} \|\widehat{g}_f(x)\|. \tag{20}$$

Пусть x – вектор, такой, что

$$\lim_{k\to\infty} k_x(B_k^*) > 0.$$

Тогда из (10), (11), (17) получаем:

$$\lim_{k\to\infty} |(\xi_{k+1}, B_k^* x)| = 0. \tag{21}$$

Разложим $\widehat{g}_f(x_k)$ по базису некоторой ортонормированной системы координат $e_1, \ldots, e_n$:

$$\widehat{g}_f(x_k) = \sum_{i=1}^{n} c_i^{(k)} e_i. \tag{22}$$

Разобъем $\{1, 2, \ldots, k\}$ на два подмножества: I_1 и I_2, где I_1 – подмножество индексов, для которых $\lim\limits_{k\to\infty} k_{e_i}(B_k^*) = 0$, I_2 – дополнительное подмножество. Из (21) следует, что $\lim\limits_{k\to\infty} |(\xi_{k+1}, B_k^* e_i)| = 0$ при $i \in I_2$, откуда (см. (13))

$$\lim_{k\to\infty} |(\widehat{g}_{\varphi_k}(y_k), B_k^* e_i)| = 0, \quad i \in I_2. \tag{23}$$

Далее,

$$\begin{aligned} \widehat{g}_{\varphi_k}(y_k) &= B_k^* \widehat{g}_f(x_k) = \sum_{i\in I_1} c_i^{(k)} B_k^* e_i + \sum_{i\in I_2} c_i^{(k)} B_k^* e_i; \\ \|\widehat{g}_{\varphi_k}(y_k)\|^2 &= \sum_{i\in I_1} c_i^{(k)} (B_k^* e_i, \widehat{g}_{\varphi_k}(y_k)) + \sum_{i\in I_2} c_i^{(k)} (B_k^* e_i, \widehat{g}_{\varphi_k}(y_k)). \end{aligned}$$

Для индексов $i \in I_1$ сходимость членов суммы к 0 следует из (20) ($|c_i^{(k)}| \le \varrho$) и из того, что $\lim\limits_{k\to\infty} B_k^* e_i = 0$ $(i \in I_1)$. Для $i \in I_2$ сходимость членов суммы к 0 следует также из ограниченности $|c_i^{(k)}|$ и из (23). Таким образом, лемма доказана.

Теорема 1. *Пусть выпуклая функция $f(x)$ обладает следующим свойством: существует постоянная $M > 1$ такая, что если $f[(1-\alpha)x_1 + +\alpha x_2]$ $(0 \le \alpha \le 1)$ строго убывает по α, то выполняется неравенство*

$$f'_{x_1 - x_2}(x_1) \cdot \|x_1 - x_2\| \le M[f(x_1) - f(x_2)]. \tag{24}$$

Тогда, если при применении алгоритма ОГСРП *с*

$$\alpha_{k+1} = \frac{M+1}{M-1}; \tag{25}$$

$$h_{k+1} = \frac{2M[f(x_k) - m]}{(M+1)\|\widehat{g}_{\varphi_k}(y_k)\|} \tag{26}$$

m выбрано большим или равным m^*, то последовательность $\{h_k\}$ является ограниченной и для произвольного $\varepsilon > 0$ найдется $\overline{k}$ такое, что $f(x_{\overline{k}}) < m + \varepsilon$ (счет прекращается, если на некотором шаге $f(x_k) \le m$) : если m выбрано меньшим m^*, то последовательность $\{h_k\}$ является неограниченной.

Доказательство. Если $m \leq m^*$, то рассмотрим функцию

$$f_m(x) = \begin{cases} f(x), & \text{если } f(x) \geq m; \\ m, & \text{если } f(x) < m. \end{cases}$$

Функция $f_m(x)$ удовлетворяет условиям теоремы 2 работы [1] с $N = 1$. Так как для $f_m(x)$ ОГСРП сходится по функционалу, то для некоторого $\overline{k}$ $f(x_{\overline{k}}) < m + \varepsilon$. Далее,

$$\frac{[f_m(x_k) - m]}{\|\widehat{g}_{\varphi_k}(y_k)\|} \leq \|A_k(x_k - x_k^*)\|,$$

где x_k^* – произвольная точка, для которой $f_m(x) = m$. Из теоремы 2 [1] следует, что x_k^* можно определять таким образом, что $\{\|A_k(x_k - x_k^*)\|\}$ будет невозрастающей по k. Отсюда следует ограниченность $\{h_k\}$.

Если m выбрано меньшим m^*, то возможны два случая:

1) $\sup\limits_k f(x_k) < \infty$. Тогда выполняются условия леммы 1:

$$\lim_{k\to\infty} \frac{f(x_k) - m}{\|\widehat{g}_{\varphi_k}(y_k)\|} \geq \lim_{k\to\infty} \frac{m^* - m}{\|g_{\varphi_k}(y_k)\|} = \infty,$$

так как

$$\lim_{k\to\infty} \|\widehat{g}_{\varphi_k}(y_k)\| = 0.$$

2) $\sup f(x_k) = \infty$.

$$\sup_k \frac{f(x_k) - m}{\|\widehat{g}_{\varphi_k}(y_k)\|} \geq \sup_k \frac{f(x_k) - m^*}{\|\widehat{g}_f(x_k)\|}. \tag{27}$$

Рассмотрим луч, исходящий из точки x_k и соединяющий ее с некоторой точкой минимума $f(x)$. Пусть x_k^* – первая точка минимума на этом луче. Применим неравенство (24):

$$\frac{f(x_k) - m^*}{\|\widehat{g}_f(x_k)\|} \geq \frac{1}{M} \cdot \frac{f'_{x_k - x_k^*}(x_k) \cdot \|x_k - x_k^*\|}{\|\widehat{g}_f(x_k)\|} =$$
$$= \frac{(\widehat{g}_f(x_k), x_k - x_k^*)\,\|x_k - x_k^*\|}{M\,\|x_k - x_k^*\|\,\|\widehat{g}_f(x_k)\|} = \frac{(\widehat{g}_f(x_k), x_k - x_k^*)}{\|\widehat{g}_f(x_k)\|}. \tag{28}$$

Но выражение $(\widehat{g}_f(x_k), x_k - x_k^*)/\,\|\widehat{g}_f(x_k)\|$ есть расстояние от некоторой точки минимума x_k^* до опорной гиперплоскости к поверхности уровня

$f(x) = f(x_k)$. Так как $f(x)$ определена во всем пространстве, область ее минимумов ограничена, а последовательность $f(x_k)$ неограничена, то

$$\sup \frac{(\widehat{g}_f(x_k), x_k - x_k^*)}{\|\widehat{g}_f(x_k)\|} = \infty.$$

Отсюда и из (27), (28) следует, что $\sup h_k = \infty$. Доказательство теоремы завершено.

Теорема 1 позволяет нам построить алгоритм минимизации функции $f(x)$, удовлетворяющей условию этой теоремы, при неизвестном m^*. Этот алгоритм будет состоять из последовательности этапов.

1 этап. Выбираем $x_0^{(1)}$, константы H, $\Delta_1 > 0$ и применяем ОГСРП, как описано в условиях теоремы 1, принимая $m = m_1 = f(x_0^{(1)}) - \Delta_1$. В силу теоремы 1 на некотором шаге $k = k_1$ должно произойти хотя бы одно из двух событий:

а) $f(x_{\overline{k}_1}^{(1)}) < f(x_0) - \dfrac{\Delta_1}{2}$;

б) $h_{\overline{k}_1+1}^{(1)} > H$.

Запоминаем $\min\limits_{0 \le k \le \overline{k}_1} f(x_k^{(1)})$. Пусть он достигается на k_1^*.

(r + 1)-й этап (r = 1, 2, ...). Если на r-ом этапе произошло событие $f\big(x_{\overline{k}_r}^{(r)}\big) < f\big(x_0^{(r)}\big) - \dfrac{\Delta_r}{2}$, то выбираем $x_0^{(r+1)} = x_{\overline{k}_r}$; $\Delta_{r+1} = \Delta_r$. Если на r-ом этапе это событие не произошло, а произошло событие $h_{\overline{k}_r+1}^{(r)} > H$, то выбираем $x_0^{(r+1)} = x_{k_r^*}$, $\Delta_{r+1} = \dfrac{\Delta_r}{2}$. В обоих случаях выбираем $m = m_{r+1} = f(x_0^{(r+1)}) - \Delta_{r+1}$ и применяем ОГСРП, как описано в теореме 1.

На некотором шаге $k = \overline{k}_{r+1}$ произойдет хотя бы одно из двух событий:

а) $f(x_{\overline{k}_{r+1}}^{(r+1)}) < f(x_0^{(r+1)}) - \dfrac{\Delta_{r+1}}{2}$;

б) $h_{\overline{k}_{r+1}+1} > H$.

Запоминаем $\min\limits_{0 \le k \le \overline{k}_{r+1}} f(x_k^{(r+1)})$ и номер k_{r+1}^*, на котором он достигается.

Переходим к $(r+2)$-му этапу и далее по индукции.

Покажем, что $\lim_{r\to\infty} f(x_0^{(r)}) = m^*$. Так как $f(x_0^{(r)})$ – невозрастающая и ограниченная снизу последовательность, то $\lim_{r\to\infty} f(x_0^{(r)})$ существует. Из описания алгоритма и доказательства теоремы 1 ясно, что при достаточно малом Δ_r случай б) может произойти только тогда, когда $m_r < m^*$. С другой стороны, случай а) может произойти после случая б) только конечное число раз подряд. Таким образом, $\lim_{r\to\infty} \Delta_r = 0$. Так как начиная с достаточно малых Δ_r, если на r-ом этапе произошел случай б), $f(x_0^{(r)}) - \Delta_r$ служит оценкой снизу для m^*, и $\lim_{r\to\infty} f(x_0^{(r)}) = m^*$.

4. О скорости сходимости метода ОГСРП

Теорема 2. *Пусть в процессе выполнення* ОГСРП *соблюдаются следующие условия:*

$$\text{а)}\quad \alpha_k = \alpha > 1 \quad (k = 1, 2, \ldots); \tag{29}$$

$$\text{б)}\quad \inf_{\{x^* : f(x^*) = m^*\}} \|A_k (x_k - x^*)\| \le r \tag{30}$$

(r – *положительная константа*). *Тогда существует бесконечная подпоследовательность* $\{x_{k_i}\}$ $(k_i < k_{i+1};\quad i = 1, 2, \ldots)$ *такая, что* $f(x_{k_i}) - m^* \le C \cdot \alpha^{-k_i/n}$.

Доказательство. Операторы A_k можно представить в виде произведения ортогонального и симметричного положительно определенного оператора [2]: $A_k = O_k \cdot S_k$.

Собственные числа оператора S_k $\lambda_j^{(k)} \ge 1$, $j = 1, 2, \ldots, n$. (Мы будем производить нумерацию собственных чисел в порядке их возрастания). Рассмотрим ортонормированную систему собственных векторов оператора S_k : $\{s_1^{(k)}, \ldots, s_n^{(k)}\}$, так что

$$S_k s_j^{(k)} = \lambda_j^{(k)} s_j^{(k)} \quad (i = 1, 2, \ldots, n).$$

Так как $A_k = R_\alpha(\xi_k) \cdot R_\alpha(\xi_{k-1}) \ldots R_\alpha(\xi_1)$, причем произведение собственных чисел оператора $R_\alpha(\xi) = \alpha$, то произведение собственных чисел оператора S_k равно α^k.

$$\prod_{j=1}^{n} \lambda_j^{(k)} = \alpha_k.$$

Отсюда $\lambda_n^{(k)} \le \alpha^{k/n}$.

Пусть $\lambda_j^{(k)} = c_j^{(k)}\alpha^{k/n}$. Тогда $c_n^{(k)} \geq 1$, $c_1^{(k)} \leq 1$, так как $\prod\limits_{j=1}^{n} c_j^{(k)} = 1$. Если последовательность $\{c_n^{(k)}\}$ ограничена числом $c_{\max}$, то выполняется $\lambda_1^{(k)} \geq c_{\max}^{-(n-1)} \cdot \alpha^{k/n}$, и для произвольного $x \in E_n$ и $k = 1, 2, \ldots$ имеем $k_x\left(A_k\right) \geq c_{\max}^{-(n-1)} \cdot \alpha^{k/n}$. Но тогда из (30) вытекает, что

$$\inf_{\{x^*:\, f(x^*)=m^*\}} \|x_k - x^*\| \leq r \cdot c_{\max}^{n-1} \cdot \alpha^{-k/n}. \tag{31}$$

Из выпуклости $f(x)$ следует равномерная ограниченность производных по направлению в ограниченной области, откуда получаем

$$f\left(x_k\right) - m^* \leq c \cdot \alpha^{-n/k}$$

Таким образом, если последовательность $\{c_n^{(k)}\}$ ограничена, то теорема имеет место.

Рассмотрим случай, когда $\{c_n^{(k)}\}$ не ограничена. Тогда найдется бесконечная последовательность чисел $k_1 < k_2 < \ldots < k_i < \ldots$ таких, что $c_n^{(k_i+1)} > c_n^{(k_i)}$, т. е.

$$\lambda_n^{(k_i+1)} > \sqrt[n]{\alpha} \cdot \lambda_n^{(k_i)}; \tag{32}$$

$$\|A_{k_i+1} s_n^{(k_i+1)}\| > \sqrt[n]{\alpha} \cdot \lambda_n^{(k_i)}. \tag{33}$$

Рассмотрим разложение единичного вектора $s_n^{(k_i+1)}$ по векторам $s_1^{(k_i)}, \ldots, s_n^{(k_i)}$:

$$\begin{aligned} s_n^{(k_i+1)} &= \sum_{j=1}^{n} \mu_j s_j^{(k_i)}; \quad \sum_{j=1}^{n} \mu_j^2 = 1; \\ A_{k_i} s_n^{(k_i+1)} &= \sum_{j=1}^{n} \lambda_j^{(k_i)} \mu_j O_{k_i} s_j^{(k_i)} = \sum_{j=1}^{n} \lambda_j^{(k_i)} \mu_j s_j'. \end{aligned}$$

Так как O_{k_i} – ортогональный оператор, то система векторов $\{O_{k_i} s_j^{(k_i)}\} = \{s_j'\}$, $j = 1, 2, \ldots, n$, является ортонормированной.

$$A_{k_i+1} s_n^{(k_i+1)} = R_\alpha\left(\xi_{k_i+1}\right) \cdot A_{k_i} s_n^{(k_i+1)} = R_\alpha\left(\xi_{k_i+1}\right) \left(\sum_{j=1}^{n} \lambda_j^{(k_i)} \mu_j s_j'\right);$$

$$\|A_{k_i+1} s_n^{(k_i+1)}\| = \sqrt{\|A_{k_i} s_n^{(k_i+1)}\|^2 + (\alpha^2 - 1)\left(\xi_{k_i+1}, A_{k_i} s_n^{(k_i+1)}\right)^2} \leq$$

$$\le \sqrt{\lambda_n^{(k_i)^2} + (\alpha^2 - 1)\left(\sum_{j=1}^{n} \lambda_j^{(k_i)} \mu_j\left(\xi_{k_i+1}, s_j'\right)\right)^2} =$$
$$= \lambda_n^{(k_i)} \sqrt{1 + (\alpha^2 - 1)\left\{\sum_{j=1}^{n} \frac{\lambda_j^{(k_i)} \mu_j}{\lambda_n^{(k_i)}}\left(\xi_{k_i+1}, s_j'\right)\right\}^2}. \tag{34}$$

Пусть ϱ – выражение, стоящее в фигурной скобке (34):

$$\|A_{k_i+1} s_n^{(k_i+1)}\| \le \lambda_n^{(k_i)} \sqrt{1 + (\alpha^2 - 1)\,\varrho^2} \le$$
$$\le \lambda_n^{(k_i)} \sqrt{\left(1 + \frac{\alpha^2}{2}\varrho^2\right)^2} = \lambda_n^{(k_i)} \left(1 + \frac{\alpha^2 - 1}{2}\varrho^2\right).$$

Пусть $c_1 = \sqrt{\dfrac{2\,(\sqrt[n]{\alpha} - 1)}{\alpha^2 - 1}}$. Если для всех $j = 1, 2, \ldots, n$

$$\left|\left(\xi_{k_i+1}, s_j'\right)\right| \le c_1 \cdot \frac{\lambda_n^{(k_i)}}{\lambda_j^{(k_i)} n},$$

то $\varrho \le c_1$, $\|A_{k_i+1} s_n^{(k_i+1)}\| \le \sqrt[n]{\alpha}\lambda_n^{(k_i)}$, что противоречит (33). Значит, найдется такое j, что $|\left(\xi_{k_i+1}, s_j'\right)| > \dfrac{c_1 \lambda_n^{(k_i)}}{n\lambda_j^{(k_i)}}$, т. е.

$$\left|\left(s_j', \widehat{g}_{\varphi_{k_i}}(y_{k_i})\right)\right| > \frac{c_1 \lambda_n^{(k_i)}}{n\lambda_j^{(k_i)}} \left\|\widehat{g}_{\varphi_{k_i}}(y_{k_i})\right\|.$$

Но $(s_j', \widehat{g}_{\varphi_{k_i}}(y_{k_i}))$ – обобщенная производная по направлению s_j', которая не превышает

$$\sup_{x \in R_r} \widehat{g}_f(x) \cdot \frac{1}{\lambda_j^{(k_i)}},$$

где

$$R_r = \left\{x \;:\; \inf_{\{x^*:\, f(x^*) = m^*\}} \|x - x^*\| \le r\right\}.$$

Отсюда

$$\|\widehat{g}_{\varphi_{k_i}}(y_{k_i})\| \le \sup_{x \in R_r} \widehat{g}_f(x) \cdot n\lambda_j^{(k_i)} \cdot \left(\lambda_j^{(k_i)} \cdot c_1 \lambda_n^{(k_i)}\right)^{-1} =$$

$$= c_2 \left(\lambda_n^{(k_i)}\right)^{-1} \leq c_3 \cdot \alpha^{-k_i/n};$$

$$f\left(x_{k_i}\right) - m^* \leq r \|\widehat{g}_{\varphi_{k_i}}\left(y_{k_i}\right)\| \leq c \cdot \alpha^{-k_i/n}.$$

Что и требовалось доказать.

Заключение

Как видно из теоремы 2, если процесс ОГСРП обладает определенными свойствами, то он обеспечивает сходимость к минимуму по функционалу со скоростью геометрической прогрессии. В частности, необходимые свойства процесса ОГСРП обеспечиваются в условиях теоремы 1, если $m = m^*$ (см. теоремы 2 и 3 из статьи [1]). m^* нам известно, когда мы сводим к задаче минимизации задачу решения систем уравнений или неравенств. Таким образом, метод ОГСРП должен оказаться эффективным при решении систем нелинейных уравнений и неравенств. Существенной особенностью этого метода является то, что гарантированный знаменатель геометрической прогрессии $q = \dfrac{1}{\sqrt[n]{\alpha}}$ не зависит от степени «овражности» функции $f(x)$, что позволяет с успехом использовать его для решения систем уравнений, близких к вырожденным.

При увеличении n эффективность метода, вообще говоря, уменьшается как в связи с тем, что q становится близким к 1, так и из-за того, что требуется большая память для запоминания матрицы B_k. Предложенную в данной статье модификацию алгоритма, когда m^* неизвестно, можно применять при решении минимаксных задач, при решении общей задачи выпуклого программирования, сводя ее к решению последовательности систем выпуклых неравенств или пользуясь методом штрафных функций, при использовании схем разложения.

Литература

1. Шор Н. З. Использование операции растяжения пространства в задачах минимизации выпуклых функций // Кибернетика. – 1970. – № 1. – С. 6–12.

2. Гантмахер Ф. Р. Теория матриц. – М.: ГИТТЛ, 1953.

Некоторые вопросы сходимости обобщенного градиентного спуска

Н. З. Шор, П. Р. Гамбурд
Кибернетика. – 1971. – № 6. – С. 82–84.

В работе [1] дано доказательство того, что метод обобщенного градиентного спуска при некоторых предположениях сходится со скоростью геометрической прогрессии.

В данной заметке дается доказательство этого факта при значительно ослабленных ограничениях, накладываемых на функционал. Кроме того, уточняется показатель сходимости для функций со слабо вытянутыми линиями уровня. Теоремы 1 и 2 содержатся в диссертации Н. З. Шора [4]. Теоремы 3, 4 получены П. Р. Гамбурдом.

Приведем некоторые определения [1]. Пусть $u = f(x)$ – выпуклая функция, определенная на n-мерном евклидовом пространстве E_n. Ω – область минимумов этой функции. Вектор $g_f(x_0)$, удовлетворяющий условию

$$f(x) - f(x_0) \geq (g_f(x_0), x - x_0), \quad x \in E_n, \tag{1}$$

называется обобщенным градиентом функции $f(x)$ в точке x_0. Процедура построения последовательности $\{x_k\}$ вида

$$x_{k+1} = x_k - h_{k+1}(x_k)\, g_f(x_k), \quad k = 0, 1, 2, \ldots, \tag{2}$$

называется методом обобщенного градиентного спуска. Обозначим через $x^*(x)$ точку из Ω, удовлетворяющую условию

$$\|x - x^*(x)\| = \min_{y \in \Omega} \|x - y\|\,. \tag{3}$$

Определим два семейства множеств:

$$\begin{aligned} L_x &= \{y : f(y) = f(x)\}, \\ L_{g_f(x)} &= \{y : (g_f(x), y - x) = 0\}. \end{aligned}$$

($L_{g_f(x)}$ определено для $g_f(x) \neq 0$, т. е. $x \notin \Omega$). Пусть $x^* \in \Omega$, $x \notin \Omega$. Представим вектор $x - x^*$ следующим образом:

$$x - x^* = a_{x^*}(x) \frac{g_f(x)}{\|g_f(x)\|} + r_x, \quad \text{где} \quad (r_x, g_f(x)) = 0. \tag{4}$$

В [1] показано, что

$$a_{x^*}(x) = \min_{y \in L_{g_f(x)}} \|y - x^*\| \geq \min_{y \in L_x} \|y - x^*\| = b_{x^*}(x). \tag{5}$$

Теорема 1. *Пусть для всех $x \in E_n$ выполняется следующее неравенство:*

$$(g_f(x), x - x^*(x)) \geq \cos\varphi \, \|g_f(x)\| \, \|x - x^*(x)\| \,, \tag{6}$$

где $\dfrac{\pi}{4} \leq \varphi < \dfrac{\pi}{2}$. *Тогда при заданном x_0, если выберем величину*

$$h_1 \geq \|x^*(x_0) - x_0\| \cos\varphi \tag{7}$$

и определим $\{h_k\}$ в соответствии с рекуррентной формулой

$$h_{k+1} = h_k \sin\varphi, \quad k = 1, 2, \ldots, \tag{8}$$

и $\{x_k\}$ по формуле

$$x_{k+1} = x_k - h_{k+1} \frac{g_f(x_k)}{\|g_f(x_k)\|}, \tag{9}$$

то либо при некотором $k = \overline{k}$, $g_f(x_k) = 0$, либо при всех $k = 0, 1, 2, \ldots$ будет выполняться неравенство

$$\|x_k - x^*(x_k)\| \leq \frac{h_{k+1}}{\cos\varphi}. \tag{10}$$

Доказательство. Для $k = 0$ неравенство (10) выполняется. Пусть оно выполняется для $k = r$ и $g_f(x_r) \neq 0$. Докажем его выполнение для $k = r + 1$. Из (3) и (7)

$$\|x_{r+1} - x^*(x_{r+1})\|^2 \leq$$

$$\leq \|x_r - x^*(x_r)\|^2 - 2h_{r+1}\left(x_r - x^*(x_r), \frac{g_f(x_r)}{\|g_f(x_r)\|}\right) + h_{r+1}^2. \tag{11}$$

Воспользовавшись последовательно неравенствами (6) и (10), получаем

$$\|x_{r+1} - x^*(x_{r+1})\|^2 \leq \|x_r - x^*(x_r)\|^2 (1 - 2\cos^2\varphi) + h_{r+1}^2 \leq$$

$$\leq \frac{h_{r+1}^2}{\cos^2\varphi}(1 - 2\cos^2\varphi) + h_{r+1}^2 = h_{r+1}^2 \left(\frac{\sin\varphi}{\cos\varphi}\right)^2. \tag{12}$$

(Здесь мы использовали то, что для $\frac{\pi}{4} \le \varphi < \frac{\pi}{2}$, $1 - 2\cos^2\varphi \ge 0$.)
Из (8) и (12) сразу получаем $\|x_{r+1} - x^*(x_{r+1})\| \le \frac{h_{r+2}}{\cos\varphi}$, что и требовалось доказать.

Замечание. Из доказательства очевидно, что для справедливости теоремы 1 достаточно выполнения условия (6) лишь для точек последовательности $\{x_k\}$.

Из теоремы 1 непосредственно следует теорема 2 [1].

Теорема 2. *Пусть выпуклая функция $f(x)$ определена на E_n, x^* – единственная точка минимума и заданы: точка $x_0 \in E_n$, числа $\sigma \ge \sqrt{2}$ и $h_1 \ge \frac{\|x_0 - x^*\|}{\sigma}$. Рассмотрим множество $Y = \{y : \|y - x^*\| \le \sigma h_1\}$. Если для любой пары точек $x, z \in Y$ и такой, что $f(x) = f(z) \ne f(x^*)$ выполняется условие*

$$\frac{\|x - x^*\|}{\|z - x^*\|} \le \sigma, \tag{13}$$

то последовательность $\{x_k\}$, образованная с помощью рекуррентных формул $x_{k+1} = x_k - h_{k+1} \cdot \frac{g_f(x_k)}{\|g_f(x_k)\|}$ и $h_{k+1} = h_k \frac{\sqrt{\sigma^2 - 1}}{\sigma}$, сходится к x^* со скоростью геометрической прогрессии: $\|x_k - x^*\| \le h_{k+1}\sigma$, за исключением того случая, когда для некоторого $k = \overline{k}$ $g_f(x_k) = 0$, т. е. $x_{\overline{k}} = x^*$.

Доказательство. Достаточно показать, что

$$(g_f(x_k), x_k - x^*) \ge \frac{1}{\sigma} \|g_f(x_k)\| \, \|x_k - x^*\|. \tag{14}$$

Воспользовавшись (4) и (5), получаем

$$(g_f(x_k), x_k - x^*) = a_{x^*}(x_k) \, \|g_f(x_k)\| \ge b_{x^*}(x_k) \, \|g_f(x_k)\|.$$

Из определения $b_{x^*}(x_k)$ и (13) получим $b_{x^*}(x_k) \ge \frac{1}{\sigma} \|x_k - x^*\|$. Откуда вытекает справедливость (14).

Замечание. Величина σ отражает в какой-то мере степень «вытянутости» поверхностей уровня L_x. В качестве примера рассмотрим частный случай минимизации положительно определенных квадратичных форм: $f(x) = (Ax, x)$.

Пусть $\underline{\lambda}, \overline{\lambda}$ – наименьшее и наибольшее собственные числа оператора A. Определим минимальное значение величины

$$\frac{(g_f(x), x)}{\|g_f(x)\| \, \|x\|} = \frac{(Ax, x)}{\|Ax\| \, \|x\|}, \quad \|x\| \ne 0.$$

Для этого зафиксируем произвольный несобственный вектор x_0 и рассмотрим действие оператора $P_{x_0}A$ на подпространство $E(x_0)$ размерности 2, порожденное векторами x_0 и Ax_0, где P_{x_0} – оператор проектирования на подпространство $E(x_0)$. Легко увидеть, что $P_{x_0}A$ определенный на подпространстве $E(x_0)$, является положительно определенным оператором и его собственные числа λ_1, λ_2 (записанные в неубывающем порядке) удовлетворяют следующим неравенствам: $\underline{\lambda} \leq \lambda_1 \leq \lambda_2 \leq \overline{\lambda}$.

Пусть x_1, x_2 – ортонормированная система собственных векторов оператора $P_{x_0}A$ в $E(x_0)$, и пусть $\|x_0\| = 1$. Представим x_0 в виде $x_0 = \alpha_1 x_1 + \alpha_2 x_2$, причем $\alpha_1^2 + \alpha_2^2 = 1$. Тогда

$$\frac{(Ax_0, x_0)}{\|Ax_0\|\,\|x_0\|} = \frac{(P_{x_0}Ax_0, x_0)}{\|P_{x_0}Ax_0\|\,\|x_0\|} - \frac{\alpha_1^2\lambda_1 + \alpha_2^2\lambda_2}{\sqrt{\alpha_1^2\lambda_1^2 + \alpha_2^2\lambda_2^2}}.$$

Минимизируя правую часть при условии $\alpha_1^2 + \alpha_2^2 = 1$, получаем неравенство $\dfrac{(Ax_0, x_0)}{\|Ax_0\|\,\|x_0\|} \geq \dfrac{2\sqrt{\lambda_1\lambda_2}}{\lambda_1 + \lambda_2}$, которое превращается в равенство при $\alpha_1^2 = \dfrac{\lambda_2}{\lambda_1 + \lambda_2}$, $\alpha_2^2 = \dfrac{\lambda_1}{\lambda_1 + \lambda_2}$. Отсюда сразу получается, что

$$\min_{x \in E_n, x \neq 0} \frac{(Ax, x)}{\|Ax\|\,\|x\|} = 2\frac{\sqrt{\overline{\lambda}\underline{\lambda}}}{\overline{\lambda} + \underline{\lambda}}$$

и достигается при $x = \sqrt{\frac{\overline{\lambda}}{\underline{\lambda}+\overline{\lambda}}}s_1 + \sqrt{\frac{\underline{\lambda}}{\underline{\lambda}+\overline{\lambda}}}s_2$, где s_1,s_2 – собственные нормированные вектора оператора A, соответствующие минимальному и максимальному собственным числам.

Обратимся теперь к теореме 1. Для нашего примера мы можем принять $\cos\varphi = \dfrac{2\sqrt{\underline{\lambda}\overline{\lambda}}}{\underline{\lambda} + \overline{\lambda}}$, при этом $\sin\varphi = \dfrac{\overline{\lambda} - \underline{\lambda}}{\overline{\lambda} + \underline{\lambda}}$ и, если $\dfrac{2\sqrt{\underline{\lambda}\overline{\lambda}}}{\underline{\lambda} + \overline{\lambda}} \leq \dfrac{1}{\sqrt{2}}$, обобщенный градиентный спуск, описанный в условии этой теоремы, дает процесс, сходящийся к точке минимума по норме со скоростью геометрической прогрессии со знаменателем $q = \dfrac{\overline{\lambda} - \underline{\lambda}}{\overline{\lambda} + \underline{\lambda}} = \dfrac{\rho - 1}{\rho + 1}$, где ρ – число обусловленности ($\rho = \overline{\lambda}/\underline{\lambda}$) [2], а по функционалу – со знаменателем q^2.

Замечание. Наши рассуждения почти полностью сохраняют силу, если A – неотрицательно определенный оператор, $\|A\| = 0$, только под $\underline{\lambda}$ нужно понимать наименьшее положительное собственное число.

Сопоставим полученный выше результат с результатом Л. В. Канторовича [3] о скорости сходимости метода наискорейшего спуска для квадратичных функционалов. Этот метод сходится по функционалу со скоростью геометрической прогрессии со знаменателем q^2.

Однако схема метода наискорейшего спуска для квадратичных функционалов (или для решения систем линейных уравнений с симметричной положительно определенной матрицей) гораздо сложнее, чем схема, описанная в условиях теоремы 1, требует $\approx 3n$ рабочих ячеек – в три раза больше, чем последняя схема, что имеет значение для задач большого объема (см. [2], §70).

Таким образом, если мы имеем возможность получить достаточно точную оценку сверху $\widehat{\rho}$ для числа обусловленности ρ и примем $\sin\varphi = \dfrac{\widehat{\rho}-1}{\widehat{\rho}+1}$, то алгоритм, описанный в теореме 1, может оказаться практически вполне конкурентно способным по сравнению с методом наискорейшего спуска.

Заметим, что для квадратичных функционалов $f(x) = (Ax, x)$ величина σ, фигурирующая в теореме 2, может быть взята равной $\sqrt{\rho}$, при этом мы получаем сходимость со скоростью геометрической прогрессии со знаменателем $\overline{q} = \dfrac{\sqrt{\sigma^2-1}}{\sigma} = \dfrac{\sqrt{\rho-1}}{\sqrt{\rho}}$.

Теорема 3. *Пусть для всех $x \in E_n$ выполняется условие (6), где $0 \le \varphi < \dfrac{\pi}{3}$. Тогда при заданном x_0, если мы выберем величину*

$$h_0 \ge \|x^*(x_0) - x_0\|, \tag{15}$$

число $q < 1$ и удовлетворяющее условию

$$1 - 2\cos\varphi q \le 0, \tag{16}$$

определим $\{h_k\}$ в соответствии с рекуррентной формулой

$$h_{k+1} = h_k q, \quad k = 0, 1, 2, \dots, \tag{17}$$

а $\{x_k\}$ по формуле (9), то либо при некотором $k = \overline{k}$, $\quad g_f(x_k) = 0$, либо для всех $k = 0, 1, 2, \dots$ будет выполняться неравенство

$$\|x_k - x^*(x_k)\| \le h_k. \tag{18}$$

Доказательство. Для $k = 0$ (18) выполняется. Пусть оно выполняется для $k = r$ и $g_f(x_r) \ne 0$. Докажем его выполнение для $k = r + 1$. Воспользовавшись последовательно (11), (6), (18), получаем

$$
\begin{aligned}
& \|x_{r+1} - x^*(x_{r+1})\|^2 \le \\
\le \;& \|x_r - x^*(x_r)\|^2 - 2h_{r+1}\cos\varphi\,\|x_r - x^*(x_r)\| + h_{r+1}^2 \le \\
\le \;& \|x_r - x^*(x_r)\|^2 - 2\cos\varphi\, q\,\|x_r - x^*(x_r)\|^2 + h_{r+1}^2 = \\
= \;& (1 - 2\cos\varphi\, q)\,\|x_r - x^*(x_r)\|^2 + h_{r+1}^2 \;\le\; h_{r+1}^2,
\end{aligned}
$$

(в заключении мы воспользовались условием (16)), т. е. то, что требовалось доказать.

Аналогично теореме 2 доказывается теорема 4.

Теорема 4. *Пусть выпуклая функция* $f(x)$ *определена на* E_n, x^* – *единственная точка минимума, задана точка* $x_0 \in E_n$*, числа* $1 \le \sigma < 2$ *и* $q < 1$ *такое, что* $1 - 2\dfrac{1}{\sigma}q \le 0$*. Если* $f(x)$ *удовлетворяет условию* (13), *то последовательность* $\{x_k\}$*, образованная с помощью рекуррентных формул*

$$
x_{k+1} = x_k - h_{k+1}\frac{g_f(x_k)}{\|g_f(x_k)\|}, \quad h_{k+1} = h_k q, \quad k = 0, 1, 2, \ldots,
$$

сходится к x^* *со скоростью геометрической прогрессии* $\|x_k - x^*\| \le$ $\le h_k$*, за исключением того случая, когда для некоторого* $k = \overline{k}$, $g_f(x_{\overline{k}}) = 0$*, т. е.* $x_{\overline{k}} = x^*$.

Литература

1. Шор Н. З. Обобщенный градиентный спуск // Труды I зимней школы по математическому программированию (г. Дрогобыч). – М., 1969. – Вып. 3. – С. 578–585.

2. Фаддеев Д. К., Фаддеева В. Н. Вычислительные методы линейной алгебры. – М.: Физматгиз, 1960.

3. Канторович Л. В. Об одном эффективном методе решения экстремальных задач для квадратичных функционалов // ДАН СССР. – 1945. – **48**. – № 7.

4. Шор Н. З. Методы минимизации недифференцируемых функций и их приложения. Автореферат докторской диссертации. – К., 1970. – 43 с.

О решении минимаксных задач методом обобщенного градиентного спуска с растяжением пространства

Н. З. Шор, Л. П. Шабашова
Кибернетика. – 1972. – № 5. – С. 54–58.

Обширный класс практически важных задач из области оптимального планирования, оптимального управления и исследования операций сводится к задаче отыскания минимакса и максимина. В данной работе рассматриваются задачи следующего вида:

найти

$$\min_{x\in E_n} F(x) = \min_{x\in E_n} \max_{i\in\{1,2,\ldots,m\}} f_i(x),$$

где $f_i(x),\ i = 1, 2, \ldots, m$ – выпуклые функции, определенные на евклидовом пространстве E_n.

К такой постановке можно свести общую задачу выпуклого программирования, задачи чебышевского приближения, задачи математического программирования при использовании метода штрафных функций, задачи минимизации кусочно-линейных функций и другие задачи оптимизации.

В связи с важностью минимаксных задач большое внимание в настоящее время уделяется разработке методов их решения. Ряд таких методов для определенных классов задач предложен в работах В. Ф. Демьянова, Б. Н. Пшеничного, Е. Г. Евтушенко и др. [4]–[7].

В принципе задача отыскания минимакса может быть сведена к решению следующей задачи выпуклого программирования:

найти $\min y$ при ограничениях

$$f_i(x) - y \le 0, \quad i = 1, 2, \ldots, m.$$

Для решения этой задачи могут быть использованы итеративные методы возможных направлений [8]. Однако такой путь решения задач связан с определенными вычислительными трудностями:

во-первых, нахождение минимума при данных ограничениях методами возможных направлений связано с необходимостью решения на каждой итерации ряда вспомогательных задач линейного или квадратичного программирования;

во-вторых, для обеспечения сходимости методов возможных направлений нужно использовать меры предосторожности против «зигзагообразного» движения;

в-третьих, на каждом шаге метода возможных направлений нужно определять минимум по направлению – довольно трудоемкая процедура, так как приходится определять минимум негладкой функции в условиях, когда определение значения функции требует значительной вычислительной работы. И, наконец, скорость сходимости этих методов слабо изучена как теоретически, так и экспериментально.

Более простым с вычислительной точки зрения является метод обобщенного градиентного спуска, который в принципе позволяет решать выпуклые минимаксные задачи [1]. Однако скорость сходимости этого метода недостаточно высока, особенно при решении задач овражного типа, которые характерны при отыскании минимаксов. В этом случае минимум достигается при равенстве значений нескольких функций, причем число таких функций в регулярном случае определяет размерность оврага (размерность оврага равна $(n - m_1 + 1)$, где m_1 – число функций, на которых достигается минимаксное значение).

Ускорение сходимости метода обобщенного градиентного спуска может быть достигнуто за счет растяжения пространства в направлении обобщенного градиента или в направлении разности обобщенных градиентов, вычисленных в двух последовательных точках [2], [3]. При использовании обобщенного градиентного спуска с растяжением пространства определение минимизирующей последовательности связано с вычислением обобщенного градиента функции $F(x)$ в каждой точке последовательности и применением оператора растяжения пространства.

Обобщенный градиент $g_F(x)$, вычисленный в точке x_0, должен удовлетворять условию

$$F(x) - F(x_0) \geq (g_F(x_0), x - x_0)$$

для любого $x \in E_n$. В случае минимаксных задач $g_F(x_0)$ можно брать равным $g_{f_{i^*}}(x_0)$, где f_{i^*} – одна из функций, на которой достигается максимум. Оператор растяжения пространства $R_\alpha(\xi)$ действует на произвольный вектор $x \in E_n$, представленный в виде

$$x = \gamma_\xi(x)\xi + d_\xi(x),$$

при условии $(\xi, d_\xi(x)) = 0$, $\|\xi\| = 1$, таким образом:

$$R_\alpha(\xi)x = \alpha\gamma_\xi(x)\xi + d_\xi(x),$$

где ξ – вектор единичной длины, в направлении которого производится растяжение пространства, α – коэффициент растяжения пространства ($\alpha \geq 0$).

В работе рассматривается применение обобщенного градиентного спуска с растяжением пространства в направлении разности обобщенных градиентов в двух последовательных точках минимизирующей последовательности к решению минимаксных задач и нахождению минимума негладких функций «овражного» типа. Описываются две модификации метода.

1. Первый алгоритм

Движение начинаем из произвольной точки $x_0 \in E_n$. При этом полагаем $\widetilde{g}_0 = (0, 0, \ldots, 0) \in E_n$ и $B_0 = I_n$, где I_n – единичная матрица размерности $n \times n$. (Смысл вектора $\widetilde{g}$ и матрицы B будет раскрыт ниже.)

Предположим, что выполнено k итераций ($k = 1, 2, \ldots$), в результате чего определены точки $x_1, x_2, \ldots, x_k$, а также вектор $\widetilde{g}_k$ и матрица B_k. Для определения точки x_{k+1} производим следующие действия.

1. Определяем значение функции в k-ой точке последовательности:

$$F\left(x_k\right) = \max_{i=\{1,2,\ldots,m\}} f_i\left(x_k\right) = f_{i^*}\left(x_k\right).$$

2. Вычисляем обобщенный градиент функции $F(x)$ в точке x_k:

$$g_F\left(x_k\right) = g_{f_{i^*}}\left(x_k\right).$$

3. Находим обобщенный градиент $\overline{g}_k$ функции $\varphi(y)$ в точке y_k в преобразованном пространстве по формуле

$$\overline{g}_k = g_{\varphi_k}\left(y_k\right) = B_k^* g_F\left(x_k\right),$$

которая получается при растяжении пространства E_n путем линейного преобразования вида $y_k = A_k x_k$ и исходя из условия

$$\varphi_k(y) = F(x) = F\left(A_k^{-1} y\right) = F\left(B_k y\right),$$

где A_k – оператор преобразования пространства; B_k – оператор, обратный оператору A_k; B_k^* – оператор, сопряженный оператору B_k.

4. Определяем разность обобщенных градиентов

$$r_k = \overline{g}_k - \widetilde{g}_k.$$

5. Находим отношение длин векторов r_k и $\widetilde{g}_k$:

$$\beta_k = \frac{\|r_k\|}{\|\widetilde{g}_k\|},$$

которое характеризует изменение направления обобщенного градиента в точке x_{k-1}, относительно направления обобщенного градиента в точке x_k в растянутом пространстве.

6. Сравниваем β_k с заданной величиной q_1:

а)если $\beta_k \leq q_1$, то определяем следующую точку минимизирующей последовательности

$$x_{k+1} = x_k - h_k B_k \frac{\widetilde{g}_k}{\|\widetilde{g}_k\|},$$

соответствующую перемещению в растянутом пространстве точки y в том же направлении и с той же величиной шага h_k, что и на k-ой итерации, после чего переходим к выполнению $(k+2)$-ой итерации;

б) если $\beta_k > q_1$, т. е. при значительном изменении направления обобщенного градиента, переходим к выполнению последующих этапов.

7. Определяем вектор ξ_{k+1}, в направлении которого будет производиться растяжение пространства:

$$\xi_{k+1} = \frac{r_k}{\|r_k\|},$$

8. Находим оператор B_{k+1}, обратный оператору A_{k+1} по формуле

$$B_{k+1} = B_k \cdot R_{1/\alpha}\left(\xi_{k+1}\right),$$

которая является следствием последовательного применения операторов растяжения пространства на каждой итерации в направлении нормированных векторов $\xi_1, \xi_2, \ldots, \xi_{k+1}$ с коэффициентом растяжения α, т. е.

$$A_{k+1} = R_\alpha\left(\xi_{k+1}\right) \cdot R_\alpha\left(\xi_k\right) \cdot \ldots R_\alpha\left(\xi_1\right) I_n.$$

Отсюда следует

$$B_{k+1} = A_{k+1}^{-1} = A_k^{-1} \cdot R_\alpha^{-1}\left(\xi_{k+1}\right) = B_k \cdot R_{1/\alpha}\left(\xi_{k+1}\right).$$

9. Определяем обобщенный градиент функции $\varphi(y)$ в точке y_k, соответствующей точке x_k, с учетом $(k+1)$ оператора растяжения пространства:

$$\widetilde{g}_{k+1} = B_{k+1}^* g_F\left(x_k\right);$$

в направлении антиградиента осуществляется движение на данной итерации.

Таблица 1.

№ итерации	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(5)}$	$F(x)$
Нулевое приближение	0	0	0	0	1	80
1	0,1119585	0,223917	0,1119585	0,1119585	1,111958	63,0894
6	1,097278	0,788586	0,871869	0,2815889	0,8040123	34,399
11	1,069281	0,7126785	0,8504548	0,8685811	1,102541	25,83837
16	1,063638	0,9135513	1,160926	1,101065	1,073642	24,69413
21	1,076656	0,9905223	1,366388	0,910783	1,116449	22,78248
26	1,129564	0,980844	1,46487	0,936742	1,115325	22,6504
31	1,13501	0,9733332	1,469467	0,9394724	1,11657	22,609
36	1,11975	0,9814588	1,475895	0,9139938	1,126114	22,60392
41	1,128545	0,978718	1,482843	0,9253958	I,123494	22,60168
46	1,122855	0,9793679	1,474063	0,919058	1,124103	22,60064
51	1,124632	0,9793914	1,477879	0,9206665	1,124184	22,60023

10. Изменяем длину шага: $h_{k+1} = h_k \cdot q_2$, где $0 < q_2 \leq 1$.

11. Определяем очередную точку минимизирующей последовательности:

$$x_{k+1} = x_k - h_{k+1} B_{k+1} \frac{\widetilde{g}_{k+1}}{\|\widetilde{g}_{k+1}\|},$$

которая соответствует точке перемещения в растянутом пространстве в направлении $-\dfrac{\widetilde{g}_{k+1}}{\|\widetilde{g}_{k+1}\|}$ на величину h_{k+1} из точки $y_k = A_k x_k$. После этого переходим к выполнению $(k+2)$-ой итерации $(k = 1, 2, \ldots)$.

Для исследования эффективности описанного алгоритма были составлены стандартные автокодовые программы для ЭВМ «Минск-22» и проведены многочисленные расчеты по решению минимаксных задач. Для проведения численных экспериментов были взяты задачи следующего вида: найти

$$\min_{x \in E_n} F(x),$$

где $F(x) = \max_{i \in \{1,2,\ldots,m\}} f_i(x)$, при $f_i(x) = A^{(i)} \sum_{j=1}^{n} \left(x^{(j)} - a_{ij}\right)^2$, где $x = \{x^{(1)}, \ldots, x^{(n)}\}$, $\{A^{(i)}\}$ – m-мерный вектор; $\{a_{ij}\}$ – матрица $m \times n$. В качестве примера приведем результаты расчетов конкретной задачи: $n = 5$; $m = 10$; $\{A^{(i)}\} = \{1;\ 5;\ 10;\ 2;\ 4;\ 3;\ 1.7;\ 2.5;\ 6;\ 3.5\}$.

Величины a_{ij} задаются следующей матрицей:

$$\begin{pmatrix} 0 & 2 & 1 & 1 & 3 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 2 & 4 & 2 & 2 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 \\ 0 & 1 & 1 & 2 & 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 3 & 2 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}.$$

Задача решалась при следующих значениях параметров: $\alpha = 3$; $q_1 = 0,9$; $q_2 = 0,95$; $h_0 = 1$. В качестве начального приближения была взята точка $x_0 = (0, 0, 0, 0, 1)$, при этом $F(x_0) = 80$. Значения минимизирующей последовательности и значения функции $F(x)$ в этих точках на соответствующих итерациях приведены в табл. 1. Точное значение x^*, при котором достигается минимум

$$F(x) = \max_{i \in \{1,\ldots,10\}} f_i(x),$$

равно:

$x_1^* = 1,12434$; $x_2^* = 0,97945$; $x_3^* = 1,47770$; $x_4^* = 0,92023$; $x_5^* = 1,12429$.

При этом

$$F\left(x^*\right) = \min_{x \in E_n} \max_{1 \le i \le 10} f_i(x) = 22,60016.$$

Как видно из табл. 1, за 51 итерацию получено приближение, у которого верны шесть знаков для $F(x)$ и $3,5$ знаков для $x^{(j)}$. Этот результат следует признать хорошим, если учесть, что функция имела двумерный овраг.

Изучение вычислительной схемы данного алгоритма и практика вычислений показали, что данный алгоритм вполне пригоден для решения минимаксных задач, с точки зрения численной реализации он является простым и обеспечивает вполне достаточную скорость сходимости. Как показали эксперименты, скорость сходимости и время решения задач зависят от выбора параметров α и q_2. Коэффициент растяжения пространства в принципе может быть любым, больше единицы. Однако, если α слишком близко к единице, то эффект растяжения слабо сказывается, что приводит к медленной сходимости, особенно для функций с многомерным оврагом. Наоборот, при больших α происходит быстрое убывание нормы матрицы и, как следствие, быстрое уменьшение длины шага в основном пространстве, что приводит к увеличению числа шагов без растяжения пространства, это также уменьшает скорость сходимости. Численный эксперимент показал, что наиболее целесообразно значение α выбирать в пределах от 2 до 4. Параметр q_2, в принципе, можно брать равным 1. Если мы выбираем $q_2 < 1$, то это приводит к уменьшению длины шага равномерно по всем направлениям (в то время как матрица неравномерно уменьшает шаг по различным направлениям). Если q_2 слишком отличен от 1, то увеличивается число «холостых» шагов (без растяжения пространства), что приводит к замедлению сходимости. С другой стороны, для многомерных задач во многих случаях целесообразно выбирать $q_2 < 1$, так как «средний» показатель уменьшения шага за счет уменьшения нормы матрицы равен $\dfrac{1}{\sqrt[n]{\alpha}}$ и при больших n может оказаться слишком близким к 1.

Вышеописанный численный пример решался при различных q_2 до получения минимального значения $F(x) = 22,60016$ при постоянных параметрах $\alpha = 3$; $q_1 = 0,9$. При $q_2 = 1$ было затрачено 69 итераций, из них с растяжением пространства – 52, отношение $\dfrac{52}{69} \approx 0,75$. При $q_2 = 0,95$ было затрачено 57 итераций, из них с растяжением пространства – 41, $\dfrac{41}{57} \approx 0,72$. При $q_2 = 0,9$ было затрачено 112 итераций, из них с растяжением пространства – 39, $\dfrac{39}{112} \approx 0,35$. Дальнейшее уменьшение

q_2 оказалось явно нецелесообразным. Эти эксперименты показывают, что, по-видимому, для каждой конкретной задачи существует асимптотически «оптимальное» q_2. Можно строить алгоритмы, которые автоматически настраиваются на это q_2. Для этого можно анализировать отношение числа итераций с растяжением пространства к общему числу итераций, и если это отношение становится меньше определенного числа, то увеличивать q_2, и наоборот, если оно слишком близко к 1, то уменьшать q_2. Аналогично можно регулировать и параметр α.

Параметр q_1 рекомендуем выбирать в пределах 0,9–1.

Применение предложенного алгоритма целесообразно при решении задач, где требуется высокая точность. При этом в оперативной памяти ЭВМ необходимо хранить матрицу B размерности $n \times n$. Поэтому в силу ограниченности объема памяти алгоритм может применяться для нахождения минимакса функций от сравнительно небольшого числа переменных (так, например, оперативная память БЭСМ-6 позволяет решать задачи размерности до 150).

При решении задач высокой размерности возникают трудности запоминания матрицы B. В таких задачах обычно минимум достигается при равенстве значений нескольких функций, число которых m_1 намного меньше n – размерности задачи. В связи с этим предлагается второй алгоритм, позволяющий избежать запоминания матриц и сократить время счета.

Дадим описание второго алгоритма.

2. Второй алгоритм

Как и в первом алгоритме, движение начинаем из произвольной точки $x_0 \in E_n$, полагая $\widetilde{g}_0 = (0, 0, ...0) \in E_n$. Предположим, что сделано k итераций ($k = 1, 2, \ldots$), в результате чего определены точки $x_1, x_2, \ldots, x_k$, вектор $\widetilde{g}_k$ и последовательность векторов $\xi_1, \xi_2, \ldots, \xi_k$, где $\xi_\nu \in E_n$, ($\nu = 1, 2, \ldots, k$), последовательность векторов единичной длины, в направлении которых производится на каждой итерации растяжение пространства E_n.

Для определения точки x_{k+1} необходимо сделать следующие вычисления.

1. $F(x_k) = \max_{1 \le i \le m} f_i(x_k)$ – максимальное значение функции в точке x_k.

2. $g_F(x_k)$ – обобщенный градиент функции $F(x)$ в точке x_k:

$$g_F(x_k) = g_{f_{i^*}}(x_k); \quad \Big(F(x_k) = f_{i^*}(x_k)\Big).$$

3. $\overline{g}_k = R_{1/\alpha}(\xi_k) R_{1/\alpha}(\xi_{k-1}) \dots R_{1/\alpha}(\xi_1) g_F(x_k)$ – обобщенный градиент функции в преобразованном пространстве. В данном алгоритме вместо матрицы B_k запоминается последовательность нормированных векторов $\xi_1, \xi_2, \dots, \xi_k$. Обобщенный градиент $\overline{g}_k$, вычисленный в точке $y_k = A_k x_k$, получается вследствие последовательного применения операторов растяжения пространства $R_{1/\alpha}(\xi_1)$ сначала к вектору $g_F(x_k)$, затем оператор растяжения $R_{1/\alpha}(\xi_2)$ действует на вектор $R_{1/\alpha}(\xi_1) \cdot g_F(x_k)$ и т.д. и, наконец, оператор растяжения пространства с коэффициентом растяжения α^{-1} и в направлении вектора ξ_k $\left(R_{1/\alpha}(\xi_k)\right)$ действует на вектор, который получился в результате последовательного действия операторов растяжения в направлении $\xi_1, \xi_2, \dots, \xi_k$.

Выражение для определения $\overline{g}_k$ получается из определения обобщенного градиента при растяжении пространства E_n путем последовательного применения операторов растяжения в направлении векторов $\xi_1, \xi_2, \dots, \xi_k$:

$$y_k = R_\alpha(\xi_k) R_\alpha(\xi_{k-1}) \dots R_\alpha(\xi_1) x_k$$

и из условия $\overline{g}_k = g_{\varphi_k}(y_k)$, где $\varphi_k(y_k) = F\left(A_k^{-1} y_k\right)$:

$$\varphi_k(y_k) = F\left(R_{1/\alpha}(\xi_1) \cdot R_{1/\alpha}(\xi_2) \dots R_{1/\alpha}(\xi_k) y_k\right).$$

4. $r_k = \overline{g}_k - \widetilde{g}_k$ – разность обобщенных градиентов.

5. $\beta_k = \dfrac{\|r_k\|}{\|\widetilde{g}_k\|}$.

6. Сравниваем β_k с заданной величиной q_1:

а) если $\beta_k \le q_1$, то $x_{k+1} = x_k - h_k R_{1/\alpha}(\xi_1) \cdot R_{1/\alpha}(\xi_2) \dots R_{1/\alpha}(\xi_k) \dfrac{\widetilde{g}_k}{\|\widetilde{g}_k\|}$, т. е. движение производится в том же направлении и с той же длиной шага, что и на k-ой итерации, после чего переходим к выполнению $(k+2)$-ой итерации;

б) если $\beta_k > q_1$, то переходим к выполнению следующего действия.

7. $\xi_{k+1} = \dfrac{r_k}{\|r_k\|}$, вектор ξ_{k+1} запоминается.

8. $\widetilde{g}_{k+1} = R_{1/\alpha}(\xi_{k+1}) \cdot R_{1/\alpha}(\xi_k) \dots R_{1/\alpha}(\xi_1) g_F(x_k)$. Учитывая выражение для получения обобщенного градиента $\overline{g}_k$, получаем окончательно

$$\widetilde{g}_{k+1} = R_{1/\alpha}(\xi_{k+1}) \overline{g}_k,$$

где $\widetilde{g}_{k+1}$ – обобщенный градиент функции $\varphi_{k+1}(y)$ в точке y_k, соответствующей точке x_k с учетом на $(k+1)$-ой итерации растяжения пространства с помощью оператора растяжения в направлении вектора ξ_{k+1} с коэффициентом растяжения α.

9. $h_{k+1} = h_k q_2$, где $0 < q_2 \leq 1$.

10. Определяем очередную точку x_{k+1} итерационной последовательности, которая получается в результате перемещения в растянутом пространстве в направлении $-\widetilde{g}_{k+1}$ из точки

$$y_k = R_\alpha(\xi_k) \cdot R_\alpha(\xi_{k-1}) \ldots R_\alpha(\xi_1)\, x_k$$

в точку

$$y_{k+1} = y_k - h_{k+1} \frac{\widetilde{g}_{k+1}}{\|\widetilde{g}_{k+1}\|}.$$

Как указывалось выше,

$$y_{k+1} = R_\alpha(\xi_{k+1}) \ldots R_\alpha(\xi_1)\, x_{k+1},$$

откуда

$$x_{k+1} = R_{1/\alpha}(\xi_1) \cdot R_{1/\alpha}(\xi_2) \ldots R_{1/\alpha}(\xi_k)\, y_{k+1}.$$

Из этих соотношений выводится выражение для получения точки x_{k+1}:

$$x_{k+1} = x_k - h_k R_{1/\alpha}(\xi_1) \cdot R_{1/\alpha}(\xi_2) \ldots R_{1/\alpha}(\xi_{k+1}) \frac{\widetilde{g}_{k+1}}{\|\widetilde{g}_{k+1}\|}.$$

Переходим к выполнению действий на $(k+2)$-ой итерации. Так как память для запоминания векторов $\xi_1, \xi_2, \ldots, \xi_k, \ldots$ ограничена, приходится описанный выше процесс обрывать после определенного числа шагов r $(r < n)$. После этого матрица B восстанавливается, значение x_i принимается за начальное значение для нового внутреннего цикла, при этом начальный шаг берется в 2–3 раза меньше, чем в начале предыдущего цикла. Процесс продолжается до тех пор, пока не будут достигнуты приемлемые значения минимизируемой функции или пока приращение функции, полученное в результате проведения очередного цикла вычислений, не станет достаточно малым. Таким образом, второй алгоритм является промежуточным между обычными алгоритмом обобщенного градиентного спуска и алгоритмом с растяжением пространства.

Этот алгоритм применялся для решения ряда задач, связанных с проблемой построения оптимальных эталонов. Постановка задачи, исходные данные, подпрограммы вычисления функций и градиентов были любезно предоставлены В. А. Ковалевским и сотрудниками руководимого им отдела ИК АН УССР для проведения численных экспериментов. Решалась сложная максиминная задача, число переменных – 1200, число функций, участвующих в операции взятия минимума, – 1536.

Эта задача для сравнения эффективности алгоритмов решалась также обычным методом обобщенного градиентного спуска без растяжения пространства с использованием различных методик регулировки длины шага:

$$1)\ h_{k+1} = h_k \cdot 0,95 \qquad \text{и} \qquad x_{k+1} = x_k + h_{k+1} \frac{g_F(x_k)}{\|g_F(x_k)\|};$$

$$2)\ h_{k+1} = h_k \cdot 0,95 \qquad \text{и} \qquad x_{k+1} = x_k + h_{k+1} \cdot g_F(x_k);$$

$$3)\ h_{k+1} = \frac{F(x^*) - F(x_k)}{\|g_F(x_k)\|}, \qquad x_{k+1} = x_k + h_{k+1} \frac{g_F(x_k)}{\|g_F(x_k)\|},$$

где $F(x^*)$ – предполагаемое значение в точке максимума. При этом наиболее хорошие результаты решения задачи были получены при первом способе регулировки шага. В табл. 2 приведены результаты счета, полученные при такой методике выбора шага.

Для решения задачи с помощью описанного здесь второго алгоритма, применение которого обусловлено значительной размерностью задачи ($n = 1200$), была составлена программа на языке ФОРТРАН. Решение проводилось на ЭВМ «БЭСМ-6», при этом полагалось $\alpha = 3$.

Анализ результатов решения задачи, полученных обоими методами и приведенных в табл. 2, показывает, что для достижения примерно одних и тех же значений оптимизируемой функции вблизи точки максимума при использовании второго алгоритма требуется значительно меньшее число итераций и меньшее количество машинного времени, чем при использовании метода обобщенного градиентного спуска без растяжения пространства. Так, например, приемлемое значение оптимизируемой функции, равное 3,88, было получено обобщенным градиентным методом без растяжения пространства на 71-ой итерации, при использовании же второй модификации алгоритма уже на 25-ой итерации было получено значение $F(x)$, равное 3,96.

Кроме того, применение второго алгоритма позволило получить значение максимизируемой функции (5,07 на 39 итерации), с хорошей точностью совпадающее с теоретическим максимумом, которую при использовании обобщенного градиентного спуска достичь не удалось, так как это требовало слишком большого числа итераций и продолжительного времени счета.

В заключение авторы выражают благодарность сотрудникам ИК АН УССР М. И. Шлезингеру и Л. С. Коровякиной за помощь в проведении численных экспериментов.

Таблица 2.

№ итерации	Значение оптимизируемой функции, полученное обобщенным градиентным методом с помощью первой регулировки шага	Значение оптимизируемой функции, полученное с помощью второго алгоритма
1	−30,92	−30,92
5	−7,009	−11,26
10	−4,548	−8,703
15	−0,3739	1,143
20	1,456	2,785
25	1,911	3,965
30	1,911	4,551
35	2,315	4,870
40	3,350	5,010
45	3,403	5,197
50	3,826	5,277
55	3,779	5,312
60	4,025	5,333
65	3,732	5,347
71	3,882	

Литература

1. Шор Н. З. Обобщенный градиентный спуск // Труды I Зимней школы по математическому программированию (г. Дрогобыч). – М., 1969. – Вып. 3. – С. 578–585.

2. Шор Н. З. О скорости сходимости метода обобщенного градиентного спуска с растяжением пространства // Кибернетика. – 1970. – № 2. – С. 80–85.

3. Шор Н. З., Журбенко Н. Г. Метод минимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов // Кибернетика. – 1971. – № 3. – С. 51–59.

4. Демьянов В. Ф. Минимизация выпуклой функции максимина // Журн. вычисл. математики и мат. физики. – 1971. – Т. 11. – № 2.

5. Грачев Н. И., Евтушенко Ю. Г. Численный метод решения минимаксных задач // Журн. вычисл. математики и мат. физики. – 1971. – Т. 11. – № 2.

6. Демьянов В. Ф. К решению некоторых минимаксных задач. I, II // Кибернетика. – 1966. – № 6. – 1967. – № 3.

7. Пропой А. И. К теории максиминных задач // Журн. вычисл. математики и мат. физики. – 1971. – Т. 11. – № 1.

8. Зойтендейк Г. Методы возможных направлений. – М.: Мир, 1963.

О классе почти-дифференцируемых функций и одном методе минимизации функций этого класса

Н. З. Шор

Кибернетика. – 1972. – № 4. – С. 65–70.

1. Введение

В большинстве работ, посвященных исследованию методов градиентного типа, при обосновании сходимости предполагается непрерывность градиента [2], [3] или выпуклость минимизируемой функции [4]–[7]. Но часто в прикладных задачах нельзя постулировать гладкость минимизируемой функции или ее выпуклость, Например, в практике планирования и проектирования элементы функции затрат обычно задаются в виде кусочно-гладких не обязательно выпуклых функций от параметра, характеризующего производительность (пропускную способность) того или иного устройства. Отсутствие свойств выпуклости и гладкости характерно для широкого круга минимаксных (одноэтапных и многоэтапных) задач.

Все это побудило нас рассмотреть класс функций, достаточно широкий, чтобы охватить кусочно-гладкие функции и функции, встречающиеся в минимаксных задачах, и в то же время настолько узкий, чтобы для этих функций можно было определить некоторый аналог градиента и иметь возможность использовать методы градиентного типа для нахождения локальных минимумов.

Этот класс функций мы назовем почти-дифференцируемыми (ПД-функциями). Постулируя возможность вычисления почти-градиента, рассмотрим затем метод минимизации, аналогичный методу обобщенного градиентного спуска с растяжением пространства [6], [7], который был разработан для минимизации выпуклых функций. Опишем возможности применения этого метода к решению систем нелинейных уравнений и задач нелинейного программирования.

2. Почти-дифференцируемые функции

Определение 1. Непрерывную функцию $f(x)$, заданную в открытой области S n-мерного евклидова пространства E_n, назовем почти-дифференцируемой, если: а) в любой ограниченной области она удовлетворяет условию Липшица; б) эта функция почти везде дифференцируема; в) градиент непрерывен на множестве $M \subseteq S$, где он определен.

Определение 2. Почти-градиентом (ПД-градиентом) функции $f(x)$ в точке x мы назовем вектор, являющийся предельной точкой некоторой последовательности градиентов $g(x^{(1)}), g(x^{(2)}), \ldots, g(x^{(k)}), \ldots$, где $\left\{x^{(k)}\right\}_{k=1}^{\infty}$ – последовательность точек, сходящаяся к x, и такая, что во всех точках этой последовательности функция $f(x)$ дифференцируема.

Теорема 1. *Множество $G(x)$ почти-градиентов в любой точке x почти-дифференцируемой функции является непустым, ограниченным и замкнутым.*

Доказательство. Из свойства а) следует ограниченность последовательности $\{g(x^{(k)})\}$, а отсюда – непустота множества $G(x)$ и его ограниченность. Замкнутость является следствием определения почти-градиентов, как предельных точек. В самом деле, пусть $\{\widehat{g}^{(k)}(x)\}_{k=1}^{\infty}$ – последовательность почти-градиентов, сходящаяся к g^*. Тогда для каждого k найдется $x^{(k)}$ такое, что $\|x - x^{(k)}\| \leq \dfrac{1}{k}$; $\|\widehat{g}^{(k)}(x) - g(x^{(k)})\| \leq \dfrac{1}{k}$, причем в точке $x^{(k)}$ $f(x)$ дифференцируема:

$$\begin{aligned}\|g^* - g(x^{(k)})\| &\leq \|g^* - \widehat{g}^{(k)}(x)\| + \|\widehat{g}^{(k)}(x) - g(x^{(k)})\| \leq \\ &\leq \|g^* - \widehat{g}^{(k)}(x)\| + \frac{1}{k}.\end{aligned}$$

$\lim \|g^* - g(x^{(k)})\| = 0$, откуда следует, что g^* является почти-градиентом $f(x)$ в точке x.

Теорема 2. *Произвольная выпуклая функция $f(x)$, определенная в E_n, является почти-дифференцируемой, а ее любой почти-градиент в точке x совпадает с некоторым обобщенным градиентом.*

Доказательство. В силу теоремы Радемейстера [1] произвольная выпуклая функция, определенная в E_n, почти везде непрерывно дифференцируема. Из общих свойств выпуклой функции известно, что ее производные по направлению в любой ограниченной области равномерно ограничены. Таким образом, $f(x)$ – почти-дифференцируемая функция.

Пусть почти-градиент $\widehat{g}_f(\overline{x}) = \lim\limits_{k\to\infty} g_f(x_k)$, где $\{g_f(x_k)\}$ – градиенты в точках x_k, $\lim\limits_{k\to\infty} x_k = \overline{x}$. Из свойства выпуклости $f(x)$ вытекает для всех $x \in E_n$:

$$f(x) - f(x_k) \geq (g_f(x_k),\, x - x_k).$$

Устремив k к ∞ получаем:

$$f(x) - f(\overline{x}) \geq (\widehat{g}_f(\overline{x}), x - \overline{x}),$$

т. е. $\widehat{g}_f(\overline{x})$ является обобщенным градиентом функции $f(x)$ в точке x. Теорема доказана.

Теорема 3. *Если две функции $f_1(x)$ и $f_2(x)$ почти-дифференцируемы, то $f_3(x) = f_1(x) + f_2(x)$; $f_4(x) = f_1(x) \cdot f_2(x)$ также почти-дифференцируемы.*

Доказательство. Рассмотрим N_3 – объединение множеств N_1 и N_2, в которых функции $f_1(x)$ и $f_2(x)$ не дифференцируемы. Мера N_3 равна 0. В остальных точках $f_3(x)$, $f_4(x)$ дифференцируемы и их градиенты подсчитываются по правилам дифференцирования суммы и произведения двух функций. Отсюда следует утверждение теоремы.

Теорема 4. *Если $f_1(x)$ и $f_2(x)$ – почти-дифференцируемы, то $f(x) = \max(f_1(x),\, f_2(x))$ – также почти-дифференцируема.*

Доказательство. Введем обозначение:

$$\varphi^+(x) = \begin{cases} 0, & \varphi(x) \leq 0, \\ \varphi(x), & \varphi(x) > 0. \end{cases}$$

Тогда

$$f(x) = f_1(x) + [f_2(x) - f_1(x)]^+ .$$

Из теоремы 2 следует, что $f_2(x) - f_1(x)$ почти-дифференцируема. Далее, в силу непрерывности $f_1(x)$ и $f_2(x)$ множество

$$M_{12} = \{x : f_1(x) > f_2(x)\}$$

является открытым, и поэтому функция $f(x)$ будет дифференцируема во всех точках этого множества, в которых дифференцируема функция $f_1(x)$. Аналогичное утверждение можно сделать для множества $M_{21} = \{x : f_1(x) < f_2(x)\}$. Что касается множества $M_0 = \{x : f_1(x) = f_2(x) = f(x)\}$, то мера границы этого множества равна 0, а для всех внутренних точек дифференцируемость вытекает из дифференцируемости в соответствующей точке $f_1(x)$. Отсюда вытекает утверждение теоремы.

3. Об одном методе нахождения локальных минимумов ПД-функций

Мы рассмотрим метод минимизации ПД-функций, соответствующий методу ОГСРП для выпуклых функций, рассмотренному в работах [6], [7], с тем изменением, что роль обобщенного градиента будет играть ПД-градиент.

Пусть $f(x)$ – ПД-функция, определенная в E_n. Мы будем предполагать, что имеется алгоритм, позволяющий с произвольной точностью вычислить некоторый ПД-градиент $\widehat{g}_f(x)$ функции $f(x)$ в произвольной точке $x \in E_n$, заданы алгоритмы вычисления двух последовательностей положительных чисел $\{h_k\}$ и $\{\beta_k\}$, $k = 1, 2, \ldots$, которые нам в дальнейшем понадобятся, а также начальное приближение x_0 и начальная матрица $B_0 = I$ $\;(I$ – единичная матрица) размерности $n \times n$.

При этих условиях определим бесконечно-шаговый процесс ПДРП, $(k+1)$-й шаг которого описывается следующим образом $(k = 0, 1, 2, \ldots)$: вычисляем

1) $\widehat{g}_f(x_k)$ (если $\widehat{g}_f(x_k) = 0$, процесс останавливается);

2) $\widetilde{g}_k = B_k^* \cdot \widehat{g}_f(x_k)$ (B_k^* – матрица, транспонированная B_k);

3) $\xi_{k+1} = \dfrac{\widetilde{g}_k}{\|\widetilde{g}_k\|}$;

4) $x_{k+1} = x_k - B_k \cdot h_{k+1} \cdot \xi_{k+1}$;

5) $B_{k+1} = B_k \cdot R_{\beta_{k+1}}(\xi_{k+1})$, где $R_\beta(\xi)$ – оператор растяжения пространства в направлении ξ с коэффициентом растяжения β. (Определение этого оператора, а также комментарии к аналогичной схеме имеются в [6]).

Докажем теорему, аналогичную теореме 2 из [7].

Теорема 5. *Пусть x^* – точка локального минимума функции $f(x)$ и в процессе реализации алгоритма ПДРП выполняются следующие условия:*

1) $\|A_k(x_k - x^*)\| \le r$, *где* $A_k = B_k^{-1}, \quad r > 0;$

2) $1 < \alpha_k \le \alpha^*; \quad \alpha_k = \dfrac{1}{\beta_k}; \quad k = 1, 2, \ldots$

Тогда существует такая подпоследовательность $x_{k_1}, x_{k_2}, \ldots, x_{k_p}, \ldots$, что

$$\|\widetilde{g}_{k_p}\| \le c\biggl(\prod_{j=1}^{k_p} \alpha_j\biggr)^{-1/n}, \quad p = 1, 2, \ldots$$

Доказательство. Повторяем рассуждения, которые мы проводили при доказательстве теоремы 2 [7] с заменой обобщенного градиента на ПД-градиент.

Теорема 6. *Если при предположениях теоремы 5 дополнительно потребовать, чтобы существовала положительная неубывающая функция $\delta(\varepsilon)$ для всех $0 < \varepsilon \le r$ такая, что*

$$\inf_{\{x:\varepsilon \le \|x-x^*\| \le r\}} (\widehat{g}_f(x), x - x^*) \ge \delta(\varepsilon) \cdot \|x - x^*\|\,, \tag{1}$$

L – константа Липшица $f(x)$ для $\{x : \|x - x^*\| \le r\}$, то существует подпоследовательность $\{x_{k_p}\}_{p=1}^{\infty}$, такая что

$$f(x_{k_p}) - f(x^*) \le L \cdot \varphi\Bigg(rc\Bigg(\prod_{j=1}^{k_p} \alpha_j\Bigg)^{-1/n}\Bigg),$$

где $\varphi(q)$ – корень уравнения $z\delta(z) = q$.

Замечание. Условие типа (1) предложил использовать Л. Г. Баженов [10].

Доказательство. Используя теорему 5, получаем:

$$\left(\widehat{g}_f(x_{k_p}), \frac{x_{k_p} - x^*}{\|A_{k_p}(x_{k_p} - x^*)\|}\right) = \left(B^*_{k_p}\widehat{g}_f(x_{k_p}), \frac{A_{k_p}(x_{k_p} - x^*)}{\|A_{k_p}(x_{k_p} - x^*)\|}\right) =$$

$$= \left(\widetilde{g}_{k_p}, \frac{A_{k_p}(x_{k_p} - x^*)}{\|A_{k_p}(x_{k_p} - x^*)\|}\right) \le \|\widetilde{g}_{k_p}\| \le c \cdot \Bigg(\prod_{j=1}^{k_p} \alpha_j\Bigg)^{-1/n}.$$

Но

$$\left(\widehat{g}_f(x_{k_p}), \frac{x_{k_p} - x^*}{\|A_{k_p}(x_{k_p} - x^*)\|}\right) \ge \frac{\delta\big(\|x_{k_p} - x^*\|\big) \cdot \|x_{k_p} - x^*\|}{\|A_{k_p}(x_{k_p} - x^*)\|} \ge$$

$$\ge \frac{\delta\big(\|x_{k_p} - x^*\|\big) \cdot \|x_{k_p} - x^*\|}{r}.$$

Таким образом,

$$\frac{\delta\big(\|x_{k_p} - x^*\|\big) \cdot \|x_{k_p} - x^*\|}{r} \le c \cdot \Bigg(\prod_{j=1}^{k_p} \alpha_j\Bigg)^{-1/n};$$

$$f(x_{k_p}) - f(x^*) \le L \left\| x_{k_p} - x^* \right\| \quad \le \quad L \cdot \varphi\Bigg(rc \Bigg(\prod_{j=1}^{k_p} \alpha_j \Bigg)^{-1/n} \Bigg),$$

где $\varphi(q)$ – корень уравнения $z \cdot \delta(z) = q$.

Теорема 7. *Пусть $f(x)$ – почти-дифференцируемая функция, определенная в некоторой сферической окрестности S_r точки x^* локального минимума ($S_r = \{x : \|x - x^*\| \le r\}$), и в тех точках x, где функция дифференцируема, ее производные по направлению $\mu(x) = \dfrac{x - x^*}{\|x - x^*\|}$ удовлетворяют следующему неравенству:*

$$N(f(x) - f(x^*)) \le f'_{\mu(x)}(x) \cdot \|x - x^*\| \le M\left(f(x) - f(x^*)\right), \tag{2}$$

где $0 < N < M$.

Тогда, если мы в ПДРП-алгоритме примем

(1) $x_0 \in S_r$;

(2) $h_{k+1} = \dfrac{2MN}{M+N} \cdot \dfrac{f(x_k) - f(x^*)}{\|\widetilde{g}_k\|}$;

(3) $\alpha_{k+1} = \alpha = \dfrac{M+N}{M-N}$,

то найдутся константа $\overline{c}$ и подпоследовательность индексов $\{k_p\}_{p=1}^{\infty}$ такая, что

$$f(x_{k_p}) - f(x^*) \le \overline{c} \cdot \alpha^{-k_{\overline{p}}/n}.$$

Доказательство. Почти повторяя доказательство теоремы 2 из [6], получаем, что для всех $k \quad \|A_k(x_k - x^*)\| \le r$. При этом неравенства (2) в тех точках, где функция недифференцируема, заменяются неравенствами:

$$N(f(x) - f(x^*)) \le (\widehat{g}_f(x), \mu(x))\,\|x - x^*\| \le M(f(x) - f(x^*)).$$

Замечание. В тексте [6] (стр. 10–11) имеются опечатки: при рассмотрении случаев I и II нужно M и N поменять местами.

После этого применяем теорему 5, учитывая, что $\alpha_k = \alpha$. Имеем:

$$N(f(x_{k_p}) - f(x^*)) \le (\widehat{g}_f(x_{k_p}), x_{k_p} - x^*) = (A^*_{k_p}\widetilde{g}_{k_p}, x_{k_p} - x^*) =$$

$$= (\widetilde{g}_{k_p}, A_{k_p}(x_{k_p} - x^*)) \le \left\|\widetilde{g}_{k_p}\right\| \cdot r \le rc \cdot \alpha^{-k_p/n},$$

$$f(x_{k_p}) - f(x^*) \le \frac{rc}{N}\alpha^{-k_p/n} = \overline{c} \cdot \alpha^{-k_p/n},$$

что и требовалось доказать.

Теорема 7 может быть уточнена.

Теорема 8. *Пусть соблюдаются условия теоремы 7. Тогда для произвольного δ, $0 < \delta < 1$, найдется c_δ такое, что для всех $k = 1, 2, \ldots$*

$$\min_{0 \le i \le k} [f(x_i) - m^*] \le c_\delta \cdot \alpha^{-k(1-\delta)/n}.$$

Доказательство. Представим $A_k = B_k^{-1}$ в виде произведения ортогонального и симметричного положительно определенного оператора:

$$A_k = O_k \cdot S_k,$$

и пусть $\lambda^{(k)}$ – максимальное собственное число оператора S_k. Предположив, что

$$\frac{\lambda^{(k+1)}}{\lambda^{(k)}} \ge 1 + \delta', \quad \delta' > 0, \tag{3}$$

мы так же, как при доказательстве теоремы 2 из [7], получим:

$$||\widetilde{g}_k|| \le \frac{\overline{c}}{\lambda^{(k)} \cdot c(\delta')}, \tag{4}$$

где $c(\delta') = \sqrt{2\delta'/(\alpha^2 - 1)}$. Пусть δ' достаточно мало, во всяком случае, меньше, чем $\sqrt[n]{\alpha} - 1$, и на $(k+1), \ldots, (k+p)$-й итерациях не выполняется (3). Получим тогда

$$\lambda^{(k+p)} \le \lambda^{(k)} \cdot (1 + \delta')^p.$$

Так как $\lambda^{(k+p)} \ge (\sqrt[n]{\alpha})^{k+p}$, то число p удовлетворяет неравенству:

$$p \ln(1 + \delta') + \ln \lambda^{(k)} - \frac{(k+p)\ln\alpha}{n} \ge 0,$$

$$p \le \frac{\ln \lambda^{(k)} - \dfrac{k \ln \alpha}{n}}{\dfrac{\ln \alpha}{n} - \ln(1 + \delta')}.$$

Пусть

$$\delta = \frac{\ln(1 + \delta') \cdot n}{\ln \alpha}.$$

Тогда

$$p \le \frac{n \ln \lambda^{(k)} - k \ln \alpha}{(1-\delta)\ln \alpha}. \tag{5}$$

Теперь рассмотрим ситуацию:

$$\frac{\lambda^{(k)}}{\lambda^{(k-1)}} \ge 1+\delta'; \quad \frac{\lambda^{(i+1)}}{\lambda^{(i)}} < 1+\delta' \quad \text{для } i = k, \ldots, k+m-1.$$

Из неравенства (5) имеем для $q \le m-1$:

$$\ln \lambda^{(k)} \ge \frac{q \ln \alpha(1-\delta) + k \ln \alpha}{n} \ge \frac{(q+k)\cdot \ln \alpha \cdot (1-\delta)}{n}. \tag{6}$$

Далее получаем, используя (4):

$$\min_{0\le i\le k+q} \|\widetilde{g}_i\| \le \|\widetilde{g}_{k-1}\| \le \frac{a_{\delta'}}{\lambda^{(k-1)}}, \quad \text{где} \quad a_{\delta'} = \overline{c}/c(\delta').$$

Но $\dfrac{a_{\delta'}}{\lambda^{(k-1)}} \le \dfrac{\alpha a_{\delta'}}{\lambda^{(k)}} \le \alpha a_{\delta'} \cdot \alpha^{-(q+k)(1-\delta)/n}$. Так как δ' однозначно определяется δ, то, обозначив $\alpha a_{\delta'} = \widetilde{c}_\delta$, получаем: для всех k, $k = 1, 2, \ldots$,

$$\min_{0\le i\le k} \|\widetilde{g}_i\| \le \widetilde{c}_\delta \cdot \alpha^{-k(1-\delta)/n}.$$

Далее, используя (2),

$$\min_{0\le i\le k}[f(x_i) - f(x^*)] \le \min_{0\le i\le k} \|\widetilde{g}_i\| \cdot \frac{r}{N} \le c_\delta \cdot \alpha^{-k(1-\delta)/n},$$

где $c_\delta = \widetilde{c}_\delta \cdot \dfrac{r}{N}$.

Следствие. Так как δ может быть сколь угодно малым, то из предыдущей теоремы получаем следующую асимптотическую оценку:

$$\overline{\lim_{k\to\infty}} \sqrt[k]{\min_{1\le i\le k}[f(x_i) - f(x^*)]} \le \frac{1}{\sqrt[n]{\alpha}},$$

т. е. метод сходится по функционалу асимптотически со скоростью геометрической прогрессии со знаменателем $\dfrac{1}{\sqrt[n]{\alpha}}$.

4. Применение к решению систем нелинейных уравнений и задач нелинейного программирования

Для применения теоремы 8 нужно знать при выборе $\{h_k\}$ и α значение $f(x^*)$ и уметь оценивать константы M и N. Наиболее просто это получается при решении «полных» систем нелинейных уравнений.

Пусть имеется система нелинейных уравнений:

$$\varphi_i(x) = 0; \quad x \in E_n; \quad i = 1, \ldots, n. \tag{7}$$

где $\varphi_i(x)$ – почти-дифференцируемые функции. Решение этой системы можно разными способами свести к решению задачи безусловной минимизации.

В частности, получаются следующие задачи:

минимизировать

$$\psi_1(x) = \max_{1 \le i \le n} |\varphi_i(x)|\,;$$

минимизировать

$$\psi_2(x) = \sum_{i=1}^{n} |\varphi_i(x)|\,;$$

минимизировать

$$\psi_3(x) = \sum_{i=1}^{n} \varphi_i^2(x).$$

Легко, увидеть (см. теоремы 3, 4), что функции $\psi_1(x)$, $\psi_2(x)$, $\psi_3(x)$ также будут почти-дифференцируемыми. Локальный минимум этих функций x^*, соответствующий решению системы (7), дает $\psi_j(x^*) = 0$; $j = 1, 2, 3$.

Нас будет интересовать поведение функций $\psi_j(x^*)$ в окрестности таких локальных минимумов. Так как $\varphi_i(x)$ – ПД-функции, то типичной следует считать ситуацию, когда в точке x^* эти функции дифференцируемы и значение якобиана

$$\det I(x^*) = \det \left\{ \frac{\partial \varphi_i}{\partial x_j}(x^*) \right\}_{i,j=1}^{n} \neq 0.$$

Назовем такой случай *регулярным.* Докажем теорему.

Теорема 9. *В регулярном случае найдется для любого $\delta > 0$ такая окрестность S_δ точки минимума x^*, что при $x \in S_\delta$*

$$(1-\delta)\psi_j(x) \le (\widehat{g}_{\psi_j}(x),\, x - x^*) \le (1+\delta)\psi_j(x)$$

для $j = 1, 2, 3$.

Доказательство. Доказательство проведем для случая $j = 1$, так как в случае $j = 2, 3$ оно тривиально. Так как в точке x^* $\varphi_i(x)$, $i = 1, \ldots, n$, дифференцируемы, то производные по любому направлению η, $\|\eta\| = 1$, удовлетворяют соотношениям:

$$\begin{aligned} &\varphi_{i\eta}(x^*) = (g_{\varphi_i}(x^*), \eta); \\ &\psi'_{1\eta}(x^*) = \max\left|(g_{\varphi_i}(x^*), \eta)\right|. \end{aligned}$$

Для любого η, $\|\eta\| = 1$, $\psi'_{1\eta}(x^*) > 0$. В самом деле, если для некоторого η $\psi'_{1\eta}(x^*) = 0$, то $(g_{\varphi_i}(x^*), \eta) = 0$, $i = 1, \ldots, n$, но тогда $\det I(x^*) = 0$, что противоречит регулярности решения x^*.

Так как производные по направлению в точке x^* представляют значения непрерывной функции, определенной на компактном множестве – единичной сфере, то множество значений ограничено и замкнуто. Следовательно, найдется такое $a > 0$, что $\psi'_{1\eta}(x^*) \ge a$ для всех η, $\|\eta\| = 1$. Для любого $\varepsilon > 0$ выберем такую окрестность O_ε точки x^*, что для всех $x \notin O_\varepsilon$, в которых $\varphi_i(x)$ дифференцируемы, выполняются следующие неравенства:

$$\begin{aligned} &|\varphi_i(\overline{x}) - (g_{\varphi_i}(x^*), \overline{x} - x^*)| \le \varepsilon \|\overline{x} - x^*\|; \\ &\|g_{\varphi_i}(\overline{x}) - g_{\varphi_i}(x^*)\| \le \varepsilon, \quad i = 1, \ldots, n. \end{aligned} \tag{8}$$

Выделим индексы $\overline{i}$, i^* такие, что:

$$\psi_1(\overline{x}) = |\varphi_{\overline{i}}(\overline{x})|; \qquad \psi'_{1\mu(x)}(\overline{x}) = |\varphi'_{\overline{i}\mu(\overline{x})}(\overline{x})|;$$

$$\psi'_{1\mu(\overline{x})}(\overline{x}^*) = |\varphi'_{i^*\mu(\overline{x})}(x^*)|; \qquad \mu(\overline{x}) = \frac{x - \overline{x}^*}{\|\overline{x} - x^*\|}.$$

Принимая во внимание (8), получаем:

$$\begin{aligned} \psi_1(\overline{x}) = |\varphi_{\overline{i}}(\overline{x})| \quad &\le \quad \left(|\varphi'_{\overline{i}\mu(\overline{x})}(x^*)| + \varepsilon\right) \|\overline{x} - x^*\| \le \\ &\le \quad \left(|\varphi'_{\overline{i}\mu(\overline{x})}(\overline{x})| + 2\varepsilon\right) \|\overline{x} - x^*\|; \end{aligned}$$

$$\psi_1(\overline{x}) \ge \Big(|\varphi'_{\overline{i}\mu(\overline{x})}(x^*)| - \varepsilon\Big) \|\overline{x} - x^*\| \ge \Big(|\varphi'_{\overline{i}\mu(\overline{x})}(\overline{x})| - 2\varepsilon\Big) \|\overline{x} - x^*\|;$$

$$\frac{\psi_1(\overline{x})}{1+\dfrac{2\varepsilon}{\psi'_{1\mu(\overline{x})}(\overline{x})}} \le \psi'_{1\mu(\overline{x})}(\overline{x}) \cdot \|\overline{x}-x^*\| \le \frac{\psi_1(\overline{x})}{1-\dfrac{2\varepsilon}{\psi'_{1\mu(\overline{x})}(\overline{x})}};$$

$$\psi'_{1\mu(\overline{x})}(\overline{x}) \ge \psi'_{1\mu(\overline{x})}(x^*) - \varepsilon \ge a - \varepsilon.$$

Так как ε может быть сколь угодно малым, то для любого δ можно найти такое ε, что будут выполняться условия:

$$(1-\delta')\psi_1(\overline{x}) \le (\widehat{g}_{\psi_1}(\overline{x}), \overline{x}-x^*) \le (1+\delta'')\psi_1(\overline{x}); \quad \delta', \delta'' < \delta,$$

что и требовалось доказать.

Из доказанной теоремы следует, что в регулярном случае, находясь в достаточно малой окрестности точки x^* и применяя ПДРП–алгоритм, мы можем работать в соответствии с теоремой 7 с большим α и $\gamma = \dfrac{2MN}{M+N}$, близкими к 1. В самом деле, если $N \ge 1-\delta$; $M \le 1+\delta$, то, приняв $\alpha = \dfrac{M+N}{M-N} \ge \dfrac{1}{\delta}$, $\gamma = \dfrac{2MN}{M+N} \approx 1$, мы получим при $\delta \to 0$ $\alpha \to \infty$. Если мы имеем задачу нелинейного программирования:

$$\text{найти } \min f(x)$$

при ограничениях

$$\varphi_i(x) \le 0; \quad i=1,\dots,k; \quad \varphi_j(x) = 0; \quad j = k+1,\dots,k+m,$$

то в условиях, когда справедливо обобщенное правило множителей Лагранжа [11], мы можем поиск локального минимума этой задачи свести, используя метод штрафных функции [8], к задаче безусловной оптимизации функции

$$L(x) = f(x) + \sum_{i=1}^{k} S_i(x)\varphi_i(x) + \sum_{j=k+1}^{k+m} S \cdot |\varphi_j(x)|,$$

где $S_i(x) = \begin{cases} 0, & \varphi_i(x) \le 0, \\ S, & \varphi_i(x) > 0, \end{cases}$ S – достаточно большое число [8].

Дополнительная трудность будет состоять в том, что $L(x^*)$ неизвестно. Однако мы можем построить алгоритм с последовательным подбором $L(x^*)$ подобно алгоритму, изложенному в [7] (стр. 83).

В данной статье мы показали возможность применения методов градиентного типа с растяжением пространства к задачам минимизации почти-дифференцируемых функций. Применительно к другому классу функций (так называемых π-дифференцируемых функций) близкие результаты имеются в [9]. Другой метод минимизации почти-дифференцируемых функций описан в [10].

Литература

1. Буземан Г. Выпуклые поверхности. – М.: Наука, 1969.
2. Любич Ю. И., Майстровский Г. Д. Общая теория релаксационных процессов для выпуклых функционалов // Успехи матем. наук. – 1970. – № 1.
3. Поляк Б. Т. Градиентные методы минимизации функционалов // Журн. вычисл. математики и мат. физики. – 1963. – № 4.
4. Шор Н. З. Обобщенный градиентный спуск // Труды I Зимней школы по математическому программированию. (Дрогобыч). – М.: ЦЭМИ АН СССР, 1969. – Вып. 3.
5. Поляк Б. Т. Один общий метод решения экстремальных задач // Докл. АН СССР. – 1967. – **174**. – № 1.
6. Шор Н. З. Использование операции растяжения пространства в задачах минимизации выпуклых функций // Кибернетика. – 1970. – № 1. – С. 6–12.
7. Шор Н. З. О скорости сходимости метода обобщенного градиентного спуска с растяжением пространства // Кибернетика. – 1970. – № 2. – С. 80–85.
8. Еремин И. И. О методе «штрафов» в выпуклом программировании // Кибернетика. – 1967. – № 4.
9. Шор Н. З. Методы минимизации недифференциальных функций и их приложения. Автореферат диссертации на соискание степени доктора физ.-мат наук. – АН УССР. – К., 1970. – 43 с.
10. Баженов Л. Г. Об условиях сходимости метода минимизации почти-дифференцируемых функций // Кибернетика. – 1972. – № 4.
11. Пшеничный Б. Н. Необходимые условия экстремума. – М.: Наука, 1969.

Об одной модификации алгоритмов минимизации градиентного типа с растяжением пространства для решения задач большой размерности

Н. З. Шор, В. И. Гершович
Кибернетика. – 1981. – № 5. – С. 67–70.

1. Для решения задачи безусловной минимизации функции $f(x)$, определенной в n-мерном евклидовом пространстве E^n и обладающей почти-градиентом [1] $g(x)$ (если $f(x)$ дифференцируема, $g(x)$ совпадает с обычным градиентом), рассмотрим следующую общую схему итеративного алгоритма.

Выбираем начальную точку $x_0 \in E^n$, вектор g_0 и симметрическую матрицу H_0 порядка $n \times n$. Полагаем $x_1 = x_0$.

На k-й итерации ($k = 1, 2, \ldots$), имея $x_k \in E^n$, вектор g_{k-1} и матрицу H_{k-1} порядка $n \times n$, поэтапно вычисляем:

1) почти-градиент $g_k = g(x_k)$;

2) векторы $d_k^t = D^t(x_0, x_1, \ldots, x_k,\ g_0, g_1, \ldots, g_k,\ H_0, H_1, \ldots, H_{k-1})$, $t = \overline{1, T}$;

3) матрицу $H_k = H_{k-1} + \sum\limits_{t=1}^{T} \xi_k^t d_k^t (d_k^t)^*$, где ξ_k^t – скаляры;

4) направление $p_k = -H_k g_k$;

5) новую точку $x_{k+1} = x_k + s_k p_k$, где s_k – скалярный шаговый множитель.

Многие современные алгоритмы минимизации могут быть получены исходя из описанной схемы при надлежащем выборе вектор-функций D^t и процедур, порождающих числовые последовательности $\{\xi_k^t\}$ и $\{s_k\}$. К таким алгоритмам относятся проективный вариант метода сопряженных градиентов, некоторые модификации квазиньютоновских методов и методов с переменной метрикой. В основе перечисленных методов, по существу, лежит идея квадратичной аппроксимации минимизируемой функции, поэтому при их применении обычно предполагаются определенные свойства гладкости. Примечательно, что исходя из этой схемы могут быть получены и методы обобщенного градиентного спуска с растяжением пространства [1], основанные на идеях, отличных от аппроксимации гессиана. Эти методы применимы и в случае минимизации негладких функций. В наиболее часто используемой их модификации – r-алгоритме [1] – на этапах 2, 3 общей схемы вычисления проводятся соответственно по формулам:

$$d_k = \gamma_k H_{k-1} e_{k-1},$$

где $\gamma_k^2 = \left(1 - \alpha_k^{-2}\right) \left(e_{k-1}, H_{k-1} e_{k-1}\right)^{-1}$, $e_{k-1} = g_k - g_{k-1}$, $\alpha_k > 1$ – коэффициент растяжения пространства;

$$H_k = H_{k-1} - d_k d_k^*.$$

2. Направление движения, вычисляемое по формуле $p_k = -H_k g_k$, можно рассматривать как результат применения к антиградиенту линейного преобразования, задаваемого матрицей H_k. Естественно поставить задачу о нахождении структуры (совокупности собственных значений и собственных векторов) преобразования H_k. В случае произвольной матрицы H_k эта задача достаточно сложна, но она существенно упрощается, если известна структура матрицы H_{k-1} [2]. Ее решение основывается на следующем результате.

Лемма. *Пусть $a^1, a^2, \ldots, a^n$ – ортонормированная система из собственных векторов симметрической матрицы A, соответствующих собственным значениям $\mu^1, \mu^2, \ldots, \mu^n$. Пусть $d = \sum\limits_{j=1}^{n} d^j a^j$.*

Положим $B = A - dd^*$. Тогда:

а) если для некоторого j $d^j = 0$, то $\lambda = \mu^j$ – собственное значение для B и соответствующий собственный вектор $b = a^j$;

б) $D(\lambda) = \det(B - \lambda I) = \left[1 - \sum\limits_{j=1}^{n} \dfrac{(d^j)^2}{\mu^j - \lambda}\right] \times \prod\limits_{j=1}^{n} (\mu^j - \lambda)$ (заметим, что $D(\lambda)$ определен при $\lambda = \mu^j$);

в) если $D(\lambda) = 0$, т. е. λ – собственное значение матрицы B, и $\lambda \neq \mu^j$ $(j = \overline{1,n})$, то соответствующий собственный вектор $b = \sum\limits_{j=1}^{n} \dfrac{d^j}{\mu^j - \lambda} a^j$.

Доказательство. Матрица преобразования $(B - \lambda I)$ в базисе $a^1, a^2, \ldots, a^n$ имеет вид

$$\begin{pmatrix} (\mu^1 - d^1 d^1 - \lambda) & -d^1 d^2 & \cdots & -d^1 d^n \\ -d^2 d^1 & (\mu^2 - d^2 d^2 - \lambda) & & -d^2 d^n \\ \vdots & \vdots & \ddots & \vdots \\ -d^n d^1 & -d^n d^2 & \cdots & (\mu^n - d^n d^n - \lambda) \end{pmatrix}.$$

Подсчет ее определителя дает утверждение п. б) леммы. Справедливость пп. а) и в) проверяется непосредственно.

Применим этот результат для изучения структуры последовательности симметрических матриц, порождаемых r-алгоритмом по формулам:

$$H_0 = I, \quad H_k = H_{k-1} - d_k d_k^*, \quad k = 1, 2, \ldots$$

При всех k H_k имеет n вещественных собственных значений и полную ортонормированную систему из собственных векторов. Из геометрического смысла операторов H_k (каждый из них есть композиция операторов сжатия [1]) ясно, что все их собственные значения локализованы в интервале (0, 1]. Пусть для H_{k-1} известны собственные значения $0 < \lambda_{k-1}^1 \leq \lambda_{k-1}^2 \leq \ldots \leq \lambda_{k-1}^n \leq 1$ и соответствующая ортонормированная система из собственных векторов $b_{k-1}^1, b_{k-1}^2, \ldots, b_{k-1}^n$. Пусть $d_k = \sum\limits_{j=1}^{n} d_k^j b_{k-1}^j$. Положим $\lambda_k^0 = 0$. Непосредственно из леммы следует теорема.

Теорема. *Верны следующие утверждения:*

а) *если для некоторого j $d_k^j = 0$, то $\lambda = \lambda_{k-1}^j$ – собственное значение для H_k и соответствующий собственный вектор $b = b_{k-1}^j$;*

б) *если $\lambda_{k-1}^j = \lambda_{k-1}^{j+1} = \ldots = \lambda_{k-1}^{j+r-1}$, т. е. собственному значению λ_{k-1}^j соответствует инвариантное подпространство M размерности $r > 1$, и $(d_k^j)^2 + (d_k^{j+1})^2 + \ldots + (d_k^{j+r-1})^2 > 0$, то $\lambda = \lambda_{k-1}^j$ – собственное значение для H_k и ему соответствует инвариантное подпространство L размерности $(r-1)$; при этом $L \subset M$ и $L \perp d_k$;*

в) *если для некоторого $j < n$ $\lambda_{k-1}^j < \lambda_{k-1}^{j+1}$, то H_k имеет собственное значение λ, совпадающее с единственным корнем уравнения*

$1-\sum\limits_{j=1}^{n} \dfrac{(d_k^j)^2}{\lambda_{k-1}^j-\lambda}=0$ в интервале $(\lambda_{k-1}^j, \lambda_{k-1}^{j+1})$; при этом соответствующий собственный вектор $b=\beta\sum\limits_{j=1}^{n} \dfrac{d_k^j}{\lambda_{k-1}^j-\lambda} b_{k-1}^j$ (здесь β – некоторая нормировочная постоянная).

Отметим, что теорема полностью решает вопрос о структуре преобразования H_k при заданных структуре преобразования H_{k-1} и векторе d_k. Это позволяет реализовать r-алгоритм, работая с преобразованием H_k $(k=0,1,\ldots)$ не в обычной матричной форме, а задавая его совокупностью собственных значений и векторов.

3. Знание структуры матриц H_k $(k=0,1,\ldots)$ полезно с нескольких точек зрения. Так, в методах гладкой оптимизации квазиньютоновского типа, использующих квадратичную аппроксимацию, появляется возможность работать одновременно с приближениями к гессиану и его обратному, получать направления «отрицательной кривизны» и т. п. Здесь остановимся на возможностях, появляющихся при решении с помощью r-алгоритма задач большой размерности.

Пусть среди собственных значений матрицы H_k лишь m первых отличны от единицы:

$$0<\lambda_k^1\le\lambda_k^2\le\ldots\le\lambda_k^m<\lambda_k^{m+1}=\lambda_k^{m+2}=\ldots=\lambda_k^n=1.$$

Тогда для произвольного вектора v справедливо:

$$H_k v=\sum_{j=1}^{n}\lambda_k^j(b_k^j,v)b_k^j=v-\sum_{j=1}^{m}(1-\lambda_k^j)(v,b_k^j)b_k^j.$$

Таким образом, если преобразование H_k «действует» (т. е. отлично от тождественного) в подпространстве размерности m, для его задания вместо матрицы порядка $n\times n$ можно запоминать лишь m n-мерных векторов. Это существенно, например, в ситуации, когда значительная сложность получения информации о минимизируемой функции позволяет рассчитывать только на небольшое (по сравнению с n) количество шагов итеративного процесса.

Пусть H_k действует в подпространстве размерности m и задается m векторами. Тогда размерность пространства, в котором действует H_{k+1} не превосходит $(m+1)$ (непосредственное следствие теоремы) и для задания H_{k+1} может понадобиться уже $(m+1)$ вектор. Пусть по каким-либо причинам на всех последующих шагах нежелательно дальнейшее увеличение количества запоминаемой информации. Тогда, чтобы сделать возможным переход к H_{k+1}, действующему в m-мерном подпространстве, необходимо «сжать» информацию об H_k на один вектор, сохранив наиболее «существенную» ее часть. При задании оператора H_k совокупностью его собственных значений и векторов (в отличие от других [1], [3] способов задания) такое «сжатие» проводится просто и естественно: отбрасывается информация о собственном векторе, соответствующем собственному значению λ_k^m (наиболее близкому к единице). Таким образом, H_k заменяется своим проектором на подпространство, определяемое векторами $b_k^1, b_k^2, \ldots, b_k^{m-1}$.

Перейдем к описанию алгоритма минимизации. В его формулировке структуры линейных преобразований задаются записями вида

$$H = \begin{Bmatrix} \lambda^1 & \lambda^2 & \ldots & \lambda^m \\ b^1 & b^2 & \ldots & b^m \end{Bmatrix},$$

означающими, что $b^1, b^2, \ldots, b^m$ – ортонормированная система из m собственных векторов преобразования H, соответствующих собственным числам $\lambda^1, \lambda^2, \ldots, \lambda^m$ таким, что $0 < \lambda^1 \leq \lambda^2 \leq \ldots \leq \lambda^m \leq 1$ $(m \leq n)$; для произвольного вектора v $Hv = v - \sum\limits_{j=1}^{n} (1 - \lambda^j)(v, b^j) b^j$.

Зафиксируем натуральное m, вещественные $R \in (0, 1]$ и $Q > 1$ такие, что $m \leq n$ и $RQ \leq 1$.

Выбираем начальную точку $x_0 \in E^n$, вектор $g_0 = 0$ и преобразование

$$H_0 = \begin{Bmatrix} 1 & 1 & \ldots & 1 \\ b_0^1 & b_0^2 & \ldots & b_0^m \end{Bmatrix},$$

где b_0^j – j-й столбец единичной матрицы. Полагаем $x_1 = x_0$.

Перед k-й итерацией имеем $x_k \in E^n$, вектор g_{k-1} и преобразование

$$H_{k-1} = \begin{Bmatrix} \lambda_{k-1}^1 & \lambda_{k-1}^2 & \ldots & \lambda_{k-1}^m \\ b_{k-1}^1 & b_{k-1}^2 & \ldots & b_{k-1}^m \end{Bmatrix},$$

k-я итерация $(k = 1, 2, \ldots)$ состоит из нескольких этапов:

1) вычисляем почти-градиент $g_k = g(x_k)$;

2) формируем информацию, необходимую для очередного пересчета структуры преобразования; если $m = n$, переходим к п. 2 в;

а) заменяем H_{k-1} его проектором, полагая

$$H_{k-1} = \begin{Bmatrix} \lambda_{k-1}^1 & \lambda_{k-1}^2 & \dots & \lambda_{k-1}^{m-1} \\ b_{k-1}^1 & b_{k-1}^2 & \dots & b_{k-1}^{m-1} \end{Bmatrix};$$

б) если $\lambda_{k-1}^{m-1} < R$, полагаем $\lambda_{k-1}^j = Q\lambda_{k-1}^j, \quad j = \overline{1,(m-1)}$;

в) вычисляем вектор $d_k = \gamma_k H_{k-1} e_{k-1} = \gamma_k H_{k-1}(g_k - g_{k-1})$, где $\gamma_k^2 = (1 - \alpha_k^{-2})(e_{k-1}, H_{k-1}e_{k-1})^{-1}$; $\alpha_k > 1$ – коэффициент растяжения пространства;

г) вычисляем $d_k^j = (d_k, b_{k-1}^j)$, $j = \overline{1,(m-1)}$; если $m = n$, то вычисляем $d_k^m = (d_k, b_{k-1}^m)$ и переходим к этапу 3; вычисляем $d_k^m = \|d_k - \sum\limits_{j=1}^{m-1} d_k^j b_{k-1}^j\|$ и вектор $b_{k-1}^m = (d_k - \sum\limits_{j=1}^{m-1} d_k^j b_{k-1}^j)/d_k^m$;

д) полагаем

$$H_{k-1} = \begin{Bmatrix} \lambda_{k-1}^1 & \lambda_{k-1}^2 & \dots & \lambda_{k-1}^{m-1} & 1 \\ b_{k-1}^1 & b_{k-1}^2 & \dots & b_{k-1}^{m-1} & b_{k-1}^m \end{Bmatrix};$$

3) находим структуру преобразования $H_k = H_{k-1} - d_k d_k^*$, т. е. получаем:

$$H_k = \begin{Bmatrix} \lambda_k^1 & \lambda_k^2 & \dots & \lambda_k^m \\ b_k^1 & b_k^2 & \dots & b_k^m \end{Bmatrix};$$

4) вычисляем направление $p_k = -H_k g_k$;

5) вычисляем новую точку $x_{k+1} = x_k + s_k p_k$, где s_k – скалярный шаговый множитель.

Заметим, что в случае $m = n$ реализуется схема, эквивалентная схеме r-алгоритма. При $1 < m < n$ получаем алгоритм (назовем его r_m-алгоритмом), занимающий промежуточное положение между ОГС [1] и r-алгоритмом и в некотором смысле аппроксимирующий последний. В отличие от r-алгоритма, работающего, вообще говоря, с линейным преобразованием всего E^n, r_m-алгоритм работает с линейным преобразованием, которое действует в изменяющемся в процессе счета подпространстве размерности $m < n$. При этом (если не учитывать подсчет градиента) основные вычислительные затраты на одном шаге в

r-алгоритме связаны с умножением матрицы на вектор и ее пересчетом и имеют порядок $O\ (n^2)$ операций, а в r_m-алгоритме – с вычислением с ограниченной точностью корней характеристического уравнения $((O(mn))$ и пересчетом системы собственных векторов $(O(m^2n))$.

Для решения задач большой размерности описанный подход представляется перспективным по следующим соображениям. Высокая скорость сходимости, которую практически обеспечивает r-алгоритм, обусловлена его способностью ослаблять овражность, «растягивая» картину поверхностей уровня минимизируемой функции. В прикладных задачах большой размерности (имеющих, как правило, овражный характер) размерность оврага обычно невелика по сравнению с n, поэтому ослабить овражность может и алгоритм, который работает с линейным преобразованием, действующим в подпространстве размерности меньше n. Ослабление овражности связано с уменьшением углов между антиградиентами и направлениями на точки минимума. Следовательно, применение описанной методики в алгоритмах градиентного типа, уменьшающих шаг со скоростью геометрической прогрессии [1], должно привести к значительному улучшению точности решения при заданном количестве шагов итеративного процесса. Для экспериментальной проверки предложенного алгоритма минимизировалась плохо обусловленная функция $f(x) = \sum\limits_{i=1}^{20}(x^{(i)} - 1)^2/2^i$. Использовались: ОГС (шаг уменьшался геометрически со знаменателем 0,99); r_m-алгоритм (с $m = 5,\ 10,\ 15,\ 20$; $Q = 1{,}1$; $R = 1/Q$); r-алгоритм. В r_m- и r-алгоритмах полагали коэффициент растяжения $\alpha_k \equiv 3$. Во всех случаях в качестве начальной выбиралась точка $x_0 = \{x_0^{(i)} = 0, \quad i = \overline{1,20}\}$. Полученные результаты приведены в табл. 1. Численные эксперименты показывают, что при увеличении m (величины используемой памяти) в r_m-алгоритме точность решения улучшается (при заданном количестве итераций). При этом не увеличивается число обращений к информации о минимизируемой функции, что указывает на преимущества r_m-алгоритма по сравнению с ОГС, которые будут особенно существенными при решении задач большой размерности.

В настоящее время исследуются процедуры регулировки шаговых множителей в r_m-алгоритме в целях получения практически наиболее эффективного алгоритма. Результаты будут опубликованы позднее.

Таблица 1:

Алгоритм минимизации	Число итераций k	Число вычислений градиента	Достигнутое значение функции	$\|x_k - x^*\|$
ОГС	100	100	$0,5 \times 10^{-2}$	3,1
ОГС	150	150	$0,1 \times 10^{-2}$	3,0
r_5	100	147	$0,5 \times 10^{-4}$	2,2
r_{10}	100	149	$0,1 \times 10^{-5}$	0,8
r_{15}	100	136	$0,1 \times 10^{-10}$	$0,4 \times 10^{-3}$
r_{20}	100	129	$0,5 \times 10^{-14}$	$0,3 \times 10^{-4}$
r	100	135	$0,2 \times 10^{-13}$	$0,4 \times 10^{-4}$

Литература

1. Шор Н. З. Методы минимизации недифференцируемых функций и их приложения. – К.: Наук. думка. – 1979. – 200 с.

2. Гершович В. И. Об одном способе представления операторов преобразования пространства в ускоренных вариантах обобщенных градиентных методов // Тез. докл. III Всесоюз. семинара «Численные методы нелинейного программирования»: – Харьков. – 1979. – С. 64–66.

3. Шор Н. З., Шабашова Л П. О решении минимаксных задач методом обобщенного градиентного спуска с растяжением пространства // Кибернетика. – 1972. – № 1. – с. 82–88.

Ч А С Т Ь I I

МАТРИЧНЫЕ ЗАДАЧИ ОПТИМИЗАЦИИ

О методе получения оценок в квадратичных экстремальных задачах с булевыми переменными

Н. З. Шор, А. С. Давыдов
Кибернетика. – 1985. – № 2. – С. 48–54.

Рассматривается класс экстремальных задач следующего вида:

$$\min f_0(x), \quad x = \{x_1, \dots, x_n\} \tag{1}$$

при ограничениях

$$\begin{gathered} f_i(x) \le 0; \quad i \in I^+, \quad f_i(x) = 0; \quad i \in I, \\ I^+ \cup I = \{1, \dots, m\}, \end{gathered} \tag{2}$$

$$x_j \in \{0, 1\}, \quad j = 1, \dots, n, \tag{3}$$

где $f_\nu(x)$ – квадратичные функции от переменных $x_1, \dots, x_n$, $\nu = 0, 1, \dots, m$.

Условие (3) можно заменить квадратичными равенствами:

$$x_j^2 - x_j = 0, \quad j = 1, \dots, n. \tag{3'}$$

Заметим, что благодаря (3′) без ограничения общности можно считать, что $f_\nu(x)$, $\nu = 0, \dots, m$, не содержат линейных членов.

Пусть f^* – оптимальное значение целевой функции в задаче (1)–(3). Условимся, что $f^* = +\infty$, если задача (1)–(3) не имеет допустимых решений.

Используем стандартный метод множителей Лагранжа [3] для получения оценок снизу для f^*. Получаем следующую функцию Лагранжа

$$L(x, u) = f_0(x) + \sum_{i=1}^{m} u_i f_i(x) + \sum_{j=1}^{n} \overline{u}_j (x_j^2 - x_j).$$

Здесь $u = \{u_1, \dots, u_m;\ \overline{u}_1 \dots, \overline{u}_n\}$ – $(m+n)$-мерный вектор множителей Лагранжа.

Пусть U – подмножество $(n+m)$-мерного линейного пространства множителей Лагранжа, вырезаемое ограничениями $u_i \geq 0$ для тех i, у которых в (2) стоит знак «$\leq$ » $(i \in I^+)$. Рассмотрим

$$\Psi(u) = \inf_x L(x, u). \tag{4}$$

При $u \in U$ для любого допустимого x $L(x,u) \leq f_0(x)$. Следовательно, $\Psi(u) \leq f^*$, $\forall\, u \in U$. Отсюда $\Psi^* = \sup\limits_{u \in U} \Psi(u) \leq f^*$.

При любом $u \in U$ $L(x,u)$ – квадратичная функция, которую можно записать в виде

$$L(x, u) = (K(u)x,\ x) + l(x, u),$$

где $K(u)$ – симметричная матрица $n \times n$, $l(x,u)$ – линейная по x функция. Заметим, что элементы матрицы $K(u) = \{k_{ij}(u)\}_{i,j=1}^n$ линейно зависят от u.

Пусть U^+ – подмножество U значений u, при которых $K(u)$ является положительно определенной матрицей, $\overline{U}^+ = \{u \,|\, u \in U, K(u)$ – неотрицательно–определенной матрицей$\}$. При $u \in U^+$ задача (4) имеет единственный минимум $x(u)$, который получается путем решения соответствующей системы линейных уравнений. При $u \notin \overline{U}^+$ матрица $K(u)$ имеет отрицательные собственные числа и $\Psi(u) = -\infty$. Таким образом, поиск $\sup_{u \in U} \Psi(u)$ следует ограничить подмножеством $\overline{U}^+$.

Лемма. *U^+ является не пустым выпуклым множеством.*

Доказательство. Зафиксируем произвольные допустимые значения $u_1, \ldots, u_m$. Легко видеть, что за счет выбора достаточно больших $\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_n$ диагональные элементы матрицы $K(u)$ можно сделать сколь угодно большими, оставляя без изменения остальные элементы. Это доказывает существование значений u, при которых матрица $K(u)$ является положительно определенной, т. е. $U^+ \neq \emptyset$.

Докажем выпуклость U^+. Пусть $u^{(1)}$, $u^{(2)} \in U^+$, α_1, $\alpha_2 \geq 0$; $\alpha_1 + \alpha_2 = 1$. Рассмотрим $\overline{u} = \alpha_1 u^{(1)} + \alpha_2 u^{(2)}$. Так как элементы матрицы $K(u)$ являются линейными функциями от u, то $K(\overline{u}) = \alpha_1 K\left(u^{(1)}\right) + \alpha_2 K(u^{(2)})$. Для любого $x \neq 0$ $\left(K(u^{(1)})x, x\right) > 0$, $\left(K\left(u^{(2)}\right)x, x\right) > 0$. Отсюда

$$(K(\overline{u})x, x) = \alpha_1 (K(u^{(1)})x, x) + \alpha_2 (K(u^{(2)})x, x) > 0,$$

что доказывает положительную определенность $K(\overline{u})$ и, следовательно, выпуклость множества U^+. Лемма доказана.

Функция $\Psi(u)$, как инфимум по x линейных по u функций, является вогнутой на U^+. Таким образом, задача определения

$$\Psi^* = \sup_{u\in\overline{U}^+} \Psi(u) = -\inf_{u\in\overline{U}^+}(-\Psi(u)) \tag{5}$$

сводится к задаче выпуклого программирования. Заметим, что множество U^+ является открытым, поэтому экстремум в (5), если он существует, может достигаться и на границе множества U^+ , т. е. во множестве $\overline{U}^+$.

Пусть $-\Psi(u) = \varphi(u)$, $\varphi(u)$ – выпуклая функция. Будем решать задачу определения $\varphi^* = \min\limits_{u\in\overline{U}^+} \varphi(u)$ в предположении, что минимум существует. Для решения этой задачи можно воспользоваться методами недифференцируемой оптимизации. Заметим, что условие $u \in \overline{U}^+$ содержит два типа ограничений:

1) $u_i \geq 0$ при $i \in I^+$, где I^+ – подмножество индексов ограничений (2), которые записаны в форме ($\leq$);

2) $K(u)$ – неотрицательно-определенная матрица, что эквивалентно бесконечной системе линейных по u неравенств

$$\sum_{i,j=1}^{n} y_i y_j k_{ij}(u) \geq 0, \quad \forall y \, \|y\| = 1, \quad y = \{y_1, \ldots, y_n\},$$

которую можно переписать в компактном виде:

$$\lambda(u) = \min_{\|y\|=1} \sum_{i,j=1}^{n} k_{ij}(u) y_i y_j \geq 0, \tag{6}$$

где $\lambda(u)$ – минимальное собственное число матрицы $K(u)$, вогнутая функция от u.

Таким образом, получаем задачу

$$\min \varphi(u) \mid u_i \geq 0, \quad i \in I^+; \quad \lambda(u) \geq 0. \tag{7}$$

Так как функция $\varphi(u)$ и ее субградиент не определены за пределами области U^+, нельзя воспользоваться методом негладких точных внешних штрафных функций. Приходится применять для получения приближенного решения барьерные штрафные функции [6], например, в форме

$$S_{\varepsilon,\delta}(u) = \varphi(u) + \sum_{i \in I^+} \frac{\varepsilon}{u_i} + \frac{\delta}{\lambda(u)},$$

где ε и δ – достаточно малые положительные числа. Для минимизации $S_{\varepsilon,\delta}(u)$ можно использовать один из обобщенных градиентных методов недифференцируемой оптимизации (методы минимизации гладких функций в данном случае не пригодны, так как $S_{\varepsilon,\delta}(u)$ – недифференцируемая функция). Остановимся на способе вычисления обобщенных градиентов $g_\varphi(u)$, $g_\lambda(u)$ функций $\varphi(u)$ и $\lambda(u)$ при $u \in U^+$. Так как $\varphi(u) = -\psi(u)$, $g_\varphi(u) = -g_\psi(u)$. В соответствии с известными результатами выпуклого анализа [3], $g_\psi(u)$ получается как вектор невязок в ограничениях (2) при $x = x(u)$, т. е. $g_\psi(u) = \{f_i(x(u))\}_{i=1}^m$.

В соответствии с формулой (6) для $\lambda(u)$ обобщенный градиент функции $\lambda(u)$ в точке $u = \overline{u}$ вычисляется следующим образом:

а) находится минимальное собственное число $\lambda(\overline{u})$ матрицы $K(\overline{u})$ и соответствующий ему собственный вектор $y(\overline{u})$, $\|y(\overline{u})\| = 1$ (если $\lambda(\overline{u})$ – кратное собственное число, то в качестве $y(\overline{u})$ можно выбрать произвольный нормированный собственный вектор, соответствующий $\lambda(\overline{u})$; в этом случае в точке $\overline{u}$ $\lambda(\overline{u})$ не является дифференцируемой);

б) в качестве $g_\lambda(\overline{u})$ берется градиент к линейной по u функции

$$e(u) = \sum_{i,j=1}^{n} y_i(\overline{u}) y_j(\overline{u}) k_{ij}(u).$$

Зная способы вычисления $g_\lambda(u)$ и $g_\varphi(u)$, легко построить алгоритм вычисления субградиента функции $S_{\varepsilon,\delta}(u)$ и применить для ее минимизации один из известных субградиентных методов с растяжением пространства (например, r-алгоритм [3]). Заметим, что при определении шагов по направлению в r-алгоритме нужно позаботиться о том, чтобы не переходить границу области U^+.

Другой способ решения задачи (7) связан с возможностью применения метода эллипсоидов ([1, 4]) или его модификаций [2].

Суть метода эллипсоидов при решении задач выпуклого программирования состоит в следующем: перед началом очередной итерации оптимум локализован в некотором эллипсоиде; берется точка $\overline{x}$ – центр эллипсоида. Если эта точка является допустимой, то с использованием субградиента целевой функции строится отсекающая гиперплоскость и вокруг оставшейся части эллипсоида описывается эллипсоид по возможности меньшего объема (если $\overline{x}$ не является допустимой, то для построения отсекающей гиперплоскости используется субградиент к ограничению, которое в данной точке не выполняется).

Таким образом, после завершения очередной итерации получаем в качестве области локализации оптимума опять эллипсоид, но меньшего объема. Модификации метода эллипсоидов предложены для ускорения сходимости этого метода и состоят в том, что погружение области локализации оптимума в эллипсоид происходит после двух или большего числа отсечений. Отметим, что метод эллипсоидов может быть реализован в форме обобщенного градиентного метода с растяжением пространства [1].

Как следует из сказанного, для реализации метода эллипсоидов и его модификаций при решении задачи (7) необходимо уметь находить $g_{\varphi}(u)$ для $u \in U^{+}$ или $g_{\lambda}(u)$ для $u \notin U^{+}$. На способах вычисления этих обобщенных градиентов останавливались выше.

Процедура вычисления $g_{\lambda}(u)$ основана на решении задачи нахождения собственных чисел и векторов симметричных матриц, которая является довольно трудоемкой, если требуется высокая точность результатов.

Воспользуемся способом, предложенным в [5]. Он позволяет строить отсекающие гиперплоскости, которые отделяют данную точку $\overline{u} \notin U^{+}$ от области U^{+} без использования алгоритма нахождения минимального собственного числа и соответствующего ему собственного вектора.

Пусть $\overline{u} \notin U^{+}$. Это значит, что матрица $K(\overline{u})$ не является положительно определенной. Пусть $M_1(\overline{u}), M_2(\overline{u}), \ldots, M_n(\overline{u})$ – последовательность главных миноров этой матрицы. По критерию Сильвестра найдется такой номер t, $1 \leq t \leq n$, что $d_i = \det M_i(\overline{u}) > 0$, $i = 1, \ldots, t-1$, $d_t = \det M_t(\overline{u}) \leq 0$.

При $t = 1$ $k_{11}(\overline{u}) \leq 0$, и так как при $u \in U^{+}$ $k_{11}(u) > 0$ и $k_{11}(u)$ является линейной функцией от u, то гиперплоскоеть $k_{11}(u) = 0$ отсечет точку $\overline{u}$ от множества U^{+}.

Пусть $t > 1$. Рассмотрим подматрицы $M_t^i(\overline{u})$, которые получаются при вычеркивании t-го столбца и i-й строки из матрицы $M_t(\overline{u})$. Пусть

$$\begin{aligned} \xi_t &= (-1)^i \det M_t^i(\overline{u}), & i = 1, \ldots, t, \\ \xi_t &= \det M_{t-1} \cdot (-1)^t \neq 0, & \xi = \{\xi_i\}_{i=1}^{t}, \end{aligned}$$

$$r_{\overline{u}}(u) = \sum_{i,j=1}^{t} k_{ij}(u)\xi_i\xi_j = \sum_{j=1}^{t} \xi_j \left(\sum_{i=1}^{t} k_{ij}(u)\xi_i \right). \tag{8}$$

При $u \in U^+$ $r_{\overline{u}}(u) > 0$, так как $M_t(u)$ является положительно определенной матрицей.

Пусть $u = \overline{u}$. $\sum\limits_{i=1}^{t} k_{ij}(\overline{u})\xi_i$ в силу определения вектора ξ с точностью до знака совпадает с определителем матрицы, получающейся из матрицы $M_i(\overline{u})$ заменой t–го столбца на i–й. Если $t \neq i$, то получающаяся матрица имеет два одинаковых столбца и ее определитель равен 0.

Таким образом,

$$r_{\overline{u}}(\overline{u}) = \sum_{i,j=1}^{t} k_{ij}(\overline{u})\xi_i\xi_j = \xi_t \cdot (-1)^t d_t = d_{t-1}d_t \leq 0. \qquad (9)$$

Заметим, что $r_{\overline{u}}(u)$ – линейная по u функция. Уравнение $r_{\overline{u}}(u) = 0$ задает гиперплоскость, которая отсекает согласно (9) точку $\overline{u}$ от множества U^+. Необходимый для определения $r_{\overline{u}}(u)$ вектор ξ вычисляется с использованием методов линейной алгебры.

Теперь можно описать алгоритм, основанный на методе эллипсоидов [1, 4] для решения задачи получения оценки (7): $\min \varphi(u)$ при ограничениях $u_i \geq 0,\ i \in I^+;\ \lambda(u) \geq 0$.

Начальная информация. Пусть нам априори известны точка $u^{(0)}$ и число $R_0 > 0$ такое, что оптимальное решение задачи (7) u^* удовлетворяет соотношению $\left\|u^* - u^{(0)}\right\| \leq R_0$.

Задана также матрица преобразования пространства $B_0 = I$ (единичная $(m+n) \times (m+n)$ матрица).

(k + 1)–й шаг алгоритма. $k = 0, 1, 2, \ldots$ Перед началом этого шага имеем точку $u^{(k)}$, число R_k – радиус области локализации оптимума и матрицу B_k преобразования пространства. Проверяем неравенства $u_i \geq 0$, $i \in I^+$. Если для некоторого $\overline{i} \in I^+$ $u_{\overline{i}} < 0$, то берем в качестве $g(u^{(k)})$ вектор, у которого все координаты равны 0, кроме координаты $\overline{i}$, равной -1. В противном случае вычисляем, как описано выше, величины $d_1, d_2, \ldots, d_n$, где $d_i = \det M_t(u^{(k)})$. Если $d_1 \leq 0$, то берем $g(u^{(k)})$ равным антиградиенту функции $k_{11}(u)$. Если $d_1, \ldots, d_{t-1} > 0$, а $d_t \leq 0$, $1 < t \leq n$, то формируем, как уже указано, функцию $r_{u^{(k)}}(u)$ и за $g(u^{(k)})$ принимаем антиградиент этой функции.

Если $d_i > 0$, $i = 1, \ldots, n$, т. е. $u \in U^+$, то, решая относительно x систему линейных уравнений

$$\frac{\partial L(x, u^{(k)})}{\partial x_j} = 0, \quad j = 1, \ldots, n,$$

получаем $x(u^{(k)})$ и формируем компоненты вектора $g(u^{(k)})$ следующим образом:

$$g_i(u^{(k)}) = -f_i(x(u^{(k)})), \quad i = 1, \dots, m,$$
$$g_{m+j}(u^{(k)}) = -x_j^2(u^{(k)}) + x_j(u^{(k)}), \quad j = 1, \dots, n.$$

Получив $g(u^{(k)})$, используем его стандартным способом в методе эллипсоидов [1]. Находим:

$$\begin{aligned}
\xi^{(k)} &= \frac{B_k^T g(u^{(k)})}{\left\| B_k^T g(u^{(k)}) \right\|}, \\
u^{(k+1)} &= u^{(k)} - h_k B_k \xi^{(k)}, \\
B_{k+1} &= B_k \cdot R_\beta(\xi^{(k)}), \quad \beta = \sqrt{\frac{m+n-1}{m+n+1}}, \\
h_k &= \frac{R_k}{m+n-1}, \quad R_{k+1} = R_k \cdot q, \quad q = \frac{m+n}{\sqrt{(m+n)^2-1}}
\end{aligned}$$

и переходим к $(k+2)$–му шагу.

Учитывая, что метод эллипсоидов сходится довольно медленно, на практике целесообразно использовать его ускоренные модификации, например, описанные в [2], при этом способ построения отсекающих гиперплоскостей остается прежним.

Квадратичные экстремальные задачи с булевыми переменными часто встречаются в приложениях (квадратичная задача о назначениях, задача о наилучшей аппроксимации в евклидовой метрике заданной точки линейной целочисленной комбинацией векторов из заданного семейства и др.).

Однако важно отметить, что многие задачи, которые, как правило, формулируются как задачи булева линейного программирования, могут быть переформулированы как квадратичные задачи с булевыми переменными, при этом предлагаемые в данной статье оценки функционалов могут оказаться более точными, чем соответствующие оценки линейного программирования. Особенно это характерно для моделей, отражающих те или иные условия несовместности (задачи о максимальном внутренне устойчивом множестве, о раскраске графов и др.). Так, если булевы переменные x_i и x_j не могут одновременно принимать значения 1, то это условие можно записать в виде $x_i x_j = 0$, т. е. в форме (2) (сравните с линейным ограничением $x_i + x_j \leq 1$). В тех случаях, когда получение оценок линейного программирования требует больших затрат времени ЭВМ, более точные оценки, получаемые из квадратичных моделей, могут оказаться практически более эффективными при реализации метода ветвей и границ.

Отметим, что вопрос частного случая задачи (1)–(3) (ограничения (2) отсутствуют) о получении оценок рассмотрен в работе [7].

Литература

1. Шор Н. З. Метод отсечения с растяжением пространства для решения задач выпуклого программирования // Кибернетика. – 1977. – № 1. – С. 94–95.

2. Шор Н. З., Гершович В. И. Об одном семействе алгоритмов для решения задач выпуклого программирования // Кибернетика. – 1979. – № 4. – С. 62–67.

3. Шор Н. З. Методы минимизации недифференцируемых функций и их приложения. – К.: Наук. думка, 1979. – 199 с.

4. Юдин Д. Б., Немировский А. С. Информационная сложность и эффективные методы решения выпуклых экстремальных задач // Экономика и мат. методы, 1976. – **12**. – № 2. – С. 357–369.

5. Grötshel M., Lovász L., Shrijver A. The ellipsoid method and its consequences in combinatorial optimization // Combinatorica. – 1981. – **1**. – № 2. – P. 169–197.

6. Großmann C., Kaplan A. Strafmethoden und modifizierte Lagrangefunktionen in der nichtlinearen Optimierung. – Leipzig: Teubner Verlagsgesselschaft, 1979. – 200 S.

7. Körner F., Richter Cl. Zur effektiven lösung von booleschen, quadratischen Optimierugs problemen // Numer. Math. – 1982. – **40**. – P. 99–109.

Квадратичные оптимизационные задачи

Н. З. Шор
Известия АН СССР. Техническая кибернетика. – 1987. – № 1. – С. 128–139.

1. Постановка задачи

Под квадратичной оптимизационной задачей (к.з.) будем понимать задачу квадратичного программирования следующего вида:

определить

$$K^* = \inf K_0(x), \quad x \in M \subseteq E^n; \tag{1.1}$$

при ограничениях

$$K_i(x) \leq 0, \quad i \in I; \quad K_i(x) = 0; \quad j \in J; \quad I \cap J = \emptyset, \tag{1.2}$$

где $K_\nu(x) = (A_\nu x, x) + (l_\nu, x) + c_\nu$ – квадратичные функции, определенные на n-мерном евклидовом пространстве E^n, A_ν – симметричные матрицы размерности $n \times n$, l_ν – n-мерные векторы, c_ν – константы, $\nu \in \{0\} \cup I \cup J$, I, J – непересекающиеся конечные подмножества натуральных чисел. M либо совпадает с E^n, либо является выпуклым замкнутым многогранным подмножеством E^n.

Если задача (1.1), (1.2) является несовместной, будем считать, что $K^* = +\infty$. К.з. является далеко идущим обобщением выпуклой задачи квадратичного программирования, методам решения которой посвящена огромная литература (см., например, [1, 2, 19]). Среди к.з. можно выделить достаточно интересные частные случаи:

а) выпуклые к.з. (в.к.з.), когда $J = \emptyset$, A_ν являются неотрицательно-определенными для $\nu \in \{0\} \cup I$;

б) к.з. с линейными ограничениями (к.з.л.о.), когда $A_\nu = 0$ для $\nu \in I \cup J$. Если A_0 – неотрицательно-определенная матрица, то получаем выпуклую задачу квадратичного программирования; в общем случае к.з.л.о. – многоэкстремальная задача и принадлежит к классу NP-трудных задач; в частности, к этому классу относятся задачи вогнутого квадратичного программирования;

в) квадратичные оптимизационные задачи с булевыми переменными (б.к.з.), когда на компоненты вектора x накладываются условия $x_k \in \{0; 1\}$, $k = 1, \ldots, n$. Эти условия эквивалентны квадратичным равенствам $x_k^2 - x_k = 0$, $k = 1, \ldots, n$, т. е. б.к.з. можно рассматривать как частный случай квадратичных оптимизационных задач.

Отметим, что любую задачу нелинейного программирования с полиномиальными целевой функцией и ограничениями можно свести к к.з., вводя новые переменные и квадратичные подстановки следующего вида: $x_i x_j = z_1$ либо $x_i^2 = z_2$ и т. п., понижающие порядок полиномов до квадратичного.

К.з. имеют очень широкую область приложений. Так среди к.з.л.о. отметим задачи проектирования и размещения чипов, задачи линейной дополнительности [18]; среди б.к.з. – квадратичную задачу о назначениях, экстремальные задачи на графах (максимальное внутренне устойчивое множество, задачи раскраски и разрезания графов и др.), играющие большую роль при решении задач оптимального проектирования: размещения, компоновки, разбиения на блоки элементов сложных устройств; задачи выбора, упаковки и др.

По многоэкстремальным к.з. в достаточно общей постановке имеется небольшое число публикаций ([9], [16]). В общем случае, ввиду сложности проблемы, при решении к.з. не избежать перебора, поэтому естественно применять схему метода «ветвей и границ». Для реализации этой схемы следует разработать эффективный и достаточно общий метод получения оценок снизу для целевого функционала в задаче (1.1), (1.2). В данной статье будет исследован двойственный подход к получению нижних оценок, основанный на использовании функции Лагранжа. Такой подход впервые описан в [9]. Он приводит к необходимости решения специальных задач выпуклого программирования: задано некоторое параметрическое семейство симметричных матриц; среди множества значений параметров, при которых матрицы являются неотрицательно-определенными, необходимо найти значение, на котором некоторая функция от параметров достигает максимума. Оказывается, что эти задачи можно достаточно эффективно решать, используя методы негладкой оптимизации. Отметим, что экстремальные задачи на классе неотрицательно-определенных матриц появляются в многочисленных приложениях, не связанных непосредственно с к.з.; разработка эффективных алгоритмов их решения наталкивается на значительные трудности (см. [14], [15]). Поэтому раздел, связанный с построением алгоритмов негладкой оптимизации для решения экстремальных задач на классе неотрицательно-определенных матриц, представляет и самостоятельный интерес.

Далее в работе более детально будут рассмотрены задачи с линейными ограничениями и булевы квадратичные задачи. Раздел, относящийся к б.к.з., основан на совместных работах С. И. Стеценко и автора [4], [8], [20].

2. Алгоритмы получения нижних оценок функционала в квадратичных оптимизационных задачах

Рассмотрим квадратичную оптимизационную задачу (1.1), (1.2). Пусть $|I| = m_1$; $|J| = m_2$; $N = m_1 + m_2$, $\lambda = \{\lambda_1, \ldots, \lambda_{m_1}; \lambda_{m_1+1}, \ldots, \lambda_N\}$ – вектор множителей Лагранжа, у которого первые m_1 компонент соответствуют ограничениям в форме неравенств, а остальные – ограничениям в форме равенств в (1.2), Λ^+ – подмножество N-мерного пространства E^N, состоящее из произвольных векторов, в которых первые m_1 компонент неотрицательны. Введем функцию Лагранжа к.з. (1.1), (2.2) $L(x, \lambda)$, определенную на прямом произведении $M \times \Lambda^+$, следующего вида

$$L(x, \lambda) = K_0(x) + \sum_{i=1}^{N} \lambda_i K_i(x).$$

Пусть T – множество допустимых решений задачи (1.1), (1.2). Если $\overline{x} \in T$, $\overline{\lambda} \in \Lambda^+$, то $\overline{\lambda}_i K_i(\overline{x}) \le 0$ для $\{i : 1 \le i \le m_1\}$; $\overline{\lambda}_i K_i(\overline{x}) = 0$ для $i = m_1 + 1, \ldots, N$. Таким образом, если $\overline{x} \in T$, $\overline{\lambda} \in \Lambda^+$, то $L(\overline{x}, \overline{\lambda}) \le K_0(\overline{x})$, отсюда

$$\Psi(\overline{\lambda}) = \inf_{x \in M} L\left(x, \overline{\lambda}\right) \le \inf_{x \in T} L\left(x, \overline{\lambda}\right) \le \inf_{x \in T} K_0(x) = K^*$$

при любом $\overline{\lambda} \in \Lambda^+$. Значит,

$$\Psi(\lambda) = \inf_{x \in M} L(x, \lambda)$$

служит оценкой снизу для K^* при $\lambda \in \Lambda^+$. Для получения наиболее точной оценки такого типа нужно решить задачу нахождения

$$\Psi^* = \sup_{\lambda \in \Lambda^+} \Psi(\lambda).$$

Функция Лагранжа $L(x,\lambda)$ представима в следующем виде $L(x,\lambda) = (A(\lambda)x,\, x) + (l(\lambda),\, x) + c(\lambda)$, где

$$A(\lambda) = A_0 + \sum_{i=1}^{N} \lambda_i A_i, \quad l(\lambda) = l_0 + \sum_{i=1}^{N} \lambda_i l_i, \quad c(\lambda) = c_0 + \sum_{i=1}^{N} \lambda_i c_i.$$

Заметим, что элементы матрицы $A(\lambda) = \{a_{ij}(\lambda)\}_{i,j=1}^{n}$ линейно зависят от λ. Пусть $D(\overline{D})$ – подмножества E^N, состоящие из таких λ, что $A(\lambda)$ является положительно-определенной (соответственно – неотрицательно-определенной) матрицей. Легко показать, что D и $\overline{D}$ – выпуклые множества, если они не пусты, при этом $\overline{D}$ является замыканием D. На множестве D функция $\Psi(\lambda)$ принимает конечные значения и является вогнутой. В обобщенном смысле она является вогнутой и на $\overline{D}$.

Если $\lambda \in \overline{D}$, то задача нахождения

$$\Psi(\lambda) = \inf_{x \in M} L(x,\lambda)$$

является задачей выпуклого квадратичного программирования и может быть решена практически достаточно эффективно конечно-шаговым алгоритмом. Если $\lambda \notin \overline{D}$, то задача нахождения $\Psi(\lambda)$ является, вообще говоря, многоэкстремальной и трудно разрешимой. Поэтому вместо

$$\Psi^* = \sup_{\lambda \in \Lambda^+} \Psi(\lambda)$$

мы будем рассматривать более слабую оценку

$$\overline{\Psi}^* = \sup_{\lambda \in \overline{D} \cap \Lambda^+} \Psi(\lambda) \le \Psi^*.$$

Если $M = E^n$, то при $\lambda \notin \overline{D}$, $\quad \Psi(\lambda) = -\infty$. Поэтому в этом важном случае $\overline{\Psi}^* = \Psi^*$. Ниже для простоты будем рассматривать случай $M = E^n$. При $\lambda \in D$ матрица $A(\lambda)$ невырождена, и для нахождения $\Psi(\lambda)$ нужно решить систему линейных уравнений $2A(\lambda)x(\lambda) + l(\lambda) = 0$, откуда $x(\lambda) = -\dfrac{1}{2}A^{-1}(\lambda)\, l(\lambda)$,

$$\Psi(\lambda) = -\frac{1}{4}(A^{-1}(\lambda)\, l(\lambda), l(\lambda)) + c(\lambda). \tag{2.1}$$

В области $D \cap \Lambda^+$ $\Psi(\lambda)$ является вогнутой непрерывно дифференцируемой, при этом ее градиент есть

$$g_\Psi(\lambda) = \{K_i(x(\lambda))\}_{i=1}^{N}. \tag{2.2}$$

Задача нахождения
$$\overline{\Psi}^* = \sup_{\lambda \in D \cap \Lambda^+} \Psi(\lambda) \tag{2.3}$$
является задачей выпуклого программирования. Рассмотрим возможные алгоритмы ее решения.

а) Метод эллипсоидов ([5, 11]). Суть метода эллипсоидов при решении задач выпуклого программирования $\min f_0(\lambda) \mid f_i(\lambda) \leq 0$, $i = 1, \ldots, m$, состоит в следующем: перед началом очередной итерации оптимум локализован в некотором эллипсоиде; берется точка $\overline{\lambda}$ – центр эллипсоида. Если эта точка является допустимой, то с использованием субградиента целевой функции строится отсекающая гиперплоскость, и вокруг оставшейся части эллипсоида описывается новый эллипсоид по возможности меньшего объема. Если $\overline{\lambda}$ не является допустимой, то строится гиперплоскость, отделяющая $\overline{\lambda}$ от области допустимых значений λ, и вокруг части эллипсоида, содержащей допустимые точки, описывается новый эллипсоид. В качестве такой отделяющей гиперплоскости можно брать, в частности, гиперплоскость, проходящую через точку $\overline{\lambda}$ ортогонально к субградиенту ограничения, которое в данной точке не выполняется [11]. В обоих случаях ($\overline{\lambda}$ – допустимая или недопустимая точка) после завершения очередной итерации получаем в качестве области локализации оптимума опять эллипсоид, но меньшего объема. Модификации метода эллипсоидов, предложенные для ускорения его сходимости, состоят в том, что погружение области локализации оптимума в новый эллипсоид происходит после двух или большего числа отсечений. Отметим, что метод эллипсоидов может быть реализован в форме метода обобщенного градиентного спуска с растяжением пространства [5]. Как следует из сказанного выше, для реализации метода эллипсоидов и его модификаций при решении задачи (2.3) необходимо уметь находить $g_\Psi(\lambda)$ для $\lambda \in D \cap \Lambda^+$ (это можно делать по формуле (2.2)) и субградиент к функции-ограничению, если $\lambda \overline{\in} D \cap \Lambda^+$. (Если λ принадлежит границе области D, то $g_\Psi(\lambda)$ необязательно определен, поэтому в этом случае также будем использовать субградиент к ограничениям.)

Условие $\lambda \in D \cap \Lambda^+$ содержит два типа ограничений:

а) $\lambda_i \geq 0$ для $i \in I$;

б) $A(\lambda)$ – положительно-определенная матрица, что эквивалентно бесконечной системе линейных по λ неравенств

$$\sum_{i,j=0}^{n} a_{ij}(\lambda) x_i x_j > 0, \quad \forall x \in E^n; \quad x = \{x_1, \ldots, x_n\}, \quad x \neq 0,$$

которую можно переписать в компактном виде

$$\mu(\lambda) = \min_{\{x | \|x\|=1\}} \sum_{i,j=1}^{n} a_{ij}(\lambda) x_i x_j > 0, \tag{2.4}$$

где $\mu(\lambda)$ – минимальное собственное число матрицы $A(\lambda)$.

Замечание. Из выражения (2.4) очевидно, что $\mu(\lambda)$ – вогнутая функция λ. В качестве суперградиента функции $\mu(\lambda)$ в точке $\overline{\lambda}$ $g_\mu(\overline{\lambda})$ можно взять градиент к линейной по λ функции

$$\Psi_{\overline{\lambda}}(\lambda) = \sum_{i,j=1}^{n} a_{ij}(\lambda) y_i(\overline{\lambda}) y_j(\overline{\lambda}), \tag{2.5}$$

где $y_i(\overline{\lambda})$, $i = 1, \ldots, n$, – компоненты нормированного собственного вектора матрицы $A(\overline{\lambda})$, соответствующего минимальному собственному числу $\mu(\overline{\lambda})$. $\mu(\lambda)$ – негладкая функция, ее субградиент терпит разрыв в тех точках λ, в которых матрица $A(\lambda)$ имеет кратное минимальное собственное число.

Для проверки принадлежности $\overline{\lambda}$ множеству $D \cap \Lambda^+$ рассматриваем сначала условие а). Если для некоторого $i = i^* \in I$; $\overline{\lambda}_{i^*} < 0$, тогда в качестве отсекающей гиперплоскости в методе эллипсоидов можно взять $\lambda_{i^*} = 0$.

Если условие а) выполняется, т. е. $\overline{\lambda} \in \Lambda^+$, то проверяем условие б). Для этого определяем $\mu(\overline{\lambda})$ и соответствующий собственный вектор матрицы $A(\overline{\lambda})$. Если $\mu(\overline{\lambda}) \le 0$, то в качестве отсекающей гиперплоскости можно взять следующую $(g_\mu(\overline{\lambda}), \lambda - \overline{\lambda}) = 0$, где $g_\mu(\overline{\lambda})$ совпадает с градиентом функции $\Psi_{\overline{\lambda}}(\lambda)$ (см. (2.5)). Описанная выше процедура проверки условия б) и построения отсекающей гиперплоскости в методе эллипсоидов, если это условие не выполняется, связана с нахождением минимального собственного числа матрицы $A(\lambda)$, что практически эквивалентно по сложности решению задачи на собственные значения для матрицы $A(\lambda)$. Решение этой задачи с высокой точностью требует при большом n значительного времени счета. В [13] в аналогичной ситуации предлагается другой способ построения отсекающей гиперплоскости в методе эллипсоидов. Рассмотрим его применительно к нашей задаче.

Пусть $\mu(\overline{\lambda}) \le 0$. Это значит, что матрица $A(\overline{\lambda})$ не является положительно-определенной. Пусть $R_1(\overline{\lambda}), \ldots, R_n(\overline{\lambda})$ – последовательность главных миноров этой матрицы. По критерию Сильвестра найдется такой номер t, $1 \le t \le n$, что $d_i = \det R_i(\overline{\lambda}) > 0$ для $i = 1, \ldots, t-1$; $d_t = \det R_t(\overline{\lambda}) \le 0$. Если $t = 1$, $d_t = a_{11}(\overline{\lambda}) \le 0$, и так как при $\lambda \in D$

$a_{11}(\overline{\lambda}) > 0$ и $a_{11}(\lambda)$ является линейной функцией от λ, то гиперплоскость $a_{11}(\lambda) = 0$ отсечет точку $\overline{\lambda}$ от множества D. Пусть $t > 1$. Рассмотрим подматрицы $R_t^i(\overline{\lambda})$, которые получаются вычеркиванием t-го столбца и i-й строки $(i \le t)$ из матрицы $R_t(\overline{\lambda})$. Пусть

$$\xi_i = (-1)^i \det R_t^i(\overline{\lambda}), \quad i = 1, \ldots, t, \quad \xi = \{\xi_i\}_{i=1}^t.$$

Введем функцию, линейную по λ

$$\rho_\xi(\lambda) = \sum_{i,j=1}^t a_{ij}(\lambda)\xi_i\,\xi_j = \sum_{j=1}^t \xi_i \left(\sum_{i=1}^t a_{ij}(\lambda)\xi_i \right).$$

Если $\lambda \in D$, $\rho_\xi(\lambda) > 0$, т. к. матрица $a_{ij}(\lambda)$ является положительно-определенной. При $\lambda = \overline{\lambda}$

$$S_j = \sum_{i=1}^t a_{ij}(\overline{\lambda})\xi_i$$

в силу определения вектора ξ с точностью до знака совпадает с определением матрицы, получающейся из матрицы $R_t(\overline{\lambda})$ заменой t-го столбца на j-й. Если $t \ne j$, то эта матрица будет иметь два одинаковых столбца, т. е. $S_j = 0$. Таким образом,

$$\rho_\xi(\overline{\lambda}) = \xi_t \sum_{i=1}^t a_{it}(\overline{\lambda})\xi_t = \xi_t(-1)^t d_t = d_{t-1} d_t \le 0.$$

Значит, гиперплоскость $\rho_\xi(\lambda) = 0$ отделяет точку $\overline{\lambda}$ от области D. Описанный способ построения отделяющей гиперплоскости при рациональном $\overline{\lambda}$ и рациональных коэффициентах квадратичных форм в задаче (1.1), (1.2) требует полиномиального числа операций. Недостатком этого способа являются слишком высокие требования к разрядности ЭВМ, на которой проводятся вычисления.

Для проверки положительной определенности матрицы $A(\overline{\lambda})$ не обязательно использовать правило Сильвестра, можно использовать процедуру выделения полных квадратов, соответствующую следующему разложению матрицы: $A(\overline{\lambda}) = BRB^T$, где $B(B^T)$ – треугольные матрицы, R – диагональная матрица. Процесс выделения полных квадратов ведем до тех пор, пока все диагональные элементы неразложенной части матрицы остаются положительными. Если процесс разложения проходит до конца, то $A(\overline{\lambda})$ – положительно-определенная матрица. Пусть процесс разложения заканчивается на k-ом шаге, $0 \le k \le n$. Тогда квадратичная форма с точностью до нумерации переменных представима в следующем виде:

$$(A(\overline{\lambda})x, x) = \sum_{i=1}^{k} r_{ii}(\overline{\lambda})\left(x_i + \sum_{\{j: i<j\le n\}} b_{ij}(\overline{\lambda})x_j\right)^2 + \Phi_{\overline{\lambda}}(x_{k+1}, \ldots, x_n), \quad (2.6)$$

где $r_{ii}(\overline{\lambda}) > 0, \quad i = 1, \ldots, k; \quad \Phi_{\overline{\lambda}}$ – квадратичная функция, у которой коэффициент при x_{k+1}^2 не превышает 0. Рассмотрим систему уравнений с треугольной матрицей

$$x_i + \sum_{\{j: i<j\le n\}} b_{ij}(\overline{\lambda})x_j = 0; \quad i = 1, \ldots, k; \quad (2.7)$$

$$x_{k+1} = 1; \quad x_{k+2} = x_{k+3} = \ldots = x_n = 0, \quad (2.8)$$

и пусть $y(\overline{\lambda}) \ne 0$ – решение этой системы. В пространстве E^N векторов λ рассмотрим гиперплоскость

$$(A(\lambda)y(\overline{\lambda}),\ y(\overline{\lambda})) = 0. \quad (2.9)$$

(Заметим, что $A(\lambda)$ линейно зависит от λ, т. е. выражение (2.9) действительно задает гиперплоскость.) Из выражения (2.6) и определения $y(\overline{\lambda})$ (см. (2.7), (2.8)) следует, что $(A(\overline{\lambda})y(\overline{\lambda}),\ y(\overline{\lambda})) \le 0$; с другой стороны, для $\lambda \in D$ $(A(\lambda)y(\overline{\lambda}),\ y(\overline{\lambda})) > 0$. Таким образом, гиперплоскость (2.9) отделяет точку λ от области D, т. е. ее можно использовать как отсекающую в методе эллипсоидов. На возможность такого способа построения отсекающей гиперплоскости указал автору А. С. Немировский. Хотя этот способ по форме отличается от предыдущего ([13]), но, по-видимому, связан с близкими вычислительными проблемами неустойчивости алгоритма выделения полных квадратов, когда минимальное собственное число или несколько собственных чисел близки к 0.

б) Метод штрафных функций. Функция $\Psi(\lambda)$ за пределами области D в обычном смысле не определена: $\Psi(\lambda) = -\infty$, если $\lambda \notin \overline{D}$. Поэтому для решения задачи (2.3) нельзя применить метод внешних штрафных функций. Однако остается возможность применения барьерных штрафных функций, например, следующего вида:

$$S(\lambda, s) = \Psi(\lambda) - \frac{s_0}{\mu(\lambda)} - \sum_{i\in I} \frac{s_i}{\lambda_i},$$

где $S = \{s_0; s_i,\ i \in I\}$ – вектор с достаточно большими положительными компонентами. $S(\lambda, s)$ – вогнутая функция λ внутри области $D \cap \Lambda^+$. В любой внутренней точке λ области $D \cap \Lambda^+$ $\quad S(\lambda, s) < \Psi(\lambda) \leq \overline{\Psi}$. Поэтому в качестве нижней оценки функционала задачи (1.1), (1.2) можно брать

$$\sup_{\lambda \in D \cap \Lambda^+} S(\lambda, s).$$

Так как $\mu(\lambda)$ – негладкая функция, то $S(\lambda, s)$ – также негладкая функция, стремящаяся к $-\infty$ при подходе к границе множества $D \cap \Lambda^+$ по внутренним точкам. Поэтому для нахождения

$$\sup_{\lambda \in D \cap \Lambda^+} S(\lambda, S)$$

нужно начинать из некоторой допустимой внутренней точки $\lambda^{(0)}$ и использовать один из монотонных методов негладкой оптимизации, обзор которых содержится в [6], [21].

Для решения задачи (2.3) можно применять также итеративный процесс типа метода «центров», решая на k-й итерации при заданной внутренней точке допустимой области $\lambda^{(k-1)}$ следующую задачу

$$\max_{\lambda} \left[\min \left[\Psi(\lambda) - \Psi(\lambda^{(k-1)});\ \mu(\lambda);\ \lambda_i,\ i \in I \right] \right]. \tag{2.10}$$

В качестве $\lambda^{(k)}$ для следующей итерации нужно брать приближенное решение задачи (2.10), максимизацию в (2.10) можно осуществлять произвольным методом негладкой оптимизации. Специальный вариант метода «центров» описан также в [21].

в) Специальная форма r-алгоритма. Приведенные выше методы получения нижних оценок целевого функционала в задаче (1.1), (1.2) недостаточно эффективны. Методы эллипсоидов при $N \geq 5$ сходятся медленно, методы барьерных функций и «центров» требуют вычисления минимальных собственных чисел матрицы $A(\lambda)$. Поэтому в качестве основного метода для получения оценок нами был выбран модифицированный метод обобщенных градиентов с растяжением пространства в направлении разности двух последовательных градиентов (r-алгоритм). Напомним, что в области $D \cap \Lambda^+$ нужно найти супремум вогнутой функции

$$\Psi(\lambda) = -\frac{1}{4}(A^{-1}(\lambda) l(\lambda), l(\lambda)) + c(\lambda),$$

причем λ входит в матрицу $A(\lambda)$, вектор $l(\lambda)$ в $c(\lambda)$ линейно. Ограничение $\lambda \in \Lambda^+$ легко учитывается путем введения негладкой функции штрафа вида

$$r(\lambda) = \sum_{i \in I} S(\lambda_i),$$

где

$$S(\lambda_i) = \begin{cases} 0, & \lambda_i \geq 0; \\ s\lambda_i, & \lambda_i < 0; \end{cases}$$

s – достаточно большое положительное число. Таким образом, будем рассматривать задачу нахождения $\sup\limits_{\lambda \in D} \overline{\Psi}(\lambda)$, где $\overline{\Psi}(\lambda) = \Psi(\lambda) + r(\lambda) =$ $= -\dfrac{1}{4}(A^{-1}(\lambda)l(\lambda), l(\lambda)) + c(\lambda) + r(\lambda)$. Допустим, что $\lambda^{(0)}$ – начальная точка области D. $\sup\limits_{\lambda \in D} \overline{\Psi}(\lambda)$ может достигаться либо в области $D(\lambda^* \in D)$, либо на границе этой области $(\lambda^* \in \overline{D} \backslash D)$, причем второй случай практически встречается наиболее часто. Поэтому имеет смысл рассмотреть поведение функции $\Psi(\lambda)$ вблизи границы области D. Функция $c(\lambda)$ линейна, а $r(\lambda)$ – кусочно-линейна, поэтому в любой ограниченной области они ограничены. Особенности функции $\overline{\Psi}$ на множестве $\overline{D} \backslash D$ возникают в связи с поведением нелинейного выражения $(A^{-1}(\lambda)l(\lambda), l(\lambda))$. Пусть $\overline{\lambda} \in \overline{D} \backslash D$. Матрица $A(\overline{\lambda})$ вырождена. Обозначим размерность инвариантного подпространства $E_0(\overline{\lambda})$ матрицы $A(\overline{\lambda})$, соответствующего нулевому собственному значению, через $m(\overline{\lambda})$, $m(\overline{\lambda}) \geq 1$. Если проекция $l(\overline{\lambda})$ на $E_0(\overline{\lambda})$ отлична от 0, то

$$\lim_{\lambda \in D; \lambda \to \overline{\lambda}} (A^{-1}(\lambda)l(\lambda), l(\lambda)) = +\infty, \quad \text{т. е.} \quad \widetilde{\Psi}(\lambda) = \lim_{\lambda \in D; \lambda \to \overline{\lambda}} \overline{\Psi}(\lambda) = -\infty.$$

Пусть $S_\varepsilon(\lambda)$ – ε-окрестность точки λ. Тогда, для того, чтобы

$$\Psi^+(\overline{\lambda}) = \lim_{\varepsilon \to 0} \sup_{D \cap S_\varepsilon(\overline{\lambda})} \Psi(\lambda) > -\infty,$$

необходимо, чтобы $l(\overline{\lambda})$ было ортогонально $E_0(\lambda)$. Это условие эквивалентно разрешимости уравнения

$$A(\overline{\lambda})x + l(\overline{\lambda}) = 0. \tag{2.11}$$

Пусть Q – подмножество симметричных матриц $A(\lambda)$, $\lambda \in \overline{D} \backslash D$, q – размерность этого множества, рассматриваемого как многообразие в пространстве симметричных матриц размерности $n(n+1)/2$. Условие разрешимости уравнения (2.11) может быть выражено в виде одного или нескольких алгебраических уравнений относительно $\overline{\lambda}$, и, вообще говоря, понижает размерность. Таким образом, размерность многообразия матриц $A(\overline{\lambda})$, принадлежащих Q, для которых $\Psi^{+}(\overline{\lambda}) > -\infty$ меньше q при $q \geq 1$. Это означает, что при движении из области D по некоторому направлению, ведущему в граничную точку $\overline{\lambda} \in \overline{D} \backslash D$, «вероятность» того, что функция $\Psi(\lambda)$ будет монотонно возрастать на открытом интервале, заканчивающемся точкой $\overline{\lambda}$, равна 0.

Функция $\overline{\Psi}(\lambda)$ как бы «самоэкранируется» от выхода за пределы области своего определения при использовании алгоритмов монотонного подъема, т. е., как правило, отпадает необходимость в использовании барьерных штрафных функций. Появляется возможность использования монотонных вариантов r-алгоритма [6], [10]. Выход за пределы области D контролируется путем применения процедуры треугольного разложения матрицы $A(\lambda)$. Если на данном шаге произошел выход за пределы области D, то шаг дробится до тех пор, пока не будет получено допустимое значение. На основе указанной модификации r-алгоритма С. И. Стеценко разработал программу для решения некоторых булевых квадратичных задач и провел успешные вычислительные эксперименты. Более подробно об этом говорится в разд. 4.

3. Квадратичные экстремальные задачи с линейными ограничениями

Задачу минимизации квадратичной функции при линейных ограничениях (к.з.л.о.) запишем в следующем виде:

найти

$$\inf K_0(x),\ K_0(x) = (A_0 x,\, x) + (b,\, x) \tag{3.1}$$

при следующих ограничениях

$$l_j(x) = (l_j, x) + c_j \leq 0; \quad x = \{x_1, \ldots, x_n\} \in E^n, \tag{3.2}$$

$l_j = \{l_j^{(1)}, \ldots, l_j^{(n)}\} \in E^n,\ c_j$ – константы, $j = 1, \ldots, m$. Невыпуклые задачи квадратичного программирования возникают в ряде приложений. В частности, к ним сводятся проблемы линейной дополнительности, играющие большую роль в анализе макроэкономических моделей, задачи проектирования электронных модулей (чипов) на интегральных схемах и другие. В литературе описаны несколько подходов к решению такого рода задач, в частности, связанные с процедурой отсечения Бендерса (см. [18]), исследования в рамках общих схем решения многоэкстремальных задач или схем вогнутого программирования. В одной из недавних работ [18] в основе алгоритма лежит представление квадратичной формы в виде суммы выпуклой и вогнутой квадратичных функций и метод ветвей и границ. Для получения нижних оценок целевой функции вогнутая часть аппроксимируется линейной или кусочно-линейной функцией. В последнем случае задача оценки оказывается задачей смешанного типа: с непрерывными и булевыми переменными. Несмотря на использование мощной ЭВМ Cray 1S ($\sim\ 10^8$ операций в секунду), решение задач средней размерности (300 выпуклых переменных, 30 вогнутых, 50 ограничений) наталкивается на серьезные трудности. Так, достигнутая за счетное время 1500 с точностью всего порядка 3%. Поэтому актуальным остается вопрос о разработке новых практически эффективных алгоритмов.

Непосредственное применение предложенного нами выше двойственного подхода к получению нижних оценок целевой функции в общем случае к.з.л.о. ничего не дает, так как при изменении множителей Лагранжа меняется лишь линейная по x часть функции Лагранжа, квадратичная часть остается без изменений. Поэтому, если A_0 не является неотрицательно-определенной, то при любых множителях Лагранжа будем иметь тривиальную оценку: $\Psi(\lambda) = -\infty$.

Однако мы можем использовать линейные ограничения (3.2) для образования из них квадратичных следствий

$$K_{pj}(x) = -l_p(x)l_j(x) \leq 0; \quad p,\, j = 1, \dots, m. \tag{3.3}$$

Рассмотрим к.з. (3.1)–(3.3). Эта задача эквивалентна задаче (3.1), (3.2), однако формально содержит дополнительные ограничения, которые позволяют воздействовать на квадратичную часть функции Лагранжа. Обозначим множители Лагранжа, соответствующие ограничениям (3.2), через λ_{pj} $(p,\, j = 1, \dots, m)$. Покажем, что при достаточно общих условиях множество $D \cap \Lambda^+$ задачи (3.1)–(3.3) не будет пустым.

Теорема. *Если выпуклая оболочка векторов l_j, $j = 1, \dots, m$, в ограничениях (3.2) содержит в качестве внутренней точки 0, то найдется неотрицательный вектор множителей Лагранжа $\{\lambda_{pj}\}_{p,j=1}^m$, при котором матрица $A(\lambda)$ становится положительно-определенной.*

Доказательство. Из условия теоремы следует, что произвольный вектор $y \in E^n$ может быть представлен в виде линейной комбинации с неотрицательными коэффициентами векторов l_j, $j = 1, \ldots, m$. Пусть e_i – орт i-й координаты, $i = 1, \ldots, n$. Тогда существует представление

$$e_i = \sum_{p=1}^{m} u_{ip}^{(+)} l_p; \qquad -e_i = \sum_{q=1}^{m} u_{iq}^{(-)} l_q, \tag{3.4}$$

где $u_{ip}^{(+)}$, $u_{iq}^{(-)}$ – неотрицательные числа, p, $q = 1, \ldots, m$; $i = 1, \ldots, n$. Из (3.4) получаем соотношения

$$x_i = \sum_{p=1}^{m} u_{ip}^{(+)} (l_p, x); \quad -x_i = \sum_{q=1}^{m} u_{iq}^{(-)} (l_q, x), \quad i = 1, \ldots, n.$$

Запишем тождества

$$x_i^2 = \sum_{p,q=1}^{m} u_{ip}^{(+)} u_{iq}^{(-)} \overline{K}_{pq}(x), \qquad i = 1, \ldots, n,$$

где $\overline{K}_{pq}(x) = -(l_p, x)(l_q, x)$ и отличается от $K_{pq}(x)$ лишь линейными членами ($p, q = 1, \ldots, m$). Взяв вектор множителей Лагранжа

$$\{\lambda_{pq}\}_{p,q=1}^{m} = \left\{ \sum_{i=1}^{n} r_i u_{ip}^{(+)} u_{iq}^{(-)} \right\}_{p,q=1}^{m},$$

где $r_i > 0$, $i = 1, \ldots, n$, мы увеличим диагональный элемент матрицы $A_0 a_{ii}^0$ на r_i, оставляя недиагональные элементы без изменения. При достаточно больших r_i матрица функции Лагранжа станет положительно-определенной.

Следствие. *Если многогранное множество W, образованное ограничениями (3.2), непусто и ограничено, то найдутся неотрицательные множители Лагранжа $\{\lambda_{pq}\}_{p,q=1}^{m}$, при которых функция Лагранжа будет неотрицательно-определенной по x.*

Доказательство. Рассмотрим произвольный вектор $s \in E^n$, $\|s\| = 1$ и допустимую точку $\overline{x}_0$. Тогда $\max\limits_{1 \le j \le m} (s, l_j) > 0$, так как в противном случае весь луч $x(t) = \overline{x}_0 + st$, $t \ge 0$, состоял бы из допустимых точек, что противоречит ограниченности W. Отсюда следует, что

$$\min_{\{s, \|s\|=1\}} \max_{1 \le j \le m} (s, l_j) > 0,$$

но это условие выражает тот факт, что выпуклая оболочка $\{l_j\}_{j=1}^m$ содержит 0 в качестве своей внутренней точки, т. е. условие теоремы выполняется.

Таким образом, искусственный прием, связанный с введением ограничений (3.3), позволяет применить разработанную методику получения нижних оценок и к задачам с линейными ограничениями.

4. Булевы квадратичные задачи

Среди оптимизационных задач вида (1.1), (1.2) булевы квадратичные задачи (б.к.з.) выделяются наличием в числе ограничений (1.2) специальных булевых ограничений вида $x_j^2 - x_j = 0$; $j = 1, \ldots, n$. Вводя эти ограничения с произвольными множителями Лагранжа в целевую функцию, можно произвольным образом менять диагональные элементы матрицы $A(\lambda)$, не изменяя недиагональные элементы. Легко подобрать множители Лагранжа так, чтобы $A(\lambda)$ стала положительно-определенной; значит, множество D в задачах б. к. з. непусто.

Типичным примером б. к. з. являются задачи нахождения максимального взвешенного внутренне устойчивого множества графов, т. е. совокупности вершин графа, попарно не связанных ребрами, с максимальным суммарным весом.

Пусть задан граф $G(V, E)$, где V – множество вершин, E – множество ребер; $V = \{1, \ldots, n\}$; ребра, соединяющие вершины $i, j \in V$, будем обозначать (i, j). Сопоставим каждой вершине $i \in V$ булеву переменную x_i и вес $c_i > 0$. Тогда задача нахождения максимального взвешенного внутренне устойчивого множества (МВ ВУМ) состоит в нахождении

$$\alpha_c(G) = \max \sum_{i=1}^{n} c_i x_i \tag{4.1}$$

при ограничениях

$$x_i x_j = 0; \qquad \forall (i,\ j) \in E; \tag{4.2}$$

$$x_i^2 - x_i = 0; \quad i = 1, \ldots, n. \tag{4.3}$$

Задача МВ ВУМ играет большую роль в многочисленных приложениях: теория информации и кодирование, проектирование различных устройств при определенных условиях несовместности; она тесно связана с известными задачами выбора, разбиения множеств, раскраски графов и другими комбинаторными задачами, имеющими, в свою очередь, массу приложений. С другой стороны, она принадлежит к классу NP-полных задач (при целочисленных весах вершин). О сложности этой задачи говорит то, что даже в частном случае, когда все $c_i = 1$, $i = 1, \ldots, n$, не найден полиномиальный алгоритм, который бы гарантировал (сколь-угодно большую) фиксированную относительную погрешность решения по функционалу для любого графа $D(V, E)$.

Обычно для решения задач нахождения МВ ВУМ используются методы «ветвей и границ», практическая эффективность которых во многом определяется алгоритмами получения оценок целевой функции.

Для задачи МВ ВУМ в связи с исследованиями по теории кодирования в [12, 13], [17] были предложены несколько эквивалентных оценок. Например, следующая: пусть $\overline{m}_G$ – класс неотрицательно-определенных матриц $B = \{b_{ij}\}_{i,j=1}^{n}$, для которых $b_{ij} = 0$, если $i \neq j$; $(i,j) \in E$; $\sum\limits_{i=1}^{n} b_{ii} \leq 1$; остальные элементы произвольные.

Тогда

$$\nu(G) = \max_{B \in \overline{m}_G} \sum_{i,j=1}^{n} \sqrt{c_i c_j}\, b_{ij} \geq \alpha_c(G). \tag{4.4}$$

В [4] показано, что описанный в данной статье двойственный подход к получению оценок в к.з. дает для задачи (4.1)–(4.3) оценку, совпадающую по величине с $\nu(G)$. Автор совместно с С. И. Стеценко разработал алгоритм, основанный на методе «ветвей и границ», в котором в качестве оценок используются описанные в данной статье оценки вида (2.3), и провели вычислительные эксперименты. Нахождение оценок производилось с помощью специальной модификации метода субградиентного типа с растяжением пространства – r-алгоритма (см. разд. 3). Эксперименты производились на графах с числом вершин от 30 до 60 при средней степени вершин 4, ребра генерировались случайным образом, веса представляли собой случайные целые числа, большие 20 и меньшие 40. Как видно из [4], размерность задачи по x равна $|V|$ – числу вершин, а по λ – числу ребер. Во всех экспериментах было получено оптимальное решение, причем первая оценка отличалась от оптимального значения не более чем на 5%, благодаря чему количество вершин в дереве ветвлений было небольшим (от 3 – 5 при $|V| = 30$ до нескольких десятков при $|V| = 60$). Время счета на ЕС-1060 менялось от 1,5 – 2 мин. при $|V| = 30$ до 1 часа при $|V| = 60$. Эксперименты показали устойчивую работу алгоритма оценки и хорошее качество оценок.

5. Экстремальные задачи на классе неотрицательно-определенных матриц

Получение оценок в к.з. (1.1), (1.2) приводит к специальным экстремальным задачам на классе положительно-определенных матриц. К такого же типа задачам приводит вычисление оценки Ловаса (4.4), статистические задачи обработки результатов экзаменов [14, 15] и ряд других задач обработки результатов наблюдений, связанных с корреляционными матрицами. Так как множество неотрицательно-определенных матриц представляет собой выпуклое множество с негладкой границей, при решении такого рода задач по существу оказываются полезными методы недифференцируемой оптимизации. Попытки решения этих задач классическими методами нелинейного программирования наталкиваются на серьезные трудности [14], [15], [13].

Нами накоплен опыт решения ряда экстремальных задач на классе неотрицательно-определенных матриц в связи с проблемами оценки хроматического класса графов и хроматического числа графов специального вида ([3], [8], [20]). Вычислительные эксперименты показали, что r-алгоритм в сочетании с методом точных негладких штрафных функций, если целевая функция определена на всем пространстве, является достаточно надежным и эффективным средством решения такого рода задач.

Таким образом, недифференцируемая оптимизация открывает хорошие перспективы решения широкого круга практически важных экстремальных задач на классе неотрицательно-определенных матриц.

Литература

1. Зойтендейк Г. Методы возможных направлений. – М.: Изд-во иностр. лит., 1963.

2. Кюнци Г. П., Крелле В. Нелинейное программирование. – М.: Сов. радио, 1963.

3. Стеценко С. И. Вычислительные эксперименты по решению квадратичных задач с булевыми переменными. – В кн.: Методы решения сложных задач математического программирования. (Сб. научных трудов ИК АН УССР.) Киев: ИК АН УССР, 1985.

4. Стеценко С. И., Шор Н. З. Связь оценок Ловаса с двойственными оценками в квадратичных булевых задачах. – В кн.: Методы решения нелинейного и дискретного программирования. (Тр. семинара НС по кибернетике АН УССР.) Киев: ИК АН УССР, 1984.

5. Шор Н. З. Метод отсечения с растяжением пространства для решения задач выпуклого программирования // Кибернетика. – 1977. – № 1. – С. 94–95.

6. Шор Н. З. Методы минимизации недифференцируемых функций и их приложения. – К.: Наук. думка, 1979. – 199 с.

7. Шор Н. З., Гершович В. И. Метод эллипсоидов, его обобщения и приложения. // Кибернетика. – 1982. – №5. – С. 61–69.

8. Шор Н. З., Стеценко С. И. Получение оценок для хроматического класса графов и использованием моделей булева квадратичного программирования // Тезисы Республиканского семинара по дискретной оптимизации (г. Ужгород). – К., 1985. – С. 135–136.

9. Шор Н. З., Давыдов А. С. О методе получения оценок в квадратичных экстремальных задачах с булевыми переменными // Кибернетика. – 1985. – № 2. – С. 48–50.

10. Шор Н. З., Алиев И. Т. Об одной монотонной модификации r-алгоритма. – В кн.: Методы решения сложных задач математического программирования. (Сб. научных трудов ИК АН УССР.) Киев: ИК АН УССР, 1985. – С. 45–50.

11. Юдин Д. Б., Немировский А. С. Информационная сложность и эффективные методы решения выпуклых экстремальных задач. // Экономика и мат. методы. – 1976. – Т. 12. – № 2.

12. Eliece Mc., Rodemich E. R., Rumsey Jr. H. G. The Lovàsz bound and generalization. // Journal of Combinatorics, Information and System Sciences. – 1978. – V. 3.

13. Grötschel M., Lovasz L., Schriyver A. Polynomial algorithms for perfect graphs: (Rep.) // Institut fur okonometrie und operations research. – WP 81176-OR. – Bonn, 1981. – 40 p.

14. Fletcher R. A nonlinear programming problem in Statistics (educational testing) // SIAM J. Sci. Stat. Comput. – 1981. – V. 2. – № 3.

15. Fletcher R. Semi–definite matrix constraints in optimization // SIAM J. of Control and Optimization. – 1985. – V. 23. – № 4.

16. Korner F., Richter Cl. Zur effektiven lösung von booleschen quadratischen optimierugs problemen // Numer. Math. – 1982. – V. 40.

17. Lovász L. On the Shannon Capacity of a graph // IEEE Trans. Inform. Theory – 1979. – V. 25. – № 1.

18. Pardalos P. M., Glick J. H., Rosen G. B. Global minimization of indefinite quadratic problems. // Report CS–85–31. – The Pennsilvania State University. Department of Computer Sciences, 1985.

19. Wolfe P. The simplex method for quadratic programming // Econometrica. – 1959. – № 27.

20. Shor N. Z., Stecenko S. J. Quadratic boolean problems and Lovász's bounds // Abstracts JFIP conference on system modelling and optimization. – Budapest, 1985.

21. Kiwiel K. C. Methods of descent for nondifferentiable optimization. – Lecture Notes in Mathematics. – Berlin: Springer–Verlag, 1985. – 1133:362.

Об одном подходе к получению глобальных экстремумов в полиномиальных задачах математического программирования

Н. З. Шор
Кибернетика. – 1987. – № 5. – С. 102–106.

Рассмотрим евклидово пространство E^n с векторами $z = \{z_1, z_1, \ldots, z_n\}$. Под полиномиальной задачей математического программирования будем понимать задачу следующего вида: найти

$$\inf P_0(z) \tag{1}$$

при ограничениях

$$P_i(z) = 0, \quad i = 1, \ldots, m, \tag{2}$$

где $P_0(z),\ P_i(z),\ i = 1, \ldots, m$ – полиномиальные функции от z.

То, что ограничения (2) заданы в форме равенств, существенно не ограничивает общность задачи, так как любое полиномиальное неравенство вида $R(z) \leq 0$ можно свести к полиномиальному равенству $R(z) + t^2 = 0$, где t – дополнительная переменная.

Вводя новые переменные и используя квадратичные подстановки вида $z_i^2 = y_i,\ z_{jk} = z_j z_k,\ y_i^2 = v_i$ и т. п., можно снизить степень полиномов в (1), (2) до квадратичной, рассматривая их как функции от расширенного множества переменных, при этом появляются новые квадратичные равенства, соответствующие указанным выше подстановкам. Таким образом, любую задачу вида (1)–(2) можно свести к квадратичной экстремальной задаче: найти

$$\inf_x K_0(x), \quad x \in E^{\overline{n}}, \quad \overline{n} \geq n, \tag{3}$$

при ограничениях:

$$K_i(x) = 0, \quad i = 1, \ldots, \overline{m}; \quad \overline{m} \geq m. \tag{4}$$

где $K_\nu(x),\ \nu = 0, 1, \ldots, \overline{m}$ – квадратичные функции. В работе [1] для получения оценок снизу оптимального значения целевой функции предложен метод, который будем называть двойственным, так как он использует множители Лагранжа.

Рассмотрим функцию Лагранжа: $L(x,u) = K_0(x) + \sum_{i=1}^{\overline{m}} u_i K_i(x) = (A(u)x,x) + (l(u),x) + c(u)$. Здесь $A(u)$ – симметричные матрицы $\overline{n}\times\overline{n}$, $l(u)$ – векторы размерности $\overline{n}$, $c(u)$ – константы, зависящие от вектора множителей Лагранжа $u = \{u_1,\ldots,u_{\overline{m}}\}$. Пусть Ω $(\overline{\Omega})$ – множество таких значений $u = \{u_1,\ldots,u_{\overline{m}}\}$, при которых матрица $A(u)$ положительно определенная (неотрицательно определенная). При $u \in \Omega$ $\min_x L(x,u) = \psi(u)$ достигается в некоторой точке $x(u)$, являющейся решением линейной системы уравнений

$$2A(u)x + l(u) = 0. \tag{5}$$

Определим $\psi(u) = L(x(u),u)$ при $u \in \Omega$. Для любого допустимого $\overline{x}$ $L(\overline{x},u) = K_0(\overline{x})$, поэтому $\psi(u) \le K_0(x)$ для произвольного u. Отсюда $\psi^* = \sup_{u\in\Omega}\psi(u) \le K_0(\overline{x})$ для произвольного допустимого $\overline{x}$, т. е. $\psi^* \le f^*$, где f^* – оптимальное значение целевой функции в задаче (3)–(4) (если система уравнений (4) несовместна, будем считать, что $f^* = +\infty$), $\psi(u)$ – вогнутая функция на $\overline{\Omega}$, $\overline{\Omega}$ – выпуклое, либо пустое множество. В последнем случае будем считать, что $\psi^* = -\infty$.

Задача нахождения $\psi^* = \sup_{u\in\Omega}\psi(u)$ относится к выпуклому программированию. Для нее разработаны довольно эффективные алгоритмы, в частности, хорошие практические результаты дал r-алгоритм со специальной регулировкой шагового множителя [2]. Задачи вида (1)–(2) и соответствующие им (3)–(4) могут быть невыпуклыми и многоэкстремальными. Поэтому нельзя в общем случае гарантировать равенство $\psi^* = f^*$. Одним из подходов к получению глобального оптимума может быть метод ветвей и границ с использованием указанных выше двойственных оценок. Примером такого подхода может служить алгоритм нахождения максимального взвешенного независимого множества вершин графа, описанный в [3]. Однако метод ветвей и границ не для всех задач достаточно эффективен. Интересно исследовать такие классы полиномиальных невыпуклых задач, для которых $\psi^* = f^*$. Отметим, что при переходе от полиномиальных задач к квадратичным можно использовать различные квадратичные подстановки и в зависимости от этого будет меняться ψ^*. Кроме того, можно генерировать новые ограничения, которые являются алгебраическими следствиями прежних, оставляя область допустимых решений без изменения. При этом двойственные оценки не убывают, а в некоторых случаях возрастают, и увеличиваются шансы совпадения ψ^* и f^*.

Легко доказывается следующая теорема.

Теорема 1. *Если* $\psi^* = \sup\limits_{u\in\Omega} \psi(u)$ *достигается на множестве* Ω, *то* $\psi^* = f^*$.

Доказательство. Градиент функции $\psi(u)$ при $u \in \Omega$ совпадает с вектором невязок

$$g_\psi(u) = \{K_i(x(u))\}_{i=1}^{\overline{m}} .$$

Если максимум $\psi(u)$ достигается в некоторой точке $u^* \in \Omega$, то $K_i(x(u^*)) = 0$, $i = 1, \ldots, \overline{m}$, т. е. $x(u^*)$ – допустимая точка, при этом $f^* \leq K_0(x(u^*)) = \psi(u^*) = \psi^*$. Но ψ^* – нижняя оценка для f^*. Отсюда следует, что $f^* = \psi^*$, что и требовалось доказать.

Таким образом, если $\psi^* < f^*$, то супремум $\psi(u)$ достигается на границе области Ω. Пусть u^* – точка на $\overline{\Omega} \setminus \Omega$, в любой окрестности которой при заданном произвольно $\varepsilon > 0$ найдутся точки $u \in \Omega$, для которых справедливо $\psi^* - \psi(u) < \varepsilon$. При этом ранг матрицы $A(u^*)$ должен совпадать с рангом присоединенной матрицы $(A(u^*)|l(u^*))$, т. е. система (5) должна иметь при $u = u^*$ решение. Поясним сказанное на примерах.

I. Найти $\min\,[(Ax,\, x) + (c, x)]$ при ограничениях $(x,\, x) - 1 = 0$, $x \in E^n$.

Приведем матрицу A к диагональному виду путем ортогонального преобразования. Получаем в новых переменных $y = \{y_1, \ldots, y_n\}$ следующую задачу: найти $\min K_0(y)$, где

$$K_0(y) = \sum_{i=1}^{n} \lambda_i y_i^2 + \sum_{i=1}^{n} a_i y_i, \tag{6}$$

при ограничениях

$$K_1(y) = \sum_{i=1}^{n} y_i^2 - 1 = 0. \tag{7}$$

Пусть $\lambda_1, \ldots, \lambda_n$ – собственные числа матрицы A, записанные в неубывающем порядке с учетом их кратности, и минимальное собственное число имеет кратность 1. Рассмотрим функцию Лагранжа задачи (6)–(7):

$$L(y, u) = K_0(y) + uK_1(y).$$

При $u > -\lambda_1$ $L(y, u)$ является положительно определенной. $L(y, u)$ можно переписать в следующем виде:

$$L(y, u) = \sum_{i=1}^{n} (\lambda_i + u) \left(y_i + \frac{a_i}{2(\lambda_i + u)} \right)^2 - \sum_{i=1}^{n} \frac{a_i^2}{4(\lambda_i + u)} - u.$$

При $u+\lambda_1>0$

$$\psi(u)=\min_x L(x,u)=-\frac{1}{4}\sum_{i=1}^{n}\frac{a_i^2}{\lambda_i+u}-u.$$

Если $a_1\neq 0$, то при $u>-\lambda_1$, $u\to -\lambda_1$, $\psi(u)\to -\infty$, при $u\to +\infty$, $\psi(u)\to -\infty$. Таким образом, $\max\psi(u)$ достигается в области положительной определенности: $-\lambda_1<u^*<+\infty$, и ему соответствует глобальный экстремум задачи (6)–(7). Если $a_1=0$, то $\max\limits_u\psi(u)$ может достигаться как на границе области положительной определенности ($u+\lambda_1=0$), так и внутри, при этом $\psi^*=f^*$. Аналогичные рассуждения проходят и в случае кратных собственных чисел матрицы A.

II. Рассмотрим задачу минимизации полинома 4-й степени от одной переменной

$$\min(x^4+ax^2+bx).$$

Произведем подстановку: $x^2-y=0$. Получаем следующую квадратичную задачу:

$$\min(y^2+ay+bx)$$

при ограничении $x^2-y=0$. Функция Лагранжа имеет следующий вид (при $u>0$):

$$\begin{aligned} L(x,y,u) &= y^2+ay+bx+u(x^2-y)= \\ &= \left(y+\frac{a-u}{2}\right)^2+u\left(x+\frac{b}{2u}\right)^2-\left(\frac{a-u}{2}\right)^2-\frac{b^2}{4u}. \end{aligned}$$

При $u>0$ (в области положительной определенности по y,x функции Лагранжа) $\psi(u)=-\left(\dfrac{a-u}{2}\right)^2-\dfrac{b^2}{4u}$. При $b\neq 0$, $u\to +0$, $\psi(u)\to -\infty$. При $u\to +\infty$, $\psi(u)\to -\infty$. Значит, $\max\limits_{u>0}\psi(u)$ достигается в точке u^*, $0<u^*<+\infty$ области $\Omega: u>0$ и ему будет соответствовать единственный глобальный оптимум $x=-\dfrac{b}{2u^*}$. При $b=0$, $a\geq 0$, имеем $u^*=a$, $\psi(u^*)=0$, $x^*=0$. При $b=0$, $a<0$, $\sup\limits_{u\geq 0}\psi(u)$ достигается в точке на границе Ω, $u^*=0$, при этом $y^*={x^*}^2=-\dfrac{a}{2}$, $\psi^*=-\dfrac{a^2}{4}$, что соответствует двум различным глобальным минимумам: $x^*=\pm\sqrt{-\dfrac{a}{2}}$. Таким образом, во всех случаях $\sup\limits_{u>0}\psi(u)=\psi^*=f^*$.

Оказывается, что результат, полученный для многочленов 4-й степени можно обобщить на многочлены от одной переменной произвольной четной степени. Рассмотрим многочлен четной степени $2n$, $n \geq 1$, от переменной x_1, у которого коэффициент при старшей степени равен 1:

$$P_{2n}(x_1) = x_1^{2n} + \sum_{k=1}^{2n} a_{2n-k} x_1^{2n-k}. \tag{8}$$

Введем обозначения: $x_k = x_1^k;\ k = 0, 1, \ldots, n$. В этих обозначениях (заметим, что $x_0 = 1$) задача минимизации $P_{2n}(x_1)$ превращается в квадратичную экстремальную задачу следующего вида:

найти минимум

$$K_{2n}(x) = x_n^2 + \sum_{k=1}^{n-1} a_{2n-k} x_n x_{n-k} + \sum_{i=0}^{n} a_{n-i} x_{n-i} \tag{9}$$

при ограничениях:

$$x_p x_q - x_r x_s = 0, \quad p + q = r + s \leq 2n - 2, \quad p \geq q, \quad p > r \geq s, \tag{10}$$

где $p,\ q,\ r,\ s$ – неотрицательные целые числа. Выражение, стоящее в левой части равенств вида (10), обозначим $R(p, q;\, r, s)$, множество различных допустимых четверок $(p, q;\, r, s)$ в (10) обозначим Q_{2n}.

Замечание. Очевидно, что среди равенств вида (10) часть является алгебраически избыточными, некоторые равенства являются даже линейными следствиями других, именно их можно отбросить без ущерба для дальнейших рассуждений. Мы сознательно сохраняем все множество выражений $R(p, q;\, r, s)$ для упрощения обозначений и доказательств.

Каждому равенству вида $R(p, q;\, r, s) = 0$ сопоставим множитель Лагранжа $\lambda(p, q;\, r, s)$. Совокупность множителей Лагранжа образует вектор λ.

Пусть $x = (x_1, \ldots, x_n)$, $x_0 = 1$.

Образуем функцию Лагранжа квадратичной задачи (9)–(10):

$$L(x, \lambda) = K_{2n}(x) + \sum_{(p,q;\, r,s) \in Q_{2n}} \lambda(p, q;\, r, s) R(p, q;\, r, s).$$

Пусть $\Omega(P_{2n})$ ($\overline{\Omega}(P_{2n})$) – множество таких векторов λ, при которых квадратичная по x функция $L(x, \lambda)$ является положительно (неотрицательно) определенной. Функция $\psi(\lambda) = \min\limits_x L(x, \lambda)$ определена при $\lambda \in \Omega(P_{2n})$. Если супремум $\psi(\lambda)$ при $\lambda \in \overline{\Omega}(P_{2n})$ совпадает с $f^* = \min\limits_{x_1} P(x_1)$, то будем говорить, что P_{2n} обладает ω-свойством. Если этот супремум достигается на $\lambda^* \in \Omega(P_{2n})$, то будем говорить, что полином P_{2n} обладает сильным ω-свойством, при этом (см. теорему 1) $\psi(\lambda^*)$ совпадает с f^*, т. е. P_{2n} обладает ω-свойством.

Покажем, что ω-свойство сохраняется при сдвиге начала координат. Итак, справедлива

Теорема 2. *(О сдвиге). Если $P(x_1)$ обладает ω-свойством, то и полином $P(x_1) = P(x_1 + a)$ обладает ω-свойством при произвольном a.*

Доказательство. Введем переменную $\overline{x}_1 = x_1 - a$. Тогда $\overline{P}(\overline{x}_1) = P(x_1)$. Таким образом, переход от полинома P к $\overline{P}$ сводится к подстановке: $x_1 = \overline{x}_1 + a$ в $P(x_1)$. Если определить $\overline{x}_k = \overline{x}_1{}^k$, $k = 1, \ldots, n$, то для x_k получим следующие выражения:

$$\overline{x}_k = (\overline{x}_1 + a)^k = \sum_{i=0}^{k} C_k^i a^i \overline{x}_{k-i}. \tag{11}$$

Для продолжения доказательства нам понадобятся две леммы.

Лемма 1. *Пусть задана система равенств вида*

$$y_i - y_j = 0; \quad i, j = 1, \ldots, k; \quad i \neq j. \tag{12}$$

Тогда любая линейная форма $l(y) = \sum\limits_{i=1}^{k} c_i y_i$ при условии $\sum\limits_{i=1}^{k} c_i = 0$ может быть записана в виде линейной комбинации левых частей равенств (12).

Доказательство. При $k = 1$ лемма тривиальна. Для $k > 1$ легко ее доказать индукцией от n к $n + 1$, используя простое тождество

$$\sum_{i=1}^{n+1} c_i y_i = \sum_{i=1}^{n-1} c_i y_i + (c_n + c_{n+1}) y_n + c_{n+1}(y_{n+1} - y_n).$$

Лемма 2. *Любое выражение вида $x_{\overline{s}} x_{\overline{q}} - x_{\overline{p}} x_{\overline{r}}$; $\overline{s} + \overline{q} = \overline{p} + \overline{r}$, где $\overline{s}$, $\overline{q}$, $\overline{p}$, $\overline{r}$ – целые неотрицательные числа, при подстановках вида (11) переходят в выражения от переменных $\overline{x}_0$, $\overline{x}_1, \ldots, \overline{x}_n$, представляющие собой линейную комбинацию выражений вида $\overline{x}_s \overline{x}_q - \overline{x}_p \overline{x}_r$, $s + q = p + r$; s, q, p, r – целые неотрицательные числа.*

Доказательство. Используем равенства (11):

$$x_{\overline{s}}x_{\overline{q}} - x_{\overline{p}}x_{\overline{r}} = \left(\sum_{i=0}^{\overline{s}} a^i C_{\overline{s}}^i \overline{x}_{\overline{s}-i}\right)\left(\sum_{j=0}^{\overline{q}} a^j C_{\overline{q}}^j \overline{x}_{\overline{q}-j}\right) -$$

$$-\left(\sum_{i=0}^{\overline{p}} a^i C_{\overline{p}}^i \overline{x}_{\overline{p}-i}\right)\left(\sum_{j=0}^{\overline{r}} a^j C_{\overline{r}}^j \overline{x}_{\overline{r}-j}\right) =$$

$$= \sum_{t=0}^{\overline{s}+\overline{q}} a^t \left(\sum_{i,j:i+j=t} C_{\overline{s}}^i C_{\overline{q}}^j \overline{x}_{\overline{s}-i}\,\overline{x}_{\overline{q}-j} - \sum_{i,j:i+j=t} C_{\overline{p}}^i C_{\overline{r}}^j\,\overline{x}_{\overline{p}-i}\overline{x}_{\overline{r}-j}\right).$$

Как следует из биномиальных тождеств и равенства $\overline{s}+\overline{q}=\overline{p}+\overline{r}$,

$$\sum_{i,j:i+j=t} C_{\overline{s}}^i C_{\overline{q}}^j - \sum_{i,j:i+j=t} C_{\overline{p}}^i C_{\overline{r}}^j = C_{\overline{s}+\overline{q}}^t - C_{\overline{p}+\overline{r}}^t = 0.$$

Используя лемму 1 многократно при различных t, получаем, что выражение $x_{\overline{s}}x_{\overline{q}}-x_{\overline{p}}x_{\overline{r}}$ представимо в виде линейной комбинации выражений вида $\overline{x}_s\overline{x}_q - \overline{x}_p\overline{x}_r$, $s+q=p+r$, s, q, p, r – целые неотрицательные числа. Лемма доказана.

Продолжим доказательство теоремы.

Полином $P(x_1)$ обладает ω-свойством. Это значит, что найдется вектор множителей Лагранжа λ^* такой, что функция Лагранжа $L(x,\lambda)$ квадратичной задачи (9)–(10) при $\lambda=\lambda^*$ является неотрицательно определенной по x и $\psi(\lambda^*) = f^*$, где $f^* = \min\limits_{x_1} P(x_1)$. Подставим в $L(x,\lambda^*)$ вместо компонент вектора x выражения из (11). (Формулы (11) можно рассматривать как переход от одной системы координат к другой.) Получим квадратичную относительно $\overline{x}$ функцию $\overline{L}(\overline{x},\lambda^*)$, и так как положительно (неотрицательно) определенная квадратичная функция при неособых линейных преобразованиях координат переходит в положительно (неотрицательно) определенную функцию, то $\overline{L}(\overline{x},\lambda^*)$ будет неотрицательно определенной. Заметим, что области значений $L(x,\lambda^*)$ и $\overline{L}(\overline{x},\lambda^*)$ совпадают, откуда $\overline{\psi}^* = \min\limits_{\overline{x}} \overline{L}(\overline{x},\lambda^*) = f^*$. С другой стороны, используя лемму 2, приходим к выводу, что выражение $\overline{L}(\overline{x},\lambda^*)$ может быть записано в форме функции Лагранжа квадратичной задачи, соответствующей нахождению $\min \overline{P}(\overline{x_1})$:

$$\min \overline{K}(\overline{x}) \tag{13}$$

при ограничениях

$$\overline{x}_p\overline{x}_q - \overline{x}_r\overline{x}_s = 0, \quad (p, q;\, r, s) \in Q_{2n}, \tag{14}$$

где $\overline{K}(\overline{x})$ – результат подстановки в $K_{2n}(x)$ (9) вместо компонент вектора x выражений (11); при этом соответствующие множители Лагранжа линейно выражаются через компоненты λ^*. Таким образом, для задачи (13), (14) существует вектор множителей Лагранжа $\overline{\lambda}^*$, при котором соответствующая квадратичная функция от $\overline{x}$ неотрицательно определена и ее минимум совпадает с f^* значением минимума полиномов $P(x_1)$ и $\overline{P}(\overline{x}_1)$. Значит, полином $\overline{P}$ также обладает ω-свойством. Теорема 2 о сдвиге доказана.

Перейдем к доказательству основного результата.

Теорема 3. *Любой полином $P_{2n}(x_1)$ вида (8) четной степени обладает ω-свойством.*

Доказательство. Для $n = 2$ ω-свойство доказано (см. пример 2). Проведем доказательство теоремы индукцией по n. Пусть ω-свойство справедливо для полиномов вида (8) степени $2n$. Докажем его справедливость для полиномов степени $2(n+1)$. По теореме 2 (о сдвиге) ω-свойство сохраняется при сдвиге аргумента, поэтому можно предположить, не уменьшая общности, что глобальный минимум находится в 0 и значение полинома в 0 равно 0. Отсюда будет следовать, что у рассматриваемого полинома $P^0_{2n+2}(x_1)$ коэффициенты при младших степенях $a_0 = a_1 = 0$ и $a_2 \geq 0$, т. е. он представим в виде $P^0_{2n+2}(x_1) = x_1^2 \cdot P_{2n}(x_1)$, причем полином $P_{2n}(x_1)$ принимает неотрицательные значения на всей оси. По предположению индукции для $P_{2n}(x_1)$ выполняется ω-свойство, т. е. для соответствующей квадратичной задачи найдется вектор множителей Лагранжа λ^*, при котором функция Лагранжа $L_{2n}(x, \lambda^*)$ будет принимать неотрицательные значения при любом x, т. е. может быть записана в виде $L^*_{2n}(x) = \sum\limits_{i=1}^{n} l_i(x) + r^2$, где $l_i(x)$ – линейные функции от $(x_1, \ldots, x_n)$, r^2 – неотрицательная константа.

Рассмотрим выражение $x_1^2 L^*_{2n}(x)$. Ему соответствует квадратичная форма

$$L^*_{2n+2}(\overline{x}) = \sum_{i=1}^{n} \overline{l}_i^{\,2}(\overline{x}) + r^2 x_1^2,$$

где $\overline{l}_i(\overline{x})$ – линейная функция от переменных $x_1, \ldots, x_n, x_{n+1}$, следующим образом получающаяся из l_i: если $l_i(x) = \sum\limits_{j=1}^{n} l_{ij}x_j + l_{i0}$, то $\overline{l}_i(\overline{x}) = \sum\limits_{j=0}^{n} l_{ij}x_{j+1}$. Легко видеть, что $L^*_{2n+2}(\overline{x})$ получается из функции Лагранжа квадратичной задачи, соответствующей минимизации полинома $P^0_{2n+2}(x_1)$ при тех же значениях множителей Лагранжа, которые образуют вектор λ^*, однако эти значения относятся к «преобразованным» ограничениям: ограничениям вида $x_px_q - x_rx_s = 0$, относящимся к задаче минимизации $P_{2n}(x_1)$, будут соответствовать ограничения $x_{p+1}x_{q+1} - x_{r+1}x_{s+1} = 0$ в задаче минимизации $P^0_{2n+2}(x_1)$. Минимальное значение $L^*_{2n+2}(\overline{x})$ достигается при $\overline{x} = 0$ и равно 0. Таким образом, при определенных значениях множителей Лагранжа квадратичной задачи, соответствующей минимизации полинома $P^0_{2n+2}(x_1)$, получается точная оценка, т. е. полином $P^0_{2n+2}(x_1)$ обладает ω-свойством. По теореме о сдвиге этим же свойством будет обладать произвольный полином $P_{2n+2}(x_1)$ степени $2n + 2$ с коэффициентом, равным 1, при старшем члене. Теорема 3 доказана.

Сделаем несколько замечаний.

1. Сведение задачи минимизации $P_{2n}(x_1)$ к квадратичной задаче неоднозначно, это связано с неоднозначностью представления x_1^k, например, $x_1^5 = x_1x_4 = x_2x_3 = x_5$ и т. п. В зависимости от принятого представления изменяются оптимальные множители Лагранжа, однако ω-свойство от конкретного представления не зависит.

2. Минимальное число ограничений, которые нужно учитывать при формулировке эквивалентной квадратичной задачи при минимизации $P_{2n}(x_1)$, равно $(n-1)$, так как нужно определить переменные $x_2, \ldots, x_n$. Остальные ограничения избыточны, т. е. их добавление не сужает области допустимых решений. Роль избыточных ограничений состоит в расширении числа двойственных переменных функции Лагранжа, что приводит, вообще говоря, к более точным оценкам. Добавление ограничения, которое является линейной комбинацией имеющихся, не отражается на точности двойственных оценок, так как «вклад» этого ограничения в функцию Лагранжа эквивалентен определенному изменению множителей Лагранжа при имеющихся ограничениях.

3. Избыточность числа ограничений приводит, как правило, к неоднозначности вектора оптимальных множителей Лагранжа λ^*. По геометрическому смыслу оптимальных множителей Лагранжа – это коэффициенты разложения по градиентам ограничений ортогонального к многообразию, вырезаемого ограничениями, антиградиента целевой

функции в оптимальном точке. При неоднозначности разложения совокупность допустимых векторов коэффициентов образует линейное многообразие. Пересечение этого многообразия с $\overline{\Omega}(P)$ и дает множество оптимальных векторов Лагранжа.

Весьма правдоподобным кажется предположение, что при минимизации полиномиальной функции многих переменных $P(x_1, \ldots, x_n)$ путем построения эквивалентной квадратичной задачи при использовании фиксированного множества избыточных ограничений получается точная двойственная оценка значения глобального минимума (аналог теоремы 3). Если это предположение удастся доказать, то в арсенале методов нахождения глобального минимума полиномиальных функций от нескольких переменных появится метод, основанный на двойственных оценках. Этот метод можно в различных формах комбинировать с методом ветвей и границ и методом релаксации ограничений. Еще больший интерес представляет обобщение намеченной в данной статье теории на общий класс полиномиальных задач с ограничениями вида (1), (2). Можно ли в общем случае дать способ генерации избыточной системы ограничений, чтобы гарантировать, что двойственная оценка окажется точной? В настоящее время идет экспериментальное исследование эффективности предложенной методики двойственных оценок на ряде классов экстремальных задач теории графов и невыпуклых квадратичных задач с линейными ограничениями [3].

Литература

1. Шор Н. З., Давыдов А. С. О методе получения оценок в квадратичных экстремальных задачах с булевыми переменными // Кибернетика. – 1985. – № 2. – С. 48–50.

2. Шор Н. З. Методы минимизации недифференцируемых функций и их приложения. – К.: Наук. думка, 1979. – 199 с.

3. Стеценко С. И. Об алгоритмах решения некоторых задач выбора // Исследование методов решения экстремальных задач. – К.: ИК АН УССР, 1986. – С. 12–16.

Об одном классе оценок глобального минимума полиномиальных функций

Н. З. Шор
Кибернетика. – 1987. – № 6. – С. 9–11.

Пусть E^n – n-мерное евклидово пространство, на котором определена полиномиальная вещественная функция $P(x) = P(x_1, \ldots, x_n)$. Рассмотрим задачу нахождения $f^* = \inf\limits_{x \in E^n} P(x_1, \ldots, x_n)$. Если $f^* > -\infty$, т. е. $P(x)$ ограничена снизу, то назовем ее ОС-функцией. Так как полином нечетной степени от одной переменной может принимать сколь угодно большие по абсолютной величине положительные и отрицательные значения, то для того, чтобы $P(x)$ была ОС-функцией, необходимо, чтобы для любого $i \leq n$ старшие степени s_i переменных x_i были четными, а значения коэффициентов при них при любых значениях остальных переменных принимали неотрицательные значения. ОС-функциями должны также являться функции $P\,(x;\, x_i = c)$ от $(n-1)$-переменной, которые получаются из $P(x)$, если придать произвольной переменной x_i $(i \leq n)$ определенное значение $x_i = c$. Проверка того, что данная полиномиальная функция $P(x)$ является ОС-функцией, – далеко не тривиальная задача.

Пусть $s_i = 2l_i$; $i = 1, \ldots, n$. Рассмотрим целочисленные векторы $\alpha = \{\alpha_1, \ldots, \alpha_n\}$ с неотрицательными элементами и одночленные выражения вида

$$R[\alpha] = x^{\alpha_1} \ldots x_n^{\alpha_n}, \quad \text{где} \quad 0 \leq \alpha_i \leq l_i; \quad i = 1, \ldots, n. \tag{1}$$

Имеем систему тождественных соотношений

$$R[\alpha_1^{(1)}, \ldots, \alpha_n^{(1)}]R[\alpha_1^{(2)}, \ldots, \alpha_n^{(2)}] - \\ - R[\alpha_1^{(3)}, \ldots, \alpha_n^{(3)}]R[\alpha_1^{(4)}, \ldots, \alpha_n^{(4)}] = 0 \tag{2}$$

(или $R[\alpha^{(1)}]R[\alpha^{(2)}] - R[\alpha^{(3)}]R[\alpha^{(4)}] = 0$) при $\alpha_i^{(1)} + \alpha_i^{(2)} = \alpha_i^{(3)} + \alpha_i^{(4)} \leq s_i$, $i = 1, \ldots, n$. Любой полином $P(x) = P(x_1, \ldots, x_n)$ со старшими степенями при x_i, равными $s_i = 2l_i$, $i = 1, \ldots, n$, может быть неоднозначно, вообще говоря, записан в виде квадратичной функции от переменных $R[\alpha^{(\nu)}]$ (см. (1)) следующего вида:

$$P(x) = K(R,\lambda) = \sum_{i,j} c_{ij} R[\alpha^{(i)}]R[\alpha^{(j)}] +$$
$$+ \sum_{(k,l;\, m,n)} \lambda_{k,l;m,n} \left(R[\alpha^{(k)}]R[\alpha^{(l)}] - R[\alpha^{(m)}]R[\alpha^{(n)}] \right),$$

где $R[\alpha^{(i)}]R[\alpha^{(j)}]$ – некоторое представление одночлена полинома $P(x)$ в виде произведения одночленов вида (1); c_{ij} – соответствующие коэффициенты; $\lambda_{k,l;m,n}$ – произвольные множители при левых частях тождественных соотношений (2). С другой стороны, $K(R,\lambda)$ можно рассматривать как функцию Лагранжа квадратичной экстремальной задачи: минимизировать

$$K(R) = \sum_{i,j} c_{ij} R[\alpha^{(i)}]R[\alpha^{(j)}]$$

при ограничениях (2), при этом $\lambda_{k,l;m,n}$ могут интерпретироваться как множители Лагранжа. Таким образом, задача безусловной минимизации полинома сведена к задаче минимизации квадратичной функции $K(R)$ при квадратичных ограничениях (2).

Пусть множество тех $\lambda = \{\lambda_{k,l;m,n}\}$, для которых квадратичная функция $K(R,\lambda)$ является неотрицательно определенной по R, не пусто. Обозначим его $\overline{\Lambda}^+$. Внутреннюю часть этого множества (совокупность тех λ, при которых $K(R,\lambda)$ положительно определена) обозначим Λ^+. Заметим, что при $\lambda \in \Lambda^+$ $\inf_R K(R,\lambda)$ достигается в некоторой точке $R^*(\lambda)$. Легко видеть, что $\Lambda^+(\overline{\Lambda}^+)$ – выпуклые множества. Пусть

$$\psi(\lambda) = \inf_R K(R,\lambda). \tag{3}$$

Так как $K(R,\lambda)$ линейна по λ, то $\psi(\lambda)$ в области своего существования (в частности, на Λ^+) – вогнутая функция. Так как при любом λ и допустимом векторе R $K(R,\lambda) = K(R)$, то

$$\psi(\lambda) = \inf_R K(R,\lambda) \le f^*. \tag{4}$$

Если $P(x)$ не является ОС-функцией, то Λ^+ пусто. Пусть $\overline{f} = \sup_{\lambda\in\overline{\Lambda}^+} \psi(\lambda)$. Из (4) следует, что $\overline{f} \le f^*$.

Если найдется $\lambda^* \in \overline{\Lambda}^+$ такой, что $\psi(\lambda^*) = f = f^* > -\infty$, то будем говорить, что $P(x)$ и соответствующая квадратичная задача вида (3), (2) обладает ω-свойством.

Если полином $P(x)$ обладает ω-свойством, то для получения оценки снизу для f^* с произвольной точностью нужно решать задачу выпуклого программирования: найти $\sup_{\lambda\in\overline{\Lambda}^+} \psi(\lambda) = -\inf_{\lambda\in\overline{\Lambda}^+}(-\psi(\lambda))$.

В [1, 2] показано, что для решения задачи такого типа можно применять метод эллипсоидов. Практически хорошо работают при решении подобных задач другие методы недифференцируемой оптимизации, например, определенные модификации r-алгоритма.

Таким образом, для полиномиальных функций, обладающих ω-свойством, получаем непереборный алгоритм оценки глобального экстремума. Отметим также, что если $\sup_{\lambda\in\overline{\Lambda}^+}\psi(\lambda)$ достигается на $\lambda^*\in\Lambda^+$, то градиент функции $\psi(\lambda)$ в этой точке, равный вектору невязок в ограничения (2), при $R = R(\lambda^*)$ равен 0, т. е. все ограничения (2) выполняются. Но это значит, что $\overline{K}(R(\lambda^*),\lambda^*) = K(R(\lambda^*)) = f^*$, т. е. $\overline{f} = f^*$ и ω-свойство выполняется. Таким образом, если ω-свойство не выполняется, то $\sup_{\lambda\in\overline{\Lambda}^+}\psi(\lambda)$ должен достигаться на границе области неотрицательной определенности.

Оказалось, что вопрос о том, обладает ли данная полиномиальная ОС-функция ω-свойством, тесно связан с возможностью ее представления в виде суммы квадратов полиномов. Докажем следующую теорему.

Теорема 1. *Для того, чтобы полином $P(x) = P(x_1,\ldots,x_n)$, принимающий минимальное значение в точке 0 и равный при этом 0, обладал ω-свойством, необходимо и достаточно, чтобы он был представим в виде суммы квадратов полиномов.*

Доказательство. 1. Необходимость. Пусть полином обладает ω-свойством. Тогда найдется $\lambda^*\in\overline{\Lambda}^+$, при котором $\overline{K}\,(R,\lambda^*)$ неотрицательно определена как функция от R и $\overline{K}\,(0,\lambda^*) = 0$. Но любую неотрицательно определенную функцию можно представить в виде суммы квадратов линейных функций. Подставив в функцию $\overline{K}\,(R,\lambda^*)$, представленную в форме суммы квадратов, вместо переменных $R[\alpha^{(\nu)}]$ их выражения в форме одночленов, получим искомое представление $P(x)$ в форме суммы квадратов полиномов.

2. Достаточность. Пусть $P(x)$ представим в виде суммы квадратов полиномов $P_1(x),\ldots,P_m(x)$:

$$P(x) = \sum_{i=1}^{m} P_i^2(x).$$

Полиномы $P_k(x)$, $k = 1,\ldots,m$, не должны содержать свободных членов, иначе $P(0) > 0$, а это противоречит $P(0) = 0$. Подставив вместо каждого одночлена в полиномах $P_k(x)$ соответствующее $c_a R[\alpha]$, получим однородную неотрицательно определенную форму относительно R, минимум которой при $R = 0$ равен 0. Сравнив коэффициенты при

$R[\alpha^{(i)}]R[\alpha^{(j)}]$ в представлении $P(x)$ в форме суммы квадратов $\overline{K}(R)$ с коэффициентом при $R[\alpha^{(i)}+\alpha^{(j)}]$ в представлении $P(x) = K(R) = \overline{K}(R,0)$, найдем $\overline{\lambda}$, при котором $\overline{K}(R) = K(R,\overline{\lambda})$. Таким образом, найдется $\overline{\lambda}$ такой, что $\psi(\overline{\lambda}) = \min\limits_{R} K(R,\overline{\lambda}) = K(0,\overline{\lambda}) = 0$. Значит, ω-свойство выполняется.

Теорема доказана.

Пусть $\min P(x_1,\ldots,x_n) = f^*$ и достигается при $x = x^*$.

Рассмотрим полином $P_0(x) = P(x-x^*) - f^*$. $P_0(x)$ принимает свое минимальное значение в точке 0 и это значение равно 0.

Графики полиномов $P_0(x)$ и $P(x)$ совпадают при соответствующем сдвиге в пространстве x и по f. Легко показать, что сдвигу в пространстве координат x соответствует линейное невырожденное преобразование в пространстве координат $R[\alpha]$. Так как при таких преобразованиях сохраняется неотрицательная (положительная) определенность квадратичных функций, то и сохраняется ω-свойство, а также возможность представления функций в форме суммы квадратов. Таким образом, из теоремы 1 вытекает более общая.

Теорема 2. *Для того, чтобы ограниченный снизу полином $P(x)$ обладал ω-свойством, необходимо и достаточно, чтобы полином $P(x) - f^*$ был представим в виде суммы квадратов полиномов.*

К сожалению, не всякая полиномиальная функция, принимающая неотрицательные значения, представима в виде суммы квадратов полиномов. Этот вопрос рассматривался еще в 1888 году молодым в то время Д. Гильбертом [3]. Он изучал представимость в форме суммы квадратов однородных полиномов (форм) четной степени m с числом переменных n (m – максимальная сумма степеней переменных, входящих в одночлен данного полинома; для однородных полиномов сумма степеней переменных, входящих в произвольный одночлен, равна m).

Гильберт показал, что при n=3, $m \geq 6$ и $n \geq 4$, $m \geq 4$ найдутся неотрицательные формы, не представимые в виде суммы квадратов форм.

Лишь для некоторых общих классов форм, а именно:

1) $m = 2$, n – произвольное число (квадратичные формы);

2) $n = 2$; m – четное произвольное, т. е. произвольные формы от двух переменных (им соответствуют произвольные полиномы от одной переменной);

3) $m = 4$, $n = 3$, т. е. произвольные биквадратные формы с тремя переменными (им соответствуют произвольные полиномы 4-й степени с

двумя переменными) – вопрос о представимости в виде суммы квадратов неотрицательных форм решается положительно.

Отсюда возникла следующая задача, известная как 17-я проблема Гильберта [4]. Рассматривается поле рациональных функций от вещественных переменных. Если рациональная функция в области своего определения неотрицательна, то всегда ли она представима в виде суммы квадратов рациональных функций? Окончательный положительный ответ на этот вопрос был дан в работе Э. Артина в 1927 г. [5].

Для неотрицательных полиномиальных форм $P(x)$ это означает, что всегда можно найти такую представимую в виде суммы квадратов неотрицательную форму $P_0(x)$, что $P_0(x)\,P(x)$ представим в виде суммы квадратов полиномов. Интересным остается вопрос о конструктивном способе выбора $P_0(x)$.

После краткого отступления вернемся к рассмотрению некоторых важных частных случаев экстремальных полиномиальных задач, для которых ω-свойство справедливо.

Системы полиномиальных уравнений. Пусть задана система уравнений

$$P_i(x) \equiv P_i(x_1, \ldots, x_n) = 0; \quad i = 1, \ldots, m. \tag{5}$$

Рассмотрим $P(x) = \sum\limits_{i=1}^{m} P_i^2(x)$.

Если система (5) совместна, то $f^* = \min P(x) = 0$; $P(x) - f^* = P(x)$ представим в виде суммы квадратов, т. е. $P(x)$ обладает ω-свойством. Множество решений системы (5) в случае ее нелинейности может быть пустым, континуальным или дискретным. Лишь в исключительных случаях решение единственно. Поэтому при минимизации $P(x)$ путем сведения к квадратичной задаче оптимальная оценка будет достигаться, как правило, на границе Λ^+ при вырожденной матрице квадратичной формы, что не позволит непосредственно получить хотя бы одно из решений. Для получения конкретного решения нужно использовать дополнительные средства для отделения корней (например, дополнительные неравенства на переменные, использование ε-возмущения коэффициентов функции $P(x)$ для получения приближения к решению и т. п.).

Интересным является вопрос – всегда ли в случае несовместности системы (5) при минимизации

$$P(x) = \sum_{i=1}^{m} P_i^2(x) \qquad \sup_{\lambda \in \Lambda^+} \psi(\lambda) > 0?$$

Задача о линейной дополнительности. Данная задача [6] ставится следующим образом: определить вектор $x = \{x_1, \ldots, x_n\} \in E^n$, удовлетворяющий системе неравенств $x \geq 0$; $Ax + b \geq 0$ (A – матрица $n \times n$, b – n-мерный вектор), и такой, что $(x,\, Ax + b) = 0$. Пусть a_i, $i = 1, \ldots, n$, – строки матрицы A. Введем дополнительные переменные v_i, $i = 1, \ldots, n$, и рассмотрим задачу: найти

$$f^* = \min_{x \geq 0; v \geq 0} \Big\{ \sum_{i=1}^{n} \left[(a_i,\, x) + b_i - v_i\right]^2 + c \sum_{i=1}^{n} x_i v_i \Big\}, \tag{6}$$

где $v = \{v_1, \ldots, v_n\}$, $c > 0$. Легко видеть, что $f^* \geq 0$ и равно 0 тогда и только тогда, когда соответствующая задача о линейной дополнительности имеет решение. Задача (6) путем подстановок $x_i = y_i^2$; $v_i = \omega_i^2$, $i = 1, \ldots, n$, сводится к задаче безусловной минимизации функции $P(y,\, \omega)$, являющейся полиномом 4-й степени от компонент y, ω и представимой в виде суммы квадратов. Таким образом, если задача линейной дополнительности имеет решение, то $P(y,\, \omega)$ обладает ω-свойством. Задача о линейной дополнительности является частным случаем задачи минимизации произвольной квадратичной функции при линейных ограничениях. Если эти ограничения имеют форму $x \geq 0$, то можно добавить квадратичные ограничения $x_i \cdot x_j \geq 0$, $i, j = 1, \ldots, n$, являющиеся следствием ограничений $x \geq 0$, применить к расширенной задаче метод множителей Лагранжа и построить двойственные оценки, следуя описанной в статье методике. Хотя эти оценки не всегда являются точными, однако они обычно оказываются более точными, чем оценки, полученные путем линеаризации вогнутых составляющих квадратичной функции [7].

Частный случай теоремы 2 для полиномов от одной переменной приводится в [8].

Литература

1. Шор Н. З., Давыдов А. С. О методе получения оценок в квадратичных зкстремальных задачах с булевыми переменными // Кибернетика. – 1985. – № 2. – С. 48–50.

2. Шор Н. З. Квадратичные оптимизационные задачи // Техн. кибернетика. – 1987. – № 1. – С. 128–139.

3. Hilbert D. Űber die darstellung definiter Formen als Summen von Formen quadraten // Math. Ann. – 1988. – 22. – P. 342–350.

4. Проблемы Гильберта / Под ред. П. С. Александрова. – М.: Наука, 1969. – 238 с.

5. Artin E. Über die Zerlegung definiter Funktionen in Quadrate. – 1927. – **5**. – S. 100–115.

6. Stone R. Linear complementarity problems with an invariant number of Solutions // Math. Programming – 1968. – **34**. – № 3. – P. 265–291.

7. Pardalos P. M., Glick J. H., Rosen J. B. Global minimisation of indefinite quadratic problems. – S. I. 1985. – 14 p. – (Reprint / Pennsyl. State University. – NCS–85–31).

8. Шор Н. З. Об одном подходе к получению глобальных экстремумов в полиномиальных задачах математического программирования // Кибернетика. – 1987. – № 5. – С. 102–106.

Использование точных штрафов при построении описанных эллипсоидов минимального объема

Н. З. Шор, С. И. Стеценко
Кибернетика. – 1989. – № 2. – С. 117–119.

Задача аппроксимации областей достижимости или произвольных совокупностей точек эллипсоидами широко используется в оптимальном управлении, игровых задачах и других приложениях [1]. Исследуем задачу нахождения эллипсоида минимального объема, включающего заданный набор точек $x_1, \ldots, x_m$ пространства R^n. Эллипсоид E в R^n можно задать неравенством

$$(K(x-a), x-a) \leq 1, \tag{1}$$

где K – симметричная, положительно определенная матрица, $K \in R^{n\times n}$, $a = (a^1, \ldots, a^n)$ – центр эллипсоида, $x = (x^1, \ldots, x^n)$ – точка эллипсоида. Объем эллипсоида, заданного неравенством (1), равен $V = \omega_n \cdot (\det K)^{-1/2}$, где ω_n – объем n-мерного единичного шара, константа при фиксированном n. Условие принадлежности набора точек $x_1, \ldots, x_m$ эллипсоиду E вида (1) можно записать как систему неравенств

$$(K(x_j-a),\, x_j-a) \leq 1, \quad j = 1, \ldots, m. \tag{2}$$

Предположим, что ранг системы векторов $\{x_2 - x_1, \ldots, x_m - x_1\}$ равен n, в противном случае матрица оптимального описанного эллипсоида будет вырожденной, и для решения исходной задачи нужно будет перейти к пространству R^t, $t < n$, где $t = \operatorname{rang}\{x_2 - x_1, \ldots, x_m - x_1\}$. Задачу построения эллипсоида минимального объема, включающего точки $x_1, \ldots, x_m \in R^n$, можно записать в виде

$$\max_{K,a}(\ln \det K), \tag{3}$$

$$(K(x_j-a),\, x_j-a) \leq 1, \quad j = 1, \ldots, m, \tag{4}$$

$$K \succ 0, \tag{5}$$

где $K = (k_{ij})_{i,j=1}^{n}$ – симметричная матрица $n \times n$, $x_j \in R^n$, $j = 1, \ldots, m$, $a \in R^n$, $\operatorname{rang}\{x_2 - x_1, \ldots, x_m - x_1\} = n$, $K \succ 0$ означает принадлежность K классу положительно определенных матриц.

В [2] эта задача (без логарифма в целевой функции) была сведена к максимизации строго квазивогнутой функции на выпуклом множестве и для ее решения предлагалось использовать метод эллипсоидов [3]. Учитывая, что скорость сходимости метода эллипсоидов резко падает при увеличении размерности задачи p (знаменатель геометрической прогрессии, определяющий скорость сходимости по функционалу, $q \approx 1 - 1/\,2p^2$), целесообразно исследовать другие процедуры нахождения оптимальных описанных эллипсоидов.

Обозначим $y = (k_{11}, \ldots, k_{1n}, k_{22}, \ldots, k_{2n}, \ldots, k_{nn}, a^1, \ldots, a^n)$. Размерность вектора y равна $p = n(n+3)/\,2$. Обозначим $f_0(y) = -\ln\det K$, $f_j(y) = (K(x_j - a),\, x_j - a) - 1$, $j = 1, \ldots, m$, $M^+ = \{y : K \succ 0\}$. Задачу (3)–(5) можно переписать в виде

$$V(0) = \min_y \left\{ f_0(y) : f_j(y) \le 0, \quad j = 1, \ldots, m; \quad y \in M^+ \right\}. \tag{6}$$

Используем для учета ограничения (4) негладкую штрафную функцию. Положим

$$F(y) = \max\{0, f_1(y), \ldots, f_m(y)\}, \tag{7}$$

$$\Phi_N(y) = f_0(y) + N \cdot F(y). \tag{8}$$

Для оценки параметра N воспользуемся теоремой 2.14 из [4]. Пусть $f_0, f_1, \ldots, f_m$ – непрерывные функции, M – некоторое множество. Положим

$$V(z) = \inf_y \left\{ f_0(y) : f_j(y) \le z^j, \quad j = 1, \ldots, m; \quad y \in M \right\}. \tag{9}$$

Теорема 2.14. [4] *Пусть* $\inf\limits_{\lambda > 0} \dfrac{V(\lambda \cdot \overline{1}) - V(0)}{\lambda} = -L > -\infty$ *и* $N > L$. *Тогда точки минимума задачи* $P(0)$ *и задачи* $\inf\limits_y \{\Phi_N(y) : y \in M\}$ *совпадают.*

Здесь $P(0)$ – задача минимизации, стоящая в правой части соотношения (9) при $z = (0, 0, \ldots, 0)$, $\Phi_N(y)$ строится по функциям $f_0, f_1, \ldots, f_m$ согласно (7), (8); $\overline{1} = (1, 1, \ldots, 1)$.

Условие $(K(x_j - a),\, x_j - a) - 1 \le \lambda$ эквивалентно условию

$$\left(\frac{K}{1+\lambda} - (x_j - a),\, x_j - a\right) \le 1 \quad \text{или} \quad (K^{(1)}(\lambda)\,(x_j - a),\, x_j - a) \le 1,$$

что соответствует подстановке $k_{ij} \to k_{ij}^{(1)}/(1+\lambda)$ в первоначальной задаче. Учитывая, что $\det K$ является однородной функцией степени n своих элементов, получаем, что при $\lambda > -1$ $V(\lambda \cdot \bar{1}) - V(0) = -\ln(1+\lambda)^n$, откуда

$$\inf_{\lambda>0} \frac{V(\lambda \cdot \bar{1}) - V(0)}{\lambda} = \inf_{\lambda>0} (-n)\frac{\ln(1+\lambda)}{\lambda} = -n > -\infty.$$

Таким образом, справедлива следующая лемма.

Лемма. *При $N > n$ точки минимума задачи* (6) *и задачи* $\inf_y\{\Phi_N(y) : y \in M^+\}$, *где* $\Phi_N(y)$ *определена согласно* (7), (8), *совпадают.*

В [2] показано, что если начальная точка 0 лежит внутри эллипсоида E, т. е. $(Ka,a) < 1$, неравенство (1) можно переписать в виде $(Bx,x) + 2(b,x) \le 1$, где

$$B = K/(1-(Ka,a)),\ b = -aK/(1-(Ka,a)). \tag{10}$$

В этом случае объем эллипсоида E равен $V = \omega_n \cdot (\varphi(B,b))^{-1/2}$, где $\varphi(B,b) = (\det B)^{n+1}/(-\det C)^n$, $C = \begin{pmatrix} B & b^T \\ b & -1 \end{pmatrix}$, а неравенства (2) примут вид

$$h_j(B,b) = (Bx_j,x_j) + 2(b,x_j) - 1 \le 0, \quad j = 1,\dots,m. \tag{11}$$

В нашей задаче для выполнения условия $0 \in E$ можно, например, перенести начало координат 0 в точку, являющуюся средним арифметическим точек $x_1,\dots,x_m$, и пересчитать координаты точек в смещенном базисе. Поэтому, не уменьшая общности, задачу (3)–(5) можно переписать в эквивалентном виде

$$\max_{B,b}\ \ln\left[(\det B)^{n+1} \Big/ \left(-\det\begin{pmatrix} B & b^T \\ b & -1 \end{pmatrix}\right)^n\right], \tag{12}$$

$$(Bx_j,x_j) + 2(b,x_j) \le 1, \quad j = 1,\dots,m, \tag{13}$$

$$B \succ 0, \tag{14}$$

или, с учетом принятых обозначений,

$$V(0) = \Big\{ \min_{B,b} [-\ln \varphi(B,b)] : \quad h_j(B,b) \le 0, \quad j = 1, \dots, m; \quad B \succ 0 \Big\}.$$

Ограничения (13), (14) образуют выпуклое множество пространства R^n. Как показано в [2], функция $\varphi(B,b)$ является строго квазивогнутой при $B \succ 0$. Аналогичным свойством, как легко показать, будет обладать функция $\ln \varphi(B,b)$. Следовательно, задача (12)–(14) имеет единственное решение. Учитывая, что

$$\begin{aligned} f_j(y) &= (1-(Ka,a)) \left[\frac{(Kx_j, x_j)}{1-(Ka,a)} - \frac{2(Ka,x_j)}{1-(Ka,a)} - 1 \right] \le \\ &\le (Bx_j, x_j) + 2(b, x_j) - 1 \end{aligned}$$

при $0 < (Ka,a) < 1$, вместо штрафной функции $\Phi_N(y)$ можно записать штрафную функцию

$$\overline{\psi}_N(B,b) = -\ln \varphi(B,b) + N \cdot \max\Big\{0,\ (1-(Ka,\, a))h_j(B,\, b);\ j = 1, \dots, m\Big\},$$

которая соответствует задаче (3)–(5) при замене переменных (10). Штрафная добавка $\overline{\psi}_N(B,b)$ при замене компонент $(1-(Ka,a))h_j(B,b)$ на $h_j\,(B,b)$ при условии $0 < (Ka,a) < 1$ может лишь увеличиться, поэтому в штрафной функции

$$\psi_N(B,b) = -\ln \varphi(B,b) + N \cdot \max\Big\{0,\ h_1(B,b), \dots,\ h_m(B,b)\Big\}, \tag{15}$$

которая соответствует квазивогнутой задаче (12)–(14), можно воспользоваться тем же штрафным множителем, что и в штрафной функции для задачи (3)–(5). В силу леммы при $N > n$ единственное решение, совпадающее с $V(0)$, будет иметь задача

$$\min_{B,b} \Big\{ \psi_N(B,b) : B \succ 0 \Big\}. \tag{16}$$

Функция $\psi_N(B,b)$ негладкая, для ее минимизации можно использовать r-алгоритм [3], при этом субградиенты $\psi_N(B,b)$ вычисляются по формулам

$$\widehat{g}_{ij} = \partial \psi_N(B,b)/\partial b_{ij} = 2(-(n+1)b_{ij}^{-1} + nc_{ij}^{-1} + Nx_r^i x_r^j),\ 1 \le i < j \le n,$$

$$\widehat{g}_{ii} = \partial \psi_N(B,b)/\partial b_{ii} = -(n+1)b_{ii}^{-1} + nc_{ii}^{-1} + \overline{N}(x_r^i)^2,\ i = 1, \dots, n,$$

$$\widehat{g}_i = \partial\psi_N(B,b)/\partial b^i = 2\left(nc^{-1}_{i\,n+1} + \overline{N}x^i_r\right),\ i = 1,\ldots,n,$$

где $B^{-1} = (b^{-1}_{ij})^n_{i,j=1}$ – матрица, обратная к B, $C^{-1} = (c^{-1}_{ij})^{n+1}_{i,j=1}$ – матрица, обратная к $C = \begin{pmatrix} B & b^T \\ b & -1 \end{pmatrix}$,

$$\overline{N} = \begin{cases} N, & \text{если } h_r(B,b) = \max\limits_{1\le j\le m} h_j(B,b) > 0, \\ 0 & \text{в противном случае .} \end{cases}$$

При использовании монотонных вариантов r-алгоритма (16) можно решать как задачу безусловной минимизации. Выполнение ограничения $B \succ 0$ в процессе поиска решения гарантируется тем, что при движении матрицы B изнутри допустимой области к ее границе $\varphi(B,b) \to 0$, $\ln\varphi(B,b) \to -\infty$, а значит, $\psi_N(B,b) \to \infty$. Таким образом, наличие логарифма в целевой функции представляет естественный «барьер», не позволяющий при минимизации $\psi_N(B,b)$ выйти за пределы области $B \succ 0$.

Выход за границу области в процессе применения r-алгоритма возможен из-за большого значения шагового множителя. Для того, чтобы этого не произошло, предлагается использовать в r-алгоритме следующую регулировку шагового множителя. Предположим, $B_k \succ 0$. На $(k+1)$-м шаге вычисляем $B_{k+1} = B_k + h_k\widehat{g}_k$. Проверяем, является ли матрица B_{k+1} положительно определенной, например, путем ее треугольной факторизации. Если B_{k+1} не является положительно определенной, то шаг h_k делится пополам $1,2,\ldots,s$ раз до тех пор, пока матрица $B'_{k+1} = B_k + \dfrac{h_k}{2^s}\widehat{g}_k$ не будет положительно определена.

Предлагаемый метод построения оптимального описанного эллипсоида, включающего заданный набор точек, был использован для решения более сложной задачи – построения эллипсоида минимального объема вокруг пересечения конечного числа заданных эллипсоидов. Приближенный итеративный алгоритм решения такой задачи последовательно генерирует точки из пересечения эллипсоидов и строит по ним эллипсоид, приближающийся к оптимальному. При проведении численных экспериментов многократно решались задачи построения оптимальных эллипсоидов, включающих заданный набор точек. Описанный выше метод оказался достаточно эффективным и надежным на практике.

Литература

1. Черноусько Ф. Л., Овсеевич А. И., Решетняк Ю. Н. и др. Алгоритмы гарантированного эллипсоидального оценивания и фильтрации для динамических систем. – М. – 1987. – 43 с. – (Препр. / Ин-т проблем механики АН СССР; № 293).

2. Калинин В. Н., Шикин Е. В. О построении эллипсоидов экстремального объема // Изв. АН СССР. Техн. кибернетика. – 1987. – № 4. – С. 60–65.

3. Гершович В. И., Шор Н. З. Метод эллипсоидов, его обобщения и приложения // Кибернетика. – 1982. – № 5. – С. 61–69.

4. Пшеничный Б. Н. Метод линеаризации. – М.: Наука, 1983. – 136 с.

Алгоритмы построения оптимальных описанных эллипсоидов на основе методов негладкой оптимизации

Н. З. Шор, А. С. Давыдов, С. И. Стеценко

Методы решения экстремальных задач и смежные вопросы. – 1989. – С. 25–30.

Описан приближенный алгоритм построения эллипсоида мини мального объема, включающего пересечение заданных эллипсоидов. Алгоритм основан на методе нахождения оптимального описанного эллипсоида по заданному набору точек, где оптимальность понимается в смысле минимума объема эллипсоида.

Построение эллипсоида минимального объема, включающего заданный набор точек. Исследуем задачу нахождения эллипсоида минимального объема, включающего заданный набор точек $x_1, \ldots, x_m$ пространства R^n. Уравнение эллипсоида E в R^n можно задать неравенством

$$(x-a)A(x-a)^T \leq 1, \tag{1}$$

где A – симметричная, положительно определенная матрица; $A \in R^{n\times n}$; $a \in R^n$ – центр эллипсоида; $x \in R^n$ – точка эллипсоида. Объем эллипсоида, заданного неравенством (1), равен $w_n \cdot (\det A)^{-1/2}$, где w_n – константа при фиксированном n . Задачу построения эллипсоида минимального объема, включающего точки $x_1, \ldots, x_m \in R^n$, можно записать в виде

$$\max_{A,a} \left\{ \ln\det A : (x_j - a)A(x_j - a)^T \leq 1, \quad j = 1, \ldots, m; \quad A \succ 0 \right\}, \tag{2}$$

где $A \succ 0$ означает принадлежность A классу положительно определенных матриц.

В [1] показано, что, если начальная точка 0 лежит внутри эллипсоида E, неравенство (1) можно переписать в виде $xBx^T + 2bx^T \leq 1$, где $B = A/(1 - aAa^T)$; $b = -aA/(1 - aAa^T)$. В этом случае объем эллипсоида $Qw_n \cdot (\varphi(B,b))^{-1/2}$, где $\varphi(B,b) = (\det B)^{n+1}/(-\det C)^n$,

$C = \begin{pmatrix} B^n, & b^T \\ b, & -1 \end{pmatrix}$, а условие принадлежности набора точек $x_1, \ldots, x_m$ эллипсоиду E можно записать как систему неравенств

$$h_j(B, b) = x_j B x_j^T + 2b x_j^T - 1 \leq 0, \quad j = 1, \ldots, m.$$

Для выполнения условия $0 \in E$ можно, например, перенести начало координат 0 в точку$(x_1 + x_2)/2$ и пересчитать координаты всего набора точек в смещенном базисе. Поэтому, не уменьшая общности, задачу (2) можно переписать в эквивалентном виде

$$\varphi^* = \max_{B,b} \{\ln \varphi(B, b) : \quad h_j(B, b) \leq 0, \ \ j = 1, \ldots, m; \ \ B \succ 0\}. \qquad (3)$$

В [1] доказано, что функция $\varphi(B, b)$ строго квазивогнута при $B \succ 0$. Нетрудно показать, что аналогичным свойством будет обладать функция $\ln \varphi(B, b)$. Поэтому (3) как задача максимизации строго квазивогнутой функции на выпуклом множестве будет иметь единственное решение.

Для учета ограничений $h_j(B, b) \leq 0$, $j = 1, \ldots, m$, воспользуемся методом негладких штрафных функций. Положим

$$F(B, b) = \max\{0, h_1(B, b), \ldots, h_m(B, b)\}.$$

Обозначим

$$\Phi_N(B, b) = -\ln \varphi(B, b) + N \cdot F(B, b).$$

В [2] показано, что при $N > n$ единственное решение, совпадающее с φ^*, будет иметь задача

$$\min_{B,b} \{\Phi_N(B, b) : B \succ 0\}. \qquad (4)$$

Функция $\Phi_N(B, b)$ негладкая, для ее минимизации можно использовать r-алгоритм, при этом субградиенты $\Phi_N(B, b)$ вычисляются по формулам:

$$\widehat{g}_{ij} = \partial \Phi_N(B, b)/\partial b_{ij} = 2(-(n+1)b_{ij}^{-1} + n c_{ij}^{-1} + \overline{N} x_r^i x_r^j), \quad 1 \leq i < j \leq n,$$

$$\widehat{g}_{ii} = \partial \Phi_N(B, b)/\partial b_{ii} = -(n+1)b_{ii}^{-1} + n c_{ii}^{-1} + \overline{N}(x_r^i)^2, i = 1, \ldots, n,$$

$$\widehat{g}_{i} = \partial \Phi_N(B, b)/\partial b_{i} = 2(n c_{in+1}^{-1} + \overline{N} x_r^i), \quad i = 1, \ldots, n,$$

где $B^{-1} = (b_{ij}^{-1})_{i,j=1}^{n}$ – матрица, обратная к B; $C^{-1} = (c_{ij}^{-1})_{i,j=1}^{n+1}$ – матрица, обратная к C;

$$N = \begin{cases} N, & \text{если } h_r(B,b) = \max\limits_{1 \le j \le m} h_j(B,b) > 0, \\ 0 & \text{в противном случае.} \end{cases}$$

При использовании монотонных вариантов r-алгоритма (4) можно решать как задачу безусловной минимизации. Выполнение ограничения $B \succ 0$ в процессе поиска решения гарантируется тем, что при приближении матрицы B к границе допустимой области (изнутри ее) $\varphi(B,b) \to 0$, $\ln\varphi(B,b) \to -\infty$, а значит, $\Phi_N(B,b) \to \infty$. Таким образом, наличие логарифма в целевой функции представляет «барьер», не позволяющий при минимизации $\Phi_N(B,b)$ выйти за пределы области $B \succ 0$ при специальной регулировке шагового множителя r-алгоритма.

Построение эллипсоида минимального объема, включающего пересечение заданных эллипсоидов. Пусть заданы k эллипсоидов $E_j = (A_j, a_j)$, $j = 1,\ldots,k$ $(k \ge 2)$. Требуется построить эллипсоид $E^* = (A^*, a^*)$ минимального объема, включающий $\bigcap_{j=1}^{k} E_j$. Опишем конструктивный алгоритм приближенного решения данной задачи.

Алгоритм итеративный, на каждой итерации строится эллипсоид минимального объема, включающий некоторое сгенерированное множество точек из $\bigcap_{j=1}^{k} E_j$, и находится подмножество точек из $\bigcap_{j=1}^{k} E_j$, не принадлежащих построенному эллипсоиду. Объединение этого подмножества и имеющегося множества точек является базовым для построения эллипсоида на следующей итерации. Опишем основные пункты алгоритма.

Поиск начальной точки. Проводится поиск точки $x_0 \in \bigcap_{j=1}^{k} E_j$. Для этого решается негладкая задача выпуклого программирования

$$f^* = \min_x \max\left\{(x - a_j)A_j(x - a_j)^T - 1, \quad j = 1,\ldots,k\right\}. \tag{5}$$

Если $f^* > 0$, то $\bigcap_{j=1}^{k} E_j = \emptyset$ и алгоритм останавливается.

Генерация начального базового множества точек. Необходимо сгенерировать некоторое множество точек границы $\bigcap_{j=1}^{k} E_j$. Для этого из точки x_0 проведем секущие до пересечения с эллипсоидами, при этом направления секущих зададим изотропными векторами с началом в x_0. Напомним, что случайный вектор $s = (s_1,\ldots,s_n)$ называется изотропным, если точка $s/|s|$ распределена равномерно на поверхности сферы $|s| = 1$. Уравнение секущей представим в виде $x(z) = x_0 + zs$. Для того, чтобы найти точки пересечения $x(z)$ с эллипсоидом E_j, нужно решить квадратное уравнение $c_1 z^2 + 2c_2 z + c_3 = 0$, где $c_1 = sA_j s^T$, $c_2 = sA_j(x_0 - a_j)^T$, $c_3 = (x_0 - a_j)A_j(x_0 - a_j)^T$. Обозначим z_j^1 и z_j^2 корни квадратного уравнения ($z_j^1 \geq 0$, $z_j^2 < 0$). Найдем $\overline{z}$ – минимальный элемент множества $\{z_1^1, \ldots, z_k^1\}$ и $\tilde{z}$ – максимальный элемент $\{z_1^2, \ldots, z_k^2\}$. Нетрудно проверить, что точки $x_1 = x_0 + \overline{z}s$ и $x_2 = x_0 + \tilde{z}s$ принадлежат границе $\bigcap_{j=1}^k E_j$. Проводя N_0 секущих из точки x_0, получим $2N_0$ точек границы $\bigcap_{j=1}^k E_j$. Обозначим начальное базовое множество X_1, $|X_1| = 2N_0 = M_1$.

Построение описанного эллипсоида. Требуется построить эллипсоид минимального объема $E^i = (a^i, A^i)$, включающий все точки из X_i, $|X_i| = M_i$. Метод и алгоритм такого построения описаны выше. Для записи уравнения полученного эллипсоида необходимо произвести такие преобразования: $a = bB^{-1}$, $A = B/(1 - bB^{-1}b^T)$.

После построения описанного эллипсоида проводится сжатие базового множества точек X_i. Назовем величину $\varphi^i(x) = 1 - (x - a^i)A^i(x - a^i)^T$ отклонением точки x от эллипсоида E^i. При сжатии в множестве X_i оставляем точки, отклонения которых от эллипсоида E^i меньше заданной величины, т. е. точки, лежащие вблизи границы E^i. Обозначим сжатое множество X_i^0

Генерация нового подмножества точек. Найдем точки $\bigcap_{j=1}^k E_j$, которые не принадлежат E^i, в два приема. Сначала из точки a^i проведем N секущих до пересечения с E_j, $j = 1, \ldots, k$. Нахождение точек пересечения секущих с границей $\bigcap_{j=1}^k E_j$ описано выше. Точки с максимальными положительными отклонениями от построенного эллипсоида E^i (их множество обозначим $\overline{X}_i$) присоединим к множеству X_i^0.

Затем решим задачу: найти точку $x \in \bigcap_{j=1}^k E_j$, максимально удаленную от E^i. Ее можно записать в виде

$$\max_x \{ \left[(x - a^i)A^i(x - a^i)^T - 1 \right] : \varphi_j(x) \leq 0, \quad j = 1, \ldots, k \}, \tag{6}$$

где $\varphi_j(x) = (x - a_j)A_j(x - a_j)^T - 1$. Используя метод точных негладких штрафных функций, перепишем задачу (6) в виде

$$\max_x \{ \left[(x - a^i)A^i(x - a^i)^T - 1 \right] - s \sum_{j=1}^k \max\{0, \varphi_j(x)\} \}, \tag{7}$$

где s – достаточно большой штрафной множитель. Так как задачи (6) и, соответственно, (7) многоэкстремальные, задавая в качестве начальных приближений точки из $X_i^0 \cup \overline{X}_i$, получим, решив (7), локальные максимумы задачи (6). Обозначим множество получаемых локальных максимумов, отклонение которых от эллипсоида E^i больше ε (ε – заданная точность построения описанного эллипсоида), $\widetilde{X}_i$.

Сформируем базовое множество для следующей итерации:

$$X_{i+1} = X_i^0 \cup \overline{X}_i \cup \widetilde{X}_i, \quad |X_{i+1}| = M_{i+1}.$$

Преобразование пространства. После построения нового описанного эллипсоида $E^1 = (A^1, a^1)$ проведем линейное преобразование пространства, в результате которого эллипсоид преобразуется в единичный шар. Так как секущие из центра эллипсоида проводятся через точки, равномерно распределенные на поверхности сферы, то в случае сильно вытянутого по каким-то направлениям эллипсоида точки пересечения секущих с эллипсоидом будут неравномерно распределены на его поверхности. Указанное преобразование пространства выравнивает плотность распределения этих точек.

Матрицу A^1 можно представить в виде $A^1 = R\Lambda R^T$, где Λ – диагональная матрица собственных чисел A^1; R – соответствующая матрица собственных векторов. Обозначим $v = R\overline{\Lambda}$, где $\overline{\Lambda}^2 = \Lambda$. Тогда $A^1 = vv^T$. Сделаем замену переменных $y = xv$. Уравнение эллипсоида E^1 в новых переменных превратится в уравнение единичного шара с центром в точке $a^1 v$. Уравнение эллипсоида E_j перепишем в виде

$$(y - a_j v)v^{-1} A_j (v^{-1})^T (y - a_j v)^T \leq 1, \quad j = 1, \ldots, k.$$

Для возврата к исходным переменным необходимо сделать замену $x = v^{-1} y$.

Критерий останова алгоритма. Алгоритм останавливается, когда не удается за определенное число попыток сгенерировать ни одной точки $\bigcap_{j=1}^k E_j$, не принадлежащей построенному эллипсоиду.

Реализация алгоритма. Описанный алгоритм требует решения трех негладких задач математического программирования:

(5) – поиск начальной точки;

(4) – построение оптимального описанного эллипсоида по точкам;

(7) – поиск максимально удаленной от эллипсоида точки из $\bigcap_{j=1}^k E_j$.
Эти задачи можно успешно решить с помощью r-алгоритма.

При решении (4) используется специальная регулировка шага r-алгоритма. Выход за границу области положительной определенности матрицы B возможен из-за большого значения шагового множителя. Предположим, $B_r \succ 0$. На $(r+1)$-м шаге вычисляем $B_{r+1} = B_r + h_r g_r$. Факторизуя матрицу B_{r+1}, проверим, является ли она положительно определенной. Если условие $B_{r+1} \succ 0$ не выполняется, то шаг h_{r_1} делится пополам $1, 2, \ldots, s$ раз до тех пор, пока матрица $B_{r+1} = B_r + \dfrac{h_r}{2^s}\overline{g}_r$ не будет положительно определенной.

Алгоритм реализован для случая пересечения двух эллипсоидов. Программа написана на языке ФОРТРАН, на СМ 1420 был просчитан ряд тестовых примеров для пространства размерностью $n = 2 \div 5$.

Численные эксперименты показали хорошую точность вычисления параметров оптимального описанного эллипсоида.

Литература

1. Калинин В. Н., Шикин Е. В. О построении эллипсоидов экстремального объема // Изв. АН СССР. Техн. кибернетика. – 1987. – № 4. – С. 60–65.

2. Шор Н. З., Стеценко С. И. Использование точных штрафов при построении описанных эллипсоидов минимального объема // Кибернетика. – 1988. – № 2. – С 75–78

Использование алгоритма субградиентного типа с растяжением пространства для построения эллипсоида максимального объема, вписанного в многогранник

Н. З. Шор, О. А. Березовский
Кибернетика. – 1989. – № 6. – С. 119–120.

Важность задачи построения n-мерного эллипсоида E_M максимального объема, вписанного в заданный n-мерный многогранник $M \subseteq R^n$, определяется тем, что центр $a(E_M)$ такого эллипсоида близок по своим свойствам центру тяжести C_M тела M. В частности, положение центра эллипсоида $a(E_M)$ так же, как и C_M, инвариантно по отношению к линейному преобразованию координат. Далее, как показано в [1], при проведении через точку $a(E_M)$ произвольной отсекающей гиперплоскости объём эллипсоида, вписанного в новый многогранник, не превышает 0,843 объёма E_M (аналогичная константа для C_M равна $1/e$ [2]). Это дает возможность строить алгоритмы минимизации определенных на многограннике M выпуклых функций, основанные на методе последовательных отсечений, проходящих через точки $a(E_{M_k})$, где $\{M_k\}_{k=1}^{\infty}$ – последовательность локализующих точку минимума многогранников, $M_1 = M$ [1].

Нахождение центра тяжести C_M требует с возрастанием размерности n и числа отсекающих плоскостей (граней) m экспоненциально растущего объема вычислений. Практически, уже при $n = 4$ нахождение C_M при достаточно большом m представляет с вычислительной точки зрения весьма тяжелую задачу. Тем большее значение приобретает проблема построения максимального по объему вписанного эллипсоида. В последние годы эта задача была рассмотрена в ряде работ [1, 3].

Пусть многогранник M задается системой линейных неравенств

$$(b_i x) \leq 1, \quad i = 1, \dots k, \quad b_i,\, x \in R^n. \tag{1}$$

Уравнение эллипсоида E_M запишем в виде

$$(K(x-a), x-a) \leq 1, \tag{2}$$

где $K = \{k_{ij}\}_{i,j=1}^{n}$ – симметричная положительно определенная матрица с неизвестными компонентами k_{ij}, $i, j = 1, \ldots, n$, $k_{ij} = k_{ji}$, $a = \{a_1, \ldots, a_n\}$ – неизвестный центр эллипсоида E_M.

Далее покажем, как задача нахождения максимального вписанного эллипсоида E_M может быть сведена к задаче выпуклого программирования, а затем, с использованием точной негладкой функции штрафа, к задаче безусловной минимизации негладкой выпуклой функции. Для решения последней задачи применялась специальная модификация r-алгоритма [4], реализующего метод субградиентного типа с растяжением пространства в направлении разности двух последовательных субградиентов.

Условие того, что эллипсоид E_M (2) содержится в многограннике M, можно записать в виде

$$\min_{x \in M_i} (K(x-a), x-a) \geq 1, \quad M_i = \{x : (b_i, x) \geq 1\}, \quad i = 1, \ldots, k. \tag{3}$$

Объем эллипсоида E_M пропорционален $(\det K)^{-1}$, поэтому задача получения максимального по объему вписанного эллипсоида может быть сформулирована в следующей форме: $\min(\ln \det K)$ при условии (3), $K \in \Omega^+$, Ω^+ – класс положителько определенных матриц.

Условие (3) можно выразить в явном виде, используя метод множителей Лагранжа с учетом того, что задачи вида (3) представляют собой задачи выпуклого программирования при $K \in \Omega^+$.

Пусть

$$L_i(x, \lambda) = (K(x-a), x-a) + \lambda(1 - (b_i, x)).$$

Решение i-й задачи вида (3) сводится к решению системы уравнений

$$\begin{cases} \dfrac{\partial L_i}{\partial x} = 2Kx - 2Ka - \lambda b_i = 0, \\ \dfrac{\partial L_i}{\partial \lambda} = (b_i, x) - 1 = 0. \end{cases}$$

Отсюда имеем

$$x = a + K^{-1} b_i (1 - (a, b_i)) / (K^{-1} b_i, b_i), \quad i = 1, \ldots, k. \tag{4}$$

Из (3), (4), учитывая, что $(b_i, a) < 1$, получаем

$$\sqrt{(K^{-1} b_i, b_i)} + (b_i, a) - 1 \leq 0, \quad i = 1, \ldots, k. \tag{5}$$

Пусть $K^{-1} = Q = \{q_{ij}\}_{i,j=1}^n$, тогда задача построения максимального по объему вписанного эллипсоида может быть представлена следующим образом: найти

$$V_0 = \inf_y \{f_0(y) : f_i(y) \leq 0, \quad i = 1, \ldots, k; \quad y \in \overline{\Omega}\}, \tag{6}$$

где

$$y = \left\{\{q_{ij}\}_{i,j=1}^n, \{a_i\}_{i=1}^n\right\}, \quad f_0(y) = -\ln\det Q,$$

$$f_i = \sqrt{(Qb_i, b_i)} + (b_i, a) - 1, \quad i = 1, \ldots, k, \quad \overline{\Omega} = \{y : Q \in \Omega^+\}.$$

Используем для учета ограничений в (6) негладкую штрафную функцию в форме функции максимума:

$$\begin{aligned} \Phi_N(y) &= f_0(y) + N \cdot F(y), \\ F(y) &= \max\{0,\ f_1(y), \ldots,\ f_k(y)\}. \end{aligned} \tag{7}$$

В монографии Б. Н. Пшеничного [5] приводится следующая теорема.

Пусть

$$V(z) = \inf\Big\{ f_0(y) : f_i(y) \leq z_i; \quad i = 1, \ldots, k; \quad y \in M \Big\},$$

где $f_i(y)$, $i = 0, \ldots, k$, – непрерывные функции, $M \subseteq R^n$ – некоторое множество, I – вектор, все компоненты которого равны единице.

Теорема. *Пусть* $\inf\limits_{\lambda>0} \dfrac{V(\lambda I) - V(0)}{\lambda} = -L > -\infty$ *и* $N > L$, *тогда точки минимума задач* $V(0)$ *и* $\inf\limits_y \Phi_N(y)$, *где* $\Phi_N(y)$ *определяется по формуле* (7), *совпадают.*

Применим сформулированную теорему для определения штрафного коэффициента N в задаче определения максимального по объему вписанного эллипсоида. Условие $\sqrt{(Qb_i, b_i)} + (b_i,\, a) - 1 \leq \lambda$ эквивалентно условию $\sqrt{\dfrac{(Qb_i, b_i)}{(1+\lambda)^2}} + \dfrac{(b_i,\, a)}{1+\lambda} - 1 \leq 0$, которому соответствует неравенство $\sqrt{(\widetilde{Q}b_i, b_i)} + (b_i, \widetilde{a}) - 1 \leq 0$, где $\widetilde{Q} = \dfrac{Q}{(1+\lambda)^2}$, $\widetilde{a} = \dfrac{a}{1+\lambda}$.

Учитывая, что $\det Q$ является однородной функцией степени n, получаем, что при $\lambda > -1$

$$V(\lambda f) - V(0) \;=\; -\ln(1+\lambda)^{2n} = -2n\ln(1+\lambda),$$

Таблица 1.

n	k	t	m	$\delta \times 10^6$
2	4	1,04	50	0,3
3	6	2,69	77	0,5
4	8	6,50	94	1,5
5	10	20,16	129	0,7
10	20	308,19	251	0,3

$$\inf_{\lambda>0} \frac{V(\lambda f) - V(0)}{\lambda} = \inf_{\lambda>0} \frac{-2n\ln(1+\lambda)}{\lambda} = -2n > -\infty.$$

Таким образом, справедлива следующая теорема.

Теорема 1. *При $N > 2n$ точки минимума задачи* (6) *и задачи: найти* $\inf \Phi_N(y)$ *при $y \notin \Omega$, совпадают.*

Пусть $Q = P^2$, $P = \{p_{ij}\}_{i,j=1}^n$ – положительно определенная матрица. При этом задача (6) принимает следующий вид: найти $\inf\limits_{p,a}(-2\ln\det P)$ при ограничениях $\|Pb_i\| + (b_i, a) - 1 \le 0$, $i = 1, \ldots, k$, $P \in \Omega^+$. В силу теоремы 1 она сводится к задаче минимизации точной штрафной функции:

$$\overline{\Phi}_N(P, a) = -2\ln\det P + N\max\{0, \ \max_i (\|Pb_i\| + (b_i, a) - 1)\}$$

при условии, что $P \in \Omega^+$ принадлежит классу положительно определенных матриц и $N > 2n$. Так как Ω^+ – выпуклое множество в пространстве симметричных матриц размерности n, $\overline{\Phi}_N(P, a)$ – выпуклая функция, то соответствующая задача является задачей выпуклого программирования. Учет ограничения $P \in \Omega^+$ облегчается тем, что при приближении к границе Ω^+ $\det P$ стремится к 0, т. е. $-2\ln\det P \to +\infty$, что создает естественный «барьер» от выхода матрицы P за пределы области при использовании монотонного метода минимизации $\overline{\Phi}_N(P, a)$, начиная от некоторой допустимой точки (P_0, a_0), $P_0 \in \Omega^+$.

Так как $\overline{\Phi}_N(P, a)$ – выпуклая недифференцируемая функция, то для ее минимизации естественно применять один из эффективных методов субградиентного типа. Среди таких методов быстрой практической сходимостью обладает r-алгоритм [4]. На его основе авторами построен алгоритм минимизации $\overline{\Phi}_N(P, a)$. Эта модификация r-алгоритма учитывает некоторые особенности задачи, что достигается регулировкой шага. Каждая итерация включает проверку условий $P \in \Omega^+$ и $(b_i, a) \leq 1,\ i = 1, \ldots, k$. В случае их нарушения шаг дробится, пока $\{P, a\}$ не возвратится в допустимую область. Кроме того, так как $P \notin \overline{\Omega}^+ \setminus \Omega^+$, для лучшей сходимости вычислительного процесса P не допускается к границе Ω^+, что достигается дроблением шага в случае, если $\dfrac{|\ln\det P_m|}{|\ln\det P_{m+1}|} > c$, где c – некоторая константа, m – номер итерации.

Некоторые результаты численных экспериментов на ЕС ЭВМ приводятся в табл. 1, где в первой строке указывается размерность пространства n, во второй – количество исходных ограничений k, в третьей – время счета t в сек, за которое получен результат с точностью

$$\max_j \frac{\left(P_{jj}^m - P_{jj}^{m-1}\right)^2}{\left(P_{jj}^{m-1}\right)^2} < 10^{-12}, \quad j = 1, \ldots, n,$$

в четвертой – количество итераций m, в пятой – δ,

$$\delta = \frac{\|a^m - a(E_M)\|}{\max\limits_{x_1, x_2 \in M} \|x_1 - x_2\|}.$$

Отметим, что в [1], [3] для нахождения вписанного максимального по объему эллипсоида E_M предлагается использовать метод эллипсоидов [4], который, однако, обладает медленной сходимостью.

Литература

1. Тарасов С. П., Хачиян Л. Г., Эрлих И. И. Метод вписанных эллипсоидов // Докл. АН СССР. – 1988. – 298. – № 5. – С. 1081–1085.

2. Левин А. Ю. Об одном алгоритме минимизации выпуклых функций // Там же. – 1965. – 160. – № 6. – С. 1245.

3. Калинин В. Н., Шикин Е. В. О построении эллипсоидов экстремального объема // Техн. кибернетика. – 1987. – № 4. – С. 60–65.

4. Шор Н. З. Методы минимизации недифференцируемых функций и их приложения. – К.: Наук. думка, 1979. – 200 с.

5. Пшеничный Б. Н. Метод линеаризации. – М.: Наука – 1983. – 136 с.

Алгоритм последовательного сжатия пространства для построения описанного эллипсоида минимального объема

Н. З. Шор, С. И. Стеценко

Исследование методов решения экстремальных задач. – 1990. – С. 25–29.

В основе алгоритма – последовательные сжатия пространства по максимальному вектору. Алгоритм протестирован на большом числе задач.

Рассмотрим задачу построения эллипсоида минимального объема, включающего набор точек $X = \{x_1, \ldots, x_m\}$ пространства R^n. Эллипсоид $E = (A, a)$ в R^n представляется неравенством $(A(x-a), x-a) \leq 1$, где A – симметричная положительно определенная матрица; $A \in R^{n\times n}$; $a = \left(a^1, \ldots, a^n\right)$ – центр эллипсоида; $x = \left(x^1, \ldots, x^n\right)$ – точка эллипсоида. Показано [1], что эту задачу можно записать в виде

$$\max_A \ln\det A, \tag{1}$$

$$(A(x_j - a), x_j - a) \leq 1, \quad j = 1, \ldots, m, \tag{2}$$

$$A \succ 0, \tag{3}$$

где $A \succ 0$ – условие положительной определенности матрицы A. Размерность вектора переменных (A, a) равна $n(n+3)/2$.

Для решения задачи (1)–(3) предложен [1] метод точных негладких штрафных функций, найдена оценка величины штрафного множителя. Для максимизации негладкой функции штрафа использован метод градиентного типа с растяжением пространства — r-алгоритм [2] со специальной регулировкой шагового множителя. Операции с матрицей растяжения пространства требуют $O\left(n^4\right)$ операций на каждой итерации алгоритма.

Опишем другой подход к решению указанной задачи. Вместо нахождения эллипсоида с неизвестным центром в R^n будем искать эллипсоид в $(n+1)$-мерном пространстве R^{n+1} с центром в начале координат. Переведем точки $x_1, x_2, \ldots, x_m$ в пространство R^{n+1}, зафиксировав

$x_i^{n+1} = 1,\ i = 1, \dots, m$. Новая задача примет вид

$$\max_{\overline{B}}\ \ln\ \det \overline{B}, \tag{4}$$

$$\left(\overline{B}\overline{x}_j, \overline{x}_j\right) \le 1, \quad j = 1, \dots, m, \tag{5}$$

$$\overline{B} \succ 0, \tag{6}$$

где $\overline{x}_j = (x_j, 1), \quad j = 1, \dots, m,\ \overline{B}$ – симметричная положительно определенная матрица, $\overline{B} \in R^{(n+1)\times(n+1)}$.

Решив задачу (4)–(6), найдем оптимальный эллипсоид $\overline{E}$ в $(n+1)$-мерном пространстве с центром в $\overline{O}$. Оптимальный эллипсоид $E \in R^n$ получится в сечении эллипсоида $\overline{E}$ гиперплоскостью $x^{n+1} = 1$. Пусть матрица $\overline{B}$ имеет вид $\overline{B} = \begin{pmatrix} B & b^T \\ b & \widehat{b} \end{pmatrix}$, где B – главный минор матрицы $\overline{B}$ размерностью $n \times n$; $b = \left(\overline{b}_{n+1,1}, \dots, \overline{b}_{n+1,n}\right)$, $\widehat{b} = \overline{b}_{n+1,n+1}$. Тогда оптимальное решение задачи (1)–(3) находится по формулам

$$a = -B^{-1}b, \quad A = B/\big(1 - \widehat{b} + \left(B^{-1}b, b\right)\big). \tag{7}$$

Функция $\ln\det\overline{B}$ строго вогнута в области положительной определенности матриц $\overline{B}$. Так как ограничения (5) линейны, а условие положительной определенности матрицы – выпуклое неравенство, задача (4)–(6) является задачей выпуклого программирования. В случае, когда ранг совокупности точек $\{x_2 - x_1, \dots, x_m - x_1\}$ совпадает с размерностью пространства n, задача (4)–(6) имеет единственное решение.

Опишем алгоритм решения задачи (4)–(6). Идея алгоритма состоит в следующем. Пусть W – выпуклое замыкание точек $\overline{O},\ x_1, \dots, x_m$. Нужно построить эллипсоид $\overline{E}$ с центром в $\overline{O}$, содержащий W, такой, чтобы отношение объемов $V(\overline{E})/V(W)$ было минимальным. Отношение объемов не меняется при невырожденных линейных преобразованиях пространства. Всегда найдется такое преобразование, что оптимальный эллипсоид станет сферой. Будем приближаться к оптимальному эллипсоиду путем построения последовательности описанных сфер в преобразованных пространствах.

Опишем вокруг W сферу минимального радиуса, который соответствует норме максимального вектора. Допустим, что такой вектор x_j единственный. Тогда, если провести сжатие вдоль x_j с коэффициентом $q < 1$, то объем W уменьшится в q раз. Если при этом образ x_j в новом пространстве окажется опять максимальным по норме, то объем сферы уменьшится в q^{n+1} раз, т. е. отношение объемов уменьшится в q^n

раз. На самом деле темп изменения этого отношения будет меньше при появлении «конкурирующих» точек. Идея сжатия по максимальному вектору приводит к итерационному процессу, на каждом шаге которого ищется максимальный вектор, проводится сжатие пространства вдоль этого вектора и вычисляется матрица преобразования исходного пространства.

Пусть за k шагов получены: X_k – матрица, составленная из вектор-столбцов $x_1^k, \ldots, x_m^k$ – образов точек $x_1, \ldots, x_m$ в преобразованном пространстве; A_k – матрица преобразования пространства.

На $(k+1)$-м шаге вычисляем

$$\begin{aligned}
\tau_{k+1} &= \max_{1 \le i \le m} \left\| x_i^k \right\| = \left\| x_j^k \right\|, \\
\xi_{k+1} &= x_j^k / \left\| x_j^k \right\|, \\
X_{k+1} &= R_{\alpha_{k+1}} \left(\xi_{k+1} \right) X_k, \\
A_{k+1} &= R_{\alpha_{k+1}} \left(\xi_{k+1} \right) A_k.
\end{aligned}$$

Здесь $R_\alpha\left(\xi\right)$ – оператор сжатия пространства в направлении ξ с коэффициентом α, $\alpha < 1$ [2]. Коэффициенты $\alpha_k = 1 - \beta_k$ выбираются из условия

$$\beta_k > 0, \quad \sum_{k=1}^{\infty} \beta_k = +\infty, \quad \beta_k \underset{k \to \infty}{\longrightarrow} 0. \tag{8}$$

Матрица эллипсоида, являющегося прообразом сферы радиуса τ_k преобразованного пространства в исходном пространстве, вычисляется по формуле $B_k = A_k A_k^T \, / \, \tau_k^2$. Трудоемкость одной итерации предлагаемого алгоритма равна $O(n(n+m))$.

Доказательство сходимости к оптимальному эллипсоиду проведем для случая $m = n$. Пусть $x_1, \ldots, x_n$ – линейно независимые векторы, $\|x_i\| = 1$, $i = 1, \ldots, n$. Проведем сжатие пространства в направлении одного из векторов набора x_j, $j \in \{1, \ldots, n\}$ с коэффициентом $1 - \beta$. Получим набор векторов

$$x_i^{(j)}\left(\beta\right) = R_{1-\beta}\left(x_j\right) x_i = x_i - \beta \left(x_i, x_j\right) x_j, \quad i = 1, \ldots, n.$$

Пусть X $\left(X^{(j)}(\beta)\right)$ – матрицы, составленные из вектор-столбцов $\{x_i\}_{i=1}^n$ $(\{x_i^{(j)}(\beta)\}_{i=1}^n)$ соответственно, т. е. $X^{(j)}\left(\beta\right) = R_{1-\beta}\left(x_j\right) X$, откуда $\det X^{(j)}\left(\beta\right) = \left(1 - \beta\right) \det X$. Определим коэффициент обусловленности

n-матрицы A со столбцами $\{a_i\}_{i=1}^n$ следующим образом:

$$\alpha(A) = \sqrt{\det\left(A^T A\right)} \Bigg/ \prod_{i=1}^{n} \|a_i\| \, .$$

Учитывая, что

$$\|x_i^{(j)}(\beta)\|^2 = 1 - 2\beta\left(x_i, x_j\right)^2 + \beta^2\left(x_i, x_j\right)^2 = 1 - 2\beta\left(x_i, x_j\right)^2 + O\left(\beta^2\right),$$

$$\begin{aligned}\alpha(x^{(j)}(\beta))/\alpha(X) &= \frac{\sqrt{\det\left(\left[X^{(j)}(\beta)\right]^T X^{(j)}(\beta)\right)}}{\sqrt{\det\left(X^T X(1-\beta)\prod_{i=1, i\neq j}^{n}\|x_i\|\right)}} = \\ &= 1 + \beta \sum_{i=1, i\neq j}^{n} \left(x_i, x_j\right)^2 .\end{aligned} \qquad (9)$$

Используя (9), докажем следующие утверждения.

Лемма 1. *В результате использования операции сжатия пространства по вектору x_j с коэффициентом $1-\beta$ обусловленность системы увеличивается на величину порядка* $\beta\sum_{i=1, i\neq j}^{n}(x_i, x_j)^2$.

Лемма 2. *При условиях (8) $\alpha\left(X_k\right) \underset{k\to\infty}{\longrightarrow} 1$, где X_k – последовательность матриц, получаемых в алгоритме.*

Теорема. *Если $m = n$, $\{x_i\}_{i=1}^n$ – линейно независимы, то*

$$\lim_{k\to\infty} \frac{\left(x_i^k, x_j^k\right)}{\left\|x_i^k\right\| \left\|x_j^k\right\|} = 0 \text{ для } i \neq j, \ \frac{\left\|x_i^k\right\|}{\left\|x_j^k\right\|} \to 1, \ \lim_{k\to\infty} B_k = B^*,$$

B^* – матрица оптимального эллипсоида.

Тестирование алгоритма проводилось следующим образом. С помощью датчика случайных чисел генерировались точки на поверхности сферы единичного радиуса с центром в $\overline{1} = (1, 1, \ldots, 1)$ пространства R^n, $n = 2, 3, \ldots, 10$. По этим точкам строился эллипсоид минимального объема. Сначала проводилось сжатие пространства с постоянным коэффициентом $\alpha = 1 - \beta$, где $\beta = 0.2$. Когда уменьшение длин максимальных векторов замедлялось, коэффициент сжатия выбирался $\alpha_k = 1-\beta_k$, где $\beta_k = 1/k$, $k = 1, 2, \ldots$.

Результаты численных экспериментов приведены в таблице. Приняты такие условные обозначения: n – размерность пространства; m – минимальное число точек, по которым удавалось восстановить исходную сферу с точностью 0.01 по детерминанту и координатам центра; ε – точность решения задачи (4)–(6) по функционалу; k – число итераций; t – время решения в секундах. Эксперименты проводились на ЭВМ IBM PC AT с ординарной точностью.

n	2	3	4	5	6	7	8	9	10
m	30	50	70	70	90	100	110	120	130
ε	0.001	0.001	0.001	0.01	0.01	0.01	0.01	0.01	0.01
k	200	300	300	400	400	500	600	700	700
t	17	35	62	91	127	203	280	348	383

Литература

1. Шор Н. З., Стеценко С. И. Использование точных штрафов при построении описанных эллипсоидов минимального объема // Кибернетика. – 1989. – № 2. – С. 117–119.

2. Шор Н. З. Методы минимизацин недифференцируемых функций и их приложения. – К.: Наук. думка, 1979. – 200 с.

Построение эллипсоида максимального объема, вписанного в многогранник, с использованием последовательного растяжения пространства

Н. З. Шор, О. А. Березовский
Кибернетика и вычислительная техника. – 1992. – Вып. 93. – С. 1–6.

Предлагается алгоритм построения эллипсоида максимального объема, вписанного в многогранник, с использованием последовательности преобразований пространства. С помощью данного алгоритма реализуется метод вписанных эллипсоидов для минимизации выпуклых функций.

Эллипсоид $E = (K, a)$ в пространстве R^n задается неравенством

$$(K(x-a), x-a) \leq 1,$$

где K – симметричная положительно определенная $n \times n$ матрица; a – n-мерный вектор центра эллипсоида; x – точка эллипсоида, $x \in R^n$.

В работе [1] авторы уже обращались к задаче построения оптимально вписанного в многогранник $M = \{x : (c_i, x) \leq 1, \quad i = 1, \ldots, m\}$ эллипсоида, где она рассматривалась в следующей постановке:

$$\min(\ln \det K)$$

при условии

$$\max_x (c_i, x) \leq 1, \quad i = 1, \ldots, m, \text{ при ограничении } (K(x-a), x-a) \leq 1;$$

$K \in \Omega^+$ – условие положительной определенности матрицы. Данная проблема сводилась к задаче минимизации выпуклой негладкой штрафной функции при условии $K \in \Omega^+$ и оценке штрафного множителя. В программной реализации предложенного алгоритма решения рассматриваемой задачи использовался метод субградиентного типа с растяжением пространства – модификация r-алгоритма [2], требующая на каждой итерации $O(n^4)$ операций. В [3] предложен метод внутренних точек для построения оптимальных вписанных эллипсоидов.

Ниже рассматривается другой подход к решению задачи построения оптимально вписанного эллипсоида, идея которого аналогична идее алгоритма, предложенного в [4] для построения эллипсоида минимального объема, описанного вокруг заданного набора точек. Требуется построить эллипсоид E^*, вписанный в многогранник M так, чтобы отношение объемов $V(M)/V(E^*)$ было минимально. Для этого будем искать линейно преобразованное пространство, в котором искомый эллипсоид представляет собой сферу. Результат достигается путем построения последовательности вписанных сфер в соответствующим образом преобразованных пространствах.

Впишем в многогранник M сферу максимального объема с центром в заданной точке a. Допустим, что она касается только одной j-й гиперплоскости. Тогда, если провести растяжение пространства с коэффициентом α вдоль вектора $\xi = c_j / \|c_j\|$ и сместить центр a вдоль вектора $-\xi$, то объем M увеличивается в α раз. Если же при этом в преобразованном пространстве активным опять окажется j-е ограничение, то объем сферы увеличится не менее чем в α^n раз, т. е. отношение объемов в этом случае уменьшится в α^{n-1} раз, что верно и для прообразов многогранника и сферы в исходном пространстве, поскольку отношение объемов не меняется при невырожденных линейных преобразованиях пространства. В действительности темп изменения этого отношения будет меньше при появлении «конкурирующих» ограничений.

Опишем общую схему алгоритма.

Заданы: матрица C_0, составленная из векторов столбцов c_0^i, $i = 1, \ldots, m$, и m-мерный вектор b, которые задают многогранник $M = \left\{x : (c_0^i, x) + b^i \leq 0,\ i = 1, \ldots, m;\ x \in R^n\right\}$. Выбираем неособенную $n \times n$ матрицу P_0 (в общем случае $P_0 = I_n$ – единичная матрица) и точку a_0 начального приближения центра эллипсоида, $a_0 \in M$. Пусть в результате вычислений после k шагов имеем матрицу преобразования пространства P_k, очередное приближение a_k центра искомого эллипсоида в этом преобразованном пространстве и матрицу C_k образов векторов c_0^i, $i = 1, \ldots, m$, в этом преобразованном пространстве. Тогда $(k+1)$-й шаг описывается следующим образом. Вычисляем:

1) $r_{k+1} = \min\limits_{1 \leq i \leq m} \rho_{k+1}^i = \rho_{k+1}^j$, где $\rho_{k+1}^i = (-b^i - (c_k^i, a_k)) / \left\|c_k^i\right\|$ – расстояние от точки a_k до i-й гиперплоскости;

2) $\gamma_{k+1}^i = r_{k+1} / \rho_{k+1}^i, \quad i = 1, \ldots, m;$

3) $z_{k+1} = \sum\limits_{i \in I} \gamma_{k+1}^i c_k^i, \quad I = \left\{i : \gamma_{k+1}^i \geq \gamma^*, \quad i = 1, \ldots, m\right\},$

4) $a'_k = a_k + (\alpha_{k+1} - 1) r_{k+1} z_{k+1} \ / \ \|z_{k+1}\|$;

$\xi_{k+1} = c^j_k / \|c^j_k\|$;

5) $a_{k+1} = R_{1/\alpha_{k+1}}(\xi_{k+1}) a'_k$;

$C_{k+1} = R_{\alpha_{k+1}}(\xi_{k+1}) C_k$;

$P_{k+1} = P_k R_{\alpha_{k+1}}(\xi_{k+1})$.

Здесь $R_\alpha(\xi)$ – оператор растяжения пространства в направлении ξ с коэффициентом растяжения α, $\alpha > 1$ [2]. Коэффициенты $\alpha_k = 1 - \beta_k$ выбираются из условия

$$\beta_k > 0, \quad \sum_{k=1}^{\infty} \beta_k = +\infty, \quad \beta_k \underset{k \to \infty}{\longrightarrow} 0. \tag{1}$$

Матрица и центр искомого эллипсоида, который является прообразом сферы радиуса r_k преобразованного пространства в исходном пространстве, вычисляются по формулам

$$(K^*_k)^{-1} = P_k P_k^T r_k^2;$$

$$a^*_k = P_k a_k.$$

Покажем, что данный процесс можно интерпретировать как метод субградиентного спуска с изменяемой метрикой.

В работе [1] указано, что, если все правые части ограничений, определяющих многогранник, равны 1, задача построения оптимального вписанного эллипсоида сводится к минимизации штрафной функции

$$S(P, a) = -\ln \det P + n \max_i \left(\|Pc_i\| + (c_i, a) - 1 \right). \tag{2}$$

Вычислим субградиент этой функции в точке $\{I_n, \overline{a}\}$:

$$\frac{\partial S}{\partial P}(P, a) = -P^{-1} + n \frac{P c_j c_j^T}{\|P c_j\|},$$

где j – индекс, на котором достигается максимум в (2).

Отсюда

$$\frac{\partial S}{\partial P}(I_n, \overline{a}) \quad = \quad -I_n + n \frac{c_j c_j^T}{\|c_j\|},$$

$$\frac{\partial S}{\partial a}(I_n,\overline{a}) \quad = \quad nc_j.$$

Рассмотрим шаг субградиентного спуска из точки $\{I_n,\overline{a}\}$:

$$\begin{aligned} P_0 &= I_n, \quad a_0=\overline{a}, \\ P_1 &= P_0 - h\frac{\partial S}{\partial P}(I_n,\overline{a}), \\ a_1 &= \overline{a} - h\frac{\partial S}{\partial a}(I_n,\overline{a}), \end{aligned}$$

где $h>0$ – некоторый шаговый множитель.

$$\begin{aligned} P_1 = I_n - h\left(-I_n + n\frac{c_j c_j^T}{\|c_j\|}\right) &= (1+h)\left(I_n - \frac{hn\,\|c_j\|}{1+h}\frac{c_j c_j^T}{\|c_j\|^2}\right) = \\ &= (1+h)R_{\beta_0}(\xi_j), \end{aligned}$$

где $\xi_j = \dfrac{c_j}{\|c_j\|}$, $\beta_0 = 1-\dfrac{hn\,\|c_j\|}{1+h}$. $R_\beta(\xi)$ – оператор растяжения пространства в направлении ξ с коэффициентом растяжения β; $a_1=\overline{a}-hnc_j$.

Сделаем замену переменных:

$$(1+h)R_{\beta_0}(\xi_j)P = \overline{P}, \quad P = \overline{P}\frac{R_{1/\beta_0}(\xi_j)}{1+h}. \tag{3}$$

Тогда

$$S(P,a) = -\ln\det\overline{P} + \ln[\beta_0(1+h)] + n\max_i\left\{\frac{\|\overline{P}R_{1/\beta_0}(\xi_j)c_i\|}{1+h} + (c_i,a) - 1\right\}.$$

Обозначим $\overline{c}_i = R_{1/\beta_0}(\xi_j)\,c_i/(1+h)$.
Тогда

$$\begin{aligned} S(P,a) = -\ln\det\overline{P} &+ \ln[\beta_0(1+h)] + \\ &+ n\max_i\{\|\overline{P}\,\overline{c}_i\| + (\overline{c}_i, R_{\beta_0}(\xi_j)a(1+h)) - 1\}. \end{aligned}$$

После замены переменных P_1 соответствует $\overline{P}_1 = I_n$, т. е. в преобразованном пространстве мы опять находимся в точке I_n, и новый цикл вычислений, связанный с шагом субградиентного спуска в преобразованном пространстве, не будет отличаться от предыдущего.

Описанный выше алгоритм фактически реализует субградиентный метод с преобразованием пространства по формуле (3). Условие $\beta_k \to 0$

соответствует условию $h_k \to 0$ в субградиентном процессе. Кроме того, темп изменения метрики пространства с увеличением k стремится к 0, т. е. при больших k описанный выше алгоритм практически не отличается от субградиентного процесса.

Доказана следующая теорема: при условии (1) описанный выше алгоритм выдает последовательности $\{P_k, r_k, a_k\}$ такие, что

$$P_k P_k^T r_k^2 \to (K^*)^{-1}, \quad P_k a_k \to a^*.$$

Отметим, что в отличие от алгоритма, описанного в [1], рассмотренный выше требует на каждой итерации всего $O(n(n+m))$ операций. При тестировании, которое показало удовлетворительные результаты, γ^* присваивалось значение 0,99, а последовательность β_k выбиралась следующим образом: сначала растяжение пространства проводилось с постоянным коэффициентом (например, $\beta_0 = 0{,}2$), а затем, когда $\det K$ переставал уменьшаться, коэффициент растяжения выбирался $\alpha_k = 1 - \ \beta_k, \quad \beta_k = \dfrac{1}{k+1}, \quad k = 1, 2, \ldots$.

Построение эллипсоида максимального объема, вписанного в многогранник, связано с широким кругом приложений, например, таких, как аппроксимация областей достижимости в теории управления и дифференциальных играх, локализация областей возможных значений параметров динамических систем [5], в статистике, планировании экспериментов, при решении задач выпуклого программирования, поиска седловых точек и др. В частности, оно используется в методе вписанных эллипсоидов [6], являющимся методом отсечения для решения задач выпуклого программирования. Основная идея этого метода заключается в следующем. Требуется минимизировать выпуклую функцию $f(x)$ на выпуклом множестве $G_0 \subset R^n$. В соответствии с общей схемой методов отсечений на k-й итерации строим вписанный в G_{k-1} эллипсоид $E_k = (K_k, a_k)$ максимального объема и определяем текущий локализатор решения задачи

$$G_k = \{x : x \in G_0, (\mathrm{grad} f(a_i), x - a_i) \le 0, \quad i = 1, \ldots, k\}.$$

При этом отношение объемов

$$V(E_{k+1})/V(E_k) \le 0{,}843. \tag{4}$$

Учитывая, что отношение объемов оптимальных описанных вокруг выпуклого тела G и вписанных в него эллипсоидов не превышает n^n, последовательность $\{V(G_k)\}_{k=0}^{\infty}$ мажорируется сверху геометрической прогрессией со знаменателем 0,843.

Неравенство (4) допускает следующее обобщение в случае, если максимальный эллипсоид, в центре которого рассекается G_k, находится приближенно с относительной погрешностью по объему $\eta \in (0,1]$:

$$V(E_{k+1}) \;/\; V(E_k) \le 0,843\eta^{-2},$$

откуда при фиксированном η, например $\eta = 0{,}99$, вытекает гарантированная оценка

$$N(\varepsilon) \le 6,64n \ln \frac{1}{\varepsilon} \tag{5}$$

числа итераций $N(\varepsilon)$, достаточных для достижения относительной точности минимизации ε [6]. Таким образом, как и известный метод центров тяжестей, метод вписанных эллипсоидов является оптимальным по порядку числа итераций методом выпуклого программирования в классе методов, использующих лишь локальную информацию о поведении минимизируемой функции в окрестностях точек-приближений.

Ниже предлагается алгоритм, реализующий метод вписанных эллипсоидов [6] с использованием рассмотренного в этой работе приближенного построения оптимального вписанного эллипсоида. Необходимо особо подчеркнуть, что при переходе к следующей итерации метода пространство не восстанавливается, т. е. построение следующего эллипсоида начинается в том же преобразованном пространстве, в котором предыдущий эллипсоид представлял из себя сферу. Благодаря этому исключается сильная «вытянутость» допустимой области, что позволяет уменьшить количество итераций для решения внутренней задачи построения оптимального эллипсоида, упрощает выбор последовательности $\{\beta_k\}_{k=1}^{\infty}$ и способствует улучшению работы алгоритма.

Общая схема алгоритма выглядит следующим образом. Заданы: матрица C_0, составленная из векторов столбцов c_0^i, $i = 1,\ldots,m$, и вектор b, $b \in R^m$, которые задают допустимую область в виде многогранника $M = \{x : (c_0^i, x) + b^i \le 0,\ i = 1,\ldots,m\}$, минимизируемая функция $f(x)$, вектор активности ограничений $s_0 = 0$, неособенная матрица $P_0 = I_n$ и точка a_1^H начального приближения центра эллипсоида на первой итерации метода, $a_1^H \in M$. После k итераций имеем матрицу преобразования P_k $(P_{k+1}^H = P_k)$, матрицу C_k и вектор b, задающие гиперплоскости ограничений в этом преобразованном пространстве, точку a_{k+1}^H начального приближения центра эллипсоида на $(k+1)$-й итерации в преобразованном пространстве,

$$\left(P_k a_{k+1}^H\right) \in M_k = \{x : x \in M,\ (\mathrm{grad} f(P_i a_i),\ x - P_i a_i) \le 0,\ i = 1,\ldots,k\},$$

и вектор активности ограничений s_k (s_k^i равно количеству итераций, на протяжении которых i-е ограничение не участвовало в построении оптимальных эллипсоидов).

Опишем $(k+1)$-й шаг алгоритма:

1) строим оптимально вписанный в M_k эллипсоид

$$E_{k+1} = (K_{k+1}, P_{k+1}a_{k+1}),$$

при этом проверяя активность ограничений (если хоть раз растяжение пространства происходило по направлению направляющей i-й гиперплоскости, $s_{k+1}^i = 0$, в противном случае $s_{k+1}^i = s_k^i + 1$); P_{k+1} – матрица преобразования пространства, в котором эллипсоид E_{k+1} представляет собой сферу радиуса 1;

2) проверяем активность ограничений: если $s_{k+1}^i > s_{\max}$, то i-е ограничение выбрасывается;

3) определяем отсекающую гиперплоскость $(c_{k+1}^{m+k+1}, x) + b^{m+k+1} \le 0$ в преобразованном пространстве

$$c_{k+1}^{m+k+1} = P_{k+1}^T \operatorname{grad} f(P_{k+1}a_{k+1}), \quad b^{m+k+1} = -(c_{k+1}^{m+k+1}, a_{k+1});$$

4) вычисляем начальное приближение центра эллипсоида для следующей итерации

$$a_{k+2}^H = a_{k+1} - \delta r_{k+1} c_{k+1}^{m+k+1} / \|c_{k+1}^{m+k+1}\|, \quad r_{k+1} = 1.$$

При тестировании выбиралось $s_{\max} = 8$, $\delta = 0,5$ и для внутренней задачи построения эллипсоида $\beta_k = 1/(1+k)$. В качестве тестовых примеров использовались одно- и многомерные «овражные» функции в пространстве размерностью до 10. Количество итераций, необходимых для достижения относительной точности ε, соответствовало оценке (5).

Литература

1. Шор Н. З., Березовский О. А. Использование алгоритма субградиентного типа с растяжением пространства для построения эллипсоида максимального объема, вписанного в многогранник // Кибернетика. – 1989. – № 6. – С. 119–120.

2. Шор Н. З. Методы минимизации недифференцируемых функций и их приложения. – К.: Наук. думка, 1979. – 200 с.

3. Нестеров Ю. Е., Немировский А. С. Самосогласование функций в полиномиальные алгоритмы в выпуклом программировании. – М.: Центр. эконом.-мат. ин-т АН СССР, 1989. – 194 с.

4. Шор Н. З., Стеценко С. И. Алгоритм последовательного сжатия пространства для построения описанного эллипсоида минимального объема // Исследование методов решения экстремальных задач. – К.: Ин-т кибернетики АН УССР, 1990. – С. 25–29.

5. Черноусько Ф. Л. Оценивание разового состояния динамических систем (метод эллипсоидов). – М.: Наука, 1988. – 320 с.

6. Тарасов С. П., Хачиян Л. Г., Эрлих И. И. Метод вписанных эллипсоидов // Докл. АН СССР. – 1988. – Т. 298. – № 3. – С. 1081–1085.

New algorithms for constructing optimal circumscribed and inscribed ellipsoids

NAUM Z. SHOR and O. A. BEREZOVSKI [1]

Optimization Methods and Software. – 1992. – **1***. – P.283–299.*

This paper is an overview of some methods used to solve the problem of constructing an ellipsoid of minimal volume containing a set of m points and the problem of constructing an ellipsoid of maximal volume inscribed in a polyhedron defined by a system of m linear inequalities in n-dimensional Euclidean space (using nonsmooth convex penalty functions and successive space transformation). The type 1 algorithms require $O(n^3)$ or $O(m^3)$ arithmetical operations per iteration (depending on whether the prime or dual algorithm is considered); the type 2 algorithms require $O(nm)$ arithmetical operations per iteration.

1. Introduction

The solution of problems of constructing optimal with respect to volume circumscribed and inscribed ellipsoids is of great interest due to the broad range of their applications. Primarily, these problems can be used to define the «upper» and «lower» approximations of complex sets, for example, the regions of attainability in the theory of control and differential games; to refine the regions of localization of probable values of dynamic system parameters using *a priori* information and results of measurements [1]; then in statistics, in planning of experiments to determine parameters of regression models [3], in numerical mathematics to describe sets of possible solutions of linear and nonlinear equations with perturbed coefficients etc.

In recent years, interest in the problems of constructing inscribed and circumscribed ellipsoids has increased considerably owing to their possible applications in mathematical programming for approximating regions of localization of optimal points in extremal problems. Thus, the widespread

[1] The authors are grateful to the referees for useful comments.

«ellipsoid method» and its modifications are based on constructing ellipsoids of minimal volume, circumscribed around regions of optimum localization; the centres of optimal inscribed ellipsoids were used in [4] to construct fast converging cutting plane algorithms for finding the minimum of a convex function on a polyhedron. Geometrical ideas linked with the approximation of regions of localization of extremums by ellipsoids are used in the «interior points» methods [5]. Based on these methods, polynomial algorithms are built for solving important classes of convex programming problems.

Consider the problems of constructing in a sense optimal inscribed and circumscribed ellipsoids. Let W_n be a set of convex compact bodies of n-dimensional Euclidean space $I\!R^n$ including interior points. For each $M \in W_n$ let $O_1(M)$ be a set of ellipsoids circumscribed around M (including M) and $O_2(M)$ be a set of n-dimensional ellipsoids inscribed in M (included in M); $V(M)$ is an n-dimensional volume of compact $M \subset I\!R^n$. Let $E^*(M) \in O_1(M)$ be a circumscribed ellipsoid of minimal volume, and $E_*(M) \in O_2(M)$, an inscribed ellipsoid of maximal volume for a convex body M. The existence and uniqueness of $E_*(M)$ and $E^*(M)$ were first justified by Lewner in an oral presentation, so these ellipsoids are sometimes called «Lewner's» ellipsoids. Behrend [10] for $n = 2$ and John [11] for arbitrary n substantiated the existence and the uniqueness of optimal circumscribed and inscribed ellipsoids for any $M \in W_n$. John proved also that for each $M \in W_n$ the pair of homotetic ellipsoids $E^+(M)$ (circumscribed) and $E_+(M)$ (inscribed) with a common centre and coefficient of gomotety $1/n$ exists. Examples of such pairs can be:

(a) ellipsoid $E^*(M)$ and ellipsoid $E^*_+(M)$ with the coefficient of homotety $1/n$ homotetic to it with respect to the centre of $E^*(M)$;

(b) ellipsoid $E_*(M)$ and ellipsoid $E^*_+(M)$ with the coefficient of homotety n homotetic to it with respect to the centre of $E_*(M)$.

If set M is a central-symmetric body then the constant n in John's theorem can be replaced by $n^{1/2}$ [11].

It should be noted that the problems connected with inscribed and circumscribed ellipsoids are closely linked and are in a sense equivalent. This follows from the properties of the so-called polar transformation $\mathrm{po}M = \{y \ : \ (x, y) \leq 1, \ x \in M\}$ introduced by Minkovski for the set $M \in W_n$, $0 \in$ int M :

$\mathrm{po}(\mathrm{po}M) = M$;

$\mathrm{po}\,(\bigcap_{i=1}^m M_i) = \mathrm{conv}\,(\bigcup_{i=1}^m \mathrm{po}M_i)$;

if $M_1 \subset M_2$ then $\mathrm{po}M_2 \subset \mathrm{po}M_1$;

if M is an ellipsoid then po$M$ is an ellipsoid too;

if $M = (K, 0)$ is an ellipsoid with centre at the point 0 then po$M = \left(K^{-1}, 0\right)$ and $V(M) \cdot V(\text{po}M) = w_n^2$ where w_n is the volume of the sphere $\{x \;:\; \|x\| \leq 1\}$;

if E^* is an ellipsoid of minimal volume, circumscribed around M (the ellipsoid centre is fixed at the origin), then poE^* is the ellipsoid of maximal volume inscribed in po$M$ (with the same centre).

The transition from the algorithms for solving problems of constructing ellipsoids with fixed ellipsoid centres to solving problems with any centres is based on the next lemma.

Lemma 1. *Let the $(n+1)$-dimensional ellipsoid E_{n+1} be intersected by hyperplane P and this hyperplane does not contain the centre of E_{n+1}. Let E_n be an n-dimensional ellipsoid defined by $E_n = E_{n+1} \bigcap P$. Then the ratio of volumes*

$$\frac{V\left(E_n\right)}{V\left(E_{n+1}\right)} = \frac{w_n}{w_{n+1}} \cdot \frac{\left(1 - h^2/h_*^2\right)^{n/2}}{h_*},$$

where h_* is the distance from the centre of E_{n+1} to the hyperplane tangent to the ellipsoid E_{n+1} and parallel to P; and h is the distance from the centre of E_{n+1} to P.

Corollary 1. *To solve the problem of constructing ellipsoid $E_n = (K, b_*) \in \in I\!R^n$ optimal with respect to volume for the set M_n it is sufficient to find the solution of the problem of constructing an optimal ellipsoid $E_{n+1} = (\overline{K}, 0) \in I\!R^{n+1}$ provided:*

- (a) *condition $E_n \subseteq M_n$ (or $E_n \supseteq M_n$) for sets from $I\!R^n$ is transferred to the respective sets of the hyperplane $x_{n+1} = h$;*
- (b) *hyperplane $x_{n+1} = h_*$ is tangential to E_{n+1}, $h_* > h$.*

And

$$b_* = -h\,\widetilde{K}^{-1}\,r, \tag{1}$$

$$K = \widetilde{K}/(1 - k\,h^2 + h^2\,r^t\,\widetilde{K}^{-1}\,r), \tag{2}$$

where $\widetilde{K}$, r, r^t and k are corresponding blocks of

$$\overline{K} = \left[\begin{array}{c|c} \widetilde{K} & r \\ \hline r^t & k \end{array}\right]. \tag{3}$$

A reduction of the general extremal ellipsoid problem to the case of a fixed centre was obtained by Titterington [14].

Now one knows the polynomial-time solvability algorithms for constructing an optimal circumscribed ellipsoid ([5, 13, 15]). The algorithms considered below are simpler than these algorithms and can be competitive at large n and m when great accuracy of defining the parameters of the extremal ellipsoids is not required.

Consider the mathematical models of the problems of constructing inscribed and circumscribed ellipsoids optimal with respect to volume.

2. Solutions of the problems of finding extremal ellipsoids using the penalty function

2.1. The problem of constructing an ellipsoid of maximal volume inscribed in a polyhedron

Consider an ellipsoid represented as follows:

$$E = (K, b) = \{x \ : \ (x - b_*)^t \, K \, (x - b_*) \leq 1\}$$

($K = \{k_{ij}\}_{i,j=1}^n$ is a symmetric positive definite $n \times n$ matrix, b_* is an n-dimensional vector of the ellipsoid centre) the volume of which is $V(E) = \nu_0 (\det K)^{-1/2}$, where ν_0 is the volume of an n-dimensional unit sphere. The problem of finding $E_*(\overline{M})$ where $\overline{M}$ is defined by a finite system of linear inequalities (without loss of generality if $0 \in \overline{M}$ then one may write this system of inequalities as $(c_i, x) \leq 1, \ i = 1, \ldots, k$) may be represented as follows ([4, 5, 13]):

find

$$V = \min\{f_0(y) \mid f_i(y) \leq 0, \quad i = 1, \ldots, k; \quad y \in \overline{\Omega}\}, \tag{4}$$

where

$$\begin{aligned} f_0(y) &= -\ln \det Q, \quad K^{-1} = Q = \{q_{ij}\}_{i,j=1}^n \\ y &= \{\{q_{ij}\}_{i,j=1}^n, \{b_*\}\}, \\ f_i(y) &= \sqrt{c_i^t \, K^{-1} c_i} + (c_i, b_*) - 1, \\ \overline{\Omega} &= \{y \ : \ Q \in \Omega^+\}. \end{aligned}$$

To take into account the restrictions in (4) we use a penalty function in the maximum form. To define penalty coefficient N the following theorem is used [9].

Let $f_i(x),\ i=1,\dots,m$, be continuous functions and $M \subset \mathbb{R}^n$.
Define

$$V(y)=\inf\{f_0(x)\ :\ f_i(x)\le y_i,\quad i=1,\dots,m;\quad x\in\mathbb{R}^n\},\tag{5}$$

where V(0) corresponds to the initial problem,

$$\Phi_N(x) = f_0(x)+NF(x),\tag{6}$$

$$F(x) = \max\{0,\ f_1(x),\dots,\ f_m(x)\}.\tag{7}$$

Further we need

Theorem 1. *Let* $\inf_{\lambda>0}\dfrac{V(\lambda)-V(0)}{\lambda}=-L>-\infty$ *and* $N>L$. *Then the solution of the problem of finding* $\inf\Phi_N(x)$ *is equal to the solution of the problem* $V(0)$.

Remark 1. Since $V(\lambda)$ is a decreasing function on λ (the lower bound is higher in a narrower region than in a wider region) then $L\ge 0$.
Remark 2. If $f_i(x),\ i=1,\dots,m$ are convex functions and the Slaiter conditions are fulfilled then $L=\sum_{i=1}^m u_i^*$ where u_i^* is the vector component of optimal Lagrange multipliers obtained from the Kuhn-Tucker theorem. If $f_0(x)$ is strictly convex then N can be equal to L.

The inequality $\sqrt{c_i^t\,Q\,c_i}+(c_i,b_*)-1\le\lambda$, is equivalent to $\sqrt{c_i^t\,\widehat{Q}\,c_i}+$ $+\left(c_i,\widehat{b}_*\right)-1\le 0$, where $\widehat{Q}=Q/(1+\lambda)^2$, $\widehat{b}_*=b_*/(1+\lambda)$. Then for $\lambda>-1$

$$V(\lambda)-V(0)=-\ln(1+\lambda)^{2n}=-2n\ln(1+\lambda),$$

$$\inf_{\lambda>0}\frac{V(\lambda)-V(0)}{\lambda}=\inf_{\lambda>0}\left(-2n\frac{\ln(1+n)}{\lambda}\right)=-2n>-\infty.$$

Thus we have proved

Theorem 2. *If* $N>2n$ *then the solution of the problem of finding*

$$\inf\{\Phi_N(y)\ :\ y\in\overline{\Omega}\}$$

is the solution of problem (4).

Finally, using Theorem 2 and variable substitution $P = \{p_{ij}\}_{i,j=1}^{n}$, $Q = P^t P$, problem (4) is reduced to the problem of minimization of the nonsmooth penalty function

$$\Phi_N^1(P, b_*) = -2 \ln \det P + N \max\{0, \max_i(\|Pc_i\| + (c_i, b_*) - 1)\} \qquad (8)$$

under the conditions $P \in \Omega^+$ and $N > 2n$. Note that N can equal $2n$ as it follows from remark 2.

2.2. The problem of constructing an ellipsoid of minimal volume containing a finite set of points

Let M be a convex polyhedron represented as the convex hull of a finite number of points $\{b_i\}_{i=1}^m$ in the space $\mathbb{R}^n$. Then the problem of constructing E^* may be modified to

$$\min\{f_0(y) \ : \ E \in O_1(M)\}, \qquad (9)$$

where

$$\left.\begin{aligned}
& f_0(y) = -\ln \det K, \\
& y = \{\{k_{ij}, b_*\}_{i,j=1}^n, b_*\} \\
& E \in O_1(M) \text{ is expressed by a system of inequalities} \\
& (b_i - b_*)^t K (b_i - b_*) \le 1, \quad i = 1, \ldots, m.
\end{aligned}\right. \qquad (9')$$

Consider the problem of finding an optimal $(n+1)$-dimensional ellipsoid $\overline{E} = \left(\overline{K}, 0\right)$ (the ellipsoid centre is fixed at the origin) containing a set of points $\{\overline{b}_1, \ldots, \overline{b}_m\}$ where $\overline{b}_i^{\,t} = (b_i^t, 1)$ and $b_i \in M$:

$$V = \min\{f_0(y) | f_i(y) \le 0, \quad i = 1, \ldots, m, \quad y \in \overline{\Omega}\}, \qquad (10)$$

where

$$\begin{aligned}
& f_0(y) = -\ln \det \overline{K}, \\
& f_i(y) = \overline{b}_i^{\,t}\, \overline{K}\, \overline{b}_i - 1, \\
& y = \{\{\overline{k}_{ij}\}_{i,j=1}^{n+1}\}, \quad \overline{K} = \{\overline{k}_{ij}\}_{i,j=1}^{n+1}, \\
& \overline{\Omega} = \{y \ : \ \overline{K} \in \Omega^+\}.
\end{aligned}$$

Using the solution of problem (10) one may easily calculate the parameters of the n-dimensional elllipsoid $E^* = (K, b_*)$ from problem (9), (9′) by formulas (1), (2). This is possible because E^* is obtained by the intersection of $\overline{E}$ and the hyperplane $x_{n+1} = 1$ (Corollary 1 of Lemma 1 for $h = 1$).

To solve problem (10) one uses the penalty function in maximum form (6), (7) and defines the penalty coefficient N by using Theorem 1. Similarly to Theorem 2 we prove

Theorem 3. *If $N > n+1$ then the solution of the problem of finding* $\inf\{\Phi_N(y) \ : \ y \in \overline{\Omega}\}$ *is equal to the solution of problem* (10).

So the problem of constructing a minimal ellipsoid containing a finite set of points is reduced to that of minimization of the nonsmooth penalty function

$$\overline{\Phi}_N^2(\overline{K}) = -\ln\det\overline{K} + N\max\{0,\ \max_i(\overline{b}_i^t\,\overline{K}\,\overline{b}_i - 1)\} \tag{11}$$

provided $\overline{K} \in \Omega^+$ and $N > n+1$.

As Ω^+ is a convex set and $\Phi_N^1(P, b_*)$ and $\Phi_N^2(\overline{K})$ are convex nonsmooth functions, (8) and (11) are convex programming problems. Taking into account the constraints of type $Z \in \Omega^+$ ($Z = P$ for (8) and $Z = \overline{K}$ for (11)) is essentially facilitated due to $-\ln\det Z \to +\infty$ when Z approaches the boundary of Ω^+. The above feature creates a natural «barrier» to prevent the exit of matrix Z out of the region when the monotonic minimization method is used.

The dimension of problem (11) of finding the minimal ellipsoid is a function of the space dimension $(n+1)(n+2)/2$. When the quantity m of constraints is less than this value it is expedient to use another problem, the solution of which is the solution of problem (9), (9′) and which is the problem of minimization of a nonsmooth function of m variables.

Theorem 4. *The solution of the problem of finding the minimal ellipsoid $(K, 0)$ containing the set $\{b_i,\, i = 1, \ldots, m\}$ with a fixed centre at the origin is equal to the solution of the following problem:*

$$f(p_*) = \min\{f(p) \ : \ p \in P\}, \tag{12}$$

where $f(p) = -\ln\det(BWB^t)$, B is an $n \times m$ matrix the columns of which are vectors $b_i \in I\!R^n$, $i = 1, \ldots, m$, $W = \operatorname{diag}[p]$ is a diagonal matrix with $w_{ii} = p_i$, $i = 1, \ldots, m$, $P = \{p \ : \ p \geq 0,\ \sum_{i=1}^m p_i = 1\} \subset I\!R^m$. Under this $K = (BW_*B^t)^{-1}/n$.

Proof. When $b_* = 0$, the Lagrange function of problem (9), (9′) is

$$
\begin{aligned}
L(K,u) &= -\ln\det K + \sum_{i=1}^{m} u_i \left(b_i^t\, K\, b_i - 1\right). \\
\partial L(K,u)/\partial k_{lj} &= -q_{lj} + \sum_{i=1}^{m} u_i b_i^l\, b_i^j = 0, \quad (K^{-1} = \{q_{lj}\}_{l,j=1}^{n}).
\end{aligned}
$$

Therefore,

$$K^{-1} = B \cup B^t, \quad \text{where } U = \operatorname{diag}[u]. \tag{13}$$

Moreover, from Remark 2 of theorem 1 we have

$$\sum_{i=1}^{m} u_i = n. \tag{14}$$

Since (9), (9′) is a convex programming problem, its solution is equal to the solution of a dual problem which may be written considering (13), (14) and a complementary condition as follows:

$$\max_u \Big\{ \ln\det(B \cup B^t) \ : \ U = \operatorname{diag}[u], \quad \sum_{i=1}^{m} u_i = n, \quad u \geq 0, \quad u \in I\!R^n \Big\}.$$

By substituting variables $p = u/n$ we obtain problem (12). The theorem is proved.

Theorem 5. *The solution of problem* (9), (9′) *is equal to the solution of the problem*

$$f_0(p) = \min\{f_0(p) \mid p \in P\}, \tag{15}$$

where

$$f_0(p) = -\ln\det(\overline{B} W \overline{B}_i^{\,t}),$$

$\overline{B}_i^{\,t} = (1_m^t | B^t)$, *$B$ is a matrix composed of the vectors b_i as columns,*

$$W = \operatorname{diag}[p],$$

$$P = \{p \mid p \geq 0, \quad \sum_{i=1}^{m} p_i = 1, \quad p \in I\!R^m\}$$

And $b_* = Bp_*$, $K = (\widetilde{B}_* W \widetilde{B}_*)^{-1}/n$, where $\widetilde{B}$ is an $n \times m$ matrix composed of vectors $(b_i - b_*)$ as columns, $W_* = \operatorname{diag}[p_*]$.

Proof. As mentioned above, for solving problem (9), (9′) it is sufficient to solve problem (10) and calculate the parameters of the desirable ellipsoid by formulas (1), (2). By Theorem 4 the solutions of problems (10) and (15) coincide and

$$K^{-1} = (n+1)\overline{B}W\overline{B}^t = (n+1)\left(\begin{array}{c|c} BWB^t & \sum\limits_{i=1}^m p_i b_i \\ \hline \sum\limits_{i=1}^m p_i b_i^t & 1 \end{array}\right). \qquad (16)$$

Using formulas (1), (2) for transforming $\overline{E} = \left(\overline{K}, 0\right) \in I\!R^{n+1}$ to $E = (K, b_*) \in I\!R^n$ it is easy to receive from (16):

$$\begin{aligned} K^{-1} &= n\left(BWB^t - b_*\, b_*^t\right) = n\widetilde{B}_* W \widetilde{B}_*^t, \\ b_* &= \sum_{i=1}^m p_i^*\, b_i. \end{aligned}$$

The theorem is proved.

To solve problem (16), we use the penalty function in form (6), (7). We also take into account that the solution of problem (16) does not change if one takes the restriction $\sum\limits_{i=1}^m p_i = 1$ in the form $\sum\limits_{i=1}^m p_i \le 1$. Similarly to Theorem 2 we can prove

Theorem 6. *If $N > n+1$ then the solution of the problem of finding $\Phi_N(p)$ at $p > 0$ is equal to the solution of problem* (15).

Thus we have obtained another problem the solution of which allows us to readily calculate (by formulas (1), (2)) the parameters of the minimal circumscribed ellipsoid. The latter problem is one of minimizing the nonsmooth convex penalty function

$$\Phi_N^3(p) = -\ln\det(\overline{B}W\overline{B}^t) + N\max\{0, \sum_{i=1}^m p_i - 1\} \qquad (17)$$

provided $p \ge 0$ and $N > n+1$.

To minimize the functions of type (8), (11), (17) we used modifications of an r-algorithm which are monotonic and rapidly converge [12]. The realization of one iteration of these algorithms requires $O\left(n^3\right)$ or $O\left(m^3\right)$ arithmetical operations (depending on whether the prime or dual algorithm is used) [16]. Below we propose another approach to construct optimal with respect to volume ellipsoids using successive space transformations. Iterative algorithms realizing the idea of this approach are convenient when high accuracy for defining the parameters of extremal ellipsoids is not necessary. They require only $O\,(nm)$ arithmetical operations per iteration.

3. "Simple" algorithms for constructing optimal with respect to volume ellipsoids

3.1. The problem of constructing an ellipsoid of minimal volume containing a finite set of points

Consider the problem of minimizing the penalty function

$$f(K) = -\ln\det K + n\max_i\{(Ka_i, a_i) - 1\} \tag{18}$$

to which the original problem of constructing an optimal with respect to volume ellipsoid containing the set of points $\{a_i\}_{i=1}^m$ with centre fixed at point 0 is reduced. Let B be a non-degenerate $n \times n$ matrix. By substituting $a_i = Ba'_i,\ i = 1, \dots, m$, we obtain:

$$\begin{aligned} f(K) &= -\ln\det K + n\max_i\Big[(KBa'_i, Ba'_i) - 1\Big] = \\ &= -\ln\det K' + C(B) + n\max_i\Big[(K'a'_i, a'_i) - 1\Big], \end{aligned}$$

where

$$\begin{gathered} K' = B^t KB, \quad C(B) = -2\ln\det B, \quad a'_i = B^{-1}a_i, \\ f'\,(K') = f(K) - C(B) = -\ln\det K' + n\max_i\Big[(K'a'_i, a'_i) - 1\Big]. \end{gathered} \tag{19}$$

Thus, the problem of minimization (18) is reduced to that of minimizing $f'\,(K')$, which differs from the former only by the values of $\{a_i\}_{i=1}^m$.

Let $K_0 = d_0 I_n,\ d_0 > 0$. Carry out one step of a subgradient descent for function (18) from point K_0. Calculate the subgradient:

$$g_f\,(K_0) = g_f\,(d_0 I_n) = -I_n/d_0 + n a_{i^*} a_{i^*}^t, \quad a_{i^*} = \arg\max_i \|a_i\|\,.$$

Let $h_0 > 0$ be a step multiplier. Then

$$K_1 \;=\; K_0 - h_0 g_f\,(x_0) = d_0 I_n - h_0\left(I_n/d_0 + n a_{i^*} a_{i^*}^t\right) =$$

$$= \quad (d_1 + h_0/d_0)\left[I_n - nh_0 \left\|a_{i^*}\right\|^2 \xi_0\xi_0^t \;/\; (d_0 + h_0/d_0)\right] =$$

$$= \quad d_1 R_{1-\mu}\left(\xi_0\right),$$

where $\xi_0 = a_{i^*} / \left\|a_{i^*}\right\|$, $d_1 = \left(d_0 + h_0\right)/d_0$, $\mu_1 = nh_0 \left\|a_{i^*}^2\right\| / d_1$; $R_\alpha(\xi)$ is the operator of space dilation in the direction ξ with the dilation coefficient α [12].

Let $\beta_1 = 1 - \sqrt{1-\mu_1}$, $B^{(1)} = R_{1-\beta_1}\left(\xi_{i^*}\right)$, $\overline{K} = d_1 I_n$. Use of the substitution $B^{(1)}a_i = a_i^{(1)}$, $i = 1, \ldots, m$, results in a new family of points $\{a_i^{(1)}\}_{i=1}^m$. When this substitution takes place, (18) reduces to the problem of minimizing the function

$$f^{(1)}(\overline{K}) = -\ln\det\overline{K} + n\max_i\left[\left(\overline{K}a_i^{(1)}, a_i^{(1)}\right) - 1\right] \tag{20}$$

and the approximation K_1 for (18) passes to the approximation $\overline{K}_1$ for (20). Herewith, the following correlation takes place:

$$K_1 = B^{(1)}K^{(1)}B^{(1)t}.$$

Let N iterations similar to the iterations of the subgradient descent be made but each time for a new function

$$f^{(r)}(\overline{K}) = -\ln\det\overline{K} + n\max_i\left[\left(\overline{K}a_i^{(r)}, a_i^{(r)}\right) - 1\right],$$

where $a_i^{(r)} = R_{1-\beta_r}\left(\xi_{r-1}\right) a_i^{(r-1)}$,

$$\xi_{r-1} = a_{i^*(r-1)}^{(r-1)} \Big/ \left\|a_{i^*(r-1)}^{(r-1)}\right\|, \quad \left\|a_{i^*(r-1)}^{(r-1)}\right\| = \min_i \left\|a_i^{(r-1)}\right\|,$$

β_r is a positive constant, $\beta_r < 1$; the step of the subgradient descent is carried out from the point $\overline{K}_r = d_r I_n$, $d_r = d_{r-1} + h_{r-1} \,/\, d_{r-1}$, $h_{r-1} > 0$ is a step multiplier, $r = 1, \ldots, N$. After N steps we have:

$$\overline{K}_N \;=\; d_N I_N, \qquad \{a_i^{(N)}\}_{i=1}^m = \left\{B^{(N)}a_i\right\}_{i=1}^m,$$
$$B_N \;=\; R_{1-\beta_N}\left(\xi_{N-1}\right)\ldots R_{1-\beta_1}\left(\xi_0\right).$$

As

$$K \;=\; B^{(N)}K^{(N)}B^{(N)t}. \tag{21}$$

$\overline{K}_N = d_N I_N$ corresponds to the point $K_N = d_N B_N B_N^t$ of the original space.

Lemma 2. *Let* $\overline{K}_N(d) = dB_N B_N^t$. *Then* $\min_{d>0} f(\overline{K}_N(d))$ *is obtained at* $d = 1/\|a_{i^*(N)}^{(N)}\|$, $i^*(N) = \arg\min_i \|a_i^{(N)}\|$.

Proof.

$$f\left(\overline{K}_N(d)\right) = -n\ln(d) - 2\sum_{i=1}^{N} \ln\left(1-\beta_i\right) + n\max_i\{d\left\|B_N a_i\right\|^2 - 1\}.$$

This function is convex and reaches its minimum at the stationary point:

$$-n/d + n\max\left\|B_N a_i\right\|^2 = 0, \quad d = 1/\left\|B_N a_i\right\|^2 = 1/\|a_i^{(N)}\|^2.$$

The lemma is proved.

Based on the above considerations, it is natural to use the following modification of the subgradient process with a variable metric to minimize the penalty function $f(K)$ [7]. In general, the problem of constructing an optimal with respect to volume ellipsoid containing the set of points $S = \{a_i\}_{i=1}^m \in I\!E^n$ with tne centre fixed at point 0 is represented as follows: a sphere $\overline{S}_0$ of minimal radius is chosen with its centre at point 0 and including S as the initial approximation. Then the space $I\!E^n$ is pressed in the direction of the maximal norm point from S; in the transformed space, the set $S^{(1)}$ corresponds to the set S and one chooses sphere $\overline{S}_1$ of the minimal radius with the centre 0 and including $S^{(1)}$, etc. The sequence of spheres in the transformed spaces corresponds to the sequence of ellipsoids $E_0, E_1, \dots, E_N, \dots$ containing S in the origin space. Under certain conditions the sequence E_N converges to the optimal circumscribed ellipsoid.

Formally the algorithm from [7] is written as follows:

0-step.

$B_0 = I_n$ (unit matrix $n \times n$), $\{a_i^{(0)}\}_{i=1}^m = \{a_i\}_{i=1}^m$, $r_0 = \max_i \|a_i\|$.

After N iterations we have the resultant matrix B_N of the space transformation, $\{a_i^{(N)}\}_{i=1}^m = \{B_N a_i\}_{i=1}^m$ and $r_N = \max_i \|a_i^{(N)}\| = \|a_{i^*(N)}^{(N)}\|$.

(N+1)-step. Calculate:

$$\text{1.} \qquad \xi_{N+1} = a_{i^*(N)}^{(N)} \,/\, \|a_{i^*(N)}^{(N)}\|. \tag{22}$$

$$\alpha_i^{(N+1)} = R_{1-\beta}\left(\xi_{N+1}\right) a_i^{(N)}, \quad i = 1, \dots, m. \tag{23}$$

$$B_{N+1} = R_{1-\beta}\left(\xi_{N+1}\right) B_N. \tag{24}$$

$$\text{2.} \qquad r_{N+1} = \max_{i=\overline{1,m}} \|a_i^{(N+1)}\| = \|a_{i^*(N+1)}^{(N+1)}\|. \tag{25}$$

Theorem 7. *Let* $K_N = B_N^t B_N / r_N^2$, *and* $\mathrm{conv}\{0, \{\mathrm{a_i}\}_{\mathrm{i=1}}^{\mathrm{m}}\}$ *be a convex body of dimension* n. *If:*

(1) $\quad 0 < \beta_N < 1, \quad N = 1, 2, \dots.$

(2) $\quad \sum_{N=1}^{\infty} \beta_N = +\infty, \quad \beta_N \to 0 \;$ *when* $\; N \to +\infty,$

then $\lim_{N\to+\infty} K_N = K^*$, *where* K^* *is the matrix of a unique optimal ellipsoid corresponding to the minimum of the penalty function* $f(K)$ (18).

Divide the proof into stages. First, consider the lemma proving the boundedness of the sequence $\{K_N\}_{N=0}^{\infty}$.

Lemma 3. *The sequence* $\{c_N\}_{N=1}^{\infty}$ *of the ratios of large axis lengths to small axis lengths for the ellipsoids*

$$E_N = \{x \in \mathbb{E}^n \; : \; \|B_N x\| \le 1\}$$

is limited.

Proof. B_N can be represented as $B_N = S_N O_N$ where S_N is the symmetric positive definite matrix, O_N is the orthogonal matrix; $\lambda_1^{(N)}, \dots, \lambda_n^{(N)}$ are the eigenvalues of matrix S_N which are written in non-increasing order with regard to the multiplicity, $e_1^{(N)}, \dots, e_n^{(N)}$ are eigenvectors corresponding to them which form the orthonormal system: $\nu_j^{(N)} = Oe_j, \; j = 1, \dots, n.$

The lemma statement means the boundedness of the sequence

$$\{c_N\}_{N=1}^{\infty} = \{\lambda_1^{(N)} / \lambda_n^{(N)}\}_{N=1}^{\infty}$$

that is equivalent to the boundedness of the sequence ratios

$$\{q_N\}_{N=1}^{\infty} = \left\{ \left(\prod_{j=1}^{n} \lambda_j^{(N)} \right)^{1/n} \Big/ \; \lambda_n^{(N)} \right\}.$$

Note that $B_{N+1} = R_{1-\beta_N}(\xi_N) B_N$, i.e. $\prod_{j=1}^{n} \lambda_j^{(N+1)} = \prod_{j=1}^{n} \lambda_j^{(N)} (1 - \beta_N)$. The proof of the lemma is reduced to the proof of the following statement: if $\lambda_1^{(N)}/\lambda_n^{(N)}$ is sufficiently large then $\lambda_n^{(N+1)} > \lambda_n^{(N)} (1 - \beta_N)^{1/n}$, i.e. $q_{N+1} < q_N$.

Consider the decomposition of vectors $a_i,\ i=1,\dots,m$, by the eigenvectors of matrix S_N:

$$a_i=\sum_{j=1}^{n}\alpha_{ij}e_j,\quad i=1,\dots,m,$$

$$\sum_{j=1}^{n}\alpha_{ij}^2\le R_0^2,$$

where $\quad R_0=\max\limits_i\|a_i\|=\|a_{i^*}\|,\ a_i^{(N)}=B_Na_i=\sum\limits_{j=1}^{n}\lambda_i^{(N+1)}\alpha_{ij}\,\nu_j^{(N)}.$

Consider the vector $z:\ \|z\|=R_0,\quad z=\sum_{j=1}^{n}\beta_je_j,\quad \sum_{j=1}^{n}\beta_j^2=R_0^2.$ $B_Nz=\sum_{j=1}^{n}\lambda_j\beta_j\nu_j^{(N)}$. If $z=R_0e_n$ then $\|Bz\|=\lambda_n^{(N)}R_0$ for any z, $\|z\|=R_0,\quad \|Bz\|\ge\lambda_n^{(N)}R_0.\quad Z_\varepsilon^{(N)}=\{z:\|z\|=R_0,\ \|Bz\|\le$ $\le\lambda_n^{(N)}R_0(1+\varepsilon),\ \varepsilon>0\}$.

Let $\max\limits_i|\alpha_{i1}|=\overline{\alpha}$. Since $\operatorname{conv}\{0,\{\alpha_i\}_{i=1}^m\}$ has interior points, $\overline{\alpha}>0$, $\|a_i^{(N)}\|\ge\lambda_i^{(N)}\alpha$, from which $\max_i\|a_i^{(N)}\|=\|a_{i*}^{(N)}\|\ge\lambda_1^{(N)}\overline{\alpha}$. Let $z\in Z_\varepsilon^{(N)}$ and its decomposition is $z=\sum_{j=1}^{n}\beta_je_j$. Then $B_Nz=\sum\limits_{j=1}^{n}\lambda_j^{(N)}\beta_j\nu_j^{(N)}$. From this

$$\sum_{j=1}^{n}\big(\lambda_j^{(N)}\big)^2\beta_j^2\le\big(\lambda_n^{(N)}\big)R_0^2(1+\varepsilon)^2. \tag{26}$$

Let $J=\{j\ :\ \lambda_j^{(N)}/\lambda_n^{(N)}\ge2,\ j=1,\dots,n\}$. Rewrite (26) in the form of

$$\sum_{j=1}^{n}\Big[\big(\lambda_j^{(N)}\big)^2-\big(\lambda_n^{(N)}\big)^2\Big]\beta_j^2\le\big(\lambda_n^{(N)}\big)^2R_0^2\,\varepsilon(2+\varepsilon).$$

Hence

$$\sum_{j\in J}(3/4)\big(\lambda_j^{(N)}\big)^2\beta_j^2\le\ \varepsilon(2+\varepsilon)\big(\lambda_n^{(N)}\big)^2R_0^2,$$

$$\sum_{j\in J}\big(\lambda_j^{(N)}\big)^2\beta_j^2\le(4/3)\varepsilon(2+\varepsilon)\big(\lambda_n^{(N)}\big)^2R_0^2.$$

Estimate $|\cos(\alpha^{(N)}_{i^*(N)}, Bz)|$ when $z \in Z^{(N)}_\varepsilon$.

$$
\begin{aligned}
&\left|\cos^2\left(\alpha^{(N)}_{i^*(N)}, Bz\right)\right| \le \left|\sum_{j=1}^{n} \lambda_j \alpha_{i^*(N)} \lambda_j \beta_j\right| \,/\, \left(\lambda_1 \,|\overline{\alpha}|\, \lambda_n R_0\right) \le \\
\le\ & \Big(\big|\sum_{j\in J} \lambda_j \alpha_{i^*(N)} \lambda_j \beta_j\big| + \big|\sum_{j\notin J} \lambda_j \alpha_{i^*(N)} \lambda_j \beta_j\big|\Big) \,/\, (\lambda_1 \,|\overline{\alpha}|\, \lambda_n R_0) \le \\
\le\ & \lambda_1^{(N)} R_0 \lambda_n^{(N)} \sqrt{(4/3)\,\varepsilon(2+\varepsilon)} R_0 + 4(\lambda_n^{(N)})^2 R_0^2 \le \\
\le\ & c_1\sqrt{\varepsilon} + c_2\big(\lambda_n^{(N)}/\lambda_1^{(N)}\big).
\end{aligned} \tag{27}
$$

Let $x \in I\!\!E^n$. $\qquad \|R_{1-\beta}(x)\|^2 = \|x\|^2 - (2\beta - \beta^2)(x,\xi)^2,$

$$
\big(\|R_{1-\beta}(x)\|/\|x\|\big)^2 = 1 - (2\beta - \beta^2)\cos^2(\widehat{x,\xi}). \tag{28}
$$

Let $\lambda_n^{(N+1)}$ be the minimal eigenvalue of matrix S_{N+1} provided that S_{N+1} satisfies the condition $B_{N+1} = S_{N+1}O_{N+1}$ where O_{N+1} is an orthogonal matrix.

$$
B_{N+1} = R_{1-\beta_n}(\xi_N)\, B_N, \quad \xi_N = a^{(N)}_{i^*(N)} \,/\, a^{(N)}_{i^*(N)}. \tag{29}
$$

By virtue of ratio (29) $\min \|B_{N+1}z\|$ is obtained at vector y which satisfies $B_N\, y \in Z^{(N)}_{\beta_N}$. In the estimation (29) take $\beta_N = \varepsilon$. In view of (28) the following estimate is valid:

$$
\begin{aligned}
&\left(\prod_{j=1}^{n} \lambda_j^{(N+1)}\right)^{1/n} \Big/\ \lambda_n^{(N+1)} = \\
&\qquad = (1-\beta_N)^{1/n} \prod_{j=1}^{n} \lambda_j^{(N)} \Big/ \left(\lambda_n^{(N)}\big[1 - (2\beta - \beta^2)\cos^2(\alpha^{(N)}_{i^*} B_N\, y)\big]\right).
\end{aligned}
$$

Taking into account (27) it is easy to see that the expression in square brackets becomes larger than $(1-\beta_N)^{1/n}$ provided that β_N and $\lambda_n^{(N)}/\lambda_1^{(N)}$ are sufficiently small, i.e.

$$
\prod_{j=1}^{n} \lambda_j^{(N+1)} \Big/ \lambda_n^{(N+1)} \le \prod_{j=1}^{n} \lambda_j^{(N)} \Big/ \lambda_n^{(N)}.
$$

Thus, the sequence $\big\{\prod_{j=1}^{n} \lambda_j^{(N)}\big\}_{N=1}^{\infty}$ cannot unlimitedly increase, which completes the proof.

Note that the sequence of small axes of ellipsoids generated via (22)–(25) $E_N = \{x : (K_N x, x) \leq 1\}$ is upper bounded, because at least one point of the family $\{a_i\}_{i=1}^m$ is on the ellipsoid surface. It follows from the proved lemma that the sequence of volumes $\{V(E_N)\}_{N=1}^{\infty}$ is upper bounded.

Now to proceed to the proof of the theorem.

The proof will be fulfilled by means of contradiction. As the sequences of matrices $\{K_N\}_{N=1}^{\infty}$ and $\{K_N^{-1}\}_{N=1}^{\infty}$ are bounded they have cluster points. In the case of convergence of the whole process to the solution $K_N \to K^*$, $K_N^{-1} \to (K^*)^{-1}$. If it is not so, then fix some cluster point $\overline{K} \neq K^*$. For any $\delta > 0$ and $\overline{\beta} > 0$ one can choose sufficiently large $\overline{N}$ in the sequence $\{K_N\}_{N=1}^{\infty}$ so that

$$\begin{aligned} &\text{a)} \quad \beta_N < \overline{\beta} \quad \text{when} \quad N > \overline{N} \quad (\text{as} \quad \beta_N \to 0); \\ &\text{b)} \quad \|K_N - \overline{K}\| < \delta. \end{aligned} \tag{30}$$

By substitution:

$$K = B_{\overline{N}} R\, B_{\overline{N}}^T \tag{31}$$

one obtains the following problem for matrix variables R

$$\min_{R>0} f(R) \quad \text{where} \quad f(R) = -\ln\det R + n \max_{i \in \overline{1,m}} \left\{ \left(R a_i^{(N)}, a_i^{(N)} \right) - 1 \right\}.$$

Mark: $R_0 = I_n$, $\{b_i\}_{i=1}^m = \{a^{(\overline{N})}\}_{i=1}^m$, $\overline{\beta}_k = \beta_{k+N}$, $k = 1, 2, \ldots$. Use the above subgradient process with variable metric to minimize $f(R)$. Step multipliers are chosen such that the approximation R_k on the k-step is multiple of the matrix $R_{\left(1-\overline{\beta}_k\right)^2}(\xi_k)$.

Let $\eta_1^{(k)}$ be the direction of subgradient descent at the point R_k and $\eta_2^{(k)}$ be the direction of the movement corresponding to subgradient descent at a respective point of the space with the variable metric at the linear transformation

$$R_k = \overline{B}_k \overline{R}_k \overline{B}_k^t, \tag{32}$$

where $\overline{B}_k = \overline{R}_{\beta_k}(\xi_k) \ldots \overline{R}_{\beta_1}(\xi_k)$.

Let $Q_l = \prod_{i=1}^l R_{1-\beta_i}(\xi_i)$. $\det Q_l = \prod_{i=1}^l (1-\beta_i)$ and as $\lambda_{\max}(Q_l) \leq 1$: $\rho(Q_l) = \lambda_{\min}(Q_l)/\lambda_{\max}(Q_l) \geq \prod_{i=1}^l (1-\beta_i) > 1 - \sum_{i=1}^l \beta_i$.

The following inequalities are correct for symmetric positive definite matrices A [11]:

$$\cos(Ax, x) = (Ax, x)/\|Ax\|\,\|x\| \geq 2\sqrt{\rho(A)}/(1+\rho(A)) \tag{33}$$

$$\sin(Ax,x) \quad \leq \quad \big(1-\rho(A)\big)/\big(1+\rho(A)\big). \tag{34}$$

Use the formula for calculating a subgradient when linear transformations of arguments take place [11]. Then

$$\sin\big(\eta_1^{(l)},\eta_2^{(l)}\big) = \sin\{\eta_1^{(l)}, Q_l\, Q_l^t\, \widehat{\eta}_1^{(l)} Q_l^t\, Q_l\} \leq$$
$$\leq \big[1-\big(1-\sum_{i=1}^{l}\overline{\beta}_i\big)^4\big] \;/\; \big[1+\big(1-\sum_{i=1}^{l}\overline{\beta}_i\big)^4\big] \leq\; 4\sum_{i=1}^{l}\overline{\beta}_i \tag{35}$$

when $\sum_{i=1}^{l}\overline{\beta}_i$ sufficiently small.

Further consider the following construction in the space of variables R. Let $\overline{R}$ be the point corresponding to the cluster point $\overline{K}$ in the space K, $R^* = \arg\min_{R>0} f(R)$. Let $S(a,r)$ be a sphere of the radius r with centre a in the space R. Consider

$$S_0 = S\left(I_n, \|I_n - R^*\|/2\right),$$

$$R(1) = \arg\min_{R\in S_0, R>0} f(R),$$

$$R(\lambda) = (1-\lambda)\, I_n + \lambda R(1), \quad 0 \leq \lambda \leq 1.$$

Let $f(R_0) - f(R^*) = 2\Delta > 0$. As the function $f(R)$ is convex $f\big(R(0)\big) - f\big(R(\lambda)\big) \geq \Delta\lambda$.

Consider $P(\lambda) = M_\lambda \bigcap S_\lambda$, where

$$M_\lambda = \big\{R \;:\; f(R) \leq f(R_0) - \Delta\lambda/2\big\},$$

$$S_\lambda = S\big(R(\lambda),\, \lambda\big\|R(0) - R(1)\big\|\big);$$

$$\overline{S}_\lambda = S\big(K(\lambda),\, \Delta\lambda/(2c)\big),$$

where $c = \max\|g_f(R)\|;\ R \in S\big(R(1),\, \big\|R(1) - R(0)\big\|\big)$. It is easy to see that $\overline{S}_\lambda \subseteq P(\lambda)$, and if $R \in S_\lambda$ and $R \notin P(\lambda)$ then

$$\cos\left(g_f(R),\, R - R(\lambda)\right) \geq \Delta\lambda \;/\; \big[2c\big\|R(0) - R(\lambda)\big\|\big] \geq 2t \geq 0, \tag{36}$$

where t is a positive constant independent of λ.

By virtue of estimation (35), one can choose the constant $L > 0$ such that if $\sum_{i=1}^{l}\overline{\beta}_i \leq L$ then the angle between the direction of the anti-subgradient at points R_i, $i = 1,\ldots,l$, and the direction of movement by algorithm (22)–(25) is not larger than the value $\varphi(L)$ dependent only on L. And when $L \to 0$ $\varphi(L) \to 0$. By virtue of inequalities (36), if points

$R_1, \ldots, R_l$ belong to S_λ and do not belong to P_λ then when L is sufficiently small

$$\cos\bigl(R_{i+1} - R_i,\, R_i - R(\lambda)\bigr) \geq t, \quad i = 1, \ldots, l-1. \tag{37}$$

Let $R_i \in S_\lambda,\ R_i \notin P(\lambda)$ and (37) are correct. Then

$$\|R_{i+1} - R(\lambda)\|^2 = \|R_i - R(\lambda) + \bigl(R_{i+1} - R_i\bigr)\|^2 \leq$$

$$\leq \|R_i - R(\lambda)\|^2 - \|R_{i+1} - R_i\| \left(2t \left\|R_i - R(\lambda)\right\| - \|R_i - R_{i+1}\|\right).$$

If $\qquad \|R_i - R_{i+1}\|^2 \leq t \left\|R_i - R(\lambda)\right\| \qquad$ then

$$\|R_{i+1} - R(\lambda)\|^2 \leq \|R_i - R(\lambda)\|^2 - t\,\|R_i - R_{i+1}\|\,\|R_i - R(\lambda)\|.$$

Let for $i = 1, \ldots, l, \qquad R_i \notin \overline{S}(\lambda)$. Then

$$0 \leq \|R_{i+1} - R(\lambda)\|^2 \leq \lambda^2 \|K(0) - K(1)\|^2 -$$

$$-\bigl[t\,\Delta\lambda/(2c)\bigr] \sum_{i=1}^{l} \|R_{i+1} - R_i\| =$$

$$= \lambda\bigl(\lambda\|R(0) - R(1)\|^2 - \bigl[t\,\Delta/(2c)\bigr] \sum_{i=1}^{l} \|R_{i+1} - R_i\|\bigr).$$

When L is fixed and λ is sufficiently small the latter expression is negative. This contradiction is due to the supposition that all $R_i,\ i = 1, \ldots, l$, belong to $\overline{S}_\lambda$. Thus one can find (when $\overline{\beta}$ and λ are sufficiently small) $i \leq l$ such that $R_i \notin \overline{S}(\lambda)$, i.e. $f(R) \leq f\,(R_0) - \lambda\Delta/2$. As $K_{\overline{N}}$ can be close to the cluster point $\overline{K}$ as possible (at definite $\overline{N}$) we showed that if $\overline{K} \neq K^*$ then there are cluster points with a significantly smaller value of the penalty function f, i.e. there is a subsequence $\{K_{N_t}\}_{t=1}^{\infty},\ N_{t+1} > N_t$, that converges to K^*. On the other hand as $\beta_n \to 0$ then it is easy to prove that although the sequence $\{f(K_N)\}_{N=0}^{\infty}$ is not monotonic in general

$$\sup_r \bigl[f\bigl(K_{\overline{N}+r}\bigr) - f\bigl(K_{\overline{N}}\bigr)\bigr] \xrightarrow{\overline{N} \to \infty} 0.$$

The consequence of this is the convergence $f\,(K_N) \to f\,(K^*)$ and, as K^* is unique, that $K_N \to K^*$. The theorem is proved.

3.2. The problem of constructing an ellipsoid of the maximal volume inscribed in a polyhedron

Let polyhedron M_n be defined by the finite system of linear inequalities which, without loss of generality, if $0 \in \text{int} M_n$, are represented as $(c_i, x) \leq 1$, $i = 1, \ldots, m$, $x \in I\!R^n$. Similar to the case of constructing an optimal circumscribed ellipsoid, consider the problem in the $(n+1)$-dimensional space (Corollary 1 of Lemma 1):

a) the condition $E_n \subseteq M_n$ represents the intersection of the hyperplane $(\widetilde{c}_0, x) \leq 1$ where $\widetilde{c}_0 = (0_n : 1)$ $(0_n \in I\!R_n)$ and of
$M_{n+1} = \{x \ : \ (\widetilde{c}_i, x) \leq 1, \ \widetilde{c}_i = (c_i, 0)\,, \ i = \overline{1,m}, \ x \in I\!R^{n+1}\}$;

b) define the hyperplane $(d, x) \leq 1$, where $d = h_* \widetilde{c}_0$ $(h_* \in I\!R^1, \ h_* > 1)$ that puts upper bounds to the optimal ellipsoid.

The proposed algorithm is analogous to the above "simple" algorithm.

0-step. $P_0 = I_0, \quad \{c_i^{(0)}\}_{i=0}^m = \{\widetilde{c}_i\}_{i=0}^m$.

Let after k steps we have matrix P_k which is inverse to the matrix of space transformation, $\{c_i^{(k)}\}_{i=0}^m$.

(k + 1)-step. Calculate:

$$\begin{array}{lrcl} 1. & A_0(i) & = & (c_0^{(k)}, c_0^{(k)})(c_i^{(k)}, c_i^{(k)}) - (c_0^{(k)}, c_i^{(k)})^2, \\ & A_1(i) & = & (c_0^{(k)} - c_i^{(k)}, c_0^{(k)}) \ / \ A_0(i), \\ & A_2(i) & = & (c_i^{(k)} - c_0^{(k)}, c_i^{(k)}) \ / \ A_0(i), \end{array}$$

$r_{k+1}^2 = \min_{i=\overline{1,m}}(A_1(i) + A_2(i)) = A_1(i^*) + A_2(i^*)$ (the square of the minimal distance from the point $0_{n+1} \in I\!R^{n+1}$ to the intersection of the hyperplane $(c_0^{(k)}, x) \leq 1$ and the body $M_{n+1}^{(k)}$).

$$\begin{array}{lrcl} 2. & z_{k+1} & = & A_1 c_{i*}^{(k)} + A_2 c_0^{(k)}, \\ 3. & B & = & (z_{k+1}, c_0^{(k)}) \ / \ (c_0^{(k)}, c_0^{(k)}), \\ & \alpha'_{k+1} & = & (r_{k+1} - \overline{B})^{-1/2} \big(h_*^2 \ / \ (c_0^{(k)}, c_0^{(k)}) - B\big)^{-1}, \\ & \xi'_{k+1} & = & c_0^{(k)} \ / \ \|c_0^{(k)}\|. \\ 3'. & P'_{k+1} & = & P_k R_{1/\alpha'_{k+1}}(\xi'_{k+1}), \\ & c_i'^{(k+1)} & = & R_{1/\alpha'_{k+1}}(\xi'_{k+1}) c_i^{(k)}, \quad i = \overline{1,m}, \\ & z'_{k+1} & = & R_{\alpha'_{k+1}}(\xi'_{k+1}) z_{k+1}. \end{array}$$

Transform the space so that the sphere $(I_{n+1}/(z'_{k+1}, z'_{k+1}), 0_{n+1})$ (I_{n+1} is a unit matrix $(n+1)\times(n+1)$) of maximal volume tangents to the hyperplane $(c_0'^{(k+1)}, x'') = 1$ in the new space.

$$
\begin{array}{lrcl}
4. & \xi_{k+1} & = & z'_{k+1} \,/\, \|z'_{k+1}\|, \\
4'. & c_i^{(k+1)} & = & R_{1/\alpha_{k+1}}(\xi_{k+1}) c_i'^{(k+1)}, \quad i = \overline{0,m}, \\
 & P_{k+1} & = & P'_{k+1} R_{1/\alpha_{k+1}}(\xi_{k+1}).
\end{array}
$$

The following theorem is true. Its proof is similar to those of Theorem 7.

Theorem 8. *Let* $K_k^{-1} = P'_k P_k'^t \|z'_k\|^2$ *and* $\mathrm{conv}\{0, \{\mathrm{c_i}\}_{\mathrm{i=1}}^{\mathrm{m}}\}$ *be a convex body of dimension* n. *If:*

(1) $\quad 0 < \beta_k < 1, \quad k = 1, 2, \ldots, .$

(2) $\quad \sum\limits_{j=1}^{n} \beta_k = +\infty, \quad \beta_k \to 0 \quad \textit{when} \quad k \to +\infty,$

where $\beta_k = \alpha_k - 1$ *then* $\lim_{k\to+\infty} K_k = K^*$, *where* K^* *is the matrix of unique optimal ellipsoid in the space* $I\!R^{n+1}$. *Further the parameters of the desirable ellipsoid in* $I\!R^n$ *are calculated by formulas* (1), (2).

In practice, it is more advantageous to use another algorithm that is based on the same ideas but considers the problem in the space R^n and does not require the ellipsoid tangents of the limiting hyperplane.

0-step. $\quad P_0 = I_0, \quad a_0 \in \mathrm{int}\, M, \quad M = \{x \,:\, (c_i, x) \le 1,\ i = \overline{1,m}\}$ (for example $a_0 = 0$), $\quad \{c_i^{(0)}\}_{i=1}^m = \{c_i\}_{i=0}^m$.

Let after k steps we have matrix P_k, which is inverse to the matrix of space transformation, $\{c_i^{(k)}\}_{i=1}^m$, a_k.

(k + 1)-step. Calculate:

$$
\begin{array}{lrcl}
1. & r_{k+1} & = & \min\limits_{i=\overline{1,m}} \rho_i^{(k+1)} = \rho_{i*}^{(k+1)}, \quad \rho_i^{(k+1)} = (-1 - c_i^{(k)}, a_k) \,/\, \|c_i^{(k)}\|. \\
2. & \gamma_i^{(k+1)} & = & r_{k+1} \,/\, \rho_i^{(k+1)} \\
3. & z_{k+1} & = & \sum\limits_{i\in I} \gamma_i^{(k+1)} c_i^{(k)}; \\
 & I & = & \{i :\ \gamma_i^{(k+1)} \ge \gamma^*, \quad i = \overline{1,m}\}, \quad \gamma^* \le 1 \quad (\gamma^* = 0.99); \\
 & a'_k & = & a_k + (\alpha_{k+1} - 1) r_{k+1} z_{k+1} \,/\, \|z_{k+1}\|.
\end{array}
$$

$$
\begin{aligned}
4. \quad \xi_{k+1} &= c_{i*}^{(k)} \;/\; \|c_{i*}^{(k)}\|. \\
4'. \quad a_{k+1} &= R_{1/\alpha_{k+1}}(\xi_{k+1})\, a'_k; \\
c_i^{(k+1)} &= R_{1/\alpha_{k+1}}(\xi_{k+1})\, c_i^{(k)}, \quad i = \overline{1,m}; \\
P_{k+1} &= P_k R_{1/\alpha_{k+1}}(\xi_{k+1}).
\end{aligned}
$$

The following theorem is true which is also proved similarly to Theorem 7.

Theorem 9. *Let* $K_k^{-1} = P_k P_k^t r_{k+1}^2$, $a_k = P_k a_k$ *and* $conv\{0, \{a_i\}_{i=1}^m\}$ *be a convex body of dimension* n. *If:*

(1) $\quad 0 < \beta_k < 1, \quad k = 1, 2, \ldots.$

(2) $\quad \sum_{k=1}^{\infty} \beta_k = +\infty, \quad \beta_k \to 0 \quad when \quad k \to +\infty,$

where $\beta_k = \alpha_k - 1$ *then* $\lim_{k\to+\infty} K_k = K^*$ *and* $\lim_{k\to+\infty} a_k = a^*$, *where* (K^*, a^*) *is the desirable optimal ellipsoid in the space* $I\!R^n$.

To test algorithm (22)–(25) problems were generated by using special stochastic procedures. The algorithm parameters were regulated in the following way; in k_0 initial iterations the space was pressed at a fixed coefficient $\alpha = 1 - \beta$, $\beta = 0.2$ and, starting from the $k_0 + 1$ iteration, α was chosen as $\alpha_{k_0+k} = 1 - \beta_{k_0+k}$, $\beta_{k_0+k} = 1/(20 + k)$, $k = 1, 2, \ldots$. Some results of these computational tests are given in Table 1, where n is the dimension of the space, m is the number of points, ε is the relative precision on the criterion function, and k is the number of iterations [7].

Table 1.

n	2	3	4	5	6	7	8	9	10
m	30	50	70	70	90	100	110	120	130
ε_k	0.001	0.001	0.001	0.01	0.01	0.01	0.01	0.01	0.01
k	200	300	300	400	400	500	600	700	700

The above proposed algorithms for constructing optimal ellipsoids may be used successfully for the approximation of convex sets when the information is not known fully *a priori* and can be added during the process of optimization. In particular, this takes place in the method of "inscribed ellipsoid" for minimizing convex functions [4]. "Simple" algorithms for constructing an optimal ellipsoid inscribed in a polyhedron were tested and gave good results for finding cuts in this method. Note that in the "inscribed ellipsoid" method a relative precision on the order of 0.01 for a volume does not influence virtually the efficiency of the algorithms on the whole [8]. Therefore, we used only 100 iterations in the interior problem for approximation of the optimal inscribed ellipsoid ($n \leq 10$). It should be very interesting to compare the efficiency of the proposed algorithms and the polynomial-time algorithms.

Bibliography

1. F. L. Chernousko, (1988). *Evaluation of Phase Conditions of Dynamic Systems (Ellipsoidal Method).* Nauka, Moscow 320 p. (in Russian).

2. B. Grunbaum, (1971). *Studies on Combinatoric Geometry and the Theory of Convex Bodies.* Nauka, Moscow 96 p. (in Russian).

3. F. F. Fedorov, (1971). *The Theory of Optimal Experiment.* Nauka, Moscow 184 p. (in Russian).

4. S. P. Tarasov, L. G. Khachian and I. I. Erlix, (1988). *The inscribed ellipsoid method. Dokl. Akad. Nauk SSSR*, v. 298 **5**, 1081–1085. (in Russian).

5. U. N. Nesterov and A. S. Nemirovskii, (1989). *Selfadjustment Functions and Polynomial Algorithms in Convex Programming.* CEMI Akad. Nauk SSSR, Moscow, 134 p. (in Russian).

6. V. L. Zaguskin, (1958) *On circumscribed and inscribed ellipsoids of extremum volume. Uspekhi Mat. Nauk. 13*, v. 13 **6**, 89–92. (in Russian).

7. N. Z. Shor and S .I Stecenko, (1990). *An algorithm with successive space pressing for constructing minimal on volume circumscribed ellipsoid.* In: Research into the Methods for Solving Extremal Problems. Inst. Kibern. Akad Nauk Ukrain. SSR, Kiev, 25–29p. (in Russian).

8. N. Z. Shor and and O. A. Berezovski, (1992) *Constructing the maximal inscribed ellipsoid for a polytope using successive space dilation.* Kibernetika i Vych. Tehnika (Naukova Dumka, Kiev). **93**. (in Russian).

9. B. N. Pshenichny, (1983). *Method of Linearization.* Nauka, Moscow, 135–136. (in Russian).

10. F. Behrend, (1937). *Über einige Affininvarianten konvexer Bereiche. Math. Ann.* **115**, 713–717.

11. F. John, (1948). *Extremum problems with inequalities as subsidiary conditions* . Studies and essays, presented to R.Courant on his 60th birthday. N.Y. 187–204.

12. N. Z. Shor, (1990). *Minimization Methods for Non–differentiable Functions.* Springer series in Computational Mathematics, Springer-Verlag, vol. 3, 162 p.

13. L. Khachiyan and M. Todd, (1990). *On the complexity of approximating the maximal inscribed ellipsoid for a polytope.* Cornell University, Ithaka, N.Y.: Technical report No. 893, 39 p.

14. D. M. Titterington, (1976). *Optimal design: some geometric aspects of D-optimality.* Biometrica, **62**, 313–320.

15. S. P. Tarasov, L. G. Khachiyan and I.I. Erlix, (1986). In: Metody i Sredstva Avtomatizatsii Proektirovanija. Moscow (in Russian).

16. N.Z. Shor and O.A. Berezovski, (1989). *Application of subgradient type method with space dilation for constructing maximal volume ellipsoid.* Kibernetika, **6**, 119–120.(in Russian).

Алгоритмы построения инвариантного эллипсоида минимального объема для устойчивой динамической системы

Н. З. Шор, О. А. Березовский
Кибернетика и системный анализ. – 1995. – № 3. – С. 130–137.

В последние годы наблюдался определенный интерес к задачам построения оптимальных по объему вписанных и описанных эллипсоидов [1]–[3], который явился предвестником сегодняшней активизации исследовательской деятельности вокруг задач матричной оптимизации (задачи математического программирования, целевую функцию или функции ограничений которых можно сформулировать в терминах матричного исчисления). Причиной повышенного к ним внимания является широкая область приложения. Экстремальные эллипсоиды используются для аппроксимации «сверху» и «снизу» сложных множеств, например областей достижимости в теории управления и дифференциальных играх, и для локализации областей возможных значений параметров динамических систем по априорным данным и результатам измерений [1]. Они также находят применение в статистике, планировании экспериментов для определения параметров регрессионных моделей [4], в вычислительной математике при описании множеств возможных решений линейных и нелинейных уравнений с возмущенными или неточно заданными параметрами, в математическом программировании для аппроксимации областей, локализующих оптимальную точку экстремальной задачи [5]–[7], и т. д.

Данная работа является продолжением исследований авторов в области построения экстремальных эллипсоидов (см., например, [8]) и посвящена следующей задаче, связанной с изучением проблем идентификации, устойчивости и управления динамическими системами.

Пусть задана некоторая динамическая система, движение которой описывается устойчивой линейной системой дифференциальных уравнений

$$\dot{x} = Ax, \quad x \in E^n. \tag{1}$$

Эллипсоид E с центром в точке 0 называется инвариантным для устойчивой системы вида (1), если при условии, что $x(t_0) \in E$, имеет место $x(t) \in E$ для всех $t \geq t_0$. Будем рассматривать эллипсоид $E \in E^n$ заданным в виде

$$E = (K, 0) = \left\{x : x^T K x \leq 1\right\},$$

где $K = \{k_{ij}\}_{i,j=1}^n$ – симметричная положительно определенная матрица $n \times n$; при этом объем эллипсоида E равен $V(e) = V_0(\det K)^{-1/2}$, где V_0 – объем единичного n-мерного шара. Тогда условие инвариантности эллипсоида $E = (K, 0)$ можно записать так: $A^T K + KA \preceq 0$ (под выражением $D \prec 0$ ($D \preceq 0$) будем понимать отрицательную (неположительную) определенность матрицы D) [9].

Сформулируем исследуемую задачу: требуется построить инвариантный эллипсоид минимального объема для устойчивой динамической системы, описанной уравнением (1), при заданном конечном наборе состояний $M = \{b_i\}_{i=1}^m \subset E^n$ в момент времени t_0. Ей можно поставить в соответствие задачу математического программирования

$$\min\{f_0(K) : M \subset E^n\}, \tag{2}$$

где $f_0(K) = -\ln\det K$, а условию $M \subset E$, которое определяет множество инвариантных эллипсоидов для системы (1), содержащих M, соответствует система неравенств

$$b_i^T K b_i^T \leq 1, \quad i = 1, \ldots, m, \tag{3}$$

$$\lambda_{\max}(W) \leq 0, \quad W = A^T K + KA \tag{4}$$

($\lambda_{\max}(W)$ – максимальное собственное число матрицы W).

Как видим, задача (2)–(4) отличается от общеизвестной задачи нахождения эллипсоида минимального объема с центром в точке 0, описанного вокруг конечного набора точек (например, [3]), наличием дополнительного ограничения (4) на отрицательную определенность матрицы Ляпунова (физический смысл ограничения (4) сводится к тому, что из любой точки эллипсоида E траектория движения должна быть направлена внутрь этого эллипсоида).

Для учета ограничений (3) и (4) рассматриваемой задачи используем негладкую штрафную функцию в форме функции максимума. В принципе ее можно вводить отдельно для каждого типа ограничений. Однако, если несколько усилить ограничение (4), представив его в виде

$$(1/\varepsilon)\lambda_{\max}(W) \leq -1, \tag{5}$$

где ε – некоторое положительное число, близкое к нулю, появляется возможность использовать одну штрафную функцию $F(K)$ как для учета ограничений (3), так и ограничения (5):

$$F(K) = \max\left\{0,\ \max_i(b_i^T K b_i - 1),\ (1/\varepsilon)\lambda_{\max}(W) + 1\right\}, \tag{6}$$

причем в этом случае легко оценить штрафной множитель. Видно, что при изменении правой части ограничений (3) и ограничения (5) на некоторое δ, оптимальное значение целевой функции меняется на $n\delta + o(\delta)$. Отсюда вытекает (см. [10]) следующая теорема.

Теорема 1. *При $N \geq n$ точки минимума задачи (2), (3), (5) и задачи: найти*

$$\inf_{K>0} \Phi(K), \tag{7}$$

где

$$\Phi(K) = -\ln\det K + N\max\left\{0,\ \max_i(b_i^T K b_i - 1),\ \frac{1}{\varepsilon}\lambda_{\max}(W) + 1\right\}, \tag{8}$$

совпадают.

Так как класс положительно определенных матриц образует выпуклое множество и функция $\Phi(K)$ (8) принадлежит классу выпуклых недифференцируемых функций, задача (7) является задачей выпуклого программирования. Отметим, что учет ограничения $K \succ 0$ существенно облегчается благодаря следующему факту: при приближении к границе положительной определенности $-\ln\det K$ стремится к $+\infty$, что создает естественный «барьер» для выхода матрицы K за пределы допустимой области при использовании монотонного метода минимизации.

Таким образом, задача определения инвариантного эллипсоида минимального объема для заданной устойчивой линейной системы дифференциальных уравнений при определенном наборе начальных условий сводится к задаче минимизации выпуклой негладкой штрафной функции размерности $n(n+1)/2$ (вектор переменных задачи составляют только элементы верхнего треугольника симметричной матрицы K). Для ее решения предлагается использовать r-алгоритм [7] как обладающий быстрой практической сходимостью и хорошо зарекомендовавший себя для решения задач выпуклой негладкой оптимизации. На каждой l-й итерации этот алгоритм требует значение субградиента, который вычисляется по формуле

$$g_l = K_l^{-1} + N \cdot V_l,$$

где

$$V_l = \begin{cases} 0, & \text{если } F(K_l) = 0, \\ b_j b_j^T, & \text{если } F(K_l) = b_j^T K_l b_j - 1, \\ Auu^T + (Auu^T)^T, & \text{если } F(K_l) = (1/\varepsilon)\lambda_{\max}(W_l) + 1. \end{cases}$$

Здесь $j = \arg\max\limits_i b_i^T K_l b_i$, u – собственный вектор матрицы W_l, соответствующий $\lambda_{\max}(W_l)$. Для удобства субградиент записан в виде матрицы $n \times n$ для задачи, когда минимизация проводится по всем элементам матрицы K без учета ее симметричности. На практике при переходе к задаче размерности $n(n+1)/2$ для диагональных элементов формула остается в силе, а для элементов верхнего треугольника умножается на два.

Ниже предлагается другой подход для решения первоначальной задачи построения минимального по объему инвариантного эллипсоида для устойчивой системы (1), использующий последовательное преобразование пространства и расширяющий класс так называемых «простых» алгоритмов ([8, 11]). Итерационные алгоритмы, реализующие идеи этого подхода, удобно использовать, когда нет необходимости в слишком большой точности определения параметров экстремальных эллипсоидов либо когда информация не задана сразу полностью, а корректируется в процессе счета.

Рассмотрим задачу минимизации штрафной функции $f(P)$:

$$f(P) = -2\ln\det P + N\max\left\{\max_i(\|Pb_i\|^2 - 1),\ (1/\varepsilon)\lambda_{\max}(W) + 1\right\}, \quad (9)$$

где $W = A^T P^T P + P^T P A$, к которой сводится исходная задача (она получается из задачи (7) путем замены переменных $K = P^T P$). Субградиент функции $f(P)$ вычисляется по формуле $g(P) = -2P^{-1} + 2NV$, где

$$V = \begin{cases} Pb_j b_j^T, & \text{если } \|Pb_j\|^2 \geq (1/\varepsilon)\lambda_{\max}(W) + 2, \\ PAuu^T + P(Auu^T)^T & \text{в противном случае.} \end{cases}$$

Здесь $j = \arg\max\limits_i \|Pb_i\|^2$ – собственный вектор матрицы W, соответствующий $\lambda_{\max}(W)$.

Пусть B – невырожденная матрица $n \times n$. Проведем преобразование фазового пространства, заданное матрицей B: $x = B^{-1}x'$. Тогда функция $f(P)$ в новом пространстве примет вид $f'(P') = -2\ln\det P' + D(B) +$

$$+ N\max\{\max_i(\|P'b_i'\|^2-1),\ (1/\varepsilon)\lambda_{\max}(W')+1\},\ \text{где}\ \ D(B)=2\ln\det B,$$

$$P'=PB^{-1},\quad b_i'=Bb_i,\quad W'=A'^TP'^TP'+P'^TP'A',\quad A'=BAB^{-1} \quad (10)$$

(поскольку $\dot{x}=Ax$ переходит в $\dot{x}'=BAB^{-1}x'$).

Таким образом, задача минимизации функции $f(P)$ (9) свелась к задаче минимизации $f'(P')$, которая отличается от предыдущей лишь значениями множества векторов $\{b_i\}_{i=1}^m$ и элементов матрицы W.

Рассмотрим процесс субградиентного спуска для функции (9), каждый шаг которого проводится в преобразованном фазовом пространстве, в котором текущее приближение искомой матрицы принимает вид единичной матрицы. Пусть после k шагов такого процесса имеем некоторое фазовое пространство, переход к которому от исходного пространства задается матрицей преобразования $B_k:\ x=B_k^{-1}x'$. Для этого пространства имеем преобразованные семейство начальных фазовых состояний $\{b_i^{(k)}\}_{i=1}^m$ и матрицу A_k (переход определяется соотношениями (10)). Отмасштабируем текущее пространство так, чтобы сфера минимального объема, описанная вокруг множества $\{b_i^{(k)}\}_{i=1}^m$, была единичного радиуса. Эта процедура сводится к операциям:

$$\widetilde{A}_k=A_k,\quad \widetilde{B}_k=B_k/r_k,\quad \widetilde{b}_i^{(k)}=b_i^{(k)}/r_k,\quad i=1,\ldots,m,$$

где $r_k=\max\limits_i\|b_i^{(k)}\|$. Итак, после k итераций имеем приближение матрицы переменных $P_k=\widetilde{B}_k$, которая в текущем фазовом пространстве равна единичной – $P_k^{(k)}=I_n$. Проведем для функции (9) один шаг субградиентного спуска из точки $P_k^{(k)}$ в этом пространстве. Вычислим субградиент в $P_k^{(k)}$:

$$g_f(P_k^{(k)})=g_f(I_n)=-I_n+NV_k,$$

где

$$V_k=\begin{cases}\widetilde{b}_i^{(k)}\widetilde{b}_j^{(k)T}, & \text{если } \|\widetilde{b}_j^{(k)}\|^2=1\ge(1/\varepsilon)\lambda_{\max}(W_k)+2,\\ \widetilde{A}_kuu^T+(\widetilde{A}_kuu^T)^T & \text{в противном случае.}\end{cases}$$

Здесь $j=\arg\max_i\|b_i^{(k)}\|$, u – собственный вектор матрицы $W_k=\widetilde{A}_k^T+\widetilde{A}_k$, соответствующий $\lambda_{\max}(W_k)$.

Пусть $h_k>0$ – шаговый множитель. Тогда

$$P_{k+1}^{(k)} = P_k^{(k)} - h_k g_f(P_k^{(k)}) = I_n - h_k(-I_n + NV_k) =$$

$$= (1+h_k)\,[I_n - Nh_kV_k] = (1+h_k)\times$$

$$\times \begin{cases} I_n - Nh_k\widetilde{b}_i^{(k)}\widetilde{b}_j^{(k)T}/(1+h_k), & \text{если } -1 \ge (1/\varepsilon)\lambda_{\max}(W_k), \\ I_n - Nh_k(\widetilde{A}_k uu^T + (\widetilde{A}_k uu^T)^T)/(1+h_k) & \text{в противном случае,} \end{cases}$$

$$= (1+h_k)\begin{cases} R_{1-\beta_k}(\xi_k), & \text{если } -1 \ge (1/\varepsilon)\lambda_{\max}(W_k), \\ I_n - \beta_k(\widetilde{A}_k uu^T + (\widetilde{A}_k uu^T)^T) & \text{в противном случае,} \end{cases}$$

где $\xi_k = \widetilde{b}_j$ (напомним, что пространство отмасштабировано так, что $\|\widetilde{b}_j\| = 1$, $\beta_k = Nh_k/(1+h_k)$, а $R_\alpha(\xi)$ – оператор растяжения пространства в направлении ξ с коэффициентом растяжения α [7]).

Зная новое приближение матрицы переменных $P_{k+1}^{(k)}$ в пространстве, переход к которому задается матрицей $\widetilde{B}_k$, перейдем от него к новому пространству с помощью преобразования, заданного матрицей

$$S_k = \begin{cases} R_{1-\beta_k}(\xi_k), & \text{если } -1 \ge (1/\varepsilon)\lambda_{\max}(W), \\ I_n - \beta_k(\widetilde{A}_k uu^T + (\widetilde{A}_k uu^T)^T) & \text{в противном случае.} \end{cases}$$

(оно соответствует оператору преобразования исходного пространства $B_{k+1} = S_kB_k$). В этом пространстве $A_{k+1} = S_k\widetilde{A}_kS_k^{-1}$, $B_{k+1} = S_k\widetilde{B}_k$, $b_i^{(k+1)} = S_k\widetilde{b}_i^{(k)}$, $i = 1,\ldots,m$, $W_k = \widetilde{A}_k^T + \widetilde{A}_k$. Проведя операцию масштабирования, получаем снова пространство, в котором $P_{k+1} = \widetilde{B}_{k+1}$ принимает вид единичной матрицы и можно проделать шаг субградиентного спуска по той же схеме, что и на k-м шаге.

Все эти рассуждения аналогичны тем, которые приведены в работе [8] при рассмотрении задачи построения оптимального по объему эллипсоида, описанного вокруг конечного набора точек, однако для данной задачи преобразование пространства задается не только оператором растяжения, но и матрицей вида $\left[I - \gamma(Auu^T + (Auu^T)^T)\right]$.

Формально алгоритм записывается следующим образом:

0-й шаг. Матрица $A_0 = A$, $B_0 = I_n$ (n-мерная единичная матрица), $\{b_i^{(0)}\}_{i=1}^m = \{b_i\}_{i=1}^m$.

После k шагов имеем B_k – результирующую матрицу преобразования пространства, и A_k.

(k + 1)-й шаг.

1. $r_k = \max\limits_i \|b_i^{(k)}\| = \|b_j^{(k)}\|$.

2. Масштабирование: $\widetilde{A}_k = A_k, \quad \widetilde{B}_k = B_k/r_k, \quad \widetilde{b}_i^{(k)} = b_i^{(k)}/r_k$, $i = 1, \ldots, m$.

3. $q = (1/\varepsilon)\lambda_{\max}(W_k), \quad W_k = \widetilde{A}_k^T + \widetilde{A}_k$.

4. Если $-1 \geq q$, то $\xi_{k+1} = \widetilde{b}_j^{(N)}$, $A_{k+1} = R_{1-\beta_k}(\xi_{k+1})\widetilde{A}_k R_{1-1/\beta_k}(\xi_{k+1})$,
$$B_{k+1} = R_{1-\beta_k}(\xi_{k+1})\widetilde{B}_k, \quad b_i^{(k+1)} = R_{1-\beta_k}(\xi_{k+1})\widetilde{b}_i^{(k)}, \quad i = 1, \ldots, m.$$

В противном случае

$$A_{k+1} = \left[I_n - \beta_k \left(\widetilde{A}_k u_k^T u_k + \left(\widetilde{A}_k^T u_k^T u_k\right)^T\right)\right] \widetilde{A}_k \times$$
$$\times \left[I_n - \beta_k \left(\widetilde{A}_k u_k^T u_k + \left(\widetilde{A}_k^T u_k^T u_k\right)^T\right)\right]^{-1},$$
$$B_{k+1} = \left[I_n - \beta_k \left(\widetilde{A}_k u_k^T u_k + \left(\widetilde{A}_k^T u_k^T u_k\right)^T\right)\right] \widetilde{B}_k,$$
$$b_i^{(k+1)} = \left[I_n - \beta_k \left(\widetilde{A}_k u_k^T u_k + \left(\widetilde{A}_k^T u_k^T u_k\right)^T\right)\right] \widetilde{b}_i^{(k)}, \quad i = 1, \ldots, m.$$

Описание алгоритма завершено.

Аналогично теореме о сходимости «простого» алгоритма построения минимального по объему эллипсоида, описанного вокруг заданного конечного набора точек (см. [8]), можно доказать сходимость приведенного алгоритма при определенных условиях.

Теорема 2. *Пусть $P_k = B_k/r_k$ и conv$\{0, \{b_i\}_{i=1}^m\}$ представляет собой n-мерное выпуклое тело. Тогда если:*

1) $0 < \beta_k < 1, \quad k = 1, 2, \ldots,$

2) $\sum\limits_{k=1}^{\infty} \beta_k = +\infty, \quad \beta_k \to 0$ *при* $k \to +\infty$, *то* $\lim\limits_{k \to +\infty} P_k = P^*$,

где $K^* = P^{*T} \cdot P^*$ – матрица экстремального эллипсоида, соответствующего минимуму штрафной функции $f(P)$ (7).

Итак, для построения экстремального инвариантного эллипсоида предложены алгоритмы двух типов, первый из которых использует негладкую штрафную функцию, а второй – последовательное преобразование пространства (как и для экстремальных эллипсоидов в работе [8]). Однако для данной задачи, в отличие от рассмотренных в [8], алгоритм второго типа не дает существенного упрощения вычислений на одной итерации по отношению к алгоритму первого типа. Рассмотренные в [8] алгоритмы первого типа требовали $O(n^4 + mn)$ $(O(m^3)$ для двойственной задачи) арифметических операций на каждой итерации, а второго – $O(nm)$ арифметических операций (следует оговориться, что необходимое для достижения определенной точности решения количество итераций для алгоритмов второго типа значительно больше, чем для первого). Для рассматриваемой в статье задачи имеем $O(n^4 + nm)$ и $O(n^3 + nm)$ арифметических операций соответственно (считаем, что вычисление максимального собственного числа производится с заданной точностью относительно следа матрицы). Конечно, вычисление обратной матрицы специального вида $(I_n - \gamma(Auu^T + (Auu^T)^T)$ требует на порядок меньше итераций, если воспользоваться известным утверждением – когда $D = C + ab^T$, то $D^{-1} = C^{-1} - (C^{-1}ab^TC^{-1})/(1 + b^TC^{-1}a)$, но трудоемкая операция нахождения максимального собственного числа и соответствующего ему собственного вектора присутствует как в алгоритме первого типа, так и второго. Однако использование на практике идеи последовательного преобразования пространства для решения исследуемой задачи нельзя безоговорочно отбрасывать. Подтверждением этому является предложенный ниже алгоритм.

Исходная задача состоит в том, чтобы построить эллипсоид минимального объема с центром в точке 0, содержащий все траектории движения, выходящие из точек множества M для $t \geq t_0$ и описанные устойчивой системой (1). Если аппроксимировать каждую из этих траекторий конечной последовательностью точек, то получим задачу построения эллипсоида минимального объема с центром в точке 0, описанного вокруг заданного набора точек. Сложность при таком подходе состоит в существенном увеличении трудоемкости задачи (размерности аппроксимируемого эллипсоидом множества точек) для нахождения инвариантного эллипсоида с приемлемой точностью.

Ниже предлагается реализующий этот подход алгоритм, который позволяет рассматривать только часть точек траекторий, что делает его достаточно работоспособным. В основе этого алгоритма лежит «простой» метод построения описанного вокруг заданой совокупности точек

$S^0 \in E^n$ эллипсоида минимального объема [11], который заключается в последовательном сжатии пространства в направлении максимально удаленной от начала координат точки аппроксимируемого множества S^0 в текущем преобразованном пространстве. Внешне этот метод выглядит следующим образом. В качестве начального приближения выбирается сфера $\overline{S}^0$ минимального радиуса с центром в точке 0, включающая S^0. Затем пространство E^n сжимается в направлении активной точки, т. е. точки, принадлежащей множеству S^0 и лежащей на этой сфере. В преобразованном пространстве множеству S^0 соответствует S^1, для которого опять строится сфера минимального радиуса $\overline{S}^1$, включающая S^1, и т. д. Последовательности сфер в преобразованных пространствах соответствует последовательность описанных вокруг S^0 эллипсоидов $E_0, E_1, \ldots, E_k$ в первоначальном пространстве. При определенных условиях, как доказано в [8], пределом последовательности E_k является оптимальный описанный эллипсоид. При реализации этого метода не обязательно просматривать все точки на всех шагах – необходимо лишь, чтобы каждая точка могла быть просмотрена неограниченное число раз.

Используем возможность этого метода в процессе его работы корректировать аппроксимируемое множество точек. Основная идея предлагаемого алгоритма построения инвариантного эллипсоида состоит в том, чтобы, начиная с исходного $M^0 = M$ (множества точек в момент времени t_0), на каждой последующей итерации дополнять это множество только теми точками траекторий, которые лежат рядом с активной точкой на данной итерации и могут повлиять на вид экстремального эллипсоида.

Перейдем непосредственно к описанию алгоритма.

0-й шаг. $B_0 = I_n$ (B_k – матрица линейного преобразования, соответствующая переходу от начального пространства к текущему на k-м шаге алгоритма; на нулевой итерации равна n-мерной единичной матрице); $M^0 = \left\{M_i^0\right\}_{i=1}^m$, где $M_i^0 = \{b_i\}$ (M_i^k представляет собой набор точек траектории, выходящей из точки b_i, в текущем пространстве; на нулевом шаге это множество состоит только из начальной точки в момент времени t_0), $m(0) = |M^0| = m$, $r_0 = \max_{j=\overline{1,m(0)}} \|b_j\|$.

После k шагов имеем B_k – результирующую матрицу преобразования пространства, $M^k = \left\{M_i^k\right\}_{i=1}^m$, $r_k = \max_{j=\overline{1,m(k)}} \|b_j^{(k)}\| = \|b_{j^*(k)}^{(k)}\|$, $m(k) = |M^k|$, $b_{j^*(k)}^{(k)} \in M_{i^*(k)}^{(k)}$.

(k +1)–й шаг.

1. $\xi_{k+1} = b^{(k)}_{j^*(k)} / \| b^{(k)}_{j^*(k)} \|$.

2. $b_j^{(k+1)} = R_{1-\beta_k}(\xi_{k+1}) b_i^{(k)}$, $j = \overline{1, m(k)}$, $B_{k+1} = R_{1-\beta_k}(\xi_{k+1}) B_k$, где $R_\alpha(\xi) = I_n + (\alpha - 1)\xi\xi^T$ – оператор растяжения пространства в направлении вектора ξ с коэффициентом α [7].

3. Добавляя новые точки к $M^k_{i^*(k)}$ ($\forall i \neq i^*(k) \quad M_i^{k+1} = M_i^k$), строим множество M^{k+1}. Для этого осуществляем следующие операции:

а) находим ближайшие предшествующую и последующую по времени к $b^{(k)}_{j^*(k)}$ точки из $M^k_{i^*(k)}$(множества точек траектории, выходящей из точки $b_{i^*(k)}$ в текущем пространстве после k шагов); эти точки характеризуются моментами времени t_1, t_2 и $t_{j^*(k)}$ соответственно;

б) вычисляем точки $b_1 = x\left(\left(t_1 + t_{j^*(k)}\right)/2\right)$ и $b_2 = x\left(\left(t_2 + t_{j^*(k)}\right)/2\right)$ (например, по разностной схеме); при $t_{j^*(k)} = t_0$ точка b_1 не строится, и если $x(t_{j^*(k)})$ является последней по времени точкой множества $M^k_{i^*(k)}$, то $b_2 = x(t_{i^*(k)} + \tau)$, где τ – некоторый постоянный шаг (например, $\tau = 0{,}01$);

в) переводим полученные точки b_1, b_2 в текущее пространство и добавляем их к $M^k_{i^*(k)} : M^k_{i^*(k+1)} = \{M^k_{i^*(k)},\ B_k b_1,\ B_k b_2\}$.

После выполнения п. 3 имеем

$$M^{k+1} = \left\{M_i^k\right\}_{i \neq i^*(k)} \cup M^k_{i^*(k+1)}, \quad m(k+1) = |M^{k+1}|.$$

4. $r_{k+1} = \max\limits_{j=\overline{1,m(k+1)}} \|b_j^{(k+1)}\| = \|b^{(k+1)}_{j^*(k+1)}\|$.

Описание алгоритма завершено.

Чтобы показать сходимость данного алгоритма, приведем следующие рассуждения. Рассмотрим алгоритм, который отличается от формально записанного выше только п. 3 (способом построения множества M^{k+1}). Пусть, начиная с исходного $M^0 = M$ (множества точек в момент времени t_0), на каждой последующей итерации это множество дополняется некоторым конечным подмножеством $\widetilde{M}^k$ точек траекторий, выходящих из всех начальных состояний M^0 динамической системы – $M^{k+1} = M^k + \widetilde{M}^k, \quad |\widetilde{M}^k| < +\infty$. Тогда при условии, что с увеличением количества шагов k итерационного алгоритма имеет место более точная аппроксимация множеством M^k всех $\left|M^0\right|$ траекторий (т. е. максимальное расстояние между соседними точками для каждой траектории стремится к нулю при $k \to +\infty$), справедливо утверждение: если

1) $0 < \beta_k < 1, \quad k = 1, \dots ,$

2) $\sum\limits_{k=1}^{\infty} \beta_k = +\infty, \quad \beta_k \to 0 \quad \text{при} \quad k \to +\infty,$

то $\lim\limits_{k \to +\infty} K_k = K^*$, где $K_k = B_k^T B_k / r_k^2$, а K^* – матрица инвариантного эллипсоида E^* минимального объема для динамической системы (1). Однако понятно, что для построения эллипсоида E^* достаточно иметь только те точки траекторий M^*, которые лежат на поверхности E^* и непосредственно влияют на вид эллипсоида. А значит, и для сходимости алгоритмов, реализующих этот подход, достаточно, чтобы, начиная с некоторого конечного N в последовательности множеств M^k, $k = N, N+1, N+2, \dots$, можно было выделить последовательность подмножеств, которая стремится к M^* при $k \to +\infty$. Формально описанный выше алгоритм этому условию удовлетворяет.

Рассмотренные в статье алгоритмы были программно реализованы и проверены на ряде численных экспериментов, которые показали их работоспособность.

Литература

1. Черноусько Ф. Л. Оценивание фазового состояния динамических систем (метод эллипсоидов). – М.: Наука, 1988. – 320 с.

2. Khachiyan L., Todd M. On the complexity of approximating the maximal inscribed ellipsoid for a polytope: Techn. rep. N 893. – Ithaca, New York: Cornel. Univer., 1990. – 39 p.

3. Shor N. Z., Stecenko S. I., Berezovski O. A. Algorithms of constructing optimal inscribed and circumscribed ellipsoids // Abstr. of 14–th IFIP Conf. «System Modelling and Optimization». – Leipzig, 1989. – Heft 3.

4. Федоров Ф. Ф. Теория оптимального эксперимента. – М.: Наука, 1971. – 184 с.

5. Нестеров Ю. Е., Немировский А. С. Самосогласованные функции и полиномиальные алгоритмы в выпуклом программировании. – М.: ЦЭМИ АН СССР, 1989. – 194 с.

6. Тарасов С П., Хачиян Л. Г., Эрлих И. И. Метод вписанных эллипсоидов //Докл. АН СССР. – 1988. – Т. 298. – № 5. – С. 1081–1085.

7. Shor N. Z. Minimization methods for non-differentiable functions // Springer series, Computational mathematics. – Berlin: Springer-Verlag, 1985. – V. 3. – 162 p.

8. Shor N. Z., Berezovski O. A. New algorithms for constructing optimal circumscribed and inscribed ellipsoids // Optimization Methods and Software. – 1992. – № 1. – P. 283–299.

9. Ланкастер П. Теория матриц. – М.: Наука, 1982. – 269 с.

10. Пшеничный Б. Н. Метод линеаризации. – М.: Наука, 1983: – 136 с.

11. Шор Н. З., Стеценко С. И. Алгоритм последовательного сжатия пространства для построения описанного эллипсоида минимального объема // Исследование методов решения экстремальных задач. – К.: Ин–т кибернетики АН УССР, 1990. – С. 25–30.

СОДЕРЖАНИЕ

НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК УКРАИНЫ
ИНСТИТУТ КИБЕРНЕТИКИ ИМ. В. М. ГЛУШКОВА

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ И МОЛОДЕЖИ
РЕСПУБЛИКИ МОЛДОВА

АКАДЕМИЯ ТРАНСПОРТА,
ИНФОРМАТИКИ И КОММУНИКАЦИЙ

Н. З. ШОР

МЕТОДЫ МИНИМИЗАЦИИ НЕГЛАДКИХ ФУНКЦИЙ И МАТРИЧНЫЕ ЗАДАЧИ ОПТИМИЗАЦИИ

Сборник избранных трудов

Издательство „Эврика“, тел. 63-90-29
мун. Кишинэу, Республика Молдова
Мунчештское шоссе, 121a

Подписано в печать 15.10.2009
Формат 60x90/16
Усл. печ. листов 15
Тираж 150
Заказ № 7

Типография Академии наук Молдовы
Г.Кишинев, ул. П. Мовилэ, 8

АКАДЕМИК Н. З. ШОР

Шор Наум Зуселевич – основоположник направления недифференцируемой оптимизации в математическом программировании, профессор, академик НАН Украины. Родился 1 января 1937 г. в Киеве. После окончания в 1958 г. Киевского национального университета имени Тараса Шевченко работал в Институте кибернетики НАН Украины. Автор 10 монографий и более 200 статей по математическому программированию, вычислительной математике и теории графов. Лауреат Государственных премий УССР (1973 г.), СССР (1981 г.), Украины (1993 г., 2000 г.), премии им. В. М. Глушкова НАН Украины (1987 г.), премии им. В. С. Михалевича НАН Украины (1997 г.).

Н. З. ШОР • МЕТОДЫ МИНИМИЗАЦИИ НЕГЛАДКИХ ФУНКЦИЙ И МАТРИЧНЫЕ ЗАДАЧИ ОПТИМИЗАЦИИ