

Институт проблем управления им. В.А. Трапезникова
Российской академии наук
(ИПУ РАН)

К 85-летию ИПУ РАН

М. В. Хлебников
М. В. Балашов, А. А. Тремба

ОПТИМИЗАЦИЯ И УПРАВЛЕНИЕ

Рекомендовано
федеральным учебно-методическим объединением
в системе высшего образования по укрупненным
группам специальностей и направлений подготовки
27.00.00 *Управление в технических системах*
в качестве учебного пособия при реализации основных
профессиональных образовательных программ высшего
образования по направлениям подготовки высшего
образования: 27.04.03 *Системный анализ и управление*;
27.04.04 *Управление в технических системах*;
27.06.01 *Управление в технических системах*

Москва
Издательство URSS
2024

ББК 22.16 22.182 22.192

Рецензенты:

директор Института системного анализа ФИЦ ИУ РАН,
академик РАН, доктор техн. наук, профессор *Ю.С. Попков*

директор Института проблем управления им. В.А. Трапезникова
РАН, академик РАН, доктор техн. наук, профессор *Д.А. Новиков*

**Хлебников Михаил Владимирович,
Балашов Максим Викторович,
Тремба Андрей Александрович**
Оптимизация и управление. — М.: ЛЕНАНД, 2024. — 448 с.

ISBN 978-5-00237-041-2

В учебном пособии в современной форме излагаются основы безусловной и условной оптимизации и рассматривается их применение при решении ряда важнейших задач теории автоматического управления: синтеза стабилизирующих и робастно стабилизирующих регуляторов, линейно-квадратичного управления по состоянию и по выходу, подавления неслучайных внешних возмущений, гарантирующей фильтрации, синтеза ПИД-регуляторов. В приложении содержится используемый математический аппарат. Многочисленные примеры иллюстрируют специфику применения теоретического материала.

Пособие предназначено для использования при реализации основных профессиональных образовательных программ высшего образования по направлениям подготовки высшего образования 27.04.03 «Системный анализ и управление», 27.04.04 «Управление в технических системах», 27.06.01 «Управление в технических системах». Книга также будет полезна научным и инженерно-техническим работникам.

© М.В. Хлебников, М.В. Балашов,
А.А. Тремба, 2024
© ИПУ РАН, 2024

*Посвящается светлой памяти
нашего учителя и друга
Бориса Теодоровича Поляка*

Оглавление

От авторов	8
Предисловие	9
Введение	13
Список обозначений	17
1. Безусловная оптимизация	24
1.1. Схемы методов оптимизации. Скорость сходимости	24
1.2. Методы нулевого порядка	27
1.2.1. Одномерная оптимизация и выпуклость	28
1.2.2. Метод золотого сечения	33
1.2.3. Выбор длины шага при минимизации выпуклых функций в \mathbb{R}^n	37
1.2.4. Метод случайного блуждания	39
1.2.5. Метод имитации отжига	42
1.2.6. Метод Нелдера–Мида	44
1.3. Методы первого порядка	47
1.3.1. Метод градиентного спуска	47
1.3.2. Условие Лежанского–Поляка–Лоясевича	51
1.3.3. Метод тяжелого шарика	56
1.3.4. Шаг Армихо	58
1.3.5. Метод сопряженных градиентов	60
1.3.6. Субградиентный спуск	66
1.4. Метод Ньютона в задачах безусловной минимизации . .	72
1.4.1. Метод Ньютона в задачах минимизации	72
1.4.2. Комбинированный метод Ньютона	78
1.5. Примеры экстремальных задач в безусловной минимизации	88

2. Условная оптимизация	103
2.1. Устойчивость задач минимизации. Модуль выпуклости	103
2.2. Конечношаговые методы	109
2.2.1. Линейное программирование	109
2.2.2. Квадратичное программирование	119
2.3. Методы первого порядка	122
2.3.1. Метод проекции градиента	122
2.3.2. Метод проекции субградиента	134
2.3.3. Метод условного градиента	135
2.4. Метод отсекающей гиперплоскости. Метод эллипсоидов	143
2.5. Метод Ньютона в задачах условной минимизации	147
2.6. Метод штрафных функций	150
2.7. Полуопределенное программирование	153
2.7.1. Линейные матричные неравенства	153
2.7.2. Задача полуопределенного программирования	157
2.7.3. Лемма Шура и ее применение	160
2.7.4. Лемма Питерсена	165
2.8. Примеры экстремальных задач в условной минимизации	167
3. Оптимизация в теории автоматического управления	176
3.1. Линейные матричные неравенства в задачах управления	176
3.1.1. Квадратичная устойчивость	176
3.1.2. Квадратичная стабилизация	182
3.1.3. Линейно-квадратичное управление по состоянию	193
3.1.4. H_∞ -оптимизация	199
3.2. Робастная стабилизация и управление	207
3.2.1. Структурированная матричная неопределенность	207
3.2.2. Робастная квадратичная стабилизация	210
3.2.3. Робастный линейно-квадратичный регулятор	214
3.2.4. H_∞ -оптимизация: робастный вариант	217
3.3. Линейно-квадратичное управление по выходу	219
3.3.1. Подходы к решению	219
3.3.2. Постановка задачи	222
3.3.3. Примеры	224
3.3.4. Свойства минимизируемой функции	228
3.3.5. Алгоритм решения линейно-квадратичной задачи	240
3.3.6. Связь с общими задачами оптимизации	244
3.4. Подавление внешних возмущений	246

3.4.1. Постановка задачи	246
3.4.2. Метод инвариантных эллипсоидов и задача анализа	247
3.4.3. Задача синтеза	253
3.4.4. Примеры	265
3.5. Гарантирующая фильтрация	275
3.5.1. Постановка задачи	277
3.5.2. Гарантирующий фильтр	278
3.5.3. Вычисление оптимальной матрицы фильтра	280
3.5.4. Непрерывный случай	290
3.5.5. Примеры и обсуждение	293
3.6. Синтез ПИД-регуляторов	300
3.6.1. Постановка задачи и ее сведение к параметрической линейно-квадратичной задаче	301
3.6.2. Примеры	308
Приложение	328
A. Сведения из математического анализа	328
A.1. Компактные множества	328
A.2. Полунепрерывные функции. Теорема Вейерштрасса	329
A.3. Производные, градиенты, разложения	330
A.4. Липшицевы функции и функции с липшицевым градиентом	333
A.5. Теорема о неявной функции и обратном отображении	335
B. Сведения из линейной алгебры	336
B.1. Элементарные свойства матриц	336
B.2. Матричные нормы	339
B.3. Уравнения и неравенства Ляпунова и их свойства	342
B.4. Матричные многообразия	350
B. Элементы теории автоматического управления	351
B.1. Пространство состояний	351
B.2. Передаточная функция	353
B.3. Нормы сигналов	360
B.4. Коэффициенты усиления	363
B.5. Нормы передаточных функций	365
B.6. Одномерные системы управления	367
B.7. Программное управление. Управляемость	373

В.8.	Управление по обратной связи	380
В.9.	Наблюдаемость	383
Г.	Выпуклость множеств и функций. Элементы негладкого анализа	385
Г.1.	Выпуклость. Теорема Каратеодори. Непрерывность выпуклых функций	385
Г.2.	Метрическая проекция точки на выпуклое замкнутое множество. Теоремы об отделимости	390
Г.3.	Выпуклые конусы	393
Г.4.	Теоремы о крайних точках и лучах	395
Г.5.	Субдифференциал. Теорема Моро–Рокафеллара	397
Г.6.	Сильно выпуклые функции	401
Г.7.	Сильно выпуклые множества и проксимально гладкие множества. Слабо выпуклые и слабо вогнутые функции	404
Г.8.	Метрика Хаусдорфа. Погрешности многогранных аппроксимаций	409
Г.9.	Выживание решений*	411
Г.10.	Неподвижные точки и точки совпадения отображений*	415
Д.	Сведения из теории экстремальных задач	418
Д.1.	Безусловная оптимизация	419
Д.2.	Гладкая задача условного экстремума. Метод множителей Лагранжа	420
Д.3.	Выпуклая задача условного экстремума	422
Д.4.	Замечание о необходимых условиях минимума функции на проксимально гладком множестве	423
Литература		425
Предметный указатель		437

От авторов

В 1983 году вышла в свет знаменитая монография Б.Т. Поляка «Введение в оптимизацию», до сих пор не потерявшая своей актуальности. Вместе с тем за прошедшие 40 лет появились и новые результаты, и новый математический аппарат, а какие-то моменты с течением времени начали смотреться под несколько иным углом зрения. Также нам представлялось важным проследить тесную связь между оптимационными задачами и задачами, возникающими в теории управления. Борис Теодорович Поляк принимал активное участие в обсуждениях общей направленности и содержания настоящей книги, однако эта деятельность оборвалась с его уходом из жизни в феврале 2023 года.

Мы благодарны Институту проблем управления им. В.А. Трапезникова Российской академии наук, в котором работаем долгое время, за предоставленную возможность заниматься теоретическими исследованиями. Славные научные традиции Института служили для нас образцами для подражания.

Мы также благодарны нашим коллегам К.З. Биглову, А.В. Дмитрику, А.В. Назину, Л.Б. Рапопорту, М.Е. Шайкину, сделавшим много критических замечаний и предложений (мы постарались их учесть при подготовке рукописи, но ответственность за окончательный результат целиком лежит на нас), а также В.П. Михайлову, взявшему на себя труд по изготовлению рисунков, и Я.И. Квинто за работу по корректуре и редактированию рукописи. Особая наша признательность — директору ИПУ РАН академику Российской академии наук Дмитрию Александровичу Новикову, инициатору создания этой книги, за постоянную поддержку и многочисленные советы.

М.В. Хлебников
М.В. Балашов
А.А. Тремба

Предисловие

Обсудим содержание двух базовых категорий, вынесенных в название настоящей книги, — «оптимизация» и «управление» — и взаимосвязь между ними.

Под *управлением* понимается воздействие на *управляемую систему с целью¹* обеспечения требуемого ее *поведения²* [16, 17]. Система управления включает в себя три составные части: управляемую систему, *управляющую систему* и связи между ними.

Оптимизация — это нахождение среди *допустимого множества* возможных вариантов (альтернатив, решений) наилучших (например, экстремизирующих *критерий эффективности*) [17].

Рассмотрим систему управления, представленную на рис. 0.0.1, где пунктирными линиями обозначены сущности, которые в зависимости от конкретной рассматриваемой модели могут присутствовать или отсутствовать.

Управляемая система характеризуется своим *состоянием* (если не оговорено особо, будем считать, что выход совпадает с состоянием), которое может изменяться во времени и зависеть в общем случае в явном виде от:

- предшествующих состояний;
- времени;
- управления (управляющего воздействия);
- возмущений, отражающих влияние внешней среды.

Эта зависимость называется *моделью управляемой системы*. Обозначим через $x(t) \in X$ состояние управляемой системы в момент времени $t \geq t_0$, принадлежащее множеству допустимых состояний X ; че-

¹ Цель — то, к чему стремятся, что надо осуществить (все определения в сносках приводятся в соответствии с [17]).

² Поведение системы — последовательное (во времени), хотя бы частично наблюдаемое, поддающееся измерению, объективной фиксации изменение ее состояния.

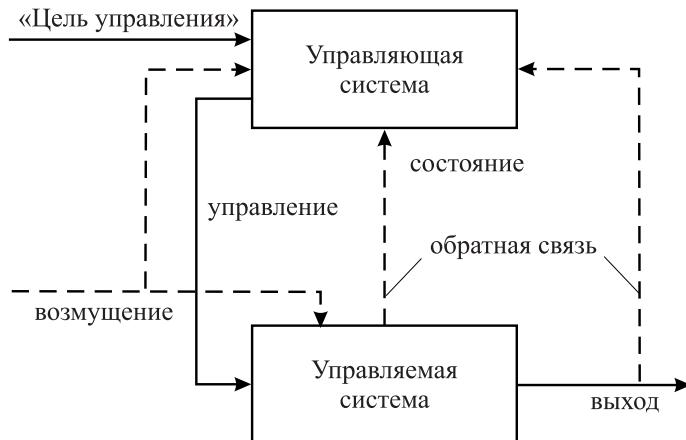


Рис. 0.0.1. Базовая структура системы управления [15].

рез x_t — ее предшествующую *траекторию* (поведение управляемой системы — последовательность всех ее состояний $\{x(\tau)\}, \tau \in [t_0, t]$); через x_t^* — *целевую траекторию*; через $u(t) \in U$ — управляющее воздействие, принадлежащее множеству допустимых управлений U ; через u_t — последовательность управляющих воздействий $\{u(\tau)\}, \tau \in [t_0, t]$; через $\omega(t) \in \Omega$ — значение возмущения в момент времени $t \geq t_0$, принадлежащее множеству возможных возмущений Ω ; через ω_t — последовательность возмущений $\{\omega(\tau)\}, \tau \in [t_0, t]$.

Общей моделью управляемой системы называется совокупность (для всех $t \geq t_0$) отображений

$$(x_t, u_t, \omega_t, u(t), \omega(t)) \mapsto x(t), \quad x(t_0) = x_0. \quad (0.0.1)$$

Достаточно часто управляемая система не обладает глубокой «памятью». Примером служит *динамическая система*

$$\dot{x} = f(x(t), u(t), t), \quad x(t_0) = x_0,$$

являющаяся хрестоматийным объектом для *теории автоматического управления* [21].

Пусть для некоторого конечного момента времени $T > t_0$ на множестве траекторий управляемой системы, множестве допустимых последовательностей управляющих воздействий и множестве возможных целевых траекторий задан действительнозначный функционал

$K(x_T, u_T, x_T^*)$, называемый *критерием*³ *эффективности*⁴ управления системой (0.0.1). Примером такого критерия может быть $-\|x_T - x_T^*\|$ — близость траектории управляемой системы к целевой, взятая со знаком минус.

Задача управления заключается в нахождении *оптимальной* последовательности управляющих воздействий u_T^* , то есть допустимой последовательности⁵ $u(t) \in U$, $t \in (t_0, T]$, обладающей максимальной эффективностью:

$$u_T^* \in \operatorname{Arg} \max_{u_T} K(x_T, u_T, x_T^*) \quad \text{при условии (0.0.1).} \quad (0.0.2)$$

Задача управления (0.0.2) является задачей оптимизации. И, в первую очередь, в этом заключается глубокая взаимосвязь управления и оптимизации: целенаправленность любого управления подразумевает его «оптимальность» по отношению к соответствующей цели. На протяжении последних 150 лет теория управления и математические методы оптимизации развиваются параллельно, обогащая друг друга, давая новые возможности и формулируя новые вызовы (а иногда и сливаясь, например, в виде теории оптимального управления)⁶.

Во-вторых, при совместном решении задач управления, вычислений и связи (так называемая парадигма C³ – Control, Computing, Communication) характерные времена изменений состояний управляемой системы и внешней среды накладывают соответствующие требования на темп вычислений, осуществляемых в управляющей системе, то есть на вычислительную сложность (с учетом специфики аппаратной реализации) используемых алгоритмов оптимизации.

В-третьих, существует еще одна важная «точка контакта» оптимизации и управления: при описании целенаправленного поведения управляемой системы, включающей человека, предполагается (см. [16]), что последний характеризуется своим некоторым критерием эфек-

³ Критерий — средство для вынесения суждения; стандарт для сравнения; правило для оценки; мера степени близости к цели.

⁴ Эффективность — степень соотношения полученного результата и характеристик, использованных для его достижения ресурсов.

⁵ В качестве отступления отметим, что в начальный момент времени управление может искаститься на весь плановый горизонт T или как функция времени — так называемое *программное управление* (или *разомкнутое управление*), или как функция состояния управляемой системы — так называемое *позиционное управление* (или управление по отклонению), или как функция внешних возмущений — так называемое *управление по возмущению* [15].

⁶ В третьей главе можно найти многочисленные примеры, когда подходы, методы и алгоритмы решения частной задачи управления могут быть обобщены на достаточно широкий класс оптимизационных задач.

тивности $k(x, u)$ и выбирает самостоятельно свое допустимое состояние \hat{x} , экстремизируя этот критерий, то есть модель управляемой системы имеет вид

$$\hat{x}(t) \in P(u(t)) \doteq \operatorname{Arg} \max_{x \in X} k(x, u(t)).$$

Соответствующая задача управления будет заключаться в максимизации гарантированной эффективности:

$$u_T^* \in \operatorname{Arg} \max_{u_T} \min_{x_T \in P(u_T)} K(x_T, u_T, x_T^*),$$

являясь нетривиальной оптимационной задачей.

В-четвертых, решение той или иной конкретной задачи управления может сводиться к нахождению «оптимальных» параметров управляемой системы (например, матрицы обратной связи линейно-квадратичного регулятора или параметров ПИД-регулятора в задаче подавления внешних возмущений — см. третью главу) посредством постановки и решения соответствующей оптимационной задачи.

Таким образом, можно выделить следующие аспекты взаимосвязи между управлением и оптимизацией:

1. Любая задача управления является задачей оптимизации.
2. Системы управления реального времени имеют свои ограничения на используемые в них методы оптимизации.
3. Модель управляемой системы может включать задачу оптимизации.
4. Решение конкретной задачи управления (в частности задачи синтеза, подбора параметров управляемой системы) может сводиться к тем или иным оптимационным задачам.

Ниже в настоящей книге, помимо изложения основ методов оптимизации (первые две главы), приводится ряд примеров применения этих методов оптимизации при решении задач автоматического управления линейными динамическими системами (глава 3).

Д.А. Новиков, академик РАН

Введение

Две науки — управление и оптимизация — до какого-то момента развивались независимо. Управление имело дело с регуляторами, устойчивостью, передаточными функциями, обратной связью, годографом Найквиста и т. п. Оптимизация же зародилась еще в античности. Потом, при изучении гладких функций, возникла теория экстремума (правило Ферма, метод Ньютона, метод множителей Лагранжа); значительно позже — уже в XX веке — сначала возник теоретический интерес к таким задачам, как линейное и нелинейное программирование, а затем были созданы и численные методы их решения. При мерно в это же время — с конца 1950-х годов — в управлении произошел большой поворот: возникла концепция оптимального управления (принцип максимума Понтрягина, линейно-квадратичный регулятор), и стала видна связь этих двух дисциплин.

В последующие годы эта связь стала только усиливаться. Так, например, проблема синтеза оптимальных регуляторов все чаще рассматривается как прямая задача оптимизации в пространстве коэффициентов регулятора, а с появлением аппарата линейных матричных неравенств и программных средств для их решения многие задачи стабилизации, синтеза оптимальных регуляторов, подавления возмущений в управлении стали записываться в формате линейных матричных неравенств и решаться с помощью методов выпуклой оптимизации. Еще один пример: задача об устойчивости динамической системы на основе функции Ляпунова очень близка к конструированию методов безусловной оптимизации: в первом случае задан метод и для него ищется функция Ляпунова, а во втором задана функция, для которой надо построить метод оптимизации.

С другой стороны, обе науки продолжают находиться в процессе интенсивного развития. При этом продолжают меняться как их аппарат, так и используемые математические модели и методы. В связи с этим имеющиеся учебники нередко не успевают за этими революцион-

ными переменами и быстро устаревают. В последнее время нехватка современных учебных пособий стала особенно заметной.

Нам хотелось написать книгу, не только затрагивающую некоторые точки соприкосновения управления и оптимизации, но и в какой-то мере восполняющую этот пробел. С одной стороны, она должна содержать аппарат, адекватный современной теории, а с другой стороны, мы постарались сделать ее по возможности широко доступной. Поэтому используемый математический аппарат ограничен основными понятиями математического анализа, линейной алгебры, конечно-мерного функционального анализа. В наши намерения также входило сделать книгу по возможности краткой, содержащей лишь базовые задачи, идеи и методы. Мы постарались избежать повторов известных монографий и сделали акценты на ряде важных понятий и концепций, которые пока слабо отражены в учебной литературе.

Учебное пособие состоит из трех глав. Первые две главы охватывают круг вопросов, связанных соответственно с безусловной и условной оптимизацией. В частности, в главе 1 «Безусловная оптимизация» вводятся базовые понятия и обсуждаются методы нулевого порядка (случайного блуждания, имитации отжига, симплексный метод и др.), градиентные методы (градиентный спуск, метод тяжелого шарика, метод сопряженных градиентов и др.), а также метод Ньютона. Завершает главу ряд примеров задач безусловной оптимизации.

В главе 2 «Условная оптимизация» читатель познакомится с линейным и квадратичным программированием, методами проекции градиента и субградиента, методом условного градиента, а также такими методами, как метод отсекающей гиперплоскости и метод эллипсоидов, метод внутренней точки и др. Нам также казалось важным уделять большое внимание полуопределенному программированию. Глава завершается примерами задач условной оптимизации.

Во второй части книги (глава 3 «Оптимизация в теории автоматического управления») обсуждаются оптимационные подходы к задачам, относящимся к числу важнейших в теории автоматического управления: синтез стабилизирующих и робастно-стабилизирующих регуляторов, H_∞ -оптимизация, линейно-квадратичное управление по состоянию и по выходу, задача подавления неслучайных ограниченных внешних возмущений, синтез гарантирующего фильтра, задача настройки ПИД-регуляторов.

При написании разделов 1.3.3, 1.3.5, 1.3.6 и 2.2 мы опирались на монографию Б.Т. Поляка «Введение в оптимизацию». Мы часто использовали и ссылались на результаты С. Бойда, Р. Калмана, Ф. Клар-

ка, А.С. Немировского, Ю.Е. Нестерова, Я. Питерсена, Б.Т. Поляка, Р.Т. Рокафеллара, а также многих других исследователей, чьи имена можно найти в библиографии.

Книга содержит ряд особенностей. В первых двух главах мы постарались в рамках доступного объема как можно шире охватить алгоритмы для решения невыпуклых экстремальных задач, познакомив российского читателя с такими понятиями, как условие Поляка – Лоясевича, проксимальная гладкость и т. п., в контексте численных методов оптимизации. Заключительная глава книги содержит ряд совсем новых результатов, самым существенным образом опирающихся на содержание первых двух глав. В виде монографии этот материал был практически недоступен российскому (да и не только) читателю. Символом «*» помечены задачи повышенной сложности и разделы для более подготовленных читателей. В конце книги приведены приложения, содержащие основные математические сведения, не всегда приводимые в стандартных студенческих курсах; они будут полезны для понимания материала⁷. Книгу завершает список рекомендуемой литературы, который содержит как классические учебники и монографии на русском языке, так и зарубежные публикации по освещаемым разделам оптимизации и теории управления; в них также можно найти более подробные ссылки на современные направления исследований.

Мы рассчитываем, что материал, изложенный в книге, может быть положен в основу годового курса для студентов магистратуры математико-инженерных специальностей. Он может быть полезен и для аспирантов с достаточной математической подготовкой для самостоятельного изучения теории оптимизации и управления. Наконец, инженеры и исследователи, которые связаны с реальными задачами оптимизации и управления, смогут почертнуть в книге нужные теоретические сведения и найти в приведенной библиографии ссылки на актуальные и современные источники.

В учебное пособие не включены методические материалы, нередко приводимые в современных учебных курсах (контрольные вопросы, основные положения раздела, задачи — это привело бы к заметному увеличению объема книги), хотя оно и содержит некоторое количество упражнений. Вероятно, разработка этих материалов станет возможной после накопления практического опыта работы с учебным пособием.

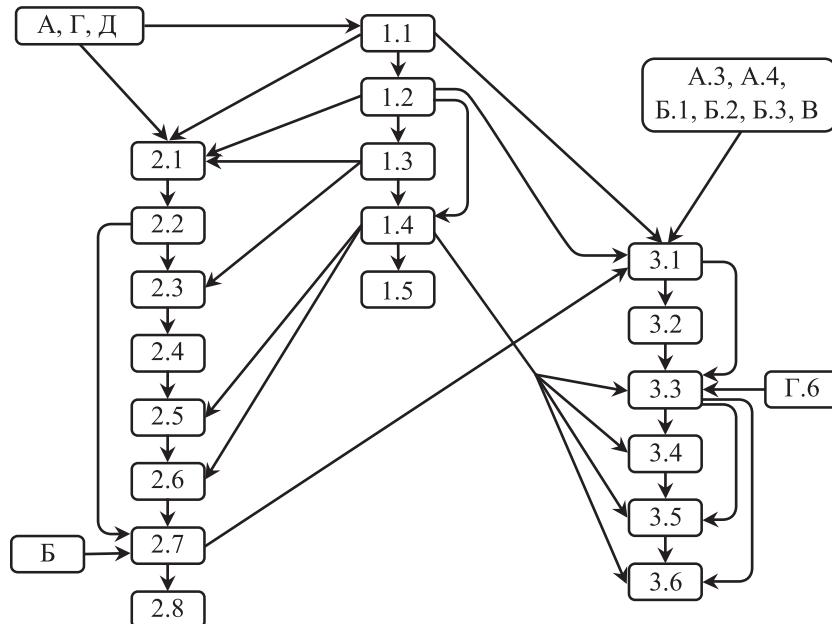
Книга целиком посвящена конечномерным задачам (это во многом обусловлено объемом работы), причем для достаточно хороших функ-

⁷При написании разделов Г.3 и Г.10 были использованы материалы, предоставленные А.В. Арутюновым.

ций, которые главным образом имеют липшицев градиент. В первых двух главах рассматриваются почти исключительно верхние оценки сложности алгоритмов. Мы практически не касаемся нижних оценок сложности, а также невыпуклого негладкого случая. Тем не менее рассматриваемые градиентные методы имеют много практических применений. Важные приложения к задачам автоматического управления собраны в третьей главе.

В книге нет бесконечномерных задач и постановок: задач оптимального управления, вариационного исчисления, а также экстремальных задач, возникающих в анализе. Мы также не затрагиваем дискретную и комбинаторную оптимизацию. Мы использовали главным образом классические градиенты (Гато, Фреше) и постарались не злоупотреблять современным вариационным анализом, хотя в приложении Г.7 можно найти некоторые идеи дальнейших обобщений.

Для удобства читателя приведем схему взаимосвязей разделов. Например, для чтения раздела 3.5 достаточно ознакомиться с приложениями А.3, А.4, Б.1–Б.3, В, Г.6 и разделами 1.1, 3.1 и 3.3. Это лишь наша рекомендация для начинающих изучать предмет; подготовленный читатель может читать разделы в любом порядке.



Список обозначений

\forall	квантор «для любого», «для всех»;
\exists	квантор «существует»;
\doteq	равно по определению;
\blacksquare	конец доказательства;
\blacktriangledown	конец замечания или примера;
\emptyset	пустое множество;
\mathbb{N}	множество натуральных чисел;
\mathbb{N}_0	множество $\mathbb{N} \cup \{0\}$;
\mathbb{R}	множество вещественных чисел;
\mathbb{R}_+	множество неотрицательных вещественных чисел;
\mathbb{C}	множество комплексных чисел;
\bar{z}	комплексное сопряжение числа $z \in \mathbb{C}$: если

$$z = x + jy, \quad j^2 = -1,$$

то

$$\bar{z} = x - jy;$$

$|x|$ абсолютная величина числа: для $x \in \mathbb{R}$

$$|x| = \begin{cases} x, & x \geq 0, \\ -x, & x < 0; \end{cases}$$

для $z \in \mathbb{C}$

$$|z| = \sqrt{z \cdot \bar{z}};$$

$\lceil x \rceil$ округление числа $x \in \mathbb{R}$ вверх до ближайшего целого:

$$\lceil x \rceil - 1 < x \leq \lceil x \rceil;$$

\mathbb{R}^n

пространство n -мерных вектор-столбцов с вещественными элементами:

$$x = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}, \quad \mathbf{x}_i \in \mathbb{R};$$

 $\mathbf{x}_i, y_i, \mathbf{a}_i, \mathbf{b}_i, \dots$

i -е компоненты векторов x, y, a, b, \dots соответственно (обозначаются монотонным шрифтом);

 $\{e_k\}_{k=1}^n$

стандартный ортонормированный базис в \mathbb{R}^n : у вектора e_k на k -м месте стоит единица, остальные элементы равны нулю;

 x^T

транспонирование вектора $x \in \mathbb{R}^n$:

$$x^T = (\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n);$$

 (x, y)

скалярное произведение векторов $x, y \in \mathbb{R}^n$:

$$(x, y) = x^T y = \sum_{i=1}^n \mathbf{x}_i y_i;$$

 $\|x\|$

норма вектора (если не указана явно, подразумевается евклидова норма); в частности, для $x \in \mathbb{R}^n$:

— l_p -норма:

$$\|x\|_p = \left(\sum_{i=1}^n |\mathbf{x}_i|^p \right)^{1/p}, \quad 1 \leq p \leq \infty,$$

— в том числе при $p = 1$

$$\|x\|_1 = \sum_{i=1}^n |\mathbf{x}_i|,$$

— при $p = 2$

$$\|x\| = \|x\|_2 = \left(\sum_{i=1}^n |\mathbf{x}_i|^2 \right)^{1/2} \quad (\text{евклидова норма}),$$

— при $p = \infty$

$$\|x\|_\infty = \max_{1 \leq i \leq n} |\mathbf{x}_i|;$$

$\mathcal{B}_r(x_0)$ замкнутый евклидов шар радиуса $r \geq 0$ в \mathbb{R}^n :

$$\mathcal{B}_r(x_0) = \{x \in \mathbb{R}^n : \|x - x_0\| \leq r\};$$

$\mathbb{R}^{n \times m}$ пространство матриц размера $n \times m$ с вещественными элементами

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}, \quad a_{ij} \in \mathbb{R};$$

I_n единичная матрица n -го порядка (если не указано явно — соответствующей размерности);

$\text{diag}\{a_1, \dots, a_n\}$ диагональная матрица с элементами a_i , $i = 1, \dots, n$;

A^\top транспонирование матрицы: если

$$A = (a_{ij}) \in \mathbb{R}^{n \times m},$$

то

$$A^\top = (a_{ji}) \in \mathbb{R}^{m \times n};$$

\mathbb{S}^n пространство симметричных матриц размера $n \times n$ с вещественными элементами:

$$\mathbb{S}^n = \{A \in \mathbb{R}^{n \times n} : A^\top = A\};$$

$\mathcal{S}_{n,k}$ многообразие Штифеля с $k \leq n$:

$$\mathcal{S}_{n,k} = \{X \in \mathbb{R}^{n \times k} : X^\top X = I_k\};$$

$\mathcal{G}_{n,k}$ вложение многообразия Грассмана с $k \leq n$ в \mathbb{S}^n :

$$\mathcal{G}_{n,k} = \{XX^\top : X \in \mathcal{S}_{n,k}\};$$

(a_{ij}) матрица с элементами

$$a_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m;$$

$\lambda_i(A)$ i -е собственное значение квадратной матрицы $A \in \mathbb{R}^{n \times n}$, $i = 1, \dots, n$;
 $\rho(A)$ спектральный радиус матрицы A :

$$\rho(A) = \max_i |\lambda_i(A)|;$$

$\sigma_i(A)$ i -е сингулярное число матрицы $A \in \mathbb{R}^{n \times m}$:

$$\sigma_i(A) = \lambda_i^{1/2}(A^T A), \quad i = 1, \dots, m;$$

$\det A$ определитель матрицы $A \in \mathbb{R}^{n \times n}$:

$$\det A = \prod_{i=1}^n \lambda_i(A);$$

$\text{tr } A$ след матрицы $A \in \mathbb{R}^{n \times n}$:

$$\text{tr } A = \sum_{i=1}^n a_{ii} = \sum_{i=1}^n \lambda_i(A);$$

$\text{rank } A$ ранг матрицы $A \in \mathbb{R}^{n \times m}$;
 $\langle A, B \rangle$ стандартное (оно же фробениусово) скалярное произведение в пространстве матриц $\mathbb{R}^{n \times m}$: для $A, B \in \mathbb{R}^{n \times m}$

$$\langle A, B \rangle = \text{tr } A^T B;$$

$\|A\|$ норма матрицы (если не указана явно, подразумевается стандартная операторная норма); в частности, для матрицы $A \in \mathbb{R}^{n \times m}$:

— стандартная операторная норма, она же спектральная норма (2-норма):

$$\|A\|_2 = \max_{\|x\| \leq 1} \|Ax\| = \max_i \lambda_i^{1/2}(A^T A) = \max_i \sigma_i(A)$$

(цифра 2 в индексе обычно опускается);
— строчная норма (1-норма):

$$\|A\|_1 = \max_{1 \leq i \leq n} \sum_{j=1}^m |a_{ij}|;$$

— фробениусова норма:

$$\|A\|_F = \sqrt{\langle A, A \rangle} = \sqrt{\text{tr } A^T A} = \left(\sum_{i=1}^n \sum_{j=1}^m |a_{ij}|^2 \right)^{1/2};$$

$A \succ 0$

матрица $A \in \mathbb{S}^n$ положительно определена:

$$x^T A x > 0 \quad \forall x \in \mathbb{R}^n, x \neq 0;$$

$A \succcurlyeq 0$

матрица $A \in \mathbb{S}^n$ неотрицательно определена:

$$x^T A x \geq 0 \quad \forall x \in \mathbb{R}^n;$$

$A \prec 0$ ($\preccurlyeq 0$)

матрица $A \in \mathbb{S}^n$ отрицательно (неположительно) определена;

$\text{Arg} \min_{x \in \mathcal{Q}} f(x)$

множество минимумов функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ на множестве $\mathcal{Q} \subset \mathbb{R}^n$:

$$\text{Arg} \min_{x \in \mathcal{Q}} f(x) = \left\{ x_* \in \mathcal{Q}: f(x_*) = \inf_{x \in \mathcal{Q}} f(x) \right\};$$

$\arg \min_{x \in \mathcal{Q}} f(x)$
dom f

элемент множества $\text{Arg} \min_{x \in \mathcal{Q}} f(x)$;
эффективное множество функции
 $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$:

$$\text{dom } f = \{x \in \mathbb{R}^n: f(x) < +\infty\};$$

$\mathcal{L}_f(\mu)$

нижнее множество Лебега (множество Лебега, множество уровня) для функции $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$:

$$\mathcal{L}_f(\mu) = \{x \in \mathbb{R}^n: f(x) \leq \mu\}, \mu \in \mathbb{R};$$

$f'(x_0)$

градиент (Гато, Фреше) функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ в точке $x_0 \in \mathbb{R}^n$;

$f'(x_0)[e]$

производная функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ в точке $x_0 \in \mathbb{R}^n$ по направлению e ;

$\frac{\partial f(x_0)}{\partial \mathbf{x}_k}$

частная производная функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ в точке $x_0 \in \mathbb{R}^n$:

$$\frac{\partial f(x_0)}{\partial \mathbf{x}_k} = f'(x_0)[e_k];$$

$f'(x_0, e)$

односторонняя производная функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ в точке $x_0 \in \mathbb{R}^n$ по направлению e ;

$F'(x_0)$ матрица Якоби для функции $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ в точке $x_0 \in \mathbb{R}^n$:

$$F'(x_0) = \left(\frac{\partial F_i(x_0)}{\partial \mathbf{x}_j} \right) \in \mathbb{R}^{m \times n};$$

$f''(x_0)$ матрица Гессе для функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ в точке $x_0 \in \mathbb{R}^n$:

$$f''(x_0) = \left(\frac{\partial^2 f(x_0)}{\partial \mathbf{x}_i \partial \mathbf{x}_j} \right) \in \mathbb{R}^{n \times n};$$

$\partial f(x_0)$ субдифференциал функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ в точке $x_0 \in \mathbb{R}^n$;

$\partial_F^+ f(x_0)$ супердифференциал Фреше функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ в точке $x_0 \in \mathbb{R}^n$;

$\partial_F^- f(x_0)$ субдифференциал Фреше функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ в точке $x_0 \in \mathbb{R}^n$;

$\partial \mathcal{Q}$ граница множества $\mathcal{Q} \subset \mathbb{R}^n$;

$\text{cl } \mathcal{Q}$ замыкание множества $\mathcal{Q} \subset \mathbb{R}^n$;

$\text{int } \mathcal{Q}$ внутренность множества $\mathcal{Q} \subset \mathbb{R}^n$;

$\varrho(x, \mathcal{Q})$ функция расстояния от точки $x \in \mathbb{R}^n$ до множества $\mathcal{Q} \subset \mathbb{R}^n$:

$$\varrho(x, \mathcal{Q}) = \inf_{a \in \mathcal{Q}} \|x - a\|;$$

$\text{diam } \mathcal{Q}$ диаметр множества $\mathcal{Q} \subset \mathbb{R}^n$:

$$\text{diam } \mathcal{Q} = \sup_{x, y \in \mathcal{Q}} \|x - y\|;$$

$P_{\mathcal{Q}}x$ метрическая проекция точки $x \in \mathbb{R}^n$ на множество $\mathcal{Q} \subset \mathbb{R}^n$:

$$P_{\mathcal{Q}}x = \{a \in \mathcal{Q}: \|x - a\| = \varrho(x, \mathcal{Q})\};$$

$\delta_{\mathcal{Q}}(\cdot)$ модуль выпуклости множества $\mathcal{Q} \subset \mathbb{R}^n$;

$\text{co } \mathcal{Q}$ выпуклая оболочка множества $\mathcal{Q} \subset \mathbb{R}^n$;

$\text{lin } \mathcal{Q}$ линейная оболочка множества $\mathcal{Q} \subset \mathbb{R}^n$;

$\text{extr } \mathcal{Q}$ крайние точки множества $\mathcal{Q} \subset \mathbb{R}^n$;

$\text{rectr } \mathcal{Q}$ крайние лучи множества $\mathcal{Q} \subset \mathbb{R}^n$;

$s(p, \mathcal{Q})$	опорная функция множества $\mathcal{Q} \subset \mathbb{R}^n$ в точке p :
	$s(p, \mathcal{Q}) = \sup_{x \in \mathcal{Q}} (p, x);$
$h(\mathcal{P}, \mathcal{Q})$	расстояние в метрике Хаусдорфа между множествами $\mathcal{P}, \mathcal{Q} \subset \mathbb{R}^n$;
$N(\mathcal{Q}, x)$	нормальный конус к выпуклому замкнутому множеству \mathcal{Q} в точке $x \in \mathcal{Q}$;
$N_P(\mathcal{Q}, x)$	конус проксимальных нормалей к замкнутому множеству \mathcal{Q} в точке $x \in \mathcal{Q}$;
$O^+ \mathcal{Q}$	асимптотический (рецессивный) конус замкнутого выпуклого множества $\mathcal{Q} \subset \mathbb{R}^n$;
$T(x_0)$	касательное подпространство в задаче условного экстремума;
\mathbb{P}	вероятность события;
\mathbb{E}	математическое ожидание;
$\mathbb{E}_{\sim \xi}$	математическое ожидание по распределению случайной величины ξ ;
$\mathcal{U}(\mathcal{Q})$	равномерное распределение на измеримом ограниченном множестве \mathcal{Q} ;
$\mathcal{N}_n(\mu, \Sigma)$	многомерное нормальное распределение в \mathbb{R}^n со средним μ и ковариационной матрицей Σ .

Глава 1

Безусловная оптимизация

1.1. Схемы методов оптимизации. Скорость сходимости

Пусть \mathbb{R}^n — n -мерное вещественное евклидово пространство со скалярным произведением $(x, y) = x^T y$ и нормой $\|x\| = \sqrt{(x, x)}$.

Под задачей оптимизации будем понимать поиск минимума числовой функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$, заданной на подмножестве конечномерного евклидового пространства $\mathcal{Q} \subset \mathbb{R}^n$.

Точка $x_* \in \mathcal{Q}$ называется локальным минимумом (максимумом), если для любой точки $x \in \mathcal{Q} \cap \{x \in \mathbb{R}^n : \|x - x_*\| \leq \delta\}$, $\delta > 0$, выполнено неравенство $f(x_*) \leq f(x)$ ($f(x_*) \geq f(x)$).

Локальный минимум (максимум) называется строгим, если неравенство $f(x_*) < f(x)$ ($f(x_*) > f(x)$) выполнено для всех точек $x \in \mathcal{Q} \cap \{x \in \mathbb{R}^n : 0 < \|x - x_*\| \leq \delta\}$.

Глобальным минимумом (максимумом) функции f на множестве \mathcal{Q} называется такая точка $x_* \in \mathcal{Q}$, что $f(x_*) \leq f(x)$ ($f(x_*) \geq f(x)$) для всех $x \in \mathcal{Q}$.

В случае $\mathcal{Q} = \mathbb{R}^n$ задача называется задачей безусловной оптимизации (она рассматривается в данной главе), а при $\mathcal{Q} \neq \mathbb{R}^n$ будем говорить о задаче условной оптимизации (ей посвящена глава 2).

Итак, рассмотрим задачу

$$\inf_{x \in \mathbb{R}^n} f(x) \quad (1.1.1)$$

безусловной минимизации функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$.

Задача максимизации функции f может быть сведена к задаче

(1.1.1) с учетом

$$\sup_{x \in \mathbb{R}^n} f(x) = -\inf_{x \in \mathbb{R}^n} (-f(x)).$$

Определим множество минимумов функции f как

$$\Omega = \operatorname{Arg} \min_{x \in \mathbb{R}^n} f(x) = \{x \in \mathbb{R}^n : f(x) = f_*\},$$

где $f_* = \inf_{x \in \mathbb{R}^n} f(x)$. В дальнейшем, если не оговорено обратное, мы предполагаем, что $f_* \in \mathbb{R}$ и $\Omega \neq \emptyset$. Произвольную точку множества Ω будем обозначать через $x_* \in \Omega$.

Под *методом* (или *алгоритмом*) *оптимизации* понимается способ решения задачи (1.1.1). Цель метода оптимизации — получить достаточно хорошее приближение $\hat{x} \in \mathbb{R}^n$ к некоторому решению задачи (1.1.1) по точке:

$$\hat{x} \approx x_*$$

или по функции:

$$\hat{x}: f(\hat{x}) \approx f_*.$$

Метод оптимизации повторяет действия, называемые *шагами*, или *итерациями*. Стартуя с заданной начальной точки $x_1 \in \mathbb{R}^n$, алгоритм на каждом шаге вычисляет новую точку, формируя последовательность $x_1, x_2, \dots, x_k, \dots$. Остановка алгоритма осуществляется при достижении заданного критерия, например, точности $\varepsilon > 0$:

$$\varrho(x_k, \Omega) = \inf_{y \in \Omega} \|x_k - y\| \leq \varepsilon, \quad (1.1.2)$$

$$f(x_k) - f_* \leq \varepsilon. \quad (1.1.3)$$

В качестве решения принимается значение $\hat{x} = x_k$.

Критерий остановки по условию (1.1.2) редко используется на практике (для него требуется заранее знать множество Ω решений задачи) и служит для анализа методов оптимизации. При этом минимум функции иногда бывает известен, и проблема состоит в нахождении самой точки минимума. Так происходит, например, при формулировке задачи решения системы уравнений $F(x) = 0$, где $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $m \leq n$, в виде задачи оптимизации

$$\min_{x \in \mathbb{R}^n} \|F(x)\|^2,$$

которая при разрешимой системе уравнений имеет минимум, равный нулю.

Часто используются косвенные критерии приближения к минимуму, такие как достижение малого значения градиента $f'(x_k) \approx 0$ или стабилизация итерационного процесса в r последних итерациях, то есть при малом расстоянии $\|x_k - x_{k-i}\|$ или малом изменении функции $|f(x_k) - f(x_{k-i})|$, $1 \leq i \leq r \leq k$. Можно также ограничиться фиксированным числом итераций или останавливать работу алгоритма по истечении определенного времени, выдавая в качестве решения последнюю или наилучшую точку.

Выделим несколько типичных схем (типов) методов оптимизации.

1. Методы «направление + шаг», или одношаговые методы:

$$x_{k+1} = x_k + \alpha_k p_k, \quad k = 1, 2, \dots \quad (1.1.4)$$

В этой схеме каждую итерацию метода можно разделить на два независимых этапа: выбор направления p_k и выбор длины шага α_k . Выбор длины шага является задачей одномерной оптимизации, для которой существуют эффективные методы решения. Этот подход является самым распространенным способом получения очередной точки.

2. Методы на основе модельной функции:

$$x_{k+1} = \arg \min \varphi_k(x), \quad \varphi_k(x) \approx f(x), \quad k = 1, 2, \dots,$$

где функции $\varphi_k(x)$ аппроксимируют («моделируют») минимизируемую функцию в окрестностях соответствующих точек x_k . При этом предполагается, что функции φ_k достаточно просты, и вспомогательные задачи оптимизации $\min \varphi_k(x)$ легко решаются. Часто в качестве $\varphi_k(x)$ используют квадратичные модельные функции.

3. Многошаговые методы:

$$x_{k+1} = g(x_k, x_{k-1}, \dots), \quad k = 2, 3, \dots \quad (1.1.5)$$

Здесь очередная точка формируется с помощью некоторой функции g , которая, например, может использовать несколько предыдущих точек.

Из определения инфимума всегда существует минимизирующая последовательность $\{x_k\}_{k=1}^{\infty}$, такая, что

$$f(x_1) \geq f(x_2) \geq \dots \geq f(x_k) \geq \dots$$

Монотонность последовательности $\{f(x_k)\}$ не гарантирует сходимости к минимуму в смысле формул (1.1.2) и (1.1.3). Сходимость последовательности $\{x_k\}_{k=1}^{\infty}$ надо исследовать отдельно. Напомним, что

монотонно убывающая последовательность $\{f(x_k)\}_{k=1}^{\infty}$ либо сходится, либо имеет предел $-\infty$, если f не ограничена снизу.

Скорость сходимости метода оптимизации описывает зависимость расстояния до решения $\varrho(x_k, \Omega)$ или же точности решения $f(x_k) - f_*$ от номера итерации k .

Рассмотрим числовую последовательность $\{r_k\}$, $r_k \rightarrow +0$. Если в качестве r_k выступает расстояние до решения $r_k = \varrho(x_k, \Omega)$, то говорят о *сходимости по точке*, а если точность $r_k = f(x_k) - f_*$, — то о *сходимости по функции*.

Квадратичная скорость сходимости соответствует неравенству

$$r_{k+1} \leq C_2 r_k^2, \quad k \geq k_0,$$

где константы $C_2 > 0$ и $k_0 \in \mathbb{N}$ таковы, что $C_2 r_{k_0} < 1$. Это самая быстрая практическая скорость сходимости, ей соответствует дважды экспоненциальное уменьшение $r_k \leq \frac{1}{C_2} (C_2 r_{k_0})^{(2^{k-k_0})}$ при $k \geq k_0$.

При *линейной скорости сходимости* происходит кратное уменьшение величины r_k на каждом шаге:

$$r_{k+1} \leq C_1 r_k, \quad 0 < C_1 < 1, \quad k \geq 1.$$

Ее иногда называют *сходимостью со скоростью геометрической прогрессии*, так как предыдущее соотношение можно записать в виде

$$r_k \leq C_1^{k-1} r_1.$$

Сублинейная скорость сходимости соответствует соотношениям вида

$$r_k \leq \frac{C_0}{k^p}, \quad C_0 > 0, \quad k \geq 1, \quad p > 0.$$

1.2. Методы нулевого порядка

Во многих задачах оптимизируемая функция задается не явно, а с помощью некоторой процедуры вычисления ее значений. Вид процедуры может быть неизвестен: это может быть алгоритм, вычисление с помощью модели или натурного эксперимента на реальном объекте и т. п., либо она может быть столь сложна, что аналитическое вычисление градиента невозможно. Во всех этих случаях единственная информация, которая известна, — это значения функции $f(x)$ в конкретной точке, а сама функция представляет собой так называемый «черный ящик». Методы, использующие только эту информацию, называются

методами нулевого порядка (*прямыми методами, безградиентными методами*, или методами без вычисления производной).

Методы оптимизации нулевого порядка можно условно разделить на два класса: переборные и поисковые. Идея *переборных методов* заключается в вычислении функции на множестве точек и выборе среди них минимального значения. Такой перебор разумен только на ограниченном множестве, и одномерная минимизация на отрезке — как единственное исключение в рамках безусловной минимизации — будет рассмотрена ниже. Эта задача очень важна как вспомогательная, например, для выбора длины шага в схеме (1.1.4). *Поисковые методы* формируют последовательность точек, сходящихся к минимуму, возможно, локальному.

1.2.1. Одномерная оптимизация и выпуклость

Рассмотрим задачу минимизации функции $f(x)$ на отрезке $[a, b]$:

$$\min_{x \in [a, b]} f(x).$$

Простейший способ решить ее — выбрать N точек на отрезке $[a, b]$ и в качестве решения задачи выбрать *рекордное значение*

$$\hat{f} = \min_{1 \leq i \leq N} f(x_i)$$

в соответствующей *рекордной точке* $x_{\hat{i}}$, где $\hat{i} = \arg \min_{1 \leq i \leq N} f(x_i)$.

Мелкостью набора точек $\{x_i\}$ на отрезке $[a, b]$ называется величина

$$\rho = \max_{x \in [a, b]} \min_{1 \leq i \leq N} |x - x_i|.$$

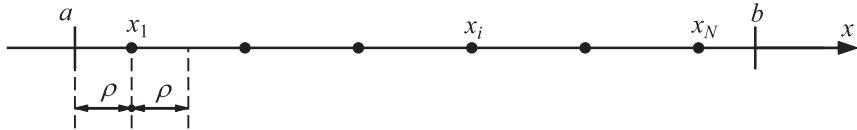
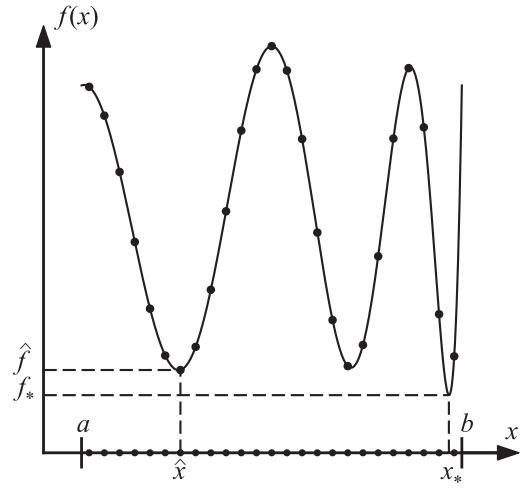
Без предположений о функции разумно распределить N точек на отрезке $[a, b]$ равномерно (см. рис. 1.2.1):

$$x_i = a + \left(i - \frac{1}{2}\right) \frac{b - a}{N}, \quad i = 1, \dots, N.$$

Для получившейся так называемой *равномерной сетки* из N точек на отрезке $[a, b]$ мелкость равна

$$\rho = \frac{b - a}{2N}. \tag{1.2.1}$$

На равномерной сетке достигается минимальная мелкость среди всех сеток с N точками. Отметим, что без знания свойств функции рекордные значения могут быть далеки от истинных, см. рис. 1.2.2 и 1.2.3.

Рис. 1.2.1. Равномерная сетка с мелкостью ρ на отрезке $[a, b]$.Рис. 1.2.2. Пример многоэкстремальной функции, у которой точка \hat{x} , соответствующая рекордному значению \hat{f} , находится далеко от точки минимума x_* .

Если функция f удовлетворяет условию Липшица¹

$$|f(x) - f(y)| \leq L|x - y| \quad \forall x, y \in [a, b],$$

то «скорость» изменения функции характеризуется константой Липшица L . Если функция дифференцируема, то это условие означает ограниченность производной $|f'(x)| \leq L$ для всех $x \in [a, b]$.

Если функция f липшицева с константой L , то по рекордному значению \hat{f} и мелкости ρ можно оценить минимум f_* :

$$\hat{f} \leq f_* + \rho L,$$

¹Любая липшицева функция непрерывна. Существуют непрерывные функции, у которых ни в одной точке нет производной; см., например, [25, теорема 7.18]. Такие функции не являются липшицевыми.

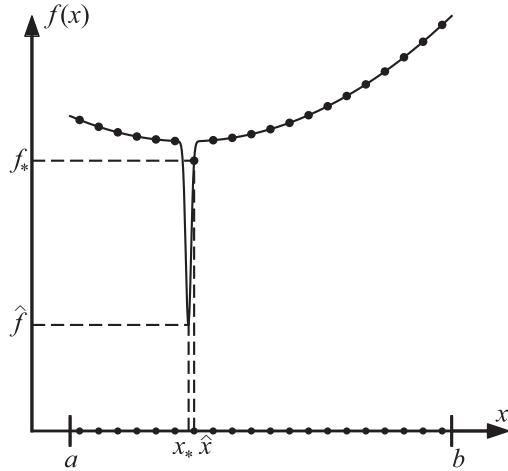


Рис. 1.2.3. Пример функции с большой константой Липшица L , у которой рекордное значение \hat{f} значительно отличается от минимума f_* .

так что

$$\hat{f} - \rho L \leq f_* \leq \hat{f}.$$

Из обратно пропорциональной зависимости мелкости равномерно распределенных точек (1.2.1) от их числа N следует очевидное соображение: чем больше пробных точек взять, тем выше получится точность. Подчеркнем, что оценить расстояние от рекордной точки до точки минимума без дополнительной информации о функции принципиально невозможно, см. рис. 1.2.2.

Напомним определение выпуклой функции, данное в приложении Г.1; см. рис. 1.2.4:

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \quad \forall \lambda \in [0, 1] \quad \forall x_1, x_2 \in [a, b].$$

Оказывается, что для выпуклых функций множество минимумов (если оно непусто) задачи (1.1.1) устроено очень просто. При этом легко построить простые алгоритмы для решения задачи (1.1.1).

Упражнение 1.2.1. Пусть функция $f: \mathbb{R} \rightarrow \mathbb{R}$ выпукла. Докажите, что любой ее локальный минимум является глобальным.

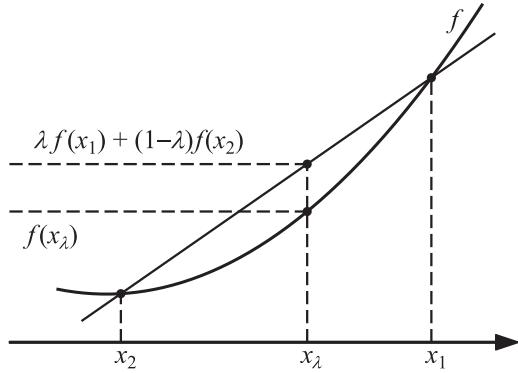


Рис. 1.2.4. Свойство выпуклой функции для промежуточных точек
 $x_\lambda = \lambda x_1 + (1 - \lambda)x_2, \lambda \in [0, 1]$.

Упражнение 1.2.2. Докажите, что при $\lambda \notin (0, 1)$ выпуклая функция f не меньше линейной функции:

$$\begin{aligned} f(\lambda x_1 + (1 - \lambda)x_2) &\geq \\ &\geq \lambda f(x_1) + (1 - \lambda)f(x_2) \quad \forall \lambda \leq 0 \text{ или } \lambda \geq 1, \forall x_1, x_2 \in [a, b]. \end{aligned}$$

Более слабое свойство функции, показанное на рис. 1.2.5 (справа), называется *квазивыпуклостью*, см. также пример 1.5.2:

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \max\{f(x_1), f(x_2)\} \quad \forall \lambda \in [0, 1] \quad \forall x_1, x_2 \in [a, b].$$

Понятия выпуклости и квазивыпуклости функций естественным образом обобщаются на многомерный случай, см. раздел Г.

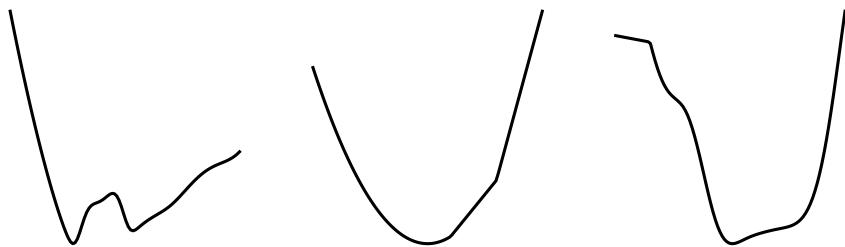


Рис. 1.2.5. Примеры невыпуклой (слева), выпуклой (посередине) и квазивыпуклой (справа) функций.

Упражнение 1.2.3. Покажите, что все выпуклые функции являются квазивыпуклыми.

Рассмотрим множества Лебега

$$\mathcal{L}_f(\mu) \doteq \{x \in [a, b]: f(x) \leq \mu\}$$

для функции $f(x)$.

Множество $\mathcal{Q} \in \mathbb{R}^n$ называется *выпуклым*, если выполнено включение

$$(1 - \lambda)x_1 + \lambda x_2 \in \mathcal{Q} \quad \forall x_1, x_2 \in \mathcal{Q} \quad \forall \lambda \in [0, 1],$$

см. раздел Г.1. Выпуклость множества \mathcal{Q} означает, что любой отрезок с концами $x_1, x_2 \in \mathcal{Q}$ содержится в \mathcal{Q} .

Для квазивыпуклой функции множества Лебега выпуклы, то есть в одномерном случае являются промежутками, см. рис. 1.2.6. Если функция к тому же полунепрерывна снизу, то все эти множества замкнуты, см. теорему А.3. Напомним, что замкнутое выпуклое множество на прямой есть либо пустое множество, либо точка, либо отрезок.

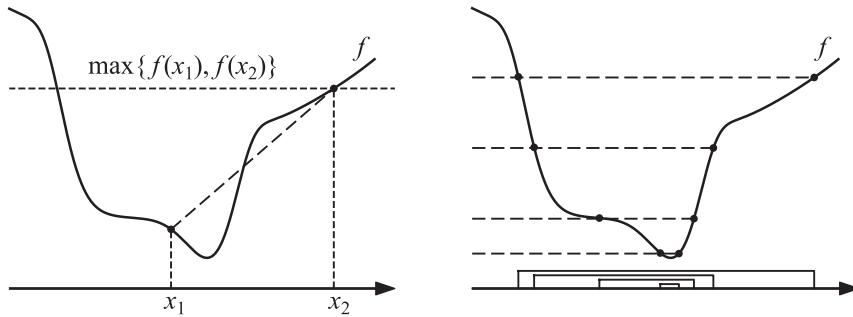


Рис. 1.2.6. Невыпуклость (слева) и множества Лебега (справа) квазивыпуклой функции.

Упражнение 1.2.4. Если множество Лебега состоит из одной точки, то как это множество связано с минимумом функции? Для каких значений параметра μ множество Лебега $\mathcal{L}_f(\mu)$ непусто?

Упражнение 1.2.5. Где находится максимум (квази)выпуклой функции на отрезке?

Упражнение 1.2.6. Приведите пример квазивыпуклой функции $f: [a, b] \rightarrow \mathbb{R}$, $a, b \in \mathbb{R}$, у которой некоторое лебегово множество незамкнуто.

1.2.2. Метод золотого сечения

Если функция выпуклая, то для ее минимизации эффективно применять метод золотого сечения. Он часто используется, например, для выбора длины шага в схеме (1.1.4).

Лемма 1.2.1. *Пусть $f: \mathcal{Q} \rightarrow \mathbb{R}$ — непрерывная выпуклая функция. Если для трех произвольных точек $a < x < b$, $a, x, b \in \mathcal{Q} \subset \mathbb{R}$, в средней точке значение функции не больше, чем в крайних, то есть*

$$f(x) \leq f(a), \quad f(x) \leq f(b),$$

то отрезок $[a, b]$ содержит точку минимума $x_ \in \operatorname{Arg} \min_{x \in \mathbb{R}} f(x)$.*

Упражнение 1.2.7. *Докажите лемму 1.2.1, используя результат упражнения 1.2.2.*

Отметим, что выпуклая функция $f: \mathbb{R} \rightarrow \mathbb{R}$ непрерывна, см. теорему Г.2. Из этой же теоремы следует, что функция, определенная на замкнутом множестве $\mathcal{Q} \neq \mathbb{R}$, может иметь разрыв только на границе.

Упражнение 1.2.8. *Предложите примеры разрывных выпуклых функций $f: [a, b] \rightarrow \mathbb{R}$, имеющих точку минимума. Предложите примеры разрывных выпуклых функций $f: [a, b] \rightarrow \mathbb{R}$, не имеющих точек минимума.*

Лемма 1.2.1 верна и для разрывных выпуклых функций f при условии, что множество точек минимума $\operatorname{Arg} \min_x f(x)$ непусто. Чтобы не оговаривать этот случай отдельно, далее будем рассматривать непрерывные выпуклые функции.

Если значения $f(x_1)$ и $f(x_2)$ выпуклой функции $f(x)$ вычислены в двух внутренних точках $a < x_1 < x_2 < b$ отрезка $[a, b]$, то из геометрически очевидного соображения (см. упражнение 1.2.2) можно понять, где находится минимум. А именно, если $f(x_1) > f(x_2)$, то точка минимума x_0 не может быть на полуинтервале $[a, x_1]$ и не может быть на полуинтервале $(x_2, b]$, если $f(x_1) < f(x_2)$, см. рис. 1.2.7. В случае $f(x_1) = f(x_2)$ хотя бы одна из точек минимума принадлежит $[x_1, x_2]$.

Таким образом, после вычисления двух значений функции область локализации минимума выпуклой функции может быть уменьшена. Это рассуждение верно для трисекции отрезка любыми двумя точками. Если брать точки x_1, x_2 вблизи середины $\frac{a+b}{2}$ отрезка $[a, b]$, то отрезок может быть сокращен почти вдвое.

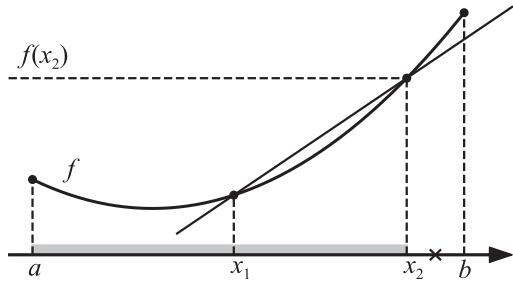


Рис. 1.2.7. Идея трисекции отрезка для выпуклой функции.

Эффективно выбирать точки x_1 и x_2 так, чтобы на следующем шаге производить не два вычисления функции f , а одно. Этого можно достичь выбором специального соотношения между отрезками, которые образуются после отсечения. Идея состоит в том, что после первого шага алгоритма оставшаяся из двух точек x_1 или x_2 попадет внутрь области локализации решения и будет делить эту область в том же соотношении, см. рис. 1.2.8.

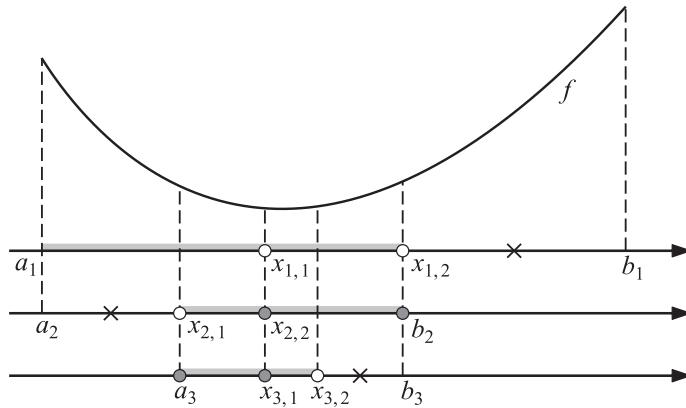


Рис. 1.2.8. Метод золотого сечения. Выколоты точки отрезка, в которых происходит вычисление функции.

Обозначим отношение отрезков через $\psi = \frac{x_2-a}{b-a}$ с симметричным расположением точек на отрезке, $x_1 - a = b - x_2$. Тогда условие сохра-

нения этого отношения можно записать в виде

$$\psi = \frac{x_2 - a}{b - a} = \frac{x_1 - a}{x_2 - a} = \frac{b - x_2}{x_2 - a} = \frac{b - a}{x_2 - a} - 1 = \frac{1}{\psi} - 1.$$

Полученное уравнение

$$\psi^2 + \psi - 1 = 0$$

имеет единственное решение меньше единицы:

$$\psi = \frac{\sqrt{5} - 1}{2} = 0,61803\dots$$

Точка x_1 располагается симметрично: $x_1 = a + (1 - \psi)(b - a)$.

Обратное отношение большего отрезка к меньшему $\frac{b-a}{x_2-a} = \frac{1}{\psi}$ соответствует «золотому сечению» $\varphi = \frac{1+\sqrt{5}}{2} = 1,61803\dots$, что и дало название методу.

Алгоритм 1 (Метод золотого сечения).

1. Задать точность $\varepsilon > 0$. Задать исходный интервал

$$[a_1, b_1] = [a, b],$$

инициализировать счетчик итераций $k = 1$.

2. На отрезке $[a_1, b_1]$ выбрать точки $x_1 = a_1 + (1 - \psi)(b_1 - a_1)$ и $x_2 = a + \psi(b_1 - a_1)$ и вычислить $f(x_1), f(x_2)$.
3. Если $f(x_1) > f(x_2)$, то:
 - обновить интервал:

$$[a_{k+1}, b_{k+1}] \leftarrow [x_1, b_k];$$

- обновить левую внутреннюю точку:

$$x_1 \leftarrow x_2;$$

- выбрать правую внутреннюю точку:

$$x_2 \leftarrow a_{k+1} + \psi(b_{k+1} - a_{k+1})$$

и вычислить $f(x_2)$.

4. Если $f(x_1) \leq f(x_2)$, то:

- обновить интервал:

$$[a_{k+1}, b_{k+1}] \leftarrow [a_k, x_2];$$

- обновить правую внутреннюю точку:

$$x_2 \leftarrow x_1;$$

- выбрать левую внутреннюю точку:

$$x_1 \leftarrow a_{k+1} + (1 - \psi)(b_{k+1} - a_{k+1})$$

и вычислить $f(x_1)$.

5. Если $b_{k+1} - a_{k+1} \leq 2\varepsilon$, то завершить работу алгоритма и вернуть в качестве решения середину отрезка:

$$\hat{x} = \frac{b_{k+1} + a_{k+1}}{2},$$

в противном случае увеличить счетчик итераций k и перейти к шагу 3.

Метод золотого сечения находит точку \hat{x} , находящуюся рядом с минимумом в пределах требуемой точности $\varrho(\hat{x}, \Omega) \leq \varepsilon$.

После каждой итерации алгоритма размер области локализации решения $[a_k, b_k]$ уменьшается в ψ раз, поэтому после k итераций имеем

$$b_{k+1} - a_{k+1} = \psi(b_k - a_k) = \psi^k(b - a),$$

то есть скорость сходимости линейная. Для метода золотого сечения зависимость между числом шагов и точностью имеет явный вид. После N шагов будет получена гарантированная точность

$$\varepsilon_N = \frac{b_{N+1} - a_{N+1}}{2} = \frac{1}{2}(b - a)\psi^N.$$

И обратно, для достижения точности ε необходимо сделать

$$N_\varepsilon = \left\lceil \log_\psi \frac{\varepsilon}{b - a} \right\rceil = \left\lceil \log_{1/\psi} \frac{b - a}{\varepsilon} \right\rceil \approx \left\lceil 0,4812 \ln \frac{b - a}{\varepsilon} \right\rceil$$

шагов.

Упражнение 1.2.9. Приведите пример квазивыпуклой функции, минимум которой не удается найти методом золотого сечения.

Указание: Продемонстрировать случай, при котором область локализации решения нельзя уменьшить, то есть нельзя исключить ни один из отрезков $[a, x_1]$ и $(x_2, b]$.

Упражнение 1.2.10. Какое дополнительное требование следует предъявить к квазивыпуклой функции, чтобы метод золотого сечения сходился, как и для выпуклой функции?

1.2.3. Выбор длины шага при минимизации выпуклых функций в \mathbb{R}^n

Если функция $f(x)$, $x \in \mathbb{R}^n$, выпуклая, то в схеме «направление + шаг» (1.1.4) функция $g(\alpha) = f(x_k + \alpha p_k)$ тоже будет выпуклой. Эта функция уже зависит от скалярного параметра — длины шага α .

Предлагается выбирать «оптимальную» длину шага из условия минимума функции вдоль выбранного направления:

$$x_{k+1} = x_k + \alpha_* p_k, \quad (1.2.2)$$

$$\alpha_* \in \operatorname{Arg} \min_{\alpha \in \mathbb{R}} g(\alpha) = \operatorname{Arg} \min_{\alpha \in \mathbb{R}} f(x_k + \alpha p_k). \quad (1.2.3)$$

Пусть функция $g(\alpha)$, определенная на \mathbb{R} , *коэрцитивна*, то есть $g(\alpha) \rightarrow \infty$ при $|\alpha| \rightarrow \infty$. Из теоремы Г.2 и коэрцитивности следует, что функция g непрерывна и множество ее точек минимума непусто.

Для применения метода одномерной минимизации нулевого порядка (например, метода золотого сечения) необходимо найти отрезок $[a, b]$, содержащий минимум функции $g(\alpha)$:

$$[a, b] \cap \operatorname{Arg} \min_{\alpha \in \mathbb{R}} g(\alpha) \neq \emptyset.$$

В таком случае подзадача (1.2.3) превращается в задачу одномерной минимизации на отрезке:

$$\min_{\alpha \in [a, b]} g(\alpha) = \min_{\alpha \in [a, b]} f(x_k + \alpha p_k).$$

Для выпуклой функции g , согласно лемме 1.2.1, достаточно найти три точки такие, что в средней точке значение функции будет меньшим, чем в крайних. Если вдоль направления p_k выпуклая функция f убывает, то с удвоением шага $\alpha_j = \alpha_0 \sum_{i=0}^j 2^i = \alpha_0(2^{j+1} - 1)$, $j = 0, 1, \dots$, будет обнаружен интервал, содержащий минимум; см. рис. 1.2.9. Увеличение шага останавливается при выполнении условия

$$f(x_k + \alpha_j p_k) \geq g(x_k + \alpha_{j-1} p_k).$$

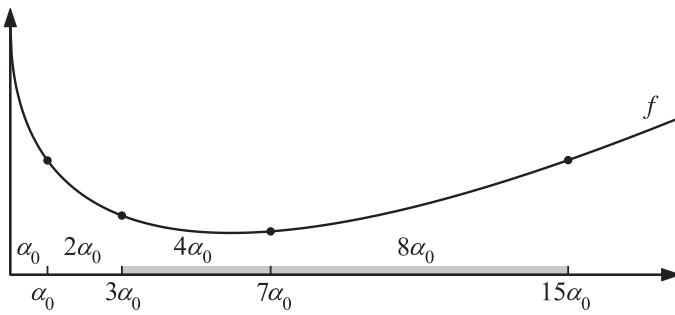


Рис. 1.2.9. Нахождение интервала, содержащего минимум выпуклой одномерной коэрцитивной функции $g: \mathbb{R} \rightarrow \mathbb{R}$.

Приведем в общем виде соответствующий алгоритм для выпуклой функции $g: \mathbb{R} \rightarrow \mathbb{R}$.

Алгоритм 2 (Поиск отрезка, содержащего минимум одномерной выпуклой функции).

1. Выбрать начальную длину шага $\alpha_0 > 0$.
2. Если $g(0) < g(\alpha_0)$, то перейти к шагу 4. В противном случае положить $k = 1$.
3. Для $\alpha_k = \alpha_0(2^{k+1} - 1)$ вычислить $g(\alpha_k)$.
Если $g(\alpha_k) \geq g(\alpha_{k-1})$, то вернуть интервал

$$[\alpha_0(2^{k-1} - 1), \alpha_0(2^{k+1} - 1)]$$

в качестве искомого и завершить работу.

Если $g(\alpha_k) < g(\alpha_{k-1})$, то увеличить счетчик итераций k и повторить шаг 3.

4. Если $g(0) < g(-\alpha_0)$, то вернуть интервал $[-\alpha_0, \alpha_0]$ в качестве искомого и завершить работу. В противном случае положить $k = 1$.
5. Для $\alpha_k = \alpha_0(2^{k+1} - 1)$ вычислить $g(-\alpha_k)$.
Если $g(-\alpha_k) \leq g(-\alpha_{k-1})$, то вернуть интервал

$$[-\alpha_0(2^{k+1} - 1), -\alpha_0(2^{k-1} - 1)]$$

в качестве искомого и завершить работу.

Если $g(-\alpha_k) < g(-\alpha_{k-1})$, то увеличить счетчик итераций k и повторить шаг 5.

Если p_k — направление убывания функции f в точке x_k , то можно ограничиться шагами 1–3 алгоритма с $\alpha \geq 0$. Для дифференцируемой функции в качестве направления убывания используется антиградиент $p_k = -f'(x_k)$, задающий направление наискорейшего спуска. В этом случае итерации (1.2.2), (1.2.3) называются методом *полной релаксации*. Однако на практике для дифференцируемых функций вспомогательная подзадача одномерной минимизации (1.2.3) не решается, а сразу происходит переход к новой точке. Исключением является метод сопряженных градиентов (см. раздел 1.3.5), для которого нахождение минимума одномерной функции $g(\alpha)$ существенно.

1.2.4. Метод случайного блуждания

В *методе случайного блуждания* итерации производятся по схеме (1.1.4), в которой выбор направления осуществляется случайным образом, а длина шага находится из условия

$$f(x_k + \alpha p_k) < f(x_k),$$

например, с помощью одномерной минимизации. Если такую точку найти не удается, то шаг не делается, а выбирается новое направление p_k .

Если функция дифференцируема, то выбор направления можно делать множеством способов на основе градиента f' (например, по антиградиенту $p_k = -f'(x_k)$). С помощью градиента можно характеризовать *допустимые направления* p_k как удовлетворяющие неравенству

$$(p_k, f'(x_k)) < 0. \quad (1.2.4)$$

Согласно определению градиента, для любого допустимого направления можно выбрать такой шаг, для которого $f(x_k + \alpha p_k) < f(x_k)$.

Для недифференцируемой функции или функции, градиент которой недоступен, в качестве направлений можно использовать:

- орты e_j (так называемый *покоординатный спуск*), перебирая их циклически:

$$p_k = e_j, \quad j = (k \bmod n) + 1$$

($k \bmod n$ означает остаток от деления числа k на число n). Покоординатный спуск иногда можно осуществить в явном виде.

- разностные аппроксимации антиградиента, например, симметричную аппроксимацию

$$p_k = - \sum_{i=1}^n \frac{f(x_k + \gamma_k e_i) - f(x_k - \gamma_k e_i)}{2\gamma_k} e_i,$$

с параметрами $\gamma_k > 0$ (метод Кифера–Вольфовича), она требует $2n$ вычислений функции; или одностороннюю аппроксимацию, которая требует $n + 1$ вычисление функции:

$$p_k = - \sum_{i=1}^n \frac{f(x_k + \gamma_k e_i) - f(x_k)}{\gamma_k} e_i.$$

Коэффициенты γ_k можно брать не общими для всех ортов, а различными для каждой из координат — в виде $\gamma_{k,i}$.

Можно выбирать направление случайным образом. Напомним, что $s \sim \mathcal{U}(\mathcal{Q})$ означает равномерное распределение случайной величины s на ограниченном множестве \mathcal{Q} . В частности, если $\mathcal{Q} = \{q_1, q_2, \dots, q_N\}$, то

$$\mathbb{P}(s = q_j) = \frac{1}{N}, \quad j = 1, \dots, N.$$

Если $\mathcal{Q} = [a, b]$ — числовой отрезок, то плотность равномерного распределения равна

$$\mathbf{p}(t) = \begin{cases} \frac{1}{b-a}, & a \leq t \leq b, \\ 0, & t < a \text{ или } t > b, \end{cases}$$

и

$$\mathbb{P}(s \leq t) = \int_{-\infty}^t \mathbf{p}(\tau) d\tau = \begin{cases} 0, & t \leq a, \\ \frac{t-a}{b-a}, & a < t \leq b, \\ 1, & t > b. \end{cases}$$

Многомерное нормальное распределение случайной величины $r \in \mathbb{R}^n$ со средним μ и ковариационной матрицей $\Sigma \succ 0$ обозначается через $r \sim \mathcal{N}_n(\mu, \Sigma)$ и задается плотностью

$$\mathbf{p}(r) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma}} e^{-\frac{1}{2}(r-\mu)^T \Sigma^{-1} (r-\mu)}.$$

Можно предложить два способа выбора случайного направления:

— случайный покоординатный спуск:

$$p_k = s_k e_{j_k}, \quad s_k \sim \mathcal{U}(\{-1, 1\}), \quad j_k \sim \mathcal{U}(\{1, 2, \dots, n\}). \quad (1.2.5)$$

— метод случайного блуждания:

$$p_k \sim \mathcal{U}(\{p \in \mathbb{R}^n : \|p\| = 1\}).$$

Равномерное распределение на евклидовом шаре можно получить с помощью нормально распределенной величины r_k :

$$p_k = \frac{r_k}{\|r_k\|}, \quad r_k \sim \mathcal{N}_n(0, I). \quad (1.2.6)$$

Метод (случайного) покоординатного спуска обладает следующими особенностями. Если функция дифференцируема и $f'_j(x_k) \neq 0$, то направление $-\text{sign}(f'_j(x_k))e_j$ вдоль соответствующей оси является допустимым в смысле (1.2.4). Если x_k не является точкой локального минимума, то из условия (1.2.4) следует, что случайно выбранное направление (1.2.6) с вероятностью $\frac{1}{2}$ является направлением убывания. Для таких точек хотя бы одна из компонент градиента $f'_j(x_k)$ ненулевая, и, таким образом, одно из направлений $\pm e_j$ будет направлением убывания функций и вдоль него можно совершить координатный спуск.

Однако метод покоординатного спуска может не обнаружить направление убывания для недифференцируемых функций, «застряв» таким образом в точке, не являющейся точкой минимума.

Пример 1.2.1. Рассмотрим функцию $f(x) = |\mathbf{x}_1 - \mathbf{x}_2| + \frac{1}{5}|\mathbf{x}_1 + \mathbf{x}_2|$, минимум которой равен нулю при $x = (0; 0)$. Она недифференцируема на прямых $\mathbf{x}_1 = -\mathbf{x}_2$ и $\mathbf{x}_1 = \mathbf{x}_2$.

Рассмотрим точки на прямой $\mathbf{x}_1 = \mathbf{x}_2$. Легко видеть, что для всех таких точек во всех четырех возможных направлениях покоординатного спуска $\pm e_1, \pm e_2$ функция возрастает. Можно предположить, что при численной реализации метода покоординатного спуска попасть в такую точку практически невозможно, однако это не так. Например, если для начальной точки $(1; 2)$ изменять только первую координату \mathbf{x}_1 , то есть делать шаг вдоль направления убывания $e_1 = (1, 0)$, то минимум вдоль этого направления будет достигнут в точке $(2; 2)$. Аналогично при изменении второй координаты \mathbf{x}_2 минимум вдоль направления $-e_2 = (0, -1)$ достигается в точке $(1; 1)$, см. рис. 1.2.10. Это же верно для любой другой начальной точки.

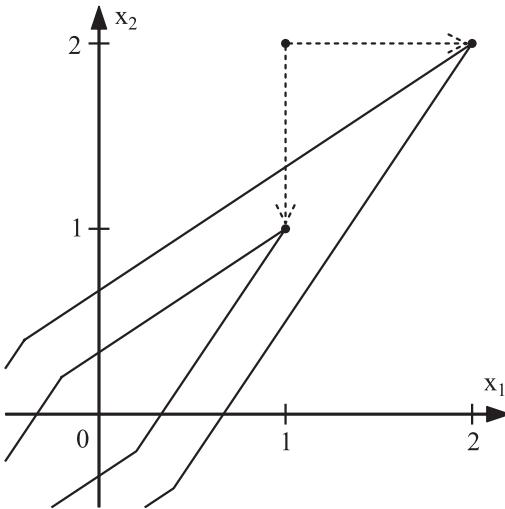


Рис. 1.2.10. Демонстрация проблемы покоординатного спуска для недифференцируемых функций.

Таким образом, при выборе длины шага по методу полной релаксации метод покоординатного спуска будет остановлен после первого же шага. ▼

Для выпуклых недифференцируемых функций можно использовать направление субградиента, однако не каждый субградиент является направлением убывания, см. пример на рис. 1.3.5. В таких случаях можно искать направление убывания случайным образом в соответствии с (1.2.6).

1.2.5. Метод имитации отжига

Одной из проблем оптимизации невыпуклых функций является многоэкстремальность (см. рис. 1.2.3): методы оптимизации, основанные на построении минимизирующей последовательности $\{f(x_k)\}$, могут сходится к локальному минимуму \tilde{x} :

$$\exists \varepsilon > 0: f(x) \geq f(\tilde{x}) \quad \forall x \in \{x \in \mathbb{R}^n : \|x - \tilde{x}\| \leq \varepsilon\}.$$

При этом в точке локального минимума направления убывания функции отсутствуют.

Идея *метода имитации отжига* состоит в том, чтобы иногда разрешать увеличение значений функции, что позволяет выйти из локальных минимумов.

Алгоритм 3 (Метод имитации отжига).

1. Выбрать начальную точку x_1 , длину шага $\alpha > 0$, нормировочный коэффициент C и последовательность «обобщенных температур» $\{T_i\}$, $T_i > 0$, $i = 1, 2, \dots$. Инициировать счетчик итераций $k = 1$ и вычислить $f(x_1)$.
2. Выбрать случайное направление

$$p_k = \frac{r_k}{\|r_k\|}, \quad r_k \sim \mathcal{N}_n(0, I),$$

и точку-кандидат

$$y = x_k + \alpha p_k.$$

3. Если $f(y) < f(x_k)$, то перейти в эту точку безусловно:

$$x_{k+1} \leftarrow y,$$

увеличить счетчик итераций k и перейти к шагу 5.

4. Если $f(y) \geq f(x_k)$ то с вероятностью $P_k = e^{-\frac{f(y)-f(x_k)}{T_k}}$ перейти в точку y :

$$\mathbb{P}(x_{k+1} \leftarrow y) = P_k.$$

В противном случае остаться на месте $(x_{k+1} \leftarrow x_k)$ и увеличить счетчик итераций k .

5. Проверить критерий остановки; если он выполняется, то завершить работу, вернув x_k в качестве решения. В противном случае вернуться к шагу 2.

Критерием остановки может быть достижение заранее заданного числа шагов N либо «застревание» в точке локального минимума, если в течение нескольких шагов x_k не обновляется.

Последовательность «обобщенных температур» T_i выбирается монотонно убывающей, например $T_i = \frac{T_1}{i}$ или $T_i = 1 - \frac{i}{N}$ (если число шагов N задано). Чем меньше T_i , тем меньше вероятность перехода в новую точку с увеличением функции, что соответствует эмпирическому правилу «охлаждения», стабилизации процесса.

Вероятностный переход в шаге 4 реализуется с помощью генерации

случайной величины $u_k \sim \mathcal{U}([0, 1])$, равномерно распределенной на отрезке $[0, 1]$, и проверки ее попадания на отрезок $[0, P_k]$: если $u_k \leq P_k$, то переход делается. Функция вычисления вероятности перехода к «худшей» точке может быть выбрана и иной, но с соблюдением следующего правила: чем больше разность $f(y) - f(x_k)$, тем меньше вероятность перехода в y .

Вероятность $e^{-\frac{f(y)-f(x_k)}{T_k}}$ является вероятностью перехода медленно останавливающей равновесной физической системы в состояние с большей свободной энергией Гиббса. Этот процесс, называемый процессом отжига, и дал название методу.

В отличие от рассмотренного выше в разделе 1.2.2 *детерминированного метода золотого сечения*, поведение метода имитации отжига, как и методов случайного покоординатного спуска (1.2.5) и метода случайного поиска (1.2.6), определяется случайными величинами. Методы такого рода называются *стохастическими*, или *случайными*. С одной стороны, для них невозможно предсказать результат вычислений, а свойства сходимости и скорость сходимости формулируются в терминах математического ожидания и пр. С другой стороны, недетерминированность метода (если она внесена искусственно, она иногда называется *рандомизацией*) часто приводит к успеху.

Для улучшения и большей устойчивости стохастических методов используются два приема: запуск из разных точек и многократный запуск метода. При этом метод можно независимо (параллельно) запускать на разных вычислительных устройствах, что существенно увеличивает шансы найти глобальный минимум.

Наконец, отметим, что метод имитации отжига успешно применяется в дискретных и дискретно-непрерывных задачах оптимизации. В них переменная x уже не обязательно представляет собой числовой вектор, а может описывать некоторую конфигурацию или состояние оптимизируемой системы, качество которой характеризуется функцией $f(x)$. Для применения метода имитации отжига надо немного изменить шаг 2 алгоритма 3. В качестве точки-кандидата y надо выбрать некоторую соседнюю, «близкую» точку/состояние/конфигурацию.

Способ выбора точки-кандидата и понятие «блзкого» определяется из предметных соображений в конкретной задаче.

1.2.6. Метод Нелдера – Мида

Этот метод использует алгоритмический подход к вычислению очередной точки (1.1.5). Идея состоит в систематическом поиске области,

где значения функции f уменьшаются. Основная проблема — выбор направления/следующей точки — решается с помощью анализа значений функции в *нескольких* точках.

Формируется набор из $n+1$ точек $x_i \in \mathbb{R}^n$, не лежащих в одной гиперплоскости. В n -мерном пространстве они являются вершинами некоторого симплекса с ненулевым объемом. Шаг метода состоит в изменении одной из точек таким образом, чтобы уменьшать *максимальное* значение функции f среди всех точек. При успешном шаге в наборе появляются новая точка, заменяющая точку с максимальным значением функции. Геометрически эти изменения соответствуют отражению, растяжению и сжатию симплекса. Ожидаемое монотонное уменьшение функции f на каждом шаге на практике приводит к сходимости метода. При этом для многоэкстремальных функций имеет место сходимость к локальному минимуму.

Существуют несколько вариаций метода Нелдера–Мида; в основном они отличаются способом изменения симплекса при неудачном пробном шаге (см. шаги 4^г и 5).

Алгоритм 4 (Метод Нелдера–Мида).

1. Параметры алгоритма:
коэффициент отражения $\alpha > 0$,
коэффициент сжатия $\beta \in (0, 1)$,
коэффициент растяжения $\gamma > 1$,
коэффициент гомотетии $\delta \in (0, 1)$.

Выбираются $n+1$ начальные точки — вершины симплекса x_i , $i = 1, \dots, n+1$, в них вычисляются значения функции

$$f_1 = f(x_1), \dots, f_{n+1} = f(x_{n+1}).$$

2. Среди вершин выбираются две с наибольшим и одна с наименьшим значениями функции f :

$$x_h, x_g, x_\ell: \quad f_h \geq f_g \geq \dots \geq f_\ell.$$

Если все значения одинаковые ($f(x_h) = f(x_\ell)$), то дальнейшая работа метода невозможна, можно вернуть любую из вершин симплекса в качестве решения.

Вычисляется центр тяжести всех точек, кроме «наихудшей»

точки x_h :

$$x_c = \frac{1}{n} \sum_{i \neq h} x_i.$$

3. Точка x_h отражается относительно x_c с коэффициентом α :

$$x_r = x_c + \alpha(x_c - x_h) = (1 + \alpha)x_c - \alpha x_h$$

и в ней вычисляется функция $f_r = f(x_r)$.

4. Изменение симплекса.

Проверка местоположения величины f_r в ряду f_h, f_g, f_ℓ (где $f_h \geq f_g > f_\ell$).

- a) Если $f_\ell \geq f_r$, то направление $x_c - x_h$ удачное и можно嘗試 увеличить шаг. Формируется пробная точка

$$x_e = x_c + \gamma(x_r - x_c) = (1 - \gamma)x_c + \gamma x_r$$

(растяжение) и вычисляется $f(x_e)$.

- если $f_r \geq f_e$, то расширяем симплекс, меняя вершину x_h на x_e (при этом обновляется порядок вершин: $x_h \leftarrow x_g, x_\ell \leftarrow x_e$). Переходим на шаг 6;
- если $f_e > f_r$, то расширение не подходит, меняем вершину x_h на x_r (при этом обновляется порядок вершин: $x_h \leftarrow x_g, x_\ell \leftarrow x_r$). Переходим на шаг 6.

- b) Если $f_g > f_r > f_\ell$, то меняем вершину x_h на x_r (при этом обновляется порядок вершин: $x_h \leftarrow x_g$ и меняется x_g). Переходим на шаг 6.
- c) Если $f_h > f_r \geq f_g$, то меняем вершину на $x_h \leftarrow x_r$. Переходим на шаг сжатия 5.

- g) Если $f_r \geq f_h$, то переходим на шаг сжатия 5.

5. Формируется пробная точка

$$x_s = x_c + \beta(x_h - x_c) = \beta x_h + (1 - \beta)x_c$$

(сжатие) и вычисляется $f_s = f(x_s)$.

- a) Если $f_h \geq f_s$, то меняем вершину x_h на x_s .
- b) Если $f_s > f_h$, то выполняется сжатие всего симплекса (гомотетия к вершине x_ℓ):

$$x_i \leftarrow x_\ell + \delta(x_i - x_\ell).$$

6. Проверка сходимости:

- а) если размер симплекса достаточно мал — например,
по критерию

$$\max_i \left\| x_i - \frac{1}{n+1} \sum_{j=1}^{n+1} x_j \right\|,$$

то закончить работу и вернуть в качестве решения

$$x_{\min} = x_\ell;$$

- б) в противном случае вернуться к шагу 2.

Замечание 1.2.1. После сжатия симплекса на шаге 56 можно дополнительно проверять уменьшение максимального значения функции в его вершинах $\max_i f(x_i)$ относительно максимального значения исходного симплекса f_h . Если оно не уменьшается, то следует выбирать меньший коэффициент δ , так как для непрерывных функций при $\delta \rightarrow 0$ имеем $x_i \rightarrow x_\ell$ и $f(x_i) \rightarrow f(x_\ell) < f_h$. ▼

Метод Нелдера–Мида отличают простота алгоритма и нетребовательность к ресурсам, однако его поведение существенно зависит от выбора начального симплекса и констант.

1.3. Методы первого порядка

1.3.1. Метод градиентного спуска

Рассмотрим функцию $f: \mathbb{R}^n \rightarrow \mathbb{R}$ с L_1 -липшицевым градиентом. Поскольку функция в точке $x \in \mathbb{R}^n$ локально убывает в направлении антиградиента $-f'(x)$, то понятен способ действия — двигаться по антиградиенту. Этот способ принято называть *градиентным спуском*. Из точки x нужно перейти к точке $y = x - \alpha f'(x)$. Параметр $\alpha > 0$ называется *шагом* метода. Он должен выбираться не слишком большим, чтобы не оказаться далеко от точки x , где направление убывания f может уже не совпадать с направлением вектора $-f'(x)$.

В то же время шаг α не должен быть слишком малым, так как в этом случае новая точка будет долго оставаться в малой окрестности точки x , см. рис. 1.3.1.

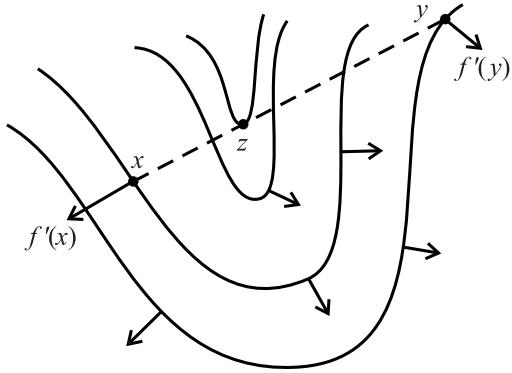


Рис. 1.3.1. Градиентный спуск. Изображены линии уровня, функция f возрастает по направлению нормалей к ним. Из точки x имеет смысл отходить по лучу $\{x - \alpha f'(x): \alpha \geq 0\}$ не далее чем до точки z . Далее, например в точке y , значение функции $f(y) > f(z)$, а направление убывания $-f'(y)$ в точке y не совпадает с вектором $-f'(x)$.

Рассмотрим оптимальный способ действий. Выберем произвольно $x_1 \in \mathbb{R}^n$ и определим последовательность

$$x_{k+1} = x_k - \alpha_k f'(x_k), \quad \alpha_k > 0. \quad (1.3.1)$$

Из условия верхней квадратичной аппроксимации для функции f (см. оценку (A.5)) имеем

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + (f'(x_k), x_{k+1} - x_k) + \frac{L_1}{2} \|x_{k+1} - x_k\|^2 = \\ &= f(x_k) - \alpha_k \|f'(x_k)\|^2 + \frac{\alpha_k^2 L_1}{2} \|f'(x_k)\|^2, \end{aligned}$$

откуда

$$f(x_{k+1}) \leq f(x_k) - \left(\alpha_k - \frac{\alpha_k^2 L_1}{2} \right) \|f'(x_k)\|^2. \quad (1.3.2)$$

Легко видеть, что в рамках формулы (1.3.2), не имея дополнительной информации о функции, надо взять $\alpha_k = \frac{1}{L_1}$ для всех k , чтобы гарантированно минимизировать значение $f(x_{k+1})$ в формуле (1.3.2) по сравнению с $f(x_k)$. При $\alpha_k = \frac{1}{L_1}$ получаем

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L_1} \|f'(x_k)\|^2. \quad (1.3.3)$$

В случае выпуклой функции, у которой множество Лебега

$$\{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$$

ограничено, градиентный спуск имеет сублинейную скорость сходимости по функции:

$$f(x_k) - f_* \leq \frac{C}{k}$$

для некоторой константы $C > 0$, которая зависит от функции f ; напомним, что $f_* = \min_{x \in \mathbb{R}^n} f(x)$. Мы докажем более общий факт в теореме 2.3.1, поэтому здесь доказательство не приводим.

Упражнение 1.3.1. Рассмотрите функции $f_m(x) = x^{2m}$, $m \in \mathbb{N}$. Все они выпуклые, имеют глобальный минимум в точке $x_* = 0$ и имеют непрерывный по Липшицу градиент в любой ограниченной окрестности нуля. Для $x_1 > 0$ рассмотрите итерации (1.3.1) для некоторого $\alpha_k = \alpha > 0$ и покажите, что скорость сходимости $\|x_k - x_*\| \rightarrow 0$ при минимизации функций f_m может быть сколь угодно медленной.

Рассмотрим сильно выпуклую функцию $f: \mathbb{R}^n \rightarrow \mathbb{R}$ (см. определение Г.10 в разделе Г.6). Покажем, что для такой функции градиентный спуск сходится к глобальному минимуму f с линейной скоростью.

Упражнение 1.3.2. Докажите, что для сильно выпуклой функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ существует единственная точка $x_* \in \mathbb{R}^n$ такая, что $f(x_*) = \inf_{x \in \mathbb{R}^n} f(x)$. Верно ли это для строго выпуклой функции?

Теорема 1.3.1. Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R}$ — функция с L_1 -липшицевым градиентом¹ и сильно выпуклая с константой $\varkappa > 0$. Пусть $f_* = \min_{x \in Q} f(x)$. Тогда при любом выборе $x_1 \in \mathbb{R}^n$ алгоритм (1.3.1) с $\alpha_k = 1/L_1$ сходится с линейной скоростью по функции и по точке:

$$f(x_{k+1}) - f_* \leq \left(1 - \frac{\varkappa}{2L_1}\right)^k (f(x_1) - f_*),$$

$$\|x_{k+1} - x_k\| \leq \sqrt{\frac{2}{L_1} \left(1 - \frac{\varkappa}{2L_1}\right)^{k-1} (f(x_1) - f_*)}.$$

¹ Поскольку непрерывность по Липшицу градиента Гато влечет существование градиента Фреше, то для липшицево дифференцируемой функции f градиент f' может пониматься и в смысле Гато, и в смысле Фреше, см. раздел А.3.

Доказательство. Из формулы (1.3.3) имеем

$$\|f'(x_k)\|^2 \leq 2L_1(f(x_k) - f(x_{k+1})).$$

Из условия сильной выпуклости (Г.21) вытекает неравенство

$$\|f'(x_k)\|^2 \geq \varkappa(f(x_k) - f_*).$$

Из двух последних оценок получаем

$$\varkappa(f(x_k) - f_*) \leq 2L_1((f(x_k) - f_*) - (f(x_{k+1}) - f_*)),$$

или

$$f(x_{k+1}) - f_* \leq \left(1 - \frac{\varkappa}{2L_1}\right)(f(x_k) - f_*). \quad (1.3.4)$$

Из формулы (1.3.4) вытекает сходимость со скоростью геометрической прогрессии.

Для сходимости по точке воспользуемся оценкой

$$\|x_{k+1} - x_k\|^2 = \frac{1}{L_1^2} \|f'(x_k)\|^2$$

и в силу (1.3.3) имеем

$$\begin{aligned} \|x_{k+1} - x_k\|^2 &= \frac{1}{L_1^2} \|f'(x_k)\|^2 \leq \frac{2}{L_1} (f(x_k) - f(x_{k+1})) \leq \frac{2}{L_1} (f(x_k) - f_*), \\ \|x_{k+1} - x_k\| &\leq \sqrt{\frac{2}{L_1} \left(1 - \frac{\varkappa}{2L_1}\right)^{k-1} (f(x_1) - f_*)}. \end{aligned}$$

Теорема доказана. ■

Упражнение 1.3.3. Докажите, что последовательность $\{x_k\}$ из теоремы 1.3.1 сходится к некоторому решению, то есть что $x_k \rightarrow x_* \in \Omega$.

Упражнение 1.3.4. Допустим, что в условиях теоремы 1.3.1 выбран постоянный шаг $\alpha_k = \alpha \in (0, \frac{1}{L_1}]$. Повторите рассуждения теоремы 1.3.1 и получите оценки скорости сходимости по функции и по точке.

Метод градиентного спуска можно применять к дифференцируемым функциям, однако доказать гарантированную скорость сходимости этого метода в общем случае не удается.

1.3.2. Условие Лежанского – Поляка – Лоясевича

Как следует из вышеизложенного, для выпуклых функций существуют эффективные методы поиска безусловного минимума. Но уже на заре зарождения численных методов минимизации функций возник вопрос об отказе от выпуклости. Достаточно эффективным (и используемым до сих пор) оказался следующий подход. Известно, что сильно выпуклые функции обладают многими полезными для методов оптимизации свойствами. Осталось только заметить, какие аналитические свойства сильно выпуклых функций «отвечают» за хорошую работу соответствующего алгоритма, вычленить эти свойства и рассмотреть уже не обязательно выпуклые функции, а обладающие указанными свойствами.

Оказывается, что одним из таких свойств является условие (Г.21). В 1963 году математики Т. Лежанский [105], Б.Т. Поляк [60] и С. Лоясевич [108] независимо (и из разных посылок) пришли к следующему свойству функции.

Определение 1.3.1. Функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ удовлетворяет *условию Лежанского – Поляка – Лоясевича* (или *условию LPL*), если f дифференцируема по Фреше, множество $\Omega = \text{Arg } \min_{x \in \mathbb{R}^n} f(x)$ ее глобальных минимумов непусто и существует константа $\mu > 0$ такая, что для всех $x \in \mathbb{R}^n$

$$\|f'(x)\|^2 \geq \mu(f(x) - f(x_*)) \quad \forall x_* \in \Omega,$$

или

$$\|f'(x)\|^2 \geq \mu(f(x) - f_*), \quad f_* = f(\Omega).$$

Следуя математической номенклатуре [99], мы часто будем называть свойство из определения 1.3.1 *свойством* или *условием Поляка – Лоясевича*, хотя все три автора пришли к формулировке определения 1.3.1 независимо.

Из формулы (Г.21) вытекает, что дифференцируемая сильно выпуклая функция обладает свойством Поляка – Лоясевича с константой $\mu = \varkappa$. Однако существуют и невыпуклые функции, удовлетворяющие определению 1.3.1.

Упражнение 1.3.5. Пусть функция f удовлетворяет определению 1.3.1. Докажите, что если $f'(x) = 0$ для некоторого $x \in \mathbb{R}^n$, то x – точка глобального минимума f .

Упражнение 1.3.6. Приведите пример функции $f: \mathbb{R} \rightarrow \mathbb{R}$, которая удовлетворяет определению 1.3.1 и не является выпуклой.

Упражнение 1.3.7. Верно ли, что всякая функция $f: \mathbb{R} \rightarrow \mathbb{R}$, которая удовлетворяет определению 1.3.1, является выпуклой в окрестности точки минимума (как на рис. 1.3.2)?

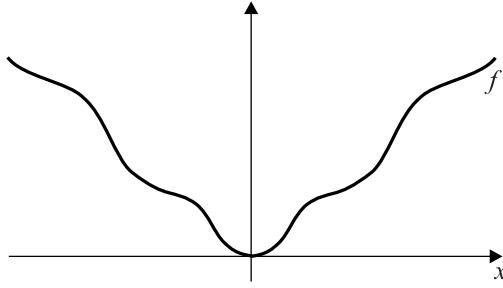


Рис. 1.3.2. Функция, удовлетворяющая условию LPL.

Выберем произвольно $x_1 \in \mathbb{R}^n$ и определим последовательность

$$x_{k+1} = x_k - \alpha_k f'(x_k), \quad \alpha_k > 0. \quad (1.3.5)$$

При $\alpha_k = \frac{1}{L_1}$, как показано в разделе 1.3.1, получаем

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L_1} \|f'(x_k)\|^2. \quad (1.3.6)$$

Теорема 1.3.2. Предположим, что $f: \mathbb{R}^n \rightarrow \mathbb{R}$ — функция с L_1 -липшицевым градиентом, удовлетворяющая условию Поляка — Лоясевича с константой $\mu > 0$. Тогда при любом выборе $x_1 \in \mathbb{R}^n$ алгоритм (1.3.5) с $\alpha_k = 1/L_1$ сходится с линейной скоростью по функции и по точке:

$$\begin{aligned} f(x_{k+1}) - f_* &\leq \left(1 - \frac{\mu}{2L_1}\right)^k (f(x_1) - f_*), \quad f_* = \min_{x \in \mathbb{R}^n} f(x), \\ \|x_{k+1} - x_k\| &\leq \sqrt{\frac{2}{L_1} \left(1 - \frac{\mu}{2L_1}\right)^{k-1} (f(x_1) - f_*)}. \end{aligned}$$

Доказательство. Из формулы (1.3.6) имеем

$$\|f'(x_k)\|^2 \leq 2L_1(f(x_k) - f(x_{k+1})).$$

Из условия LPL следует, что

$$\|f'(x_k)\|^2 \geq \mu(f(x_k) - f_*).$$

Из двух последних оценок получаем неравенство

$$\mu(f(x_k) - f_*) \leq 2L_1 \left((f(x_k) - f_*) - (f(x_{k+1}) - f_*) \right),$$

или

$$f(x_{k+1}) - f_* \leq \left(1 - \frac{\mu}{2L_1}\right) (f(x_k) - f_*). \quad (1.3.7)$$

Из формулы (1.3.7) вытекает сходимость со скоростью геометрической прогрессии.

Для сходимости по точке воспользуемся равенством

$$\|x_{k+1} - x_k\|^2 = \frac{1}{L_1^2} \|f'(x_k)\|^2$$

и в силу (1.3.6) имеем

$$\begin{aligned} \|x_{k+1} - x_k\|^2 &= \frac{1}{L_1^2} \|f'(x_k)\|^2 \leq \frac{2}{L_1} (f(x_k) - f(x_{k+1})) \leq \frac{2}{L_1} (f(x_k) - f_*), \\ \|x_{k+1} - x_k\| &\leq \sqrt{\frac{2}{L_1} \left(1 - \frac{\mu}{2L_1}\right)^{k-1} (f(x_1) - f_*)}, \end{aligned}$$

что и завершает доказательство. ■

Упражнение 1.3.8. Пусть в условиях теоремы 1.3.2 выбран постоянный шаг $\alpha_k = \alpha \in (0, \frac{1}{L_1}]$. Повторите рассуждения, приведенные в доказательстве теоремы 1.3.2, и получите оценки скорости сходимости по функции и по точке.

Упражнение 1.3.9. Пусть Ω — множество глобальных минимумов f и $\{x_k\}$ — последовательность из теоремы 1.3.2. Докажите, что найдутся точка $x_* \in \Omega$ и числа $C > 0$, $q \in (0, 1)$ такие, что $\|x_k - x_*\| \leq Cq^k$.

Пример 1.3.1. Рассмотрим систему уравнений $g(x) = 0$. Здесь $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $m \leq n$, — непрерывно дифференцируемая вектор-функция с липшицевым градиентом. Будем считать, что $\text{rank } g'(x) = m$ для всех x . Через $g'(x) = \left(\frac{\partial g_i(x)}{\partial x_j}\right)_{1 \leq i \leq m, 1 \leq j \leq n}$ обозначена матрица Якоби. Пусть система $g(x) = 0$ имеет решение. Задача состоит в поиске корня уравнения $g(x) = 0$.

Рассмотрим $f(x) = \frac{1}{2} \|g(x)\|^2$ с градиентом $f'(x) = g(x)g'(x)$ (строка $g(x)$ умножается на матрицу Якоби $g'(x)$). Для каждого x квадратичная форма $k(h) = (h, g'(x)g'^T(x)h)$ строго положительно определена

в силу условия полного ранга g' . Пусть $x_1 \in \mathbb{R}^n$ — первая точка итераций (1.3.5) и

$$\min_{x \in B_R(x_1)} \sigma(g'(x)g'^T(x)) = \lambda_0 > 0,$$

где $\sigma(g'(x)g'^T(x))$ — спектр матрицы $g'(x)g'^T(x)$ и L_1 — константа Липшица $f'(x) = g(x)g'(x)$ на $B_R(x_1)$. В этом случае для всех $x \in B_R(x_1)$ $\|f'(x)\|^2 = g(x)g'(x)g'^T(x)g^T(x) \geq \lambda_0\|g\|^2 = 2\lambda_0 f(x)$, то есть функция f удовлетворяет условию LPL с константой $\mu = 2\lambda_0$. Предположим, что

$$R \geq \frac{C}{1-q}, \quad C = \sqrt{\frac{2}{L_1} f(x_1)}, \quad q = \sqrt{1 - \frac{\lambda_0}{L_1}}. \quad (1.3.8)$$

Упражнение 1.3.10. Покажите, что при указанном выше соотношении между R , λ_0 , L_1 и x_1 каждая точка x_k алгоритма (1.3.5) содержится в шаре $B_R(x_1)$.

Указание. Докажите, что итерации $\{x_k\}$, соответствующие процедуре (1.3.5) с $\alpha_k = 1/L_1$, сходятся к точке минимума x_* , где $f(x_*) = 0$. При этом для всякого номера $m \geq 1$ в силу теоремы 1.3.2

$$\|x_m - x_1\| \leq \sum_{k=1}^{m-1} \|x_{k+1} - x_k\| \leq \sum_{k=1}^{\infty} Cq^{k-1} \leq R.$$

Сравните результат с упражнением 1.3.9.

В силу теоремы 1.3.2 с $\mu = 2\lambda_0$ итерации $\{x_k\}$ (1.3.5) для нахождения корня $g(x) = 0$ сходятся к корню с линейной скоростью. При этом условие (1.3.8) выполнено при выборе такой начальной точки $x_1 \in \mathbb{R}^n$, что $f(x_1) \leq \frac{L_1 R^2}{2}(1-q)^2$. ▼

Рассмотрим случайный покоординатный спуск (1.2.5). Предположим, что функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ для каждого вектора e_i ортонормированного базиса удовлетворяет условию: существует такое число $L_1 > 0$, что для всех $1 \leq i \leq n$, $x \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$, выполнено неравенство

$$f(x + \alpha e_i) \leq f(x) + \alpha \frac{\partial f(x)}{\partial x_i} + \frac{L_1}{2} \alpha^2. \quad (1.3.9)$$

Напомним, что $\frac{\partial f(x)}{\partial x_i} = \lim_{t \rightarrow 0} \frac{f(x+te_i) - f(x)}{t}$. Функция с L_1 -липшицевым градиентом удовлетворяет условию (1.3.9).

Теорема 1.3.3 ([99]). Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ удовлетворяет условию (1.3.9), а также условию Поляка – Лоясевича с константой $\mu > 0$. Рассмотрим покоординатный спуск

$$x_k \in \mathbb{R}^n, \quad \alpha \in \left(0, \frac{2}{L_1}\right), \quad x_{k+1} = x_k - \alpha \frac{\partial f(x_k)}{\partial \mathbf{x}_{i_k}} e_{i_k}. \quad (1.3.10)$$

Если выбирать номер i_k равномерно распределенным на $\{1, \dots, n\}$, то алгоритм (1.3.10) имеет линейную скорость сходимости в смысле математического ожидания:

$$\mathbb{E}(f(x_k) - f_*) \leq \left(1 - \frac{\mu q(\alpha)}{n}\right)^{k-1} (f(x_1) - f_*), \quad (1.3.11)$$

$$\text{где } f_* = \min_{x \in \mathbb{R}^n} f(x) \text{ и } q(\alpha) = \alpha - \frac{L_1}{2} \alpha^2.$$

Доказательство. Аналогично доказательству формулы (1.3.2)

$$f(x_{k+1}) \leq f(x_k) - q(\alpha) \left\| \frac{\partial f(x_k)}{\partial \mathbf{x}_{i_k}} \right\|^2.$$

Взяв математическое ожидание от обеих частей предыдущего неравенства по i_k , получаем

$$\begin{aligned} \mathbb{E}f(x_{k+1}) &\leq f(x_k) - q(\alpha) \mathbb{E} \left\| \frac{\partial f(x_k)}{\partial \mathbf{x}_{i_k}} \right\|^2 = \\ &= f(x_k) - \frac{q(\alpha)}{n} \sum_{i=1}^n \left\| \frac{\partial f(x_k)}{\partial \mathbf{x}_i} \right\|^2 = f(x_k) - \frac{q(\alpha)}{n} \|f'(x_k)\|^2. \end{aligned}$$

Используя условие LPL и повторяя рассуждения доказательства теоремы 1.3.2, получаем оценку (1.3.11). ■

Отметим, что за счет вычисления на каждом шаге частной производной $\frac{\partial f(x_k)}{\partial \mathbf{x}_i}$ вместо градиента $f'(x_k) = \left(\frac{\partial f(x_k)}{\partial \mathbf{x}_1}, \dots, \frac{\partial f(x_k)}{\partial \mathbf{x}_n}\right)$ общее количество операций для достижения одной точности при случайному покоординатном спуске в ряде случаев меньше, чем при градиентном спуске [117]. Тем не менее заметим, что с ростом n знаменатель прогрессии $1 - \frac{\mu q(\alpha)}{n}$ в оценке (1.3.11) стремится к единице и теоретическая оценка сходимости дает сомнительный прогноз в пространстве \mathbb{R}^n большей размерности.

1.3.3. Метод тяжелого шарика

В градиентном методе на каждом шаге никак не используется информация, полученная на предыдущих итерациях. Естественным желанием является учесть предысторию процесса для ускорения сходимости. Методы, в которых новое приближение x_{k+1} зависит от s предыдущих, то есть

$$x_{k+1} = \varphi(x_k, x_{k-1}, \dots, x_{k-s+1}),$$

называются s -шаговыми. Одним из s -шаговых методов для $s = 2$ является *метод тяжелого шарика*:

$$x_{k+1} = x_k - \alpha f'(x_k) + \beta(x_k - x_{k-1}), \quad (1.3.12)$$

где $\alpha > 0$ и $\beta \geq 0$ — некоторые параметры. В частности, при $\beta = 0$ получаем стандартный градиентный метод.

Своё название метод получил из-за следующей физической аналогии. Движение тела («тяжелого шарика») в потенциальном поле с потенциалом $f(x)$ при наличии вязкости или силы трения описывается дифференциальным уравнением второго порядка:

$$x''(t) = -f'(x(t)) - \mu x'(t). \quad (1.3.13)$$

Ясно, что из-за потери энергии на трение тело в конце концов окажется в точке минимума потенциала $f(x)$. Таким образом, тяжелый шарик решает соответствующую задачу минимизации. Если рассмотреть разностный аналог уравнения (1.3.13), то придем к итерационному методу (1.3.12).

Введение инерционного члена $\beta(x_k - x_{k-1})$ может привести к ускорению сходимости. Это видно, например, из рис. 1.3.3. Вместо зигзагообразного движения в градиентном методе в данном случае получается более плавная траектория по «дну оврага». Эти эвристические соображения подкрепляются следующей теоремой.

Теорема 1.3.4. *Пусть x_* — невырожденная точка минимума функции f , то есть $0 \prec f''(x_*)$. Тогда при $L \geq l > 0$,*

$$0 \leq \beta < 1, \quad 0 < \alpha < \frac{2(1+\beta)}{L}, \quad lI_n \preceq f''(x_*) \preceq LI_n \quad (1.3.14)$$

найдется такое $\varepsilon > 0$, что при любых x_1, x_2 : $\|x_i - x_\| \leq \varepsilon$, $i = 1, 2$, метод (1.3.12) сходится к x_* со скоростью геометрической прогрессии: для всех $k \in \mathbb{N}$ выполнена оценка*

$$\|x_k - x_*\| \leq c(\delta)(q + \delta)^k, \quad 0 \leq q < 1, \quad 0 < \delta < 1 - q. \quad (1.3.15)$$

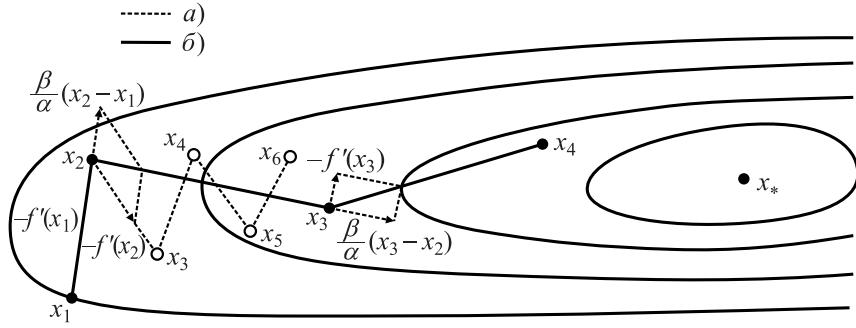


Рис. 1.3.3. Избегание оврагов в методе тяжелого шарика. Траектория а) соответствует методу тяжелого шарика, а траектория б) — обычному градиентному спуску.

Величина q минимальна и равна $q_0 = (\sqrt{L} - \sqrt{l}) / (\sqrt{L} + \sqrt{l})$ при

$$\alpha_0 = \frac{4}{(\sqrt{L} + \sqrt{l})^2}, \quad \beta_0 = q_0^2. \quad (1.3.16)$$

Доказательство. Наметим схему доказательства. Увеличим раз мерность, что позволит свести многошаговый процесс ($s = 2$) к одино шаговому. Введем $2n$ -мерный вектор $z_k = (x_k - x_0, x_{k-1} - x_0)^T$. Тогда итерационный процесс (1.3.12) можно записать в виде

$$z_{k+1} = Az_k + o(z_k),$$

где квадратная $2n \times 2n$ матрица A имеет вид

$$A = \begin{pmatrix} (1 + \beta)I_n - \alpha B & -\beta I_n \\ I_n & 0 \end{pmatrix}, \quad B = f''(x_*).$$

Пусть $l = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n = L$ — собственные значения матрицы B . Тогда собственные значения ϱ_j , $j = 1, \dots, 2n$, матрицы A совпадают с собственными значениями матриц 2×2 вида

$$\begin{pmatrix} 1 + \beta - \alpha \lambda_i & -\beta \\ 1 & 0 \end{pmatrix}.$$

Следовательно, они являются корнями уравнений

$$\varrho^2 - \varrho(1 + \beta - \alpha \lambda_i) + \beta = 0, \quad i = 1, \dots, n. \quad (1.3.17)$$

Можно показать, что при условии (1.3.14) любой корень ϱ любого из уравнений (1.3.17) по модулю меньше 1: $|\varrho| < 1$.

В силу теоремы A.6 о локальной линейной сходимости справедлива оценка (1.3.15). Если вычислить $\min_{\alpha, \beta} \max_{1 \leq j \leq 2n} |\varrho_j|$, то можно получить оптимальное q_0 и соответствующие α_0, β_0 из (1.3.16). ■

Сравним скорость сходимости, даваемую одношаговым и двухшаговым методами при оптимальном выборе параметров. Одношаговый метод дает $q_1 \approx 1 - \frac{l}{2L}$ ($l/2$ равна константе сильной выпуклости \varkappa , а $L = L_1$ — константа Липшица для градиента), тогда как двухшаговый дает $q_2 \approx 1 - 2\sqrt{\frac{l}{L}}$. Таким образом, по порядку величины q двухшаговый метод в $\sqrt{L/l}$ раз быстрее одношагового.

Упражнение 1.3.11. Докажите предыдущее утверждение.

Отметим также, что подбор оптимальных значений α и β в (1.3.12) непрост: формулами (1.3.16) обычно воспользоваться не удается, так как спектр матрицы $f''(x_*)$ априори неизвестен.

1.3.4. Шаг Армихо

В рассмотренных выше разделах шаг метода градиентного спуска выбирался равным $\frac{1}{L_1}$ или меньше этой величины. Однако в реальных задачах значение константы Липшица градиента f' или даже ее разумная оценка сверху бывают неизвестны. В таких случаях можно выбирать шаг градиентного спуска по схеме, которая называется *правилом Армихо*.

Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ дифференцируема. Тогда при малых $\alpha_k > 0$ в силу дифференцируемости функции f имеем

$$f(x_k - \alpha_k f'(x_k)) = f(x_k) - \alpha_k \|f'(x_k)\|^2 + o(\alpha_k), \quad \alpha_k \rightarrow +0.$$

Зафиксируем число $\alpha \in (0, 1)$, число $d > 0$ и $\beta \in (0, 1)$. Если брать в правой части вместо α_k значение $\alpha\alpha_k$, то при достаточно малых α_k из предыдущей формулы вытекает неравенство

$$f(x_k - \alpha_k f'(x_k)) \leq f(x_k) - \alpha\alpha_k \|f'(x_k)\|^2,$$

поскольку $\alpha\alpha_k \|f'(x_k)\|^2 < \alpha_k \|f'(x_k)\|^2 + o(\alpha_k)$ при $\alpha_k \rightarrow +0$.

Но что значит «достаточно малое» α_k и как его найти? Предлагается брать $\alpha_k = d\beta^m$, где $m \in \mathbb{N} \cup \{0\}$ выбирается из условия

$$m = \min\{l = 0, 1, 2, \dots : f(x_k - d\beta^l f'(x_k)) \leq f(x_k) - \alpha d\beta^l \|f'(x_k)\|^2\}.$$

Как видно из вышеизложенного, рано или поздно наступит момент, когда неравенство в определении m станет справедливым, поэтому m определено корректно, см. рис. 1.3.4.

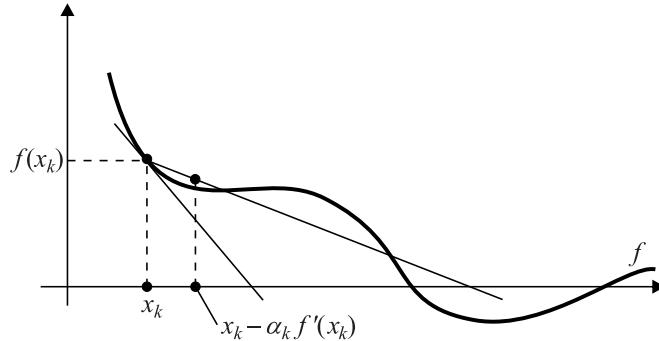


Рис. 1.3.4. Выбор шага α_k Армихо.

Теорема 1.3.5. Предположим, что $f: \mathbb{R}^n \rightarrow \mathbb{R}$ — функция с L_1 -липшицевым градиентом, удовлетворяющая условию Поляка — Лоясевича с некоторой константой μ . Тогда метод градиентного спуска с шагом Армихо сходится с линейной скоростью по функции (1.3.18) и по точке (1.3.19).

Доказательство. Из формулы (1.3.2) получаем, что $\alpha_k = d\beta^{m_k}$, где $m_k = m$ есть минимальное неотрицательное целое число, для которого $\alpha\alpha_k \leq \alpha_k - \frac{L_1}{2}\alpha_k^2$, то есть

$$\alpha d\beta^m \leq d\beta^m - \frac{L_1}{2}d^2\beta^{2m}.$$

Решая последнее неравенство, получаем, что для любого k

$$m_k \leq m = \max \left\{ 0, \left[\log_\beta \frac{2(1-\alpha)}{dL_1} \right] + 1 \right\}.$$

Таким образом, $\alpha_k = d\beta^{m_k} \geq d\beta^m = \alpha_0 > 0$ для всех k .

Из оценки для α_k и формулы (1.3.2) получаем

$$\alpha\alpha_k \|f'(x_k)\|^2 \leq \left(\alpha_k - \frac{L_1}{2}\alpha_k^2 \right) \|f'(x_k)\|^2 \leq f(x_k) - f(x_{k+1}).$$

Из условия Поляка–Лоясевича следует, что

$$\mu(f(x_k) - f_*) \leq \|f'(x_k)\|^2.$$

Из двух последних оценок вытекает сходимость по функции:

$$f(x_{k+1}) - f_* \leq (1 - \mu\alpha\alpha_0)(f(x_k) - f_*). \quad (1.3.18)$$

Сходимость по точке доказывается так же, как в теореме 1.3.2. Пусть $q(\alpha) = \alpha - \frac{L_1}{2}\alpha^2$. Тогда $\alpha\alpha_0 \leq \alpha\alpha_k \leq q(\alpha_k)$ и

$$\begin{aligned} \|x_{k+1} - x_k\|^2 &= \\ &= \alpha_k^2 \|f'(x_k)\|^2 \leq \frac{\alpha_k^2}{q(\alpha_k)} (f(x_k) - f(x_{k+1})) \leq \frac{d^2}{q(\alpha_k)} (f(x_k) - f_*) \end{aligned}$$

и

$$\|x_{k+1} - x_k\| \leq \sqrt{\frac{d^2}{\alpha\alpha_0} (1 - \mu\alpha\alpha_0)^{k-1} (f(x_1) - f_*)}, \quad (1.3.19)$$

что и завершает доказательство. ■

Теорема 1.3.5 показывает, что, если функция имеет непрерывный по Липшицу градиент и удовлетворяет условию Поляка–Лоясевича, то шаг Армихо гарантирует линейную сходимость градиентного спуска. При этом для функционирования алгоритма не важны значения константы Липшица L_1 и константы μ из условия Поляка–Лоясевича. Эти константы фигурируют только в доказательстве сходимости алгоритма.

Как видно из вышеизложенного, шаг Армихо можно применять и для минимизации дифференцируемой функции. Однако оценить скорость сходимости градиентного спуска в этом случае не удается.

1.3.5. Метод сопряженных градиентов

В общем случае двухшаговый инерционный метод имеет непостоянные параметры:

$$x_{k+1} = x_k - \alpha_k f'(x_k) + \beta_k(x_k - x_{k-1}), \quad k = 2, 3, \dots \quad (1.3.20)$$

Их можно выбрать из решения двумерной задачи оптимизации

$$\{\alpha_k, \beta_k\} = \arg \min_{\alpha, \beta} f(x_k - \alpha_k f'(x_k) + \beta_k(x_k - x_{k-1})). \quad (1.3.21)$$

Для случая квадратичной функции

$$f(x) = \frac{1}{2}(Qx, x) - (c, x), \quad Q \succ 0, \quad (1.3.22)$$

эта задача может быть решена явно. Если $f'(x_k) \neq 0$, $k \geq 2$, то

$$\begin{aligned} \alpha_k &= \frac{\|r_k\|^2(Qp_k, p_k) - (r_k, p_k)(Qr_k, p_k)}{(Qr_k, r_k)(Qp_k, p_k) - (Qr_k, p_k)^2}, \\ \beta_k &= \frac{\|r_k\|^2(Qr_k, p_k) - (r_k, p_k)(Qr_k, r_k)}{(Qr_k, r_k)(Qp_k, p_k) - (Qr_k, p_k)^2}, \\ r_k &= f'(x_k) = Qx_k - c, \quad p_k = x_k - x_{k-1}. \end{aligned} \quad (1.3.23)$$

Если $r_k = f'(x_k) = 0$, то x_k — точка минимума $f(x)$.

Для квадратичной функции при специальном выборе второй точки с помощью метода скорейшего спуска

$$\begin{aligned} x_2 &= \arg \min_{x=x_1-\alpha f'(x_1)} f(x), \\ x_2 &= x_1 - \frac{\|r_1\|^2}{(Qr_1, r_1)} r_1, \quad r_1 = f'(x_1) = Qx_1 - c \end{aligned} \quad (1.3.24)$$

метод (1.3.20), (1.3.21) сходится к минимуму за конечное число шагов.

Лемма 1.3.1. Значения градиентов r_1, r_2, \dots в методе (1.3.20), (1.3.23), (1.3.24) попарно ортогональны.

Доказательство. Воспользуемся индукцией по k .

Пусть $(r_i, r_j) = 0$ при $1 \leq i < k$, $k \geq 3$, и $r_i \neq 0$, $i = 1, \dots, k$. Орто-гональность градиентов r_1, r_2, r_3 следует непосредственно из определения метода. Тогда, умножая (1.3.20) слева на Q , получаем

$$r_{i+1} = r_i - \alpha_k Qr_i + \beta_k(r_i - r_{i-1}),$$

Из $r_i \neq 0$ для $i \leq k$ следует, что $\alpha_k \neq 0$. Поэтому Qr_i есть линейная комбинация r_{i+1}, r_i, r_{i-1} , в том числе Qr_k есть линейная комбинация r_{k+1}, r_k, r_{k-1} , и в силу предположения индукции $(Qr_i, r_j) = 0$, $|i - j| > 1$, $i < k$, $j \leq k$. Следовательно, для $i \leq k - 2$ верно

$$(r_{k+1}, r_i) = (r_k - \alpha_k Qr_k + \beta_k(r_k - r_{k-1}), r_i) = 0.$$

Непосредственно из (1.3.20), (1.3.23) следует

$$(r_{k+1}, r_k) = 0, \quad (r_{k+1}, p_k) = 0.$$

Наконец, из (1.3.24) $p_2 = -\frac{\|r_1\|}{(Qr_1, r_1)}r_1$ и из (1.3.20) при $k = 1$ имеем $p_k = -\alpha_{k-1}r_{k-1} + \beta_{k-1}p_{k-1}$.

Последовательно применяя это соотношение, получаем, что p_k есть линейная комбинация r_0, r_1, \dots, r_{k-1} , причем r_{k-1} входит с коэффициентом $-\alpha_{k-1} \neq 0$. Поэтому из $(r_{k+1}, p_k) = 0$ и $(r_{k+1}, r_i) = 0$, $i \leq k-2$, следует $(r_{k+1}, r_{k-1}) = 0$.

Итак, для всех $i \leq k$ имеем $(r_{k+1}, r_i) = 0$, что и завершает доказательство. \blacksquare

Поскольку в \mathbb{R}^n не может существовать более n ортогональных ненулевых векторов, то для некоторого $k \leq n+1$ будет $r_k = 0$. Таким образом, доказана

Теорема 1.3.6. *Метод (1.3.20), (1.3.23), (1.3.24) дает точку минимума квадратичной функции (1.3.22) за число итераций, не превосходящее n .*

Из теоремы Куна–Таккера Д.3 следует, что если L — некоторое подпространство в \mathbb{R}^n и $f(x)$ — выпуклая дифференцируемая функция, то условие

$$(f'(x_0), a) = 0 \text{ для всех } a \in L$$

необходимо и достаточно, чтобы x_0 было минимумом $f(x)$ на L . Отсюда и из леммы 1.3.1 следует, что x_k — точка минимума квадратичной функции $f(x)$ на подпространстве, проходящем через x_1 и порожденном r_1, \dots, r_{k-1} . Этот факт является важнейшим свойством метода сопряженных градиентов.

Последовательные направления движения p_k удовлетворяют соотношению

$$(Qp_i, p_j) = 0, \quad i \neq j. \quad (1.3.25)$$

Поскольку $p_i = x_i - x_{i-1}$, то для квадратичной функции имеем

$$Qp_i = Qp_i - Qp_{i-1} = r_i - r_{i-1}.$$

С другой стороны, из доказательства леммы 1.3.1 следует, что p_k является линейной комбинацией r_1, \dots, r_{k-1} . Поэтому для $i > k$ имеем

$$(Qp_i, p_k) = \left(r_i - r_{i-1}, \sum_{j=0}^{k-1} \mu_j r_j \right) = 0,$$

а из симметричности матрицы справедливо $(Qp_i, p_k) = (Qp_k, p_i)$, то есть равенство верно для любых $i \neq k$.

Векторы p_i , связанные соотношением (1.3.25), называются *сопряженными*, или *Q-ортогональными*.¹ Это и объясняет название метода: в нем строятся сопряженные направления, являющиеся, согласно (1.3.20), линейными комбинациями градиентов на предыдущих шагах.

В общем случае знание произвольных сопряженных направлений $s_i \neq 0, i = 1, \dots, n, (Qs_i, s_j) = 0, i \neq j$, позволяет легко решить систему

$$Qx = c, \quad Q \succ 0. \quad (1.3.26)$$

Если искать решение в виде $x = \sum_{i=1}^n \alpha_i s_i$, то, подставляя его в систему (1.3.26), умножая скалярно на s_i и используя *Q-ортогональность*, получим соотношение для коэффициентов:

$$\alpha_i = \frac{(c, s_i)}{(Qs_i, s_i)}.$$

Полученному решению можно придать рекуррентную форму: задать $x_1 = 0$ и построить $x_{k+1} = x_k + \alpha_k s_k$ с использованием указанного выражения для α_k . Коэффициенты α_k можно определить иначе как

$$\alpha_k = \arg \min_{\alpha} f(x_k + \alpha s_k).$$

При этом можно запускать итерации с любого начального условия. Таким образом, знание системы сопряженных направлений позволяет найти минимум квадратичной функции с помощью n одномерных минимизаций.

В методе сопряженных градиентов сопряженные направления не выбираются заранее, а строятся рекуррентно. Рассмотрим итерационный процесс, предложенный Р. Флетчером и С. Ривзом:

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k p_k, \quad \alpha_k = \arg \min_{\alpha} f(x_k + \alpha p_k), \\ p_k &= -r_k + \beta_k p_{k-1}, \quad r_k = f'(x_k), \\ \beta_k &= \frac{\|r_k\|^2}{\|r_{k-1}\|^2}, \quad \beta_1 = 0, \quad k = 1, 2, \dots \end{aligned} \quad (1.3.27)$$

Лемма 1.3.2. Для квадратичной функции вида (1.3.22) методы (1.3.20), (1.3.23), (1.3.24) и (1.3.27) при одинаковом x_1 определяют одну и ту же последовательность точек x_k .

¹Они ортогональны в метрике $(x, y)_Q \doteq (Qx, y)$, задаваемой матрицей Q .

Поскольку p_k в (1.3.27) и (1.3.23) отличаются лишь скалярными (ненулевыми) множителями, а r_k совпадают, то процесс (1.3.27) обладает теми же свойствами, что и (1.3.20), (1.3.23): векторы p_i являются сопряженными, а градиенты r_i взаимно ортогональны.

Из леммы 1.3.2 и теоремы 1.3.6 следует, что метод (1.3.27) дает точку минимума квадратичной функции в \mathbb{R}^n за число итераций, не превосходящее n .

Для неквадратичных задач метод (1.3.27) существенно проще, чем метод (1.3.20), (1.3.21), так как требует решения лишь одномерной (а не двумерной) вспомогательной задачи минимизации. Разумеется, в неквадратичном случае теряется свойство конечности метода и (1.3.27) превращается в, вообще говоря, бесконечный итерационный двухшаговый метод.

Обычно для неквадратичных задач метод сопряженных градиентов применяется в несколько иной форме: в него вводится процедура обновления, а именно, время от времени шаг делается не по формуле (1.3.27), а как в начальной точке, то есть по градиенту. Наиболее естественно производить обновление через число итераций, равное размерности пространства.

Алгоритм 5 (Метод сопряженных градиентов, вариант Флетчера – Ривза).

1. Выбрать начальную точку x_1 , инициализировать счетчик итераций $k = 1$.
2. Вычислить градиент функции

$$r_k = f'(x_k)$$

и выбрать направление

$$p_k = -r_k + \beta_k p_{k-1},$$

$$\beta_k = \begin{cases} 0, & \text{mod } (k-1, n) = 0, \\ \frac{\|r_k\|^2}{\|r_{k-1}\|^2}, & \text{mod } (k-1, n) \neq 0. \end{cases} \quad (1.3.28)$$

3. Решить задачу одномерной минимизации

$$x_{k+1} = \arg \min_{\alpha \geq 0} f(x_k + \alpha p_k).$$

4. Проверить выполнение критерия остановки (например, малость r_k). Если он выполняется, то завершить работу и вернуть x_{k+1} в качестве решения. В противном случае вернуться на шаг 2.

Метод сопряженных градиентов с обновлением сходится глобально, при этом в окрестности минимума — с квадратичной скоростью.

Теорема 1.3.7. *Пусть x_* — невырожденная точка минимума и в ее окрестности $f''(x)$ удовлетворяет условию Липшица. Тогда существуют $\varepsilon > 0$, $c > 0$ такие, что для алгоритма 5 с начальной точкой $\|x_1 - x_*\| \leq \varepsilon$ справедлива оценка:*

$$\|x_{(m+1)n} - x_*\| \leq c \|x_{mn} - x_*\|^2, \quad m > 1.$$

Иными словами, n шагов метода сопряженных градиентов по скорости сходимости эквивалентны одному шагу метода Ньютона. В основе доказательства лежат идея квадратичной аппроксимации $f(x)$ и факт конечности метода для квадратичных функций (см. теорему 1.3.6). Доказательство сверхлинейной скорости сходимости метода сопряженных градиентов для неквадратичных функций может быть найдено в [83].

Существуют также схемы метода сопряженных градиентов, отличающиеся от решения двумерной задачи оптимизации (1.3.21) и варианта Флетчера–Ривза (1.3.28). В основном они отличаются выбором коэффициента β_k , включающего обновления (1.3.28):

$$\begin{aligned} \beta_k &= \frac{\|r_k\|^2}{\|r_{k-1}\|^2} && (\text{Флетчер–Ривз}), \\ \beta_k &= \max \left\{ 0, \frac{(r_k, r_k - r_{k-1})}{\|r_k\|^2} \right\} && (\text{Полак–Рибьель}), \\ \beta_k &= \frac{(r_k, r_k - r_{k-1})}{(-p_k, r_k - r_{k-1})} && (\text{Хестенес–Штифель}), \\ \beta_k &= \frac{\|r_k\|^2}{(-p_k, r_k - r_{k-1})} && (\text{Дай–Юан}). \end{aligned}$$

Для квадратичной функции получаемые последовательности $\{x_k\}$ во всех этих вариантах совпадают. В остальных случаях выбор подходящего варианта делается на эмпирической основе.

1.3.6. Субградиентный спуск

Рассмотрим задачу минимизации выпуклой негладкой функции f на \mathbb{R}^n . Под негладкостью будем понимать априорную недифференцируемость f в произвольной точке x . Тем не менее выпуклая функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ субдифференцируема в любой точке $x \in \mathbb{R}^n$, см. теорему Г.13.

Для того чтобы понять отличие субдифференцируемости от дифференцируемости функции в точке в контексте задач минимизации, рассмотрим пример.

Пусть $x = (x_1, x_2)^T \in \mathbb{R}^2$ и $f(x) = |x_1 - x_2| + \frac{1}{5}|x_1 + x_2|$, см. рис. 1.3.5. Тогда в точке $x_0 = (1; 1)$ значение f по любой из координатных осей возрастает. Однако точка $(1; 1)$ не является точкой минимума функции f (легко видеть, что единственный глобальный минимум f — это точка $(0; 0)$). Это означает, что метод покоординатного спуска неприменим для минимизации таких негладких функций.

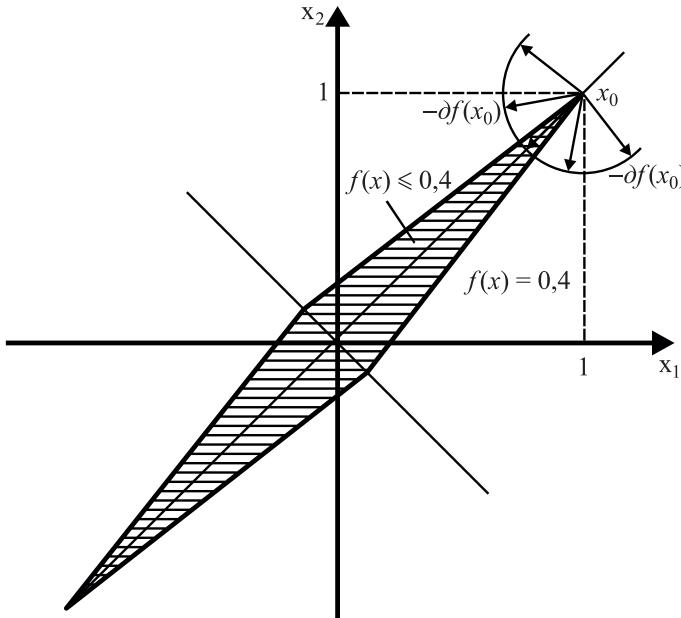


Рис. 1.3.5. Проблемы с субградиентным спуском.

Аналогичные проблемы возникают в ряде других алгоритмов, например в методе скорейшего спуска, когда на каждом шаге выбирается

направление наибыстрейшего убывания функции f и в этом направлении делается некоторый шаг.

Одна из идей минимизации негладких выпуклых функций принадлежит Н.З. Шору. Запишем аналог градиентного метода:

$$x_1 \in \mathbb{R}^n, \quad x_{k+1} = x_k - \alpha_k f'(x_k), \quad (1.3.29)$$

где $f'(x_k) \in \partial f(x_k)$ — произвольный субградиент функции f в точке x_k , см. определение Г.9.

Снова рассмотрим пример $f(x) = |\mathbf{x}_1 - \mathbf{x}_2| + \frac{1}{5}|\mathbf{x}_1 + \mathbf{x}_2|$. Вектор $(1,2; -0,8)^T$ является субградиентом функции f в точке $(1; 1)$, но движение по нему приводит к увеличению функции f при любом выборе длины шага $\alpha_k \in \mathbb{R}$. Таким образом, значение функции не обязано убывать в методе (1.3.29).

Оказывается, что в методе (1.3.29) монотонной является другая величина — евклидово расстояние от текущей точки до точки минимума. В использовании этой монотонности и заключается одна из основных идей субградиентного метода.

Второй важный нюанс состоит в правиле выбора шага α_k . Легко видеть, что фиксированный шаг $\alpha_k = \alpha > 0$, как в градиентном спуске, здесь не годится. Рассмотрим в качестве примера функцию $f(x) = \|x\|$ (евклидову норму). Тогда для всех $x \neq 0$ имеем $\partial f(x) = \{x/\|x\|\}$, то есть функция дифференцируема и ее градиент $f'(x) = \frac{x}{\|x\|}$ имеет длину 1. Тем самым в методе (1.3.29) при $\alpha_k = \alpha > 0$ получается равенство $\|x_{k+1} - x_k\| = \alpha > 0$ и последовательность $\{x_k\}$ не будет сходиться. В субградиентном методе можно уменьшать длину шага при приближении к точке минимума либо используя близость значения функции в текущей точке к минимальному, либо выбирая некоторую априорно стремящуюся к нулю последовательность $\{\alpha_k\}$. Ниже будут рассмотрены оба случая.

Рассмотрим субградиентный метод в следующей форме:

$$x_{k+1} = x_k - \alpha_k \frac{f'(x_k)}{\|f'(x_k)\|}, \quad \alpha_k \rightarrow +0, \quad \sum_{k=1}^{\infty} \alpha_k = +\infty, \quad (1.3.30)$$

где $f'(x_k) \in \partial f(x_k)$ — некоторый (произвольный) субградиент. Таким образом, длина шага стремится к нулю, а суммарная длина шагов равна бесконечности. Примерами последовательностей $\{\alpha_k\}$ в (1.3.30) для $\alpha_0 > 0$ являются

$$\alpha_k = \frac{\alpha_0}{k+c}, \quad c > 0;$$

$$\alpha_k = \begin{cases} \frac{\alpha_0}{k^\rho}, & 0 < \rho \leq 1; \\ \frac{\alpha_0}{k \ln k}, & k > 1. \end{cases}$$

Утверждение о сходимости для (1.3.30) будет относиться к величине

$$\varphi_k = \min_{1 \leq i \leq k} f(x_i), \quad (1.3.31)$$

где $\{x_i\}$ — последовательность из (1.3.30). Таким образом, φ_k — минимальное, «рекордное» значение $f(x_i)$ за k итераций.

Теорема 1.3.8. В методе (1.3.30) с выпуклой функцией f имеем

$$\varphi_k \rightarrow f_* = \inf_{x \in \mathbb{R}^n} f(x).$$

Перед доказательством отметим, что величина f_* может быть равна $-\infty$. Существование точки минимума также не предполагается.

Доказательство. От противного предположим, что $f(x_k) > f_1$ для некоторого $f_1 > f_*$ и всех k .

Выберем точку $\tilde{x} \in \mathbb{R}^n$ такую, что $f(\tilde{x}) < f_1$. Поскольку выпуклая функция непрерывна (см. теорему Г.2), найдется число $\rho > 0$ такое, что $f(x) \leq f_1$ при $\|x - \tilde{x}\| \leq \rho$. В частности, для $x_\rho = \tilde{x} + \rho \frac{f'(x_k)}{\|f'(x_k)\|}$ будет выполняться неравенство $f(x_\rho) \leq f_1$ для всякого субградиента $f'(x_k) \in \partial f(x_k)$.

С другой стороны, из субградиентного неравенства для всякого субградиента $f'(x_k) \in \partial f(x_k)$ имеем

$$\begin{aligned} f(x_\rho) &\geq f(x_k) + (f'(x_k), x_\rho - x_k) \geq f_1 + (f'(x_k), \tilde{x} - x_k) + (f'(x_k), x_\rho - \tilde{x}) = \\ &= f_1 + (f'(x_k), \tilde{x} - x_k) + \rho \|f'(x_k)\|, \end{aligned}$$

то есть $(f'(x_k), x_k - \tilde{x}) / \|f'(x_k)\| \geq \rho$.

Оценим расстояние до \tilde{x} в процессе итераций:

$$\begin{aligned} \|x_{k+1} - \tilde{x}\|^2 &= \|x_k - \tilde{x}\|^2 - 2\alpha_k \left(\frac{f'(x_k)}{\|f'(x_k)\|}, x_k - \tilde{x} \right) + \alpha_k^2 \leq \\ &\leq \|x_k - \tilde{x}\|^2 - 2\rho\alpha_k + \alpha_k^2. \end{aligned}$$

Поскольку $\alpha_k \rightarrow 0$, то найдется такой номер k_0 , что $\alpha_k \leq \rho$ при всех $k \geq k_0$. Поэтому при $k \geq k_0$

$$\|x_{k+1} - \tilde{x}\|^2 \leq \|x_k - \tilde{x}\|^2 - \alpha_k \rho.$$

Суммируя эти неравенства по k от $k = k_0$ до бесконечности, получаем

$$\rho \sum_{k=k_0}^{\infty} \alpha_k \leq \|x_{k_0} - \tilde{x}\|^2,$$

что противоречит равенству $\sum_{k=1}^{\infty} \alpha_k = +\infty$. Таким образом, неравенство $f(x_k) \geq f_1 > f_*$ невозможно. Последнее означает, что $\varphi_k \rightarrow f_*$. Теорема доказана. ■

Упражнение 1.3.12. Пусть в методе (1.3.30) множество минимумов Ω выпуклой функции f непусто и ряд $\sum_{k=1}^{\infty} \alpha_k^2$ сходится. Докажите, что тогда $x_k \rightarrow x_* \in \Omega$.

Величина φ_k , определяемая соотношением (1.3.31), не может сходиться быстро к f_* . Действительно, расстояние до точки минимума не меньше чем длина шага α_k , а эта величина не может быть слишком малой в силу расходимости ряда $\sum_{k=1}^{\infty} \alpha_k$. Кроме того, существует много кандидатов на роль последовательности $\{\alpha_k\}$ в методе (1.3.30). Выбор последовательности в конкретной задаче неочевиден. Поэтому опишем другой способ регулировки длины шага.

В некоторых задачах бывает известно минимальное значение f_* функции f . Например, для системы совместных линейных уравнений

$$(a_i, x) = b_i, \quad a_i, x \in \mathbb{R}^n, \quad 1 \leq i \leq n,$$

поиск решения может быть сведен к минимизации негладких функций

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

или

$$f(x) = \max_{1 \leq i \leq n} |(a_i, x) - b_i|.$$

В обоих случаях $f_* = 0$. Значение f_* позволяет построить следующий вариант субградиентного метода, не содержащего произвольных параметров:

$$x_{k+1} = x_k - \frac{f(x_k) - f_*}{\|f'(x_k)\|^2} f'(x_k), \quad (1.3.32)$$

где субградиент $f'(x_k) \in \partial f(x_k)$ выбирается произвольно. Геометрический смысл выбора такого шага показан на рис. 1.3.6.

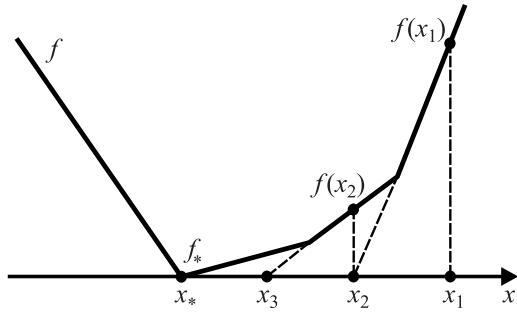


Рис. 1.3.6. Геометрический смысл шага в методе (1.3.32).

Определение 1.3.2. Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R}$ — выпуклая недифференцируемая функция, $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое множество, а $f_* = \min_{\mathcal{Q}} f$. Пусть также $\Omega = \{x \in \mathcal{Q}: f(x) = f_*\} \neq \emptyset$. Будем говорить, что выполнено *условие острого минимума*, если существуют такое число $\alpha > 0$ и $\tilde{x} \in \mathbb{R}^n$, что для всех $x \in \mathbb{R}^n$: $\|x - P_{\Omega}\tilde{x}\| \leq \varrho(\tilde{x}, \Omega)$ выполнено неравенство

$$f(x) - f_* \geq \alpha \varrho(x, \Omega). \quad (1.3.33)$$

Определение 1.3.2 было впервые дано в [63, утверждение 4]. В данном разделе считается, что $\mathcal{Q} = \mathbb{R}^n$.

Теорема 1.3.9. Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R}$ — выпуклая функция и множество точек Ω минимумов f на \mathbb{R}^n непусто. Тогда в методе (1.3.32) $x_k \rightarrow x_* \in \Omega$. Скорость сходимости задается оценкой

$$\liminf_{k \rightarrow \infty} \sqrt{k}(f(x_k) - f_*) = 0. \quad (1.3.34)$$

Если выполнено условие острого минимума (1.3.33), то

$$\|x_{k+1} - x_*\| \leq \sqrt{1 - \frac{\alpha^2}{c^2}} \|x_k - x_*\|, \quad (1.3.35)$$

$$\text{где } c = \sup_{k, f'(x_k) \in \partial f(x_k)} \|f'(x_k)\|.$$

Доказательство. Пусть \tilde{x} — произвольная точка множества Ω .

Тогда для любого $f'(x_k) \in \partial f(x_k)$ имеем:

$$\begin{aligned} \|x_{k+1} - \tilde{x}\|^2 &= \|x_k - \tilde{x}\|^2 - 2 \frac{(f'(x_k), x_k - \tilde{x})(f(x_k) - f_*)}{\|f'(x_k)\|^2} + \\ &\quad + \frac{(f(x_k) - f_*)^2}{\|f'(x_k)\|^2} \leq \|x_k - \tilde{x}\|^2 - \frac{(f(x_k) - f_*)^2}{\|f'(x_k)\|^2}. \end{aligned}$$

Отсюда следует, что $\frac{f(x_k) - f_*}{\|f'(x_k)\|} \rightarrow 0$. Из неравенства $\|x_k - \tilde{x}\| \leq \|x_1 - \tilde{x}\|$ следует, что последовательность $\{x_k\}$ ограничена и поэтому

$$c = \sup_{k, f'(x_k) \in \partial f(x_k)} \|f'(x_k)\| < +\infty. \quad (1.3.36)$$

Упражнение 1.3.13. Докажите неравенство (1.3.36).

Это означает, что $f(x_k) \rightarrow f_*$. Из ограниченности последовательности $\{x_k\}$ по теореме Больцано–Вейерштрасса найдется подпоследовательность $\{x_{k_i}\}$ такая, что $x_{k_i} \rightarrow x_* \in \Omega$ при $i \rightarrow \infty$. Заменяя в полученной выше оценке \tilde{x} на x_* , получаем, что $\{\|x_k - x_*\|\}$ монотонно убывает, а $\{\|x_{k_i} - x_*\|\}$ сходится по i к нулю. Отсюда $x_k \rightarrow x_* \in \Omega$.

Из оценки

$$\frac{(f(x_k) - f_*)^2}{\|f'(x_k)\|^2} \leq \|x_k - x_*\|^2 - \|x_{k+1} - x_*\|^2$$

вытекает

$$\sum_{k=1}^{\infty} \frac{(f(x_k) - f_*)^2}{\|f'(x_k)\|^2} < +\infty.$$

С учетом условия $\|f'(x_k)\| \leq c$ следует $\sum_{k=1}^{\infty} (f(x_k) - f_*)^2 < +\infty$. Если предположить, что $\liminf_{k \rightarrow \infty} \sqrt{k}(f(x_k) - f_*) = \beta > 0$, то для всех достаточно больших номеров k выполнена оценка $f(x_k) - f_* \geq \beta/(2\sqrt{k})$. Это противоречит сходимости ряда с общим членом $(f(x_k) - f_*)^2$, поэтому выполнено условие (1.3.34).

Пусть выполнено условие острого минимума. Тогда из первой оценки с $\tilde{x} = x_*$ следует

$$\|x_{k+1} - x_*\|^2 \leq \|x_k - x_*\|^2 - \frac{\alpha^2}{c^2} \|x_k - x_*\|^2,$$

то есть выполнено (1.3.35). ■

Знаменатель прогрессии $\sqrt{1 - \frac{\alpha^2}{c^2}}$ может быть близок к 1, если линии уровня f сильно вытянуты. Если f_* точно не известно, то можно модифицировать метод (1.3.32) следующим образом:

$$x_{k+1} = x_k - \frac{f(x_k) - f_k}{\|f'(x_k)\|^2} f'(x_k),$$

где f_k есть некоторое приближение f_* на k -м шаге. На основе поведения x_k можно пересчитывать f_k .

В заключение заметим, что итерационный процесс (1.3.32) можно применять и для минимизации гладких функций, причем скорость его сходимости имеет тот же порядок, что и скорость сходимости градиентного метода из раздела 1.3.1.

1.4. Метод Ньютона в задачах безусловной минимизации

1.4.1. Метод Ньютона в задачах минимизации

Рассмотрим общую постановку задачи безусловной выпуклой минимизации

$$\min_{x \in \mathbb{R}^n} f(x),$$

где $f: \mathbb{R}^n \rightarrow \mathbb{R}$ — дважды непрерывно дифференцируемая выпуклая функция. Необходимым и достаточным условием глобального экстремума является условие $f'(x) = 0$. Таким образом, для $F(x) = f'(x)$ получаем, что в рассматриваемом гладком случае нахождение экстремума сводится к решению системы уравнений $F(x) = 0$. Последняя система представляет собой систему из n уравнений с n неизвестными.

Через

$$F'(x) = \left(\frac{\partial F_i(x)}{\partial x_j} \right)_{i,j=1,\dots,n} = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right)_{i,j=1,\dots,n}$$

обозначим матрицу Якоби функции F в точке x ; иными словами — матрицу вторых производных $f''(x)$ (гессиан f) в точке x .

Для решения системы $F(x) = 0$ может быть применен метод Ньютона. В отличие от рассмотренных выше вариантов градиентного спуска он сходится с квадратичной скоростью.

Пусть $x_1 \in \mathbb{R}^n$. Итерации метода Ньютона имеют вид

$$x_{k+1} = x_k - \alpha_k (f''(x_k))^{-1} f'(x_k). \quad (1.4.1)$$

Здесь $\alpha_k > 0$ — шаг метода, который может зависеть от номера итерации k . Будем также предполагать, что матрица $f''(x_k)$ невырождена для всех k . С учетом $F(x) = f'(x)$ можно написать схему метода:

$$x_{k+1} = x_k - \alpha_k (F'(x_k))^{-1} F(x_k). \quad (1.4.2)$$

Нестрого поясним происхождение итерационного процесса (1.4.2). Допустим, что $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ — непрерывно дифференцируемое отображение, и мы хотим найти решение $y \in \mathbb{R}^n$ уравнения $F(y) = 0$, считая, что множество решений непусто. Если выбрать точку $x \in \mathbb{R}^n$ так, чтобы она была достаточно близка к некоторому решению y , то, пользуясь разложением по формуле Тейлора векторной функции F (см. приложение A.3), имеем

$$0 = F(y) = F(x) + F'(x)(y - x) + o(\|y - x\|), \quad x \rightarrow y.$$

Считая, что матрица Якоби $F'(x)$ обратима и $o(\|y - x\|)$ мало по сравнению с остальными членами предыдущего равенства, получаем, что

$$y \approx x - (F'(x))^{-1} F(x).$$

Таким образом, можно написать итерации (1.4.2).

Обратим внимание, что приведенная схема является лишь иллюстрацией (причем грубой) и ни в коем случае не претендует на обоснование метода. Во-первых, мы откинули $o(\cdot)$ без всякого анализа. Во-вторых, даже наличие решения и близость к нему начальной точки итераций x_1 не влечет сходимости алгоритма (1.4.2), см. пример 1.5.6.

Далее займемся выяснением достаточных условий сходимости метода Ньютона к решению соответствующей задачи.

Теорема 1.4.1. *Пусть выпуклая функция $f : \mathbb{R}^n \rightarrow \mathbb{R}$ дважды непрерывно дифференцируема с матрицей вторых производных $f''(x)$ (матрица Гессе), которая удовлетворяет условию Липшица в операторной норме с константой Липшица $L_2 > 0$. Пусть $x_* \in \mathbb{R}^n$ — точка глобального минимума f на \mathbb{R}^n и $f''(x_*) \succcurlyeq \ell I_n$ для некоторого $\ell > 0$. Тогда при $\|x_1 - x_*\| < r = \frac{2\ell}{3L_2}$ итерации метода Ньютона (1.4.1) с $\alpha_k = 1$ для всех k сходятся с квадратичной скоростью*

$$\|x_{k+1} - x_*\| \leq \frac{L_2}{2(\ell - L_2\|x_k - x_*\|)} \|x_k - x_*\|^2. \quad (1.4.3)$$

При этом $\|x_k - x_*\| < r$ для всех k .

Доказательство. Воспользовавшись формулой Ньютона–Лейбница, с учетом $f'(x_*) = 0$ получаем

$$x_{k+1} - x_* = x_k - x_* - (f''(x_k))^{-1} f'(x_k) = x_k - x_* - (f''(x_k))^{-1} \int_0^1 (f''(x_* + t(x_k - x_*)), x_k - x_*) dt = (f''(x_k))^{-1} A_k (x_k - x_*),$$

где $A_k = \int_0^1 (f''(x_k) - f''(x_* + t(x_k - x_*))) dt$.

Положим $r_k = \|x_k - x_*\|$. Тогда

$$\begin{aligned} \|A_k\| &= \left\| \int_0^1 (f''(x_k) - f''(x_* + t(x_k - x_*))) dt \right\| \leqslant \\ &\leqslant \int_0^1 \|f''(x_k) - f''(x_* + t(x_k - x_*))\| dt \leqslant \int_0^1 L_2 (1-t) r_k dt = \frac{r_k}{2} L_2. \end{aligned}$$

Из условия Липшица для f'' имеем

$$f''(x_k) \succcurlyeq f''(x_*) - L_2 r_k I_n \succcurlyeq (\ell - L_2 r_k) I_n.$$

Отсюда при условии $r_k < \ell/L_2$ матрица $f''(x_k)$ положительно определена и

$$\|(f''(x_k))^{-1}\| \leqslant (\ell - L_2 r_k)^{-1}.$$

Поэтому при $r_k < \frac{2\ell}{3L_2}$ имеем

$$r_{k+1} \leqslant \frac{L_2 r_k^2}{2(\ell - L_2 r_k)},$$

то есть (1.4.3). При этом

$$r_{k+1} \leqslant \frac{L_2 r_k^2}{2(\ell - L_2 \frac{2\ell}{3L_2})} = \frac{3L_2}{2\ell} r_k^2 < r_k < r,$$

что и завершает доказательство. ■

Упражнение 1.4.1. Докажите, что в условиях теоремы 1.4.1 минимум x_* единственен.

В случае невыпуклой функции f метод Ньютона можно применять для поиска стационарной точки, то есть такой точки x , в которой $F(x) = f'(x) = 0$. Для решения системы $F(x) = 0$ при условии непрерывной дифференцируемости F (что следует из дважды непрерывной дифференцируемости f) можно применить итерационную схему (1.4.2), см. рис. 1.4.1.

Приведем теорему, доказанную Л.В. Канторовичем, и дадим более короткое ее доказательство, следя [130].

Теорема 1.4.2 (Канторович [54, теорема 1]). Пусть $\mathcal{V}_0 \subset \mathbb{R}^n$ — открытое подмножество, $x_1 \in \mathcal{V}_0$, $F \in C^1(\mathcal{V}_0)$ и определена матрица $(F'(x_1))^{-1}$. Предположим, что выполнены условия:

- a) $\|(F'(x_1))^{-1}\| \leq M$ для некоторого $M > 0$;
- б) $\|(F'(x_1))^{-1} F(x_1)\| \leq K$ для некоторого $K > 0$;
- в) существует $L_1 > 0$ такое, что $\|F'(x) - F'(y)\| \leq L_1 \|x - y\|$ для всех $x, y \in \mathcal{V}_0$.

Пусть также $h = MKL_1 \leq \frac{1}{2}$.

Определим

$$\Omega_0 = \{x \in \mathbb{R}^n : \|x - x_1\| \leq t_0\}, \quad \text{где } t_0 = \frac{1 - \sqrt{1 - 2h}}{h} K.$$

Тогда если $\Omega_0 \subset \mathcal{V}_0$, то итерации (1.4.2) с $\alpha_k = 1$ корректно определены (то есть матрицы $F'(x_k)$ обратимы для всех k), содержатся в Ω_0 и сходятся к точке $x_* \in \Omega_0$ такой, что $F(x_*) = 0$, со скоростью

$$\|x_k - x_*\| \leq \frac{K}{2^{k-1}h} (1 - \sqrt{1 - 2h})^{2^{k-1}}. \quad (1.4.4)$$

Доказательство. Определим $t_1 = 0$ и

$$p(t) = \frac{h}{2K}t^2 - t + K, \quad s(t) = t - \frac{p(t)}{p'(t)}, \quad t_{k+1} = s(t_k).$$

Из определения $\{t_k\}$ для всех $k \geq 1$ вытекает равенство

$$t_{k+1} - t_k = \frac{h}{2(K - ht_k)} (t_k - t_{k-1})^2. \quad (1.4.5)$$

Для доказательства достаточно преобразовать предыдущее выражение и воспользоваться формулой

$$t_{k-1} = t_k - \sqrt{\frac{2K}{h} p(t_k)}. \quad (1.4.6)$$

Упражнение 1.4.2. Докажите формулу (1.4.6).

Также заметим, что $s'(t) = \frac{p(t)p''(t)}{p'^2(t)} > 0$ при $0 \leq t < t_0$, поэтому $t_k < t_{k+1} \rightarrow t_0$, это в частности вытекает из результатов примера 1.5.6.

Из квадратичной аппроксимации липшицево дифференцируемой вектор-функции (A.6) имеем

$$\|F(x) - F(y) - F'(y)(x - y)\| \leq \frac{L_1}{2} \|x - y\|^2 \quad \forall x, y \in \mathcal{V}_0.$$

Если x — внутренняя точка Ω_0 , то

$$\|x - x_1\| < t_0 \leq \frac{1}{ML_1}, \quad \text{следовательно } \|F'(x) - F'(x_1)\| < \frac{1}{M}.$$

Поэтому в силу формулы (A.8)

$$\left\| (F'(x))^{-1} \right\| \leq \frac{M}{1 - ML_1 \|x - x_1\|}.$$

Если точка $N(x) = x - (F'(x))^{-1}F(x)$ также содержится во внутренности Ω_0 , то в силу равенства $F(x) + F'(x)(N(x) - x) = 0$ получаем

$$\begin{aligned} \|N(N(x)) - N(x)\| &\leq \left\| (F'(N(x)))^{-1} F(N(x)) \right\| \leq \\ &\leq \frac{M \|F(N(x)) - F(x) - F'(x)(N(x) - x)\|}{1 - ML_1 \|x_1 - N(x)\|} \leq \\ &\leq \frac{h \|x - N(x)\|^2}{2(K - h \|x_1 - N(x)\|)}. \end{aligned} \quad (1.4.7)$$

Пусть элемент x_m существует и $\|x_m - x_{m-1}\| \leq t_m - t_{m-1}$ для всех $m \leq k$. По предположению это верно для $k = 2$. Тогда имеет место неравенство $\|x_k - x_1\| \leq t_k - t_1 < t_0$ и поэтому x_k содержится во внутренности Ω_0 .

Применяя (1.4.7) для $x = x_{k-1}$, с учетом (1.4.5) имеем для всякого k оценку $\|x_{k+1} - x_k\| \leq t_{k+1} - t_k$. Отсюда следует, что для всяких натуральных m, k выполнена оценка $\|x_{k+m} - x_k\| \leq t_0 - t_k$. Следовательно, последовательность $\{x_k\}$ фундаментальная и $x_k \rightarrow x_* \in \Omega_0$. При этом $\|x_* - x_k\| \leq t_0 - t_k$. Равенство $F(x_*) = 0$ вытекает из непрерывности F и F' в точке x_* .

Из равенства $t_2 = s(t_1)$ при $t_1 = 0$ и формулы (1.4.5) следует, что выражения $ht_2 = Kh$ и $ht_3 - ht_2 = \frac{Kh^4}{2(1-h^2)}$ возрастают по h . Умножая

(1.4.5) на h , по индукции получаем, что если выражения $ht_k - ht_{k-1}$ и ht_k возрастают по h , то $ht_{k+1} - ht_k$ и ht_{k+1} также возрастают по h . Для $h = \frac{1}{2}$ следует $s(t) = \frac{t}{2} + K$, $t_{k+1} = \frac{t_k}{2} + K$, $t_k = (2^{k-1} - 1)K/2^{k-2}$. Отсюда при $h \in (0, \frac{1}{2}]$ из монотонности по h выражения ht_k имеем $K - ht_k \geq \frac{K}{2^{k-1}}$ и $\frac{1}{K - ht_k} \leq \frac{2^{k-1}}{K}$.

Введем в рассмотрение величину $t_{k+1} = t_k - \frac{p(t_k)}{p'(t_k)}$ и заметим, что справедливо равенство

$$\begin{aligned} t_0 - t_{k+1} &= t_0 - t_k + \frac{p(t_k)}{p'(t_k)} = t_0 - t_k - \frac{\frac{h}{2K}t_k^2 - t_k + K}{1 - \frac{h}{K}t_k} = \\ &= \frac{\frac{h}{2K}t_k^2 - \frac{h}{K}t_k t_0 + (t_0 - K)}{1 - \frac{h}{K}t_k}. \end{aligned}$$

Из равенства $p(t_0) = 0$ получаем $t_0 - K = \frac{h}{2K}t_0^2$, откуда

$$\begin{aligned} t_0 - t_{k+1} &= \frac{\frac{h}{2K}t_k^2 - \frac{h}{K}t_k t_0 + \frac{h}{2K}t_0^2}{1 - \frac{h}{K}t_k} = \\ &= \frac{h}{2(K - ht_k)}(t_0 - t_k)^2 \leq \frac{h}{K} \cdot 2^{k-2}(t_0 - t_k)^2. \end{aligned}$$

В силу $t_0 - t_1 = t_0 = \frac{K}{h}(1 - \sqrt{1 - 2h})$ и оценки $\|x_k - x_*\| \leq t_0 - t_k$ получаем (1.4.4). ■

Существует также модифицированный метод Ньютона:

$$x_1 \in \mathbb{R}^n, \quad x_{k+1} = x_k - \alpha_k (F'(x_1))^{-1} F(x_k). \quad (1.4.8)$$

Его отличие от схемы (1.4.2) состоит в том, что матрица Якоби $F'(x)$ обращается только в начальной точке x_1 итераций. Скорость сходимости этого метода меньше, чем метода (1.4.2), однако вычислять обратную матрицу нужно лишь в одной точке, см. рис. 1.4.1.

Приведем теорему о сходимости (1.4.8) для отображения F .

Теорема 1.4.3 ([10, гл. X, §3, теорема 1]). Пусть отображение F дифференцируемо в шаре $\mathcal{B}_r(x_1)$, $r > 0$, а производная $F'(x)$ удовлетворяет для некоторой константы $L_1 > 0$ на этом шаре условию Липшица $\|F'(x) - F'(y)\| \leq L_1 \|x - y\|$ для всех $x, y \in \mathcal{B}_r(x_1)$. Предположим, что матрица $(F'(x_1))^{-1}$ существует и

$$M = \|(F'(x_1))^{-1}\|, \quad K = \|(F'(x_1))^{-1} F(x_1)\|, \quad h = MKL_1.$$

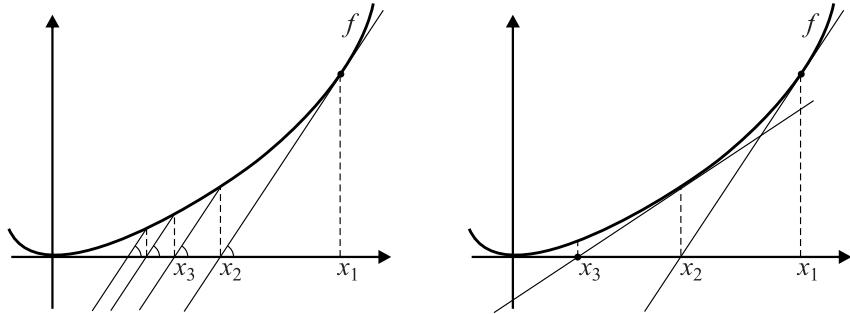


Рис. 1.4.1. Слева — итерации модифицированного метода Ньютона.
Справа — итерации по схеме (1.4.2). В обоих случаях $\alpha_k = 1$.

Тогда при условии $h < \frac{1}{4}$ в шаге $\|x - x_1\| \leq Kt_0$, где t_0 — меньший корень уравнения $ht^2 - t + 1 = 0$, уравнение $F(x) = 0$ имеет единственное решение x_* и последовательность $\{x_k\}$ (1.4.8) с $\alpha_k = 1$ сходится к x_* с линейной скоростью:

$$\|x_k - x_*\| \leq \frac{q^k}{1-q} K, \quad q = \frac{1 - \sqrt{1 - 4h}}{2} < \frac{1}{2}.$$

Доказательство последней теоремы повторяет доказательство теоремы о неявной функции, основанное на принципе сжимающих отображений.

Сделаем в заключение несколько замечаний, сравнив метод Ньютона с методом градиентного спуска. Метод Ньютона имеет сверхлинейную скорость сходимости, в отличие от линейной скорости градиентного спуска. Очень часто метод Ньютона сходится с требуемой точностью за 5–7 итераций. С другой стороны, метод Ньютона очень чувствителен к выбору начального приближения. Отметим, что градиентные методы гораздо слабее зависят от начального приближения.

Эти замечания позволяют дать следующую общую рекомендацию: начинать решение задачи можно с помощью градиентного метода, а попав в достаточно малую окрестность решения, переключаться на метод Ньютона.

1.4.2. Комбинированный метод Ньютона

Альтернативой градиентному методу может служить метод Ньютона (1.4.1) с длиной шага $\alpha_k < 1$, который называется *демпфированным*

методом Ньютона. Обозначим через $p_k = (f''(x_k))^{-1} f'(x_k)$ решение уравнения

$$f''(x_k)p = f'(x_k). \quad (1.4.9)$$

Есть несколько способов выбирать длину шага α_k в схеме

$$x_{k+1} = x_k - \alpha p_k.$$

Во-первых, для дважды дифференцируемой выпуклой функции *ниютоновское направление* $-p_k$ является направлением убывания, и поэтому можно применять правило Армихо (1.3.4).

Упражнение 1.4.3. Докажите, что для выпуклой дважды дифференцируемой функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ имеет место $f''(x) \succcurlyeq 0$ для всех $x \in \mathbb{R}^n$.

Указание: используйте разложение второго порядка для f .

Упражнение 1.4.4. Докажите, что ньютоновское направление является направлением убывания.

Во-вторых, можно использовать выбор шага в зависимости от величины градиента и p_k . Пусть матрица вторых производных липшицева (норма матрицы операторная):

$$\|f''(x) - f''(y)\| \leq L_2 \|x - y\|.$$

Рассмотрим зависимость нормы градиента в новой точке от параметра $\alpha \geq 0$. Из формулы Ньютона–Лейбница, липшицевости матрицы вторых производных, определения операторной нормы, линейности интеграла и (1.4.9) следует верхняя граница:

$$\begin{aligned} \|f'(x_{k+1})\| &= \left\| f'(x_k) + \int_0^\alpha f''(x_k - \tau p_k)(-\alpha p_k) d\tau \right\| = \\ &= \|f'(x_k) - \alpha f''(x_k)p_k + \int_0^\alpha (f''(x_k - \tau p_k) - f''(x_k))(-\alpha p_k) d\tau\| \leq \\ &\leq \|f'(x_k) - \alpha f''(x_k)p_k\| + \int_0^\alpha \|(f''(x_k - \tau p_k) - f''(x_k))\| d\tau \cdot \|\alpha p_k\| \leq \\ &\leq |1 - \alpha| \cdot \|f'(x_k)\| + \frac{L_2 \alpha^2}{2} \|p_k\|^2 \doteq \phi_k(\alpha). \end{aligned} \quad (1.4.10)$$

Минимум по α этой верхней границы достигается при

$$\alpha_k = \arg \min_{\alpha} \phi_k(\alpha) = \min \left\{ 1, \frac{\|f'(x_k)\|}{L_2 \|p_k\|^2} \right\}. \quad (1.4.11)$$

Отметим, что формула определена корректно, поскольку $p_k \neq 0$ (в противном случае из (1.4.9) следует, что $f'(x_k) = 0$, то есть x_k — стационарная точка и решение уже найдено). Норма градиента монотонно убывает, поскольку $\|f'(x_{k+1})\| \leq \min_\alpha \phi_k(\alpha) < \phi_k(0) = \|f'(x_k)\|$.

Упражнение 1.4.5. Докажите, что решение задачи минимизации $\min_\alpha \phi_k(\alpha)$ имеет явный вид (1.4.11) при всех значениях $L_2 > 0$, $f'(x_k)$ и $p_k \neq 0$. Какое решение получится при $p_k = 0$?

Выбор шага (1.4.11) интересен тем, что он определяет правило, по которому надо выбирать классический метод Ньютона с $\alpha_k = 1$ или демпфированный метод Ньютона с $\alpha_k < 1$. Будем называть подобные методы *комбинированными методами Ньютона*:

$$\alpha_k = \begin{cases} 1, & \|p_k\| \leq \sqrt{\frac{1}{L_2} \|f'(x_k)\|}, \\ \frac{\|f'(x_k)\|}{L_2 \|p_k\|^2}, & \|p_k\| > \sqrt{\frac{1}{L_2} \|f'(x_k)\|}. \end{cases}$$

Соответственно, уменьшение нормы градиента происходит с линейной скоростью сходимости:

$$\|f'(x_{k+1})\| \leq \begin{cases} \frac{L_2}{2} \|p_k\|^2, & \|p_k\| \leq \sqrt{\frac{1}{L_2} \|f'(x_k)\|}, \\ \left(1 - \frac{\alpha_k}{2}\right) \|f'(x_k)\|, & \|p_k\| > \sqrt{\frac{1}{L_2} \|f'(x_k)\|}. \end{cases} \quad (1.4.12)$$

При $\alpha_k = 1$ из $\|p_k\| \leq \sqrt{\frac{1}{L_2} \|f'(x_k)\|}$ следует, что $\|f'(x_{k+1})\| \leq \frac{1}{2} \|f'(x_k)\|$. Строго говоря, для линейной скорости сходимости нужно, чтобы в формуле (1.4.12) при $\alpha_k < 1$ множитель $1 - \frac{\alpha_k}{2}$ был отделен от единицы, то есть α_k был ограничен снизу положительным числом. Рассмотрим свойство матрицы $f''(x)$, обеспечивающее это условие.

Пусть функция f сильно выпукла, то есть $f''(x) \succ \ell I$ для некоторого $\ell > 0$ (напомним, что это число связано с константой сильной выпуклости \varkappa как $\ell = 2\varkappa$). Тогда из уравнения (1.4.9) следует

$$\|p_k\| \leq \frac{1}{\ell} \|f'(x_k)\|, \quad (1.4.13)$$

и скорость сходимости оказывается квадратичной:

$$\begin{aligned} \|f'(x_{k+1})\| &\leq \\ &\leq \begin{cases} \frac{L_2}{2\ell^2} \|f'(x_k)\|^2, & \|p_k\| \leq \sqrt{\frac{1}{L_2} \|f'(x_k)\|}, \\ \left(1 - \frac{\ell^2}{2L_2 \|f'(x_k)\|}\right) \|f'(x_k)\|, & \|p_k\| > \sqrt{\frac{1}{L_2} \|f'(x_k)\|}, \end{cases} \end{aligned} \quad (1.4.14)$$

так как

$$\begin{aligned} \left(1 - \frac{\ell^2}{2L_2\|f'(x_k)\|}\right)\|f'(x_k)\| &= \|f'(x_k)\| - \frac{\ell^2}{2L_2} \leqslant \\ &\leqslant \left(1 - \frac{\ell^2}{2L_2\|f'(x_1)\|}\right)\|f'(x_k)\|. \end{aligned}$$

Здесь использовано $\alpha_k = \frac{\|f'(x_k)\|}{L_2\|p_k\|^2}$ и монотонное убывание $\|f'(x_k)\|$.

Из этих выражений можно извлечь больше информации о характере итераций. Сравним условие $\|p_k\| \leqslant \sqrt{\frac{1}{L_2}\|f'(x_k)\|}$, при котором выполняется шаг метода Ньютона с $\alpha_k = 1$, и неравенство (1.4.13).

Очевидно, что если градиент в точке x_k достаточно мал, а именно, $\frac{1}{\ell}\|f'(x_k)\| \leqslant \sqrt{\frac{1}{L_2}\|f'(x_k)\|}$, то есть

$$\|f'(x_k)\| \leqslant \frac{\ell^2}{L_2}, \quad (1.4.15)$$

то метод Ньютона выполняется с длиной шага 1 и сопутствующей ему квадратичной скоростью сходимости, ср. с теоремой 1.4.1.

Из гарантированного убывания нормы градиента (1.4.14) на $\frac{\ell^2}{2L_2}$ на каждом шаге и (1.4.15) следует, что комбинированный метод Ньютона работает следующим образом: сначала делается не более чем $\lceil 2\frac{L_2}{\ell^2}\|f(x_1)\| \rceil - 2$ шагов демпфированного метода Ньютона (первая фаза), а затем происходит переключение на классический метод Ньютона (вторая фаза), сходящийся к решению $x_* : f'(x_*) = 0$.

При этом можно оценить расстояние от начальной точки до решения, и, как следствие, можно требовать выполнение условий липшицевости и невырожденности гессиана не на всем пространстве, а только в окрестности решения.

Для упрощения анализа сходимости продемонстрируем эффект переключения на немного другом варианте комбинированного метода Ньютона. С учетом (1.4.13) верхняя граница (1.4.10) принимает вид

$$\|f'(x_{k+1})\| \leqslant |1-t| \cdot \|f'(x_k)\| + \frac{L_2 t^2}{2\ell^2} \|f'(x_k)\|^2, \quad (1.4.16)$$

а оптимальный шаг (1.4.17) зависит только от градиента в текущей точке (применяется то же условие (1.4.15)) и равен

$$\alpha_k = \min \left\{ 1, \frac{\ell^2}{L_2\|f'(x_k)\|} \right\}. \quad (1.4.17)$$

Скорость сходимости этого метода аналогична (1.4.14), но условие переключения зависит только от нормы градиента, и по его изменению алгоритм также делится на две фазы: убывание на постоянную величину и убывание с квадратичной скоростью сходимости:

$$\|f'(x_{k+1})\| \leq \begin{cases} \frac{L_2}{2\ell^2} \|f'(x_k)\|^2, & \|f'(x_k)\| \leq \frac{\ell^2}{L_2}, \\ \|f'(x_k)\| - \frac{\ell^2}{2L_2}, & \|f'(x_k)\| > \frac{\ell^2}{L_2}. \end{cases} \quad (1.4.18)$$

Докажем сходимость комбинированного метода Ньютона в окрестности минимума в зависимости от нормы градиента в начальной точке. Для оценки скорости сходимости используем монотонно возрастающие и строго выпуклые функции, являющиеся суммой двойного экспоненциального ряда $H_m: [0, 1] \rightarrow \mathbb{R}_+$, определяемого следующим образом (см. подробнее [61]):

$$H_m(\delta) = \sum_{j=m}^{\infty} \delta^{(2^j)}, \quad m = 0, 1, 2, \dots$$

Отметим очевидные свойства этих рядов:

$$H_m(\delta) = \delta^{(2^m)} + H_{m+1}(\delta), \quad H_m(\delta^2) = H_{m+1}(\delta), \quad m = 0, 1, 2, \dots$$

Приведем без доказательства результат о верхней границе функции, связанной с H_0 .

Утверждение 1.4.1 ([125, следствие 4.3]). *Функция*

$$F_1(w) = \begin{cases} 2H_0\left(\frac{w}{2}\right), & 0 \leq w \leq \frac{1}{2}, \\ \lceil 2w \rceil - 2 + 2H_0\left(\frac{1}{2} - \frac{\lceil 2w \rceil - 2w}{4}\right), & w > \frac{1}{2}, \end{cases} \quad (1.4.19)$$

непрерывна при $w \geq 0$. Она имеет кусочно-линейную верхнюю границу $F_2(w)$:

$$F_1(w) \leq F_2(w) = \begin{cases} 2(2c-1)w, & 0 \leq w \leq \frac{1}{2}, \\ 2(w+c-1), & w > \frac{1}{2}, \end{cases}$$

$$\text{где } c = H_0\left(\frac{1}{2}\right) \approx 0,8164215.$$

Предположим, что множество точек минимума выпуклой функции f непусто, $\Omega = \{x \in \mathbb{R}^n : f(x) = \min_x f(x)\} \neq \emptyset$. Рассмотрим его окрестность радиуса $\rho > 0$ (отметим, что она тоже выпуклая):

$$\Omega_\rho \doteq \{x \in \mathbb{R}^n : \varrho(x, \Omega) \leq \rho\}.$$

Следующий результат является адаптацией теоремы 4.1 и следствия 4.3 из [125], примененных к уравнению $f'(x) = 0$.

Теорема 1.4.4 ([125, теорема 4.1, следствие 4.3]). *Пусть для дифференцируемая выпуклая функция f для всех $x \in \Omega_\rho$ имеет липшицев гессиан с константой Липшица L_2 и для всех $x \in \Omega_\rho$ выполняется условие $f''(x) \succcurlyeq \ell I$. Тогда комбинированный метод Ньютона с шагом (1.4.17) сходится к точке минимума $x_* \in \Omega$, если*

$$\|f'(x_1)\| \leq \frac{\ell^2}{L_2} \times \begin{cases} \frac{1}{2(2c-1)} \frac{L_2}{\ell} \rho, & 0 \leq \rho \leq (2c-1) \frac{\ell}{L_2}, \\ \frac{L_2}{2\ell} \rho + 1 - c, & \rho > (2c-1) \frac{\ell}{L_2}, \end{cases} \quad (1.4.20)$$

где $c = H_0\left(\frac{1}{2}\right)$. При этом скорость сходимости квадратичная:

$$\|f'(x_{k+1})\| \leq \|f'(x_1)\| - \frac{\ell^2}{2L_2} k, \quad 0 \leq k < k_{\max}, \quad (1.4.21)$$

$$\|x_{k+1} - x_*\| \leq \frac{\ell}{L_2} \left(k_{\max} - k + 2H_0\left(\frac{\bar{w}}{2}\right) \right), \quad 0 \leq k < k_{\max}, \quad (1.4.22)$$

$$\|f'(x_{k+1})\| \leq \frac{2\ell^2}{L_2} \left(\frac{\bar{w}}{2}\right)^{(2^{(k-k_{\max})})}, \quad k \geq k_{\max}, \quad (1.4.23)$$

$$\|x_{k+1} - x_*\| \leq \frac{2\ell}{L_2} H_{k-k_{\max}}\left(\frac{\bar{w}}{2}\right), \quad k \geq k_{\max}, \quad (1.4.24)$$

зде

$$k_{\max} \doteq \max \left\{ 0, \left\lceil \frac{2L_2}{\ell^2} \|f'(x_1)\| \right\rceil - 2 \right\} \geq 0, \quad (1.4.25)$$

$$\begin{aligned} \bar{w} &= \frac{L_2}{\ell^2} \|f'(x_1)\| - \frac{k_{\max}}{2} = \\ &= \min \left\{ \frac{L_2}{\ell^2} \|f'(x_1)\|, 1 - \frac{1}{2} \left\lceil \frac{2L_2}{\ell^2} \|f'(x_1)\| \right\rceil + \frac{L_2}{\ell^2} \|f'(x_1)\| \right\} \in [0, 1]. \end{aligned}$$

Доказательство. Предположим, что все точки x_k , $k = 1, 2, \dots$ принадлежат множеству Ω_ρ , для которого выполняются указанные в теореме предположения о гессиане функции. Из выпуклости Ω_ρ следует, что матрица вторых производных функции f липшицева на отрезке $[x_k, x_{k+1}]$ и верно (1.4.16). Сделаем замену переменных

$$w_k = \frac{L_2}{\ell^2} \|f'(x_{k+1})\|, \quad k = 0, 1, \dots,$$

в которых неравенство (1.4.16) при шаге (1.4.17) принимает компактный вид:

$$\begin{aligned} w_k &\leq \arg \min_{\alpha} |1 - \alpha| w_{k-1} + \frac{1}{2} \alpha^2 w_{k-1}^2 = \\ &= \begin{cases} w_{k-1} - \frac{1}{2}, & \omega_{k-1} > 1, \\ \frac{1}{2} w_{k-1}^2, & \omega_{k-1} \leq 1, \end{cases} \quad k = 1, 2, \dots \end{aligned}$$

Минимум достигается при $\alpha_k = \frac{1}{w_{k-1}} < 1$, если $w_{k-1} > 1$ (первая фаза), и $\alpha_k = 1$ (вторая фаза) в противном случае. Переход между фазами однократный из-за монотонного убывания w_k . Пусть пороговое значение 1 достигается после $\hat{k} \geq 0$ итераций с $w_{\hat{k}} \leq 1$.

В первой фазе выполняются шаги демпфированного метода Ньютона и w_k убывает не медленнее чем на $\frac{1}{2}$ на каждом шаге:

$$w_k \leq w_0 - \frac{1}{2}k, \quad 0 \leq k \leq \hat{k}. \quad (1.4.26)$$

При этом \hat{k} ограничено сверху: $\hat{k} \leq \max\{0, \lceil 2w_0 \rceil - 2\} = k_{\max}$ с учетом (1.4.25). Также ограничены длина шага:

$$\|x_{k+1} - x_k\| = \alpha_k \|p_k\| \leq \alpha_k \frac{\ell}{L_2} w_{k-1} = \frac{\ell}{L_2}, \quad 1 \leq k \leq \hat{k},$$

и расстояние от x_k до $x_{\hat{k}+1}$, пройденное в первой фазе алгоритма:

$$\|x_{\hat{k}+1} - x_{k+1}\| \leq \frac{\ell}{L_2} (\hat{k} - k), \quad 0 \leq k \leq \hat{k}. \quad (1.4.27)$$

Теперь рассмотрим вторую фазу с $w_k \leq 1$, $k \geq \hat{k}$, и классическим методом Ньютона. Скорость сходимости в этой фазе квадратичная, $w_{k+1} \leq \frac{1}{2} w_k^2$, $k \geq \hat{k}$, откуда

$$w_{\hat{k}+m} \leq 2 \left(\frac{w_{\hat{k}}}{2} \right)^{(2^m)}, \quad m \geq 0. \quad (1.4.28)$$

Из сильной выпуклости на Ω_ρ и (1.4.13) следует

$$\|x_{k+1} - x_k\| = \|p_k\| \leq \frac{1}{\ell} \|f'(x_k)\| = \frac{\ell}{L_2} w_{k-1}, \quad k \geq \hat{k} + 1,$$

и для $m_2 > m_1 \geq 0$ выполнено

$$\begin{aligned} \|x_{\hat{k}+1+m_2} - x_{\hat{k}+1+m_1}\| &\leq \sum_{i=m_1}^{m_2-1} \|x_{\hat{k}+i+2} - x_{\hat{k}+i+1}\| \leq \frac{\ell}{L_2} \sum_{i=m_1}^{m_2-1} w_{\hat{k}+i} \leq \\ &\leq \frac{2\ell}{L_2} \sum_{i=m_1}^{m_2-1} \left(\frac{w_{\hat{k}}}{2}\right)^{(2^i)} = \frac{2\ell}{L_2} \left(H_{m_1}\left(\frac{w_{\hat{k}}}{2}\right) - H_{m_2}\left(\frac{w_{\hat{k}}}{2}\right)\right). \end{aligned}$$

Поскольку $H_{m_1}\left(\frac{w_{\hat{k}}}{2}\right) \leq H_{m_1}\left(\frac{1}{2}\right) \rightarrow 0$ при $m \rightarrow \infty$, то последовательность x_k фундаментальная. Она сходится к точке x_* , в которой из непрерывности градиента

$$\|f'(x_*)\| = \lim_{k \rightarrow \infty} \|f'(x_k)\| = \lim_{k \rightarrow \infty} \frac{\ell^2}{L_2} w_{k-1} = 0,$$

при этом

$$\|x_{\hat{k}+1+m} - x_*\| \leq \frac{\ell}{L_2} \sum_{i=m}^{\infty} w_{\hat{k}+i} \leq \frac{2\ell}{L_2} H_m\left(\frac{w_{\hat{k}}}{2}\right), \quad m = 0, 1, \dots \quad (1.4.29)$$

В частности, можно оценить расстояние до решения от точки, в которой завершается первая фаза алгоритма:

$$\|x_{\hat{k}+1} - x_*\| \leq \frac{2\ell}{L_2} H_0\left(\frac{w_{\hat{k}}}{2}\right). \quad (1.4.30)$$

Отметим, что при $k = \hat{k}$ обе границы (1.4.26) и (1.4.28) совпадают между собой. Из монотонного убывания последовательности w_k , монотонности функции H_0 и свойства $H_m\left(\frac{1}{4}\right) = H_m\left(\frac{1}{2}\right) - \frac{1}{2^{(2^m)}}$ по индукции можно использовать в формулах верхнюю границу числа итераций первой фазы k_{\max} вместо \hat{k} и $\bar{w} = w_0 - \frac{1}{2}k_{\max}$ вместо неизвестного \hat{w} .

Тогда (1.4.26) переходит в (1.4.21), (1.4.28) переходит в (1.4.23), а (1.4.29) переходит в (1.4.24). При $k \leq k_{\max}$ из неравенства треугольника

$$\|x_{k+1} - x_*\| \leq \|x_{k+1} - x_{k_{\max}+1}\| + \|x_{k_{\max}+1} - x_*\|$$

и соответствующих верхних границ (1.4.27) и (1.4.30) следует (1.4.22).

Для завершения доказательства осталось проверить, что условия (1.4.20) действительно приводят к предположению $x_k \in \Omega_\rho$, а именно, что $\|x_k - x_*\| \leq \rho$, $k = 1, 2, \dots$. Из монотонного убывания верхней границы для $\|x_k - x_*\|$ видно, что достаточно оценить расстояние от начальной точки x_1 до точки x_* , к которой сходится последовательность точек x_k .

Рассмотрим два случая, в зависимости от того, с какой фазы начинается комбинированный метод Ньютона: с классического метода Ньютона ($k_{\max} = 0$ при $\bar{w} = \frac{L_2}{\ell^2} \|f'(x_1)\| \leq 1$) или с демпфированного при $\frac{L_2}{\ell^2} \|f'(x_1)\| > 1$.

В первом случае имеет место (1.4.24) и условие $\|x_1 - x_*\| \leq \rho$ гарантируется при

$$\rho \geq \frac{2\ell}{L_2} H_0\left(\frac{L_2}{2\ell^2} \|f'(x_1)\|\right) \geq \|x_1 - x_*\|, \quad \|f'(x_1)\| \leq \frac{\ell^2}{L_2}. \quad (1.4.31)$$

Во втором случае справедливо неравенство (1.4.22) и $\|x_1 - x_*\| \leq \rho$ выполняется при

$$\rho \geq \frac{\ell}{L_2} \left(k_{\max} + 2H_0\left(\frac{\bar{w}}{2}\right) \right) \geq \|x_1 - x_*\|, \quad \|f'(x_1)\| > \frac{\ell^2}{L_2}. \quad (1.4.32)$$

Выразив явно k_{\max} и \bar{w} через $\|f'(x_1)\|$, получим, что условия (1.4.31) и (1.4.32) в точности соответствуют ограничению на функцию F_1 из (1.4.19), а именно, $F_1\left(\frac{L_2}{\ell^2} \|f'(x_1)\|\right) \leq \rho$. Согласно утверждению 1.4.1, оно будет заведомо выполнено, если выполнено $F_2\left(\frac{L_2}{\ell^2} \|f'(x_1)\|\right) \leq \rho$.

Завершает доказательство тот факт, что обратная к кусочно-линейной функции F_2 также является кусочно-линейной, и неравенство $F_2\left(\frac{L_2}{\ell^2} \|f'(x_1)\|\right) \leq \rho$ эквивалентно (1.4.20). ■

Следствие 1.4.1. *Если условия теоремы 1.4.4 выполняются на всем пространстве \mathbb{R}^n (то есть $\rho = \infty$), то комбинированный метод Ньютона с шагом (1.4.17) сходится глобально с квадратичной скоростью.*

Замечание 1.4.1. В условии теоремы 1.4.4 достаточно требовать условие $f''(x) \succcurlyeq \ell I$ не для всех точек множества Ω_ρ , а только для $x \in \Omega_\rho \setminus \Omega = \{x \in \Omega_\rho : x \notin \Omega\}$. Полученный класс функций включает класс сильно выпуклых функций, но не совпадает с ним. Например, такова функция

$$f(x) = ((Qx, x) - (c, x) + d)_+, \quad Q \succ 0,$$

где $(y)_+ — так называемая *функция положительной срезки*, то есть $(y)_+ = \max\{0, y\}$. ▼$

Анализ комбинированного метода с шагом (1.4.11) проводится аналогичным образом. Отличие состоит в том, что на первых k_{\max} шагах с достаточно большим $\|f'(x_k)\|$ может реализоваться шаг $\alpha_k = 1$, и $\|x_{k+1} - x_k\| \leq \frac{1}{\ell} \|f'(x_k)\|$ может быть большим, чем $\frac{\ell}{L_2}$ при выборе шага (1.4.17).

Для применения шага (1.4.11) или (1.4.17) необходимо знать параметры задачи ℓ^2 и L_2 или их оценки. Если они неизвестны, то можно применять адаптивные варианты комбинированного метода Ньютона. В них делаются пробные шаги, но, в отличие от выбора шага типа Армихо применительно к $\|f'(x)\|$ (с константой $\gamma \in (0, 1)$, см. также раздел 1.3.4), вместо условия

$$\|f'(x_k - \alpha p_k)\| \leq (1 - \gamma\alpha) \|f'(x_k)\|$$

проверяется убывание нормы градиента согласно (1.4.12) или (1.4.18).

Свойства адаптивного алгоритма, соответствующего шагу (1.4.18), рассмотрены в [125]. Приведем алгоритм, в котором используется оценка константы Липшица L_2 с помощью последовательности параметров M_k .

Алгоритм 6 (Адаптивный комбинированный метод Ньютона).

1. Выбрать параметры $M_1 > 0$, $Q > 1$, точность $\varepsilon > 0$.
Инициализировать счетчик итераций $k = 1$ и вычислить $u_1 = \|f'(x_1)\|$.

2. Вычислить

$$p_k = (f''(x_k))^{-1} f'(x_k).$$

3. Вычислить

$$\alpha_k = \min \left\{ 1, \frac{\|f'(x_k)\|}{M_k \|p_k\|^2} \right\}, \quad u_{k+1} = \|f'(x_k - \alpha_k p_k)\|.$$

4. Если выполняются условия

$$\alpha_k < 1, \quad u_{k+1} \leq u_k - \frac{1}{2M_k} \frac{u_k^2}{\|p_k\|^2},$$

или

$$\alpha_k = 1, \quad u_{k+1} \leq \frac{M_k}{2} \|p_k\|^2,$$

то перейти к шагу 5.

В противном случае обновить $M_k \leftarrow QM_k$ и перейти к шагу 3 без изменения значения счетчика итераций k .

5. Перейти в точку

$$x_{k+1} = x_k - \alpha_k p_k.$$

6. Проверить условие остановки: если $\|f'(x_{k+1})\| \leq \varepsilon$, то завершить работу и вернуть x_{k+1} в качестве решения.

В противном случае положить $M_{k+1} = M_k$, увеличить счетчик итераций k и перейти к шагу 2.

1.5. Примеры экстремальных задач в безусловной минимизации

В заключительном разделе рассмотрим некоторые примеры задач безусловной минимизации. Все задачи решены, также сформулированы некоторые упражнения для самостоятельного решения. Наиболее сложные задачи отмечены символом «*».

Пример 1.5.1. Пусть $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{k \times n}$, $b \in \mathbb{R}^k$. Требуется найти минимум $\|Ax - b\|$.

Будем считать, что $k \leq n$ и $\text{rank } A = k$. Функция $g(x) = \|Ax - b\|$ является негладкой, поэтому «выгоднее» переформулировать задачу для функции $f(x) = \|Ax - b\|^2$. Действительно, легко видеть, что

$$\left(\min_{x \in \mathbb{R}^n} g(x) \right)^2 = \min_{x \in \mathbb{R}^n} f(x).$$

Имеем

$$f(x) = (Ax, Ax) - 2(Ax, b) + (b, b) = (A^T Ax, x) - 2(A^T b, x) + (b, b);$$

необходимое, а в силу выпуклости f и достаточное, условие экстремума есть

$$f'(x) = 2A^T Ax - 2A^T b = 0, \quad x = (A^T A)^{-1} A^T b.$$

Поэтому $\min_{x \in \mathbb{R}^n} \|Ax - b\| = \|A(A^T A)^{-1} A^T b - b\|$.

Упражнение 1.5.1. Используя условие $\text{rank } A = k$ для матрицы $A \in \mathbb{R}^{n \times k}$, $k \leq n$, докажите обратимость матрицы $A^T A$.

Отметим, что эквивалентная переформулировка задачи для случая гладкой функции, как в примере 1.5.1, является важным шагом, с которого обычно надо начинать анализ экстремальной задачи.

Задача из примера 1.5.1 может быть сформулирована для разных классов матриц, для случая, когда x или b — матрицы и т. п. ▼

Пример 1.5.2 (Квазивыпуклые функции). Назовем функцию $f: \mathbb{R}^n \rightarrow \mathbb{R}$ квазивыпуклой, если ее непустые нижние лебеговы множества

$$\mathcal{L}_f(\mu) = \{x \in \mathbb{R}^n : f(x) \leq \mu\}$$

выпуклые. Функция f называется сильно квазивыпуклой при $\mu \leq \hat{\mu}$, если ее непустые лебеговы множества $\mathcal{L}_f(\mu)$ для всех $\mu \leq \hat{\mu}$ сильно выпуклы с некоторым радиусом $R = R(\mu) > 0$, см. определение Г.11. Слова «при $\mu \leq \hat{\mu}$ » иногда будем опускать.

Упражнение 1.5.2. Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R}$ — сильно выпуклая функция с константой $\kappa > 0$, удовлетворяющая на множестве $\mathcal{L}_f(\hat{\mu})$ условию Липшица с константой $L > 0$. Докажите, что функция f сильно квазивыпукла при $\mu \leq \hat{\mu}$ с $R = L/\kappa$.

Упражнение 1.5.3. Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ квазивыпуклая. Это эквивалентно тому, что

$$f((1 - \lambda)x + \lambda y) \leq \max\{f(x), f(y)\} \quad \forall x, y \in \mathbb{R}^n \quad \forall \lambda \in [0, 1].$$

Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ имеет липшицев градиент с константой Липшица $L_1 > 0$. Тогда непустое лебегово множество $\mathcal{L}_f(f(x))$ является проксимально гладким с константой проксимальной гладкости $r(f(x)) = m(f(x))/L_1$, см. определение Г.12. Здесь положено $m(f(x)) = \inf_{z \in \partial \mathcal{L}_f(f(x))} \|f'(z)\|$. Доказательство этого факта несложно и рекомендуется в качестве упражнения, его также можно найти в [133]. Будем обозначать через $R(\mu)$ радиус сильной выпуклости множества $\mathcal{L}_f(\mu)$; см. рис. 1.5.1.

Теорема 1.5.1. Для квазивыпуклой функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ с непрерывным по Липшицу градиентом и свойством $f_* = \min_{x \in \mathbb{R}^n} f(x) \in \mathbb{R}$ рассмотрим алгоритм градиентного спуска

$$x_{k+1} = x_k - \alpha f'(x_k), \quad \alpha > 0.$$

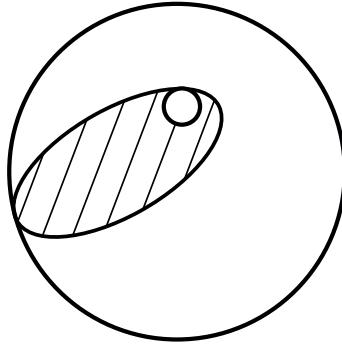


Рис. 1.5.1. Лебегово множество $\{x: f(x) \leq \mu\}$ квазивыпуклой функции из теоремы 1.5.1 (заштриховано). Это множество можно «прокатить» по внутренности сферы радиуса $R(\mu)$, а по его границе внутри лебегова множества можно «прокатить» шар радиуса $r(\mu)$.

Пусть $\emptyset \neq \Omega \subset \mathbb{R}^n$ — множество глобальных минимумов f . Если $x_1 \in \mathbb{R}^n$ и для всех $x \in \mathcal{L}_f(f(x_1)) \setminus \Omega$ выполнено

$$\frac{r(f(x))}{R(f(x))} \geq \mu > 0,$$

то $\varrho(x_k, \Omega) \rightarrow 0$ со скоростью геометрической прогрессии при условии на шаг $0 < \alpha < \min\left\{\frac{\mu}{L_1}, \frac{1}{L_1}\right\}$.

Упражнение 1.5.4. Покажите, что $\mu \leq 1$.

Перед доказательством заметим, что множество $\Omega = \mathcal{L}_f(f_*)$ является сильно выпуклым с некоторым радиусом $R(f_*) \geq 0$.

Доказательство. Для каждого k зафиксируем $z_k \in P_\Omega x_k$, отмечим, что $f'(z_k) = 0$. Пусть $\varrho_k = \|x_k - z_k\|$. Если $x_k \notin \Omega$, то в силу условия теоремы $f'(x_k) \neq 0$. Тогда

$$\begin{aligned} \varrho_{k+1}^2 &\leq \|x_k - \alpha f'(x_k) - z_k\|^2 = \\ &= \varrho_k^2 - 2\alpha(f'(x_k), x_k - z_k) + \alpha^2 \|f'(x_k) - f'(z_k)\|^2. \end{aligned}$$

Из неравенства

$$\left\| x_k - R(f(x_k)) \frac{f'(x_k)}{\|f'(x_k)\|} - z_k \right\| \leq R(f(x_k)),$$

которое следует из опорного условия 1) для сильно выпуклых множеств (см. раздел Г.7), возведением в квадрат получаем

$$(f'(x_k), x_k - z_k) \geq \frac{\|f'(x_k)\|}{2R(f(x_k))} \|x_k - z_k\|^2.$$

Отсюда, с учетом неравенства $\|f'(x_k) - f'(z_k)\| \leq L_1 \|x_k - z_k\|$, имеем

$$\varrho_{k+1}^2 \leq \varrho_k^2 \left(1 - \frac{\alpha \|f'(x_k)\|}{R(f(x_k))} + \alpha^2 L_1^2 \right).$$

Для всех номеров k в силу неравенства $\alpha < 1/L_1$ по лемме 2.3.1, которая верна для функций с липшицевым градиентом, выполняется следующее условие монотонного убывания последовательности $\{f(x_k)\}$:

$$f(x_{k+1}) \leq f(x_k) \leq f(x_1)$$

и, значит,

$$\frac{\|f'(x_k)\|}{R(f(x_k))} \geq \frac{r(f(x_k))L_1}{R(f(x_k))} \geq \mu L_1.$$

Поэтому

$$\varrho_{k+1}^2 \leq \varrho_k^2 (1 - \alpha \mu L_1 + \alpha^2 L_1^2).$$

При $0 < \alpha < \frac{\mu}{L_1}$ получаем $1 - \alpha \mu L_1 + \alpha^2 L_1^2 < 1$. ■

Упражнение 1.5.5. Докажите теорему 1.5.1, в которой условие непрерывности по Липшицу градиента f' заменено на следующее условие ограниченности градиента Фреше $f'(x)$:

$$\|f'(x)\| \leq L_1 \varrho(x, \Omega) \quad \forall x \in \mathbb{R}^n, \tag{1.5.1}$$

где $L_1 > 0$ и Ω — множество решений задачи $\min_{x \in \mathbb{R}^n} f(x)$.

Условие (1.5.1) называется *условием ограничения ошибки*. ▼

Пример 1.5.3 (Задача Ферма – Торричелли). На плоскости \mathbb{R}^2 выбраны три точки x_1, x_2, x_3 . Найти такую точку $x_0 \in \mathbb{R}^2$, что сумма расстояний $f(x) = \sum_{i=1}^3 \|x - x_i\|$ минимальна при $x = x_0$.

Рассмотрим случай, когда углы треугольника $x_1 x_2 x_3$ меньше $2\pi/3$. В примере 1.5.3 функция $f(x)$ выпуклая, но негладкая, и возвведение в квадрат не спасает положение. Также заметим, что минимизация функции $f_1(x) = \sum_{i=1}^3 \|x - x_i\|^2$ — совсем другая задача.

Упражнение 1.5.6. Решите задачу $\min_{x \in \mathbb{R}^n} f_1(x)$.

Из определения субдифференциала выпуклой функции f необходимое и достаточное условие экстремума есть $0 \in \partial f(x)$. В силу теоремы Моро–Рокафеллара (см. приложение Г.5)

$$\partial f(x) = \sum_{i=1}^3 \partial \|x - x_i\|,$$

и если точка $x \neq x_i$, $1 \leq i \leq 3$, то

$$\partial f(x) = \sum_{i=1}^3 \frac{x - x_i}{\|x - x_i\|}.$$

Пусть x_0 — такая точка треугольника $x_1x_2x_3$, что

$$\angle x_1x_0x_2 = \angle x_2x_0x_3 = \angle x_1x_0x_3 = 2\pi/3.$$

Тогда $\partial f(x_0) = \sum_{i=1}^3 \frac{x_0 - x_i}{\|x_0 - x_i\|} = 0$, что является необходимым и достаточным условием глобального минимума f . Точка x_0 называется точкой Ферма–Торричелли.

Упражнение 1.5.7. Пусть величины углов треугольника $x_1x_2x_3$ составляют менее $\frac{2}{3}\pi$. Построим на каждой стороне треугольника $x_1x_2x_3$ вовне равносторонний треугольник. Покажите, что описанные около равносторонних треугольников окружности пересекаются в точке x_0 , см. рис. 1.5.2.

Упражнение 1.5.8. Пусть угол $\angle x_1x_2x_3 \geq \frac{2}{3}\pi$. Докажите, что в этом случае минимум функции f достигается в точке $x_0 = x_2$.

▼

Пример 1.5.4 (Нули и экстремумы). Пусть многочлен $P(z)$ степени n с комплексными коэффициентами и аргументом $z \in \mathbb{C}$ имеет корни $\{z_k\}_{k=1}^n \subset \mathbb{C}$. Пусть также $\{w_m\}_{m=1}^{n-1} \subset \mathbb{C}$ — корни $P'(z) = 0$. Покажем, что $\{w_m\}_{m=1}^{n-1} \subset \text{co}\{z_k\}_{k=1}^n$. Напомним, что со — оператор взятия выпуклой оболочки. Данное утверждение называется *теоремой Люка*. Оно является обобщением теоремы Ролля для многочленов на случай комплексного аргумента.

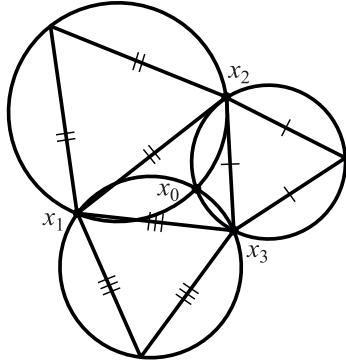


Рис. 1.5.2. Иллюстрация к упражнению 1.5.7.

Пусть $P(z) = \prod_{k=1}^n (z - z_k)$, корни $\{z_k\}$ не обязательно различны.
Тогда $P'(z) = \sum_{k=1}^n \prod_{m=1, m \neq k}^n (z - z_m)$, откуда

$$\frac{P'(z)}{P(z)} = \sum_{k=1}^n \frac{1}{z - z_k}.$$

Таким образом, если $z \in \mathbb{C}$ корень P' и $z \notin \{z_k\}_{k=1}^n$, то

$$\frac{P'(z)}{P(z)} = \sum_{k=1}^n |z - z_k|^{-2} (\bar{z} - \bar{z}_k) = 0,$$

где \bar{z} — комплексное сопряжение z , откуда

$$\sum_{k=1}^n |z - z_k|^{-2} (z - z_k) = 0.$$

Вводя числа $\lambda = \sum_{k=1}^n |z - z_k|^{-2} > 0$ и $\lambda_k = |z - z_k|^{-2}/\lambda$, $\lambda_k > 0$,
 $\sum_{k=1}^n \lambda_k = 1$, из последнего равенства получаем, что $z = \sum_{k=1}^n \lambda_k z_k$. Значит, $z \in \text{co}\{z_k\}_{k=1}^n$. \blacktriangledown

Приведенный пример, а также вычислительная практика показывают, что расположение корней производной может быть устроено достаточно непросто даже в случае многочленов. Рассмотрим еще несколько иллюстративных задач.

Упражнение 1.5.9. Пусть $P(x)$, $x \in \mathbb{R}$, — многочлен степени $k \in \mathbb{N}$ от вещественной переменной, причем $P(x) \geq 0$ для всех $x \in \mathbb{R}$. Тогда найдется точка $x_0 \in \mathbb{R}$, что $P(x_0) = \min_{x \in \mathbb{R}} P(x)$. При этом число k четно.

Таким образом, ограниченный снизу многочлен скалярной переменной достигает инфимума. Оказывается, что для многочлена уже от двух переменных это не так.

Пример 1.5.5. Рассмотрим в \mathbb{R}^2 многочлен от переменных x_1, x_2 вида

$$P(x_1, x_2) = x_1^2 + (x_1 x_2 - 1)^2.$$

С одной стороны, $P(x_1, x_2) \geq 0$ для всех (x_1, x_2) . Если $x_1 = 0$, то $P(0, x_2) = 1$. Если $x_1 \neq 0$, то $P(x_1, x_2) \geq x_1^2 > 0$. Таким образом, для всех $(x_1, x_2) \in \mathbb{R}^2$ выполнена оценка $P(x_1, x_2) > 0$. С другой стороны, для $m \in \mathbb{N}$ имеем

$$P\left(\frac{1}{m}, m\right) = \frac{1}{m^2} + \left(\frac{1}{m} \cdot m - 1\right)^2 = \frac{1}{m^2} \rightarrow +0 \quad \text{при } m \rightarrow \infty.$$

Таким образом, многочлен $P(x_1, x_2)$ ограничен снизу на \mathbb{R}^2 , но не достигает инфимума. ▼

Пример 1.5.6 (Неподвижные точки и сходимость итерационных методов). Напомним теорему Брауэра [18, глава 2, раздел 2]: Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое компактное подмножество и $F: \mathcal{Q} \rightarrow \mathcal{Q}$ — непрерывное отображение множества \mathcal{Q} в себя. Тогда F имеет неподвижную точку, то есть существует точка $x_* \in \mathcal{Q}$ такая, что $F(x_*) = x_*$.

Теорема Брауэра бывает полезна для доказательства существования решения различных задач. Если удается записать условие задачи в виде $x = F(x)$ и при этом x содержится в некотором конечномерном выпуклом компактном подмножестве, а непрерывное отображение F переводит это подмножество в себя, то решение существует. Однако существование решения не означает, вообще говоря, сходимости итерационного процесса $x_1 \in \mathcal{Q}$, $x_{k+1} = F(x_k)$. Приведем простейшие примеры в \mathbb{R} .

Пусть непрерывная функция $f: [a, b] \rightarrow [a, b]$ переводит отрезок $[a, b]$ в себя. Рассмотрим $g(x) = x - f(x)$. Тогда $g(a) = a - f(a) \leq 0$ и $g(b) = b - f(b) \geq 0$. По теореме Больцано–Коши о промежуточных значениях непрерывной на отрезке функции найдется такая точка $x_* \in [a, b]$, что $g(x_*) = x_* - f(x_*) = 0$.

Упражнение 1.5.10. Пусть $f: [a, b] \rightarrow [A, B] = f([a, b])$ — непрерывная функция и $[a, b] \subset [A, B]$. Тогда найдется точка $x_* \in [a, b]$ такая, что $f(x_*) = x_*$.

Покажем, что если непрерывная функция $f: [a, b] \rightarrow [a, b]$ монотонно возрастает, то итерации $x_1 \in [a, b]$, $x_{k+1} = f(x_k)$ сходятся к неподвижной точке $x_* \in [a, b]$.

Пусть $x_2 = f(x_1) \geq x_1$. Отсюда по индукции получаем, что

$$x_{k+1} = f(x_k) \geq f(x_{k-1}) = x_k.$$

Таким образом, последовательность $\{x_k\}$ монотонно возрастает и ограничена: $\{x_k\} \subset [a, b]$. Отсюда по теореме Вейерштрасса о монотонной последовательности следует сходимость $x_k \rightarrow x_* \in [a, b]$ и в пределе $f(x_*) = x_*$.

Если $x_2 = f(x_1) < x_1$, то снова по индукции получаем, что

$$x_{k+1} = f(x_k) \leq f(x_{k-1}) = x_k.$$

Таким образом, последовательность $\{x_k\}$ монотонно убывает и ограничена: $\{x_k\} \subset [a, b]$. Отсюда, как и выше, получаем, что $x_k \rightarrow x_* \in [a, b]$ и $f(x_*) = x_*$.

Тем не менее, если непрерывная функция $f: [a, b] \rightarrow [a, b]$ монотонно убывает, то итерации $x_1 \in [a, b]$, $x_{k+1} = f(x_k)$ могут не сходиться к неподвижной точке $x_* \in [a, b]$. Простейшим контрпримером является функция $f(x) = 1 - x$ на отрезке $x \in [0, 1]$. Она имеет единственную неподвижную точку $x_* = \frac{1}{2}$, однако для любого $x_1 \in [0, 1] \setminus \{\frac{1}{2}\}$ имеем $x_{2k-1} = x_1$, $x_{2k} = 1 - x_1$ при всяком $k \in \mathbb{N}$. Последовательность $\{x_k\}$ не сходится. ▼

Пример 1.5.7 (Аппроксимация Моро–Иосиды).* Рассмотрим выпуклую и полунепрерывную снизу функцию $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$. Оказывается, что f можно приблизить выпуклой функцией $f_\lambda(x)$ с непрерывным по Липшичу градиентом. Эта аппроксимация называется *аппроксимацией Моро–Иосиды*.

Пусть $\lambda > 0$. Тогда

$$f_\lambda(x) = \inf_{y \in \mathbb{R}^n} \left(f(y) + \frac{1}{2\lambda} \|y - x\|^2 \right). \quad (1.5.2)$$

Поскольку функция под операцией взятия инфимума в (1.5.2) сильно выпуклая и полунепрерывная снизу по y , то инфимум достигается в

единственной точке. Будем обозначать ее через $p = p(x) = \text{prox}_f^\lambda x$, то есть для всякого $x \in \mathbb{R}^n$

$$f_\lambda(x) = f(\text{prox}_f^\lambda x) + \frac{1}{2\lambda} \|x - \text{prox}_f^\lambda x\|^2. \quad (1.5.3)$$

Всюду ниже функция f будет считаться собственной, то есть

$$\text{dom } f = \{x \in \mathbb{R}^n : f(x) < +\infty\} \neq \emptyset.$$

Отметим также, что $f(x) \geq f_\lambda(x)$ для любого $\lambda > 0$. Это вытекает из формулы (1.5.2) при $y = x$.

Лемма 1.5.1. *Пусть $\lambda = 1$. Тогда для всякого $x \in \mathbb{R}^n$ равенство $p = \text{prox}_f^1 x$ верно в том и только том случае, когда для всякого $y \in \mathbb{R}^n$ выполнено неравенство $(y - p, x - p) + f(p) \leq f(y)$.*

Доказательство. Пусть $p = \text{prox}_f^1 x$. Зафиксируем $y \in \mathbb{R}^n$. Для $\alpha \in (0, 1)$ определим $p_\alpha = \alpha y + (1 - \alpha)p$. В силу (1.5.2) имеем

$$\begin{aligned} f(p) &\leq f(p_\alpha) + \frac{1}{2}\|x - p_\alpha\|^2 - \frac{1}{2}\|x - p\|^2 = \\ &= f(p_\alpha) + \frac{1}{2}\|(x - p) - \alpha(y - p)\|^2 - \frac{1}{2}\|x - p\|^2 \leq \\ &\leq \alpha f(y) + (1 - \alpha)f(p) - \alpha(x - p, y - p) + \frac{\alpha^2}{2}\|y - p\|^2, \end{aligned}$$

поэтому

$$(y - p, x - p) + f(p) \leq f(y) + \frac{\alpha}{2}\|y - p\|^2.$$

Переходя к пределу $\alpha \rightarrow +0$, получаем $(y - p, x - p) + f(p) \leq f(y)$.

Пусть для всех $y \in \mathbb{R}^n$ выполнено неравенство

$$(y - p, x - p) + f(p) \leq f(y)$$

для некоторого $p \in \mathbb{R}^n$. Тогда для любого $y \in \mathbb{R}^n$

$$\begin{aligned} f(p) + \frac{1}{2}\|x - p\|^2 &\leq \\ &\leq f(y) + \frac{1}{2}\|x - p\|^2 + (x - p, p - y) + \frac{1}{2}\|y - p\|^2 = f(y) + \frac{1}{2}\|y - x\|^2. \end{aligned}$$

Значит, p удовлетворяет (1.5.3) для $\lambda = 1$. ■

Лемма 1.5.2. $x \rightarrow \text{prox}_f^1 x$ является нерастягивающим отображением.

Доказательство. Пусть $p = \text{prox}_f^1 x$, $q = \text{prox}_f^1 y$. Тогда в силу леммы 1.5.1

$$(q - p, x - p) + f(p) \leq f(q), \quad (p - q, y - q) + f(q) \leq f(p).$$

Складывая неравенства, получаем

$$\begin{aligned} 0 &\leq (p - q, (x - p) - (y - q)), \\ \|p - q\|^2 &\leq (p - q, x - y), \quad \|p - q\| \leq \|x - y\|. \end{aligned} \tag{1.5.4}$$

Лемма доказана. \blacksquare

Упражнение 1.5.11. Пусть $\text{Arg} \min_{z \in \mathbb{R}^n} f(z) \neq \emptyset$. Докажите, что $x = \text{prox}_f^1 x$ тогда и только тогда, когда $x \in \text{Arg} \min_{z \in \mathbb{R}^n} f(z)$.

Отметим, что

$$\text{prox}_{\lambda f}^1 x = \text{prox}_f^\lambda x. \tag{1.5.5}$$

Теорема 1.5.2. Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, $\lambda > 0$ и $f_\lambda(x) = \text{an-проксимация}$ Иосиды (1.5.2). Тогда градиент

$$f'_\lambda(x) = \frac{1}{\lambda} (x - \text{prox}_f^\lambda x)$$

существует и удовлетворяет условию Липшица с константой $\frac{1}{\lambda}$.

Отметим, что из условия теоремы вытекает, что $f'_\lambda(x)$ есть градиент. Фреше.

Доказательство. Зафиксируем $x \in \mathbb{R}^n$ и $p = \text{prox}_f^1 x$. Из формулы (1.5.3) для любого $y \in \mathbb{R}^n$ и $q = \text{prox}_f^\lambda y$ по лемме 1.5.1 и в силу равенства (1.5.5) имеем

$$\begin{aligned} f_\lambda(y) - f_\lambda(x) &= f(q) - f(p) + \frac{\|y - q\|^2 - \|x - p\|^2}{2\lambda} \geq \\ &\geq \frac{2(q - p, x - p) + \|y - q\|^2 - \|x - p\|^2}{2\lambda} = \\ &= \frac{\|y - q - (x - p)\|^2 + 2(y - x, x - p)}{2\lambda} \geq \frac{(y - x, x - p)}{\lambda}. \end{aligned}$$

Аналогично получаем, что

$$f_\lambda(y) - f_\lambda(x) \leq \frac{(y-x, y-q)}{\lambda}.$$

Комбинируя две последние оценки с учетом левой оценки в формуле (1.5.4) получаем, что

$$\begin{aligned} 0 &\leq f_\lambda(y) - f_\lambda(x) - \frac{(y-x, x-p)}{\lambda} \leq \\ &\leq \frac{(y-x, (y-q)-(x-p))}{\lambda} = \frac{\|y-x\|^2 - (y-x, q-p)}{\lambda} \leq \\ &\leq \frac{\|y-x\|^2 - \|q-p\|^2}{\lambda} \leq \frac{1}{\lambda} \|y-x\|^2. \end{aligned}$$

Отсюда

$$\lim_{y \rightarrow x} \frac{f_\lambda(y) - f_\lambda(x) - (y-x, \frac{x-p}{\lambda})}{\|y-x\|} = 0.$$

Это означает, что $f'_\lambda(x) = \frac{x-p}{\lambda}$ есть градиент Фреше f_λ в точке x .

Для точек x_i и $p_i = \text{prox}_f^\lambda x_i$, $i = 1, 2$, имеем с учетом (1.5.4)

$$\begin{aligned} \left\| \frac{x_1 - p_1}{\lambda} - \frac{x_2 - p_2}{\lambda} \right\|^2 &= \\ &= \frac{\|x_1 - x_2\|^2}{\lambda^2} + \frac{\|p_1 - p_2\|^2}{\lambda^2} - \frac{2}{\lambda^2} (x_1 - x_2, p_1 - p_2) \leq \\ &\leq \frac{\|x_1 - x_2\|^2}{\lambda^2} + \frac{\|p_1 - p_2\|^2}{\lambda^2} - \frac{2}{\lambda^2} \|p_1 - p_2\|^2 \leq \frac{\|x_1 - x_2\|^2}{\lambda^2}. \end{aligned}$$

Теорема доказана. ■

Покажем, что для всякой точки $x \in \mathbb{R}^n$, где $\partial f(x) \neq \emptyset$, имеет место сходимость $f_\lambda(x) \rightarrow f(x)$ при $\lambda \rightarrow +0$. С одной стороны $f_\lambda(x) \leq f(x)$. Пусть $\xi \in \partial f(x)$. Тогда $f(y) \geq f(x) + (\xi, y-x)$ и

$$f_\lambda(x) \geq f(x) + \inf_{y \in \mathbb{R}^n} \left((\xi, y-x) + \frac{1}{2\lambda} \|y-x\|^2 \right),$$

а последний инфимум достигается в точке $y = x - \lambda\xi$ и равен $-\frac{\lambda}{2} \|\xi\|^2$. Отсюда $\lim_{\lambda \rightarrow +0} f_\lambda(x) = f(x)$.

Упражнение 1.5.12. Пусть $\mathcal{Q} \subset \text{int dom } f$ — непустое выпуклое компактное подмножество. Докажите, что $\lim_{\lambda \rightarrow +0} f_\lambda(x) = f(x)$ равномерно на \mathcal{Q} .

Операция инфимума в определении (1.5.2) называется *инфимальной конволюцией* функций f и $\frac{1}{2\lambda} \|\cdot\|^2$.

Упражнение 1.5.13. Докажите, что надграфик функции f_λ является с точностью до замыкания суммой Минковского надграфиков функций f и $\frac{1}{2\lambda} \|\cdot\|^2$ (см. определение суммы Минковского в разделе Г.1). Отсюда же вытекает выпуклость f_λ .

▼

Пример 1.5.8 (Вариационный принцип Экланда).* Если минимизировать полунепрерывную снизу функцию $f: \mathbb{R}^n \supset Q \rightarrow \mathbb{R}$ на компактном множестве Q , то в силу теоремы Вейерштрасса А.2 обязательно найдется точка $x_* \in Q$ минимума f на Q , то есть $f(x) \geq f(x_*)$ для всех $x \in Q$. Однако это не так, если множество Q не компактно, например, $Q = \mathbb{R}^n$. Даже если $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ ограничена снизу, то инфимум f может не достигаться. Из определения инфимума мы можем написать, что для любого $\varepsilon > 0$ найдется такая точка $x_\varepsilon \in \mathbb{R}^n$, что $f(x) \geq f(x_\varepsilon) - \varepsilon$ для всех $x \in \mathbb{R}^n$.

Упражнение 1.5.14. Докажите предыдущее утверждение для ограниченной снизу функции.

Это довольно скучная информация. Ниже будут сформулированы и доказаны результаты [18] о наличии и свойствах точек, в которых достигается близкое к инфимуму значение полунепрерывной снизу и ограниченной снизу функции.

Теорема 1.5.3 (Вариационный принцип Экланда). Предположим, что функция $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ собственная, полунепрерывная снизу и ограниченная снизу. Пусть $\varepsilon > 0$ и точка $x_\varepsilon \in \mathbb{R}^n$ такая, что

$$f(x_\varepsilon) \leq \inf_{x \in \mathbb{R}^n} f(x) + \varepsilon.$$

Тогда для любого $\lambda > 0$ найдется точка $y = y(\varepsilon, \lambda) \in \mathbb{R}^n$ такая, что

- 1) $f(y) \leq f(x_\varepsilon)$;
- 2) $\|y - x_\varepsilon\| \leq \lambda$;
- 3) для всякой точки $x \in \mathbb{R}^n \setminus \{y\}$ выполнено $f(x) + \frac{\varepsilon}{\lambda} \|x - y\| > f(y)$.

Доказательство. Зафиксируем $\alpha > 0$. Рассмотрим на $\mathbb{R}^n \times \mathbb{R}$ отношение частичного порядка¹ \prec вида

$$(x_1, r_1) \prec (x_2, r_2) \iff r_2 - r_1 + \alpha \|x_2 - x_1\| \leq 0. \quad (1.5.6)$$

¹Не путать с обозначением отрицательно определенной матрицы.

Нетрудно видеть, что данное отношение \prec рефлексивно и транзитивно. Кроме того, для каждого $(x_1, r_1) \in \mathbb{R}^n \times \mathbb{R}$ множество

$$\{(x, r) : (x, r) \prec (x_1, r_1)\}$$

непусто и замкнуто в $\mathbb{R}^n \times \mathbb{R}$.

Упражнение 1.5.15. Докажите предыдущее утверждение.

Покажем, что если замкнутое подмножество $\mathcal{S} \subset \mathbb{R}^n \times \mathbb{R}$ таково, что существует число μ , при котором любой элемент $(x, r) \in \mathcal{S}$ удовлетворяет условию $r \geq \mu$, то для каждого элемента $(x_1, r_1) \in \mathcal{S}$ найдется элемент $(\tilde{x}, \tilde{r}) \in \mathcal{S}$, удовлетворяющий соотношению $(x_1, r_1) \prec (\tilde{x}, \tilde{r})$ и являющийся максимальным во множестве \mathcal{S} в смысле отношения частичного порядка (1.5.6).

Для элемента $(x_1, r_1) \in \mathcal{S}$ определим по индукции последовательность точек $\{(x_k, r_k)\}_{k=1}^{\infty} \subset \mathcal{S}$. Пусть точка (x_k, r_k) известна. Определим множество

$$\mathcal{S}_k = \{(x, r) \in \mathcal{S} : (x_k, r_k) \prec (x, r)\} \quad (1.5.7)$$

и число

$$\mu_k = \inf\{r : (x, r) \in \mathcal{S}_k \text{ для некоторого } x\}. \quad (1.5.8)$$

По определению множества \mathcal{S} справедливо неравенство $\mu_k \geq \mu$. Теперь определим (x_{k+1}, r_{k+1}) как произвольный элемент \mathcal{S} , для которого

$$r_{k+1} \leq \frac{r_k + \mu_k}{2}. \quad (1.5.9)$$

Множества \mathcal{S}_k замкнуты и упорядочены по включению: $\mathcal{S}_{k+1} \subset \mathcal{S}_k$. Отсюда $\mu_{k+1} \geq \mu_k$. Из этого неравенства и формулы (1.5.9) получаем неравенство

$$0 \leq r_{k+1} - \mu_{k+1} \leq \frac{1}{2}(r_k - \mu_k) \leq 2^{-k}(r_1 - \mu).$$

Следовательно, для любого элемента $(x, r) \in \mathcal{S}_{k+1}$ из определения (1.5.8) числа μ_k и определения отношения частичного порядка (1.5.6) получаем неравенства

$$|r_{k+1} - r| \leq |r_{k+1} - \mu_{k+1}| \leq 2^{-k}(r_1 - \mu),$$

$$\|x_{k+1} - x\| \leq \frac{2^{-k}}{\alpha} |r_1 - \mu|.$$

Это значит, что диаметр множеств \mathcal{S}_k стремится к нулю при $k \rightarrow \infty$. В силу полноты $\mathbb{R}^n \times \mathbb{R}$ существует единственная точка

$$(\tilde{x}, \tilde{r}) = \bigcap_{k=1}^{\infty} \mathcal{S}_k.$$

Из определения (1.5.7) множеств \mathcal{S}_k следует, что $(x_k, r_k) \prec (\tilde{x}, \tilde{r})$ для всех k , в частности и для $k = 1$.

Предположим, что найдется точка $(x_0, r_0) \in \mathcal{S}$, для которой выполнено условие $(\tilde{x}, \tilde{r}) \prec (x_0, r_0)$. В силу транзитивности \prec получаем $(x_k, r_k) \prec (x_0, r_0)$ для всех k , откуда $(x_0, r_0) \in \bigcap_{k=1}^{\infty} \mathcal{S}_k$. Отсюда вытекает равенство $(x_0, r_0) = (\tilde{x}, \tilde{r})$ и максимальность (\tilde{x}, \tilde{r}) для отношения \prec .

Теперь положим $\mathcal{S} = \text{epi } f$, $\alpha = \varepsilon/\lambda$ и точку $(x_1, r_1) = (x_\varepsilon, f(x_\varepsilon))$. По доказанному выше в \mathcal{S} существует максимальный элемент $(y, r_\varepsilon) \in \mathcal{S}$ такой, что

$$(x_\varepsilon, f(x_\varepsilon)) \prec (y, r_\varepsilon). \quad (1.5.10)$$

Так как $(y, r_\varepsilon) \in \mathcal{S}$, то $(y, r_\varepsilon) \prec (y, f(y))$, что, вследствие максимальности элемента (y, r_ε) , означает, что $r_\varepsilon = f(y)$. Тогда из (1.5.10) имеем

$$f(y) - f(x_\varepsilon) + \alpha \|y - x_\varepsilon\| \leq 0, \quad (1.5.11)$$

откуда вытекает утверждение 1).

Максимальность элемента $(y, f(y))$ во множестве \mathcal{S} означает, что для любого $x \in \mathbb{R}^n$, $x \neq y$, и такого, что $f(x) < +\infty$, соотношение $(y, f(y)) \prec (x, f(x))$ не имеет места. Это в силу отношения (1.5.6) означает выполнение свойства 3). Наконец, поскольку $f(x_\varepsilon) \leq \inf_{x \in \mathbb{R}^n} f(x) + \varepsilon$, имеем $f(y) \geq f(x_\varepsilon) - \varepsilon$. С учетом неравенства (1.5.11) отсюда получаем оценку 3). Теорема доказана. ■

Отметим, что свойство 3) теоремы 1.5.3 означает, что у функции $g(x) = f(x) + \frac{\varepsilon}{\lambda} \|x - y\|$ в точке y достигается глобальный минимум.

Поясним геометрический смысл свойства 3). Определим конус вращения $\mathcal{K} = \{(x, r): r + \frac{\varepsilon}{\lambda} \|x - y\| < 0\}$. Конус \mathcal{K} имеет вершину $(0, 0)$ и угол φ между осью и образующей, $\operatorname{tg} \varphi = \frac{\lambda}{\varepsilon}$. Тогда сдвинутый конус

$$(y, f(y)) + \mathcal{K} = \left\{ (x, r): r - f(y) + \frac{\varepsilon}{\lambda} \|x - y\| < 0 \right\}$$

лежит ниже графика f , причем касается графика только своей вершиной $(y, f(y))$, см. рис. 1.5.3. Чем меньше ε/λ , тем более плоским

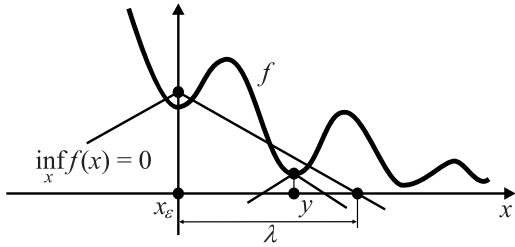


Рис. 1.5.3. Геометрический смысл вариационного принципа.

является конус \mathcal{K} и тем ближе множество $(y, f(y)) + \mathcal{K}$ к горизонтальной плоскости, проходящей через точку $(y, f(y))$.

Выбором коэффициентов ε и λ можно находить баланс между утверждениями 2) и 3) теоремы в зависимости от преследуемых целей. Если коэффициент λ/ε увеличивать, то конус \mathcal{K} становится более плоским, а точка y дает значение $f(y)$, приближающееся к инфимуму. Но при этом, так как правая часть неравенства 2) увеличивается, информация о положении точки y будет незначительной. Если коэффициент λ/ε уменьшать, то точка y близка к исходной точке x_ε , но конус \mathcal{K} острый, и из неравенства 3) можно получить мало информации.

Выбирая $\lambda = 1$ и $\lambda = \sqrt{\varepsilon}$, получаем

Следствие 1.5.1. Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ собственная, полуунепрерывная снизу и ограниченная снизу. Пусть $\varepsilon > 0$ и точка $x_\varepsilon \in \mathbb{R}^n$ такая, что

$$f(x_\varepsilon) \leq \inf_{x \in \mathbb{R}^n} f(x) + \varepsilon.$$

Для $\lambda = 1$ в теореме 1.5.3 найдется точка $y = y(\varepsilon) \in \mathbb{R}^n$ такая, что

- 1) $f(y) \leq f(x_\varepsilon)$;
- 2) для всякой точки $x \in \mathbb{R}^n \setminus \{y\}$ выполнено $f(x) > f(y) - \varepsilon \|x - y\|$.

Для $\lambda = \sqrt{\varepsilon}$ найдется точка $y = y(\varepsilon) \in \mathbb{R}^n$ такая, что

- 1) $f(y) \leq f(x_\varepsilon)$;
- 2) $\|y - x_\varepsilon\| \leq \sqrt{\varepsilon}$;
- 3) для всякой точки $x \in \mathbb{R}^n \setminus \{y\}$ выполнено $f(x) > f(y) - \sqrt{\varepsilon} \|x - y\|$.

Упражнение 1.5.16. Пусть выполнено условие теоремы 1.5.3 и функция f дифференцируема по Гато во всех точках. Покажите, что тогда найдется последовательность точек $\{x_k\}$ такая, что

$$f(x_k) \rightarrow \inf_{x \in \mathbb{R}^n} f(x), \quad f'(x_k) \rightarrow 0.$$



Глава 2

Условная оптимизация

2.1. Устойчивость задач минимизации. Модуль выпуклости

В задаче $\min_{\mathcal{Q}} f$ — задаче минимизации выпуклой функции f на выпуклом замкнутом множестве \mathcal{Q} — важными являются вопросы устойчивости. Пусть множество \mathcal{Q} в некотором смысле немного изменили. Как при этом изменится решение? Тот же вопрос возникает в случае малого в некотором смысле изменения функции f . Кроме того, рассматриваемые ниже алгоритмы дают последовательность $\{x_k\} \subset \mathcal{Q}$, которая может сходиться к решению как по функции, так и по точке. Пусть $f_* = \min_{\mathcal{Q}} f \in \mathbb{R}$ и $\Omega \neq \emptyset$ — множество минимумов выпуклой функции f на выпуклом замкнутом множестве \mathcal{Q} .

Следуя сказанному в разделе 1.1, будем говорить, что последовательность $\{x_k\}$ сходится к решению по функции, если $f(x_k) \rightarrow f_*$, и сходится к решению по точке, если $x_k \rightarrow x_* \in \Omega$.

Из сходимости по точке обычно вытекает сходимость по функции. Дело в том, что в методах оптимизации обычно рассматриваются функции, которые являются хотя бы липшицевыми на рассматриваемой области с некоторой константой Липшица $L > 0$. Если есть сходимость по точке: $x_k \rightarrow x_* \in \Omega$, то $|f(x_k) - f(x_*)| \leq L \cdot \|x_k - x_*\| \rightarrow 0$ и имеет место сходимость по функции. Обратное, как легко видеть, неверно. Приведем простейший пример.

Пример 2.1.1. Рассмотрим в \mathbb{R}^2 функцию $f(x) = -x_2 = -(e_2, x)$ и множество \mathcal{Q} , которое представляет собой прямоугольник с вершинами в точках $(0; 0)$, $(0; 1)$, $(1; 0)$ и $(1; 1)$.

Легко видеть, что решение задачи минимизации f на \mathcal{Q} есть произвольная точка $(x_1; 1)$, $x_1 \in [0, 1]$, и $f(x_1, 1) = -1$. Зафиксируем одно из решений, например $(0, 1)$. Рассмотрим несколько измененное множество $\widehat{\mathcal{Q}}$ вида

$$\widehat{\mathcal{Q}} = \text{co}\{(0; 0), (1; 0), (1; 1 + \varepsilon), (0; 1)\}, \quad \varepsilon > 0.$$

Заметим, что множества \mathcal{Q} и $\widehat{\mathcal{Q}}$ близки при малом $\varepsilon > 0$. В частности, в смысле метрики Хаусдорфа (см. раздел Г.8) $h(\mathcal{Q}, \widehat{\mathcal{Q}}) \leq \varepsilon$. Однако решение по точке задачи $\min_{\widehat{\mathcal{Q}}} f$ есть $(1; 1 + \varepsilon)$, при этом $f(1, 1 + \varepsilon) = -1 - \varepsilon$.

Мы видим, что значение минимума по функции изменилось на ε . Однако точка экстремума $(1; 1 + \varepsilon)$ возмущенной задачи оказалась далеко от зафиксированной точки экстремума $(0; 1)$ исходной задачи. Тем не менее решение возмущенной задачи близко ко множеству решений исходной задачи; ниже мы покажем, что это не случайно. ▼

Рассмотрим задачу $f_* = \min_{x \in \mathcal{Q}} f$. Будем для простоты считать, что \mathcal{Q} компактно, а лебеговы множества $\mathcal{L}_f(\mu) = \{x \in \mathbb{R}^n : f(x) \leq \mu\}$ липшицевой с константой $L > 0$ функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ имеют границу $\partial\mathcal{L}_f(\mu) = \{x \in \mathbb{R}^n : f(x) = \mu\}$, которая непрерывно в метрике Хаусдорфа зависит от уровня μ , см. раздел Г.8.

Решением задачи является множество $\partial\mathcal{L}_f(f_*) \cap \mathcal{Q}$. Пусть x_0 — точка из этого множества, $f_* = f(x_0)$. Рассмотрим возмущенную задачу $\widehat{f}_* = \min_{x \in \widehat{\mathcal{Q}}} f$ и обозначим через $h = h(\mathcal{Q}, \widehat{\mathcal{Q}})$ расстояние в метрике Хаусдорфа между компактными множествами \mathcal{Q} и $\widehat{\mathcal{Q}}$. Пусть $\widehat{x} \in \partial\mathcal{L}_f(\widehat{f}_*) \cap \widehat{\mathcal{Q}}$ и $x_1 \in \widehat{\mathcal{Q}}$ — такая точка, что $\|x_1 - x_0\| \leq h$. Тогда

$$\widehat{f}_* = f(\widehat{x}) \leq f(x_1) \leq f(x_0) + L\|x_1 - x_0\| \leq f_* + Lh.$$

Аналогично $f_* \leq \widehat{f}_* + Lh$. Тем самым $|f_* - \widehat{f}_*| \leq Lh$ и

$$\lim_{h(\mathcal{Q}, \widehat{\mathcal{Q}}) \rightarrow 0} h(\partial\mathcal{L}_f(f_*), \partial\mathcal{L}_f(\widehat{f}_*)) = 0.$$

Значениями многозначного отображения $\widehat{\mathcal{Q}} \mapsto \widehat{\mathcal{Q}} \cap \partial\mathcal{L}_f(\widehat{f}_*)$ являются пересечение значений непрерывного в метрике Хаусдорфа многозначного отображения $F(\widehat{\mathcal{Q}}) = \partial\mathcal{L}_f(\widehat{f}_*)$ с замкнутыми значениями и значений непрерывного в метрике Хаусдорфа многозначного отображения $G(\widehat{\mathcal{Q}}) = \widehat{\mathcal{Q}}$ с компактными значениями. При этом график $F(\widehat{\mathcal{Q}})$ замкнут, так как F непрерывно. В силу [18, гл. 3, §1, теорема 8], а также замечания после определения 1 из [18, гл. 3, §1], это пересечение явля-

ется ε -полунепрерывным сверху в точке \mathcal{Q} . Это значит, что

$$\begin{aligned} \forall \varepsilon > 0 \exists \delta > 0 \forall \widehat{\mathcal{Q}}: h(\mathcal{Q}, \widehat{\mathcal{Q}}) < \delta \text{ и } \widehat{\mathcal{Q}} \text{ — компактно} \\ \partial \mathcal{L}_f(\widehat{f}_*) \cap \widehat{\mathcal{Q}} \subset \mathcal{L}_f(f_*) \cap \mathcal{Q} + \mathcal{B}_\varepsilon(0), \end{aligned} \quad (2.1.1)$$

где $\widehat{f}_* = \min_{x \in \widehat{\mathcal{Q}}} f$, $f_* = \min_{x \in \mathcal{Q}} f$.

Формула (2.1.1) означает, что всякое решение возмущенной задачи близко к некоторому решению исходной задачи, если величина $h(\mathcal{Q}, \widehat{\mathcal{Q}})$ достаточно мала. Однако количественную оценку близости в общей ситуации получить затруднительно.

В настоящем разделе будут обсуждены некоторые свойства функций и множеств, которые гарантируют устойчивость задач минимизации при малом возмущении множества, а также сходимость метода по точке в случае, если доказана его сходимость по функции. При этом будут получены количественные оценки близости.

Основным инструментом будет служить модуль выпуклости функции или множества. Впервые модуль выпуклости для произвольных множеств и его применение в задачах оптимизации были рассмотрены в работе Б.Т. Поляка [62].

Следующее определение дано Б.Т. Поляком [62] и обобщает определение равномерной выпуклости шара в банаховом пространстве [5].

Определение 2.1.1. Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое компактное подмножество с непустой внутренностью. Будем говорить, что множество \mathcal{Q} *равномерно выпукло* с модулем $\delta_{\mathcal{Q}}(\cdot): [0, \text{diam } \mathcal{Q}] \rightarrow [0, +\infty)$, где $\delta_{\mathcal{Q}}(0) = 0$, $\delta_{\mathcal{Q}}(t) > 0$ при $t > 0$ и $\delta_{\mathcal{Q}}(t)$ — неубывающая функция, если для любых точек $x, y \in \mathcal{Q}$ выполнено включение

$$\frac{x+y}{2} + \delta_{\mathcal{Q}}(\|x-y\|)\mathcal{B}_1(0) \subset \mathcal{Q}.$$

Определение 2.1.1 проиллюстрировано на рис. 2.1.1.

Данное определение можно переносить на неограниченные множества, требуя, чтобы для всякого $R > 0$ множество $B_R(0) \cap \mathcal{Q}$, если оно непусто, являлось равномерно выпуклым с модулем $\delta_{\mathcal{Q}, R}(\cdot)$.

Упражнение 2.1.1. Пусть множество $\mathcal{Q} \subset \mathbb{R}^n$ является равномерно выпуклым с модулем $\delta_{\mathcal{Q}}$. Докажите, что $\lim_{t \rightarrow +0} \delta_{\mathcal{Q}}(t) = 0$.

Упражнение 2.1.2. Покажите, что выпуклый многогранник не является равномерно выпуклым множеством. Покажите, что непустое множество вида $\mathcal{Q} = \bigcap_{x \in \mathcal{X}} B_R(x)$, где $R > 0$ и $\mathcal{X} \subset \mathbb{R}^n$ есть произвольное множество центров шаров, равномерно выпукло с модулем $\delta_{\mathcal{Q}}(t) = \frac{t^2}{8R}$.

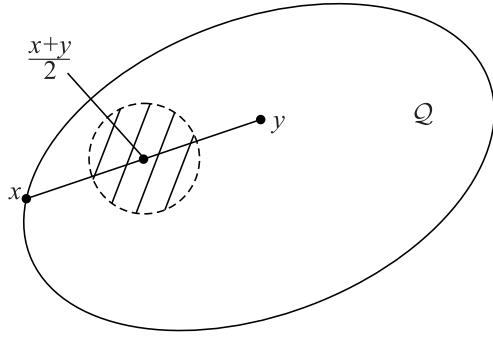


Рис. 2.1.1. Модуль выпуклости множества. Заштрихованный шар радиуса $\delta_Q(\|x - y\|)$ с центром в точке $\frac{x+y}{2}$ содержится во множестве Q .

Упражнение 2.1.3. Пусть $Q \subset \mathbb{R}^n$ — строго выпуклое компактное подмножество, то есть граница ∂Q не содержит отрезков ненулевой длины. Тогда множество Q равномерно выпукло с некоторым модулем выпуклости δ_Q со свойствами из определения 2.1.1.

Упражнение 2.1.3 показывает, что в \mathbb{R}^n классы строго выпуклых компактных множеств и равномерно выпуклых множеств совпадают.

Определение 2.1.2. Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R}$ — выпуклая функция. Будем говорить, что функция f является равномерно выпуклой с модулем $\delta_f: [0, +\infty) \rightarrow [0, +\infty)$, где $\delta_f(t)$ — неубывающая функция такая, что $\delta_f(0) = 0$ и $\delta_f(t) > 0$ при $t > 0$, если для любых точек $x_0, x \in \mathbb{R}^n$ и числа $\lambda \in [0, 1]$ выполнено неравенство

$$f((1 - \lambda)x_0 + \lambda x) \leq (1 - \lambda)f(x_0) + \lambda f(x) - \lambda(1 - \lambda)\delta_f(\|x_0 - x\|). \quad (2.1.2)$$

Определение 2.1.2 проиллюстрировано на рис. 2.1.2. Аналогичное определение можно дать для функции $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$.

Важным классом равномерно выпуклых функций является класс сильно выпуклых функций с модулем $\delta_f(t) = \frac{\varkappa}{2}t^2$, $\varkappa > 0$, см. раздел Г.6, формула (Г.15).

Упражнение 2.1.4. Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ является строго выпуклой, то есть

$$f((1 - \lambda)x_0 + \lambda x) < (1 - \lambda)f(x_0) + \lambda f(x) \quad \forall x_0, x \in \mathbb{R}^n \quad \forall \lambda \in (0, 1),$$

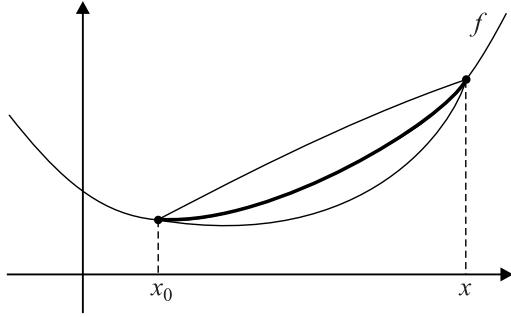


Рис. 2.1.2. Модуль выпуклости функции. Кривая, показанная жирной линией, задается выражением из правой части неравенства (2.1.2).

и множество $\text{dom } f = \{x \in \mathbb{R}^n : f(x) < +\infty\}$ является выпуклым компактом.

Тогда функция f равномерно выпукла, то есть для любых точек $x_0, x \in \text{dom } f$ выполнено неравенство (2.1.2) для некоторой функции δ_f со свойствами из определения 2.1.2.

Теорема 2.1.1. Пусть множество $\mathcal{Q} \subset \mathbb{R}^n$ – равномерно выпуклое с модулем $\delta_{\mathcal{Q}}$, функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ выпукла. Пусть также x_* – минимум функции f на множестве \mathcal{Q} и $f'(x_*) \neq 0$ для субградиента $f'(x_*) \in \partial f(x_*)$ такого, что $-f'(x_*) \in N(\mathcal{Q}, x_*)$. Тогда если последовательность $\{x_k\} \subset \mathcal{Q}$ дает сходимость $f(x_k) \rightarrow f(x_*)$, то

$$\delta_{\mathcal{Q}}(\|x_k - x_*\|) \leq \frac{f(x_k) - f(x_*)}{2\|f'(x_*)\|}. \quad (2.1.3)$$

Доказательство. Из определения 2.1.1 модуля выпуклости

$$\frac{x_k + x_*}{2} + \delta_{\mathcal{Q}}(\|x_k - x_*\|)\mathcal{B}_1(0) \subset \mathcal{Q}.$$

Учитывая необходимое условие экстремума $-f'(x_*) \in N(\mathcal{Q}, x_*)$, получаем

$$\left(-\frac{f'(x_*)}{\|f'(x_*)\|}, x_*\right) \geq \delta_{\mathcal{Q}}(\|x_k - x_*\|) + \left(-\frac{f'(x_*)}{\|f'(x_*)\|}, \frac{x_k + x_*}{2}\right),$$

откуда

$$(f'(x_*), x_k - x_*) \geq 2\|f'(x_*)\| \cdot \delta_{\mathcal{Q}}(\|x_k - x_*\|). \quad (2.1.4)$$

Из субградиентного неравенства для f имеем неравенство

$$f(x_k) - f(x_*) \geq (f'(x_*), x_k - x_*),$$

что с учетом (2.1.4) дает утверждение (2.1.3) теоремы. ■

Упражнение 2.1.5. Докажите, что в условиях теоремы 2.1.1 минимум x_* единственен.

Таким образом, сходимость алгоритма по функции на равномерно выпуклом множестве влечет его сходимость по точке.

Рассмотрим теперь вопрос устойчивости решения задачи $\min_{\mathcal{Q}} f$ при возмущении множества \mathcal{Q} .

Теорема 2.1.2. Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ равномерно выпукла с модулем δ_f . Рассмотрим две задачи $\min_{\mathcal{P}} f$ с компактными выпуклыми множествами $\mathcal{P} = \mathcal{Q}$ и $\mathcal{P} = \widehat{\mathcal{Q}}$, причем расстояние в метрике Хаусдорфа между \mathcal{Q} и $\widehat{\mathcal{Q}}$ есть $h = h(\mathcal{Q}, \widehat{\mathcal{Q}})$. Пусть $L > 0$ — константа Липшица f на множестве $\text{co}(\mathcal{Q} \cup \widehat{\mathcal{Q}})$.

Тогда решения $u_{\mathcal{Q}}$ для \mathcal{Q} и $u_{\widehat{\mathcal{Q}}}$ для $\widehat{\mathcal{Q}}$ связаны оценкой

$$\|u_{\mathcal{Q}} - u_{\widehat{\mathcal{Q}}}\| \leq h + \delta_f^{-1}(Lh). \quad (2.1.5)$$

Обратную функцию $\delta_f^{-1}(\cdot)$ для $\mu \geq 0$ будем определять следующим образом:

$$\delta_f^{-1}(\mu) = \sup\{t \geq 0 : \delta_f(t) \leq \mu\}.$$

Заметим также, что из строгой выпуклости f следует одноточечность решений обеих задач.

Доказательство. Без ограничения общности будем считать, что $f(u_{\mathcal{Q}}) \leq f(u_{\widehat{\mathcal{Q}}})$. Повторяя рассуждения, используемые при доказательстве формулы (Г.16), получаем

$$f(x) \geq f(x_0) + (f'(x_0), x - x_0) + \delta_f(\|x - x_0\|)$$

для всех $x, x_0 \in \mathbb{R}^n$ и $f'(x_0) \in \partial f(x_0)$. Пусть $u \in \widehat{\mathcal{Q}}$ — такая точка, что $\|u - u_{\mathcal{Q}}\| \leq h$.

Рассмотрим функцию $\widehat{f}(x) = f(x)$ при $x \in \widehat{\mathcal{Q}}$ и $\widehat{f}(x) = +\infty$ при $x \notin \widehat{\mathcal{Q}}$. Функция $\widehat{f}: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ равномерно выпукла с модулем δ_f и $u_{\widehat{\mathcal{Q}}}$ есть ее глобальный минимум на \mathbb{R}^n . Отсюда $0 \in \partial \widehat{f}(u_{\widehat{\mathcal{Q}}})$ и

$$\widehat{f}(u) \geq \widehat{f}(u_{\widehat{\mathcal{Q}}}) + \delta_f(\|u - u_{\widehat{\mathcal{Q}}}\|),$$

$\delta_f(\|u - u_{\hat{\mathcal{Q}}}\|) \leq \hat{f}(u) - \hat{f}(u_{\hat{\mathcal{Q}}}) = f(u) - f(u_{\hat{\mathcal{Q}}}) \leq f(u) - f(u_{\mathcal{Q}}) \leq L\|u - u_{\mathcal{Q}}\|$,
то есть $\|u - u_{\hat{\mathcal{Q}}}\| \leq \delta_f^{-1}(L\|u - u_{\mathcal{Q}}\|) \leq \delta_f^{-1}(Lh)$. Из неравенства треугольника

$$\|u_{\mathcal{Q}} - u_{\hat{\mathcal{Q}}}\| \leq \|u - u_{\mathcal{Q}}\| + \|u - u_{\hat{\mathcal{Q}}}\|$$

получаем утверждение (2.1.5) теоремы. ■

Упражнение 2.1.6. Пусть $\varepsilon > 0$. Рассмотрим задачи $\min_{\mathcal{Q}} f$ и $\min_{\mathcal{Q}} \hat{f}$, где $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое компактное подмножество, а f и \hat{f} — выпуклые на \mathbb{R}^n функции, для которых

$$\|f - \hat{f}\|_C = \sup_{x \in \mathbb{R}^n} |f(x) - \hat{f}(x)| \leq \varepsilon.$$

Докажите, что

$$\left| \min_{\mathcal{Q}} f - \min_{\mathcal{Q}} \hat{f} \right| \leq \varepsilon.$$

Заинтересованного читателя мы отсылаем к монографии [36], где можно найти другие подходы к доказательству устойчивости решения экстремальных задач.

2.2. Конечношаговые методы

2.2.1. Линейное программирование

Задача линейного программирования состоит в поиске инфимума линейной функции $(c, x) = (c_1, x_1) + (c_2, x_2)$ при наличии линейных ограничений

$$A_{11}x_1 + A_{12}x_2 = b_1, \quad A_{21}x_1 + A_{22}x_2 \leq b_2, \quad x_1 \geq 0. \quad (2.2.1)$$

Здесь $c_i, x_i \in \mathbb{R}^{n_i}$, $i = 1, 2$, $n = n_1 + n_2$, $x \in \mathbb{R}^n$, $b_i \in \mathbb{R}^{m_i}$, $i = 1, 2$, матрицы A_{ij} имеют соответствующие размеры. Неравенства для векторов понимаются покомпонентно. Часть компонент x , составляющая x_1 , неотрицательна, а компоненты, составляющие x_2 , могут иметь любой знак.

Отметим, что поиск супремума (c, x) при x , удовлетворяющем системе (2.2.1), эквивалентен поиску инфимума: надо лишь заменить вектор c на $-c$.

Определение 2.2.1. Пусть $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$. Задачу линейного программирования вида

$$\inf(c, x) \quad \text{при условии } x \in \mathcal{Q} = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\} \quad (2.2.2)$$

принято называть *канонической задачей линейного программирования*.

Покажем, что общую задачу минимизации линейной функции (c, x) при ограничениях (2.2.1) можно свести к задаче (2.2.2). Определим $x_2 = u - v$, где $u, v \in \mathbb{R}^{n_2}$ и $u, v \geq 0$, а также переменную $z \in \mathbb{R}^{m_2}$ вида $z = b_2 - A_{21}x_1 - A_{22}(u - v) \geq 0$. Тогда систему (2.2.1) можно переписать в виде

$$\inf(c, x)$$

при ограничениях

$$A_{11}x_1 + A_{12}(u - v) = b_1, \quad A_{21}x_1 + A_{22}(u - v) + z = b_2, \quad x_1, u, v, z \geq 0,$$

где $x = (x_1, u - v)^T$.

Тем самым, имея теорию и алгоритм для решения задачи (2.2.2), мы можем получить решение любой задачи линейного программирования.

Напомним, что в силу формулы (Г.3) для любого выпуклого замкнутого множества $\mathcal{Q} \subset \mathbb{R}^n$ имеем $\mathcal{Q} = \text{co}(\text{extr } \mathcal{Q}) + O^+ \mathcal{Q}$, где $O^+ \mathcal{Q}$ — асимптотический конус (Г.2). Далее будем считать, что в задаче (2.2.2) $m \leq n$ и $\text{rank } A = r \leq m$.

Рассмотрим, как устроено множество крайних точек множества \mathcal{Q} в (2.2.2).

Теорема 2.2.1. Пусть $m \leq n$ и в (2.2.2) матрица $A \in \mathbb{R}^{m \times n}$ имеет ранг $\text{rank } A = r \leq m$. Точка $x_0 = ((\mathbf{x}_0)_1, \dots, (\mathbf{x}_0)_n)^T \in \mathcal{Q}$ в (2.2.2) крайняя тогда и только тогда, когда система (2.2.2) может быть переписана в виде

$$\sum_{k=1}^r A_{i_k}(\mathbf{x}_0)_{i_k} + \sum_{j=1, j \notin \{i_k\}_{k=1}^r}^n A_j(\mathbf{x}_0)_j = b, \quad (2.2.3)$$

где столбцы $\{A_{i_k}\}_{k=1}^r$ линейно независимы, компоненты $(\mathbf{x}_0)_{i_k} \geq 0$ для всех $1 \leq k \leq r$, а во второй сумме все $(\mathbf{x}_0)_j = 0$.

Доказательство. Пусть $x_0 \in \mathcal{Q}$ удовлетворяет условию (2.2.3). Обозначим $\tilde{A} = (A_{i_1}, \dots, A_{i_r})$, $\tilde{x}_0 = ((\mathbf{x}_0)_{i_1}, \dots, (\mathbf{x}_0)_{i_r})^T$. Матрица \tilde{A} невырождена.

Допустим, что точка x_0 не крайняя. Тогда существует ненулевой вектор $y \in \mathbb{R}^n$ и $\varepsilon > 0$ такие, что $A(x_0 \pm \varepsilon y) = b$ и $x_0 \pm \varepsilon y \geq 0$. В силу $(\mathbf{x}_0)_j = 0$ во второй сумме получаем, что $y_j = 0$ при $j \neq i_k$ для

$1 \leq k \leq r$. Обозначив $\tilde{y} = (y_{i_1}, \dots, y_{i_r})^T$, получаем $\tilde{A}(\tilde{x} \pm \varepsilon \tilde{y}) = b$ и $\tilde{x} \pm \varepsilon \tilde{y} \geq 0$. Отсюда и из равенства $\tilde{A}\tilde{x} = b$ следует, что $\tilde{A}\tilde{y} = 0$ и $\tilde{y} \neq 0$. Это противоречит невырожденности матрицы \tilde{A} . Противоречие показывает, что точка x_0 — крайняя.

Пусть x_0 — крайняя точка. Если $x_0 = 0$, то в качестве столбцов матрицы \tilde{A} возьмем любые линейно независимые столбцы A .

Пусть $x_0 \in \text{extr } Q \setminus \{0\}$. Перенумеруем компоненты точки x_0 так, чтобы $(x_0)_1 > 0, \dots, (x_0)_l > 0$, а $x_k = 0$ при любом $k: l+1 \leq k \leq n$. Покажем, что $l \leq r$ и что после указанной перестановки компонент точки x_0 и соответствующей перестановки столбцов матрицы A в выражении

$$Ax_0 = \sum_{k=1}^l A_k(x_0)_k$$

столбцы $\{A_k\}_{k=1}^l$ будут линейно независимы, что и завершит доказательство теоремы.

Допустим, что столбцы $\{A_k\}_{k=1}^l$ линейно зависимы. Тогда найдется ненулевой вектор $a \in \mathbb{R}^n$ вида $a = (\alpha_1, \dots, \alpha_l, 0, \dots, 0)^T$ такой, что $\sum_{k=1}^l A_k \alpha_k = 0$. Существует число $\varepsilon > 0$ такое, что для каждого номера k от 1 до l справедливы неравенства $(x_0)_k \pm \varepsilon \alpha_k \geq 0$. Тогда

$$\begin{aligned} A(x_0 \pm \varepsilon a) &= \sum_{k=1}^l A_k((x_0)_k \pm \varepsilon \alpha_k) + \sum_{k=l+1}^n A_k(x_0)_k = \\ &= \sum_{k=1}^l A_k(x_0) \pm \varepsilon \sum_{k=1}^l A_k \alpha_k = b. \end{aligned}$$

Следовательно, $[x_0 - \varepsilon a, x_0 + \varepsilon a] \subset Q$, то есть $x_0 \notin \text{extr } Q$, и получаем противоречие. \blacksquare

Рассмотрим теперь задачу

$$\sup(b, x), \quad A^T x \leq c, \quad b, x \in \mathbb{R}^m, \quad c \in \mathbb{R}^n, \quad m \leq n. \quad (2.2.4)$$

Задача (2.2.4) называется *основной задачей линейного программирования*, см. рис. 2.2.1. Пусть \bar{x} — решение (2.2.4), а $\bar{\lambda}$ — решение (2.2.2) с вектором λ вместо x . Покажем, что $(b, \bar{x}) = (c, \bar{\lambda})$.

Запишем функцию Лагранжа для задачи (2.2.4), см. раздел Д.3, с минимизируемой функцией $-(b, x)$:

$$L(x, \lambda) = -(b, x) + \lambda(A^T x - c) = (A\lambda - b)x - (c, \lambda).$$

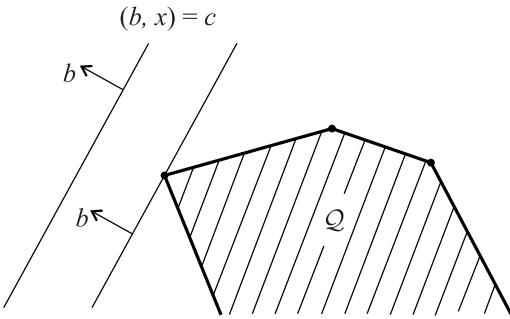


Рис. 2.2.1. Основная задача линейного программирования.

Поскольку (2.2.4) есть задача выпуклого программирования, то необходимым и достаточным условием экстремума (см. подробнее раздел Д.1) является условие

$$0 \in \partial_x L(\bar{x}, \bar{\lambda}) = A\bar{\lambda} - b.$$

С учетом полиэдральности множества ограничений в задаче (2.2.4), теоремы Д.3 и замечания после этой теоремы о полиэдральных множествах получаем

$$-(b, \bar{x}) = L(\bar{x}, \bar{\lambda}) = -(b, \bar{x}) + (A^T \bar{x} - c)\bar{\lambda} = -(c, \bar{\lambda}) + (A\bar{\lambda} - b)\bar{x} = -(c, \bar{\lambda}),$$

причем последнее выражение есть максимум $(-c, \lambda)$ по всем $\lambda \geq 0$ таким, что $A\lambda - b = 0$.

В силу этого говорят, что задачи (2.2.2) и (2.2.4) являются *двойственными*. Решение любой из них автоматически решает и двойственную задачу.

Рассмотрим структуру решения задачи линейного программирования $\sup(c, x)$ при условии $x \in \mathcal{Q}$. Множество \mathcal{Q} может быть задано линейными ограничениями (2.2.2), (2.2.4) или самого общего вида. Если существует такой вектор $v \in O^+ \mathcal{Q}$, что $(c, v) < 0$, то $\inf_{\mathcal{Q}}(c, x) = -\infty$. Если для каждого вектора $v \in O^+ \mathcal{Q}$ выполнено неравенство $(c, v) \geq 0$, то в силу теоремы Г.12 решение задачи достигается в некоторой крайней точке множества \mathcal{Q} .

Рассмотрим вопрос об устойчивости решения задач линейного программирования. В свете эквивалентности задач далее будем рассматривать задачу в основной форме (2.2.4), то есть в виде

$$\inf(c, x) \quad \text{при условии } x \in \mathcal{Q} = \{x \in \mathbb{R}^n : Ax \leq b\}, \quad b \in \mathbb{R}^m. \quad (2.2.5)$$

Условие $\inf(c, x)$ эквивалентно $\sup(-c, x)$. Ниже мы будем предполагать, что множество \mathcal{Q} не содержит прямых, то есть для \mathcal{Q} выполнена теорема Г.10 Крейна–Мильмана–Кли.

Теорема 2.2.2. *Пусть множество \mathcal{Q} в задаче (2.2.5) не содержит прямых и $\|c\| = 1$. Пусть $\mathcal{Q} \subset \Omega \neq \emptyset$ — множество решений задачи (2.2.5) и $f_0 = f(\Omega) \in \mathbb{R}$. Тогда найдется такое число $\alpha > 0$, что для всякого $x \in \mathcal{Q}$*

$$\alpha \varrho(x, \Omega) \leq (c, x) - f_0.$$

Отметим, что число α зависит от матрицы A и векторов b, c .

Доказательство. В силу формулы (Г.3) $\mathcal{Q} = \text{co extr } \mathcal{Q} + O^+ \mathcal{Q}$. Поскольку множество Ω выпукло и замкнуто, а также не содержит прямых, то $\Omega = \text{co extr } \Omega + O^+ \Omega$. Отметим, что множества $\text{extr } \mathcal{Q}$ и $\text{extr } \Omega$ конечны.

Шаг 1. Покажем, что существует такое число $\alpha_0 \in (0, 1)$, что для всех $x \in \text{co extr } \mathcal{Q}$ выполнено неравенство

$$\alpha_0 \varrho(x, \text{co extr } \Omega) \leq (c, x) - f_0. \quad (2.2.6)$$

Упражнение 2.2.1. Докажите, что $\emptyset \neq \text{extr } \Omega \subset \text{extr } \mathcal{Q}$.

Упражнение 2.2.2. Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое подмножество. Докажите, что функция $\mathbb{R}^n \ni x \mapsto \varrho(x, \mathcal{Q})$ выпукла.

Пусть $\{x_i\}_{i=1}^I = \text{extr } \mathcal{Q}$ и $\{x_i\}_{i \in J} = \text{extr } \Omega$, где J — некоторое подмножество $\{1, \dots, I\}$. Если $\{1, \dots, I\} = J$, то левая и правая части в (2.2.6) равны нулю и $\alpha_0 > 0$ можно взять любым. Далее считаем $\{1, \dots, I\} \neq J$. Определим

$$\alpha_0 = \min \left\{ 1, \min_{1 \leq i \leq I, i \notin J} \frac{(c, x_i) - f_0}{\varrho(x_i, \text{co extr } \Omega)} \right\} > 0,$$

так как для всех рассматриваемых i числитель $(c, x_i) - f_0$ и знаменатель $\varrho(x_i, \text{co extr } \Omega)$ строго положительны.

Пусть $\text{co extr } \mathcal{Q} \ni x = \sum_{i=1}^I \lambda_i x_i$, $\lambda_i \geq 0$, $\sum_{i=1}^I \lambda_i = 1$. Тогда, с учетом

упражнения 2.2.2,

$$\begin{aligned}\varrho(x, \text{co extr } \mathcal{Q}) &= \varrho\left(\sum_{i=1}^I \lambda_i x_i, \text{co extr } \mathcal{Q}\right) \leq \sum_{i=1}^I \lambda_i \varrho(x_i, \text{co extr } \mathcal{Q}) = \\ &= \sum_{i=1, i \notin J}^I \lambda_i \varrho(x_i, \text{co extr } \mathcal{Q}) \leq \alpha_0^{-1} \sum_{i=1, i \notin J}^I \lambda_i ((c, x_i) - f_0) = \\ &= \alpha_0^{-1} \sum_{i=1}^I \lambda_i ((c, x_i) - f_0) = \alpha_0^{-1} ((c, x) - f_0).\end{aligned}$$

Упражнение 2.2.3. Пусть \mathcal{P}, \mathcal{Q} — замкнутые подмножества и $x, y \in \mathbb{R}^n$. Тогда

$$\varrho(x + y, \mathcal{P} + \mathcal{Q}) \leq \varrho(x, \mathcal{P}) + \varrho(y, \mathcal{Q}).$$

Шаг 2. Определим

$$H_c = \{x \in \mathbb{R}^n : (c, x) = 0\}, \quad H_c^+ = \{x \in \mathbb{R}^n : (c, x) > 0\}.$$

Пусть $x \in \mathcal{Q} = \text{co extr } \mathcal{Q} + O^+ \mathcal{Q}$ — произвольный. Тогда $x = z + v$, где $z \in \text{co extr } \mathcal{Q}$, а $v \in O^+ \mathcal{Q}$. Поскольку конус $O^+ \mathcal{Q}$ полиэдральный, то он конечнопорожденный и может быть представлен в виде

$$O^+ \mathcal{Q} = \left\{ \sum_{i=1}^m \lambda_i u_i + u : \lambda_i \geq 0, u_i \in H_c^+, u \in O^+ \Omega \right\}.$$

Если $v \in O^+ \Omega$, то $(c, v) = 0$ и с учетом упражнения 2.2.3 и равенства $\varrho(v, O^+ \Omega) = 0$ заключаем, что

$$\frac{(c, z + v) - f_0}{\varrho(z + v, \text{co extr } \Omega + O^+ \Omega)} \geq \frac{(c, z) - f_0}{\varrho(z, \text{co extr } \Omega)},$$

где правая часть в силу шага 1 не меньше чем α_0 .

Пусть $v \in O^+ \mathcal{Q} \setminus O^+ \Omega$. Тогда $(c, v) > 0$ и

$$\frac{(c, z + v) - f_0}{\varrho(z + v, \text{co extr } \Omega + O^+ \Omega)} \geq \frac{(c, z) - f_0 + (c, v)}{\varrho(z, \text{co extr } \Omega) + \varrho(v, O^+ \Omega)}.$$

Определим $\varrho_i = (c, u_i) = \varrho(u_i, H_c) > 0$, $1 \leq i \leq m$, $\varrho_0 = \min_{1 \leq i \leq m} \varrho_i > 0$,

а также $d_0 = \max_{1 \leq i \leq m} \|u_i\| > 0$. Имеем для $v = \sum_{i=1}^m \lambda_i u_i + u \in O^+ \mathcal{Q}$, $u \in O^+ \Omega$

$$(c, v) = \sum_{i=1}^m \lambda_i (c, u_i) \geq \varrho_0 \sum_{i=1}^m \lambda_i,$$

$$\varrho(v, O^+ \Omega) \leq \sum_{i=1}^m \lambda_i \|u_i\| \leq d_0 \sum_{i=1}^m \lambda_i.$$

Отсюда $(c, v)/\varrho(v, O^+ \Omega) \geq \varrho_0/d_0 = \beta_0 \in (0, 1]$.

Из оценки шага 1 имеем

$$\frac{(c, z) - f_0}{\varrho(z, \text{co extr } \Omega)} \geq \alpha_0.$$

Из двух последних оценок получаем, что

$$\begin{aligned} \frac{(c, x) - f_0}{\varrho(x, \text{co extr } \Omega + O^+ \Omega)} &= \frac{(c, z + v) - f_0}{\varrho(z + v, \text{co extr } \Omega + O^+ \Omega)} \geq \\ &\geq \min\{\alpha_0, \beta_0\} =: \alpha, \end{aligned}$$

что и завершает доказательство теоремы. \blacksquare

Заметим, что последняя оценка в доказательстве теоремы 2.2.2 вытекает из того, что если для положительных числа $a, b, c, d, \alpha, \beta$ выполнены неравенства $\frac{a}{b} \geq \alpha, \frac{c}{d} \geq \beta$, то $\frac{a+c}{b+d} \geq \min\{\alpha, \beta\}$.

Из теоремы 2.2.2 в частности следует, что если решение $\{x_0\} = \Omega$ задачи линейного программирования единственное, то оно удовлетворяет условию острого минимума $f(x) - f_0 \geq \alpha \|x - x_0\|$ для всех $x \in \mathcal{Q}$.

Теорема 2.2.3 (Хоффман). Пусть $\mathcal{Q} \subset \mathbb{R}^n$ есть множество вида (2.2.5), которое не содержит прямых. Тогда найдется число $\alpha > 0$ такое, что

$$\alpha \varrho(x, \mathcal{Q}) \leq \sum_{i=1}^m ((a_i, x) - b_i)_+ \quad \forall x \in \mathbb{R}^n, \quad (2.2.7)$$

где $b^T = (b_1, \dots, b_m)$.

Напомним, что $(\cdot)_+$ — функция положительной срезки, см. с. 87. Отметим также, что число α зависит от векторов a_i и вектора b .

Доказательство. Рассмотрим задачу линейного программирования

$$\min \sum_{i=1}^m t_i \quad \text{при} \quad (a_i, x) - b_i = t_i - z_i, \quad t_i \geq 0, \quad z_i \geq 0, \quad 1 \leq i \leq m.$$

Здесь t_i и z_i — дополнительные скаляры. Для любого $x \in \mathbb{R}^n$ вектор (x, t, z) , где $t_i = ((a_i, x) - b_i)_+$, $z_i = (b_i - (a_i, x))_+$, $1 \leq i \leq m$, является допустимой точкой в этой задаче, а решения задачи имеют вид $(x, 0, (b - Ax)_+)$, $x \in \mathcal{Q}$. Применяя теорему 2.2.2, получаем (2.2.7). \blacksquare

Теорема 2.2.3 утверждает, что если в какой-либо точке невязка в нарушении линейных неравенств мала, то эта точка близка ко множеству допустимых точек. С помощью теоремы 2.2.3 можно получить еще одно утверждение об устойчивости в задаче линейного программирования.

Теорема 2.2.4. *Предположим, что $\mathcal{Q} \subset \mathbb{R}^n$ — множество вида (2.2.5), которое не содержит прямых, множество решений Ω непусто и $f_0 = \min_{\mathcal{Q}} f$. Тогда найдется число $\alpha > 0$ (зависящее от векторов a_i, b, c) такое, что*

$$\alpha \varrho(x, \Omega) \leq \sum_{i=1}^m ((a_i, x) - b_i)_+ + ((c, x) - f_0)_+ \quad \forall x \in \mathbb{R}^n. \quad (2.2.8)$$

Доказательство. Для доказательства запишем

$$\Omega = \{x \in \mathbb{R}^n : Ax \leq b, (c, x) \leq f_0\}$$

и воспользуемся теоремой 2.2.3. ■

Из (2.2.8) следует, что всякая минимизирующая последовательность $\{x_k\}$ вида $(c, x_k) \rightarrow f_0$, $((a_i, x_k) - b_i)_+ \rightarrow 0$ сходится ко множеству решений задачи линейного программирования в том смысле, что $\varrho(x_k, \Omega) \rightarrow +0$.

Заметим, что теоремы 2.2.2, 2.2.3, 2.2.4 можно доказать и для множества \mathcal{Q} вида (2.2.5), которое является цилиндром. Для этого надо рассмотреть $(n-1)$ -мерное сечение, проходящее через исследуемую точку x и перпендикулярное образующей \mathcal{Q} . Если сечение не содержит прямых, то применить соответствующую теорему. Если сечение является цилиндром, то рассмотреть $(n-2)$ -мерное сечение, проходящее через исследуемую точку, и т. д.

Доказательство теоремы Хоффмана для произвольного случая может быть найдено в [6, теорема 6.4].

Поскольку решение задачи линейного программирования достигается в крайних точках и из теоремы 2.2.1 следует их простое описание как решение системы линейных уравнений, то из конечности числа вершин следует возможность найти решение за конечное число шагов путем их перебора. *Симплекс-метод* использует перебор вершин с монотонным убыванием значений функции. Кроме этого, перебор идет по соседним вершинам, поэтому система линейных уравнений, которую требуется решать, мало отличается от предыдущей.

Симплекс-метод применяется для решения канонической задачи линейного программирования в форме (2.2.2).

В вершине множества ограничений $\mathcal{Q} = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$ не более m компонент вектора x могут быть положительны. Вершина x называется *невырожденной*, если число положительных компонент в ней равно m . Предполагается, что допустимое множество непусто, а все его вершины невырождены. Из теоремы 2.2.1 следует, что такие случаи соответствуют рангу матрицы A , равному m , то есть ее строки линейно независимы. Пустоту множества ограничений можно проверить с помощью вспомогательной задачи оптимизации со вспомогательной скалярной переменной w :

$$\inf_{x,w} w \quad \text{при } Ax \leq b + 1_n \cdot w,$$

где 1_n — вектор, состоящий из единиц. Если решение не ограничено снизу или меньше нуля, то множество ограничений непусто.

Опишем одну итерацию симплекс-метода. Пусть на k -м шаге получена точка x_k , являющаяся вершиной допустимого множества, и

$$I_k = \{i : x_{k,i}\}.$$

В силу предположения невырожденности I_k содержит m элементов. Разобьем вектор $x \in \mathbb{R}^n$ на две группы: $x = (u, v)$, где $u \in \mathbb{R}^m$ отвечает компонентам из I_k , а $v \in \mathbb{R}^{n-m}$ — оставшимся компонентам.

Тогда система $Ax = b$ может быть записана в виде

$$A_1 u + A_2 v = b$$

с квадратной матрицей A_1 . Из условия невырожденности и теоремы 2.2.1 матрица A_1 имеет обратную, поэтому можно выразить одни переменные через другие:

$$u = A_1^{-1}(b - A_2 v).$$

Минимизируемая функция принимает вид

$$(c, x) = (c_1, u) + (c_2, v) = (c_2 - A_2^T (A_1^{-1})^T c_1, v) + (c_1, A_1^{-1} b),$$

а исходная задача эквивалентна задаче

$$\min(c_2 - A_2^T (A_1^{-1})^T c_1, v) \quad \text{при } A_1^{-1}(b - A_2 v) \geq 0, v \geq 0. \quad (2.2.9)$$

При этом точке x_k соответствует вектор $u_k > 0, v_k = 0$. Таким образом, ограничение $A_1^{-1}(b - A_2 v_k) \geq 0$ выполняется как строгое неравенство,

поскольку $A_1^{-1}(b - A_2 v_k) = u_k > 0$. Тем самым оно может быть отброшено (как неактивное) при проверке v_k на оптимальность. Но в задаче $\min(d, v)$ при $v \geq 0$ минимум достигается в нуле тогда и только тогда, когда $d = c_2 - A_2^T (A_1^{-1})^T c_1 \geq 0$. В этом случае v_k является решением (2.2.9), а x_k является решением (2.2.2).

Если же среди компонент d есть отрицательные (например, наибольшее по модулю $d_j < 0$), то $v = 0$ не является решением (2.2.9), и, увеличивая v_j , можно уменьшить значение функции (d, v) . Делается шаг

$$v_{k+1} = \alpha_k e_j, \quad j: d_j < 0.$$

Выбор длины шага α_k производится из условия не нарушения ограничения $A_1^{-1}(b - A_2 v_{k+1}) \geq 0$, которое в точке $v_k = 0$ было неактивным, то есть

$$\alpha_k = \max\{\alpha \geq 0\} \quad \text{при } A_1^{-1}(b - \alpha A_2 e_j) \geq 0, \quad j = 1, \dots, n-m.$$

Если $\alpha_k = \infty$, то задача не имеет решения, так как функция (d, v) не ограничена снизу на допустимом множестве. Если же $\alpha_k < \infty$, то получаем новую точку $x_{k+1} = \{u_{k+1}, v_{k+1}\}$, где $u_{k+1} = A_1^{-1}(b - A_2 v_{k+1})$, $v_{k+1} = \alpha_k e_j$.

Эта точка является вершиной допустимого множества, так как она допустима и имеет m положительных компонент, j -я компонента стала положительной, а одна из компонент u_{k+1} обратилась в нуль. При этом

$$(c, x_{k+1}) = (d, v_{k+1}) + (c_1, A_1^{-1}b) < (d, v_k) + (c_1, A_1^{-1}b) = (c, x_k),$$

таким образом, метод формирует монотонно убывающую последовательность. Следовательно, возвращение в какую-либо из ранее проходимых вершин невозможно, и в силу конечности числа вершин допустимого множества симплекс-метод конечен.

В общем случае без предположения о невырожденности всех вершин допустимого множества перебор вершин можно организовать таким образом, чтобы избежать зацикливания.

Несмотря на конечность симплекс-метода, существуют примеры, в которых минимизирующая последовательность обходит все вершины допустимого множества. В. Кли и Дж. Минти показали, что для допустимого множества в виде незначительно модифицированного единичного n -мерного куба симплекс-метод обходит все его 2^n вершин [100]. Приведем его в форме [92]:

$$\max \sum_{j=1}^n 2^{n-j} x_j \quad \text{при } x \geq 0, \quad 2 \sum_{j=1}^{i-1} 2^{i-j} x_j + x_i \leq 5^i, \quad i = 1, \dots, n.$$

В «типичных» задачах симплекс-метод сходится очень быстро, и поэтому он реализован как базовый метод решения задач линейного программирования практически в каждом пакете программ для решения оптимизационных задач.

2.2.2. Квадратичное программирование

Задача минимизации квадратичной формы

$$\min_x \frac{1}{2}(Qx, x) - (c, x) \quad (2.2.10)$$

при линейных ограничениях

$$Ax \leq b \quad (2.2.11)$$

называется *задачей квадратичного программирования*. Здесь $x \in \mathbb{R}^n$, $Q \in \mathbb{S}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Квадратичная форма (Qx, x) является выпуклой в том и только том случае, когда $Q \succcurlyeq 0$; без этого условия задача в общем случае может быть многоэкстремальной. В дальнейшем будем всегда предполагать, что $Q \succcurlyeq 0$.

Линейные ограничения в (2.2.11) можно привести к виду $A_1\tilde{x} = b_1$, $\tilde{x} \geq 0$, аналогично тому, как это делается в задаче линейного программирования. В задаче могут присутствовать ограничения типа равенства $A_2x = b_2$, по которым можно исключить часть переменных.

Общие условия экстремума для выпуклых задач можно конкретизировать.

Теорема 2.2.5 ([20, глава 10, §4, теорема 1]). *Необходимым и достаточным условием экстремума для задачи (2.2.10), (2.2.11) в точке x_0 , $Ax_0 \leq b$, является существование y_0 такого, что*

$$\begin{aligned} Qx_0 - c + A^T y_0 &= 0, \\ y_0 &\geq 0, \\ (y_0)_i(Ax_0 - b)_i &= 0, \quad i = 1, \dots, m. \end{aligned}$$

Последнее условие (так называемое *условие дополняющей нежесткости*) говорит о том, что либо множитель Лагранжа $(y_0)_i$ равен нулю, либо соответствующее ограничение-неравенство находится на своей границе (либо выполняются оба эти условия одновременно).

Для функции Лагранжа определим двойственную функцию $d(y)$:

$$\begin{aligned} L(x, y) &= \frac{1}{2}(Qx, x) - (c, x) + (y, Ax - b), \\ d(y) &= \inf_x L(x, y). \end{aligned}$$

Введем в рассмотрение множество

$$\mathcal{D} = \{y : d(y) \neq -\infty\}.$$

Если квадратичная функция ограничена снизу на \mathbb{R}^n (например, при $Q \succ 0$), то она достигает своего минимума. Поэтому для всякого $y \in \mathcal{D}$ найдется $x = x(y)$ такой, что $L'_x(x, y) = 0$, то есть

$$Qx - c + A^T y = 0.$$

При этом

$$d(y) = L(x(y), y) = -\frac{1}{2}(Qx, x) - (b, y).$$

Таким образом, двойственная задача имеет вид

$$\begin{aligned} & \min_y \frac{1}{2}(Qx, x) + (b, y), \\ & Qx + A^T y = c, \\ & y \geq 0. \end{aligned}$$

В двух важных случаях можно исключить свободную переменную x : при $c = 0$, что приводит к задаче линейного программирования, и при $Q \succ 0$; в последнем случае

$$x(y) = Q^{-1}(c - A^T y),$$

а двойственная функция определена везде: $\mathcal{D} = \mathbb{R}^m$. Двойственная задача принимает вид

$$\begin{aligned} & \min_{y \geq 0} \frac{1}{2}(Q^{-1}(c - A^T y), c - A^T y) + (b, y) = \\ & = \min_{y \geq 0} \frac{1}{2}(AQ^{-1}A^T y, y) + (b - AQ^{-1}c, y). \quad (2.2.12) \end{aligned}$$

Сформулируем полученное утверждение о двойственных задачах в виде теоремы.

Теорема 2.2.6. *При $Q \succ 0$ задача (2.2.10), (2.2.11) двойственна задаче (2.2.12): их решения x_0, y_0 существуют или не существуют одновременно и связаны соотношением*

$$x_0 = Q^{-1}(c - A^T y_0).$$

При этом выполняются неравенства

$$\frac{1}{2}(Qx, x) - (c, x) \geq \frac{1}{2}(Q^{-1}(c - A^T y), c - A^T y) + (b, y)$$

для всех $Ax \leq b$, $y \geq 0$, и равенство возможно лишь при $x = x_0$, $y = y_0$.

Таким образом, при $Q \succ 0$ задача квадратичного программирования сводится к минимизации квадратичной функции на \mathbb{R}_+^m , что позволяет применять методы типа проекции градиента и пр.

С другой стороны, существует возможность решения задачи квадратичного программирования за конечное число шагов, так как минимизация квадратичной функции на линейном многообразии сводится к решению системы линейных уравнений. Поэтому, последовательно решая задачи на гранях многогранника ограничений, можно найти решение.

Например, пусть решается задача (2.2.10), (2.2.11), на k -м шаге найдена допустимая точка x_k и выделено множество активных ограничений $I_k = \{i : (a_i, x_k) = b_i\}$, где $a_i \in \mathbb{R}^n$ — транспонированная i -я строка матрицы A . Найдем решение задачи

$$\min_x \frac{1}{2}(Qx, x) - (c, x), \quad (a_i, x) = b_i, \quad i \in I_k.$$

Решение этой задачи может быть найдено из системы линейных уравнений

$$\begin{aligned} Qx - c + \sum_{j \in I_k} y_j a_j &= 0, \\ (a_i, x) &= b_i, \quad i \in I_k. \end{aligned}$$

Если окажется, что решение этой задачи \bar{x}, \bar{y} таково, что \bar{x} удовлетворяет всем остальным ограничениям $(a_i, \bar{x}) \leq b_i, i \notin I_k$, а $\bar{y} \geq 0, i \in I_k$, то, согласно теореме 2.2.6, \bar{x} является решением задачи.

Если же точка \bar{x} не удовлетворяет этим ограничениям, то берется точка $x_{k+1} = x_k + \lambda(\bar{x} - x_k)$, $\lambda \in (0, 1)$, на отрезке $[x_k, \bar{x}]$, удовлетворяющая всем ограничениям. Так как исходно точка x_k удовлетворяла всем ограничениям, то

$$\lambda = \max_{\substack{i \notin I_k \\ (a_i, \bar{x} - x_k) \neq 0}} \frac{b_i - (a_i, x_k)}{(a_i, \bar{x} - x_k)}.$$

После этого обновляются активные ограничения I_{k+1} и вычисления повторяются. Если точка \bar{x} — допустимая, но есть отрицательные \bar{y}_i , $i \in I_k$, то $x_{k+1} = \bar{x}$, а набор активных ограничений обновляется по множителям Лагранжа: $I_{k+1} = \{i \in I_k : y_i < 0\}$. При условии, что описанный процесс начинается с допустимой точки:

$$Ax_1 \leq b, \quad I_1 = \{i : (a_i, x_1) = b_i\},$$

за конечное число итераций решение будет найдено.

2.3. Методы первого порядка

2.3.1. Метод проекции градиента

Рассмотрим общую задачу минимизации

$$\min_{x \in \mathcal{Q}} f(x). \quad (2.3.1)$$

Пусть множество $\mathcal{Q} \subset \mathbb{R}^n$ замкнуто, а функция f имеет липшицев градиент с константой Липшица $L_1 > 0$. Для решения задачи можно применить следующий метод, называемый *методом проекции градиента*.

Зафиксируем $x_1 \in \mathcal{Q}$, часто выбирается $x_1 \in \partial\mathcal{Q}$. Положим $k = 1$. Если $f'(x_k) = 0$, то точка x_k оказалась стационарной, а в случае выпуклой функции f — решением задачи (2.3.1). Если $f'(x_k) \neq 0$, то выберем некоторый шаг $\alpha_k > 0$ и определим следующую точку $x_{k+1} \in P_{\mathcal{Q}}(x_k - \alpha_k f'(x_k))$. Далее переопределяем $k := k + 1$ и продолжаем итерации.

Перед дальнейшим рассмотрением алгоритма сделаем одно замечание. Итерации метода проекции градиента можно осуществить на практике, если в нашем распоряжении есть эффективный способ нахождения метрической проекции точки на множество \mathcal{Q} . Для множеств простой структуры, например единичной евклидовой сферы с центром в нуле $\mathcal{S}_1 = \{x \in \mathbb{R}^n : \|x\| = 1\}$, метрическая проекция любой точки $x \in \mathbb{R}^n \setminus \{0\}$ легко ищется и равна $P_{\mathcal{S}_1} x = \frac{x}{\|x\|}$. Если множество \mathcal{Q} является многогранником, то проекцию можно находить с помощью решения задачи квадратичного программирования, см. раздел 2.2.2. Существуют классы невыпуклых множеств, для которых легко ищется метрическая проекция точки. Это в первую очередь матричные многообразия, см. приложение B.4. Матричные многообразия активно

исследуются последнее время в связи с многочисленными приложениями [28].

В дальнейших рассуждениях мы будем предполагать, что у нас имеется простое (относительно остальных шагов алгоритма) вычислительное средство, которое позволяет найти метрическую проекцию точки на множество \mathcal{Q} . Такой вспомогательный алгоритм в оптимизации принято называть *оракулом*. Итак, мы предполагаем, что везде ниже имеется оракул, который для любой точки $x \in \mathbb{R}^n$ может вычислить хотя бы один элемент множества $P_{\mathcal{Q}}x$.

Также мы хотим обратить внимание читателя на то, что во многих теоремах мы не требуем выпуклости множества \mathcal{Q} , а иногда и выпуклости функции f .

Рассмотрим свойства изложенного алгоритма. Определение слабо вогнутой функции и суперградиента Фреше даны в разделе Г.7. Функция с L_1 -липшицевым градиентом является слабо вогнутой с константой L_1 , а ее супердифференциал Фреше в любой точке $x \in \mathbb{R}^n$ состоит из одного элемента — градиента $f'(x)$.

Лемма 2.3.1. *Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ слабо вогнута с константой $L_1 > 0$, множество \mathcal{Q} замкнуто, $x_1 \in \mathcal{Q}$, $\alpha \in (0, \frac{1}{L_1})$, суперградиент Фреше $f'(x_1) \in \partial_F^+ f(x_1)$ и $x_2 \in P_{\mathcal{Q}}(x_1 - \alpha f'(x_1))$ произвольны. Тогда*

$$f(x_1) - f(x_2) \geq \frac{1}{2} \left(\frac{1}{\alpha} - L_1 \right) \|x_2 - x_1\|^2.$$

Заметим, что в лемме 2.3.1 множество \mathcal{Q} , как и функция f , невыпукло, поэтому проекция $P_{\mathcal{Q}}(x_1 - \alpha f'(x_1))$, вообще говоря, не одноточечна, см. рис. 2.3.1.

Доказательство. Зафиксируем произвольным образом суперградиент Фреше $f'(x_1) \in \partial_F^+ f(x_1)$. Рассмотрим квадратичную функцию

$$\psi(x) = f(x_1) + (f'(x_1), x - x_1) + \frac{1}{2\alpha} \|x - x_1\|^2.$$

В силу условия слабой вогнутости f с константой L_1 из квадратичной аппроксимации f вида (Г.22) для всякого $x \in \mathbb{R}^n$ выполнена оценка

$$\psi(x) \geq f(x) + \frac{1}{2} \left(\frac{1}{\alpha} - L_1 \right) \|x - x_1\|^2.$$

Для функции $\psi(x)$ точка $x_1 - \alpha f'(x_1)$ является глобальным минимумом. Из того, что линии уровня $\psi(x) = C$ есть сферы с центром

$x_1 - \alpha f'(x_1)$, получаем $\psi(x_2) \leq \psi(x_1)$. Отсюда

$$f(x_1) = \psi(x_1) \geq \psi(x_2) \geq f(x_2) + \frac{1}{2} \left(\frac{1}{\alpha} - L_1 \right) \|x_2 - x_1\|^2,$$

что и завершает доказательство. \blacksquare

Лемма 2.3.1 утверждает, что каждый нетривиальный шаг метода проекции градиента (когда $x_{k+1} \neq x_k$) строго уменьшает значение функции f .

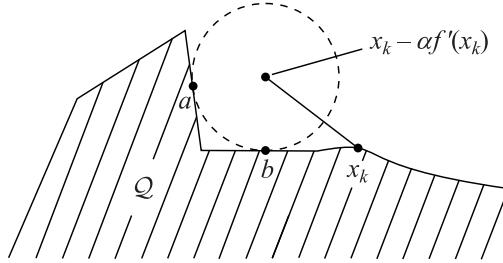


Рис. 2.3.1. Один шаг метода проекции градиента в лемме 2.3.1. Точка x_{k+1} есть один из элементов проекции $P_Q(x_k - \alpha f'(x_k))$: либо $x_{k+1} = a$, либо $x_{k+1} = b$.

Пусть дополнительно к условиям леммы 2.3.1 функция f имеет липшицев градиент с константой $L_1 > 0$. Тогда при дополнительном условии ограниченности f на \mathcal{Q} снизу ($f(x) \geq f_{\inf} \in \mathbb{R}$ для всех $x \in \mathcal{Q}$) и выпуклости \mathcal{Q} каждая сходящаяся подпоследовательность последовательности $\{x_k\}$ метода проекции градиента с $\alpha_k = \alpha \in (0, 1/L_1)$ имеет предел во множестве стационарных точек задачи (2.3.1). Точку $x_0 \in \mathcal{Q}$ такую, что $-f'(x_0) \in N(\mathcal{Q}, x_0)$, будем называть стационарной в задаче (2.3.1), где $N(\mathcal{Q}, x_0)$ — нормальный конус к выпуклому замкнутому множеству \mathcal{Q} в точке $x_0 \in \mathcal{Q}$. Определим

$$\psi_k(x) = f(x_k) + (f'(x_k), x - x_k) + \frac{1}{2\alpha} \|x - x_k\|^2.$$

Пусть $\{x_{k_m}\}$ — сходящаяся подпоследовательность, $x_{k_m} \rightarrow x_* \in \mathcal{Q}$. Запишем необходимое условие экстремума в точке x_{k_m+1} для ψ_{k_m} :

$$\psi'_{k_m}(x_{k_m+1}) \in -N(\mathcal{Q}, x_{k_m+1}).$$

Для любого $m \in \mathbb{N}$ имеем

$$\sum_{k=1}^m (f(x_k) - f(x_{k+1})) = f(x_1) - f(x_{m+1}) \leq f(x_1) - f_{\inf} = C \geq 0,$$

то есть

$$\frac{1}{2} \left(\frac{1}{\alpha} - L_1 \right) \sum_{k=1}^m \|x_{k+1} - x_k\|^2 \leq C.$$

Следовательно, ряд $\sum_{k=1}^{\infty} \|x_{k+1} - x_k\|^2$ сходится и $\lim_{k \rightarrow \infty} \|x_{k+1} - x_k\| = 0$.

Расписывая $\psi'_{k_m}(x_{k_m+1}) \in -N(\mathcal{Q}, x_{k_m+1})$ через f в виде

$$f'(x_{k_m}) + \frac{1}{\alpha}(x_{k_m+1} - x_{k_m}) \in -N(\mathcal{Q}, x_{k_m+1}),$$

переходя к пределу $m \rightarrow \infty$ и пользуясь полуунпрерывностью сверху нормального конуса $\mathcal{Q} \ni x \rightarrow N(\mathcal{Q}, x)$, получаем

$$f'(x_*) \in -N(\mathcal{Q}, x_*).$$

Если множество \mathcal{Q} невыпукло, то аналогичные результаты можно доказать с заменой нормального конуса к выпуклому множеству на соответствующий структуре множества \mathcal{Q} нормальный конус. Например, результат для проксимально гладкого множества \mathcal{Q} можно найти в приложении [Д.4](#).

В случае выпуклой функции можно уточнить результат о сходимости итераций метода проекции градиента.

Теорема 2.3.1. *Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ выпукла и непрерывна по Липшичу градиентом f' с константой Липшица L_1 , множество \mathcal{Q} выпукло и компактно, $\alpha_k = \alpha \in (0, \frac{1}{L_1})$. Если $f_* = \min_{x \in \mathcal{Q}} f(x)$, то существует такая константа $C > 0$, что при всех $k \in \mathbb{N}$*

$$f(x_k) - f_* \leq \frac{C}{k}.$$

Доказательство. Пусть $x_* \in \mathcal{Q}$ — некоторое решение. Тогда

$$\begin{aligned} f(x_k) - f(x_*) &\leq (f'(x_k), x_k - x_*) = (f'(x_k), x_k - x_{k+1}) + \\ &+ \frac{1}{\alpha} (x_k - \alpha f'(x_k) - x_{k+1}, x_* - x_{k+1}) - \frac{1}{\alpha} (x_k - x_{k+1}, x_* - x_{k+1}). \end{aligned}$$

Поскольку

$$x_k - \alpha f'(x_k) - x_{k+1} = x_k - \alpha f'(x_k) - P_{\mathcal{Q}}(x_k - \alpha f'(x_k)) \in N(\mathcal{Q}, x_{k+1}),$$

то

$$(x_k - \alpha f'(x_k) - x_{k+1}, x_* - x_{k+1}) \leq 0$$

и

$$\begin{aligned} f(x_k) - f(x_*) &\leq (f'(x_k), x_k - x_{k+1}) - \frac{1}{\alpha}(x_k - x_{k+1}, x_* - x_{k+1}) \leq \\ &\leq \left(\|f'(x_k)\| + \frac{1}{\alpha}\|x_* - x_{k+1}\| \right) \|x_k - x_{k+1}\|. \end{aligned}$$

В последней формуле выражение в скобках ограничено константой $B > 0$ в силу компактности \mathcal{Q} , а $\|x_k - x_{k+1}\| \rightarrow 0$. Итак, $f(x_k) \rightarrow f(x_*)$.

Пусть $\varphi_k = f(x_k) - f(x_*)$. Из последней оценки $\varphi_k \leq B\|x_k - x_{k+1}\|$. Из леммы 2.3.1 $\varphi_k - \varphi_{k+1} \geq D\|x_k - x_{k+1}\|^2$, где $D = \frac{1}{2}(\frac{1}{\alpha} - L_1)$. Отсюда

$$\varphi_k - \varphi_{k+1} \geq \frac{D}{B^2} \varphi_k^2.$$

Аналогично доказательству теоремы 2.3.7 пункт 2) получаем, что для некоторой константы $C > 0$ выполнена оценка $\varphi_k \leq \frac{C}{k}$. ■

Приведем результаты о линейной сходимости метода проекции градиента.

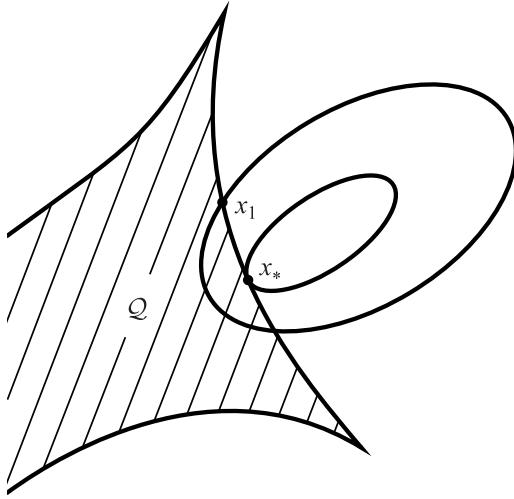


Рис. 2.3.2. Иллюстрация к теореме 2.3.2: изображены лебеговы множества $\{x: f(x) \leq f(x_1)\}$ и $\{x: f(x) \leq f(x_*)\}$.

Теорема 2.3.2 ([74]). Пусть функция f сильно выпукла с константой $\kappa > 0$ и имеет непрерывный по Липшичу градиент f' с константой Липшица L_1 . Пусть множество $\mathcal{Q} \subset \mathbb{R}^n$ — проксимально гладкое с константой $R > 0$, $x_1 \in \mathcal{Q}$ и функция f удовлетворяет условию Липшица на множестве $\mathcal{L}_f(f(x_1)) = \{x \in \mathcal{Q}: f(x) \leq f(x_1)\}$ с константой $L > 0$. Предположим, что $\frac{L}{\kappa} < R$ и $x_* \in \mathcal{Q}$ — решение задачи $\min_{x \in \mathcal{Q}} f(x)$. Зададим достаточно малое положительное число $\alpha < \min\left\{\frac{R}{L}, \frac{2}{L_1}\right\}$. Пусть $x_{k+1} = P_{\mathcal{Q}}(x_k - \alpha f'(x_k))$, $k = 1, 2, \dots$. Тогда последовательности $\{x_k\}$ и $\{f(x_k)\}$ линейно сходятся к x_* и $f(x_*)$, при этом решение x_* единственное.

Доказательство. Отметим, что из определения L для всякой точки $x \in \mathcal{L}_f(f(x_1))$ имеет место оценка

$$\|f'(x)\| \leq L.$$

Решение задачи минимизации x_0 существует в силу сверхлинейного роста f , см. (Г.16). Из пункта 3) свойств проксимально гладких множеств на с. 406, оценки $\|f'(x_k)\| \leq L$ и необходимого условия экстремума $P_{\mathcal{Q}}(x_* - \alpha f'(x_*)) = x_*$ (см. раздел Д.4, условие (Д.8)) имеем

$$\begin{aligned} \|x_{k+1} - x_*\| &= \|P_{\mathcal{Q}}(x_k - \alpha f'(x_k)) - P_{\mathcal{Q}}(x_* - \alpha f'(x_*))\| \leq \\ &\leq \frac{R}{R - \alpha L} \|x_k - x_* - (\alpha f'(x_k) - \alpha f'(x_*))\|. \end{aligned}$$

Отсюда

$$\begin{aligned} \|x_{k+1} - x_*\|^2 &\leq \left(\frac{R}{R - \alpha L}\right)^2 \left(\|x_k - x_*\|^2 - \right. \\ &\quad \left. - 2\alpha(x_k - x_*, f'(x_k) - f'(x_*)) + \alpha^2 \|f'(x_k) - f'(x_*)\|^2 \right). \end{aligned}$$

Из условия Липшица для f' и выпуклости f с учетом неравенства (А.7)

$$L_1(x_k - x_*, f'(x_k) - f'(x_*)) \geq \|f'(x_k) - f'(x_*)\|^2$$

получаем

$$\begin{aligned} \|x_{k+1} - x_*\|^2 &\leq \\ &\leq \left(\frac{R}{R - \alpha L}\right)^2 \left(\|x_k - x_*\|^2 - \alpha(2 - \alpha L_1)(f'(x_k) - f'(x_*), x_k - x_*) \right). \end{aligned}$$

Применяя свойство сильной выпуклости (Г.17), то есть

$$(f'(x_k) - f'(x_*), x_k - x_*) \geq \varkappa \|x_k - x_*\|^2,$$

имеем

$$\|x_{k+1} - x_*\| \leq \|x_k - x_*\| \frac{1}{1 - \frac{\alpha L}{R}} \sqrt{1 - 2\alpha\varkappa + \alpha^2\varkappa L_1}. \quad (2.3.2)$$

В силу леммы 2.3.1 имеем оценку $f(x_{k+1}) \leq f(x_k) \leq f(x_1)$, поэтому $x_{k+1} \in \mathcal{L}_f(f(x_1))$.

При малых $\alpha > 0$

$$\frac{1}{1 - \frac{\alpha L}{R}} \sqrt{1 - 2\alpha\varkappa + \alpha^2\varkappa L_1} \sim 1 + \alpha \left(\frac{L}{R} - \varkappa \right) < 1, \quad \alpha \rightarrow +0.$$

Линейная сходимость $\{x_k\}$ к x_* доказана.

Если предположить, что есть еще стационарная точка $w_0 \in \mathcal{Q}$, то, повторяя предыдущие рассуждения с w_0 вместо x_k , по формуле (2.3.2) получаем

$$\|w_0 - x_*\| \leq \|w_0 - x_*\| \frac{1}{1 - \frac{\alpha L}{R}} \sqrt{1 - 2\alpha\varkappa + \alpha^2\varkappa L_1},$$

причем $\frac{1}{1 - \frac{\alpha L}{R}} \sqrt{1 - 2\alpha\varkappa + \alpha^2\varkappa L_1} < 1$. Отсюда $\|w_0 - x_*\| = 0$.

Определим $D = \max_{k \in \mathbb{N}} \|x_k - x_*\|$. Из верхней квадратичной аппроксимации

$$f(x_k) - f(x_*) \leq (f'(x_*), x_k - x_*) + \frac{L_1}{2} \|x_k - x_*\|^2 \leq \|x_k - x_*\| \left(L + \frac{L_1}{2} D \right)$$

получаем линейную сходимость $f(x_k) \rightarrow f(x_*)$. ■

Заметим, что вместо сильной выпуклости f можно требовать в случае непустоты множества решений Ω выполнение условия¹

$$(f'(x) - f'(P_\Omega x), x - P_\Omega x) \geq \varkappa \|x - P_\Omega x\|^2$$

для любой точки $x \in \mathcal{Q}$ и для некоторой константы $\varkappa > 0$. В качестве $P_\Omega x$ в этом условии может выступать какой-либо элемент метрической проекции x на Ω .

¹ В англоязычной литературе подобное условие принято называть *restricted secant inequality* [99].

Упражнение 2.3.1. Докажите, что при условии теоремы 2.3.2, кроме существования точки x_* , множество минимумов f на \mathcal{Q} не пусто. Таким образом, можно не требовать существование решения x_* .

Упражнение 2.3.2. При $0 \leq \alpha \leq \min\left\{\frac{R}{L}, \frac{2}{L_1}\right\}$ найти минимум функции

$$\varphi(\alpha) = \frac{1}{1 - \frac{\alpha L}{R}} \sqrt{1 - 2\alpha\kappa + \alpha^2\kappa L_1}.$$

Следствие 2.3.1. Если в теореме 2.3.2 множество \mathcal{Q} выпукло, то $R = +\infty$ и для любого $\alpha \in (0, \frac{2}{L_1})$

$$\|x_{k+1} - x_*\| \leq \|x_k - x_*\| \sqrt{1 - 2\alpha\kappa + \alpha^2\kappa L_1}, \quad \sqrt{1 - 2\alpha\kappa + \alpha^2\kappa L_1} < 1.$$

Итак, сильная выпуклость функции, или условие restricted secant inequality, обеспечивает линейную сходимость метода проекции градиента. Рассмотрим условия линейной сходимости при невыпуклости \mathcal{Q} и условии, что функция f не лучше, чем выпуклая. При этом потребуется некоторое новое условие вместо выпуклости множества \mathcal{Q} .

Определение 2.3.1 ([48]). Будем говорить, что компактное множество $\mathcal{Q} \subset \mathbb{R}^n$ удовлетворяет *опорному условию сильной выпуклости* для вектора $p \neq 0$ с радиусом $R > 0$, если

$$\mathcal{Q} \subset \mathcal{B}_R \left(\mathcal{Q}(p) - R \frac{p}{\|p\|} \right). \quad (2.3.3)$$

Здесь $\mathcal{Q}(p) \in \mathcal{Q}$ такая точка, что $(p, \mathcal{Q}(p)) = s(p, \mathcal{Q})$.

Обычно считается, что $\|p\| = 1$, хотя это условие и необязательно.

Заметим, что множество \mathcal{Q} в определении 2.3.1 может быть даже не строго выпуклым.

Приведем простейший пример. Пусть $\mathcal{Q} = \text{co}\{\pm e_n\}$ — отрезок длины 2. Пусть единичный вектор $p \in \mathbb{R}^n$ образует угол $\alpha \in (0, \frac{1}{2}\pi)$ с вектором e_n стандартного ортонормированного базиса. Из элементарной планиметрии легко видеть, что $\mathcal{Q}(p) = e_n$ и выполнено включение $\mathcal{Q} \subset \mathcal{B}_R(e_n - Rp)$ для $R = \frac{1}{\cos \alpha}$. Для всякого единичного вектора p такого, что $\alpha = \frac{\pi}{2}$, множество \mathcal{Q} не удовлетворяет опорному условию сильной выпуклости.

Теорема 2.3.3 ([48]). Пусть компактное множество $\mathcal{Q} \subset \mathbb{R}^n$ удовлетворяет опорному условию сильной выпуклости в направлении

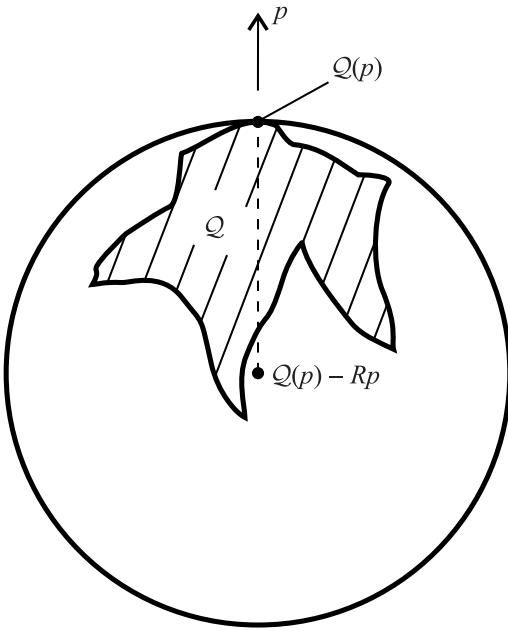


Рис. 2.3.3. Иллюстрация к определению 2.3.1.

вектора $\|p_0\| = 1$ с радиусом R . Допустим, что $r > 0$, $x_0 = \mathcal{Q}(p_0) + rp_0$ и $x \in \mathbb{R}^n$ — точка такая, что $\varrho(x, \mathcal{Q}) = \rho > 0$. Определим проекцию $a_0 = P_{\mathcal{Q}}x_0 = \mathcal{Q}(p_0)$ и зафиксируем произвольную точку $a \in P_{\mathcal{Q}}x$. Пусть множество \mathcal{Q} проксимально гладкое с константой $K > 0$. Тогда

$$\|a_0 - a\| \leq \frac{2R}{2R + r - \frac{R}{K}\rho} \|x_0 - x\|. \quad (2.3.4)$$

Доказательство. Запишем опорное условие 1) сильной выпуклости, см. с. 404

$$\left\| a_0 - \frac{x_0 - a_0}{r} R - a \right\| \leq R.$$

Возводя неравенство в квадрат, получаем

$$\|a_0 - a\|^2 - \frac{2R}{r} (x_0 - a_0, a_0 - a) = \|a_0 - a\|^2 - \frac{2R}{r} (x_0 - a_0 + a - a, a_0 - a) \leq 0.$$

Отсюда следует, что

$$(2R + r)\|a_0 - a\|^2 \leq 2R(x_0 - a, a_0 - a). \quad (2.3.5)$$

Запишем теперь опорное условие 1) для проксимальной гладкости, см. стр. 406

$$\left\| a + \frac{x-a}{\rho} K - a_0 \right\| \geq K.$$

Возводя в квадрат, получаем

$$\|a_0 - a\|^2 + \frac{2K}{\rho} (x - a, a - a_0) \geq 0,$$

откуда

$$(x - a, a_0 - a) \leq \frac{\rho}{2K} \|a_0 - a\|^2. \quad (2.3.6)$$

Из неравенства (2.3.5) следует, что

$$(2R+r)\|a_0 - a\|^2 \leq 2R(x_0 - a, a_0 - a) \leq 2R(x_0 - x, a_0 - a) + 2R(x - a, a_0 - a),$$

а с учетом формулы (2.3.6) приходим к неравенству

$$(2R+r)\|a_0 - a\|^2 \leq 2R(x_0 - a, a_0 - a) \leq 2R(x_0 - x, a_0 - a) + \frac{R\rho}{K} \|a_0 - a\|^2,$$

откуда

$$\left(2R + r - \frac{R\rho}{K}\right) \|a_0 - a\|^2 \leq 2R(x_0 - x, a_0 - a) \leq 2R\|x_0 - x\|\|a_0 - a\|.$$

Теорема доказана. ■

Замечание 2.3.1. Если в условии теоремы $\rho = 0$, то $x = a$ и из формулы (2.3.5) следует, что

$$\begin{aligned} (2R+r)\|a_0 - a\|^2 &\leq 2R(x_0 - a, a_0 - a) = \\ &= 2R(x_0 - x, a_0 - a) \leq 2R\|x_0 - x\|\|a_0 - a\|. \end{aligned}$$

Таким образом, при $\rho = 0$ оценка (2.3.4) также верна. ▼

Замечание 2.3.2. Если в условиях предыдущей теоремы множество \mathcal{Q} выпукло, то $K = +\infty$ и оценка приобретает вид

$$\|a_0 - a\| \leq \frac{2R}{2R+r} \|x_0 - x\|. \quad \blacktriangledown$$

Замечание 2.3.3. Результат предыдущей теоремы не изменится, если считать, что $\|x_0 - a_0\| \geq r$ (и $\frac{x_0 - a_0}{\|x_0 - a_0\|} = p_0$) и $\varrho(x, \mathcal{Q}) \leq \rho$. Это непосредственно вытекает из оценки (2.3.4). ▼

Отметим, что априори проверить условия приведенных ниже теорем обычно невозможно. Поэтому смысл этих теорем следующий. В силу леммы 2.3.1, алгоритм метода проекции градиента является монотонным. Однако если окажется, что в точке, являющейся решением, выполнено условие одной из теорем, то наш алгоритм будет сходиться со скоростью геометрической прогрессии.

Теорема 2.3.4. *Пусть точка $x_* \in \partial\mathcal{Q}$ — решение в задаче (2.3.1). Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ выпукла с липшицевым градиентом f' с константой Липшица $L_1 > 0$ и $f'(x_*) \neq 0$, а компактное множество \mathcal{Q} удовлетворяет опорному условию сильной выпуклости в направлении вектора $-f'(x_*)$ с радиусом $R > 0$. Кроме того, пусть множество \mathcal{Q} —proxимально гладкое с константой $K > 0$. Пусть также $L = \max_{x \in \mathcal{Q}} \|f'(x)\|$ и $\|f'(x_*)\| > \frac{RL}{K}$. Тогда алгоритм проекции градиента с $\alpha_k = \alpha \in (0, \frac{2}{L_1})$ сходится с линейной скоростью:*

$$\|x_* - x_{k+1}\| \leq \frac{2R}{2R + \alpha\|f'(x_0)\| - \alpha L \frac{R}{K}} \|x_* - x_k\|.$$

Если функция f не выпуклая, то при условии $2RL_1 + \frac{RL}{K} < \|f'(x_)\|$ и $\alpha_k = \alpha > 0$ имеет место линейная сходимость:*

$$\|x_* - x_{k+1}\| \leq \frac{2R(1 + \alpha L_1)}{2R + \alpha\|f'(x_*)\| - \alpha L \frac{R}{K}} \|x_* - x_k\|.$$

Доказательство. Обозначим

$$\beta_k = \frac{2R}{2R + \alpha\|f'(x_*)\| - \frac{R}{K}\varrho(x_k - \alpha f'(x_k), \mathcal{Q})}.$$

Тогда с учетом формулы (2.3.4) аналогично теореме 2.3.2 имеем оценку

$$\|x_{k+1} - x_*\|^2 \leq \beta_k^2 \left(\|x_k - x_*\|^2 - \alpha(2 - \alpha L_1)(f'(x_k) - f'(x_*), x_k - x_*) \right).$$

Из монотонности субдифференциала выпуклой функции (Г.4) и условия $0 < \alpha < \frac{2}{L_1}$ вытекает неравенство

$$\alpha(2 - \alpha L_1)(f'(x_k) - f'(x_0), x_k - x_0) \geq 0,$$

откуда

$$\|x_{k+1} - x_*\| \leq \beta_k \|x_k - x_*\| \leq \frac{2R}{2R + \alpha\|f'(x_*)\| - \alpha L \frac{R}{K}} \|x_* - x_k\|.$$

Пусть функция f не выпуклая. Из оценки

$$\begin{aligned} \|x_{k+1} - x_*\|^2 &\leq \beta_k^2 (\|x_k - x_*\|^2 - \\ &\quad - 2\alpha(x_k - x_*, f'(x_k) - f'(x_*)) + \alpha^2 \|f'(x_k) - f'(x_*)\|^2) \end{aligned}$$

в обозначениях теоремы вытекает неравенство

$$\|x_{k+1} - x_*\|^2 \leq \beta_k^2 (1 + \alpha L_1)^2 \|x_k - x_*\|^2 \leq \beta^2 (1 + \alpha L_1)^2 \|x_k - x_*\|^2,$$

где $\beta = \frac{2R}{2R + \alpha \|f'(x_*)\| - \alpha L_1 \frac{R}{K}}$. Из условия $2RL_1 + \frac{RL}{K} < \|f'(x_*)\|$ получаем $\beta(1 + \alpha L_1) < 1$. ■

Пусть в задаче (2.3.1) множество \mathcal{Q} и функция f выпуклы, функция f имеет непрерывный по Липшицу градиент с константой L_1 , а также удовлетворяет условию квадратичного роста на множество \mathcal{Q} , то есть для некоторой константы $\beta > 0$ и всякого $x \in \mathcal{Q}$

$$f(x) - f(P_\Omega x) \geq \beta \|x - P_\Omega x\|^2,$$

где Ω — множество решений (2.3.1). Тогда, как показано в работе [116], при выборе определенного шага $\alpha_k = \alpha \in (0, \frac{1}{L_1}]$ метод проекции градиента сходится по точке с линейной скоростью.

Приведем одну теорему о методе альтернативных проекций.

Теорема 2.3.5 ([50]). *Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклый компакт, а $\mathcal{P} \subset \mathbb{R}^n$ —proxимально гладкое множество с константой $K > 0$. Допустим, что $\mathcal{Q} \cap \mathcal{P} = \emptyset$, $\inf_{x \in \mathcal{P}, y \in \mathcal{Q}} \|x - y\| = \|x_* - y_*\| = r_0 > 0$, $x_* \in \mathcal{P}$, $y_* \in \mathcal{Q}$. Пусть множество \mathcal{Q} удовлетворяет опорному условию сильной выпуклости для вектора $r_0 = \frac{x_* - y_*}{\|x_* - y_*\|}$ с радиусом $R > 0$. Предположим, что $K > 2R + r_0$ и в задаче $\inf_{(x,y) \in \mathcal{P} \times \mathcal{Q}} \|x - y\|$ нет других, кроме (x_*, y_*) , стационарных точек при условии $x \in \mathcal{P}$, $y \in \mathcal{Q}$ и $\|x - y\| \leq r_0 \frac{K}{2R + r_0}$. Тогда при произвольном выборе $x_1 \in \mathcal{P}$, таком, что $\varrho(x_1, \mathcal{Q}) \leq r < r_0 \frac{K}{2R + r_0}$, альтернативные проекции $\{(x_k, y_k)\}_{k \geq 1}$ вида $y_k = P_\mathcal{Q} x_k$, $x_{k+1} = P_\mathcal{P} y_k$, $k = 1, 2, \dots$, сходятся к (x_*, y_*) с линейной скоростью.*

Прежде чем переходить к доказательству теоремы, дадим следующее

Определение 2.3.2. *Стационарной точкой* в задаче

$$\inf_{(x,y) \in \mathcal{P} \times \mathcal{Q}} \|x - y\|$$

будем называть такую точку $(x, y) \in \mathcal{P} \times \mathcal{Q}$, что $P_{\mathcal{Q}}x = y$ и $P_{\mathcal{P}}y = x$.

Обратим внимание, что если множество \mathcal{P} просто выпукло, то в условиях теоремы 2.3.5 точка (x_*, y_*) , где достигается расстояние между множествами \mathcal{Q} и \mathcal{P} , — единственная стационарная точка в задаче $\inf_{(x,y) \in \mathcal{P} \times \mathcal{Q}} \|x - y\|$, $K = +\infty$, и $x_* \in \mathcal{P}$ можно выбирать произвольно.

Доказательство. Если $\varrho(x_1, \mathcal{Q}) = \|x_1 - y_1\| \leq r$, то из определения проекции $\varrho(y_1, \mathcal{P}) = \|y_1 - x_2\| \leq \|y_1 - x_1\| \leq r$ и т. д., $\|x_k - y_k\| \leq r$, $\|x_{k+1} - y_y\| \leq r$ для всех k .

Запишем условие Липшица из замечания 2.3.2

$$\|y_* - y_k\| \leq \frac{2R}{2R + r_0} \|x_* - x_k\|. \quad (2.3.7)$$

Из проксимальной гладкости \mathcal{P} получаем

$$\|x_* - x_{k+1}\| \leq \frac{K}{K - r} \|y_* - y_k\|. \quad (2.3.8)$$

Совместно формулы (2.3.7) и (2.3.8) дают

$$\|x_* - x_{k+1}\| \leq \frac{K}{K - r} \frac{2R}{2R + r_0} \|x_* - x_k\|.$$

В силу условий на константы $\frac{K}{K - r} \frac{2R}{2R + r_0} < 1$.

Оценка для $\|y_k - y_*\|$ устанавливается аналогичным образом. ■

2.3.2. Метод проекции субградиента

Рассмотрим общую задачу минимизации (2.3.1) с выпуклой функцией $f: \mathbb{R}^n \rightarrow \mathbb{R}$ на выпуклом замкнутом множестве $\mathcal{Q} \subset \mathbb{R}^n$:

$$f_* = \min_{x \in \mathcal{Q}} f(x), \quad f_* \in \mathbb{R}, \quad \Omega = \operatorname{Arg} \min_{x \in \mathcal{Q}} f(x) \neq \emptyset.$$

Мы не будем предполагать что-либо о дифференцируемости функции f . Далее считаем, что величина f_* известна. Рассмотрим субградиентный спуск в духе (1.3.32):

$$x_{k+1} = P_{\mathcal{Q}}y_k, \quad y_k = x_k - \frac{f(x_k) - f_*}{\|f'(x_k)\|^2} f'(x_k), \quad f'(x_k) \in \partial f(x_k). \quad (2.3.9)$$

Отличие от формулы (1.3.32) состоит в том, что точка y_k проектируется на множество \mathcal{Q} . Покажем, что алгоритм условной минимизации (2.3.9) сходится с такой же скоростью, что и алгоритм (1.3.32).

Теорема 2.3.6. Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R}$ — выпуклая функция и множество точек Ω минимумов f на \mathcal{Q} непусто. Тогда в методе (2.3.9) $x_k \rightarrow x_* \in \Omega$, причем скорость сходимости задается равенством

$$\liminf_{k \rightarrow \infty} \sqrt{k}(f(x_k) - f_*) = 0.$$

Если выполнено условие острого минимума (1.3.33), то есть для некоторого $\alpha > 0$

$$f(x) - f_* \geq \alpha \varrho(x, \Omega) \quad \forall x \in \mathcal{Q},$$

то

$$\|x_{k+1} - x_*\| \leq \sqrt{1 - \frac{\alpha^2}{c^2}} \|x_k - x_*\|,$$

$$\text{где } c = \sup_{k, f'(x_k) \in \partial f(x_k)} \|f'(x_k)\|.$$

Доказательство. Пусть $\tilde{x} \in \Omega \subset \mathcal{Q}$ произвольная точка. Тогда для любого $f'(x_k) \in \partial f(x_k)$ имеем

$$\begin{aligned} \|x_{k+1} - \tilde{x}\|^2 &= \|P_{\mathcal{Q}}y_k - P_{\mathcal{Q}}\tilde{x}\|^2 \leq \|y_k - \tilde{x}\|^2 = \\ &= \|x_k - \tilde{x}\|^2 - 2 \frac{(f'(x_k), x_k - \tilde{x})(f(x_k) - f_*)}{\|f'(x_k)\|^2} + \\ &\quad + \frac{(f(x_k) - f_*)^2}{\|f'(x_k)\|^2} \leq \|x_k - \tilde{x}\|^2 - \frac{(f(x_k) - f_*)^2}{\|f'(x_k)\|^2}. \end{aligned}$$

Далее доказательство повторяет доказательство теоремы 1.3.9. ■

2.3.3. Метод условного градиента

Снова рассмотрим общую задачу минимизации (2.3.1)

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

Множество $\mathcal{Q} \subset \mathbb{R}^n$ есть выпуклый компакт, а функция f имеет липшицев градиент с константой Липшица $L_1 > 0$.

Предположим, что в нашем распоряжении имеется оракул, позволяющий вычислять опорный элемент множества \mathcal{Q} в направлении любого единичного вектора $p \in \mathbb{R}^n$, то есть элемент множества

$$\mathcal{Q}(p) = \partial s(p, \mathcal{Q}) = \{x \in \mathcal{Q}: (p, x) = s(p, \mathcal{Q})\}.$$

Если множество $\mathcal{Q}(p)$ не одноточечно, то произвольную точку множества $\mathcal{Q}(p)$ также будем обозначать через $\mathcal{Q}(p)$. Примером таких множеств может служить конечная сумма эллипсоидов:

$$\mathcal{Q} = \sum_{k=1}^m A_k \mathcal{B}_1(0),$$

где $A_k \in \mathbb{R}^{n \times n}$ — невырожденная матрица для всех k . В этом случае для $\|p\| = 1$

$$s(p, \mathcal{Q}) = \sum_{k=1}^m s(p, A_k \mathcal{B}_1(0)) = \sum_{k=1}^m \|A_k^\top p\|, \quad \mathcal{Q}(p) = \sum_{k=1}^m \frac{A_k A_k^\top p}{\|A_k^\top p\|}.$$

Другим примером множества \mathcal{Q} , для которого можно эффективно вычислять $\mathcal{Q}(p)$, является множество достижимости линейной управляемой системы. Действительно, для системы $\dot{x} \in Ax + \mathcal{U}$, $x(0) = 0$, множество достижимости $\mathcal{R}_0(t)$ и опорный элемент $(\mathcal{R}_0(t))(p)$ множества достижимости в направлении $\|p\| = 1$ есть

$$\mathcal{R}_0(t) = \int_0^t e^{A(t-s)} \mathcal{U} ds = \int_0^t e^{As} \mathcal{U} ds, \quad (2.3.10)$$

$$(\mathcal{R}_0(t))(p) = \int_0^t (e^{As} \mathcal{U})(p) ds = \int_0^t e^{As} \mathcal{U}(e^{A^\top s} p) ds.$$

Интеграл в (2.3.10) понимается в смысле Аумана:

$$\int_0^t e^{As} \mathcal{U} ds = \left\{ \int_0^t e^{As} u(t) ds : u(\cdot) \in L_1([0, +\infty), \mathcal{U}) \right\},$$

то есть интеграл в фигурных скобках берется по всем измеримым функциям $u(t)$ со значениями во множестве \mathcal{U} .

В рассматриваемом случае можно решать задачу следующим образом. Зафиксируем $x_1 \in \mathcal{Q}$, обычно выбирается точка на границе: $x_1 \in \partial \mathcal{Q}$. Положим $k = 1$. Определим $z_k \in \mathcal{Q}$ как решение вспомогательной задачи $\min_{z \in \mathcal{Q}} (f'(x_k), z)$, то есть $z_k \in \mathcal{Q}(-f'(x_k))$.

Если $f'(x_k) = 0$, то точка x_k — стационарная, а в случае выпуклой функции f — решение задачи (2.3.1). Если $f'(x_k) \neq 0$, то выбираем

некоторый шаг $\alpha_k \in [0, 1]$ и точку $x_{k+1} = x_k + \alpha_k(z_k - x_k)$. Дальше переопределяем $k := k + 1$ и продолжаем итерации, см. рис. 2.3.4.

Полученный алгоритм называется *методом условного градиента*, или *методом Франк–Вульфа*. Приведем теорему о сходимости метода условного градиента.

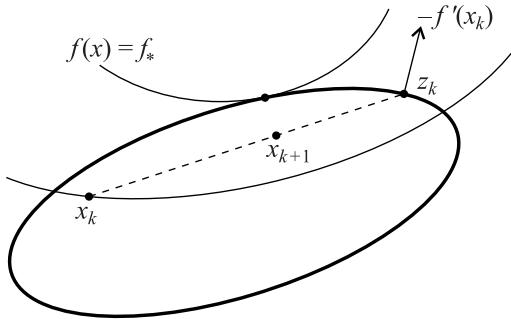


Рис. 2.3.4. Один шаг метода условного градиента. Изображены линии уровня $\{x: f(x) = f^*\}$ и $\{x: f(x) = f(x_k)\}$.

Теорема 2.3.7. Пусть в задаче (2.3.1) градиент функции f удовлетворяет условию Липшица с константой $L_1 > 0$, $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое компактное подмножество. Пусть $f_* = \min_{x \in \mathcal{Q}} f(x)$. Рассмотрим итерации метода условного градиента при

$$\alpha_k = \min \left\{ 1, \gamma_k \frac{(f'(x_k), x_k - z_k)}{\|x_k - z_k\|^2} \right\}, \quad 0 < \varepsilon_1 \leq \gamma_k \leq \frac{2 - \varepsilon_2}{L_1},$$

для некоторых фиксированных чисел $\varepsilon_1, \varepsilon_2 > 0$.

Тогда:

- 1) $\lim_{k \rightarrow \infty} (f'(x_k), x_k - z_k) = 0$ и последовательность $\{f(x_k)\}$ монотонно убывает;
- 2) если дополнительно к предыдущим условиям функция f выпукла, то $f(x_k) - f_* \leq \frac{C}{k}$ для некоторой константы $C > 0$;
- 3) если множество \mathcal{Q} сильно выпукло, функция f выпукла и выполнено условие $\inf_{x \in \mathcal{Q}} \|f'(x)\| \geq \varepsilon > 0$, то сходимость $\{f(x_k)\}$ и $\{x_k\}$ к решению линейная.

Доказательство. Из L_1 -липшицевости градиента f' имеем верх-

нюю квадратичную аппроксимацию $f(x_{k+1})$ вида (A.5):

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq (f'(x_k), x_{k+1} - x_k) + \frac{L_1}{2} \|x_{k+1} - x_k\|^2 = \\ &= -\alpha_k (f'(x_k), x_k - z_k) + \frac{L_1 \alpha_k^2}{2} \|x_k - z_k\|^2. \end{aligned}$$

Рассмотрим два случая.

Случай 1:

$$\gamma_k \frac{(f'(x_k), x_k - z_k)}{\|x_k - z_k\|^2} \geq 1.$$

Тогда $\alpha_k = 1$ и

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq -(f'(x_k), x_k - z_k) + \frac{L_1}{2} \|x_k - z_k\|^2 \leq \\ &\leq (f'(x_k), x_k - z_k) \left(\frac{L_1}{2} \frac{\|x_k - z_k\|^2}{(f'(x_k), x_k - z_k)} - 1 \right) \leq \\ &\leq (f'(x_k), x_k - z_k) \left(\frac{L_1 \gamma_k}{2} - 1 \right) \leq -\frac{\varepsilon_2}{2} (f'(x_k), x_k - z_k) \leq 0. \end{aligned}$$

Случай 2:

$$\gamma_k \frac{(f'(x_k), x_k - z_k)}{\|x_k - z_k\|^2} < 1.$$

Тогда $\alpha_k = \gamma_k \frac{(f'(x_k), x_k - z_k)}{\|x_k - z_k\|^2}$ и

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq -\gamma_k \frac{(f'(x_k), x_k - z_k)^2}{\|x_k - z_k\|^2} + \frac{L_1 \gamma_k}{2} \frac{(f'(x_k), x_k - z_k)^2}{\|x_k - z_k\|^2} \leq \\ &\leq \frac{(f'(x_k), x_k - z_k)^2}{\|x_k - z_k\|^2} \left(-\gamma_k + \frac{L_1 \gamma_k^2}{2} \right) \leq -\frac{\varepsilon_1 \varepsilon_2}{2} \frac{(f'(x_k), x_k - z_k)^2}{\|x_k - z_k\|^2} \leq 0. \end{aligned}$$

Совместно случаи 1 и 2 дают оценку $f(x_{k+1}) \leq f(x_k)$, то есть монотонное убывание $\{f(x_k)\}$. Поскольку непрерывная функция f ограничена на компакте Q , то по теореме Вейерштрасса о пределе монотонной последовательности существует конечный предел $\lim_{k \rightarrow \infty} f(x_k) \in \mathbb{R}$.

Определим невязку $\delta_k = f(x_k) - f(x_{k+1})$, из существования конечного предела $\{f(x_k)\}$ следует $\delta_k \rightarrow 0$. При выполнении случая 1 имеем

$$(f'(x_k), x_k - z_k) \leq \frac{2}{\varepsilon_2} \delta_k. \quad (2.3.11)$$

Определив $D = \text{diam } \mathcal{Q}$, $D \geq \|x_k - z_k\|$, получаем для случая 2

$$(f'(x_k), x_k - z_k) \leq \sqrt{\frac{2}{\varepsilon_1 \varepsilon_2} D^2 \delta_k}. \quad (2.3.12)$$

Из (2.3.11), (2.3.12) и условия $\delta_k \rightarrow 0$ имеем

$$\lim_{k \rightarrow \infty} (f'(x_k), x_k - z_k) = 0.$$

Пункт 1) доказан.

Перейдем к доказательству пункта 2). Пусть функция f выпукла. Тогда для решения $x_* \in \mathcal{Q}$ задачи (2.3.1) и $f_* = f(x_*)$ с учетом неравенства $(f'(x_k), z_k - x_*) \leq 0$ имеем $0 \leq f(x_k) - f_*$ и

$$\begin{aligned} f(x_k) - f_* &\leq (f'(x_k), x_k - x_*) = \\ &= (f'(x_k), x_k - z_k) + (f'(x_k), z_k - x_*) \leq \\ &\leq (f'(x_k), x_k - z_k) \rightarrow 0, \end{aligned} \quad (2.3.13)$$

откуда $f(x_k) \rightarrow f_*$.

Оценим скорость сходимости. Пусть $\varphi_k = f(x_k) - f_*$, $\delta_k = \varphi_k - \varphi_{k+1}$. Из формул (2.3.11), (2.3.12) получаем, что

$$(f'(x_k), x_k - z_k) \leq \max \left\{ \frac{2\delta_k}{\varepsilon_2}, \sqrt{\frac{2D^2}{\varepsilon_1 \varepsilon_2} \delta_k} \right\} \leq K \sqrt{\varphi_k - \varphi_{k+1}}$$

для достаточно больших k и некоторого $K > 0$. С другой стороны, из выпуклости f (см. (2.3.13)) имеем $\varphi_k \leq (f'(x_k), x_k - z_k)$. Отсюда

$$\varphi_{k+1} \leq \varphi_k - \frac{\varphi_k^2}{K^2}.$$

Разделив обе части предыдущего неравенства на $\varphi_k \varphi_{k+1}$, находим

$$\frac{1}{\varphi_k} + \frac{1}{K^2} \leq \frac{1}{\varphi_k} + \frac{\varphi_k}{K^2 \varphi_{k+1}} \leq \frac{1}{\varphi_{k+1}}.$$

Складывая предыдущие неравенства с номерами от 1 до k и сокращая одинаковые слагаемые, получаем

$$\frac{1}{\varphi_1} - \frac{1}{K^2} + \frac{k+1}{K^2} \leq \frac{1}{\varphi_{k+1}},$$

что и означает утверждение пункта 2).

Докажем пункт 3). Пусть множество \mathcal{Q} сильно выпукло с радиусом $R > 0$. Это значит, что модуль выпуклости \mathcal{Q} есть $\delta_{\mathcal{Q}}(t) = \lambda t^2$, где $\lambda = \frac{1}{8R}$, см. эквивалентные свойства сильно выпуклых множеств в приложении Г.7.

Для случая 1 имеем с учетом формулы (2.3.13)

$$f(x_k) - f(x_{k+1}) = \varphi_k - \varphi_{k+1} \geq \frac{\varepsilon_2}{2} (f'(x_k), x_k - z_k) \geq \frac{\varepsilon_2}{2} \varphi_k,$$

то есть $\varphi_{k+1} \leq (1 - \frac{\varepsilon_2}{2}) \varphi_k$.

Для случая 2 имеем

$$f(x_k) - f(x_{k+1}) = \varphi_k - \varphi_{k+1} \geq \frac{\varepsilon_1 \varepsilon_2}{2} \frac{(f'(x_k), x_k - z_k)^2}{\|x_k - z_k\|^2}.$$

Поскольку $\frac{x_k + z_k}{2} + \mathcal{B}_{\delta_{\mathcal{Q}}(\|x_k - z_k\|)}(0) \subset \mathcal{Q}$ и $\delta_{\mathcal{Q}}(t) = \lambda t^2$, то

$$\begin{aligned} \left(-f'(x_k), \frac{x_k + z_k}{2}\right) + \|f'(x_k)\| \cdot \lambda \|x_k - z_k\|^2 &\leq \\ &\leq s(-f'(x_k), \mathcal{Q}) = (-f'(x_k), z_k), \end{aligned}$$

так что

$$(f'(x_k), x_k - z_k) \geq 2\lambda\varepsilon \|x_k - z_k\|^2. \quad (2.3.14)$$

Отсюда

$$\begin{aligned} \varphi_k - \varphi_{k+1} &\geq \frac{\varepsilon_1 \varepsilon_2}{2} \frac{(f'(x_k), x_k - z_k)^2}{\|x_k - z_k\|^2} \geq \\ &\geq \lambda \varepsilon_1 \varepsilon_2 \varepsilon (f'(x_k), x_k - z_k) \geq \lambda \varepsilon_1 \varepsilon_2 \varepsilon \varphi_k, \end{aligned}$$

то есть в случае 2 имеем $\varphi_{k+1} \leq (1 - \lambda \varepsilon \varepsilon_1 \varepsilon_2) \varphi_k$.

Объединяя оба случая, получаем, что

$$\varphi_{k+1} \leq \max \left\{ 1 - \frac{\varepsilon_2}{2}, 1 - \lambda \varepsilon \varepsilon_1 \varepsilon_2 \right\} \varphi_k, \quad \varphi_k = f(x_k) - f_*$$

Из выпуклости f имеем для точки-решения x_* оценку, аналогичную (2.3.14):

$$f(x_k) - f(x_*) \geq (f'(x_*), x_k - x_*) \geq 2\lambda\varepsilon \|x_k - x_*\|^2,$$

$$\text{откуда } \|x_k - x_*\| \leq \sqrt{\frac{f(x_k) - f_*}{2\lambda\varepsilon}}. \quad \blacksquare$$

Будем говорить, что выпуклое компактное множество $\mathcal{Q} \subset \mathbb{R}^n$ *равномерно выпукло* в точке $x_0 \in \partial\mathcal{Q}$, если существует строго положительная при $t > 0$ функция $\delta_{\mathcal{Q},x_0} : [0, \text{diam } \mathcal{Q}] \rightarrow [0, +\infty)$, $\delta_{\mathcal{Q},x_0}(0) = 0$, такая, что для любой точки $x \in \mathcal{Q}$ выполнено включение

$$\frac{x_0 + x}{2} + \delta_{\mathcal{Q},x_0}(\|x_0 - x\|)\mathcal{B}_1(0) \subset \mathcal{Q}.$$

Отличие определения равномерной выпуклости в точке множества \mathcal{Q} от равномерной выпуклости множества \mathcal{Q} состоит в том, что одна из точек в определении модуля выпуклости — точка $x_0 \in \partial\mathcal{Q}$ — фиксируется. Порядок функции $\delta_{\mathcal{Q},x_0}(t) \asymp t^\sigma$ при $t \rightarrow +0$ — не менее второго: $\sigma \geq 2$.

Утверждение 2.3.1 ([132, §5]). *Пусть точка $x_* \in \mathcal{Q}$ является решением в задаче (2.3.1), $f_* = f(x_*)$, функция f выпукла, имеет липшицев градиент с константой L_1 и $\inf_{x \in \mathcal{Q}} \|f'(x)\| \geq \varepsilon > 0$. Пусть множество \mathcal{Q} равномерно выпукло в точке $x_* \in \partial\mathcal{Q}$ с модулем $\delta_{\mathcal{Q},x_*}(t) \asymp t^\sigma$ при $t \rightarrow +0$. Тогда метод условного градиента вида*

$$x_1 \in \partial\mathcal{Q}, \quad z_k \in \operatorname{Arg} \min_{z \in \mathcal{Q}} (f'(x_k), z),$$

$$\alpha_k \in \operatorname{Arg} \min_{\alpha \in [0, 1]} f(x_k + \alpha(z_k - x_k)), \quad x_{k+1} = x_k + \alpha_k(z_k - x_k)$$

сходится со скоростью

$$f(x_k) - f_* \leq \frac{C}{k^{\frac{\sigma}{\sigma-2}}}, \quad C > 0,$$

при $\sigma > 2$ и с линейной скоростью при $\sigma = 2$.

Другие случаи в задаче (2.3.1), для которых имеет место линейная сходимость метода условного градиента, — выпуклые многогранники и сильно выпуклая функция. За деталями мы отсылаем читателя к обзору [81].

В ряде случаев в методе условного градиента в задаче (2.3.1) для сильно выпуклого множества \mathcal{Q} можно обойтись без промежуточной одномерной минимизации f на отрезке $[x_k, z_k]$ и рассматривать алгоритм условного градиента вида $x_1 \in \partial\mathcal{Q}$, $x_{k+1} \in \operatorname{Arg} \max_{x \in \mathcal{Q}} (-f'(x_k), x)$.

Лемма 2.3.2. *Пусть выпуклое компактное множество $\mathcal{Q} \subset \mathbb{R}^n$ удовлетворяет условию (2.3.3) в направлении p , $\|p\| = 1$, с радиусом*

R. Тогда для всякого единичного вектора $q \in \mathbb{R}^n$ и произвольного элемента опорного множества $\mathcal{Q}(q)$ (который также обозначим через $\mathcal{Q}(q)$) выполнено неравенство

$$\|\mathcal{Q}(p) - \mathcal{Q}(q)\| \leq 2R\|p - q\|.$$

Доказательство. Зафиксируем единичный вектор $q \in \mathbb{R}^n$ и элемент множества $\mathcal{Q}(q)$. Будем полагать без ограничения общности, что $\mathcal{Q}(p) - Rp = 0$.

Пусть $(p, q) \geq 0$. Точка $\mathcal{Q}(q)$ находится во множестве $S = \mathcal{B}_R(0) \cap H_q^+$ (где $H_q^+ = \{x \in \mathbb{R}^n : (q, x) \geq (q, \mathcal{Q}(p))\}$), которое является сферическим сегментом. Пусть $\xi = [0, Rq] \cap \partial H_q^+$. Тогда $\|\xi - Rp\| \leq R\|q - p\|$ (так как ξ — метрическая проекция Rq на ∂H_q^+) и

$$\|\mathcal{Q}(q) - \mathcal{Q}(p)\| \leq \text{diam } S = 2\|\xi - Rp\|,$$

поскольку $(p, q) \geq 0$ и диаметр S реализуется на основании сегмента.

Если $(p, q) < 0$, то $\|p - q\|^2 = 2 - 2(p, q) > 2$, $\|p - q\| > \sqrt{2}$ и

$$\frac{\|\mathcal{Q}(p) - \mathcal{Q}(q)\|}{\|p - q\|} \leq \frac{2R}{\sqrt{2}} = \sqrt{2}R.$$

Лемма доказана. ■

Теорема 2.3.8. ([48]) Пусть точка $x_* \in \partial \mathcal{Q}$ — решение в задаче (2.3.1). Предположим, что $\|f'(x)\| \geq m > 0$ для всех $x \in \mathcal{Q}$, градиент f' удовлетворяет условию Липшица с константой $L_1 > 0$ и множество \mathcal{Q} удовлетворяет опорному условию сильной выпуклости для вектора $-f'(x_0)$ с константой R . Пусть также $\frac{2RL_1}{m} < 1$.

Тогда итерации

$$x_1 \in \mathcal{Q}, \quad x_{k+1} \in \text{Arg} \max_{x \in \mathcal{Q}} (-f'(x_k), x)$$

сходятся к x_* с линейной скоростью:

$$\|x_* - x_{k+1}\| \leq \frac{2RL_1}{m} \|x_* - x_k\|.$$

Доказательство. Заметим, что выпуклый компакт со \mathcal{Q} также удовлетворяет опорному условию сильной выпуклости для вектора $-f'(x_0)$ с константой R . Для вектора $p \neq 0$ через $\mathcal{Q}(p)$ будем обозначать произвольную точку из \mathcal{Q} такую, что $(p, \mathcal{Q}(p)) = s(p, \mathcal{Q})$, см. раздел Г.1. Через $(\text{co } \mathcal{Q})(p)$ будем обозначать множество точек

$$\{x \in \text{co } \mathcal{Q} : (p, x) = s(p, \text{co } \mathcal{Q})\}.$$

Упражнение 2.3.3. Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — произвольное множество. Докажите, что $s(p, \text{co } \mathcal{Q}) = s(p, \mathcal{Q})$ для любого $p \in \mathbb{R}^n$.

Из равенства

$$\|x_* - x_{k+1}\| = \|\mathcal{Q}(-f'(x_*)) - \mathcal{Q}(-f'(x_k))\|,$$

соотношений

$$\mathcal{Q}(-f'(x_k)) \in (\text{co } \mathcal{Q})(-f'(x_k)), \quad \mathcal{Q}(-f'(x_*)) = (\text{co } \mathcal{Q})(-f'(x_*))$$

(см. упражнение 2.3.4) и леммы 2.3.2 имеем

$$\begin{aligned} \|\mathcal{Q}(-f'(x_*)) - \mathcal{Q}(-f'(x_k))\| &\leqslant \\ &\leqslant 2R \left\| \frac{f'(x_*)}{\|f'(x_*)\|} - \frac{f'(x_k)}{\|f'(x_k)\|} \right\| \leqslant \frac{2R}{m} \|f'(x_*) - f'(x_k)\| \leqslant \\ &\leqslant \frac{2RL_1}{m} \|x_* - x_k\|. \end{aligned}$$

Теорема доказана. ■

Заметим, что в отличие от известных ранее результатов в теореме 2.3.8 ни функция $f(x)$, ни множество \mathcal{Q} не предполагаются выпуклыми.

Упражнение 2.3.4. Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — компактное подмножество и $p \in \mathbb{R}^n$ — единичный вектор. Докажите, что $\mathcal{Q}(p) \subset (\text{co } \mathcal{Q})(p)$.

2.4. Метод отсекающей гиперплоскости. Метод эллипсоидов

Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое компактное подмножество с непустой внутренностью и $f: \mathbb{R}^n \rightarrow \mathbb{R}$ — выпуклая функция. Как обычно, мы хотим найти минимум функции f на множестве \mathcal{Q} . В этом разделе дифференцируемость функции f не предполагается. Однако, как известно (см. теорему Г.13), субдифференциал функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ есть выпуклый компакт для любой точки $x \in \mathbb{R}^n$.

Определим центр масс множества \mathcal{Q} :

$$x_{\mathcal{Q}} = \frac{\int_{\mathcal{Q}} x d\mu}{\mu_{\mathcal{Q}}},$$

где μ — мера Лебега (в данной ситуации можно рассматривать и меру Жордана) в \mathbb{R}^n . В силу непустоты внутренности \mathcal{Q} имеем $\mu\mathcal{Q} > 0$.

Зафиксируем произвольный субградиент $f'(x_{\mathcal{Q}}) \in \partial f(x_{\mathcal{Q}})$. Из субградиентного неравенства, см. определение Г.9, для всякого $x \in \mathcal{Q}$ выполнена оценка

$$f(x) \geq f(x_{\mathcal{Q}}) + (f'(x_{\mathcal{Q}}), x - x_{\mathcal{Q}}). \quad (2.4.1)$$

Если $f'(x_{\mathcal{Q}}) = 0$, то точка $x_{\mathcal{Q}} \in \mathcal{Q}$ является решением задачи. Далее будем считать, что $f'(x_{\mathcal{Q}}) \neq 0$.

Рассмотрим гиперплоскость

$$H_{x_{\mathcal{Q}}} = \{x \in \mathbb{R}^n : (f'(x_{\mathcal{Q}}), x - x_{\mathcal{Q}}) = 0\}$$

и полупространства

$$H_{x_{\mathcal{Q}}}^+ = \{x \in \mathbb{R}^n : (f'(x_{\mathcal{Q}}), x - x_{\mathcal{Q}}) > 0\},$$

$$H_{x_{\mathcal{Q}}}^- = \{x \in \mathbb{R}^n : (f'(x_{\mathcal{Q}}), x - x_{\mathcal{Q}}) \leq 0\}.$$

Из неравенства (2.4.1) следует, что решение не может находиться во множестве $\mathcal{Q} \cap H_{x_{\mathcal{Q}}}^+$. Значит, оно находится во множестве $\tilde{\mathcal{Q}} = \mathcal{Q} \cap H_{x_{\mathcal{Q}}}^-$.

Теорема 2.4.1 (Грюнбаум – Хаммер [4]). *Пусть в принятых выше обозначениях $\tilde{\mathcal{Q}} = \mathcal{Q} \cap H_{x_{\mathcal{Q}}}^-$. Тогда*

$$\frac{\mu\tilde{\mathcal{Q}}}{\mu\mathcal{Q}} \leq 1 - \frac{1}{e}.$$

Возникает следующий алгоритм, который иногда называется *методом отсекающей гиперплоскости*.

Положим $\mathcal{Q}_0 = \mathcal{Q}$ и рассмотрим итеративную процедуру:

$$x_k = x_{\mathcal{Q}_k},$$

$$\xi_k \in \{x_i\}_{i=0}^k : f(\xi_k) = \min_{0 \leq i \leq k} f(x_i), \quad k = 0, 1, \dots, \quad (2.4.2)$$

$$\mathcal{Q}_{k+1} = \mathcal{Q}_k \cap H_{x_k}^-.$$

Покажем, что значения $f(\xi_k)$, полученные в алгоритме (2.4.2), сходятся с линейной скоростью к значению f_* минимума функции f на множестве \mathcal{Q} . Определим $\text{Var } f \doteq \max_{\mathcal{Q}} f - f_*$.

Теорема 2.4.2. В результате работы алгоритма (2.4.2) имеет место оценка

$$f(\xi_k) \leq f_* + 2 \operatorname{Var} f \cdot \left(1 - \frac{1}{e}\right)^{k/n}.$$

Доказательство. Случай $n = 1$ очевиден, поэтому рассмотрим размерность $n \geq 2$. Без ограничения общности будем считать, что точка $0 \in \mathcal{Q}_k$ есть точка глобального минимума f на \mathcal{Q}_k и тем самым на \mathcal{Q}_0 . Для $k > 0$ выберем $\alpha \in (0, 1)$ так, что

$$\alpha > \left(\frac{\mu\mathcal{Q}_k}{\mu\mathcal{Q}_0}\right)^{1/n} \geq \frac{\alpha}{2}. \quad (2.4.3)$$

Поскольку число $\left(\frac{\mu\mathcal{Q}_k}{\mu\mathcal{Q}_0}\right)^{1/n} < 1$, то такое α существует.

Из левой части двойного неравенства в (2.4.3) следует неравенство $\alpha^n \mu\mathcal{Q}_0 > \mu\mathcal{Q}_k$, то есть $\alpha\mathcal{Q}_0 \not\subseteq \mathcal{Q}_k$. Поэтому найдется точка $\tilde{x} \in \alpha\mathcal{Q}_0$ такая, что $\tilde{x} \notin \mathcal{Q}_k$. Определим также такую точку $\xi \in \mathcal{Q}_0$, что $\tilde{x} = \alpha\xi$.

Так как $0 \in \mathcal{Q}_k$ и $\alpha\xi \notin \mathcal{Q}_k$, то из выпуклости множества \mathcal{Q}_k следует, что $\xi \notin \mathcal{Q}_k$. Поэтому, с учетом равенства $f_* = f(0)$ и правой оценки в (2.4.3), получаем

$$\begin{aligned} f(\xi_k) &\leq f(\tilde{x}) = f(\alpha\xi + (1 - \alpha) \cdot 0) \leq \\ &\leq \alpha f(\xi) + (1 - \alpha)f(0) = f_* + \alpha(f(\xi) - f_*) \leq \\ &\leq f_* + 2 \operatorname{Var} f \cdot \frac{\alpha}{2} \leq f_* + 2 \operatorname{Var} f \cdot \left(\frac{\mu\mathcal{Q}_k}{\mu\mathcal{Q}_{k-1}} \cdot \frac{\mu\mathcal{Q}_{k-1}}{\mu\mathcal{Q}_{k-2}} \cdots \frac{\mu\mathcal{Q}_1}{\mu\mathcal{Q}_0}\right)^{1/n}. \end{aligned}$$

Применяя теорему 2.4.1, получаем искомое утверждение. ■

Результат теоремы 2.4.2 был получен А.Ю. Левиным [56] и Д. Ньюманом [119]. Одна из основных проблем в практической применимости теоремы 2.4.2 состоит в том, что вычисление центра масс $x_{\mathcal{Q}}$ выпуклого тела \mathcal{Q} является сложной задачей. В то же время многие другие точки, связанные со множеством \mathcal{Q} , не гарантируют выполнение утверждения 2.4.1.

Н.З. Шором был предложен следующий алгоритм, который получил название «метод эллипсоидов». Рассмотрим замкнутый евклидов шар B_0 с центром в точке $x_0 \in \mathbb{R}^n$, который содержит множество $\mathcal{Q}_0 = \mathcal{Q}$. Если $x_0 \notin \mathcal{Q}_0$, то положим $y_0 = P_{\mathcal{Q}_0}x_0$, $p_0 = \frac{x_0 - y_0}{\|x_0 - y_0\|}$, и рассмотрим полушар

$$B_0^{0,5} = B_0 \cap \{x \in \mathbb{R}^n : (p_0, x - x_0) \leq 0\},$$

который содержит целиком множество \mathcal{Q}_0 .

Рассмотрим случай $x_0 \in \mathcal{Q}_0$. Если $f'(x_0) = 0$ для некоторого субградиента $f'(x_0) \in \partial f(x_0)$, то задача решена. Если $f'(x_0) \neq 0$ для всякого элемента субдифференциала $\partial f(x_0)$, то зафиксируем субградиент $f'(x_0) \in \partial f(x_0)$ и рассмотрим полушар

$$B_0^{0,5} = B_0 \cap \{x \in \mathbb{R}^n : (f'(x_0), x - x_0) \leq 0\}.$$

Легко видеть, что в обоих случаях $B_0^{0,5} \cap \mathcal{Q}_0 \neq \emptyset$ и минимум f на \mathcal{Q}_0 содержится в $B_0^{0,5} \cap \mathcal{Q}_0 \subset B_0^{0,5}$. Покажем, что полушар $B_0^{0,5}$ можно поместить в эллипсоид B_1 , объем которого меньше, чем объем B_0 . Без ограничения общности будем считать, что $B_0 = \mathcal{B}_1(0)$, а $B_0^{0,5}$ задается соотношением

$$B_0^{0,5} = \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \leq 1, x_n \geq 0 \right\}.$$

Покажем, что множество $B_0^{0,5}$ можно вписать в эллипсоид с центром $(0; 0; \dots; 0; \frac{1}{n+1})$ и полуосами $\alpha > 0$ по координатам x_i , $1 \leq i \leq n-1$, и $\frac{n}{n+1}$ по x_n . Отметим, что такой эллипсоид содержит точку $(0; \dots; 0; 1)$. Потребуем, чтобы граница эллипсоида ∂B_1 содержала также точки единичной сферы $\partial \mathcal{B}_1(0) \cap \{x \in \mathbb{R}^n : x_n = 0\}$.

Найдем α . Рассматривая двумерное сечение эллипсоида B_1 плоскостью $Ox_1x_n = \{x \in \mathbb{R}^n : x_i = 0, i \neq 1, i \neq n\}$, получаем уравнение эллипса $\frac{x_1^2}{\alpha^2} + (x_n - \frac{1}{n+1})^2 \left(\frac{n+1}{n}\right)^2 = 1$. Эллипс проходит через точку $x_1 = 1, x_n = 0$, откуда $\alpha = \frac{n}{\sqrt{n^2-1}}$.

Объем полученного эллипсоида равен

$$\mu B_1 = \mu \mathcal{B}_1(0) \sigma_n, \quad \text{где } \sigma_n = \frac{n^n}{(n-1)^{(n-1)/2} (n+1)^{(n+1)/2}}.$$

Упражнение 2.4.1. Исследуйте функцию

$$f(x) = \frac{x^x}{(x-1)^{(x-1)/2} (x+1)^{(x+1)/2}}$$

и докажите, что при $x \geq 2$ выполнено неравенство $f(x) < 1$.

Таким образом, центры эллипсоидов B_0, B_1 и т. д. выступают вместо центров масс $x_{\mathcal{Q}_i}$. Процесс можно продолжать: за счет линейной замены координат превратить B_1 в единичный шар и т. д. Повторяя рассуждения теоремы 2.4.2, получаем линейную со знаменателем σ_n

сходимость алгоритма (2.4.2) с описанными эллипсоидами и их центрами x_i по функции.

В заключение заметим, что в описанном методе величина σ_n стремится к 1 снизу при $n \rightarrow \infty$. Так, при $n = 50$ величина σ_{50} становится больше 0,99. Это делает применение указанного алгоритма малоэффективным в пространствах большой размерности.

2.5. Метод Ньютона в задачах условной минимизации

Идея шага метода проекции градиента состоит в том, чтобы из текущей точки $x_1 \in \mathcal{Q}$ множества \mathcal{Q} отступить в точку $x_1 - \alpha f'(x_1)$, где $f'(x_1)$ — градиент функции f в точке x_1 , а $\alpha > 0$ — шаг метода. Для функции

$$\psi(x) = f(x_1) + (f'(x_1), x - x_1) + \frac{1}{2\alpha} \|x - x_1\|^2$$

точка $x_1 - \alpha f'(x_1)$ является глобальным минимумом на \mathbb{R}^n , см. лемму 2.3.1, а линии уровня $\psi(x) = C$ являются шарами, так как график ψ — параболоид вращения. Далее, проектирование на \mathcal{Q} , то есть поиск следующей точки $x_2 = P_{\mathcal{Q}}(x_1 - \alpha f'(x_1))$, аналитически есть поиск минимума функции $\psi(x)$ при $x \in \mathcal{Q}$. При этом локальное приближение исходной функции f функцией ψ в окрестности x_1 обычно является достаточно неточным, особенно в случае, когда собственные числа матрицы $f''(x_1)$ сильно различаются, а линии уровня f имеют «овражный» характер.

Возникает следующая идея. Если позволяет гладкость функции f , то вместо приближения на текущем шаге f параболоидом ψ в окрестности точки x_1 можно рассмотреть разложение f по формуле Тейлора до второго порядка малости и аппроксимировать f в окрестности точки x_1 с помощью функции

$$\varphi(x) = f(x_1) + (f'(x_1), x - x_1) + \frac{1}{2}(f''(x_1)(x - x_1), x - x_1).$$

В качестве следующей точки итераций выбирается минимум функции $\varphi(x)$ при $x \in \mathcal{Q}$. Наличие квадратичного члена $\frac{1}{2}(f''(x_1)(x - x_1), x - x_1)$ вместо $\frac{1}{2\alpha} \|x - x_1\|^2$ позволяет более точно учесть геометрическую специфику функции f . Описанная идея впервые была реализована в работе [57].

Пусть $x_k \in \mathcal{Q}$. Определим

$$f_k(x) = (f'(x_k), x - x_k) + \frac{1}{2}(f''(x_k)(x - x_k), x - x_k), \quad (2.5.1)$$

а следующую точку $x_{k+1} \in Q$ определим из условия

$$f_k(x_{k+1}) \leq f_k(x) \quad \forall x \in \mathcal{Q}.$$

Теорема 2.5.1. Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое подмножество, $x_1 \in \mathcal{Q}$, и $f: \mathbb{R}^n \rightarrow \mathbb{R}$ — дважды дифференцируемая функция, причем для всех $x, z \in \mathcal{Q}$ и $y \in \mathbb{R}^n$ выполнены условия

$$\begin{aligned} m\|y\|^2 &\leq (f''(x)y, y) \leq M\|y\|^2, \quad m > 0, \\ \|f''(x) - f''(z)\| &\leq L_2\|x - z\|, \quad L_2 > 0. \end{aligned} \quad (2.5.2)$$

Пусть также $\delta = (2L_2/m)\|x_2 - x_1\| < 1$, где x_2 есть минимум f_1 (2.5.1) на \mathcal{Q} . Тогда последовательность $\{x_k\}$, $x_{k+1} \in \operatorname{Arg} \min_{x \in \mathcal{Q}} f_k(x)$, сходится к точке $x_* \in \mathcal{Q}$ минимума f на \mathcal{Q} , причем для всякого $l \in \mathbb{N}$

$$\|x_l - x_*\| \leq \frac{m}{2L_2} \sum_{k=l}^{\infty} \delta^{2^{k-1}}.$$

Доказательство. Так как $f_k(x)$ достигает минимума на \mathcal{Q} в точке x_{k+1} , то запишем критерий экстремума для выпуклой функции f на выпуклом множестве \mathcal{Q} вида $(f'_k(x_{k+1}), x - x_{k+1}) \geq 0$ для всех $x \in \mathcal{Q}$.

Упражнение 2.5.1. Докажите приведенный выше критерий.

Следовательно,

$$(f'_k(x_{k+1}), x_{k+1} - x_k) = (f'(x_k) + f''(x_k)(x_{k+1} - x_k), x_{k+1} - x_k) \leq 0.$$

Поэтому

$$\begin{aligned} f_k(x_{k+1}) &= (f'(x_k), x_{k+1} - x_k) + \frac{1}{2}(f''(x_k)(x_{k+1} - x_k), x_{k+1} - x_k) \leq \\ &\leq -\frac{1}{2}(f''(x_k)(x_{k+1} - x_k), x_{k+1} - x_k) \leq -\frac{m}{2}\|x_{k+1} - x_k\|^2. \end{aligned}$$

Таким образом,

$$\|x_{k+1} - x_k\|^2 \leq -\frac{2}{m}f_k(x_{k+1}). \quad (2.5.3)$$

С другой стороны,

$$\begin{aligned} f_k(x_{k+1}) &= (f'(x_k), x_{k+1} - x_k) + \\ &+ \frac{1}{2}(f''(x_k)(x_{k+1} - x_k), x_{k+1} - x_k) \geq (f'(x_k), x_{k+1} - x_k). \end{aligned}$$

Далее, в силу (2.5.2),

$$f'(x_k) = f'(x_{k-1}) + f''(x_{k-1})(x_k - x_{k-1}) + r, \quad \text{где } \|r\| \leq L_2 \|x_k - x_{k-1}\|^2.$$

Поэтому

$$\begin{aligned} -f_k(x_{k+1}) &\leq -(f'(x_k), x_{k+1} - x_k) = \\ &= -(f'(x_{k-1}) + f''(x_{k-1})(x_k - x_{k-1}) + r, x_{k+1} - x_k) = \\ &= -(f'_{k-1}(x_k), x_{k+1} - x_k) + (r, x_{k+1} - x_k) \leq (r, x_{k+1} - x_k) \leq \\ &\leq \|r\| \|x_{k+1} - x_k\| \leq L_2 \|x_k - x_{k-1}\|^2 \|x_{k+1} - x_k\|. \end{aligned}$$

Объединяя последнюю оценку с неравенством (2.5.3), получаем

$$\|x_{k+1} - x_k\| \leq \frac{2L_2}{m} \|x_k - x_{k-1}\|^2.$$

Отсюда по индукции следует, что

$$\|x_{k+1} - x_k\| \leq \frac{m}{2L_2} \delta^{2^{k-1}},$$

поэтому при $l > k$ имеем

$$\|x_l - x_k\| \leq \sum_{i=k}^{l-1} \|x_{i+1} - x_i\| \leq \frac{m}{2L_2} \sum_{i=k}^{l-1} \delta^{2^{i-1}}.$$

Таким образом, последовательность $\{x_k\}$ фундаментальна. Следовательно $\lim_{k \rightarrow \infty} x_k = x_* \in \mathcal{Q}$ и

$$\|x_k - x_*\| \leq \frac{m}{2L_2} \sum_{i=k}^{\infty} \delta^{2^{i-1}}.$$

Предельным переходом для непрерывной функции f_k (2.5.1) получаем равенство

$$\min_{x \in \mathcal{Q}} \left(f(x_*) + (f'(x_*), x - x_*) + \frac{1}{2} (f''(x_*)(x - x_*), x - x_*) \right) = f(x_*), \quad (2.5.4)$$

что является необходимым и, в силу выпуклости задачи, достаточным условием минимума x_* при минимизации f на \mathcal{Q} . \blacksquare

Упражнение 2.5.2. Докажите, что условие (2.5.4) необходимо для того, чтобы точка $x_* \in Q$ была решением задачи $\min_Q f$.

Заметим, что условие $\delta = (2L_2/m)\|x_2 - x_1\| < 1$, характеризующее близость x_1 к точке минимума x_0 , допускает конструктивную проверку, в отличие от вычисления $\|x_1 - x_*\|$ при неизвестной точке x_* .

Рассмотрим пример. Для задачи нелинейного программирования с линейными ограничениями

$$\min f(x) \quad \text{при условии } Ax \leq b, \quad x \in \mathbb{R}^n, \quad b \in \mathbb{R}^m,$$

применение метода Ньютона сводится к последовательному по $k \in \mathbb{N}$ решению задач квадратичного программирования

$$\min_{Ax \leq b} \left((f'(x_k), x - x_k) + \frac{1}{2} (f''(x_k)(x - x_k), x - x_k) \right);$$

см. раздел 2.2.2.

2.6. Метод штрафных функций

Рассмотрим задачу условной оптимизации с ограничениями вида

$$\min_{g(x) \leq 0, h(x) = 0} f(x), \quad (2.6.1)$$

где $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $h: \mathbb{R}^n \rightarrow \mathbb{R}^\ell$.

На эту задачу можно посмотреть с позиций безусловной оптимизации, добавив ограничения $g(x) \leq 0$ и $h(x) = 0$ тем или иным образом к минимизируемой функции $f(x)$ в виде *штрафных членов*, построив так называемую *штрафную функцию*.

Мы рассмотрим простейший случай — штрафную функцию с *квадратичным штрафом* вида

$$f(x) + M \|g_+(x)\|^2 + N \|h(x)\|^2. \quad (2.6.2)$$

Здесь M и N — некоторые положительные параметры (*коэффициенты штрафа*), а $g_{+,i}(x) = \max\{0, g_i(x)\}$ — покомпонентная функция положительной срезки. Введение квадратичных штрафных членов позволяет сохранить дифференцируемость минимизируемой функции (конечно, при условии дифференцируемости функций $g(x)$ и $h(x)$), однако в общем случае штрафные члены могут быть линейными или иными монотонными функциями, обращающимися в нуль на допустимом множестве.

Рассмотрим семейство задач

$$\begin{aligned} \min_{x \in Q_0} f_k(x), \quad k = 1, 2, \dots, \\ f_k(x) = f(x) + M_k \|g_+(x)\|^2 + M_k \|h(x)\|^2, \end{aligned} \tag{2.6.3}$$

при общем коэффициенте штрафа $M_k \rightarrow \infty$. Здесь Q_0 — некоторое ограниченное замкнутое множество локализации минимума функции $f(x)$ (например, шар достаточно большого радиуса).

Теорема 2.6.1. *Пусть функции f , g и h непрерывны, а множество Ω решений задачи (2.6.1) непусто. Тогда всякая предельная точка последовательности*

$$x_k = \arg \min_{x \in Q_0} f_k(x), \quad k = 1, 2, \dots,$$

где множество Q_0 ограничено и замкнуто, $Q_0 \cap \Omega \neq \emptyset$, а функции $f_k(x)$ задаются соотношениями (2.6.3), является решением задачи (2.6.1).

Доказательство. Прежде всего заметим, что метод (2.6.3) корректно определен, так как функции $f_k(x)$ непрерывны, а множество Q_0 ограничено и замкнуто. Поскольку $x_k \in Q_0$, то существует хотя бы одна предельная точка \tilde{x} последовательности $\{x_k\}$.

Пусть $x_* \in Q_0 \cap \Omega$ — точка минимума в задаче (2.6.1), тогда из (2.6.3) имеем

$$f_k(x_k) \leq f_k(x_*) = f(x_*).$$

Отсюда

$$\|g_+(x_k)\|^2 + \|h(x_k)\|^2 \leq \frac{1}{M_k} (f(x_*) - f(x_k)),$$

и из ограниченности $f(x_*) - f(x_k)$ на компактном множестве при взятии предела $k \rightarrow \infty$ следует $\|g_+(\tilde{x})\| = \|h(\tilde{x})\| = 0$, то есть

$$g_i(\tilde{x}) \leq 0, \quad i = 1, \dots, m, \quad h_j(\tilde{x}) = 0, \quad j = 1, \dots, \ell.$$

Таким образом \tilde{x} — допустимая точка задачи (2.6.1).

С другой стороны,

$$\begin{aligned} f(x_k) &\leq f(x_*) - M_k (\|g_+(x_k)\|^2 + \|h(x_k)\|^2) = \\ &= f(x_*) - M_k (\|g_+(x_k)\|^2 + \|h(x_k)\|^2), \end{aligned}$$

откуда в пределе имеем $f(\tilde{x}) \leq f(x_*)$, поэтому \tilde{x} является решением задачи (2.6.1). \blacksquare

В общем случае вместо задачи условной оптимизации

$$\min_{x \in \mathcal{Q}} f(x) \quad (2.6.4)$$

можно рассмотреть задачу безусловной оптимизации со штрафным членом:

$$\min_{x \in \mathbb{R}^n} (f(x) + M g_{\mathcal{Q}}(x)), \quad (2.6.5)$$

где коэффициент M достаточно велик. Функция $g_{\mathcal{Q}}(x)$ играет роль либо *внешнего штрафа*, увеличиваясь при отдалении от \mathcal{Q} :

$$g_{\mathcal{Q}}(x) = \begin{cases} 0, & x \in \mathcal{Q}, \\ \rightarrow \infty, & \varrho(x, \mathcal{Q}) \rightarrow \infty, \end{cases}$$

(к таковым относятся штрафные члены в (2.6.2)), либо роль *внутреннего штрафа*, «запрещая» приближаться к границам \mathcal{Q} :

$$g_{\mathcal{Q}}(x) = \begin{cases} \infty, & x \notin \mathcal{Q}, \\ \rightarrow \infty, & x \in \mathcal{Q}, \varrho(x, \partial \mathcal{Q}) \rightarrow 0. \end{cases}$$

Внутренние штрафы иногда называют *барьерами*, а функцию $g_{\mathcal{Q}}(x)$ — *барьерной функцией*.

Особенность внешних штрафов в том, что если решение исходной задачи (2.6.4) находится на границе \mathcal{Q} , то решение задачи (2.6.5) со штрафным членом, как правило, находится вне \mathcal{Q} . Аналогично теореме 2.6.1 легко показать, что с увеличением коэффициента штрафа M_k предельная точка последовательности

$$x_k = \arg \min_{x \in \mathbb{R}^n} (f(x) + M_k g_{\mathcal{Q}}(x))$$

будет принадлежать множеству \mathcal{Q} . Поэтому при практическом применении внешних штрафов важно иметь возможность оценивать расстояние $\varrho(x, \mathcal{Q})$ до допустимого множества и вычислять проекцию $P_{\mathcal{Q}}(x)$ на него.

Особенность внутренних штрафов состоит в том, что при применении численных методов порождаемая с их помощью последовательность точек $\{x_k\}$ всегда принадлежит множеству \mathcal{Q} ; это налагает требование определения допустимой начальной точки $x_1 \in \mathcal{Q}$.

На методе штрафных функций основан *метод внутренней точки* (см. подробнее [13, 42]), который является основным средством решения важного класса задач выпуклой оптимизации, рассматриваемого в следующем разделе, — задач полуопределенного программирования.

2.7. Полуопределенное программирование

2.7.1. Линейные матричные неравенства

Рассмотрим линейную матричнозначную функцию $F: \mathbb{R}^\ell \rightarrow \mathbb{R}^{n \times n}$ векторного аргумента $x = (x_1 \ \dots \ x_\ell)^T \in \mathbb{R}^\ell$:

$$F(x) = F_0 + \sum_{i=1}^{\ell} x_i F_i, \quad (2.7.1)$$

где $F_i \in \mathbb{S}^{n \times n}$, $i = 0, \dots, \ell$, — известные матрицы, называемые иногда *коэффициентами*, а x_i , $i = 1, \dots, \ell$, — скалярные переменные.

Запись

$$F(x) \prec 0, \quad (2.7.2)$$

в которой неравенство понимается в смысле знакопределенности матрицы, называется *линейным матричным неравенством в канонической форме* относительно переменных x_1, \dots, x_ℓ . Строгая знакопределенность матрицы $F(x) \prec 0$ определяется соответствующей ей квадратичной формой:

$$F(x) \prec 0 \iff (F(x)y, y) = y^T F(x)y < 0 \quad \forall y \in \mathbb{R}^n, \quad y \neq 0.$$

Нестрогое неравенство определяется аналогично:

$$F(x) \preccurlyeq 0 \iff (F(x)y, y) = y^T F(x)y \leq 0 \quad \forall y \in \mathbb{R}^n.$$

Легко заметить, что матричное неравенство (2.7.2) эквивалентно числовому неравенству $\lambda_{\max}(F(x)) < 0$, но функция $\lambda_{\max}(F(x))$ нелинейна по x при $n > 1$. При единичной размерности матриц ($n = 1$) функция (2.7.1) линейная.

Строго говоря, правильнее было бы использовать термин «аффинное матричное неравенство», поскольку функция (2.7.1) аффинна, но по сложившейся традиции принято наименование «линейное матричное неравенство» (ЛМН).

Несмотря на кажущийся специальный вид линейных матричных неравенств, они представляют собой весьма богатый класс ограничений, в том числе встречающихся в теории управления. Принципиально важным свойством ЛМН является выпуклость множества его решений (она доказывается непосредственно по определению), которая позволяет формулировать многие задачи оптимального управления в виде задач выпуклого программирования.

Приведем еще несколько простых полезных свойств ЛМН, непосредственно следующих из свойств знакопределенных матриц.

Например, неравенство обратного знака сводится к неравенству вида (2.7.2) очевидным образом: $F(x) \succ 0$ тогда и только тогда, когда $-F(x) \prec 0$, а запись

$$F(x) \prec G(x)$$

для линейных функций $F(x), G(x)$ вида (2.7.1) эквивалентна матричному неравенству

$$F(x) - G(x) \prec 0.$$

Далее, несколько линейных матричных неравенств

$$F_j(x) \prec 0, \quad j = 1, \dots, m,$$

эквивалентны одному блочно-диагональному ЛМН

$$\begin{pmatrix} F_1(x) & & \\ & \ddots & \\ & & F_m(x) \end{pmatrix} \prec 0.$$

Еще одно часто используемое свойство заключается в сохранении знака строгого ЛМН при умножении слева и справа на произвольную невырожденную матрицу M :

$$F(x) \prec 0 \implies MF(x)M^T \prec 0, \quad \det M \neq 0.$$

Для нестрогого ЛМН это свойство выполняется независимо от вырожденности или невырожденности матрицы M , в том числе для неквадратных матриц (ср. со свойствами знакопределенных матриц в приложении Б.1).

Как правило, приходится иметь дело с линейными матричными неравенствами, в которых переменными являются не скаляры, как в (2.7.2), а матрицы. Типичными примерами таких ЛМН являются условие на знакопределенность матрицы: $Q \succ 0$, а также классическое неравенство Ляпунова вида

$$A^T Q + QA \prec 0$$

или

$$A^T Q + QA + R \preccurlyeq 0, \tag{2.7.3}$$

где переменной является симметричная матрица Q .

Как будет показано далее, ЛМН в матричной форме естественным образом возникают в задачах управления. При необходимости матричная форма легко может быть приведена к каноническому виду. Например, рассмотрим ЛМН (2.7.3), и пусть

$$E_1, \dots, E_\ell, \quad \ell = \frac{n(n+1)}{2},$$

есть базис в пространстве \mathbb{S}^n . Тогда, имея представление $Q = \sum_{i=1}^{\ell} \mathbf{x}_i E_i$ для $Q \in \mathbb{S}^n$, вместо (2.7.3) получаем

$$\underbrace{R}_{F_0} + \sum_{i=1}^{\ell} \mathbf{x}_i \underbrace{(E_i A + A^T E_i)}_{F_i} \preceq 0.$$

Введем следующее важное понятие.

Определение 2.7.1. Совокупность точек

$$\mathcal{D}_{\text{feas}} = \{x \in \mathbb{R}^\ell : F(x) \preceq 0\}$$

называется *допустимой областью* линейного матричного неравенства $F(x) \preceq 0$.

Очевидно, что $\mathcal{D}_{\text{feas}}$, как совокупность решений ЛМН, представляет собой выпуклую область, возможно неограниченную, возможно пустую.

Сделаем два важных замечания.

Замечание 2.7.1. В дальнейшем иногда придется решать задачи отыскания точной нижней грани функции на открытом множестве, заданном *строгими* ЛМН-ограничениями, причем это множество будет заведомо непусто (то есть $\mathcal{D}_{\text{feas}}$ содержит внутреннюю точку). Будем заменять такую задачу на задачу отыскания минимума этой функции на замыкании ограничивающего множества. ▼

Замечание 2.7.2. Как правило, в приложениях возникают именно *нестрогие* неравенства вида $F(x) \preceq 0$, и тому есть несколько причин. Прежде всего, при решении задач стабилизации (в теории автоматического управления) обычно требуется не только обеспечить устойчивость, то есть выполнение *строгого* неравенства

$$-\operatorname{Re} \lambda_i(A_c) > 0$$

для собственных значений матрицы замкнутой системы, но и гарантировать некоторую степень устойчивости, то есть выполнение *нестрогого неравенства*

$$-\operatorname{Re} \lambda_i(A_c) \geqslant \sigma > 0.$$

Аналогичным образом почти всегда от решения $P \in \mathbb{R}^{n \times n}$ системы ЛМН требуется не просто положительная определенность ($P \succ 0$), но выполнение этого свойства с некоторым «запасом»: $P \succcurlyeq \delta I$, где $\delta > 0$ задано. Это условие, в частности, означает, что эллипсоид

$$\mathcal{E} = \{x \in \mathbb{R}^n : x^T P^{-1} x \leqslant 1\},$$

определенный матрицей P , содержит шар радиуса $\delta^{1/2}$, то есть имеет заданную параметром δ «степень невырожденности».

Наконец, чуть ниже будет сформулирована одна из основных задач теории ЛМН — так называемая задача полуопределенного программирования, в самом названии которой содержится указание на нестрогость линейного матричного неравенства. ▼

Перейдем к рассмотрению следующего примера.

Пример 2.7.1. Пусть

$$F(x) = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} + x_1 \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Видим, что точка $x^0 = (0 \ 0)^T$ принадлежит допустимому множеству и собственные значения матрицы $F(x^0)$ строго отрицательны. Нарушение знакопределенности симметричной матрицы может произойти лишь если одно из собственных значений обратится в нуль, то есть

$$\det F(x) = 0 \tag{2.7.4}$$

при некотором x . Решая это нелинейное уравнение, находим, что допустимое множество

$$\mathcal{D}_{\text{feas}} = \{(x_1 \ x_2)^T : x_1^2 + x_2^2 \leqslant 1\}$$

является единичным кругом с центром в начале координат.

Из (2.7.4) видно, что при $\ell = 2$, $n = 2$, допустимая область может представлять собой любое выпуклое множество в \mathbb{R}^2 , задаваемое кривыми второго порядка. Так например, допустимым множеством неравенства

$$F(x) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} + x_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \prec 0$$

является внутренность ветви гиперболы $x_1 x_2 = -1$, расположенной во второй координатной четверти. ▼

2.7.2. Задача полуопределенного программирования

Теперь сформулируем две основные проблемы теории линейных матричных неравенств.

Определение 2.7.2. Задача разрешимости (допустимости) заключается в отыскании некоторой точки $x \in \mathcal{D}_{\text{feas}}$ или в доказательстве того, что такой точки не существует.

Забегая вперед (см. раздел 3.1.1) отметим, что типичным примером задачи допустимости является отыскание общей квадратичной функции Ляпунова для семейства линейных стационарных систем

$$\dot{x} = A_i x, \quad i = 1, \dots, k,$$

что сводится к проверке разрешимости конечного числа ЛМН

$$A_i^T Q + Q A_i \preceq 0, \quad i = 1, \dots, k,$$

относительно матричной переменной $Q \succ 0$.

Следующая важная задача ЛМН-теории — оптимизация линейного критерия на множестве, заданном линейными матричными неравенствами.

Определение 2.7.3. Задача минимизации линейной функции

$$\min c^T x, \quad c \in \mathbb{R}^n,$$

при ЛМН-ограничениях

$$F_i(x) \preceq 0, \quad i = 1, \dots, \ell,$$

относительно переменной $x \in \mathbb{R}^n$ называется задачей полуопределенного программирования, или задачей SDP (от *Semi-Definite Programming*).

Очевидно, что задача максимизации линейной функции $c^T x$ эквивалентно сводится к стандартной SDP заменой ее знака.

Замечание 2.7.3. Правильнее было бы определить задачу SDP как отыскание не минимума, а *точной нижней грани* линейной целевой функции на допустимом множестве. Действительно, при (да же и замкнутом) неограниченном D_{feas} оптимальное значение целевой

функции может быть конечным, но при этом не достигаться в конечной точке. Например, ЛМН из примера 2.7.1 задает допустимое множество как внутренность ветви гиперболы, и минимальное значение линейной функции $g(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_2$, равное нулю, не достигается в конечной точке D_{feas} . Разумеется, такая ситуация возможна и для задач с матричными переменными, но, как правило, такие проблемы далее встречаться не будут. ▼

Ясно, что задача допустимости является частным случаем задачи SDP. В самом деле, рассмотрим линейное матричное неравенство

$$F(x) \preceq 0 \quad (2.7.5)$$

и предположим, что $\hat{\gamma}, \hat{x}$ — решение задачи SDP

$$\min \gamma \quad \text{при ограничении } F(x) \preceq \gamma I,$$

где минимизация проводится по переменным $\gamma \in \mathbb{R}$ и $x \in \mathbb{R}^\ell$. Очевидно, что ЛМН (2.7.5) разрешимо тогда и только тогда, когда $\hat{\gamma} \leq 0$; при этом \hat{x} является допустимой точкой.

Для задачи SDP можно использовать следующую запись:

$$\min \gamma \quad \text{при } F(x, \gamma) \preceq 0,$$

где функция $F(x, \gamma)$ аффинна по совокупности переменных x, γ . Действительно, в задаче из определения 2.7.3 введем дополнительную скалярную переменную γ и составим неравенство

$$c^T x \leq \gamma.$$

Поскольку оно является линейным матричным неравенством со скалярными коэффициентами c_i , добавим его к исходному набору ЛМН-ограничений $F(x) \preceq 0$. Минимизация величины γ на множестве, задаваемом такой расширенной системой ЛМН, эквивалентна исходной задаче SDP.

Отсюда следует, что спецификой задачи SDP в сравнении с общей задачей минимизации выпуклой функции на выпуклом множестве является вид ограничений. Оказывается, что ЛМН-ограничения представляют весьма широкий класс выпуклых множеств. Как было показано выше, в этот класс входят множества, задаваемые линейными ограничениями (таким образом, задача линейного программирования может формулироваться как SDP), выпуклыми квадратичными ограничениями (как в элементарном примере 2.7.1), матричными ограничениями типа неравенства Ляпунова. В терминах ЛМН можно формулировать ограничения на матричную норму, выпуклые квадратичные

матричные неравенства, условия на взаимно-обратные матрицы, а также описывать широкие классы неопределенностей, присутствующих в системе управления.

Задача полуопределенного программирования является одним из основных форматов, к которому будут сводиться задачи управления, подавления возмущений, фильтрации и т. д., рассматриваемые в главе 3.

Опишем несколько полезных технических приемов, которые позволяют сформировать задачу полуопределенного программирования. Первый из них заключается в упрощении целевой функции.

Лемма 2.7.1. Задача

$$\min \operatorname{tr} L_1(P) \quad \text{при ограничении } L_2(P) \preceq 0,$$

где $P = P^T$ — матричная переменная, а $L_1(P)$, $L_2(P)$ — некоторые линейные функции, эквивалентна задаче SDP

$$\min \operatorname{tr} Z \quad \text{при ограничениях } L_2(P) \preceq 0, \quad L_1(P) \preceq Z, \quad (2.7.6)$$

где оптимизация проводится по матричным переменным $P = P^T$ и $Z = Z^T$ той же размерности, что и $L_1(P)$.

Доказательство. В силу монотонности следа матрицы имеем

$$\operatorname{tr} Z \geq \operatorname{tr} L_1(P).$$

Таким образом, решение задачи (2.7.6) дает оценку сверху для решения исходной задачи.

С другой стороны, если \widehat{P} — решение исходной задачи, то, полагая

$$Z = L_1(\widehat{P}),$$

получаем $\operatorname{tr} Z = \operatorname{tr} L_1(\widehat{P})$, и при этом второе неравенство в (2.7.6) выполнено. Таким образом, \widehat{P} , Z — решение задачи (2.7.6). ■

Нетрудно видеть, что полученный результат будет справедлив не только для функции следа, но и для любой монотонной непрерывной функции $\varphi(\cdot)$, то есть такой, что

$$F \preceq G \implies \varphi(F) \leq \varphi(G),$$

например, $\varphi(\cdot) = \|\cdot\|$.

2.7.3. Лемма Шура и ее применение

В качестве одного из наиболее эффективных технических средств в рамках ЛМН-теории часто используется так называемая *лемма Шура*, позволяющая сводить некоторые нелинейные матричные неравенства к линейным, в частности, типовые задачи с обратными матрицами и квадратичными членами. Приведем ее в следующей формулировке.

Лемма 2.7.2 (Шур [127]). Пусть

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{pmatrix} \in \mathbb{R}^{(n+m) \times (n+m)},$$

где $M_{11} \in \mathbb{S}^n$, $M_{12} \in \mathbb{R}^{n \times m}$, $M_{22} \in \mathbb{S}^m$.

Тогда

$$M \prec 0 \iff M_{22} \prec 0, \quad M_{11} - M_{12}M_{22}^{-1}M_{12}^T \prec 0.$$

Доказательство. Введем в рассмотрение вектор

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^{n+m},$$

разбиение которого соответствует разбиению матрицы M . Тогда

$$\begin{aligned} x^T M x &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \\ &= x_1^T M_{11} x_1 + x_1^T M_{12} x_2 + x_2^T M_{12}^T x_1 + x_2^T M_{22} x_2 = \\ &= (M_{22}^{-1} M_{12}^T x_1 + x_2)^T M_{22} (M_{22}^{-1} M_{12}^T x_1 + x_2) + \\ &\quad + x_1^T (M_{11} - M_{12} M_{22}^{-1} M_{12}^T) x_1. \end{aligned}$$

Таким образом, заключаем, что $M \prec 0$ тогда и только тогда, когда $M_{22} \prec 0$ и $M_{11} - M_{12} M_{22}^{-1} M_{12}^T \prec 0$. ■

Матрица $M_{11} - M_{12} M_{22}^{-1} M_{12}^T$ называется *дополнением по Шуре* к блоку M_{22} в матрице M .

Замечание 2.7.4. Лемма Шура справедлива в следующей «симметричной» формулировке:

$$M \prec 0 \iff M_{11} \prec 0, \quad M_{22} - M_{12}^T M_{11}^{-1} M_{12} \prec 0,$$

которая оказывается в равной мере полезной. ▼

Следующий результат представляет собой модификацию леммы Шура для *нестрого* матричного неравенства.

Лемма 2.7.3. *Пусть*

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{pmatrix} \in \mathbb{R}^{(n+m) \times (n+m)},$$

где $M_{11} \in \mathbb{S}^n$, $M_{12} \in \mathbb{R}^{n \times m}$, $M_{22} \in \mathbb{S}^m$.

Если $M_{22} \prec 0$, то

$$M \preccurlyeq 0 \iff M_{11} - M_{12}M_{22}^{-1}M_{12}^T \preccurlyeq 0.$$

Из леммы Шура для нестрогих неравенств и свойств определенных матриц вытекает следующее важное следствие.

Следствие 2.7.1. *Пусть матрица*

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{pmatrix} \preccurlyeq 0,$$

причем $M_{22} \prec 0$, а M_{12} — матрица максимального ранга. Тогда

$$M_{11} \prec 0.$$

Упражнение 2.7.1. Сформулируйте утверждение, аналогичное лемме 2.7.2, для случая положительно определенной матрицы M и утверждение, аналогичное лемме 2.7.3, для случая неотрицательно определенной матрицы M . Дайте им обоснование.

Основное применение леммы Шура — сведение нелинейных матричных неравенств к линейным, в частности переход от обратных матриц к прямым; приведем наиболее важные примеры.

Пример 2.7.2 (Неравенство Риккати). Рассмотрим матричное неравенство Риккати

$$A^T Q + QA - QBS^{-1}B^T Q + R \succ 0, \quad S \succ 0,$$

относительно матричной переменной $Q \succ 0$; оно часто встречается в задачах оптимального управления и является квадратичным по Q . «Разворачивая» это матричное неравенство с помощью леммы Шура (иными словами, применяя ее в «обратную сторону»), приходим к эквивалентной записи в виде ЛМН:

$$\begin{pmatrix} A^T Q + QA + R & QB \\ B^T Q & S \end{pmatrix} \succ 0,$$

откуда, в частности, следует, что множество решений неравенства Риккати выпукло (что совсем не очевидно!).

Вообще, пусть матричнозначные функции $F_{ij}(x)$, $i, j = 1, 2$, аффинны по векторной переменной $x \in \mathbb{R}^\ell$, причем

$$F_{11}(x) = F_{11}^T(x), \quad F_{22}(x) = F_{22}^T(x),$$

и требуется удовлетворить неравенствам

$$F_{22}(x) \prec 0, \quad F_{11}(x) - F_{12}(x)F_{22}^{-1}(x)F_{12}^T(x) \prec 0,$$

второе из которых, вообще говоря, нелинейно по x .

По лемме Шура получаем, что они эквивалентны одному линейному матричному неравенству

$$F(x) = \begin{pmatrix} F_{11}(x) & F_{12}(x) \\ F_{12}^T(x) & F_{22}(x) \end{pmatrix} \prec 0.$$

Кроме того, поскольку ЛМН $F(x) \prec 0$ выпукло, то и исходное неравенство оказывается выпуклым. \blacktriangledown

Пример 2.7.3 (Ограничение на норму матрицы). Пусть имеется ограничение

$$\|F(x)\|_2 \leq 1$$

на спектральную норму аффинной матричной функции $F: \mathbb{R}^\ell \rightarrow \mathbb{R}^{n \times n}$ векторной переменной $x \in \mathbb{R}^\ell$.

По лемме B.1 последнее неравенство эквивалентно матричному неравенству

$$F^T(x)F(x) \preceq I,$$

которое по лемме Шура эквивалентно ЛМН

$$\begin{pmatrix} I & F(x) \\ F^T(x) & I \end{pmatrix} \succcurlyeq 0$$

относительно переменной x .

В частности, ограничение $\|x\| \leq 1$ на величину евклидовой нормы вектора $x \in \mathbb{R}^n$ можно записать в линейно-матричном виде

$$\begin{pmatrix} 1 & x^T \\ x & I \end{pmatrix} \succcurlyeq 0. \quad \blacktriangledown$$

Приведенный пример имеет важные применения. Так, если переменными являются матрицы, то, рассматривая задачу минимизации максимального собственного значения матрицы $P \succ 0$ (или, что тоже самое, ее спектральной нормы) при некотором ЛМН-ограничении $L(P) \leq 0$, видим, что эта задача эквивалентна следующей задаче SDP:

$$\min \gamma \quad \text{при ограничениях } P \preceq \gamma I, \quad P \succ 0, \quad L(P) \leq 0$$

с переменными $P \in \mathbb{S}^n$ и $\gamma \in \mathbb{R}$.

Аналогичным образом, если требуется минимизировать величину $x_0^T P^{-1} x_0$ (иными словами, норму вектора $P^{-1/2} x_0$, где x_0 фиксировано) на решениях линейного матричного неравенства $L(P) \leq 0$, то вводя скалярную переменную γ из расчета

$$x_0^T P^{-1} x_0 \leq \gamma \quad (2.7.7)$$

видим, что по лемме Шура это неравенство представимо в виде ЛМН

$$\begin{pmatrix} \gamma & x_0^T \\ x_0 & P \end{pmatrix} \succcurlyeq 0$$

относительно $P = P^T$ и $\gamma \in \mathbb{R}$. Таким образом, минимизация величины $x_0^T P^{-1} x_0$ эквивалентна задаче SDP

$$\min \gamma \quad \text{при ограничениях} \quad \begin{pmatrix} \gamma & x_0^T \\ x_0 & P \end{pmatrix} \succcurlyeq 0, \quad P \succ 0, \quad L(P) \leq 0. \quad (2.7.8)$$

Пример 2.7.4 (Оценка обратной матрицы). Пусть имеется нелинейное ограничение

$$F(x) \preceq G^{-1}(x) \prec 0,$$

где F, G — симметричные аффинные матричнозначные функции от $x \in \mathbb{R}^\ell$. По лемме Шура получаем, что оно эквивалентно ЛМН

$$\begin{pmatrix} F(x) & I \\ I & G(x) \end{pmatrix} \preceq 0. \quad \blacktriangledown$$

Часто используется следующий прием, позволяющий свести минимизацию некоторой специальной нелинейной функции матричного аргумента при ЛМН-ограничениях к эквивалентной задаче SDP.

Лемма 2.7.4. *Задача*

$$\min \operatorname{tr} Y P^{-1} Y^T \quad \text{при ограничении } L(P, Y) \leq 0,$$

где $P = P^T$, Y — матричные переменные соответствующей размерности, а $L(P, Y)$ — некоторая линейная функция, эквивалентна задаче SDP

$$\min \operatorname{tr} Z \quad \text{при ограничениях } L(P, Y) \leq 0, \quad \begin{pmatrix} Z & Y \\ Y^T & P \end{pmatrix} \succ 0, \quad (2.7.9)$$

где оптимизация проводится по матричным переменным $P = P^T$, Y , $Z = Z^T$.

Доказательство. Заметим, что аргумент $Y P^{-1} Y^T$ целевой функции имеет структуру дополнения по Шуру к блоку Z в матрице

$$\begin{pmatrix} Z & Y \\ Y^T & P \end{pmatrix},$$

поэтому, вводя новую матричную переменную $Z = Z^T$ соответствующей размерности и требуя выполнения второго ограничения в (2.7.9), по лемме Шура имеем в силу монотонности следа матрицы:

$$\operatorname{tr} Z \geq \operatorname{tr} Y P^{-1} Y^T.$$

Таким образом, решение задачи (2.7.9) даст оценку сверху для решения исходной задачи.

С другой стороны, если \hat{P} , \hat{Y} — решение исходной задачи, то, полагая $Z = \hat{Y} \hat{P}^{-1} \hat{Y}^T$, получаем $\operatorname{tr} Z = \operatorname{tr} \hat{Y} \hat{P}^{-1} \hat{Y}^T$, и при этом второе неравенство в (2.7.9) выполнено, то есть \hat{P} , \hat{Y} , Z — решение задачи (2.7.9). ■

Отметим, что этот результат остается справедливым, если вместо следа используется любая иная монотонная функция.

Из леммы 2.7.4 вытекает важное следствие.

Следствие 2.7.2. *Задача*

$$\min \operatorname{tr} P^{-1} \quad \text{при ограничениях } L(P) \leq 0, \quad P \succ 0$$

относительно матричной переменной $P \in \mathbb{S}^n$, где $L(P)$ — некоторая линейная функция, эквивалентна задаче SDP

$$\min \operatorname{tr} Z \quad \text{при ограничениях } L(P) \leq 0, \quad \begin{pmatrix} Z & I \\ I & P \end{pmatrix} \succ 0, \quad P \succ 0,$$

где оптимизация проводится по матричным переменным $P, Z \in \mathbb{S}^n$.

2.7.4. Лемма Питерсена

Пусть $G \in \mathbb{S}^n$ — некоторая матрица; рассмотрим ее возмущение вида

$$G + M\Delta N + N^T\Delta^T M^T, \quad (2.7.10)$$

где $\Delta \in \mathbb{R}^{p \times q}$ — возмущающая матрица, а $M \in \mathbb{R}^{n \times p}$ и $N \in \mathbb{R}^{q \times n}$ — постоянные «обрамляющие» матрицы соответствующих размерностей, задающие структуру неопределенности. В этой схеме симметричное возмущение задается с помощью матрицы Δ , которая не обязана быть симметричной и даже квадратной.

Мы хотим ответить на вопрос о том, при каких условиях возмущенная матрица (2.7.10) является знакопредeterminedной при всех ограниченных по норме возмущениях Δ :

$$G + M\Delta N + N^T\Delta^T M^T \preceq 0 \quad \text{при всех } \Delta: \|\Delta\| \leq 1.$$

Мы будем называть эти условия *леммой Питерсена*; приведем этот результат в следующей формулировке.

Лемма 2.7.5 (Питерсен [122]). *Пусть $G \in \mathbb{S}^n$, а $M \in \mathbb{R}^{n \times p}$ и $N \in \mathbb{R}^{q \times n}$ — ненулевые матрицы. Неравенство*

$$G + M\Delta N + N^T\Delta^T M^T \preceq 0$$

справедливо для всех $\Delta \in \mathbb{R}^{p \times q}$: $\|\Delta\| \leq 1$ тогда и только тогда, когда существует число $\varepsilon > 0$ такое, что

$$\begin{pmatrix} G + \varepsilon M M^T & N^T \\ N & -\varepsilon I \end{pmatrix} \preceq 0.$$

Обратим внимание на то, что матричная неопределенность Δ не предполагается фиксированной; единственное требование — ее ограниченность по норме. Таким образом, полученный результат и его приводимые ниже модификации справедливы и для нестационарной неопределенности $\Delta(t)$: $\|\Delta(t)\| \leq 1$.

Следующий вариант леммы Питерсена — для случая *строгого* матричного неравенства.

Лемма 2.7.6. *Пусть $G \in \mathbb{S}^n$, $M \in \mathbb{R}^{n \times p}$, $N \in \mathbb{R}^{q \times n}$. Неравенство*

$$G + M\Delta N + N^T\Delta^T M^T \prec 0$$

справедливо для всех $\Delta \in \mathbb{R}^{p \times q}$: $\|\Delta\| \leq 1$ тогда и только тогда, когда существует число $\varepsilon > 0$ такое, что

$$\begin{pmatrix} G + \varepsilon M M^T & N^T \\ N & -\varepsilon I \end{pmatrix} \prec 0.$$

Итак, лемма Питерсена сводит проверку знакопределенности семейства (2.7.10) к задаче разрешимости линейного матричного неравенства относительно одной скалярной переменной ε .

Приведем теперь некоторые полезные технические модификации и следствия из этой леммы.

Следствие 2.7.3. *Пусть $\gamma > 0$, $G \in \mathbb{S}^n$, а $M \in \mathbb{R}^{n \times p}$ и $N \in \mathbb{R}^{q \times n}$ — ненулевые матрицы. Следующие утверждения эквивалентны:*

1) неравенство

$$G + M\Delta N + N^T \Delta^T M^T \preceq 0$$

справедливо для всех $\Delta \in \mathbb{R}^{p \times q}$: $\|\Delta\| \leq \gamma$;

2) существует число ε такое, что

$$\begin{pmatrix} G + \varepsilon MM^T & \gamma N^T \\ \gamma N & -\varepsilon I \end{pmatrix} \preceq 0;$$

3) существует число ε такое, что

$$\begin{pmatrix} \frac{1}{\gamma} G + \varepsilon MM^T & N^T \\ N & -\varepsilon I \end{pmatrix} \preceq 0;$$

4) существует число ε такое, что

$$\begin{pmatrix} G + \gamma^2 \varepsilon MM^T & N^T \\ N & -\varepsilon I \end{pmatrix} \preceq 0.$$

Оказывается, что лемма Питерсена полностью сохраняет свою силу и в том случае, когда возмущение Δ ограничено во фробениусовой норме.

Следующий полезный результат представляет собой обобщение достаточной части леммы Питерсена.

Лемма 2.7.7. *Пусть $G \in \mathbb{S}^n$, $M \in \mathbb{R}^{n \times p}$, $N \in \mathbb{R}^{q \times n}$. Тогда*

$$M\Delta N + N^T \Delta^T M^T \preceq \varepsilon MM^T + \frac{1}{\varepsilon} N^T N$$

для любого $\varepsilon > 0$ и всех

$$\Delta \in \mathbb{R}^{p \times q}: \|\Delta\| \leq 1 \text{ или } \|\Delta\|_F \leq 1.$$

2.8. Примеры экстремальных задач в условной минимизации

В заключение главы 2 рассмотрим несколько примеров задач условной минимизации. Все задачи раздела решены, также сформулированы некоторые упражнения для самостоятельного решения. Наиболее сложные задачи отмечены символом «*».

Пример 2.8.1. Рассмотрим способ нахождения стационарных точек в задаче минимизации, см. раздел [Д.2](#):

$$\min f_0(x) \quad \text{при условии} \quad f_i(x) = 0, \quad 1 \leq i \leq m < n, \quad x \in \mathbb{R}^n.$$

Мы будем предполагать, что функции $f_i \in C^2$, причем вторая производная f''_i непрерывна по Липшицу при всех $0 \leq i \leq m$. Также будем считать, что выполняется условие полного ранга

$$\operatorname{rank}(f'_1(x), \dots, f'_m(x)) = m$$

для всех x . Определим функцию Лагранжа

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

вектор множителей Лагранжа $\lambda = (\lambda_1, \dots, \lambda_m)^T \in \mathbb{R}^m$, а также функцию $f(x) = (f_1(x), \dots, f_m(x))^T$.

Точка x , удовлетворяющая ограничениям, стационарная, если $f'_0(x)$ есть линейная комбинация векторов $\{f'_i(x)\}_{i=1}^m$. Напомним, что необходимое условие экстремума в точке x в приведенной выше задаче есть стационарность точки x .

Рассмотрим вектор-функцию

$$F(x, \lambda) = (L'_x(x, \lambda), f_1(x), \dots, f_m(x))^T \in \mathbb{R}^{n+m},$$

$$\text{где } L'_x(x, \lambda) = f'_0(x) + \sum_{i=1}^m \lambda_i f'_i(x).$$

Для нахождения стационарной точки далее можно пытаться решать систему $F(x, \lambda) = 0$ методом Ньютона, см. теорему [1.4.2](#). Заметим, что производная F по (x, λ) есть

$$F'(x, \lambda) = \begin{pmatrix} f''_0(x) + \sum_{i=1}^m \lambda_i f''_i(x) & f'^T(x) \\ f'(x) & 0 \end{pmatrix},$$

и итерационный процесс имеет вид

$$\begin{pmatrix} x_{k+1} \\ \lambda_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ \lambda_k \end{pmatrix} - (F'(x_k, \lambda_k))^{-1} F(x_k, \lambda_k).$$

Если начальную точку $(x_1, \lambda_1)^T$ итераций задать не удается (то есть процесс с начальной точкой $(x_1, \lambda_1)^T$ не сходится), то можно начать с градиентного спуска для функции $\|F(x, \lambda)\|$.

Упражнение 2.8.1. Докажите, что для фиксированного x

$$\begin{aligned} \lambda(x) &= \arg \min_{\lambda \in \mathbb{R}^m} \|F(x, \lambda)\| = \\ &= \arg \min_{\lambda \in \mathbb{R}^m} \|f'_0(x) + f'^T(x)\lambda\| = \\ &= -(f'(x)f'^T(x))^{-1} f'(x) f'_0(x). \end{aligned}$$

Поэтому при градиентном спуске можно минимизировать функцию $\|F(x, \lambda(x))\|$. Когда точка $(x, \lambda(x))$ окажется в достаточно малой окрестности нуля функции F , нужно переключиться на метод Ньютона. ▼

Пример 2.8.2 (Максимальный вписанный шар). Будем искать евклидов шар максимального радиуса, вписанный в ограниченный многогранник, заданный системой неравенств

$$\mathcal{Q} = \{x \in \mathbb{R}^n : Ax \leq b\}, \quad x \in \mathbb{R}^n, \quad b \in \mathbb{R}^m,$$

где $m > n$.

Отметим, что в общем случае такой шар не единственный.

Упражнение 2.8.2. Как численно проверить, что многогранник \mathcal{Q} является ограниченным?

Будем считать, что i -я строка матрицы A есть вектор длины 1: $\|A_i\| = 1$. Это условие всегда можно получить, делая i -е неравенство $(A_i, x) \leq b_i$ системы неравенств на $\|A_i\|$, $1 \leq i \leq m$.

В силу компактности \mathcal{Q} решение задачи существует. Пусть $B_r(x_0)$ — искомый шар. Тогда расстояние от x_0 до ограничения $(A_i, x) = b_i$, $1 \leq i \leq m$, равно

$$h_i = \frac{1}{\|A_i\|}|(A_i, x_0) - b_i| = b_i - (A_i, x_0).$$

Упражнение 2.8.3. Докажите указанную выше формулу для h_i .

Теперь задача решается просто. Легко видеть, что r есть минимальное расстояние по всем i , $r = \min_{1 \leq i \leq m} h_i$. Поэтому задачу можно переформулировать как следующую задачу линейного программирования:

$$\max r \quad \text{при условии } (A_i, x) + r \leq b_i, \quad 1 \leq i \leq m. \quad (2.8.1)$$

Упражнение 2.8.4. Докажите, что любое решение задачи (2.8.1) дает радиус r и центр максимального вписанного шара x .

Задачу (2.8.1) можно считать задачей с расширенной фазовой переменной. Действительно, пусть $z = (x, r)^T \in \mathbb{R}^{n+1}$ ($x \in \mathbb{R}^n$, $r \in \mathbb{R}$), $c = (0, \dots, 0, 1)^T \in \mathbb{R}^{n+1}$, $a_i = (A_i, 1)^T \in \mathbb{R}^{n+1}$. Тогда задача поиска максимального вписанного шара есть

$$\max(c, z) \quad \text{при условии } (a_i, z) \leq b_i, \quad 1 \leq i \leq m.$$

Упражнение 2.8.5. Пусть $\mathcal{Q} \subset \mathbb{R}^n$ – строго выпуклое компактное множество, то есть его граница не содержит отрезков. Докажите, что максимальный вписанный во множество \mathcal{Q} шар единственен.

Иногда максимальный вписанный в компактное выпуклое подмножество $\mathcal{Q} \subset \mathbb{R}^n$ шар называется чебышевским. Мы придерживаемся другой терминологии и рассмотрим чебышевский шар в конце раздела, в примере 2.8.4. ▼

Пример 2.8.3 (Квазивыпуклые функции в задаче условного экстремума). Определение квазивыпуклой функции дано в разделе 1.5.2. Рассмотрим задачу минимизации квазивыпуклой функции f на выпуклом замкнутом множестве $\mathcal{Q} \subset \mathbb{R}^n$. Приведем условия метода проекции градиента.

Теорема 2.8.1. Для выпуклого замкнутого множества $\mathcal{Q} \subset \mathbb{R}^n$ и квазивыпуклой функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$ с непрерывным по Липшичу градиентом с константой Липшичу L_1 рассмотрим алгоритм метода проекции градиента:

$$x_1 \in \mathcal{Q}, \quad x_{k+1} = P_{\mathcal{Q}}(x_k - \alpha f'(x_k)), \quad \alpha > 0.$$

Пусть $\min_{x \in \mathcal{L}_f(f(x_1)) \cap \mathcal{Q}} \|f'(x)\| \geq l > 0$, лебеговы множества $\mathcal{L}_f(t)$ для всех $\min_{x \in \mathcal{Q}} f(x) \leq t \leq f(x_1)$ сильно выпуклы с радиусом $R > 0$. Пусть в

решении $x_* \in \partial\mathcal{Q}$ для некоторого $K > R$ выполнено включение

$$\text{int } \mathcal{B}_K \left(x_* + \frac{f'(x_*)}{\|f'(x_*)\|} K \right) \subset \mathcal{Q}.$$

Предположим, что $\frac{l}{R} > \frac{\|f'(x_*)\|}{K}$.

Тогда при выборе достаточно малого шага метода $\alpha < \frac{1}{L_1}$ описанный градиентный метод сходится к x_* с линейной скоростью.

Доказательство. Необходимое условие экстремума

$$x_* = P_{\mathcal{Q}}(x_* - \alpha f'(x_*))$$

выполнено для любого $\alpha > 0$. С учетом этого и в силу условия Липшица для градиента f' с константой L_1 имеем

$$\begin{aligned} \|x_{k+1} - x_*\|^2 &= \|P_{\mathcal{Q}}(x_k - \alpha f'(x_k)) - P_{\mathcal{Q}}(x_* - \alpha f'(x_*))\|^2 \leqslant \\ &\leqslant \|x_k - \alpha f'(x_k) - (x_* - \alpha f'(x_*))\|^2 \leqslant \\ &\leqslant \|x_k - x_*\|^2 - 2\alpha(f'(x_k) - f'(x_*), x_k - x_*) + \alpha^2 L_1^2 \|x_k - x_*\|^2. \end{aligned}$$

Заметим, что из условия $\alpha < 1/L_1$ в силу леммы 2.3.1 для всех k выполняется неравенство $f(x_{k+1}) \leqslant f(x_k) \leqslant f(x_1)$.

Отметим, что в силу условия теоремы $\|f'(x_k)\| \geqslant l > 0$. Из сильной выпуклости множества $\mathcal{L}_f(f(x_k))$ аналогично теореме 1.5.1 получаем

$$\left\| x_k - R \frac{f'(x_k)}{\|f'(x_k)\|} - x_* \right\| \leqslant R$$

и

$$\frac{\|f'(x_k)\|}{2R} \|x_k - x_*\|^2 \leqslant (f'(x_k), x_k - x_*).$$

Если $x_k \in \text{int } \mathcal{Q}$, то минимизация на k -м шаге аналогична шагу градиентного спуска без ограничений. При этом из формулы (1.3.2) получаем, что

$$f(x_{k-1}) - f(x_k) \geqslant \left(\alpha - \frac{\alpha^2}{2} L_1 \right) \|f'(x_{k-1})\|^2 \geqslant \left(\alpha - \frac{\alpha^2}{2} L_1 \right) l^2 > 0,$$

то есть значение функции уменьшилось на фиксированную положительную величину. Если условие $x_k \in \text{int } \mathcal{Q}$ выполняется на каждом шаге, то минимум будет достигнут не более чем за

$$\left[\frac{f(x_1) - f_0}{\left(\alpha - \frac{\alpha^2}{2} L_1 \right) l^2} \right] + 1$$

шагов.

Рассмотрим случай $x_k \in \partial\mathcal{Q}$. При этом выполняется включение $x_k \notin \text{int } \mathcal{B}_K \left(x_* + \frac{f'(x_*)}{\|f'(x_*)\|} K \right)$. Это эквивалентно неравенству

$$\left\| x_* + \frac{f'(x_*)}{\|f'(x_*)\|} K - x_k \right\| \geq K,$$

что после возвведения в квадрат и выполнения преобразований дает

$$\frac{\|f'(x_*)\|}{2K} \|x_k - x_*\|^2 \geq (f'(x_*), x_k - x_*).$$

Подставляя две последние оценки в оценку для $\|x_{k+1} - x_*\|^2$, получаем

$$\begin{aligned} \|x_{k+1} - x_*\|^2 &\leq \\ &\leq \|x_k - x_*\|^2 \left(1 - \alpha \left(\frac{\|f'(x_k)\|}{R} - \frac{\|f'(x_*)\|}{K} \right) + \alpha^2 L_1^2 \right) \leq \\ &\leq \|x_k - x_*\|^2 \left(1 - \alpha \left(\frac{l}{R} - \frac{\|f'(x_*)\|}{K} \right) + \alpha^2 L_1^2 \right). \end{aligned}$$

При достаточно малом $\alpha > 0$ выражение в круглых скобках строго меньше 1. ■

Упражнение 2.8.6. Докажите, что решение x_* в теореме 2.8.1 единствено.



Пример 2.8.4 (Чебышевский центр множества).* Исследуем свойства евклидова шара минимального радиуса, который содержит заданное выпуклое компактное подмножество $\mathcal{Q} \subset \mathbb{R}^n$. Такой шар $\mathcal{B}_R(x_0)$ называется чебышевским, его центр $x_0 = x_0(\mathcal{Q})$ называется чебышевским центром множества \mathcal{Q} , а радиус $R = R(\mathcal{Q})$ — чебышевским радиусом множества \mathcal{Q} . Чебышевский центр является важной характеристикой множества в разных задачах.

Упражнение 2.8.7. Докажите, что для любого выпуклого компактного подмножества $\mathcal{Q} \subset \mathbb{R}^n$ чебышевский шар существует.

Покажем, что чебышевский шар единственный. Допустим, существует два разных шара $\mathcal{B}_R(x_i)$, $i = 0, 1$, каждый из которых чебышевский. Тогда $\mathcal{Q} \subset \mathcal{B}_R(x_0) \cap \mathcal{B}_R(x_1)$. Из теоремы Пифагора следует, что пересечение шаров содержится в шаре с центром в точке $\frac{1}{2}(x_0 + x_1)$

и радиуса $\sqrt{R^2 - \frac{1}{4}\|x_0 - x_1\|^2} < R$, что противоречит минимальности радиуса R . Таким образом, для евклидова шара чебышевский центр единственный.

Чебышевский центр интересным делает то, что он является решением сильно выпуклых экстремальных задач.

Лемма 2.8.1. *Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое компактное множество $\mathcal{Q} \subset \mathbb{R}^n$. Определим функцию $g_{\mathcal{Q}}(a) = \max_{x \in \mathcal{Q}} \|x - a\|^2$. Тогда чебышевский центр множества \mathcal{Q} есть решение задачи*

$$\min_{a \in \mathcal{Q}} g_{\mathcal{Q}}(a), \quad (2.8.2)$$

а также задачи

$$\min_{a \in \mathbb{R}^n} g_{\mathcal{Q}}(a). \quad (2.8.3)$$

Доказательство. Сначала рассмотрим задачу (2.8.2). Минимум сильно выпуклой функции с константой сильной выпуклости $\varkappa = 2$ достигается на выпуклом замкнутом множестве \mathcal{Q} в единственной точке $a_0 \in \mathcal{Q}$. Определим $R = \sqrt{g_{\mathcal{Q}}(a_0)}$. Тогда для любой точки $x \in \mathcal{Q}$ выполнено неравенство $\|x - a_0\| \leq R$, то есть $\mathcal{Q} \subset \mathcal{B}_R(a_0)$. При этом для любого $r < R$ найдется точка $x_r \in \mathcal{Q}$ такая, что $\|x_r - a_0\| > r$.

Покажем, что чебышевский центр u для \mathcal{Q} , то есть решение задачи (2.8.3), есть решение задачи (2.8.2). Отсюда получим, что $u = a_0$ и $\mathcal{B}_R(a_0)$ — чебышевский шар для \mathcal{Q} .

В силу сильной выпуклости $g_{\mathcal{Q}}$ решение задачи (2.8.3) существует и единственno. Остается показать, что $u \in \mathcal{Q}$.

Допустим, что $u \notin \mathcal{Q}$. Тогда определим $v = P_{\mathcal{Q}}u$ и $p = (u - v)/\|u - v\|$ — внешняя нормаль ко множеству \mathcal{Q} в точке $v \in \mathcal{Q}$. Зафиксируем произвольную точку $x \in \mathcal{Q}$. Тогда $\mathcal{Q} \subset \{z \in \mathbb{R}^n : (p, z - v) \leq 0\}$, см. теорему Г.3. Из последнего включения $(u - v, x - v) \leq 0$, откуда

$$\begin{aligned} \|x - u\|^2 &= \|(x - v) - (u - v)\|^2 = \|x - v\|^2 - 2(u - v, x - v) + \|u - v\|^2 \geq \\ &\geq \|x - v\|^2 - (u - v, x - v) + \|u - v\|^2 = \|x - v - \frac{u - v}{2}\|^2 + \frac{3}{4}\|u - v\|^2. \end{aligned}$$

Таким образом,

$$g_{\mathcal{Q}}(u) \geq g_{\mathcal{Q}}\left(\frac{u + v}{2}\right) + \frac{3}{4}\|u - v\|^2 > g_{\mathcal{Q}}\left(\frac{u + v}{2}\right).$$

Последнее неравенство противоречит тому, что $u \in \mathbb{R}^n$ есть решение задачи (2.8.3). Значит, $u \in \mathcal{Q}$ и u — чебышевский центр \mathcal{Q} . ■

Чебышевский центр удовлетворяет условию Гельдера как функция множества в метрике Хаусдорфа h , см. раздел Г.8, с показателем $\frac{1}{2}$. Более точно, если $\mathcal{Q}_1, \mathcal{Q}_2$ — выпуклые компактные подмножества и для некоторого $M > 0$ выполнено включение

$$\mathcal{Q}_1 \cup \mathcal{Q}_2 \subset \mathcal{B}_M(0),$$

то для чебышевских центров x_1 и x_2 этих множеств выполнена оценка

$$\|x_1 - x_2\| \leq 2\sqrt{3Mh(\mathcal{Q}_1, \mathcal{Q}_2)} + h(\mathcal{Q}_1, \mathcal{Q}_2).$$

Доказательство можно найти в [19, теорема 2.1.1].

Покажем, что порядок $\sqrt{h(\mathcal{Q}_1, \mathcal{Q}_2)}$ неулучшаем. Зададим $R > 0$ и $\varepsilon \in (0, R)$. Рассмотрим множества на плоскости \mathbb{R}^2 :

$$\mathcal{Q}_1 = \{x \in \mathbb{R}^2 : 0 \leq x_2 \leq \varepsilon, x_1^2 + x_2^2 \leq R^2\},$$

а \mathcal{Q}_2 симметрично \mathcal{Q}_1 относительно прямой $x_2 = \varepsilon/2$. Легко видеть, что чебышевские центры этих множеств есть $x_1 = (0; 0)$, $x_2 = (0; \varepsilon)$ и $\|x_1 - x_2\| = \varepsilon$. При этом хаусдорфово расстояние есть

$$h(\mathcal{Q}_1, \mathcal{Q}_2) = \sqrt{R^2 + \varepsilon^2} - R \leq \frac{\varepsilon^2}{2R}.$$

Оценки показывают точность порядка $\sqrt{h(\mathcal{Q}_1, \mathcal{Q}_2)}$.

Пусть точка $a \in \mathbb{R}^n$ является чебышевским центром выпуклого компактного множества $\mathcal{Q} \subset \mathbb{R}^n$. Положим $f(z) = \max_{x \in \mathcal{Q}} \|x - z\|$. Поскольку $z = a$ является глобальным минимумом f , то $0 \in \partial f(a)$. Обозначим чебышевский радиус \mathcal{Q} через $R = f(a)$. В силу теоремы Г.17 об очистке

$$0 \in \partial f(a) = \text{co} \bigcup_{x \in \mathcal{Q}, \|x-a\|=R} \frac{a-x}{\|a-x\|},$$

причем из последней выпуклой оболочки можно взять $k \leq n+1$ точек $x \in \mathcal{Q}$, $\|x-a\| = R$, выпуклая оболочка которых содержит 0. Это значит, что найдутся точки $\{x_i\}_{i=1}^k \subset \mathcal{Q}$, $\|x_i - a\| = R$ и числа $\lambda_i > 0$, $\sum_{i=1}^k \lambda_i = 1$, такие, что

$$0 = \sum_{i=1}^k \lambda_i \frac{a - x_i}{\|a - x_i\|} = \sum_{i=1}^k \lambda_i \frac{a - x_i}{R}.$$

Последнее означает $a = \sum_{i=1}^k \lambda_i x_i$. Таким образом, всегда найдется симплекс (может быть, не полноразмерный) с вершинами $\{x_i\}_{i=1}^k \subset \mathcal{Q}$ такой, что $\|x_i - a\| = R$. При этом этот симплекс содержит точку a .

Покажем, что R — чебышевский радиус для указанного симплекса, а a — его чебышевский центр. Далее будем считать, что $a = 0$. Допустим, существует вектор $p \neq 0$ такой, что сдвиг вдоль вектора p не выводит симплекс из $\mathcal{B}_R(0)$. Это значит, что найдется $\alpha > 0$ такое, что $x_i + \alpha p \in \mathcal{B}_R(0)$ для всех i .

Из равенства

$$0 = \sum_{i=1}^k \lambda_i(x_i, p)$$

следует, что для некоторого i выполнено условие $(p, x_i) \geq 0$. Это означает, что при всяком $\alpha > 0$ выполнено включение $x_i + \alpha p \notin \mathcal{B}_R(0)$, которое вытекает из соотношения $x_i + \alpha p \in \{z \in \mathbb{R}^n : (x_i, z) \geq R\}$. Получаем противоречие.

Пусть d — диаметр симплекса, а чебышевский центр $a = 0$. Для каждого i от 1 до k имеем

$$(1 - \lambda_i)d^2 \geq \sum_{j \neq i} \lambda_j \|x_i - x_j\|^2 = \sum_{j=1}^k \lambda_j \|x_i - x_j\|^2.$$

В силу равенства

$$\|x_i - x_j\|^2 = \|x_i\|^2 + \|x_j\|^2 - 2(x_i, x_j) = 2R^2 - 2(x_i, x_j)$$

получаем

$$\begin{aligned} (1 - \lambda_i)d^2 &\geq \left(\sum_{i \neq j} \lambda_j \right) \cdot 2R^2 - 2 \left(x_i, \sum_{j=1}^k \lambda_j x_j \right) = \\ &= 2R^2 - 2(x_i, 0) = 2R^2. \end{aligned}$$

Суммируя полученные неравенства по $1 \leq i \leq k$, имеем

$$\sum_{i=1}^k (1 - \lambda_i)d^2 = (k - 1)d^2 \geq 2kR^2,$$

откуда, с учетом $k \leq n + 1$,

$$R \leq \sqrt{\frac{k-1}{2k}} d \leq \sqrt{\frac{n}{2(n+1)}} d. \quad (2.8.4)$$

Последнее неравенство называется *неравенством Юнга*. Совместно с предыдущими рассуждениями оно показывает, что для всякого выпуклого компактного подмножества из \mathbb{R}^n диаметра $d > 0$ чебышевский радиус R оценивается по формуле (2.8.4), см. рис. 2.8.1.

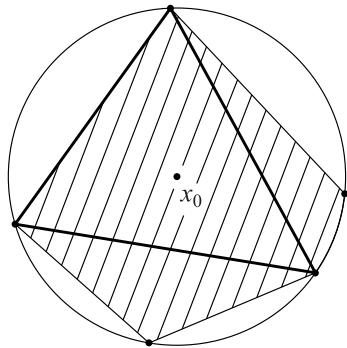


Рис. 2.8.1. Чебышевский центр x_0 . Жирными линиями показан симплекс, содержащий точку x_0 .

Упражнение 2.8.8. Покажите, что равенство в (2.8.4) достигается только в случае правильного симплекса с вершинами на сфере радиуса R .

▼

Глава 3

Оптимизация в теории автоматического управления

В этой главе будет рассмотрен ряд важнейших задач теории автоматического управления, связанных с синтезом стабилизирующих и робастно стабилизирующих регуляторов, линейно-квадратичным управлением по состоянию и по выходу, подавлением неслучайных внешних возмущений, настройкой ПИД-регуляторов и др. Все они будут сведены к конечномерным задачам оптимизации — это позволит использовать для их решения идеи и методы оптимизации, рассмотренные в предыдущих двух главах.

3.1. Линейные матричные неравенства в задачах управления

3.1.1. Квадратичная устойчивость

Обсудим важнейшее понятие устойчивости линейных динамических систем. Для нелинейных нестационарных систем имеется множество определений устойчивости: устойчивость точки равновесия и устойчивость движения; устойчивость по Ляпунову и асимптотическая устойчивость; устойчивость «в малом» и «в большом»; устойчивость по начальным условиям и по возмущению [23].

Для линейных стационарных систем картина намного проще и будем говорить просто об *устойчивости*, понимая под этим стремление к нулю траекторий системы без внешних возмущений, то есть *асимптотической устойчивости*.

тотическую устойчивость.

Система может быть неустойчивой; в этом случае можно пытаться *стабилизировать* систему: добиться устойчивости замкнутой системы путем введения обратной связи. В этом разделе ограничимся рассмотрением линейной статической обратной связи. За подробностями и полными доказательствами приводимых далее утверждений мы отсылаем читателя к монографии [22].

Начнем с определений.

Определение 3.1.1. Линейная непрерывная стационарная система

$$\dot{x} = Ax, \quad (3.1.1)$$

где t — время, $x = x(t) \in \mathbb{R}^n$, $\dot{x} = \frac{dx}{dt}$, $A \in \mathbb{R}^{n \times n}$, называется (*асимптотически*) *устойчивой*, если для любого ее начального состояния $x(0) = x_0$ выполнено условие

$$x(t) \xrightarrow[t \rightarrow \infty]{} 0.$$

Важнейшим средством исследования устойчивости является *второй (прямой) метод Ляпунова*, или *метод функций Ляпунова*.

Определение 3.1.2. Скалярная функция

$$V: \mathbb{R}^n \rightarrow \mathbb{R},$$

непрерывная вместе со своими частными производными, называется *функцией Ляпунова* для системы (3.1.1), если она обладает следующими свойствами:

- 1) $V(x) \geq 0$ для всех x , причем $V(x) = 0$ лишь для $x = 0$;
- 2) $\dot{V}(x(t)) \doteq \frac{d}{dt} V(x(t)) < 0$, где $x(t)$ — произвольное ненулевое решение системы (3.1.1).

Отметим, что второе свойство заключается в убывании функции Ляпунова на решениях системы; иными словами, ее производная, вычисленная в силу системы, отрицательна. Метод функций Ляпунова состоит в построении функции с указанными свойствами, существование которой позволяет сделать вывод об устойчивости системы.

Введем в рассмотрение функцию

$$V(x) = x^T Q x, \quad Q \succ 0, \quad (3.1.2)$$

и найдем условия, при которых она будет функцией Ляпунова для системы (3.1.1).

В силу положительной определенности матрицы Q имеем $V(x) > 0$ для всех $x \neq 0$, так что первое свойство из определения 3.1.2 выполнено. Вычислим производную функции $V(x)$ на траекториях системы (3.1.1):

$$\dot{V}(x) = \dot{x}^T Q x + x^T Q \dot{x} = x^T A^T Q x + x^T Q A x = x^T (A^T Q + Q A) x.$$

Следовательно, второе свойство из определения 3.1.2 эквивалентно условию

$$x^T (A^T Q + Q A) x < 0 \quad \text{при всех } x \neq 0,$$

то есть выполнению матричного неравенства *Ляпунова*

$$A^T Q + Q A \prec 0 \tag{3.1.3}$$

(это неравенство уже встречалось в разделе 2.7.1).

Упражнение 3.1.1. Пользуясь леммой Шура (раздел 2.7.3) покажите, что неравенство Ляпунова (3.1.3) эквивалентно ЛМН

$$AP + PA^T \prec 0, \quad P = Q^{-1},$$

также называемому неравенством Ляпунова.

Определение 3.1.3. Функцию (3.1.2), удовлетворяющую условию (3.1.3), будем называть *квадратичной функцией Ляпунова* для системы (3.1.1).

Сделаем важное замечание.

Замечание 3.1.1. Пусть $V(x) = x^T Q x$ — квадратичная функция Ляпунова для системы (3.1.1). Рассмотрим эллипсоид

$$\mathcal{E} = \{x \in \mathbb{R}^n : V(x) \leq 1\} = \{x \in \mathbb{R}^n : x^T P^{-1} x \leq 1\} \tag{3.1.4}$$

с центром в начале координат и положительно определенной матрицей $P = Q^{-1}$. Как непосредственно следует из определения 3.1.2, если начальное условие x_0 системы (3.1.1) лежит в эллипсоиде \mathcal{E} , то ее траектория будет оставаться в нем для всех моментов времени. Это свойство будем называть *инвариантностью* эллипсоида (3.1.4) относительно начальных условий (а сам эллипсоид будем называть *инвариантным*).

В последующих разделах понятие инвариантного эллипсоида для линейных систем (в том числе при наличии внешних возмущений) будет неоднократно встречаться. Конструкции, обобщающие понятие инвариантного эллипсоида и связанные с концепцией *выживающей* решений, обсуждаются в приложении Г.9. ▼

В следующей теореме устанавливается критерий устойчивости для линейной системы.

Теорема 3.1.1 ([22]). *Система (3.1.1) устойчива тогда и только тогда, когда ее матрица A гурвицева.¹*

Замечание 3.1.2. В лемме Б.4 установлено, что для гурвицевой матрицы A уравнение Ляпунова

$$AP + PA^T = -R, \quad R \succ 0,$$

имеет положительно определенное решение P ; иными словами, устойчивая система $\dot{x} = Ax$ с необходимостью обладает квадратичной функцией Ляпунова $V(x) = x^T P^{-1} x$. С учетом теоремы 3.1.1 заключаем, что непрерывная линейная система устойчива тогда и только тогда, когда у нее имеется квадратичная функция Ляпунова. ▼

Теперь можно оценить поведение решений линейной системы при наличии ограниченных внешних возмущений.

Теорема 3.1.2. *Для того чтобы любое решение $x(t)$ системы*

$$\dot{x} = Ax + w \tag{3.1.5}$$

было ограниченным для всех ограниченных внешних возмущений

$$\|w(t)\| \leq \gamma \quad \forall t \geq 0,$$

где $\gamma > 0$ — заданное число, необходимо и достаточно, чтобы матрица A была гурвицевой.

Гурвицевость матрицы A необходима и достаточна для того, чтобы решение невозмущенной системы стремилось к нулю при любом начальном приближении и чтобы решение возмущенной системы оставалось ограниченным для ограниченных возмущений. Первое свойство иногда называется *устойчивостью по начальному приближению*, второе — *устойчивостью по входу*.

В дальнейшем не будем делать различий между устойчивостью системы и устойчивостью ее матрицы A ; соответственно, величину $\sigma = \sigma(A) > 0$ естественно также называть *степенью устойчивости системы (3.1.1)* или *системы (3.1.5)*. Еще раз подчеркнем эквивалентность устойчивости линейной стационарной системы и существования у нее квадратичной функции Ляпунова.

¹Определение гурвицовой матрицы приведено в приложении Б.3.

Замечание 3.1.3. Обратим внимание на следующее обстоятельство: существование положительно определенного решения $P \succ 0$ неравенства Ляпунова

$$AP + PA^T + 2\sigma P \preccurlyeq 0$$

при некотором заданном $\sigma > 0$ эквивалентно тому, что степень устойчивости (см. приложение [Б.3](#)) матрицы A не меньше чем σ . \blacktriangledown

Упражнение 3.1.2. Дайте обоснование предыдущему утверждению, используя результаты, приведенные в приложении [Б.3](#).

Обратимся к дискретному случаю. В простейшем случае уравнение состояния невозмущенной линейной системы в дискретном времени имеет вид

$$x_{k+1} = Ax_k, \quad k = 0, 1, 2, \dots, \quad (3.1.6)$$

где $x_k \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$. Будем называть систему [\(3.1.6\)](#) (*дискретно*) *устойчивой*, если

$$x_k \xrightarrow[k \rightarrow \infty]{} 0$$

для любого начального x_0 .

Как и в непрерывном случае, функции Ляпунова являются эффективным средством исследования устойчивости.

Определение 3.1.4. Непрерывная скалярная функция

$$V: \mathbb{R}^n \rightarrow \mathbb{R}$$

называется *функцией Ляпунова* для системы [\(3.1.6\)](#), если она обладает следующими свойствами:

- 1) $V(x) \geq 0$ для всех x , причем $V(x) = 0$ лишь для $x = 0$;
- 2) $V(x_{k+1}) < V(x_k)$, $k = 1, 2, \dots$, где $x_k \neq 0$ — произвольное решение системы [\(3.1.6\)](#).

Для линейной дискретной системы квадратичная функция Ляпунова вводится аналогично непрерывному случаю. Рассмотрим функцию

$$V(x) = x^T Q x, \quad Q \succ 0; \quad (3.1.7)$$

для нее первое свойство выполнено очевидным образом. Чтобы удовлетворить второму свойству, вычислим значения $V(x)$ на траекториях системы [\(3.1.6\)](#):

$$V(x_{k+1}) = x_{k+1}^T Q x_{k+1} = x_k^T A^T Q A x_k.$$

Свойство 2) эквивалентно выполнению

$$x^T (A^T Q A - Q) x < 0 \quad \text{при всех } x \neq 0,$$

то есть выполнению *дискретного неравенства Ляпунова*

$$A^T Q A - Q \prec 0. \quad (3.1.8)$$

Упражнение 3.1.3. Пользуясь леммой Шура (раздел 2.7.3) покажите, что дискретное неравенство Ляпунова (3.1.8) эквивалентно ЛМН

$$A P A^T - P \prec 0, \quad P = Q^{-1},$$

также называемому *дискретным неравенством Ляпунова*.

Функция вида (3.1.7)–(3.1.8) называется *квадратичной функцией Ляпунова* для системы (3.1.6).

Теорема 3.1.3 ([22]). Для устойчивости линейной дискретной системы (3.1.6) необходимо и достаточно, чтобы матрица A была шуровской.

Напомним, что определения устойчивых по Шуру и по Гурвицу матриц даны в приложении Б.3. Разумеется, замечание 3.1.2 остается справедливым и в дискретном случае; при этом (дискретное) уравнение Ляпунова будет иметь вид

$$A P A^T - P = -R, \quad R \succ 0.$$

Теперь оценим поведение решений дискретной системы при наличии ограниченных внешних возмущений. Следующая теорема является дискретным аналогом теоремы 3.1.2.

Теорема 3.1.4. Для того чтобы произвольное решение x_k системы

$$x_{k+1} = Ax_k + w_k$$

было ограниченным для всех ограниченных внешних возмущений

$$\|w_k\| \leq \gamma \quad \forall k = 0, 1, \dots,$$

где $\gamma > 0$ – заданное число, необходимо и достаточно, чтобы матрица A была шуровской.

3.1.2. Квадратичная стабилизация

Теперь перейдем к так называемой *квадратичной стабилизации*; она основывается на том, что устойчивость линейной системы эквивалентна наличию у нее квадратичной функции Ляпунова вида

$$V(x) = x^T Q x, \quad Q \succ 0. \quad (3.1.9)$$

Рассмотрим линейную систему управления

$$\dot{x} = Ax + Bu, \quad x(0) = x_0, \quad (3.1.10)$$

с состоянием $x(t) \in \mathbb{R}^n$ и управлением $u(t) \in \mathbb{R}^p$, где $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$; пара (A, B) предполагается управляемой (см. приложение B.7).

Нас интересует вопрос ее стабилизации при помощи статической линейной обратной связи по состоянию

$$u = Kx, \quad K \in \mathbb{R}^{p \times n}. \quad (3.1.11)$$

Имеет место следующий результат.

Теорема 3.1.5. Пусть матрицы \widehat{P} и \widehat{Y} удовлетворяют линейным матричным неравенствам

$$AP + PA^T + BY + Y^T B^T \prec 0, \quad P \succ 0. \quad (3.1.12)$$

Тогда регулятор (3.1.11) с матрицей

$$\widehat{K} = \widehat{Y} \widehat{P}^{-1}$$

стабилизирует систему (3.1.10), а квадратичная форма

$$V(x) = x^T \widehat{P}^{-1} x$$

является функцией Ляпунова для замкнутой системы.

Доказательство. Замкнув систему (3.1.10) статической линейной обратной связью по состоянию (3.1.11), приходим к замкнутой системе

$$\dot{x} = A_c x \quad (3.1.13)$$

с матрицей $A_c = A + BK$.

Как следует из вышеизложенного, функция (3.1.9) является функцией Ляпунова для системы (3.1.13) тогда и только тогда, когда

$$A_c^T Q + Q A_c \prec 0,$$

то есть когда найдутся матрицы K и $Q \succ 0$ такие, что

$$(A + BK)^T Q + Q(A + BK) \prec 0. \quad (3.1.14)$$

Неравенство (3.1.14) нелинейно (и невыпукло!) относительно матричных переменных Q и K . Его можно свести к линейному с помощью следующего приема, который будет систематически использоваться в дальнейшем.

Домножив матричное неравенство (3.1.14) слева и справа на матрицу $P = Q^{-1}$, получим

$$(A + BK)P + P(A + BK)^T \prec 0,$$

или

$$AP + PA^T + BKP + PK^T B^T \prec 0.$$

Введем вспомогательную матричную переменную

$$Y = KP \in \mathbb{R}^{p \times n},$$

исключая K . В силу $P \succ 0$, матрица K восстанавливается единственным образом:

$$K = YP^{-1}.$$

В результате приходим к эквивалентному матричному неравенству

$$AP + PA^T + BY + Y^T B^T \prec 0,$$

линейному по переменным Y и P . ■

Задача из теоремы 3.1.5 представляет собой задачу разрешимости ЛМН, рассмотренную в разделе 2.7.2. Из доказательства видно, что для управляемых систем разрешимость линейных матричных неравенств (3.1.12) эквивалентна возможности стабилизировать систему с помощью линейной обратной связи по состоянию. Более того, теорема предлагает простую параметризацию всех стабилизирующих регуляторов через решения этого ЛМН и служит основой синтеза. Понятно, что никаких выводов о *степени устойчивости* замкнутой системы в утверждении теоремы 3.1.5 не делается; она зависит от того, какое именно допустимое решение ЛМН (3.1.12) будет выбрано.

Замечание 3.1.4. Иногда нужно не просто стабилизировать систему, а гарантировать желаемую степень $\sigma > 0$ ее устойчивости.

В этом случае первое из неравенств (3.1.12) заменится на нестрогое ЛМН

$$AP + PA^T + BY + Y^T B^T \preccurlyeq -2\sigma P$$

в формулировке теоремы 3.1.5. ▼

Таким образом, имеем следующий результат.

Следствие 3.1.1. *Пусть матрицы \hat{P} и \hat{Y} удовлетворяют линейным матричным неравенствам*

$$AP + PA^T + BY + Y^T B^T \preccurlyeq -2\sigma P, \quad P \succ 0 \quad (3.1.15)$$

при некотором фиксированном значении параметра $\sigma > 0$.

Тогда регулятор (3.1.11) с матрицей

$$\hat{K} = \hat{Y}\hat{P}^{-1}$$

стабилизирует систему (3.1.10) со степенью устойчивости σ , а квадратичная форма

$$V(x) = x^T \hat{P}^{-1} x$$

является функцией Ляпунова для замкнутой системы.

Упражнение 3.1.4. *Докажите следствие 3.1.1, используя замечание 3.1.3.*

Из теоремы 3.1.5 вытекает, что эллипсоид

$$\mathcal{E} = \{x \in \mathbb{R}^n : x^T \hat{P}^{-1} x \leq 1\} \quad (3.1.16)$$

с центром в начале координат и матрицей \hat{P} является инвариантным относительно начальных условий для замкнутой системы (3.1.10), см. замечание 3.1.1.

Если при этом начальное условие $x(0) = x_0$ системы (3.1.10) задано, то естественно потребовать, чтобы эллипсоид (3.1.16) содержал начальную точку, то есть

$$x_0^T \hat{P}^{-1} x_0 \leq 1,$$

или в ЛМН-форме:

$$\begin{pmatrix} 1 & x_0^T \\ x_0 & \hat{P} \end{pmatrix} \succcurlyeq 0, \quad (3.1.17)$$

и это неравенство следует добавить к ограничениям (3.1.12) в теореме 3.1.5.

Обычно на практике требуется, чтобы управляющие воздействия были тем или иным образом ограничены. Покажем, как можно учитывать ограничения на величину управления в рамках описанной выше квадратичной стабилизации.

Будем накладывать следующее ограничение на управление, которое должно выполняться вдоль траектории системы:

$$\|u(t)\| \leq \mu \quad \forall t \geq 0, \quad (3.1.18)$$

где $\mu > 0$ — уровень допустимых управлений.

В следующей лемме дается достаточное условие для выполнения этого ограничения, которое формулируется в виде линейного матричного неравенства относительно матричных переменных P и Y , фигурирующих в теореме 3.1.5. Это неравенство будет использоваться и далее, при рассмотрении систем с внешними возмущениями.

Лемма 3.1.1. *Пусть матрицы P и Y удовлетворяют линейным матричным неравенствам (3.1.12) (или (3.1.15)) и (3.1.17). Тогда выполнение ЛМН*

$$\begin{pmatrix} P & Y^T \\ Y & \mu^2 I \end{pmatrix} \succ 0 \quad (3.1.19)$$

гарантирует выполнение ограничения (3.1.18) на траектории системы (3.1.10), замкнутой обратной связью $K = YP^{-1}$, с начальным условием $x(0) = x_0$.

Доказательство. Поскольку $u = Kx$, ограничение (3.1.18) может быть представлено в виде

$$x^T K^T K x \leq \mu^2.$$

Рассмотрим эллипсоид

$$\mathcal{E} = \{x \in \mathbb{R}^n : x^T Q x \leq 1\} \quad (3.1.20)$$

с матрицей $Q = P^{-1} \succ 0$, удовлетворяющей (3.1.12) и (3.1.17), и потребуем выполнения

$$x^T K^T K x \leq \mu^2 \quad \text{при} \quad x^T Q x \leq 1. \quad (3.1.21)$$

Легко видеть, что если эллипсоиды \mathcal{E}_1 и \mathcal{E}_2 заданы выражением вида (3.1.20) с матрицами Q_1 и Q_2 , то вложенность $\mathcal{E}_1 \subseteq \mathcal{E}_2$ эквивалентна соотношению $Q_1 \succ Q_2$. Поэтому для выполнения (3.1.21) необходимо и достаточно, чтобы

$$\frac{1}{\mu^2} K^T K \preceq Q.$$

Далее, поскольку $K = YP^{-1}$, то полученнное матричное неравенство примет вид

$$\frac{1}{\mu^2} P^{-1} Y^T Y P^{-1} \preceq P^{-1}.$$

Домножив его слева и справа на матрицу P , получаем

$$\frac{1}{\mu^2} Y^T Y \preceq P,$$

или, по лемме Шура,

$$\begin{pmatrix} P & Y^T \\ Y & \mu^2 I \end{pmatrix} \succcurlyeq 0,$$

что и завершает доказательство. ■

Достаточность этого условия следует из (3.1.21): выполнение ограничения (3.1.18) требуется не только вдоль траектории системы, но и в любой точке эллипсоида с матрицей P , содержащего траекторию.

Таким образом, для построения квадратично стабилизирующего регулятора при наличии ограничения на управление теорема 3.1.5 модифицируется следующим образом. К неравенствам (3.1.12), гарантирующим стабилизацию, добавляются линейные матричные неравенства (3.1.17) и (3.1.19). При этом эллипсоид с полученной матрицей P_* будет инвариантным, содержащим начальное условие x_0 , а ограниченность управления вдоль траектории гарантируется условием (3.1.19).

Ясно, что не для всякого начального условия x_0 систему можно стабилизировать управлением заданного уровня. Вопрос о минимально возможном значении величины μ решается с помощью следующего очевидного развития леммы 3.1.1 с учетом замечания 3.1.4.

Лемма 3.1.2. *Пусть γ_* — решение задачи*

$$\min \gamma$$

при ограничениях

$$AP + PA^T + BY + Y^T B^T \preceq -2\sigma P, \quad P \succ 0,$$

$$\begin{pmatrix} 1 & x_0^T \\ x_0 & P \end{pmatrix} \succcurlyeq 0, \quad \begin{pmatrix} P & Y^T \\ Y & \gamma I \end{pmatrix} \succcurlyeq 0$$

относительно матричных переменных $P \in \mathbb{S}^n$, $Y \in \mathbb{R}^{p \times n}$ и скалярной переменной γ при некотором фиксированном значении параметра $\sigma > 0$.

Тогда при

$$\mu \geq \mu_{\min} = \sqrt{\gamma_*}$$

существует линейный регулятор $u = Kx$, стабилизирующий систему (3.1.10) со степенью устойчивости σ и удовлетворяющий ограничению (3.1.18).

Оптимизационная задача, сформулированная в лемме 3.1.2, представляет собой задачу полуопределенного программирования — SDP; см. раздел 2.7. Вообще говоря, второе из ограничений ($P \succ 0$) представляет собой строгое матричное неравенство, однако в силу замечания 2.7.2 будем его понимать как $P \succcurlyeq \delta I$ при некотором (небольшом) $\delta > 0$.

К задачам SDP будут сводиться многие из задач, рассматриваемых в данной главе.

Замечание 3.1.5. В качестве программной среды для численных расчетов во всех примерах нами использовалась среда MATLAB. При этом для решения задач разрешимости ЛМН, полуопределенного программирования и параметризованных задач полуопределенного программирования, с которыми читатель встретится в этом и следующем разделе, привлекались два свободно распространяемых пакета — SeDuMi [128] и YALMIP [107].

Первый из них представляет собой решатель для широкого круга задач выпуклой оптимизации, в основе которого лежат методы внутренней точки (см. раздел 2.6). Второй пакет является интерфейсом — простым и гибким языком обращения к решателям. Удобной альтернативой последнему может служить также свободно распространяемый пакет cvx [91]. ▼

Пример 3.1.1. В качестве примера рассмотрим задачу управления двухмассовой системой — системой из двух твердых тел с массами m_1 и m_2 , соединенных пружиной с коэффициентом упругости k , скользящих без трения вдоль неподвижного горизонтального стержня (рис. 3.1.1); к левому телу приложено управление u .

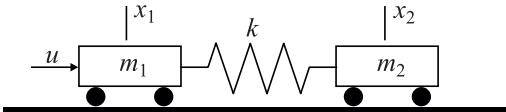


Рис. 3.1.1. Двухмассовая система из примера 3.1.1.

Эта задача часто используется как тестовая для различных методов синтеза регуляторов, чему способствуют ее реальное происхождение и разумная размерность модели; в различных постановках она будет привлекаться ниже для демонстрации работоспособности и эффективности предлагаемых методов синтеза управления.

Обозначим через x_1, v_1 координату и скорость левого тела, а через x_2, v_2 — правого тела. Тогда

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ v_1 \\ v_2 \end{pmatrix}$$

есть вектор состояния рассматриваемой динамической системы, полностью описывающий ее поведение.

Непрерывная модель колебаний системы описывается уравнениями

$$\begin{aligned} \dot{x}_1 &= v_1, \\ \dot{x}_2 &= v_2, \\ \dot{v}_1 &= -\frac{k}{m_1}x_1 + \frac{k}{m_2}x_2 + \frac{1}{m_1}u, \\ \dot{v}_2 &= \frac{k}{m_2}x_1 - \frac{k}{m_2}x_2. \end{aligned}$$

В результате приходим к системе вида (3.1.10) с матрицами

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{k}{m_1} & \frac{k}{m_2} & 0 & 0 \\ \frac{k}{m_2} & -\frac{k}{m_1} & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ \frac{1}{m_1} \\ 0 \end{pmatrix}.$$

Пусть параметры системы единичны:

$$k_1 = k_2 = m_1 = m_2 = 1,$$

а в качестве начальной точки возьмем

$$x_0 = \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix}.$$

Зададимся также желаемой степенью устойчивости замкнутой системы:

$$-\max_i \operatorname{Re} \lambda(A_c) = \sigma = 1.$$

Решая задачу допустимости для системы линейных матричных неравенств

$$AP + PA^T + BY + Y^T B^T \preceq -2\sigma P, \quad \begin{pmatrix} 1 & x_0^T \\ x_0 & P \end{pmatrix} \succcurlyeq 0, \quad P \succ 0,$$

получаем

$$\hat{P} = \begin{pmatrix} 97,4738 & 17,0132 & -107,4174 & -43,8841 \\ 17,0132 & 12,5907 & 7,2590 & -14,3540 \\ -107,4174 & 7,2590 & 440,7313 & 13,6781 \\ -43,8841 & -14,3540 & 13,6781 & 27,0149 \end{pmatrix}$$

и

$$\hat{Y} = (-145,4358 \quad -23,7736 \quad -589,3107 \quad 57,7900),$$

что дает регулятор

$$\hat{K} = (-22,3342 \quad -11,0404 \quad -5,4426 \quad -37,2519),$$

стабилизирующий замкнутую систему со степенью устойчивости

$$\sigma = \sigma(A + B\hat{K}) = 1,2165.$$

Эта величина оказывается заметно больше желаемой $\sigma = 1$, поскольку, повторимся, регулятор \hat{K} был получен в результате решения задачи допустимости.

В приведенном решении никак не учитывалась величина управляющего воздействия; решим теперь задачу заново, наложив на него явные ограничения. Для этого сначала определим величину μ_{\min} минимально допустимого ресурса управления, как это предписано леммой 3.1.2:

$$\mu_{\min} = 32,2736.$$

Положим теперь

$$\tilde{\mu} = 1,2\mu_{\min} = 38,7284$$

и воспользуемся леммой 3.1.1 с учетом замечания 3.1.4. В результате решения задачи допустимости

$$AP + PA^T + BY + Y^T B^T \preceq -2\sigma P,$$

$$\begin{pmatrix} 1 & x_0^T \\ x_0 & P \end{pmatrix} \succcurlyeq 0, \quad \begin{pmatrix} P & Y^T \\ Y & \tilde{\mu}^2 \end{pmatrix} \succcurlyeq 0, \quad P \succ 0$$

получаем

$$\tilde{P} = \begin{pmatrix} 139,9102 & 35,8094 & -146,2256 & -69,0944 \\ 35,8094 & 24,1731 & -6,0969 & -25,4615 \\ -146,2256 & -6,0969 & 349,1403 & 40,4163 \\ -69,0944 & -25,4615 & 40,4163 & 41,7251 \end{pmatrix}$$

и

$$\tilde{Y} = (58,0144 \quad -14,9543 \quad -505,0770 \quad 12,1049),$$

что дает стабилизирующий регулятор

$$\tilde{K} = (-10,1007 \quad 0,9545 \quad -4,3082 \quad -11,6806),$$

норма которого почти вчетверо меньше нормы регулятора \hat{K} . При этом степень устойчивости замкнутой системы оказывается равной

$$\tilde{\sigma} = \sigma(A + B\tilde{K}) = 1,0676,$$

причем везде внутри инвариантного эллипсоида с матрицей \tilde{P} гарантировано следующее ограничение на управление:

$$|u| = |\tilde{K}x| \leq 38,7284.$$

В действительности управление не превышает заметно меньшей величины 18,4277, поскольку лемма 3.1.1 дает лишь достаточные условия; см. рис. 3.1.2, на котором приведены графики управления для системы, замкнутой регуляторами \hat{K} (пунктир) и \tilde{K} (сплошная линия).

На рис. 3.1.3 показано изменение со временем ∞ -нормы решения системы, замкнутой регуляторами \hat{K} (пунктир) и \tilde{K} (сплошная линия).

Обратим внимание на явно выраженный *эффект всплеска* траектории системы, замкнутой регулятором \hat{K} : норма решения возрастает более чем в семь раз, прежде чем начать убывать, — довольно большая величина для исследуемой маломерной системы. В то же время при регуляторе \tilde{K} наблюдается почти вдвое меньший всплеск.

Итак, введение ограничения на управление в данном примере приводит к гораздо меньшим коэффициентам усиления, придает замкнутой системе степень устойчивости, существенно более близкую к желаемой, и заметно ослабляет эффект всплеска. \blacktriangledown

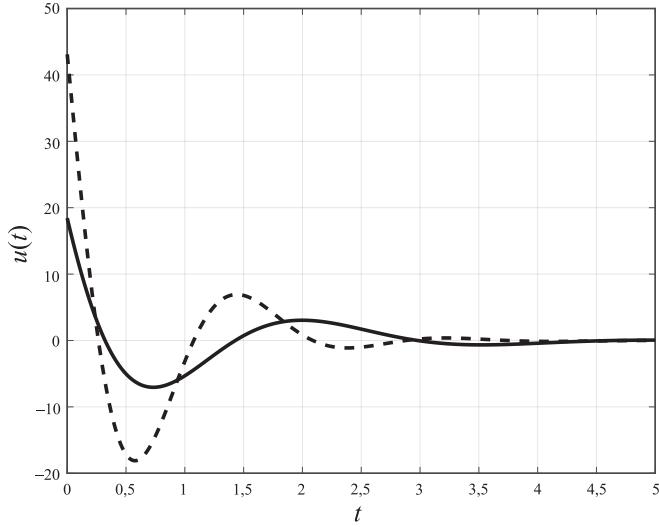


Рис. 3.1.2. Управление в системе из примера 3.1.2 при различных регуляторах.

Таким образом, отыскание стабилизирующего регулятора может формулироваться в терминах линейных матричных неравенств и сводиться к решению задачи допустимости.

Обратимся теперь к случаю дискретного времени; для него справедлив результат, аналогичный теореме 3.1.5. В самом деле, рассмотрим линейную систему управления в дискретном времени

$$x_{k+1} = Ax_k + Bu_k, \quad (3.1.22)$$

где $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{p \times n}$, с состоянием $x_k \in \mathbb{R}^n$ и управлением $u_k \in \mathbb{R}^p$, выбираемым в виде

$$u_k = Kx_k. \quad (3.1.23)$$

Пара (A, B) предполагается управляемой, см. раздел B.7.

Как показано в разделе 3.1.1, функция

$$V(x) = x^T P^{-1} x, \quad P \succ 0,$$

является функцией Ляпунова для замкнутой системы с матрицей A_c тогда и только тогда, когда

$$A_c P A_c^T - P \prec 0.$$

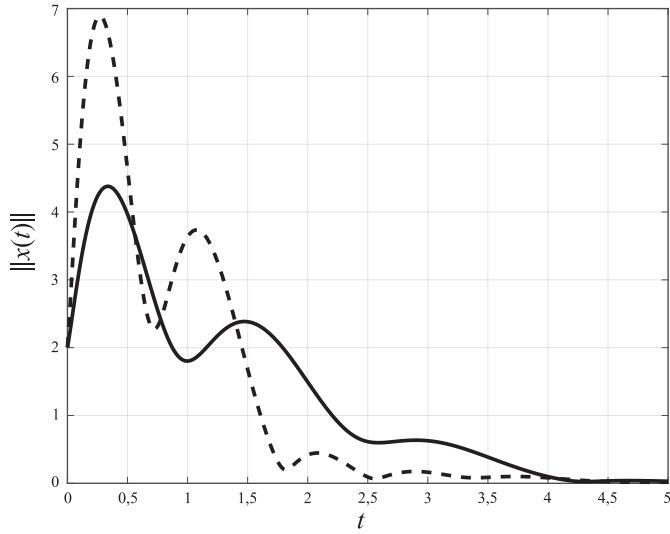


Рис. 3.1.3. Норма решения системы из примера 3.1.2 для различных регуляторов.

Преобразуя это ЛМН с помощью леммы Шура (лемма 2.7.2) и вводя, как и выше, новую матричную переменную $Y = KP$, приходим к следующему утверждению.

Теорема 3.1.6. *Пусть матрицы \hat{P} и \hat{Y} удовлетворяют линейным матричным неравенствам*

$$\begin{pmatrix} P & AP + BY \\ PA^T + Y^T B^T & P \end{pmatrix} \succ 0, \quad P \succ 0.$$

Тогда регулятор (3.1.23) с матрицей

$$\hat{K} = \hat{Y}\hat{P}^{-1}$$

стабилизирует систему (3.1.22), а квадратичная форма

$$V(x) = x^T \hat{P}^{-1} x$$

является функцией Ляпунова для замкнутой системы.

Таким образом, и в дискретном, и в непрерывном времени отыскание стабилизирующего регулятора в форме статической линейной

обратной связи по состоянию может формулироваться в терминах линейных матричных неравенств и сводиться к решению задачи разрешимости.

Описанный выше подход к стабилизации, основанный на поиске квадратичной функции Ляпунова (и потому называемый *квадратичной стабилизацией*) не дает решения в явном виде, а сводит задачу к решению линейных матричных неравенств. Такая техника оказывается особенно эффективной для задач робастной стабилизации, когда требуется стабилизировать систему в условиях неопределенности (а также при наличии внешних возмущений и в разнообразных задачах оптимизации), чему и посвящены следующие разделы.

3.1.3. Линейно-квадратичное управление по состоянию

В предыдущем разделе показано, что стабилизирующий регулятор может строиться через решения некоторых систем ЛМН. Каждая такая система имеет множество решений, и среди них можно выбирать то, которое оптимизирует тот или иной показатель качества. В этом разделе продемонстрируем эффективность техники ЛМН и простоту ее применения к решению некоторых классических задач оптимального управления и покажем, как эти задачи могут сводиться к формату полуопределенного программирования. Ограничимся непрерывными системами; дискретные аналоги приводимых ниже результатов легко могут быть получены по той же схеме с использованием той же ЛМН-техники.

Рассмотрим линейную систему управления с фиксированным начальным условием:

$$\dot{x} = Ax + Bu, \quad x(0) = x_0, \quad (3.1.24)$$

где $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, с состоянием $x(t) \in \mathbb{R}^n$ и управлением $u(t) \in \mathbb{R}^p$; пара (A, B) управляема (см. приложение B.7).

Будем искать закон управления в форме линейной обратной связи по состоянию

$$u = Kx, \quad K \in \mathbb{R}^{p \times n}, \quad (3.1.25)$$

минимизирующий следующий *квадратичный критерий качества*:

$$J = \int_0^\infty (x^T Rx + u^T Su) dt. \quad (3.1.26)$$

Здесь $R \in \mathbb{S}^n$ и $S \in \mathbb{S}^p$ — заданные положительно определенные матрицы, так что $J \geq 0$.

Примечательно, что в этой задаче расширение класса управлений до программных не приводит к улучшению используемого критерия качества.

Очевидно, что для того чтобы функционал J был конечен, необходимо и достаточно, чтобы система (3.1.24), замкнутая обратной связью (3.1.25), была устойчива. При этом управляемость пары (A, B) гарантирует существование стабилизирующих обратных связей.

Критерий (3.1.26) гарантирует малость как состояний замкнутой системы, так и применяющихся управлений (за счет больших управлений можно добиться, чтобы $x(t)$ быстро убывал, однако тогда член $u^T S u$ будет очень велик). Матрицы R и S играют роль *весовых коэффициентов*, учитывающих оба указанных фактора.

Приведенная выше задача является одной из основных в теории управления; в отечественной литературе она иногда называется *задачей об аналитическом конструировании регуляторов*, а в западной — *задачей о линейно-квадратичном регуляторе*.

Стандартный метод решения этой задачи (см. подробнее [9, 29]) основан на рассмотрении *алгебраического уравнения Риккати*

$$A^T Q + QA - QBS^{-1}B^T Q + R = 0 \quad (3.1.27)$$

относительно матрицы Q . При этом оптимальный регулятор задается выражением

$$K_* = -S^{-1}B^T Q_*,$$

а значение функционала J равно

$$J_* = x_0^T Q_* x_0,$$

где Q_* — положительно определенное решение уравнения (3.1.27).

Однако для решения этой задачи можно применить технику, использующую аппарат линейных матричных неравенств. В этом разделе и далее будет использоваться следующий результат, известный под названием *лемма Беллмана*.

Лемма 3.1.3 (Беллман [76]). *Значение функционала*

$$J = \int_0^\infty x^T R x dt, \quad R \succ 0,$$

на решениях системы

$$\dot{x} = Ax, \quad x(0) = x_0, \quad (3.1.28)$$

с устойчивой матрицей A равно

$$J = x_0^T Q x_0,$$

где матрица Q удовлетворяет уравнению Ляпунова

$$A^T Q + Q A = -R. \quad (3.1.29)$$

Доказательство. Поскольку матрица A устойчива и $R \succ 0$, то по лемме Б.4 уравнение (3.1.29) имеет единственное решение $Q \succ 0$.

Рассмотрим функцию

$$V(x) = x^T Q x$$

на решениях системы (3.1.28). Поскольку $Q \succ 0$, то $V(x) > 0$ при $x \neq 0$; при этом производная функции $V(x)$ в силу системы имеет вид

$$\dot{V}(x) = \frac{d}{dt} V(x(t)) = x^T (A^T Q + Q A) x = -x^T R x < 0,$$

откуда

$$\int_0^\infty x^T R x \, dt = - \int_0^\infty \dot{V}(x) \, dt = -V(x) \Big|_0^\infty = x_0^T Q x_0,$$

что и завершает доказательство. ■

Замечание 3.1.6. Пусть матрица $Q \succ 0$ является решением матричного неравенства Ляпунова

$$A^T Q + Q A \preccurlyeq -R. \quad (3.1.30)$$

В силу леммы Б.5 имеем $Q \succcurlyeq Q_*$, где Q_* — решение уравнения Ляпунова (3.1.29), поэтому

$$J_* = x_0^T Q_* x_0 \leq x_0^T Q x_0.$$

Следовательно, точное значение функционала J можно искать как решение задачи SDP

$$\min_Q x_0^T Q x_0 \quad \text{при ограничении (3.1.30).}$$

Вернемся к исходной задаче. Система (3.1.24), замкнутая обратной связью (3.1.25), принимает вид

$$\dot{x} = (A + BK)x,$$

и на ее решениях имеем

$$J = \int_0^\infty (x^T Rx + x^T K^T SKx) dt = \int_0^\infty x^T (R + K^T SK)x dt.$$

Воспользовавшись леммой Беллмана 3.1.3 (точнее, замечанием к ней), получаем, что минимальное значение функционала J обеспечивается решением задачи

$$\min x_0^T Q x_0$$

при ограничении

$$(A + BK)^T Q + Q(A + BK) \preceq -R - K^T SK.$$

Домножив последнее матричное неравенство слева и справа на матрицу $P = Q^{-1}$, получаем

$$(A + BK)P + P(A + BK)^T + PRP + PK^T SKP \preceq 0,$$

или, дважды применяя лемму Шура,

$$\begin{pmatrix} (A + BK)P + P(A + BK)^T & P & PK^T \\ P & -R^{-1} & 0 \\ KP & 0 & -S^{-1} \end{pmatrix} \preceq 0.$$

Окончательно, введя вспомогательную матричную переменную

$$Y = KP,$$

приходим к линейному матричному неравенству

$$\begin{pmatrix} AP + PA^T + BY + Y^T B^T & P & Y^T \\ P & -R^{-1} & 0 \\ Y & 0 & -S^{-1} \end{pmatrix} \preceq 0. \quad (3.1.31)$$

Таким образом, на решениях полученного матричного неравенства следует минимизировать величину

$$x_0^T Q x_0 = x_0^T P^{-1} x_0,$$

которая нелинейна по переменной P . Для этого введем скалярную переменную γ и, используя прием (2.7.7)–(2.7.8) из раздела 2.7.3, запишем неравенство

$$x_0^T P^{-1} x_0 \leq \gamma$$

в виде ЛМН

$$\begin{pmatrix} \gamma & x_0^T \\ x_0 & P \end{pmatrix} \succcurlyeq 0.$$

Наконец, поскольку весовая матрица R положительно определена, в силу следствия 2.7.1 имеем

$$AP + PA^T + BY + Y^T B^T \prec 0$$

для всех P и Y , удовлетворяющих (3.1.31). Это означает, что соответствующий регулятор $K = YP^{-1}$ будет стабилизирующим (см. раздел 3.1.1).

В результате приходим к следующему утверждению.

Теорема 3.1.7. Пусть P_* , Y_* – решение задачи SDP

$$\min \gamma$$

при ограничениях

$$\begin{pmatrix} AP + PA^T + BY + Y^T B^T & P & Y^T \\ P & -R^{-1} & 0 \\ Y & 0 & -S^{-1} \end{pmatrix} \preccurlyeq 0, \quad (3.1.32)$$

$$\begin{pmatrix} \gamma & x_0^T \\ x_0 & P \end{pmatrix} \succcurlyeq 0, \quad P \succ 0$$

относительно матричных переменных $P \in \mathbb{S}^n$, $Y \in \mathbb{R}^{p \times n}$ и скалярной переменной γ .

Тогда регулятор (3.1.25) с матрицей

$$K_* = Y_* P_*^{-1}$$

стабилизирует систему (3.1.24). При этом квадратичная форма

$$V(x) = x^T P_*^{-1} x$$

является функцией Ляпунова для замкнутой системы, а величина $J_* = x_0^T P_*^{-1} x_0$ дает минимальное значение функционала (3.1.26) на решениях системы (3.1.24) с начальным условием x_0 .

Оптимизационная задача, сформулированная в теореме 3.1.7, представляет собой задачу полуопределенного программирования относительно двух матричных переменных P и Y и одной скалярной переменной γ . Таким образом, классическая задача о линейно-квадратичном регуляторе сведена к формату SDP. Более того, поскольку весовые матрицы R и S (точнее, матрицы, обратные к ним) входят в матричное неравенство (3.1.32) линейно, то при необходимости они могут рассматриваться как переменные.

Пример 3.1.2. Вернемся к двухмассовой системе, рассмотренной в примере 3.1.1, и построим для нее линейно-квадратичный регулятор, полагая весовые матрицы единичными:

$$R = S = I.$$

Воспользовавшись теоремой 3.1.7, для начальной точки

$$x_0 = \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix}$$

находим

$$P_* = \begin{pmatrix} 0,5847 & 0,2782 & -0,2891 & -0,2743 \\ 0,2782 & 0,4491 & -0,0784 & -0,1546 \\ -0,2891 & -0,0784 & 0,6993 & -0,0005 \\ -0,2743 & -0,1546 & -0,0005 & 0,3745 \end{pmatrix},$$

$$Y_* = (0 \ 0 \ -1 \ 0),$$

откуда

$$K_* = (-1,7212 \ 0,3070 \ -2,1078 \ -1,1366)$$

и значение функционала

$$J_* = 16,6635.$$

При этом замкнутая система имеет вполне разумную степень устойчивости

$$\sigma(A_c) = -\max_i \operatorname{Re} \lambda_i(A + BK_*) = 0,3738. \quad \blacktriangledown$$

Замечание 3.1.7. Можно рассматривать задачу о линейно-квадратичном регуляторе в более общей постановке, когда минимизируемый функционал имеет структуру

$$J = \int_0^{\infty} (x^T Rx + 2u^T L^T x + u^T Su) dt,$$

где $R \in \mathbb{S}^n$, $L \in \mathbb{R}^{n \times p}$ и $S \in \mathbb{S}^p$ — заданные матрицы такие, что

$$\begin{pmatrix} R & L \\ L^T & S \end{pmatrix} \succ 0.$$

В этом случае задача сводится к рассмотренной выше с помощью замены

$$u = \tilde{u} - S^{-1}L^T x.$$

▼

Обратим внимание, что в формулировке теоремы 3.1.7 явно присутствуют начальные условия x_0 системы, поэтому полученный регулятор, будучи оптимальным по критерию, зависит от значения x_0 , и для других начальных условий задачу придется решать заново. Этим полученное решение отличается от классического (использующего уравнение Риккати), которое приводит к «универсальному» регулятору, дающему оптимальное значение критерия *для любого начального условия*.

Такой неприятности можно избежать, если принять во внимание следствие Б.4, согласно которому, оптимизируя *след* матрицы на решениях *неравенства Ляпунова* или Риккати, получаем решение соответствующего *уравнения*. Иными словами, решая задачу SDP вида $\max \text{tr } P$ при ограничении (3.1.32), приходим к регулятору, совпадающему с классическим, полученным через решение уравнения Риккати.

3.1.4. H_{∞} -оптимизация

Рассмотрим линейную систему управления

$$\begin{aligned} \dot{x} &= Ax + Dw, & x(0) &= 0, \\ y &= Cx, \end{aligned} \tag{3.1.33}$$

где $A \in \mathbb{R}^{n \times n}$, $D \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{\ell \times n}$ с состоянием $x(t) \in \mathbb{R}^n$, входом (внешним возмущением) $w(t) \in \mathbb{R}^m$ и выходом $y(t) \in \mathbb{R}^{\ell}$.

Напомним (см. приложение B.3), что передаточная функция системы (3.1.33) от входа w к выходу y имеет вид

$$G(s) = C(sI - A)^{-1}D.$$

При этом выражение

$$\|G(s)\|_\infty = \sup_{\omega} \|C(j\omega I - A)^{-1}D\|_2$$

определяет так называемую H_∞ -норму передаточной функции $G(s)$ (см. приложение B.4); для устойчивой системы эта величина конечна и равна коэффициенту

$$\sup_{\|w\|_2 \neq 0} \frac{\|y\|_2}{\|w\|_2}$$

усиления входного сигнала при прохождении через систему (входной и выходной сигнал измеряются в норме L_2).

Построение систем, оптимальных по критерию H_∞ -нормы, составляет предмет H_∞ -оптимизации — большого раздела теории автоматического управления.

Имеет место следующий результат.

Лемма 3.1.4. *Рассмотрим систему, заданную с помощью передаточной функции:*

$$z = G(s)w, \quad G(s) = C(j\omega I - A)^{-1}D,$$

где матрица A гурвицева, z — выход системы, а w — внешнее возмущение, ограниченное в L_2 -норме:

$$\|w\|_2^2 = \int_0^\infty w^T(t)w(t) dt \leq 1. \quad (3.1.34)$$

Тогда

$$\|G(s)\|_\infty = \sup_w \|z\|_2,$$

где супремум берется по всем входным возмущениям, удовлетворяющим (3.1.34).

Одной из задач теории H_∞ -оптимизации является вычисление значения H_∞ -нормы передаточной функции устойчивой системы. Способы вычисления этой величины при описании системы с помощью передаточных функций хорошо известны. Рассмотрим альтернативный

способ, предполагающий описание системы в пространстве состояний и основанный на технике линейных матричных неравенств. Он численно устойчив и дает точный результат при численной реализации.

Рассмотрим несколько более общую формулировку, а именно, для системы

$$\dot{x} = Ax + Dw, \quad x(0) = 0,$$

с гурвицевой матрицей $A \in \mathbb{R}^{n \times n}$, состоянием $x(t) \in \mathbb{R}^n$ и L_2 -ограниченным внешним возмущением $w(t) \in \mathbb{R}^m$, удовлетворяющим ограничению (3.1.34), будем интересоваться величиной

$$J = \int_0^\infty x^T Rx dt, \quad R \succ 0.$$

В частности, при $R = I$ этот показатель совпадает с квадратом L_2 -нормы решения:

$$\int_0^\infty x^T x dt = \|x\|_2^2.$$

Какие значения может принимать функционал J на решениях системы при всевозможных внешних возмущениях $\|w\|_2 \leq 1$? Ясно, что $J_{\min} = 0$ при $w(t) \equiv 0$ (тогда $x(t) \equiv 0$), поэтому важно найти величину $J_{\max} = \max_{\|w\|_2 \leq 1} J$.

Теорема 3.1.8. Для функционала

$$J = \int_0^\infty x^T Rx dt, \quad R \succ 0,$$

на решениях системы

$$\dot{x} = Ax + Dw, \quad x(0) = 0, \quad (3.1.35)$$

с гурвицевой матрицей $A \in \mathbb{R}^{n \times n}$, $D \in \mathbb{R}^{n \times m}$, состоянием $x(t)$ и L_2 -ограниченным внешним возмущением $w(t)$: $\|w\|_2 \leq 1$ справедлива оценка

$$J \leq \gamma_*,$$

где γ_* — решение задачи SDP

$$\min \gamma \quad npu \quad \begin{pmatrix} A^T Q + Q A + R & Q D \\ D^T Q & -\gamma I \end{pmatrix} \preccurlyeq 0, \quad Q \succ 0 \quad (3.1.36)$$

относительно матричной переменной $Q \in \mathbb{S}^n$ и скалярной переменной γ .

Доказательство. Пусть ЛМН (3.1.36) выполняется для некоторого $\gamma > 0$ и $Q \succ 0$. Это означает, что для всех $x \in \mathbb{R}^n$, $w \in \mathbb{R}^m$ верно

$$\begin{pmatrix} x \\ w \end{pmatrix}^T \begin{pmatrix} A^T Q + Q A + R & Q D \\ D^T Q & -\gamma I \end{pmatrix} \begin{pmatrix} x \\ w \end{pmatrix} \leq 0,$$

или

$$x^T (A^T Q + Q A + R) x + x^T Q D w + w^T D^T Q x - \gamma w^T w \leq 0. \quad (3.1.37)$$

Рассмотрим квадратичную форму

$$V(x) = x^T Q x$$

с некоторой матрицей $Q \succ 0$. На решениях системы (3.1.35) имеем

$$\dot{V}(x) = \dot{x}^T Q x + x^T Q \dot{x} = (Ax + Dw)^T Q x + x^T Q (Ax + Dw),$$

откуда с учетом (3.1.37)

$$\dot{V}(x) \leq -x^T R x + \gamma w^T w.$$

Интегрируя это неравенство по t от 0 до T , с учетом $V(x(0)) = 0$ получаем

$$V(x(T)) \leq - \int_0^T x^T R x dt + \gamma \int_0^T w^T w dt.$$

Поскольку $V(x(T)) \geq 0$, переходя к пределу при $T \rightarrow \infty$ (что возможно из силу гурвицевости матрицы A) имеем

$$\int_0^\infty x^T R x dt \leq \gamma \int_0^\infty w^T w dt \leq \gamma.$$

Таким образом, число γ дает верхнюю оценку для функционала J . ■

Как видим, сформулированная задача оптимального управления сведена к задаче полуопределенного программирования относительно матричной переменной Q и скалярной переменной γ .

Заметим, что если пара (A, D) управляема (см. приложение B.7), то оценка сверху для функционала J , предоставляемая величиной γ_* , является достижимой. Иными словами, существует возмущение (3.1.34), для которого $J = \gamma_*$.

Нетрудно видеть, что утверждение теоремы 3.1.8 остается справедливым и в том случае, когда $R = C^T C$ и пара (A, C) наблюдаема (см. приложение B.9). Тогда функционал J может пониматься как квадрат L_2 -нормы выхода $z = Cx$ системы (3.1.35) при L_2 -ограниченном внешнем возмущении, а величина γ_* , полученная в теореме (3.1.8), представляет собой не что иное как квадрат H_∞ -нормы передаточной функции от входа w к выходу z . Таким образом, теорема (3.1.8) предлагает еще один способ вычисления H_∞ -нормы.

Обратимся теперь к задаче синтеза и рассмотрим линейную систему управления

$$\begin{aligned}\dot{x} &= Ax + B_1 u + Dw, \quad x(0) = 0, \\ z &= Cx + B_2 u,\end{aligned}\tag{3.1.38}$$

где $A \in \mathbb{R}^{n \times n}$, $B_1 \in \mathbb{R}^{n \times p}$, $D \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{\ell \times n}$, $B_2 \in \mathbb{R}^{\ell \times p}$, с состоянием $x(t) \in \mathbb{R}^n$, регулируемым выходом $z(t) \in \mathbb{R}^\ell$, управлением $u(t) \in \mathbb{R}^p$ и внешним возмущением $w(t) \in \mathbb{R}^m$, ограниченным в L_2 -норме.

В простейшей постановке задача H_∞ -оптимизации заключается в выборе регулятора в форме статической линейной обратной связи по состоянию

$$u = Kx, \quad K \in \mathbb{R}^{p \times n},\tag{3.1.39}$$

который минимизирует величину H_∞ -нормы передаточной функции системы (3.1.38).

В соответствии с леммой 3.1.4 введем в рассмотрение функционал

$$J = \int_0^\infty z^T z dt = \|z\|_2^2$$

и будем искать его минимум по всем стабилизирующими регуляторам вида (3.1.39); решение этой задачи и дает H_∞ -оптимальный регулятор.

Обратим внимание, что в (3.1.38) управление u включено в уравнение для выхода z для того, чтобы ограничить величину используемого управления. С той же самой целью управление было введено в критерий J в задаче о линейно-квадратичном регуляторе, см. (3.1.26).

Существуют методы решения этой задачи, исходящие из описания системы в частотной области (то есть с помощью передаточных функций). Они используют довольно сложный математический аппарат

и непросты с вычислительной точки зрения. Мы же снова воспользуемся техникой линейных матричных неравенств и сведем задачу к полуопределенному программированию.

Преобразуем величину $\|z\|_2^2$:

$$\begin{aligned}\|z\|_2^2 &= \int_0^\infty (Cx + B_2 u)^T (Cx + B_2 u) dt = \\ &= \int_0^\infty (x^T C^T C x + 2u^T B_2^T C x + u^T B_2^T B_2 u) dt.\end{aligned}$$

Далее будем предполагать¹, что $B_2^T C = 0$, тогда

$$\|z\|_2^2 = \int_0^\infty (x^T C^T C x + u^T B_2^T B_2 u) dt = \int_0^\infty x^T (C^T C + K^T B_2^T B_2 K) x dt.$$

На основании теоремы 3.1.8 заключаем, что если матричное неравенство

$$\begin{pmatrix} A_c^T Q + Q A_c + C^T C + K^T B_2^T B_2 K & Q D \\ D^T Q & -\gamma I \end{pmatrix} \preccurlyeq 0, \quad (3.1.40)$$

где $A_c = A + B_1 K$, имеет решение $Q \succ 0$, то $J \leqslant \gamma$.

Умножив полученное неравенство (3.1.40) слева и справа на блочно-диагональную матрицу $\text{diag}\{P \ 0\}$, где $P = Q^{-1}$, имеем

$$\begin{pmatrix} \Omega & D \\ D^T & -\gamma I \end{pmatrix} \preccurlyeq 0,$$

где

$$\Omega = P(A + B_1 K)^T + (A + B_1 K)P + P C^T C P + P K^T B_2^T B_2 K P,$$

или, дважды применяя лемму Шура,

$$\begin{pmatrix} P(A + B_1 K)^T + (A + B_1 K)P & D & P C^T & P K^T B_2^T \\ D^T & -\gamma I & 0 & 0 \\ C P & 0 & -I & 0 \\ B_2 K P & 0 & 0 & -I \end{pmatrix} \preccurlyeq 0.$$

¹Это предположение непринципиально и вводится исключительно для упрощения выкладок. В общем случае можно воспользоваться соответствующей заменой переменных, см. замечание 3.1.7.

После замены $Y = KP$ окончательно имеем

$$\begin{pmatrix} AP + PA^T + B_1 Y + Y^T B_1^T & D & PC^T & Y^T B_2^T \\ D^T & -\gamma I & 0 & 0 \\ CP & 0 & -I & 0 \\ B_2 Y & 0 & 0 & -I \end{pmatrix} \preccurlyeq 0. \quad (3.1.41)$$

Таким образом, если линейное матричное неравенство (3.1.41) выполняется при некоторых $P \succ 0$, Y и γ , то $J \leq \gamma$, а соответствующий стабилизирующий регулятор получается как $K = YP^{-1}$.

Замечание 3.1.8. Следует отметить, что будучи H_∞ -оптимальным, полученный регулятор $K = YP^{-1}$ может обладать некоторыми недостатками. В частности, из-за нестрогости матричного неравенства (3.1.41) он может не оказаться стабилизирующим (замкнутая система будет находиться на границе устойчивости, то есть $\sigma(A+B_1K) = 0$). Чтобы этого избежать, можно предложить следующий простейший способ: в соответствии с замечанием 3.1.4 заменим левый верхний блок в (3.1.41) на

$$AP + PA^T + B_1 Y + Y^T B_1^T + 2\sigma P,$$

где $\sigma > 0$ — желаемая степень устойчивости замкнутой системы. \blacktriangledown

В этом случае приходим к следующему результату.

Теорема 3.1.9. Пусть P_* , Y_* , γ_* — решение задачи SDP

$$\min \gamma$$

при ограничениях

$$\begin{pmatrix} AP + PA^T + B_1 Y + Y^T B_1^T + 2\sigma P & D & PC^T & Y^T B_2^T \\ D^T & -\gamma I & 0 & 0 \\ CP & 0 & -I & 0 \\ B_2 Y & 0 & 0 & -I \end{pmatrix} \preccurlyeq 0,$$

$$P \succ 0$$

относительно матричных переменных $P \in \mathbb{S}^n$, $Y \in \mathbb{R}^{p \times n}$ и скалярной переменной γ при некотором фиксированном значении параметра $\sigma > 0$.

Тогда стабилизирующий регулятор (3.1.39) с матрицей

$$K_* = Y_* P_*^{-1}$$

будет минимизировать функционал $J = \sup_w \|z\|_2^2$ на решениях системы (3.1.38) со степенью устойчивости σ :

$$J_{\min} = \gamma_*,$$

а квадратичная форма $V(x) = x^T P_*^{-1} x$ будет являться функцией Ляпунова для замкнутой системы.

Таким образом, задача вновь сведена к формату SDP.

Пример 3.1.3. Вновь обратимся к двухмассовой системе из примера 3.1.1 и будем полагать, что на левое и правое тело воздействуют компоненты L_2 -ограниченного внешнего возмущения

$$w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix},$$

см. рис. 3.1.4.

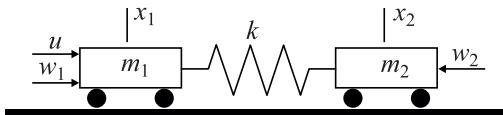


Рис. 3.1.4. Двухмассовая система из примера 3.1.3.

При этом матрицы системы (3.1.38) (при единичных значениях параметров) имеют следующий вид:

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

В качестве регулируемого выхода системы примем вектор

$$z = \begin{pmatrix} v_1 \\ v_2 \\ u \end{pmatrix},$$

так что

$$C = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Задавшись целевой степенью устойчивости $\sigma = 0,1$ замкнутой системы и решая задачу SDP из теоремы 3.1.9, получаем

$$P_* = \begin{pmatrix} 2,3458 & 2,1908 & -0,4298 & -0,4536 \\ 2,1908 & 2,2804 & -0,2048 & -0,2915 \\ -0,4298 & -0,2048 & 0,4778 & 0,0694 \\ -0,4536 & -0,2915 & 0,0694 & 0,2289 \end{pmatrix},$$

$$Y_* = (0 \ 0 \ -1 \ 0).$$

Этой паре соответствует регулятор

$$K_* = (-1,7182 \ 1,4064 \ -0,7433 \ -1,3887) \cdot 10^3,$$

доставляющий функционалу $J = \sup_w \|z\|_2^2$ на решениях замкнутой системы значение 4,5200. При этом замкнутая система имеет степень устойчивости $\sigma(A + B_1 K_*) = 0,1672$. Заметим также, что в соответствии с замечанием 2.7.2 строгое ЛМН $P > 0$ заменено на нестрогое $P \succcurlyeq P_0$ при $P_0 = 0,0001I$. \blacktriangledown

3.2. Робастная стабилизация и управление

В разделе 3.1.4 рассмотрен один из источников неопределенности в описании системы — неизвестные внешние возмущения; в остальном везде выше предполагалось, что доступно точное математическое описание системы. Эта ситуация является идеализированной: в реальных задачах неизбежно присутствует неопределенность в описании системы и используемое управление должно быть работоспособно в этих условиях. Такое управление называется *робастным*.

3.2.1. Структурированная матричная неопределенность

Рассмотрим семейство матриц со структурированной матричной неопределенностью вида

$$A(\Delta) = A + F\Delta H, \quad (3.2.1)$$

где матрица $A \in \mathbb{R}^{n \times n}$ устойчива, $F \in \mathbb{R}^{n \times q_1}$, $H \in \mathbb{R}^{q_2 \times n}$, а возмущение $\Delta \in \mathbb{R}^{q_1 \times q_2}$ (возможно, зависящее от времени) ограничено по норме:

$$\|\Delta\| \leq \gamma.$$

Без ограничения общности можно полагать, что $\gamma = 1$; этого легко добиться соответствующим масштабированием матрицы F или H .

Обратимся к идеям квадратичной устойчивости и покажем, что построение общей функции Ляпунова сводится к решению *линейных* матричных неравенств, значительно упрощая вычисления. Рассмотрим сперва задачу проверки квадратичной устойчивости при $\|\Delta\| \leq 1$.

Как показано ранее, выполнение неравенства Ляпунова

$$(A + F\Delta H)P + P(A + F\Delta H)^T \prec 0,$$

то есть

$$AP + PA^T + F\Delta HP + PH^T\Delta^T F^T \prec 0$$

с некоторой матрицей $P \succ 0$ при всех допустимых неопределенностях означает, что у семейства (3.2.1) есть общая квадратичная функция Ляпунова $V(x) = x^T P^{-1}x$. Теперь остается воспользоваться леммой Питерсена (в форме леммы 2.7.6) с

$$G = AP + PA^T, \quad M = F, \quad N = HP$$

и записать последнее матричное неравенство в виде эквивалентного ему ЛМН относительно скалярной переменной $\varepsilon > 0$ и матричной переменной $P \succ 0$:

$$\begin{pmatrix} AP + PA^T + \varepsilon FF^T & PH^T \\ HP & -\varepsilon I \end{pmatrix} \prec 0. \quad (3.2.2)$$

Если это ЛМН разрешимо, то семейство (3.2.1) робастно квадратично устойчиво, и наоборот.

Окончательно, придав ЛМН (3.2.2) вид

$$\begin{pmatrix} A(\frac{1}{\varepsilon}P) + (\frac{1}{\varepsilon}P)A^T + FF^T & (\frac{1}{\varepsilon}P)H^T \\ H(\frac{1}{\varepsilon}P) & -I \end{pmatrix} \prec 0$$

и переобозначив

$$P \doteq \frac{1}{\varepsilon}P,$$

приходим к следующему утверждению.

Теорема 3.2.1. *Разрешимость линейных матричных неравенств*

$$\begin{pmatrix} AP + PA^T + FF^T & PH^T \\ HP & -I \end{pmatrix} \prec 0, \quad P \succ 0$$

относительно матричной переменной $P \in \mathbb{S}^n$ эквивалентна квадратичной устойчивости семейства (3.2.1) при всех $\|\Delta\| \leq 1$, причем решение \hat{P} определяет общую квадратичную функцию Ляпунова.

Теперь нетрудно вычислить радиус квадратичной устойчивости семейства (3.2.1):

$$\gamma_{\max} \doteq \sup \{ \gamma : (A + F\Delta H)P + P(A + F\Delta H)^T \prec 0 \\ \text{при некотором } P \succ 0 \text{ и всех } \|\Delta\| \leq \gamma \}, \quad (3.2.3)$$

то есть максимальный размах γ_{\max} неопределенности Δ , такой, что при всех $0 < \gamma < \gamma_{\max}$ у семейства имеется общая квадратичная функция Ляпунова.

Итак, пусть $\|\Delta\| \leq \gamma$; потребуем выполнения

$$AP + PA^T + \gamma(F\Delta HP + PH^T\Delta^T F^T) \prec 0$$

при некоторой матрице $P \succ 0$ и всех неопределенностях $\|\Delta\| \leq 1$. Пользуясь леммой Питерсена (в форме леммы 2.7.6) с

$$G = AP + PA^T, \quad M = \gamma F, \quad N = HP$$

и принимая во внимание однородность неравенства (3.2.2), получаем

$$\begin{pmatrix} AP + PA^T + \gamma^2 FF^T & PH^T \\ HP & -I \end{pmatrix} \prec 0, \quad P \succ 0.$$

С учетом замечания 2.7.1 приходим к следующей теореме.

Теорема 3.2.2. *Пусть γ_* — решение задачи SDP*

$$\min(-\gamma)$$

при ограничениях

$$\begin{pmatrix} AP + PA^T + \gamma FF^T & PH^T \\ HP & -I \end{pmatrix} \preceq 0, \quad P \succ 0,$$

относительно матричной переменной $P \in \mathbb{S}^n$ и скалярной переменной γ . Тогда радиус квадратичной устойчивости семейства (3.2.1) равен $\sqrt{\gamma_}$.*

Замечание 3.2.1. Теорема 3.2.2 утверждает, что величина γ_* равна радиусу квадратичной устойчивости (3.2.3). При этом в силу нестрогости ЛМН-ограничения соответствующая матрица P_* , вообще говоря, не определяет общей квадратичной функции Ляпунова. Однако, как следует из определения радиуса квадратичной устойчивости,

для любого $\tilde{\gamma} < \gamma_*$, сколь угодно близкого к γ_* , найдется $P \succ 0$ такое, что выполнено *строгое ЛМН-условие*.

При этом общей квадратичной функции семейства (3.2.1) при всех $\|\Delta\| \leq \tilde{\gamma}$ будет отвечать решение P_* задачи SDP

$$\min(-\varepsilon)$$

при ограничениях

$$\begin{pmatrix} AP + PA^T + \varepsilon I + \tilde{\gamma}^2 FF^T & PH^T \\ HP & -I \end{pmatrix} \preccurlyeq 0, \quad P \succ 0$$

относительно матричной переменной $P \in \mathbb{S}^n$ и скалярной переменной $\varepsilon > 0$. \blacktriangledown

3.2.2. Робастная квадратичная стабилизация

Рассмотрим задачу квадратичной стабилизации в робастном варианте. А именно, для линейной системы управления

$$\dot{x} = (A + F\Delta H)x + Bu \quad (3.2.4)$$

с состоянием $x(t) \in \mathbb{R}^n$ и управлением $u(t) \in \mathbb{R}^p$, где $A \in \mathbb{R}^{n \times n}$, $F \in \mathbb{R}^{n \times q_1}$, $H \in \mathbb{R}^{q_2 \times n}$, пара (A, B) управляема, а матричная неопределенность $\Delta \in \mathbb{R}^{q_1 \times q_2}$ удовлетворяет ограничению

$$\|\Delta\| \leq 1,$$

будем искать общий регулятор вида

$$u = Kx \quad (3.2.5)$$

такой, чтобы стабилизировать систему (3.2.4) при всех допустимых неопределенностях.

Воспользуемся теоремой 3.1.5 из раздела 3.1.1 и будем предполагать существование матриц $P \succ 0$ и Y , удовлетворяющих матричному неравенству

$$(A + F\Delta H)P + P(A + F\Delta H)^T + BY + Y^T B^T \prec 0,$$

или

$$AP + PA^T + BY + Y^T B^T + F\Delta H P + PH^T \Delta^T F^T \prec 0.$$

В силу леммы Питерсена (в форме леммы 2.7.6) с

$$G = AP + PA^T + BY + Y^T B^T, \quad M = F, \quad N = HP$$

полученное матричное неравенство выполняется при всех допустимых значениях матричной неопределенности Δ тогда и только тогда, когда существует положительное число ε такое, что

$$\begin{pmatrix} AP + PA^T + BY + Y^T B^T + \varepsilon FF^T & PH^T \\ HP & -\varepsilon I \end{pmatrix} \prec 0,$$

или

$$\begin{pmatrix} A(\frac{1}{\varepsilon}P) + (\frac{1}{\varepsilon}P)A^T + B(\frac{1}{\varepsilon}Y) + (\frac{1}{\varepsilon}Y)^T B^T + FF^T & (\frac{1}{\varepsilon}P)H^T \\ H(\frac{1}{\varepsilon}P) & -I \end{pmatrix} \prec 0.$$

Переобозначая

$$P \doteq \frac{1}{\varepsilon}P, \quad Y \doteq \frac{1}{\varepsilon}Y,$$

приходим к следующему утверждению.

Теорема 3.2.3. *Пусть матрицы \hat{P} и \hat{Y} удовлетворяют линейным матричным неравенствам*

$$\begin{pmatrix} AP + PA^T + BY + Y^T B^T + FF^T & PH^T \\ HP & -I \end{pmatrix} \prec 0, \quad P \succ 0. \quad (3.2.6)$$

Тогда регулятор (3.2.5) с матрицей

$$\hat{K} = \hat{Y}\hat{P}^{-1}$$

робастно квадратично стабилизирует систему (3.2.4) при всех неопределенностях $\|\Delta\| \leq 1$, а квадратичная форма

$$V(x) = x^T \hat{P}^{-1} x$$

является общей функцией Ляпунова для замкнутой системы (при всех неопределенностях $\|\Delta\| \leq 1$).

Теорема 3.2.3 дает необходимые и достаточные условия *робастной квадратичной стабилизируемости* неопределенной системы.

Упражнение 3.2.1. *Используя замечание 3.1.4 покажите, что в случае заданной степени устойчивости $\sigma > 0$ замкнутой системы первое из условий (3.2.6) теоремы 3.2.3 заменяется на нестрогое матричное неравенство*

$$\begin{pmatrix} AP + PA^T + BY + Y^T B^T + 2\sigma P + FF^T & PH^T \\ HP & -I \end{pmatrix} \preccurlyeq 0.$$

Можно рассмотреть также и случай, когда структурированная неопределенность содержится и в матрице B системы:

$$\dot{x} = (A + F_A \Delta_A H_A)x + (B + F_B \Delta_B H_B)u,$$

где $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $F_A \in \mathbb{R}^{n \times q_1^A}$, $H_A \in \mathbb{R}^{q_2^A \times n}$, $F_B \in \mathbb{R}^{n \times q_1^B}$, $H_B \in \mathbb{R}^{q_2^B \times m}$, $\Delta_A \in \mathbb{R}^{q_1^A \times q_2^A}$, $\Delta_B \in \mathbb{R}^{q_1^B \times q_2^B}$, причем

$$\|\Delta_A\| \leq 1, \quad \|\Delta_B\| \leq 1.$$

Однако при этом соответствующий аналог теоремы 3.2.3 предоставит лишь *достаточные* условия робастной квадратичной стабилизируемости, поскольку лемма Питерсена для нескольких неопределенностей справедлива только в части достаточности. Необходимость имеет место лишь если A и B подвержены воздействию одной и той же возмущающей матрицы: $\Delta_A = \Delta_B$ (при этом, разумеется, возможно $F_A \neq F_B$, $H_A \neq H_B$).

Закончим этот раздел изучением еще одной характеристики системы (3.2.4) — *радиуса квадратичной стабилизируемости*, который определим как величину максимального размаха неопределенности, при котором еще возможна робастная квадратичная стабилизация:

$$\gamma_{\max}^{\text{st}} = \sup \left\{ \gamma : (A + BK + F\Delta H)P + P(A + BK + F\Delta H)^T \prec 0 \right. \\ \left. \text{при некотором } P \succ 0, \text{ регуляторе } K \text{ и всех } \|\Delta\| \leq \gamma \right\}.$$

Ответ дается следующим утверждением.

Теорема 3.2.4. *Пусть γ_* — решение задачи SDP*

$$\min(-\gamma)$$

при ограничениях

$$\begin{pmatrix} AP + PA^T + BY + Y^T B^T + \gamma FF^T & PH^T \\ HP & -I \end{pmatrix} \preccurlyeq 0, \quad P \succ 0$$

относительно матричных переменных $P \in \mathbb{S}^n$, $Y \in \mathbb{R}^{p \times n}$ и скалярной переменной γ . Тогда радиус квадратичной стабилизируемости системы (3.2.4) равен $\sqrt{\gamma_}$.*

Упражнение 3.2.2. *Дайте обоснование теореме 3.2.4, используя утверждения теорем 3.2.2 и 3.2.3.*

Сделаем важное замечание.

Замечание 3.2.2. В системе

$$\dot{x} = (A + F\Delta H)x + Bu$$

радиус квадратичной стабилизируемости, вообще говоря, может и не быть конечным. В частности, это заведомо имеет место при $F = B$. \blacktriangledown

В нижеследующем примере радиус квадратичной стабилизируемости системы также бесконечен.

Пример 3.2.1. Рассмотрим систему

$$\dot{x} = (A + F\Delta H)x + Bu,$$

где

$$A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$$F = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad H = \begin{pmatrix} 0 & 1 \end{pmatrix}, \quad \Delta \in \mathbb{R}.$$

Будем искать матрицу усиления в виде

$$K = \begin{pmatrix} 0 & k \end{pmatrix};$$

тогда матрица замкнутой системы приобретает вид

$$A_c(\Delta) = A + F\Delta H + BK = \begin{pmatrix} -1 & \Delta \\ 0 & k+1 \end{pmatrix}.$$

Очевидно, что она устойчива тогда и только тогда, когда $k < -1$, причем устойчивость сохраняется *при всех* $\Delta \in \mathbb{R}$.

Теперь зафиксируем некоторый уровень неопределенности $\gamma > 0$ и найдем общую квадратичную функцию Ляпунова для семейства устойчивых систем $\dot{x} = A_c(\Delta)x$ при всех $|\Delta| < \gamma$. Будем искать матрицу функции Ляпунова в виде

$$P = \begin{pmatrix} p_1 & 0 \\ 0 & p_2 \end{pmatrix} \succ 0.$$

Представим $k = -1 - \varepsilon$, $\varepsilon > 0$; тогда

$$A_c(\Delta)P + PA_c^T(\Delta) = \begin{pmatrix} -2p_1 & \Delta p_2 \\ \Delta p_2 & -2\varepsilon p_2 \end{pmatrix}.$$

Для отрицательной определенности этой матрицы получаем (например, из леммы Шура) следующее условие:

$$p_1 > \frac{\Delta^2}{4\varepsilon} p_2,$$

которое выполняется для всех допустимых $|\Delta| < \gamma$ при выборе

$$p_1 > \frac{\gamma^2}{4\varepsilon} p_2,$$

например, при $p_2 = \varepsilon$, $p_1 = \gamma^2$.

Итого, каков бы ни был размах неопределенности γ , робастно стабилизирующий регулятор может быть выбран в виде $K = (0 \ -\varepsilon)$, $\varepsilon > 0$, а в качестве матрицы общей квадратичной функции Ляпунова взята матрица $P = \text{diag}\{\gamma^2 \ \varepsilon\}$. ▼

Идеи и технические средства, использовавшиеся при построении (робастно) квадратично стабилизирующих регуляторов самым существенным образом будут применяться в дальнейшем изложении. Основными среди таких средств являются замена переменных, введенная в теореме 3.1.5, и использование леммы Питерсена для учета структурированной неопределенности при робастном синтезе.

Далее будут рассмотрены робастные варианты линейно-квадратичной задачи и задачи H_∞ -оптимизации. Еще раз напомним, что мы ограничиваемся построением квадратично стабилизирующих регуляторов; эффективная техника решения таких задач, основанная на линейных матричных неравенствах, была обсуждена в разделе 3.1. Соответственно, говоря о минимизации того или иного критерия, подразумеваем, что она производится в классе робастно *квадратично* стабилизирующих регуляторов.

3.2.3. Робастный линейно-квадратичный регулятор

Построим оптимальный линейно-квадратичный регулятор в ситуации, когда матрица системы содержит структурированную матричную неопределенность. Рассмотрим линейную систему управления

$$\dot{x} = (A + F\Delta H)x + Bu, \quad x(0) = x_0, \quad (3.2.7)$$

с состоянием $x(t) \in \mathbb{R}^n$ и управлением $u(t) \in \mathbb{R}^p$, где $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $F \in \mathbb{R}^{n \times q_1}$, $H \in \mathbb{R}^{q_2 \times n}$, а матричная неопределенность

$\Delta \in \mathbb{R}^{q_1 \times q_2}$ удовлетворяет ограничению

$$\|\Delta\| \leq 1. \quad (3.2.8)$$

Будем искать закон управления в форме линейной обратной связи по состоянию

$$u = Kx, \quad K \in \mathbb{R}^{p \times n}, \quad (3.2.9)$$

минимизируя квадратичный критерий качества

$$J = \max_{\|\Delta\| \leq 1} \int_0^\infty (x^T Rx + u^T Su) dt, \quad R, S \succ 0, \quad (3.2.10)$$

по всем квадратично стабилизирующими регуляторам.

Иными словами, встанем на позиции минимакса и будем оптимизировать качество замкнутой системы против «наихудшей» неопределенности.

Согласно теореме 3.1.7, будем предполагать существование матриц $P \succ 0$ и Y , удовлетворяющих матричному неравенству

$$\begin{pmatrix} (A + F\Delta H)P + P(A + F\Delta H)^T + BY + Y^T B^T & P & Y^T \\ P & -R^{-1} & 0 \\ Y & 0 & -S^{-1} \end{pmatrix} \preceq 0,$$

которое представимо в виде

$$\begin{aligned} & \begin{pmatrix} AP + PA^T + BY + Y^T B^T & P & Y^T \\ P & -R^{-1} & 0 \\ Y & 0 & -S^{-1} \end{pmatrix} + \\ & + \begin{pmatrix} F \\ 0 \\ 0 \end{pmatrix} \Delta \begin{pmatrix} HP & 0 & 0 \end{pmatrix} + \begin{pmatrix} PH^T \\ 0 \\ 0 \end{pmatrix} \Delta^T \begin{pmatrix} F^T & 0 & 0 \end{pmatrix} \preceq 0. \end{aligned}$$

Вновь воспользовавшись леммой Питерсена заключаем, что полученное матричное неравенство выполняется при всех допустимых значениях матричной неопределенности Δ тогда и только тогда, когда существует положительное число ε такое, что

$$\begin{pmatrix} AP + PA^T + BY + Y^T B^T + \varepsilon FF^T & P & Y^T & PH^T \\ P & -R^{-1} & 0 & 0 \\ Y & 0 & -S^{-1} & 0 \\ HP & 0 & 0 & -\varepsilon I \end{pmatrix} \preceq 0. \quad (3.2.11)$$

Далее, поскольку весовая матрица R положительно определена, в силу следствия 2.7.1 имеем

$$\begin{pmatrix} AP + PA^T + BY + Y^T B^T + \varepsilon FF^T & PH^T \\ HP & -\varepsilon I \end{pmatrix} \prec 0$$

для всех P и Y , удовлетворяющих ЛМН (3.2.11). Это означает, что соответствующий регулятор $K = YP^{-1}$ будет робастно стабилизировать систему (см. раздел 3.2.2).

Наконец, минимизируя на решениях полученного линейного матричного неравенства величину $x_0^T P^{-1} x_0$ (переписав ее в виде линейной функции, так же как и в неробастном случае), приходим к следующему утверждению.

Теорема 3.2.5. Пусть P_* , Y_* — решение задачи SDP

$$\min \gamma$$

при ограничениях

$$\begin{pmatrix} AP + PA^T + BY + Y^T B^T + \varepsilon FF^T & P & Y^T & PH^T \\ P & -R^{-1} & 0 & 0 \\ Y & 0 & -S^{-1} & 0 \\ HP & 0 & 0 & -\varepsilon I \end{pmatrix} \preccurlyeq 0,$$

$$\begin{pmatrix} \gamma & x_0^T \\ x_0 & P \end{pmatrix} \succcurlyeq 0, \quad P \succ 0$$

относительно матричных переменных $P \in \mathbb{S}^n$, $Y \in \mathbb{R}^{p \times n}$ и скалярных переменных γ и $\varepsilon > 0$.

Тогда регулятор (3.2.9) с матрицей

$$K_* = Y_* P_*^{-1}$$

будет робастно квадратично стабилизировать систему (3.2.7). При этом квадратичная форма

$$V(x) = x^T P_*^{-1} x$$

является общей функцией Ляпунова для замкнутой системы при всех неопределеностях (3.2.8), а минимальное значение функционала (3.2.10) на решениях системы (3.2.7) с начальным условием x_0 равно

$$J_* = x_0^T P_*^{-1} x_0.$$

Отметим, что при достаточно высоком уровне неопределенности $\|\Delta\|$ решение данной задачи отсутствует. Впрочем, в теореме 3.2.4 был представлен простой способ отыскания максимального уровня $\gamma_{\max}^{\text{st}}$, допускающего квадратичную стабилизацию, так что в предыдущей теореме молчаливо предполагается, что этот уровень не превышен.

3.2.4. H_∞ -оптимизация: робастный вариант

Обратимся теперь к робастной постановке задачи H_∞ -оптимизации и рассмотрим систему

$$\begin{aligned}\dot{x} &= (A + F\Delta H)x + B_1u + Dw, \quad x(0) = 0, \\ z &= Cx + B_2u,\end{aligned}\tag{3.2.12}$$

где $A \in \mathbb{R}^{n \times n}$, $B_1 \in \mathbb{R}^{n \times p}$, $D \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{\ell \times n}$, $B_2 \in \mathbb{R}^{\ell \times p}$, с состоянием $x(t) \in \mathbb{R}^n$, управлением $u(t) \in \mathbb{R}^p$, регулируемым выходом $z(t) \in \mathbb{R}^\ell$ и ограниченным в L_2 -норме внешним возмущением $w(t) \in \mathbb{R}^m$.

Как и ранее, полагаем, что неопределенность в системе имеет структурированную матричную форму: $F \in \mathbb{R}^{n \times q_1}$ и $H \in \mathbb{R}^{q_2 \times n}$ — постоянные матрицы, а матричная неопределенность $\Delta \in \mathbb{R}^{q_1 \times q_2}$ удовлетворяет ограничению

$$\|\Delta\| \leq 1.\tag{3.2.13}$$

Будем искать регулятор в форме статической линейной обратной связи по состоянию

$$u = Kx,\tag{3.2.14}$$

который минимизирует величину

$$J = \max_{\|\Delta\| \leq 1} \sup_{\|w\| \leq 1} \|z\|_2^2$$

на решениях системы (3.2.12) по всем квадратично стабилизирующими регуляторам.

Согласно теореме 3.1.9, будем предполагать существование матриц $P \succ 0$, Y и скалярной переменной γ , удовлетворяющих матричному неравенству

$$\begin{pmatrix} \Phi & D & PC^T & Y^T B_2^T \\ D^T & -\gamma I & 0 & 0 \\ CP & 0 & -I & 0 \\ B_2 Y & 0 & 0 & -I \end{pmatrix} \preccurlyeq 0,$$

где

$$\Phi = (A + F\Delta H)P + P(A + F\Delta H)^T + B_1Y + Y^T B_1^T.$$

Представив его в виде

$$\begin{aligned} & \begin{pmatrix} AP + PA^T + B_1Y + Y^T B_1^T & D & PC^T & Y^T B_2^T \\ D^T & -\gamma I & 0 & 0 \\ CP & 0 & -I & 0 \\ B_2Y & 0 & 0 & -I \end{pmatrix} + \\ & + \begin{pmatrix} F \\ 0 \\ 0 \\ 0 \end{pmatrix} \Delta (HP \quad 0 \quad 0 \quad 0) + \begin{pmatrix} PH^T \\ 0 \\ 0 \\ 0 \end{pmatrix} \Delta^T (F^T \quad 0 \quad 0 \quad 0) \preceq 0 \end{aligned}$$

и воспользовавшись леммой Питерсена, заключаем, что полученное матричное неравенство выполняется при всех допустимых значениях матричной неопределенности Δ тогда и только тогда, когда существует положительное число ε такое, что

$$\begin{pmatrix} \Omega & D & PC^T & Y^T B_2^T & PH^T \\ D^T & -\gamma I & 0 & 0 & 0 \\ CP & 0 & -I & 0 & 0 \\ B_2Y & 0 & 0 & -I & 0 \\ HP & 0 & 0 & 0 & -\varepsilon I \end{pmatrix} \preceq 0, \quad (3.2.15)$$

где

$$\Omega = AP + PA^T + B_1Y + Y^T B_1^T + \varepsilon FF^T.$$

Как и в неробастном случае, из-за нестрогости матричного неравенства (3.2.15) регулятор $K = YP^{-1}$ может оказаться не стабилизирующим. При этом, действуя согласно сказанному в замечании 3.1.8, приходим к следующему результату.

Теорема 3.2.6. *Пусть P_* , Y_* , γ_* – решение задачи SDP*

$$\min \gamma$$

при ограничениях

$$\begin{pmatrix} \Omega & D & PC^T & Y^T B_2^T & PH^T \\ D^T & -\gamma I & 0 & 0 & 0 \\ CP & 0 & -I & 0 & 0 \\ B_2Y & 0 & 0 & -I & 0 \\ HP & 0 & 0 & 0 & -\varepsilon I \end{pmatrix} \preceq 0, \quad P \succ 0,$$

где

$$\Omega = AP + PA^T + B_1Y + Y^T B_1^T + 2\sigma P + \varepsilon FF^T,$$

относительно матричных переменных $P \in \mathbb{S}^n$, $Y \in \mathbb{R}^{p \times n}$ и скалярных переменных γ и $\varepsilon > 0$ при некотором фиксированном значении параметра $\sigma > 0$.

Тогда регулятор (3.2.14) с матрицей

$$K_* = Y_* P_*^{-1}$$

робастно квадратично стабилизирует систему (3.2.12) со степенью устойчивости σ . При этом минимальное значение функционала J на ее решениях равно

$$J_* = \gamma_*,$$

а квадратичная форма

$$V(x) = x^T P_*^{-1} x$$

является общей функцией Ляпунова для замкнутой системы при всех неопределенностях (3.2.13).

3.3. Линейно-квадратичное управление по выходу

3.3.1. Подходы к решению

Задача о линейно-квадратичном регуляторе (*linear quadratic regulator, LQR*) активно исследовалась в прошлом веке, начиная с основополагающих работ Р. Калмана [96, 97] 1960-х годов. Оказалось, что на бесконечном горизонте времени оптимальное (по квадратичному критерию) управление в линейной стационарной системе имеет вид линейной статической обратной связи *по состоянию*, при этом матрица обратной связи находится как решение алгебраического матричного уравнения Риккати. Этот классический результат давно вошел в учебники по управлению [29, 32, 40]. Позже появились новые подходы к этой проблеме, основанные на технике полуопределенного программирования и предполагающие ее сведение к задаче выпуклой оптимизации с ограничениями в виде линейных матричных неравенств [2, 37, 67, 95]; соответствующие результаты были освещены в разделе 3.1.3. Линейная статическая обратная связь — это очень естественное и простое с инженерной точки зрения управление, поэтому вполне понятно желание распространить эту технику на иные задачи управления.

Ближайшая к LQR-задаче — это задача синтеза линейно-квадратичного управления *по выходу*. А именно, рассмотрим линейную систему управления с интегральным квадратичным критерием, но уже для случая, когда измерению доступно не все состояние системы, а лишь ее линейный выход. Попытки построения статической обратной связи по выходу сопряжены с существенными трудностями. Впервые эту проблему стали рассматривать У. Левин и М. Атанс [104], которые показали, что такое стабилизирующее управление может не существовать, а если оно и существует, то отсутствуют простые критерии проверки его оптимальности. Большие усилия были направлены на нахождение условий его существования, см., например, [79, 129], но эта проблема так и остается открытой. Стабилизуемость системы при помощи статической обратной связи по выходу является лишь необходимым условием ее оптимальности, и даже оно формулируется в виде системы нелинейных матричных уравнений [104].

Таким образом, синтез обратной связи по выходу требует применения численных методов. Первый из них был предложен в [104] и предполагал на каждой итерации решение системы нелинейных матричных уравнений. Метод, предложенный Б. Андерсоном и Дж. Муром [29], основан на решении линейных матричных уравнений, но его свойства не очевидны; некоторые результаты, связанные со сходимостью обоих этих подходов, содержатся в [111]. С тех пор был предложен целый ряд итеративных алгоритмов, см., например, [85, 95, 111, 112, 115, 126, 131] и ссылки в них, однако для многих из них отсутствует строгое обоснование, а другие предполагают решение на каждой итерации сложных нелинейных задач. Таким образом, задача синтеза линейно-квадратичного управления по выходу так и остается сложной и нерешенной проблемой.

Вместе с тем многообещающим инструментом синтеза обратной связи в линейно-квадратичной задаче представляется градиентный метод, см. раздел 1.3.1. При этом матрица регулятора K для управления по состоянию ($u(t) = Kx(t)$) или по полностью наблюдаемому выходу ($u(t) = Ky(t)$) рассматривается как переменная при оптимизации интегрального критерия, который может быть представлен как $f(K)$. Эта функция определена на множестве \mathcal{S} стабилизирующих обратных связей (в этом и только в этом случае величина $f(K)$ будет конечной). Множество \mathcal{S} открыто и минимум $f(K)$ достигается в его внутренней точке, поэтому можно воспользоваться градиентным методом для безусловной минимизации функции $f(K)$:

$$K_{j+1} = K_j - \gamma_j f'(K_j),$$

предполагая известным некоторый начальный стабилизирующий регулятор K_0 . Градиент $f'(K)$ функции $f(K)$ для управления по состоянию был выписан в пионерской работе Р. Калмана [96], а для управления по выходу получен У. Левиным и М. Атансом [104]. Его вычисление несложно и предполагает решение всего лишь двух уравнений Ляпунова.

Такой подход выглядит очень привлекательным, однако он сопряжен с определенными трудностями. Для управления по состоянию множество \mathcal{S} связно¹, однако (в общем случае) невыпукло [71], а $f(K)$ может быть невыпуклой функцией. Еще более сложная ситуация имеет место при управлении по выходу: множество \mathcal{S} может быть даже несвязным [53, 88]. Ввиду этого во многих работах градиентный метод применялся без строгого обоснования как чисто эвристический алгоритм.

Однако сравнительно недавно был получен ряд прорывных результатов, связанных с синтезом обратной связи по состоянию для линейно-квадратичной задачи. В статье [87], посвященной ее дискретной версии, было показано, что функция $f(K)$, несмотря на невыпуклость, удовлетворяет условию Поляка–Лоясевича (см. раздел 1.3.2), на основе которого удалось доказать глобальную сходимость градиентного метода к оптимальному регулятору. Далее в работах [113, 114] были получены аналогичные результаты и для линейно-квадратичной задачи в *непрерывном* времени.

С управлением по выходу ситуация гораздо сложнее. Как уже упоминалось, область \mathcal{S} может быть несвязной, а седловые точки и локальные минимумы функции могут быть разбросаны по разным компонентам связности; более того, одной компоненте могут принадлежать несколько локальных минимумов. Поэтому здесь трудно ожидать чего-то большего, чем сходимость к стационарной точке.

Заметим, что большинство результатов, связанных со сходимостью градиентного метода для синтеза обратной связи по состоянию в линейно-квадратичной задаче, относятся к системам в дискретном времени. Мы же сосредоточимся на непрерывном случае и докажем сходимость метода к единственной точке минимума с линейной скоростью. Близкие результаты были получены в [113, 114], но техника доказательства там совсем другая. Так, в [114] специальной заменой переменных соответствующая задача была сведена к задаче выпуклой оп-

¹Здесь и далее под связностью будем понимать линейную связность. Напомним, что множество \mathcal{Q} называется *линейно связным*, если любую пару его точек можно соединить непрерывной кривой, принадлежащей \mathcal{Q} .

тимизации, однако это возможно только для управления по состоянию; мы же предложим технику, подходящую как для случая управления по состоянию, так и по выходу. Далее будут получены результаты, связанные со сходимостью к стационарной точке для обратной связи по выходу, которые основаны на так называемом свойстве *L-гладкости*.

3.3.2. Постановка задачи

Рассмотрим линейную систему

$$\begin{aligned}\dot{x} &= Ax + Bu, \\ y &= Cx\end{aligned}\tag{3.3.1}$$

с состоянием $x(t) \in \mathbb{R}^n$, управлением $u(t) \in \mathbb{R}^p$, выходом $y(t) \in \mathbb{R}^\ell$ и известными матрицами $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $C \in \mathbb{R}^{\ell \times n}$.

Сопоставим ей квадратичный критерий

$$\mathbb{E}_{\sim x(0)} \int_0^\infty (x^T(t)Qx(t) + u^T(t)Ru(t))dt,\tag{3.3.2}$$

где математическое ожидание берется по распределению начальных условий $x(0)$ с нулевым средним и ковариационной матрицей Σ , то есть

$$\mathbb{E} x(0)x^T(0) = \Sigma,$$

а $0 \prec Q \in \mathbb{S}^n$ и $0 \prec R \in \mathbb{S}^p$ — заданные весовые матрицы.

Замкнув систему (3.3.1) статической линейной обратной связью по выходу

$$u = -Ky, \quad K \in \mathbb{R}^{p \times \ell},$$

приходим к замкнутой системе

$$\dot{x} = A_K x, \quad A_K = A - BKC,$$

а минимизируемая функция примет вид

$$f(K) = \mathbb{E}_{\sim x(0)} \int_0^\infty (x^T(t)(Q + C^T K^T R K) x(t))dt.$$

Мы используем запись $f(K)$, чтобы подчеркнуть, что значение минимизируемой функции зависит только от выбора регулятора K .

Таким образом, исходная задача свелась к следующей задаче оптимизации:

$$\inf_{K \in \mathcal{S}} f(K),\tag{3.3.3}$$

где \mathcal{S} — множество стабилизирующих регуляторов

$$\mathcal{S} = \{K \in \mathbb{R}^{p \times \ell}: \operatorname{Re} \lambda_i(A - BKC) < 0 \text{ для всех } i = 1, \dots, n\}.$$

Напомним, что функция $f(K)$ определена только на множестве \mathcal{S} стабилизирующих регуляторов.

Вопрос о существовании стабилизирующих регуляторов по выходу очень сложен, см. подробнее в [79, 129]. Однако в рамках рассматриваемого подхода этот вопрос не является значимым: некоторый стабилизирующий регулятор $K_0 \in \mathcal{S}$ предполагается заранее известным. В частности, при гурвицовой матрице A можно положить $K_0 = 0$. Регулятор K_0 будет служить начальным приближением для итеративного метода, который будет далее построен. В этом смысле наша задача состоит в том, чтобы улучшить известный регулятор K_0 по отношению к критерию качества $f(K)$.

Итак, будем предполагать выполненные следующие предположения.

Предположение 3.3.1. Существует регулятор $K_0 \in \mathcal{S}$.

Предположение 3.3.2. $\operatorname{rank} C = \ell$.

Заметим, что не накладываются требования ни управляемости, ни наблюдаемости, а лишь предполагается существование некоторого стабилизирующего регулятора.

Введем в рассмотрение *множество уровня*

$$\mathcal{S}_0 = \{K \in \mathcal{S}: f(K) \leq f(K_0)\}.$$

Теперь переформулируем исходную задачу как задачу оптимизации. При этом, чтобы обойтись без вычисления интеграла в (3.3.2), воспользуемся леммой 3.1.3 (леммой Беллмана). Это позволяет записать задачу (3.3.3) в окончательной форме:

$$\inf_K f(K), \quad f(K) = \operatorname{tr} X(K)\Sigma,$$

при ограничении

$$(A - BKC)^T X + X(A - BKC) + C^T K^T R K C + Q = 0$$

относительно матричных переменных $K \in \mathbb{R}^{p \times \ell}$ и $X \in \mathbb{S}^n$.

Это задача матричной оптимизации с нелинейным ограничением типа равенства. Запись $f(K)$ подчеркивает, что при заданном K матрица X находится из уравнения Ляпунова (3.3.2); тем самым независимой переменной является K .

Заметим, что положительно определенное решение $X \succ 0$ матричного уравнения Ляпунова (3.3.2) существует при любом $K \in \mathcal{S}$. В дальнейшем для упрощения записи иногда будем опускать зависимость X от K .

3.3.3. Примеры

Переходя к рассмотрению свойств функции $f(K)$, ее области определения \mathcal{S} и множества уровня \mathcal{S}_0 , начнем с нескольких простых примеров, чтобы показать все разнообразие возможных ситуаций.

Пример 3.3.1. Рассмотрим одномерную систему вида (3.3.1) с параметрами

$$A = 0, \quad B = \frac{1}{2}, \quad Q = R = \Sigma = C = 1, \quad K = k \in \mathbb{R}.$$

В этом случае функция $f(k) = k + \frac{1}{k}$ может быть выписана явным образом. При этом множество $\mathcal{S} = \mathbb{R}_+$ выпукло и неограничено, множество \mathcal{S}_0 ограничено, а функция $f(k)$ выпукла и неограничена на \mathcal{S} , см. рис. 3.3.1.

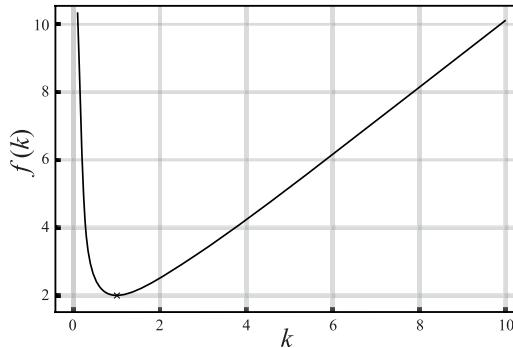


Рис. 3.3.1. Функция $f(k)$ в примере 3.3.1.



Пример 3.3.2. Пусть $n = p = 2$ и $A = B = C = I$. Тогда

$$\mathcal{S} = \left\{ K \in \mathbb{R}^{2 \times 2} : k_{11} + k_{22} < 1 + k_{11}k_{22} + k_{12}k_{21}, \quad k_{11} + k_{22} < 2 \right\}.$$

Множество \mathcal{S} невыпукло; это легко проверить, рассмотрев сечение

$$x = k_{11} = k_{12}, \quad y = k_{22} = k_{21},$$

см. рис. 3.3.2. Более того, граница множества \mathcal{S} негладкая.

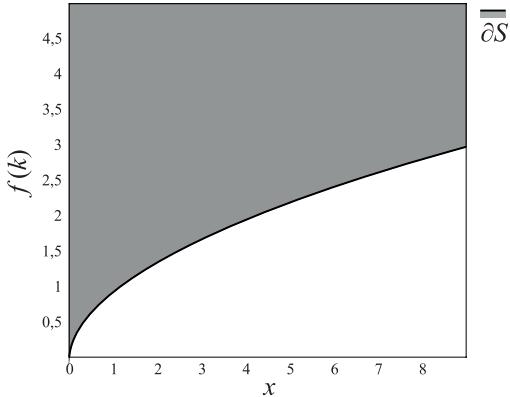


Рис. 3.3.2. Невыпуклость множества сечения множества \mathcal{S} в примере 3.3.2.

▼

Пример 3.3.3. Пусть $n = 3$, $p = 1$; рассмотрим систему (3.3.1) с матрицами

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad C = I.$$

При этом

$$\mathcal{S} = \{K \in \mathbb{R}^{1 \times 3} : k_1 > 0, k_2 k_3 > k_1\}.$$

Вновь множество \mathcal{S} невыпукло с негладкой границей. В частности, его сечение

$$x = k_1, \quad y = k_2 = k_3$$

невыпукло, см. рис. 3.3.3.

▼

Три рассмотренных примера относились к случаю управления по состоянию; теперь обратимся к задаче управления по выходу.

Пример 3.3.4. Рассмотрим систему (3.3.1) с матрицами

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -\alpha \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

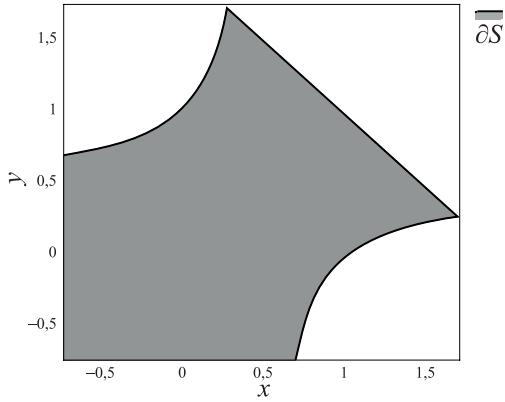


Рис. 3.3.3. Невыпуклость сечения множества \mathcal{S} в примере 3.3.3.

$$C = \begin{pmatrix} 5 & 2 & 1 \end{pmatrix}, \quad Q = \Sigma = I, \quad R = 1.$$

В этом случае

$$\mathcal{S} = \{k \in \mathbb{R}: k + \alpha > 0, (k + \alpha)(2k + 1) > 5k + 1 > 0\}.$$

При $\alpha = -1$ это множество несвязно и состоит из двух компонент. График функции $f(k)$ показан на рис. 3.3.4, он имеет по одному локальному минимуму на каждой из компонент связности.

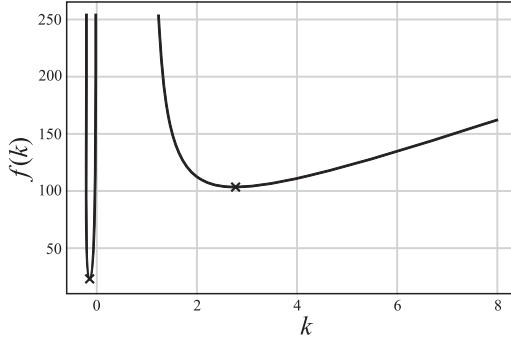


Рис. 3.3.4. График функции $f(k)$ при $\alpha = -1$ в примере 3.3.4.

При $\alpha = -1,4$ множество \mathcal{S} связано, оно представляет собой луч $k > -0,2$. График функции $f(k)$ показан на рис. 3.3.5; он имеет два

локальных минимума в единственной компоненте связности.

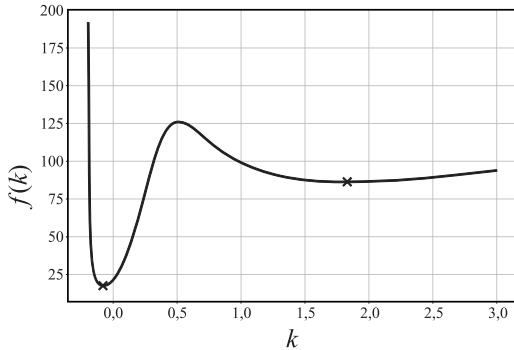


Рис. 3.3.5. График функции $f(k)$ при $\alpha = -1,4$ в примере 3.3.4.



Пример 3.3.5. Рассмотрим систему с матрицами

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -\alpha \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

$$C = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}, \quad Q = \Sigma = I, \quad R = 1.$$

Множество

$$\mathcal{S} = \left\{ K \in \mathbb{R}^{1 \times 2} : \begin{array}{l} \alpha + k_2 > 0, \quad 1 + k_1 + k_2 > 0, \\ (1 + k_2)(\alpha + k_1 - k_2) > 1 + k_1 + k_2 \end{array} \right\}$$

при $\alpha = 1,2$ связно и имеет два локальных минимума и седловую точку $K = (1,95 \quad 0,38)$, см. рис. 3.3.6.

При $\alpha = 0,9$ появляются уже две компоненты связности с единственным локальным минимумом в каждой из них, см. рис. 3.3.7.



Подведем итоги: область определения \mathcal{S} функции $f(K)$ может быть невыпуклой с негладкой границей даже при управлении по состоянию и несвязной при обратной связи по выходу. Функция $f(K)$ неограничена на этом множестве и — в рассмотренных примерах — выглядит гладкой; ниже будет показана справедливость этих свойств.

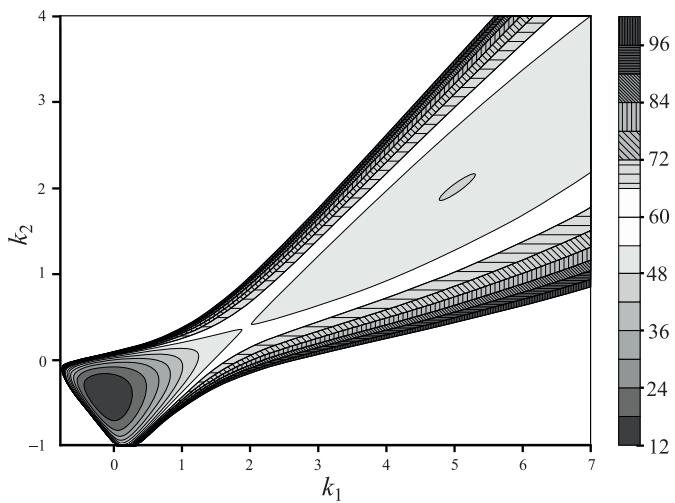


Рис. 3.3.6. Локальные минимумы $f(K)$ при $\alpha = 1,2$ в примере 3.3.5.

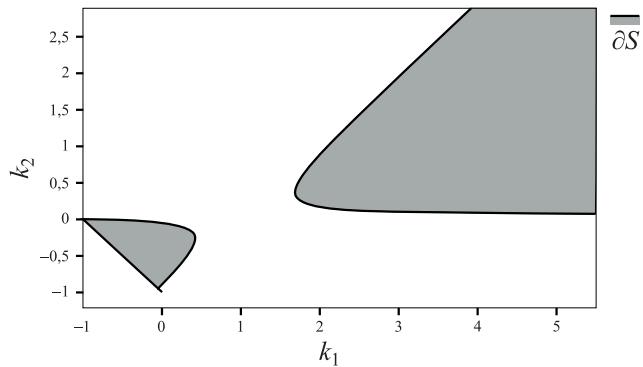


Рис. 3.3.7. Две компоненты связности множества S при $\alpha = 0,9$ в примере 3.3.5.

3.3.4. Свойства минимизируемой функции

Оказывается, что множество S для управления по состоянию является связным, см. [114], и это же верно и для множества S_0 . В самом деле, имеет место следующий результат.

Лемма 3.3.1. *Пусть $C = I$. Тогда множества \mathcal{S} и \mathcal{S}_0 связны для любого $K_0 \in \mathcal{S}$.*

Доказательство. При $C = I$ уравнение (3.3.2) принимает вид

$$(A - BK)^T X + X(A - BK) + K^T R K + Q = 0.$$

Как показано в [67], это уравнение может быть заменено неравенством

$$(A - BK)^T X + X(A - BK) + K^T R K + Q \leq 0,$$

которое после его домножения слева и справа на $P = X^{-1}$ принимает вид

$$AP + PA^T - BKP - PK^T B^T + PK^T RKP + PQP \leq 0.$$

При этом множество стабилизирующих регуляторов может быть представлено в виде

$$\mathcal{S} = \{K = R^{-1}B^T P^{-1}: AP + PA^T - BR^{-1}B^T + PQP \leq 0, P \succ 0\},$$

где первое из неравенств, определяющих величину P , может быть записано в виде эквивалентного ЛМН

$$\begin{pmatrix} AP + PA^T - BR^{-1}B^T & P \\ P & -Q^{-1} \end{pmatrix} \leq 0.$$

Следовательно, они определяют выпуклое множество; при этом его образ, задаваемый непрерывным отображением $K = R^{-1}B^T P^{-1}$, связан.

Множество \mathcal{S}_0 определяется тем же самым отображением для того же множества матриц P с дополнительным ограничением

$$\text{tr } P^{-1}\Sigma \leq f(K_0). \quad (3.3.4)$$

Это ограничение также является выпуклым: оно может быть представлено в виде совокупности линейных неравенств

$$\begin{pmatrix} H & I \\ I & P \end{pmatrix} \succ 0, \quad \text{tr } H\Sigma \leq f(K_0) \quad (3.3.5)$$

относительно P и вспомогательной матричной переменной H , откуда и вытекает связность множества \mathcal{S}_0 . ■

Упражнение 3.3.1. Используя результаты, приведенные в приложении 2.7.3, покажите, что неравенство (3.3.4) можно представить в эквивалентном виде (3.3.5).

Приведенное доказательство использует хорошо известную технику [37] матричных замен переменных, которые позволяют свести исходную проблему к выпуклой; она была использована в разделах 3.1.2, 3.1.3 и др. Эта линия исследований была развита в работах [113, 114] для обоснования использования градиентного метода. К сожалению, этот подход неприменим для задачи управления по выходу: ее принципиально нельзя свести к выпуклой путем замен переменных.

Рассмотренные выше примеры показывают, что функция $f(K)$ не ограничена на своей области определения. Исследуем ее поведение более подробно.

Определение 3.3.1. Непрерывную функцию $f: \mathcal{S} \ni K \rightarrow \mathbb{R}$, определенную на множестве \mathcal{S} , будем называть *коэрцитивной*¹, если для любой последовательности $\{K_j\} \in \mathcal{S}$ выполнены условия:

- 1) $f(K_j) \rightarrow +\infty$ при $\|K_j\| \rightarrow +\infty$;
- 2) $f(K_j) \rightarrow +\infty$ при $K_j \rightarrow K \in \partial\mathcal{S}$.

Лемма 3.3.2. Функция $f(K) = \text{tr } X \Sigma$ коэрцитивна, причем верны следующие оценки:

$$f(K) \geq \frac{\lambda_{\min}(\Sigma)\lambda_{\min}(Q)}{-2 \operatorname{Re} \lambda_{\max}(A_K)}, \quad (3.3.6)$$

$$f(K) \geq \frac{\lambda_{\min}(\Sigma)\lambda_{\min}(R)\lambda_{\min}(CC^T)}{2\|A\| + 2\|K\|_F\|B\|\|C\|} \|K\|_F^2. \quad (3.3.7)$$

Доказательство. Рассмотрим последовательность

$$\{K_j\} \in \mathcal{S}: \quad K_j \rightarrow K \in \partial\mathcal{S},$$

это означает, что $\sigma(K) = 0$. В силу непрерывности отображения σ имеем

$$\sigma(A - BK_j C) \rightarrow \sigma(A - BKC);$$

иными словами, для любого $\varepsilon > 0$ найдется число $N = N(\varepsilon)$ такое, что $|\sigma(A - BK_j C) - \sigma(A - BKC)| = \sigma(A - BK_j C) < \varepsilon$ для всех $j \geq N$.

Пусть X_j — решение уравнения Ляпунова (3.3.2), ассоциированного с регулятором K_j , тогда с учетом лемм Б.3 и Б.10 имеем

$$\begin{aligned} f(K_j) &= \text{tr } X_j \Sigma \geq \lambda_{\min}(\Sigma) \text{tr } X_j \geq \frac{\lambda_{\min}(\Sigma)\lambda_{\min}(Q + C^T K_j^T R K_j C)}{2\sigma(A_{K_j})} \geq \\ &\geq \frac{\lambda_{\min}(\Sigma)\lambda_{\min}(Q)}{2\sigma(A_{K_j})} \geq \frac{\lambda_{\min}(\Sigma)\lambda_{\min}(Q)}{2\varepsilon} \xrightarrow{\varepsilon \rightarrow 0} +\infty. \end{aligned}$$

¹ См. также раздел 1.2.3 и приложение Г.6.

С другой стороны, рассмотрим последовательность

$$\{K_j\} \in \mathcal{S}: \quad \|K_j\| \rightarrow +\infty.$$

С учетом лемм [Б.8](#) и [Б.3](#) имеем

$$\begin{aligned} f(K_j) &= \text{tr } X_j \Sigma = \text{tr}(Y_j(Q + C^T K_j^T R K_j C)) \geq \text{tr}(Y_j C^T K_j^T R K_j C) \geq \\ &\geq \lambda_{\min}(Y_j) \text{tr}(C^T K_j^T R K_j C) \geq \lambda_{\min}(Y_j) \lambda_{\min}(CC^T) \text{tr}(K_j^T R K_j) \geq \\ &\geq \lambda_{\min}(Y_j) \lambda_{\min}(CC^T) \lambda_{\min}(R) \text{tr}(K_j^T K_j) \geq \\ &\geq \lambda_{\min}(Y_j) \lambda_{\min}(CC^T) \lambda_{\min}(R) \|K_j\|_F^2, \end{aligned}$$

где Y_j — решение сопряженного уравнения Ляпунова

$$A_{K_j} Y_j + Y_j A_{K_j}^T + \Sigma = 0.$$

Далее, согласно лемме [Б.10](#)

$$\lambda_{\min}(Y_j) \geq \frac{\lambda_{\min}(\Sigma)}{2\|A_{K_j}\|}, \quad (3.3.8)$$

и для завершения доказательства остается оценить сверху величину $\|A_{K_j}\|$:

$$\begin{aligned} \|A - BK_j C\| &\leq \|A\| + \|BK_j C\| \leq \|A\| + \|BK_j C\|_F \leq \\ &\leq \|A\| + \|B\| \|C\| \|K_j\|_F; \end{aligned}$$

здесь использованы леммы [Б.2](#) и [Б.3](#).

Окончательно

$$\begin{aligned} f(K_j) &\geq \frac{\lambda_{\min}(\Sigma) \lambda_{\min}(CC^T) \lambda_{\min}(R) \|K_j\|_F^2}{2\|A - BK_j C\|} \geq \\ &\geq \frac{\lambda_{\min}(\Sigma) \lambda_{\min}(CC^T) \lambda_{\min}(R) \|K_j\|_F^2}{2\|A\| + 2\|B\| \|C\| \|K_j\|_F} \xrightarrow{\|K_j\| \rightarrow +\infty} +\infty, \end{aligned}$$

что и завершает доказательство. ■

Из оценки [\(3.3.7\)](#) немедленно вытекает

Следствие 3.3.1. Для любого $K_0 \in \mathcal{S}$ множество \mathcal{S}_0 ограничено.

С другой стороны, у функции $f(K)$ на множестве \mathcal{S}_0 существует точка минимума, как у непрерывной функции на компактом множестве, но \mathcal{S}_0 не имеет общих точек с границей множества \mathcal{S} в силу [\(3.3.6\)](#). Следовательно, справедливо

Следствие 3.3.2. Существует точка минимума $K_* \in \mathcal{S}$.

Дифференцируемость функции $f(K)$ была установлена в работе Р. Калмана [96] для задачи управления по состоянию и в работе У. Левина и М. Атанаса [104] — для обратной связи по выходу. Приведем и докажем соответствующий результат.

Лемма 3.3.3. Для любого $K \in \mathcal{S}$ градиент функции $f(K)$ дается выражением

$$\frac{1}{2}f'(K) = (RKC - B^T X)YC^T, \quad (3.3.9)$$

где Y — решение уравнения Ляпунова

$$A_K Y + YA_K^T + \Sigma = 0. \quad (3.3.10)$$

Доказательство. Придадим величине K в уравнении (3.3.2) приращение ΔK и обозначим соответствующее приращение P через ΔP :

$$(A - B(K + \Delta K)C)^T(X + \Delta X) + (X + \Delta X)(A - B(K + \Delta K)C) + C^T(K + \Delta K)^T R(K + \Delta K)C + Q = 0.$$

Линеаризуя и вычитая из него уравнение (3.3.2), имеем

$$(A - BKC)^T \Delta X + \Delta X(A - BKC) + (RKC - B^T X)^T \Delta KC + C^T \Delta K^T (RKC - B^T X) = 0.$$

Из полученного и двойственного к нему уравнения (3.3.10) находим

$$\begin{aligned} \Delta f(K) &= \text{tr}(X + \Delta X)\Sigma - \text{tr} X\Sigma = \text{tr} \Delta X\Sigma = \\ &= 2 \text{tr}(Y(RKC - B^T X)^T \Delta KC) = 2\langle (RKC - B^T X)YC^T, \Delta K \rangle, \end{aligned}$$

откуда и следует (3.3.9). ■

Необходимым условием существования точки минимума K_* функции $f(K)$ является $f'(K_*) = 0$ (поскольку K_* существует и принадлежит открытому множеству \mathcal{S}). Таким образом, приходим к системе из трех нелинейных матричных уравнений относительно K_* , а именно, $f'(K_*) = 0$, (3.3.10) и (3.3.2). В общем случае они не могут быть решены явным образом и здесь требуется применение численных методов.

Вместе с тем в случае обратной связи по состоянию (иными словами при $C = I$) можно получить явную форму решения. А именно, приравнивая градиент функции $f(K)$, найденный в соответствии с леммой 3.3.3, к нулю и учитывая, что $Y \succ 0$, имеем

$$K_* = R^{-1}B^T X_*.$$

Подставляя найденную величину K_* матрицы обратной связи в уравнение (3.3.2), приходим к уравнению Риккати относительно соответствующей ей матрицы X_* :

$$A^T X_* + X_* A - X_* B R^{-1} B^T X_* + Q = 0.$$

Строго говоря, получено не в полном смысле слова явное решение, поскольку уравнение Риккати придется решать численным образом, но методы его решения относятся к числу хорошо разработанных, см. [72, 77].

Функция $f(K)$ оказывается дважды дифференцируемой; чтобы избежать тензорной записи, ограничимся изучением действия гессиана $f(K)$ на матрицу $E \in \mathbb{R}^{p \times \ell}$, понимаемое в смысле второй производной по направлению.

Лемма 3.3.4. *Градиент функции $f(K)$ дифференцируем для любого $K \in \mathcal{S}$, причем действие гессиана функции $f(K)$ на произвольную матрицу $E \in \mathbb{R}^{p \times \ell}$ дается соотношением*

$$\frac{1}{2} \langle f''(K)[E], E \rangle = \langle RECYC^T, E \rangle - 2\langle B^T X' Y C^T, E \rangle, \quad (3.3.11)$$

где X' — решение уравнения Ляпунова

$$A_K^T X' + X' A_K - (BEC)^T X - XBE + C^T E^T RKC + C^T K^T REC = 0.$$

Доказательство. Вычислим $\langle f''(K)[E], E \rangle$, взяв производную по направлению $E \in \mathbb{R}^{p \times \ell}$ от $f'(K)[E] = \langle f'(K), E \rangle$. Для этого, линеаризуя соответствующие величины, вычислим приращение $f'(K)[E]$ по направлению E в соответствии с (3.3.9):

$$\begin{aligned} & \frac{1}{2} \Delta f'(K)[E] = \\ & = (R(K + \delta E)C - B^T(X + \Delta X))(Y + \Delta Y)C^T - (RKC - B^T X)YC^T = \\ & = (RKC - B^T X)(Y + \delta Y'(K)[E])C^T + \delta(REC - B^T X'(K)[E])YC^T - \\ & \quad - (RKC - B^T X)YC^T = \delta((RKC - B^T X)Y'(K)[E] + \\ & \quad + (REC - B^T X'(K)[E])Y)C^T, \end{aligned}$$

где

$$\begin{aligned}\Delta X &= X(K + \delta E) - X(K) = \delta X'(K)[E], \\ \Delta Y &= Y(K + \delta E) - Y(K) = \delta Y'(K)[E].\end{aligned}$$

Таким образом, обозначая $X' = X'(K)[E]$ и $Y' = Y'(K)[E]$, имеем

$$\frac{1}{2} \langle f''(K)[E], E \rangle = \langle ((RKC - B^T X)Y' + (REC - B^T X')Y)C^T, E \rangle.$$

Далее, $X = X(K)$ есть решение уравнения (3.3.2); запишем его в приращениях по направлению E :

$$\begin{aligned}(A - B(K + \delta E)C)^T(X + \delta X') + (X + \delta X')(A - B(K + \delta E)C) + \\ + C^T(K + \delta E)^T R(K + \delta E)C + Q = 0,\end{aligned}$$

или

$$\begin{aligned}(A - BKC)^T(X + \delta X') + (X + \delta X')(A - BKC) - \\ - \delta(BEC)^T X - \delta X B E C + C^T K^T R K C + \\ + \delta(C^T K R E C + C^T E R K C) + Q = 0.\end{aligned}$$

Вычитая из полученного соотношения уравнение (3.3.2), приходим к уравнению

$$\begin{aligned}A_K^T X' + X' A_K - (BEC)^T X - X B E C + \\ + C^T E^T R K C + C^T K^T R E C = 0. \quad (3.3.12)\end{aligned}$$

Далее, $Y = Y(K)$ есть решение уравнения Ляпунова (3.3.10); запишем его в приращениях по направлению E :

$$(A - B(K + \delta E)C)(Y + \delta Y') + (Y + \delta Y')(A - B(K + \delta E)C)^T + \Sigma = 0,$$

или

$$\begin{aligned}(A - BKC)(Y + \delta Y') + (Y + \delta Y')(A - BKC)^T - \\ - \delta B E C Y - \delta Y(BEC)^T + \Sigma = 0.\end{aligned}$$

Вычитая из полученного соотношения уравнение (3.3.10), приходим к уравнению

$$A_K Y' + Y' A_K^T - B E C Y - Y(BEC)^T = 0. \quad (3.3.13)$$

Из двойственных уравнений (3.3.12) и (3.3.13) по лемме Б.8 имеем соотношение

$$\operatorname{tr} X' BECY = \operatorname{tr} Y'(XB - C^T K^T R)EC,$$

или

$$\langle (RKC - B^T X)Y'C^T, E \rangle = -\langle B^T X' Y C^T, E \rangle,$$

так что

$$\begin{aligned} \frac{1}{2} \langle f''(K)[E], E \rangle &= \langle ((RKC - B^T X)Y' + (REC - B^T X')Y)C^T, E \rangle = \\ &= \langle RECY'C^T, E \rangle - \langle B^T X' Y C^T, E \rangle + \langle (RKC - B^T X)Y'C^T, E \rangle = \\ &= \langle RECYC^T, E \rangle - 2\langle B^T X' Y C^T, E \rangle. \end{aligned}$$

Лемма доказана. ■

Хотя функция $f(K)$ в общем случае невыпукла, для обратной связи по состоянию можно гарантировать локальную сильную выпуклость в области точки минимума K_* .

Следствие 3.3.3. Для задачи управления по состоянию функция $f(K)$ является сильно выпуклой в окрестности K_* .

Доказательство. Заметим, что при $K = K_*$ второе слагаемое в правой части (3.3.11) обращается в нуль. Тогда с учетом $R, Y \succ 0$ легко видеть, что

$$\langle REY, E \rangle = \operatorname{tr}(R^{\frac{1}{2}} E)Y(R^{\frac{1}{2}} E)^T > 0.$$

Таким образом, гессиан положительно определен в точке K_* , поэтому существует некоторая окрестность точки K_* , в которой функция $f(K)$ сильно выпукла. ■

Дадим следующее

Определение 3.3.2. Дважды дифференцируемая функция называется *L-гладкой*, если ее вторая производная по направлению ограничена константой L .

Функция $f(K)$ не является *L-гладкой* на множестве \mathcal{S} , однако она обладает этим свойством на множестве уровня \mathcal{S}_0 .

Теорема 3.3.1. На множестве \mathcal{S}_0 справедлива оценка

$$|\langle f''(K)[E], E \rangle| \leq L \|E\|_F^2 \quad (3.3.14)$$

с константой

$$\begin{aligned} L = \frac{2f(K_0)}{\lambda_{\min}(Q)} & \left(\lambda_{\max}(R) \|C\|^2 + \|B\| \|C\|_F \frac{\sqrt{n}f(K_0)}{\lambda_{\min}(\Sigma)} \times \right. \\ & \left. \times \left(\frac{f(K_0)\|B\|}{\lambda_{\min}(\Sigma)\lambda_{\min}(Q)} + \sqrt{\frac{f^2(K_0)\|B\|^2}{\lambda_{\min}^2(\Sigma)\lambda_{\min}^2(Q)}} + \lambda_{\max}(R) \right) \right). \end{aligned} \quad (3.3.15)$$

Доказательство. Заметим, что на множестве \mathcal{S} для действия гессиана $f''(K)$ на матрицу $E \in \mathbb{R}^{p \times \ell}$, $\|E\|_F = 1$, справедлива оценка

$$\begin{aligned} \frac{1}{2} |\langle f''(K)[E], E \rangle| & \leq \\ & \leq \lambda_{\max}(R) \lambda_{\max}(CYC^T) + \|X'\|_F \|B\| \|C\|_F \|Y\|, \end{aligned} \quad (3.3.16)$$

где X' и Y — решения уравнений Ляпунова

$$A_K^T X' + X' A_K + (RKC - B^T X)^T EC + C^T E^T (RKC - B^T X) = 0$$

и

$$A_K Y + Y A_K^T + \Sigma = 0$$

соответственно.

Действительно, из (3.3.11) вытекает, что

$$\frac{1}{2} \sup_{\|E\|_F=1} |\langle f''(K)[E], E \rangle| \leq \sup_{\|E\|_F=1} (|\langle RECYC^T, E \rangle| + 2|\langle B^T X' Y, E \rangle|).$$

Оценим с помощью леммы Б.3 первое слагаемое в правой части неравенства:

$$\begin{aligned} \langle RECYC^T, E \rangle &= \text{tr } RECYC^T E^T \leq \\ &\leq \lambda_{\max}(R) \lambda_{\max}(CYC^T) \text{tr } E^T E \leq \lambda_{\max}(R) \lambda_{\max}(CYC^T), \end{aligned}$$

а второе слагаемое оценим применяя неравенство Коши–Буняковского и лемму Б.2:

$$\begin{aligned} |\langle B^T X' Y C^T, E \rangle| &= |\langle X', BECY \rangle| \leq \|X'\|_F \|BECY\|_F \leq \\ &\leq \|X'\|_F \|B\| \|E\|_F \|C\|_F \|Y\| = \|X'\|_F \|B\| \|C\|_F \|Y\|. \end{aligned}$$

В оценке (3.3.16) величина Y зависит от K , а X' — от K и E . Требуется получить равномерную оценку, которая бы зависела только от параметров исходной задачи и от величины K_0 .

Первое слагаемое в правой части (3.3.16) может быть оценено следующим образом:

$$\begin{aligned}\lambda_{\max}(R)\lambda_{\max}(CYC^T) &\leqslant \lambda_{\max}(R) \operatorname{tr} CYC^T \leqslant \\ &\leqslant \lambda_{\max}(R) \operatorname{tr} Y \|C\|^2 \leqslant \frac{\lambda_{\max}(R)}{\lambda_{\min}(Q)} f(K_0) \|C\|^2,\end{aligned}$$

поскольку

$$\begin{aligned}\lambda_{\min}(Q) \operatorname{tr} Y &\leqslant \lambda_{\min}(Q + C^T K^T RKC) \operatorname{tr} Y \leqslant \\ &\leqslant \operatorname{tr}(Q + C^T K^T RKC) = \operatorname{tr} X\Sigma = f(K) \leqslant f(K_0).\end{aligned}$$

Поскольку $\|Y\| \leqslant \operatorname{tr} Y$, второе слагаемое в правой части (3.3.16) может быть оценено так:

$$\|X'\|_F \|B\| \|C\|_F \|Y\| \leqslant \|X'\|_F \frac{\|B\| \|C\|_F}{\lambda_{\min}(Q)} f(K_0),$$

и остается оценить величину $\|X'\|_F$.

Прежде всего покажем, что для некоторой константы α верна оценка $X' \preceq \alpha X$. Вспомним, что X' — решение уравнения

$$A_K^T X' + X' A_K + C^T K^T REC + C^T E^T RKC - (XBEC + (XBEC)^T) = 0.$$

Используя результат леммы Б.2 заключаем, что при всех $\alpha, \beta > 0$ справедлива оценка $X' \preceq \tilde{X}'$, где \tilde{X}' — решение уравнения Ляпунова

$$\begin{aligned}A_K^T \tilde{X}' + \tilde{X}' A_K + \alpha C^T K^T RKC + \\ + \frac{1}{\alpha} C^T E^T REC + \beta X^2 + \frac{1}{\beta} (BEC)^T BEC = 0,\end{aligned}$$

или

$$\begin{aligned}A_K^T \left(\frac{\tilde{X}'}{\alpha} \right) + \left(\frac{\tilde{X}'}{\alpha} \right) A_K + C^T K^T RKC + \frac{1}{\alpha^2} C^T E^T REC + \\ + \frac{1}{\alpha} \left(\beta X^2 + \frac{1}{\beta} (BEC)^T BEC \right) = 0. \quad (3.3.17)\end{aligned}$$

Покажем, что выбором констант α и β можно добиться того, чтобы $\tilde{X}'/\alpha \preceq X$. В самом деле, сравнивая уравнения (3.3.17) и (3.3.2), приходим к необходимости выполнения соотношения

$$C^T E^T \left(\frac{1}{\alpha} R + \frac{1}{\beta} B^T B \right) EC + \beta X^2 - \alpha Q \preceq 0. \quad (3.3.18)$$

Упражнение 3.3.2. Покажите, что при

$$\tilde{\alpha} = \frac{\|X\|\|B\| + \sqrt{\|X\|^2\|B\|^2 + \lambda_{\min}(Q)\lambda_{\max}(R)}}{\lambda_{\min}(Q)}, \quad \tilde{\beta} = \frac{\|B\|}{\|X\|}$$

выполняется более сильное условие

$$\frac{1}{\alpha}\lambda_{\max}(R) + \frac{1}{\beta}\|B\|^2 + \beta\|X\|^2 - \alpha\lambda_{\min}(Q) \leq 0,$$

а значит, и условие (3.3.18).

Таким образом,

$$X' \preccurlyeq \tilde{X}' \preccurlyeq \tilde{\alpha}X,$$

откуда имеем

$$\|X'\|_F \leq \sqrt{n}\|X'\| \leq \sqrt{n}\tilde{\alpha}\|X\| \leq \sqrt{n}\tilde{\alpha}\frac{f(K_0)}{\lambda_{\min}(\Sigma)}.$$

Здесь использована оценка

$$\|X\|\lambda_{\min}(\Sigma) \leq \operatorname{tr} X\Sigma = f(K) \leq f(K_0),$$

предоставляемая леммой Б.3. С учетом этой и ранее полученных оценок окончательно приходим к величине (3.3.15). ■

Несмотря на то, что граница множества S может быть негладкой, граница множества уровня функции $f(K)$ является гладкой в силу свойства L -гладкости функции $f(K)$. Сама же функция $f(K)$ может быть невыпуклой даже в случае обратной связи по состоянию, однако при этом она будет удовлетворять условию Поляка–Лоясевича, см. раздел 1.3.2.

Теорема 3.3.2. В линейно-квадратичной задаче при управлении по состоянию функция $f(K)$ на множестве S_0 удовлетворяет условию Поляка–Лоясевича

$$\frac{1}{2}\|f'(K)\|_F^2 \geq \mu(f(K) - f(K_*))$$

при

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

Доказательство. Пусть регулятор $K_* \in \mathcal{S}$ доставляет минимум функции $f(K)$, а $X = X(K)$ и $X_* = X(K_*)$ — решения уравнений Ляпунова

$$A_K^T X + X A_K + K^T R K + Q = 0$$

и

$$A_{K_*}^T X_* + X_* A_{K_*} + K_*^T R K_* + Q = 0$$

соответственно. Вычитая эти уравнения, имеем

$$A_K^T X - A_{K_*}^T X_* + X A_K - X_* A_{K_*} + K^T R K - K_*^T R K_* = 0,$$

или

$$\begin{aligned} A_{K_*}^T (X - X_*) + (X - X_*) A_{K_*} + (K - K_*)^T M + M^T (K - K_*) - \\ - (K - K_*)^T R (K - K_*) = 0, \end{aligned}$$

где $M = R K - B^T X$.

Согласно лемме [Б.2](#), для любого $\alpha > 0$ справедливо

$$(K - K_*)^T M + M^T (K - K_*) \preccurlyeq \frac{1}{\alpha} (K - K_*)^T (K - K_*) + \alpha M^T M.$$

Поэтому, положив $\alpha = \lambda_{\min}^{-1}(R)$, имеем

$$\begin{aligned} (K - K_*)^T M + M^T (K - K_*) - (K - K_*)^T R (K - K_*) \preccurlyeq \\ \preccurlyeq \alpha M^T M + (K - K_*)^T \left(\frac{1}{\alpha} I - R \right) (K - K_*) \preccurlyeq \frac{1}{\lambda_{\min}(R)} M^T M. \end{aligned}$$

Следовательно, $X - X_* \preccurlyeq Z$, где Z — решение уравнения Ляпунова

$$A_{K_*}^T Z + Z A_{K_*} + \frac{1}{\lambda_{\min}(R)} M^T M = 0,$$

так что

$$\begin{aligned} f(K) - f(K_*) &= \text{tr}(X - X_*) \Sigma \leqslant \text{tr} Z \Sigma = \frac{1}{\lambda_{\min}(R)} \text{tr}(M^T M Y_*) \leqslant \\ &\leqslant \frac{\lambda_{\max}(Y_*)}{\lambda_{\min}(R)} \text{tr}(M^T M) \leqslant \frac{\lambda_{\max}(Y_*)}{\lambda_{\min}(R) \lambda_{\min}^2(Y)} \text{tr}(Y^T M^T M Y) = \\ &= \frac{\lambda_{\max}(Y_*)}{4 \lambda_{\min}(R) \lambda_{\min}^2(Y)} \|f'(K)\|_F^2, \end{aligned}$$

где Y удовлетворяет уравнению Ляпунова

$$A_K Y + Y A_K^T + \Sigma = 0.$$

Согласно (3.3.8),

$$\lambda_{\min}(Y) \geq \frac{\lambda_{\min}(\Sigma)}{2\|A_K\|} \geq \frac{\lambda_{\min}(\Sigma)}{2(\|A\| + \|B\|\|K\|_F)} > 0.$$

Таким образом,

$$f(K) - f(K_*) \leq \frac{(\|A\| + \|B\|\|K\|_F)^2 \lambda_{\max}(Y_*)}{\lambda_{\min}(R)\lambda_{\min}^2(\Sigma)} \|f'(K)\|_F^2.$$

Далее, получим оценку для $\|K\|_F$ при $K \in \mathcal{S}$. Рассматривая (3.3.7) как квадратное уравнение относительно $\|K\|_F$, получаем следующую верхнюю оценку для его большего корня:

$$\begin{aligned} \|K\|_F &\leq \frac{2\|B\|f(K) + 2\|B\|f(K)\sqrt{1 + \frac{2\|A\|\lambda_{\min}(\Sigma)\lambda_{\min}(R)}{\|B\|^2 f(K)}}}{2\lambda_{\min}(\Sigma)\lambda_{\min}(R)} \leq \\ &\leq \frac{2\|B\|f(K)}{\lambda_{\min}(\Sigma)\lambda_{\min}(R)} + \frac{\|A\|}{\|B\|}, \end{aligned}$$

что и завершает доказательство теоремы. ■

Обратим внимание, что константа μ зависит от K_0 и стремится к нулю при $f(K) \rightarrow \infty$. Повторим, что условие Поляка–Лоясевича не выполняется на всем множестве \mathcal{S} (мы это видели в примере 3.3.1). Конечно, это условие не выполняется и для задачи управления по выходу: так, в примере 3.3.4 были две компоненты связности с различными значениями локального минимума на каждой из них, а в примере 3.3.5 наблюдалось два локальных минимума в одной компоненте связности.

3.3.5. Алгоритм решения линейно-квадратичной задачи

В этом разделе предполагается обсудить применение градиентного метода для минимизации функции $f(K)$. Это довольно нестандартная задача, поскольку функция $f(K)$ определена не на всем множестве матриц K . Более того, она неограничена на своей области определения и может быть невыпуклой.

Вместе с тем свойства функции $f(K)$, которые были установлены в теоремах 3.3.1 и 3.3.2, позволяют построить градиентный метод

$$K_{j+1} = K_j - \gamma_j f'(K_j) \quad (3.3.19)$$

ее минимизации и обосновать его сходимость. При управлении по состоянию имеет место глобальная сходимость с линейной скоростью к единственной точке минимума, а при управлении по выходу процесс будет сходиться к стационарной точке. При этом заранее известный стабилизирующий регулятор K_0 будет служить в качестве начальной точки.

Теорема 3.3.3. В линейно-квадратичной задаче градиентный метод (3.3.19) для любого $0 < \gamma_j \leq \frac{2}{L}$ (где величина L дается теоремой 3.3.1) порождает невозрастающую последовательность $f(K_j)$:

$$f(K_{j+1}) \leq f(K_j) - \gamma_j \left(1 - \frac{L\gamma_j}{2}\right) \|f'(K_j)\|_F^2.$$

При этом если $0 < \varepsilon_1 \leq \gamma_j \leq \frac{2}{L} - \varepsilon_2$, $\varepsilon_2 > 0$, то

$$f'(K_j) \rightarrow 0, \quad \min_{0 \leq j \leq k} \|f'(K_j)\|_F^2 \leq \frac{f(K_0)}{c_1 k}, \quad \text{где } c_1 = \frac{\varepsilon_1 \varepsilon_2 L}{2},$$

и в случае $C = I$ последовательность $\{K_j\}$ сходится к глобальному минимуму K_* с линейной скоростью:

$$\|K_j - K_*\| \leq cq^j, \quad 0 \leq q < 1. \quad (3.3.20)$$

Доказательство. Нетривиальной частью доказательства является обоснование того, что все итерации остаются во множестве \mathcal{S}_0 . Обозначим $K = K_0$ и рассмотрим скалярную функцию

$$\varphi(t) = f(K_t), \quad K_t = K - tf'(K), \quad t \geq 0;$$

будем предполагать, что $f'(K) \neq 0$. Функция $\varphi(t)$ дважды дифференцируема для малых t (так как K — внутренняя точка множества \mathcal{S} и $f(K)$ дважды дифференцируема на \mathcal{S}), причем

$$\varphi'(0) = -\|f'(K)\|^2 < 0.$$

Следовательно, $\varphi(t) < \varphi(0) = f(K)$ и $K_t \in \mathcal{S}_0$ для малых $t > 0$.

Множество \mathcal{S}_0 ограничено; обозначим

$$T = \max\{t: K_\tau \in \mathcal{S}_0, 0 \leq \tau \leq t\},$$

то есть T — момент первого пересечения луча K_t , $t > 0$, с границей множества \mathcal{S}_0 . Теорема 3.3.1 гарантирует L -гладкость функции $f(K_t)$ при $0 \leq t \leq T$. Это в свою очередь влечет L' -гладкость функции $\varphi(t)$ при $0 \leq t \leq T$ с $L' = L\|f'(K)\|^2$. Следовательно,

$$|\varphi(T) - \varphi(0) - \varphi'(0)T| \leq \frac{L'T^2}{2},$$

но $\varphi(T) = \varphi(0) = f(K)$, и получаем, что $T \geq \frac{2}{L}$. Это означает, что $K_t = K - tf'(K) \in \mathcal{S}_0$ для всех $0 \leq t \leq \frac{2}{L}$.

Таким образом, имеем $[K_0, K_1] \in \mathcal{S}_0$, где K_1 — результат первой итерации с начальным приближением K_0 , и это же будет справедливо для всех последующих итераций K_j , $j > 1$. Для завершения доказательства остается воспользоваться свойством L -гладкости вдоль всей траектории градиентного спуска. ■

Простейшим выбором длины шага является $\gamma_j = 1/L$, при этом в неравенстве (3.3.20) константы c и q могут быть записаны явным образом. Однако в общем случае константа L неизвестна, и весьма трудно даже получить ее оценку. Поэтому представляется разумным воспользоваться правилом Армихо (см. раздел 1.3.4); величина γ будет удовлетворять этому правилу, если

$$f(K - \gamma f'(K)) \leq f(K) - \alpha \gamma \|f'(K)\|_F^2$$

для некоторого $0 < \alpha < 1$.

Выполнения этого условия можно достичь последовательным дроблением пробного шага γ , который может быть выбран из следующих соображений. Рассмотрим одномерную функцию

$$\varphi(t) = f(K - tf'(K)).$$

Сделав одну итерацию метода Ньютона для минимизации $\varphi(t)$ при начальной точке $t_0 = 0$, имеем:

$$t_1 = \frac{\varphi'(0)}{\varphi''(0)},$$

или, вычисляя производные,

$$t_1 = \frac{\|f'(K)\|_F^2}{\langle f''(K)[f'(K)], f'(K) \rangle}. \quad (3.3.21)$$

Выражения для величин, входящих в полученную формулу, даются соотношениями (3.3.9) и (3.3.11). Заметим, что $t_1 \geq 1/L$ согласно (3.3.14),

так что полученная таким образом длина шага ограничена снизу. Полагая при $K = K_j$

$$\gamma_j = \min\{t_1, T_1\}$$

при некотором $T_1 > 0$ (это накладывает ограничение сверху на длину пробного шага), приходим к следующему алгоритму.

Алгоритм 7.

1. Задаемся параметрами $\varepsilon > 0$, $0 < \tau < 1$, $T_1 > 0$ и начальным приближением $K_0 \in \mathcal{S}$.
2. На j -й итерации задан регулятор K_j . Вычисляем матрицу

$$A_{K_j} = A - BK_jC,$$

решаем уравнения

$$A_{K_j}^T X + X A_{K_j} + Q + K_j^T R K_j = 0,$$

$$A_{K_j} Y + Y A_{K_j}^T + \Sigma = 0$$

и находим матрицы X и Y .

3. Вычисляем градиент $H_j = f'(K_j)$ из соотношения

$$f'(K) = 2(RKC - B^T X)YC^T.$$

Если $\|H_j\| \leq \varepsilon$, то K_j принимаем за приближенное решение и завершаем работу алгоритма.

4. Решаем уравнение

$$A_{K_j}^T X' + X' A_{K_j} + C^T H_j^T (RK_j C - B^T X) + \\ + (RK_j C - B^T X)^T H_j C = 0$$

и находим величину

$$\frac{1}{2} \langle f''(K_j)[H_j], H_j \rangle = \langle RH_j CYC^T, H_j \rangle - 2\langle B^T X' Y C^T, H_j \rangle.$$

5. Находим величину пробного шага

$$\gamma = \min \left\{ T_1, \frac{\|H_j\|_F^2}{\langle f''(K_j)[H_j], H_j \rangle} \right\}.$$

6. Подбираем длину шага γ_j дроблением γ до выполнения условий:
 - a. K_{j+1} — стабилизирующий регулятор (то есть, матрица $A - BK_{j+1}C$ гурвицова);
 - б. $f(K_{j+1}) \leq f(K_j) - \tau\gamma_j\|H_j\|^2$.
7. Делаем шаг градиентного метода

$$K_{j+1} = K_j - \gamma_j H_j.$$

Переходим к п.2.

Предлагаемый метод сходится в следующем смысле.

Теорема 3.3.4. В алгоритме 7 число дроблений пробного шага γ на каждой итерации равномерно ограничено и справедливы результа́ты, сформулированные в теореме 3.3.3.

3.3.6. Связь с общими задачами оптимизации

Полученные выше результаты, относящиеся к частной задаче оптимизации обратной связи, могут быть использованы и для решения более общих задач безусловной и условной минимизации. Рассмотрим некоторые из таких постановок.

Правило выбора пробного шага, предложенное в (3.3.21), может быть использовано при решении гладкой задачи безусловной оптимизации $\min_{x \in \mathbb{R}^n} f(x)$ с помощью градиентного спуска (см. раздел 1.3.1):

$$x_{j+1} = x_j - \gamma_j f'(x_j), \quad \gamma_j = \frac{\|f'(x_j)\|^2}{(f''(x_j)f'(x_j), f'(x_j))}, \quad (3.3.22)$$

в случае, когда структура задачи позволяет эффективно вычислять значение квадратичной формы в знаменателе. Это особенно удобно в практических задачах, поскольку не требует знания констант L_1 и κ и — ценой небольших вычислительных затрат — использует информацию о вторых производных. Для квадратичных функций вида $f(x) = (Hx, x)$ этот метод совпадает с методом наискорейшего спуска; для неквадратичных функций его строгое обоснование возможно в сильно выпуклом случае.

В [86] был установлен следующий результат.

Теорема 3.3.5. Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R}$ — дважды дифференцируемая, строго выпуклая функция с константой κ , а $f'(x)$ и $f''(x)$ липшицевы с константами L_1 и L_2 соответственно.

Тогда если начальное условие x_0 удовлетворяет соотношению

$$L_2 \sqrt{2L_1(f(x_0) - f(x_*))} \leq 3\kappa^2(1 - \delta), \quad \delta > 0,$$

то процесс (3.3.22) сходится к глобальному минимуму x_* с линейной скоростью:

$$f(x_j) - f(x_*) \leq (f(x_0) - f(x_*)) \left(1 - \frac{\kappa\delta}{L_1}\right)^j.$$

Демпфированная версия метода (3.3.22) (в этом случае γ_j заменяется на $\sigma\gamma_j$, где $\sigma < \frac{L_1}{\kappa}$) сходится к глобальному минимуму x_* при произвольном начальном условии x_0 :

$$f(x_j) - f(x_*) \leq (f(x_0) - f(x_*)) \left(1 - \frac{\kappa\sigma}{L_1}\right)^j.$$

Аналогичное правило выбора пробного шага может быть использовано при решения задачи условной минимизации $\min_{x \in Q} f(x)$ с помощью метода проекции градиента, см. раздел 2.3.1.

Близкий подход может быть применен и для метода сопряженных градиентов (см. подробнее раздел 1.3.5) в задачах безусловной оптимизации $\min_{x \in \mathbb{R}^n} f(x)$. Классическая версия метода предполагает процедуру одномерной минимизации для нахождения длины шага α_j ; вместо этого можно использовать следующую схему:

$$x_{j+1} = x_j + \alpha_j p_j, \quad p_j = -f'(x_j) + \beta_j p_{j-1}, \quad (3.3.23)$$

$$\alpha_j = \frac{\|p_j\|^2}{(f'(x_j)p_j, p_j)}, \quad \beta_j = \frac{\|f'(x_j)\|^2}{\|f'(x_{j-1})\|^2}, \quad \beta_0 = 0 \quad (3.3.24)$$

(существуют различные формулы для вычисления β_j , см. [20], здесь приведена простейшая). Получение результатов, связанных со сходимостью метода (3.3.23)–(3.3.24), представляется важной и интересной задачей.

Предложенный градиентный метод оптимизации обратной связи может быть также рассмотрен в рамках общей оптимизационной задачи с ограничениями типа равенства:

$$\min_{g(x,y)=0} f(x, y),$$

где $g: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$.

Будем предполагать, что может быть найдено (в явном или численном виде) решение $x(y)$ равенства $g(x, y) = 0$ при фиксированном $y \in \mathcal{S}$, и положим $F(y) = f(x(y), y)$. Тогда исходная задача сводится к задаче безусловной оптимизации

$$\min_{y \in \mathcal{S}} F(y).$$

Градиент функции $F(y)$ выписывается без каких-либо сложностей:

$$F'(y) = -\left(g'_y(x, y)\right)\left(g'_x(x, y)\right)^{-1}f'_x(x, y) + f'_y(x, y),$$

и градиентный метод с начальным приближением $y_0 \in \mathcal{S}$ обращается в так называемый *метод приведенного градиента*:

$$y_{j+1} = y_j - \gamma_j F'(y_j), \quad x_j = x(y_j).$$

Этот метод был предложен Ф. Вульфом [134] и реализован во множестве алгоритмов, см., например, [68]; при этом, как правило, $\mathcal{S} = \mathbb{R}^n$.

Однако для нелинейных ограничений типа равенства была показана лишь его локальная сходимость [20], в то время как основной интерес представляет обоснование его глобальной сходимости. Применительно к рассмотренной выше линейно-квадратичной задаче, переменной x соответствует величина Y , а переменной y — величина K . Основным инструментом для доказательства сходимости в общем случае является получение результатов, аналогичных сформулированным в теоремах 3.3.1 и 3.3.2. Если такие результаты окажутся верными, то доказательство будет являться полным аналогом приведенных выше рассуждений.

3.4. Подавление внешних возмущений

3.4.1. Постановка задачи

Задача о подавлении ограниченных внешних возмущений формулируется следующим образом. Рассмотрим линейную стационарную динамическую систему в непрерывном времени:

$$\begin{aligned} \dot{x} &= Ax + Bu + Dw, & x(0) &= x_0, \\ y &= C_1 x, \\ z &= C_2 x + B_1 u, \end{aligned} \tag{3.4.1}$$

где $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $B_1 \in \mathbb{R}^{r \times p}$, $D \in \mathbb{R}^{n \times m}$, $C_1 \in \mathbb{R}^{\ell \times n}$, $C_2 \in \mathbb{R}^{r \times n}$, с состоянием $x(t) \in \mathbb{R}^n$, управлением $u(t) \in \mathbb{R}^p$, измеряемым выходом $y(t) \in \mathbb{R}^\ell$, регулируемым выходом $z(t) \in \mathbb{R}^r$ и измеримым по t внешним возмущением $w(t) \in \mathbb{R}^m$, ограниченным в каждый момент времени:

$$\|w(t)\| \leq 1 \quad \text{для всех } t \geq 0.$$

Никаких других ограничений на возмущение $w(t)$ не накладывается; так, оно не предполагается ни случайным, ни гармоническим.

Задача состоит в выборе стабилизирующего управления в форме обратной связи по состоянию

$$u = Kx, \quad K \in \mathbb{R}^{p \times n}$$

(если оно доступно) или по выходу

$$u = Ky, \quad K \in \mathbb{R}^{p \times \ell},$$

так, чтобы уменьшить «пик» выхода $z(t)$, то есть величину $\max_t \|z(t)\|$. Точное решение этой задачи затруднительно, однако возможна минимизация верхней грани этой величины, формулируемой с помощью понятия *инвариантного эллипсоида*. Этот подход впервые был применен в монографии С. Бойда с соавторами [37] и развит в публикациях [58, 66, 69].

Подробное изложение этой техники можно найти в книге [22]. При этом задача управления по состоянию с помощью замен переменных сводится к *параметрической задаче полуопределенного программирования* (см. раздел 2.7), то есть к задаче оптимизации линейной функции при ограничениях в форме линейных матричных неравенств и наличии скалярного параметра. Для таких задач существуют удобные численные методы решения [2, 91]. В задаче управления по выходу такое сведение невозможно (впрочем, можно применять не статический, а динамический регулятор, и тогда задача записывается как выпуклая, см. [22]); для таких задач требуются новые методы решения.

3.4.2. Метод инвариантных эллипсоидов и задача анализа

В этом разделе будет кратко рассмотрена концепция метода инвариантных эллипсоидов; подробности читатель сможет найти в [22].

Рассмотрим линейную стационарную динамическую систему в непрерывном времени:

$$\begin{aligned}\dot{x} &= Ax + Dw, \quad x(0) = x_0, \\ z &= Cx,\end{aligned}\tag{3.4.2}$$

где $A \in \mathbb{R}^{n \times n}$, $D \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{n \times n}$, с состоянием $x(t) \in \mathbb{R}^n$, выходом $z(t) \in \mathbb{R}^n$ и измеримым по t внешним возмущением $w(t) \in \mathbb{R}^m$, ограниченным в каждый момент времени:

$$\|w(t)\| \leq 1 \quad \text{для всех } t \geq 0.\tag{3.4.3}$$

Будем полагать, что система (3.4.2) устойчива, то есть матрица A гурвицева, пара (A, D) управляема (см. приложение Б.3), C — матрица полного ранга.

Заметим, что здесь наложены более жесткие требования к постановке задачи, чем обычно [22, 58, 65]: предполагается, что размерности выхода и состояния совпадают, а матрица C — невырожденная. Это предположение можно было бы ослабить, но наша цель сейчас — получить наиболее простые и наглядные результаты.

Определение 3.4.1. Эллипсоид с центром в начале координат

$$\mathcal{E}_x = \{x \in \mathbb{R}^n : x^T P^{-1} x \leq 1\}, \quad P \succ 0,\tag{3.4.4}$$

называется инвариантным для динамической системы (3.4.2), (3.4.3), если из условия $x(0) \in \mathcal{E}_x$ следует $x(t) \in \mathcal{E}_x$ для всех моментов времени $t \geq 0$.

Иными словами, любая траектория системы, исходящая из точки, лежащей в эллипсоиде \mathcal{E}_x , при всех допустимых внешних возмущениях, действующих на систему, в любой момент времени будет находиться в этом эллипсоиде.

Инвариантный эллипсоид обладает свойством *притягиваемости*: любая траектория системы, исходящая из точки вне эллипса \mathcal{E}_x , стремится к эллипсу \mathcal{E}_x с течением времени.

Соответственно, если начальное состояние системы принадлежит инвариантному эллипсоиду, то имеется равномерная оценка поведения траекторий системы: в любой момент времени траектории принадлежат этому эллипсоиду при любых допустимых внешних возмущениях; если же начальные условия произвольны, имеем асимптотическую оценку поведения траекторий системы: при любых допустимых внешних возмущениях траектории будут стремиться к этому эллипсоиду с течением времени.

Инвариантные эллипсоиды можно рассматривать как характеристику влияния внешних возмущений на траектории динамической системы. А именно, в рамках задачи анализа проблема состоит в оценке степени влияния внешних возмущений на вектор выхода системы. В этой связи естественно интересоваться минимальными в том или ином смысле эллипсоидами, содержащими выход системы.

Нетрудно видеть, что если \mathcal{E}_x — инвариантный эллипсоид (3.4.4) с матрицей P , то выход системы (3.4.2) при $x_0 \in \mathcal{E}_x$ принадлежит эллипсоиду

$$\mathcal{E}_z = \{z \in \mathbb{R}^n : z^T (CPC^T)^{-1} z \leq 1\}. \quad (3.4.5)$$

Эллипсоид (3.4.5) будем называть *ограничивающим (по выходу)*. Часто в качестве критерия его минимальности рассматривается линейная функция $f(P) = \text{tr } CPC^T$, значение которой равно сумме квадратов полуосей ограничивающего эллипса. Нетрудно видеть, что справедлива оценка

$$\|z(t)\|^2 \leq \text{tr } CPC^T.$$

Таким образом, получаем верхнюю оценку выхода при ограниченной помехе, ее и будем минимизировать.

В [37] был установлен результат, дающий критерий инвариантности эллипса в терминах линейных матричных неравенств. Несколько уточняя этот критерий (см. [22]), приходим к следующей одномерной оптимизационной задаче при наличии матричных ограничений типа равенств.

Теорема 3.4.1. *Пусть матрица A гурвицева, $\sigma = \sigma(A) > 0$, пара (A, D) управляема, а матрица $P(\alpha) \succ 0$, $0 < \alpha < 2\sigma$, удовлетворяет уравнению Ляпунова*

$$\left(A + \frac{\alpha}{2} I\right)P + P\left(A + \frac{\alpha}{2} I\right)^T + \frac{1}{\alpha} DD^T = 0. \quad (3.4.6)$$

Тогда задача об оптимальном ограничивающем эллипсе сводится к минимизации скалярной функции

$$f(\alpha) = \text{tr } CP(\alpha)C^T$$

на интервале $0 < \alpha < 2\sigma$, и если α_* — точка минимума и $x(0)$ удовлетворяет условию $x^T(0)P^{-1}(\alpha_*)x(0) \leq 1$, то гарантируется равномерная оценка

$$\|z(t)\| \leq \sqrt{f(\alpha_*)}, \quad 0 \leq t < \infty.$$

Для компактности записи будем далее часто опускать зависимость P от α .

Поскольку уравнение (3.4.6) нелинейно по совокупности переменных P и α , то в литературе по этой проблематике (см., например, [66]) поиск минимального ограничивающего эллипсоида предлагалось проводить на одномерной сетке по параметру α ; при этом вместо матричного уравнения (3.4.6) рассматривалось соответствующее матричное неравенство Ляпунова. Рассмотрим более эффективный способ оптимизации, основанный на свойствах минимизируемой функции и на явном виде ее производных. Сформулируем эти свойства.

Лемма 3.4.1. Пусть матрица A гурвицева, $\sigma = \sigma(A) > 0$, пара (A, D) управляема, а матрица C такова, что $C^T C \succ 0$. Тогда функция $f(\alpha) = \text{tr} CPC^T$, где $0 < \alpha < 2\sigma$, а P — решение уравнения Ляпунова (3.4.6), обладает следующими свойствами:

a) функция $f(\alpha)$ определена, положительна и строго выпукла на интервале $0 < \alpha < 2\sigma$, а ее значения стремятся к бесконечности на концах интервала, причем существует $c > 0$ такое, что

$$f(\alpha) \geq \frac{c}{\alpha(2\sigma - \alpha)}, \quad 0 < \alpha < 2\sigma;$$

б) производная функции $f(\alpha)$ имеет вид

$$f'(\alpha) = \text{tr} Y \left(P - \frac{1}{\alpha^2} DD^T \right), \quad (3.4.7)$$

где Y — решение уравнения Ляпунова

$$\left(A + \frac{\alpha}{2} I \right)^T Y + Y \left(A + \frac{\alpha}{2} I \right) + C^T C = 0; \quad (3.4.8)$$

в) вторая производная функции $f(\alpha)$ определяется формулой

$$f''(\alpha) = 2 \text{tr} Y \left(X + \frac{1}{\alpha^3} DD^T \right), \quad (3.4.9)$$

где Y — решение уравнения Ляпунова (3.4.8), а X — решение уравнения Ляпунова

$$\left(A + \frac{\alpha}{2} I \right) X + X \left(A + \frac{\alpha}{2} I \right)^T + P - \frac{1}{\alpha^2} DD^T = 0, \quad (3.4.10)$$

причем $f''(\alpha^*) > 0$ и $f''(\alpha)$ монотонно возрастает слева и справа от α^* .

Доказательство. а) Согласно [22, лемма 1.2.3] уравнение Ляпунова (3.4.6) имеет единственное положительно определенное решение тогда и только тогда, когда матрица $A + \frac{\alpha}{2}I$ гурвицева: $\operatorname{Re} \lambda_i(A + \frac{\alpha}{2}I) < 0$, то есть при $0 < \alpha < 2\sigma$.

Оценим величину $f(\alpha) = \operatorname{tr} CP(\alpha)C^T$, используя лемму Б.10 с очевидными заменами:

$$f(\alpha) = \operatorname{tr} CP(\alpha)C^T \geq \lambda_{\min}(C^TC) \lambda_{\max}(P(\alpha)) \geq \frac{\|u^*D\|^2 \lambda_{\min}(C^TC)}{\alpha(2\sigma - \alpha)},$$

где u имеет тот же смысл, что и в лемме Б.10, а величина $\|u^*D\|^2$ положительна в силу предположения об управляемости пары (A, D) (а тем самым и пары $(A + \frac{\alpha}{2}I, D)$).

Покажем теперь, что функция $f(\alpha) = \operatorname{tr} CP(\alpha)C^T$ строго выпукла на интервале $(0, 2\sigma)$. В соответствии с леммой Б.4 решение уравнения (3.4.6) представимо в явном виде как

$$P(\alpha) = \int_0^{+\infty} e^{(A + \frac{\alpha}{2}I)t} \frac{1}{\alpha} DD^T e^{(A + \frac{\alpha}{2}I)^T t} dt = \int_0^{+\infty} \underbrace{\frac{e^{\alpha t}}{\alpha}}_{g(\alpha, t)} \underbrace{e^{At} DD^T e^{A^T t}}_{h(t)} dt.$$

Но $g(\alpha, t) > 0$, $h(t) \succ 0$ при $\alpha > 0$, поэтому на интервале $(0, 2\sigma)$ имеем

$$P(\alpha) = \int_0^{+\infty} g(\alpha, t) h(t) dt \succ 0, \quad f(\alpha) = \operatorname{tr} P(\alpha) C^T C > 0.$$

Прямым вычислением получаем

$$g''(\alpha, t) = ((\alpha t - 1)^2 + 1) \frac{e^{\alpha t}}{\alpha^3} \geq \frac{e^{\alpha t}}{\alpha^3} = \frac{1}{\alpha^2} g(\alpha, t)$$

(здесь дифференцирование производится по α), так что

$$f''(\alpha) = \int_0^{+\infty} g''(\alpha, t) h(t) dt \geq \frac{1}{\alpha^2} f(\alpha) \geq \frac{1}{4\sigma^2} f(\alpha^*) > 0.$$

Таким образом, вторая производная функции $f(\alpha)$ положительна и стремится к бесконечности на концах интервала $(0, 2\sigma)$.

Аналогичным образом прямым вычислением четвертой производной получаем

$$g^{(IV)}(\alpha, t) = ((\alpha t - 2)^2 \alpha^2 t^2 + 2(2\alpha t - 3)^2 + 6) \frac{e^{\alpha t}}{\alpha^5} \geq \frac{6}{\alpha^5} e^{\alpha t} = \frac{6}{\alpha^4} g(\alpha, t),$$

таким образом,

$$f^{(IV)}(\alpha) \geq \frac{6}{\alpha^4} f(\alpha) > 0,$$

то есть вторая производная $f''(\alpha)$ сама является выпуклой и растет на границах интервала.

б) Выведем теперь формулу для производной функции $f(\alpha)$. В уравнении (3.4.6) решение P является функцией от α . Продифференцируем это уравнение; под P' будем понимать производную по α :

$$AP' + P'A^T + \alpha P' + P - \frac{1}{\alpha^2} DD^T = 0.$$

Сравнивая уравнения для P' и Y и применяя лемму Б.8, получаем желаемую формулу:

$$f'(\alpha) = \text{tr } CP'C^T = \text{tr } Y \left(P - \frac{1}{\alpha^2} DD^T \right).$$

в) Аналогично получим выражение для второй производной $f(\alpha)$. Дифференцируя уравнение для P' по α , получаем

$$AP'' + P''A^T + \alpha P'' + 2P' + \frac{2}{\alpha^3} DD^T = 0.$$

Вновь применяя лемму Б.8 к этому уравнению и уравнению (3.4.10) (и имея в виду, что $X = P'$), получаем

$$f''(\alpha) = \text{tr } CP''C^T = 2 \text{tr } Y \left(X + \frac{1}{\alpha^3} DD^T \right),$$

что и завершает доказательство. ■

Как видно, для вычисления функции $f(\alpha)$ и двух ее производных достаточно решить три уравнения Ляпунова: (3.4.6), (3.4.8) и (3.4.10).

Указанные свойства функции позволяют применить для ее минимизации метод Ньютона (см. раздел 1.4). В самом деле, зададимся начальным приближением $0 < \alpha_0 < 2\sigma$, например $\alpha_0 = \sigma$, и применим итерационный процесс

$$\alpha_{j+1} = \alpha_j - \frac{f'(\alpha_j)}{f''(\alpha_j)}. \quad (3.4.11)$$

Следующая теорема гарантирует глобальную сходимость алгоритма.

Теорема 3.4.2. *В методе (3.4.11) справедливы оценки*

$$|\alpha_j - \alpha_*| \leq \frac{f''(\alpha_0)}{2^j f''(\alpha_*)} |\alpha_0 - \alpha_*|,$$

$$|\alpha_{j+1} - \alpha_*| \leq c |\alpha_j - \alpha_*|^2,$$

где $c > 0$ — некоторая константа (она может быть выписана явно).

Этот результат следует из утверждений леммы 3.4.1 и по существу является небольшой модификацией теорем 3 и 5 в [118], поэтому опускаем его доказательство. Первая оценка гарантирует глобальную сходимость метода (быстрее, чем геометрическая прогрессия с коэффициентом $1/2$), а вторая — квадратичную сходимость в окрестности решения. Реально требуется не более 3–4 итераций для получения решения с большой точностью (если только начальная точка не слишком близка к границам интервала).

Таким образом, имеется быстрый алгоритм для оптимизации по параметру α .

3.4.3. Задача синтеза

Перейдем к задаче синтеза и рассмотрим линейную непрерывную систему управления

$$\begin{aligned}\dot{x} &= Ax + Bu + Dw, & x(0) &= x_0, \\ y &= C_1 x, \\ z &= C_2 x,\end{aligned}$$

где $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $D \in \mathbb{R}^{n \times m}$, $C_1 \in \mathbb{R}^{\ell \times n}$, $C_2 \in \mathbb{R}^r \times n$, с состоянием $x(t) \in \mathbb{R}^n$, управлением $u(t) \in \mathbb{R}^p$, наблюдаемым выходом $y(t) \in \mathbb{R}^\ell$, регулируемым выходом $z(t) \in \mathbb{R}^r$ и внешним возмущением $w(t) \in \mathbb{R}^m$, удовлетворяющим ограничению (3.4.3);

Для того чтобы получить строгое доказательство сходимости предлагаемого далее алгоритма (теорема 3.4.3), будем далее предполагать, что размерности векторов x , w и z совпадают, а матрицы D и C_2 — квадратные невырожденные. Однако метод применим и работоспособен и в общей ситуации; рассматриваемые ниже примеры демонстрируют, что он вполне эффективен и при отсутствии таких ограничений. Вместе с тем его полное обоснование представляет собой важную задачу.

Целью является нахождение регулятора K в форме статической линейной обратной связи по выходу

$$u = Ky,$$

который стабилизирует замкнутую систему и подавляет воздействие внешних возмущений w , минимизируя размер ограничивающего эллипсоида для выхода z .

В качестве критерия оптимальности рассмотрим величину

$$\text{tr } C_2 P C_2^T + \rho \|K\|_F^2. \quad (3.4.12)$$

Первая компонента (3.4.12) определяет размер ограничивающего эллипса по критерию следа, а вторая представляет штраф за величину управления (при этом коэффициент $\rho > 0$ регулирует его важность). Наличие второй компоненты позволяет избежать появления больших значений матрицы регулятора: в частности, если управление и возмущение «приложены в одной точке» (то есть матрицы B и D совпадают), то за счет больших величин K линейный выход системы может быть сделан сколь угодно малым. Обычно штраф за управление вводится путем добавления члена с управлением в регулируемый выход (см. выражение $B_1 u$ в (3.4.1)); здесь же удобнее использовать форму показателя качества (3.4.12).

В целом вид функции (3.4.12) аналогичен критерию оптимальности в задаче о линейно-квадратичном регуляторе, где также присутствуют члены, отвечающие за величину отклонения траектории и за величину управления; разница лишь в том, что там эти члены имеют вид интегралов, а не максимальных значений.

Таким образом, задача о синтезе регулятора, подавляющего внешние возмущения, свелась к следующей оптимизационной задаче:

$$\min_{K, \alpha} f(K, \alpha), \quad f(K, \alpha) = \text{tr } C_2 P(K, \alpha) C_2^T + \rho \|K\|_F^2,$$

при ограничении

$$\left(A + B K C_1 + \frac{\alpha}{2} I \right) P + P \left(A + B K C_1 + \frac{\alpha}{2} I \right)^T + \frac{1}{\alpha} D D^T = 0 \quad (3.4.13)$$

относительно матричных переменных $0 \prec P \in \mathbb{S}^n$, $K \in \mathbb{R}^{p \times n}$ и скалярного параметра $\alpha > 0$.

Запись $f(K, \alpha)$ подчеркивает, что при заданных K и α матрица P находится из уравнения Ляпунова (3.4.13); тем самым независимыми

переменными являются K и α . Как и выше, для упрощения записи зависимость P от K и α будем обычно опускать.

Минимизацию по α можно производить достаточно эффективно, см. раздел 3.4.2; нужно лишь матрицу A заменить на $A_K \doteq A + BK_0C_1$. Поэтому будем интересоваться функцией

$$f(K) = \min_{\alpha} f(K, \alpha);$$

ее минимизацией и займемся, предварительно исследовав ее свойства.

В дальнейшем делаем следующее предположение.

Предположение 3.4.1. Пусть известен стабилизирующий регулятор K_0 , то есть такой, что матрица $A + BK_0C_1$ гурвицева.

Это предположение существенно, поскольку проблема существования стабилизирующего статического регулятора по выходу является нерешенной.

Перейдем к описанию свойств функции $f(K)$.

Лемма 3.4.2. *Функция $f(K)$ определена и положительна на множестве \mathcal{S} стабилизирующих регуляторов.*

В самом деле, если матрица A_K гурвицева, тогда $\sigma(A_K) > 0$ и для $0 < \alpha < 2\sigma(A_K)$ решение $P \succ 0$ уравнения Ляпунова (3.4.13) существует. Тем самым определена функция $f(K, \alpha) > 0$; при этом $f(K) > 0$ в силу леммы 3.4.1. Множество ее определения \mathcal{S} может быть невыпуклым и несвязным, причем его границы могут быть негладкими, см. примеры далее.

Лемма 3.4.3. *На множестве \mathcal{S} стабилизирующих регуляторов функция $f(K)$ коэрцитивна, причем справедливы следующие оценки:*

$$f(K) \geq \frac{\lambda_{\min}(DD^T)\lambda_{\min}(C_2C_2^T)}{4\sigma^2(A_K)}, \quad (3.4.14)$$

$$f(K) \geq \rho \|K\|^2.$$

Доказательство. Предположим, что $\{K_j\} \in \mathcal{S}$ — последовательность стабилизирующих регуляторов такая, что $K_j \rightarrow K \in \partial\mathcal{S}$, то есть $\sigma(A + BK_0C_1) = 0$. Это означает, что для любого $\varepsilon > 0$ найдется число $N = N(\varepsilon)$ такое, что неравенство

$$|\sigma(A + BK_jC_1) - \sigma(A + BK_0C_1)| = \sigma(A + BK_jC_1) < \varepsilon$$

справедливо для всех $j \geq N(\varepsilon)$.

Пусть P_j — решение уравнения Ляпунова (3.4.13), ассоциированного с регулятором K_j :

$$\left(A_{K_j} + \frac{\alpha_j}{2} I \right) P_j + P_j \left(A_{K_j} + \frac{\alpha_j}{2} I \right)^T + \frac{1}{\alpha_j} D D^T = 0,$$

а Y_j — решение двойственного к нему уравнения Ляпунова

$$\left(A_{K_j} + \frac{\alpha_j}{2} I \right)^T Y_j + Y_j \left(A_{K_j} + \frac{\alpha_j}{2} I \right) + C_2 C_2^T = 0.$$

Тогда

$$\begin{aligned} f(K_j) &= \text{tr } C_2 P_j C_2^T + \rho \|K_j\|_F^2 \geq \text{tr } P_j C_2 C_2^T = \text{tr} \left(Y_j \frac{1}{\alpha_j} D D^T \right) \geq \\ &\geq \frac{1}{\alpha_j} \lambda_{\min}(Y_j) \|D\|_F^2 \geq \frac{1}{\alpha_j} \frac{\lambda_{\min}(C_2 C_2^T)}{2 \|A + BK_j C + \frac{\alpha_j}{2} I\|} \|D\|_F^2 \geq \\ &\geq \frac{\lambda_{\min}(C_2 C_2^T)}{4\sigma(A + BK_j C_1) \|A + BK_j C + \frac{\alpha_j}{2} I\|} \|D\|_F^2 \geq \\ &\geq \frac{\lambda_{\min}(C_2 C_2^T)}{4\varepsilon (\|A + BK_j C\| + \varepsilon)} \|D\|_F^2 \xrightarrow[\varepsilon \rightarrow 0]{} +\infty, \end{aligned}$$

поскольку

$$0 < \alpha_j < 2\sigma(A + BK_j C_1)$$

и

$$\|A + BK_j C_1 + \frac{\alpha_j}{2} I\| \leq \|A + BK_j C_1\| + \frac{\alpha_j}{2} \leq \|A + BK_j C_1\| + \sigma(A + BK_j C_1).$$

С другой стороны,

$$f(K_j) = \text{tr } C_2 P_j C_2^T + \rho \|K_j\|_F^2 \geq \rho \|K_j\|_F^2 \xrightarrow[\|K_j\| \rightarrow +\infty]{} +\infty,$$

что и завершает доказательство. ■

Введем в рассмотрение множество уровня

$$\mathcal{S}_0 = \{K \in \mathcal{S}: f(K) \leq f(K_0)\}.$$

Из леммы 3.4.3 вытекает очевидное следствие.

Следствие 3.4.1. Для любого регулятора $K_0 \in \mathcal{S}$ множество \mathcal{S}_0 ограничено.

С другой стороны, у функции $f(K)$ на множестве \mathcal{S}_0 существует точка минимума (как у непрерывной функции на компактном множестве), но множество \mathcal{S}_0 не имеет общих точек с границей \mathcal{S} в силу (3.4.14). Впоследствии будет показано, что $f(K)$ дифференцируема на \mathcal{S}_0 . Следовательно, справедливо следствие 3.4.2.

Следствие 3.4.2. *Существует точка минимума K_* на множестве S , и в ней градиент обращается в нуль.*

Перейдем к свойствам градиента функции $f(K)$.

Лемма 3.4.4. *Функция $f(K, \alpha)$ определена на множестве стабилизирующих обратных связей K и для $0 < \alpha < 2\sigma(A_K)$. На этом допустимом множестве она дифференцируема, причем градиент дается выражениями*

$$f'_K(K, \alpha) = 2(\rho K + B^T Y P C_1^T), \quad (3.4.15)$$

$$f'_{\alpha}(K, \alpha) = \text{tr } Y \left(P - \frac{1}{\alpha^2} D D^T \right),$$

где матрица Y является решением уравнения Ляпунова

$$\left(A_K + \frac{\alpha}{2} I \right)^T Y + Y \left(A_K + \frac{\alpha}{2} I \right) + C_2^T C_2 = 0. \quad (3.4.16)$$

Минимум $f(K, \alpha)$ достигается во внутренней точке допустимого множества и определяется условиями

$$f'_K(K, \alpha) = 0, \quad f'_{\alpha}(K, \alpha) = 0.$$

При этом $f(K, \alpha)$ как функция от α строго выпукла на интервале $0 < \alpha < 2\sigma(A_K)$ и достигает минимума в его внутренней точке.

Доказательство. Итак, имеем задачу:

$$\min f(K, \alpha), \quad f(K, \alpha) = \text{tr } C_2 P C_2^T + \rho \|K\|_F^2,$$

при ограничении

$$\left(A + B K C_1 + \frac{\alpha}{2} I \right) P + P \left(A + B K C_1 + \frac{\alpha}{2} I \right)^T + \frac{1}{\alpha} D D^T = 0.$$

Дифференцирование функции f по α осуществляется в соответствии с леммой 3.4.1: для этого достаточно заменить матрицу A на $A + B K C_1$.

Для дифференцирования по K дадим ему приращение ΔK и обозначим соответствующее приращение P через ΔP :

$$\begin{aligned} & \left(A + B(K + \Delta K)C_1 + \frac{\alpha}{2}I \right)(P + \Delta P) + \\ & + (P + \Delta P) \left(A + B(K + \Delta K)C_1 + \frac{\alpha}{2}I \right)^T + \frac{1}{\alpha}DD^T = 0, \end{aligned}$$

или, после линеаризации и вычитания этого и предыдущего уравнений,

$$\begin{aligned} & \left(A + BKC_1 + \frac{\alpha}{2}I \right)\Delta P + \Delta P \left(A + BKC_1 + \frac{\alpha}{2}I \right)^T + \\ & + B\Delta KC_1 P + P(B\Delta KC_1)^T = 0. \quad (3.4.17) \end{aligned}$$

Теперь вычислим приращение функционала $f(K)$, линеаризуя соответствующие величины:

$$\begin{aligned} \Delta f(K) &= \text{tr } C_2 \Delta PC_2^T + \rho \text{tr } K^T \Delta K + \rho \text{tr } (\Delta K)^T K = \\ &= \text{tr } C_2^T C_2 \Delta P + 2\rho \text{tr } K^T \Delta K. \end{aligned}$$

Рассмотрим уравнение Ляпунова (3.4.16), двойственное к (3.4.17). По лемме B.8 из уравнений (3.4.17) и (3.4.16) имеем

$$\Delta f(K) = 2 \text{tr } C_1 PYB\Delta K + 2\rho \text{tr } K^T \Delta K = \langle 2(\rho K + B^T YPC_1^T), \Delta K \rangle.$$

Таким образом,

$$f'_K(K, \alpha) = 2(\rho K + B^T YPC_1^T),$$

что и завершает доказательство леммы. ■

Свойства гессиана функции $f(K)$ представлены следующим утверждением.

Лемма 3.4.5. *Функция $f(K)$ дважды дифференцируема, причем действие гессиана $f''(K)$ на произвольную матрицу $E \in \mathbb{R}^{p \times \ell}$ дается выражением*

$$\frac{1}{2} \langle f''(K)[E], E \rangle = \rho \langle E, E \rangle + 2 \langle B^T YP'C_1^T, E \rangle, \quad (3.4.18)$$

где P' – решение уравнения Ляпунова

$$\left(A_K + \frac{\alpha}{2}I \right)P' + P' \left(A_K + \frac{\alpha}{2}I \right)^T + BEC_1 P + P(BEC_1)^T = 0. \quad (3.4.19)$$

Доказательство. Вычислим приращение $f'(K)[E]$ по направлению $E \in \mathbb{R}^{p \times \ell}$, линеаризуя соответствующие величины:

$$\begin{aligned}\Delta f'(K)[E] &= \\ &= 2(\rho K + \rho \delta E + B^T(Y + \Delta Y)(P + \Delta P)C_1^T) - 2(\rho K + B^T Y P C_1^T) = \\ &= 2(\rho K + \rho \delta E + B^T(Y + \delta Y'(K)[E])(P + \delta P'(K)[E])C_1^T) - \\ &\quad - 2(\rho K + B^T Y P C_1^T) = 2\delta(\rho E + B^T(Y P'(K)[E] + Y'(K)[E]P)C_1^T),\end{aligned}$$

где

$$\Delta P = P(K + \delta E) - P(K) = \delta P'(K)[E],$$

$$\Delta Y = Y(K + \delta E) - Y(K) = \delta Y'(K)[E].$$

Таким образом, обозначая $P' = P'(K)[E]$ и $Y' = Y'(K)[E]$, имеем

$$\frac{1}{2}\langle f''(K)[E], E \rangle = \langle \rho E + B^T(Y P' + Y' P)C_1^T, E \rangle.$$

Далее, $P = P(K)$ есть решение уравнения (3.4.13); запишем его в приращениях по направлению E :

$$\begin{aligned}\left(A + B(K + \delta E)C_1 + \frac{\alpha}{2}I\right)(P + \delta P') + \\ + (P + \delta P')\left(A + B(K + \delta E)C_1 + \frac{\alpha}{2}I\right)^T + \frac{1}{\alpha}DD^T = 0,\end{aligned}$$

или

$$\begin{aligned}\left(A + BKC_1 + \frac{\alpha}{2}I\right)(P + \delta P') + \\ + (P + \delta P')\left(A + BKC_1 + \frac{\alpha}{2}I\right)^T + \\ + \delta(BEC_1P + P(BEC_1)^T) + \frac{1}{\alpha}DD^T = 0.\end{aligned}$$

Вычитая из полученного соотношения уравнение (3.4.6), приходим к уравнению (3.4.19).

Далее, $Y = Y(K)$ есть решение уравнения Ляпунова (3.4.16); запишем его в приращениях по направлению E :

$$\begin{aligned}\left(A + B(K + \delta E)C_1 + \frac{\alpha}{2}I\right)^T(Y + \delta Y') + \\ + (Y + \delta Y')\left(A + B(K + \delta E)C_1 + \frac{\alpha}{2}I\right) + C_2^T C_2 = 0,\end{aligned}$$

или

$$\begin{aligned} & \left(A + BKC_1 + \frac{\alpha}{2}I \right)^T (Y + \delta Y') + \\ & + (Y + \delta Y') \left(A + BKC_1 + \frac{\alpha}{2}I \right) + \\ & + \delta ((BEC_1)^T Y + YBEC_1) + C_2^T C_2 = 0. \end{aligned}$$

Вычитая из полученного соотношения уравнение (3.4.8), имеем

$$\begin{aligned} & \left(A + BKC_1 + \frac{\alpha}{2}I \right)^T Y' + Y' \left(A + BKC_1 + \frac{\alpha}{2}I \right) + \\ & + (BEC_1)^T Y + YBEC_1 = 0. \quad (3.4.20) \end{aligned}$$

Из (3.4.19) и (3.4.20) имеем соотношение

$$\text{tr } P' Y BEC_1 = \text{tr } Y' BEC_1 P,$$

так что

$$\begin{aligned} \frac{1}{2} \langle f''(K)[E], E \rangle &= \rho \langle E, E \rangle + \langle B^T (YP' + Y'P) C_1^T, E \rangle = \\ &= \rho \langle E, E \rangle + 2 \langle B^T Y P' C_1^T, E \rangle. \end{aligned}$$

Лемма доказана. ■

Градиент функции $f(K)$ не является липшицевым на множестве \mathcal{S} стабилизирующих регуляторов, однако он обладает этим свойством на его подмножестве \mathcal{S}_0 . Соответствующий результат представлен далее.

Лемма 3.4.6. *На множестве \mathcal{S}_0 градиент функции $f(K)$ липшицев с константой*

$$\begin{aligned} L &= \rho + 8\sqrt{n} \|B\|_F \|C_1\| \frac{f^2(K_0)}{\lambda_{\min}(C_2^T C_2) \lambda_{\min}^2(DD^T)} \times \\ &\times (\|A\| + \sqrt{f(K_0)} \|B\| \|C_1\|)^2 \left(\frac{f^2(K_0)}{\lambda_{\min}^2(C_2^T C_2)} + \|B\|^2 \|C_1\|^2 \right). \quad (3.4.21) \end{aligned}$$

Доказательство. Сначала покажем, что для действия гессиана функции $f(K)$ на матрицу $E \in \mathbb{R}^{p \times \ell}$ такую, что $\|E\|_F = 1$, справедлива оценка

$$\frac{1}{2} \sup_{\|E\|_F=1} |\langle f''(K)[E], E \rangle| \leq \rho + 2\|P'\|_F \|Y\| \|B\|_F \|C_1\|.$$

В самом деле, согласно (3.4.18)

$$\begin{aligned}
 & \frac{1}{2} \sup_{\|E\|_F=1} |\langle f''(K)[E], E \rangle| \leq \\
 & \leq \sup_{\|E\|_F=1} \rho \langle E, E \rangle + 2 \sup_{\|E\|_F=1} |\langle B^T Y P' C_1^T, E \rangle| = \\
 & = \rho \sup_{\|E\|_F=1} \|E\|_F^2 + 2 \sup_{\|E\|_F=1} |\langle P', Y B E C_1 \rangle| \leq \\
 & \leq \rho + 2 \|P'\|_F \sup_{\|E\|_F=1} \|Y B E C_1\|_F \leq \rho + 2 \|P'\|_F \|Y\| \|B\|_F \|C_1\|,
 \end{aligned}$$

поскольку с учетом леммы Б.2

$$\|Y B E C_1\|_F \leq \|Y\| \|B\|_F \|E\|_F \|C_1\|.$$

Теперь остается оценить сверху величину

$$\rho + 2 \|P'\|_F \|Y\| \|B\|_F \|C_1\|.$$

Имеем оценку для $\|Y\|$:

$$\begin{aligned}
 \frac{1}{\alpha} \lambda_{\min}(DD^T) \|Y\| & \leq \frac{1}{\alpha} \lambda_{\min}(DD^T) \operatorname{tr} Y \leq \frac{1}{\alpha} \operatorname{tr} Y D D^T = \operatorname{tr} Y \frac{1}{\alpha} D D^T = \\
 & = \operatorname{tr} P C_2^T C_2 = \operatorname{tr} C_2 P C_2^T = f(K) - \rho \|K\|_F^2 \leq f(K) \leq f(K_0),
 \end{aligned}$$

откуда

$$\|Y\| \leq \frac{\alpha}{\lambda_{\min}(DD^T)} f(K_0). \quad (3.4.22)$$

Оценка для α устанавливается следующим образом:

$$\begin{aligned}
 \alpha < 2\sigma(A + B K C_1) & \leq 2\|A + B K C_1\| \leq \\
 & \leq 2(\|A\| + \|B\| \|K\| \|C_1\|) \leq 2(\|A\| + \|B\| \|K\|_F \|C_1\|) \leq \\
 & \leq 2(\|A\| + \|B\| \sqrt{f(K)} \|C_1\|) \leq 2(\|A\| + \sqrt{f(K_0)} \|B\| \|C_1\|),
 \end{aligned}$$

так что

$$\|Y\| \leq 2 \frac{\|A\| + \sqrt{f(K_0)} \|B\| \|C_1\|}{\lambda_{\min}(DD^T)} f(K_0).$$

Теперь оценим сверху $\|P\|$:

$$\lambda_{\min}(C_2^T C_2) \|P\| \leq \operatorname{tr}(C_2 P C_2^T) = f(K) - \rho \|K\|_F^2 \leq f(K) \leq f(K_0),$$

откуда

$$\|P\| \leq \frac{f(K_0)}{\lambda_{\min}(C_2^T C_2)}.$$

Наконец, оценим сверху $\|P'\|_F$. С учетом леммы B.2 заметим, что

$$\begin{aligned} \lambda_{\max}(BEC_1 P + P(BEC_1)^T) &= \\ &= \|BEC_1 P + P(BEC_1)^T\| \leq \|P^2 + BEC_1(BEC_1)^T\| \leq \\ &\leq \|P\|^2 + \|B\|^2 \|C_1\|^2 \|E\|_F^2 \leq \frac{f^2(K_0)}{\lambda_{\min}^2(C_2^T C_2)} + \|B\|^2 \|C_1\|^2 = \\ &= \xi \frac{1}{\alpha} \lambda_{\min}(DD^T) \end{aligned}$$

при

$$\xi = \frac{\alpha}{\lambda_{\min}(DD^T)} \left(\frac{f^2(K_0)}{\lambda_{\min}^2(C_2^T C_2)} + \|B\|^2 \|C_1\|^2 \right).$$

Поэтому для решения P' уравнения Ляпунова (3.4.19) справедлива оценка

$$\begin{aligned} P' &\leq \xi P \leq \frac{\alpha}{\lambda_{\min}(DD^T)} \left(\frac{f^2(K_0)}{\lambda_{\min}^2(C_2^T C_2)} + \|B\|^2 \|C_1\|^2 \right) \frac{f(K_0)}{\lambda_{\min}(C_2^T C_2)} I \leq \\ &\leq 2f(K_0) \frac{\|A\| + \sqrt{f(K_0)} \|B\| \|C_1\|}{\lambda_{\min}(DD^T) \lambda_{\min}(C_2^T C_2)} \left(\frac{f^2(K_0)}{\lambda_{\min}^2(C_2^T C_2)} + \|B\|^2 \|C_1\|^2 \right) I, \end{aligned}$$

откуда

$$\begin{aligned} \|P'\|_F &\leq 2\sqrt{n} f(K_0) \frac{\|A\| + \sqrt{f(K_0)} \|B\| \|C_1\|}{\lambda_{\min}(DD^T) \lambda_{\min}(C_2^T C_2)} \times \\ &\quad \times \left(\frac{f^2(K_0)}{\lambda_{\min}^2(C_2^T C_2)} + \|B\|^2 \|C_1\|^2 \right). \quad (3.4.23) \end{aligned}$$

С учетом оценок (3.4.22) и (3.4.23) приходим к величине (3.4.21). ■

Полученные свойства минимизируемой функции и ее производных позволяют построить метод минимизации и обосновать его сходимость.

Ограничение в оптимизационной задаче (3.4.13) невыпукло по совокупности переменных P и K . В случае управления по состоянию (то есть при $C_1 = I$) задачу можно свести к эквивалентной линейной, введя вспомогательную матричную переменную и исключив матрицу регулятора K (этот прием был уже неоднократно использован выше),

однако в случае управления по выходу такой подход неприменим. Отметим еще, что даже в случае управления по состоянию минимизируемая функция не является выпуклой.

Предлагается итеративный подход к решению задачи (3.4.13); в его основе лежит применение градиентного метода по переменной K и минимизации по α , описанной в подразделе 3.4.2. Приведем принципиальную схему алгоритма.

Алгоритм 8.

1. Задаемся параметрами $\varepsilon > 0$, $\gamma > 0$, $0 < \tau < 1$ и начальным стабилизирующим приближением K_0 . Вычисляем величину $\alpha_0 = \sigma(A + BK_0C_1)$.
2. На j -й итерации заданы величины K_j , α_j . Вычисляем матрицу $A + BK_jC_1$, решаем уравнение (3.4.16) и находим Y . Вычисляем градиент $H_j = f'_K(K_j, \alpha_j)$ из уравнения (3.4.15). Если $\|H_j\| \leq \varepsilon$, то K_j принимаем за приближенное решение и завершаем работу алгоритма.
3. Делаем шаг градиентного метода

$$K_{j+1} = K_j - \gamma_j H_j.$$

Длину шага $\gamma_j > 0$ подбираем дроблением γ до выполнения условий:

- а. матрица $A + BK_{j+1}C_1 + \frac{\alpha_j}{2}I$ гурвицева;
 - б. $f(K_{j+1}) \leq f(K_j) - \tau\gamma_j\|H_j\|^2$.
4. Для полученного приближения K_{j+1} решаем задачу минимизации $f(K_{j+1}, \alpha)$ по α , получаем α_{j+1} . Переходим к п. 2.

Реализуемый при помощи алгоритма 8 метод сходится в следующем смысле.

Теорема 3.4.3. В алгоритме 8 на каждой итерации реализуется лишь конечное число дроблений γ_j , последовательность $\{f(K_j)\}$ монотонно убывает и

$$\|f'(K_j)\| \rightarrow 0$$

со скоростью геометрической прогрессии.

Доказательство. Используем обычную схему анализа градиентного метода для безусловной минимизации функций с липшицевым градиентом [20]. Прежде всего, алгоритм 8 определен корректно в начальной точке, так как K_0 является стабилизирующим регулятором

в силу предположения. Далее, при достаточно малых γ_j в алгоритме происходит монотонное уменьшение $f(K)$ (движение по антиградиенту), то есть K_j остаются в области S_0 и тем самым можно применять результаты леммы 3.4.6 о липшицевости градиента.

Таким образом, применимы результаты о сходимости градиентного метода для безусловной минимизации [20]. В частности, условие б) на шаге 3 алгоритма 8 будет выполнено после конечного числа дроблений, а в градиентном методе будет иметь место сходимость по градиенту с линейной скоростью. ■

Естественно, что трудно рассчитывать на сходимость к глобальному минимуму, поскольку область определения $f(K)$ может быть даже несвязной. Однако, по-видимому, для задачи управления по состоянию (то есть когда измеряемый выход y совпадает с x , иными словами при $C_1 = I$) можно ожидать, что минимизируемая функция удовлетворяет на множестве S_0 условию Поляка – Лоясевича (см. раздел 1.3.2), и гарантировать глобальную сходимость к единственной точке минимума подобно ситуации с линейно-квадратичным регулятором при управлении по состоянию (раздел 3.3).

Заметим еще, что способ выбора шага в алгоритме 8 отнюдь не является самым быстрым с вычислительной точки зрения. Например, весьма перспективным является способ, аналогичный предложенному в разделе 3.3 и основанный на использовании вторых производных. При этом шаг выбирается по формуле

$$\gamma_j = \frac{\|H_j\|^2}{\langle f''(K)[H_j], H_j \rangle}, \quad (3.4.24)$$

где вычисление выражения в знаменателе производится с помощью формулы (3.4.18).

Еще раз подчеркнем, что использование вторых производных требует всего лишь решения еще одного уравнения Ляпунова (3.4.19), то есть не сильно усложняет вычисления, а смысл формулы (3.4.24) состоит в том, что она реализует один шаг метода Ньютона для минимизации одномерной функции

$$\varphi(\gamma) = f(K_j - \gamma H_j).$$

Как правило, этот вариант (назовем его в нашем контексте *модифицированным алгоритмом 8*) дает более быструю сходимость, чем алгоритм 8; это подтверждается приводимыми далее примерами.

3.4.4. Примеры

Пример 3.4.1. Рассмотрим математический маятник (рис. 3.4.1), движущийся в вязкой среде, на который действует ограниченное внешнее возмущение.

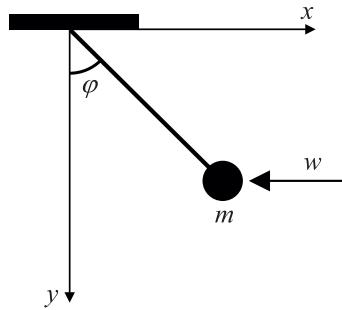


Рис. 3.4.1. Математический маятник из примера 3.4.1.

Полагая параметры единичными, приходим к линеаризованной системе

$$\begin{aligned}\dot{x}_1 &= x_2, \\ \dot{x}_2 &= -x_1 - x_2 + w, \quad |w| \leq 1,\end{aligned}$$

или в матричной форме

$$A = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}, \quad D = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Рассматривая в качестве выхода состояния системы ($C = I$), приходим к задаче

$$\min \operatorname{tr} P \quad \text{при} \quad AP + PA^T + \alpha P + \frac{1}{\alpha} DD^T = 0$$

относительно $P \in \mathbb{S}^2$ и $0 < \alpha < -2 \max \operatorname{Re} \lambda(A) = 1$.

Зная первую (3.4.7) и вторую (3.4.9) производные минимизируемой функции, с помощью метода Ньютона находим оптимальное значение $\alpha_* = 0,4618$ и соответствующую матрицу

$$P_* = \begin{pmatrix} 2,4461 & -0,5649 \\ -0,5649 & 2,1422 \end{pmatrix}$$

эллипса, содержащего траектории системы; при этом $\operatorname{tr} P_* = 4,5883$.

На рис. 3.4.2 показан график $f(\alpha) = \text{tr } P(\alpha)$, а на рис. 3.4.3 — динамика оптимизационной процедуры. По существу уже первая итерация дает решение с достаточной точностью.

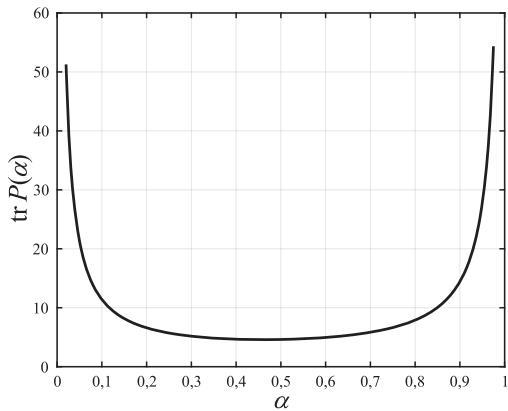


Рис. 3.4.2. График функции $f(\alpha) = \text{tr } P(\alpha)$ в примере 3.4.1.

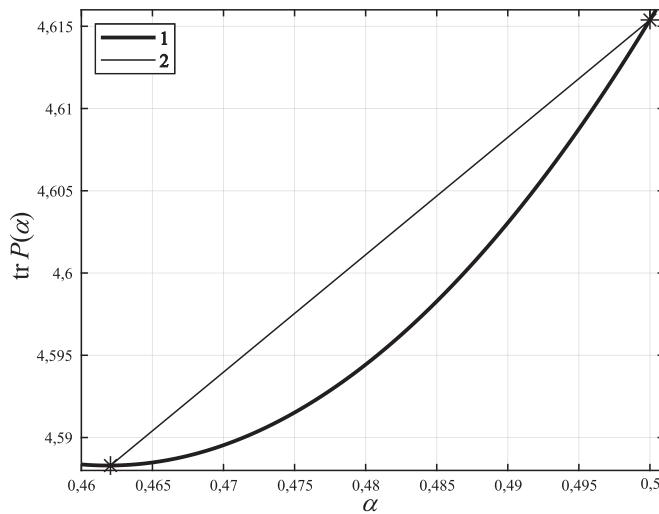


Рис. 3.4.3. Оптимизационная процедура в примере 3.4.1.



Пример 3.4.2. Рассмотрим математический маятник, на который действует ограниченное внешнее возмущение w , $|w| \leq 1$, для компенсации которого приложено управляющее воздействие u :

$$\begin{aligned}\dot{x}_1 &= x_2, \\ \dot{x}_2 &= -x_1 + u + w.\end{aligned}$$

При этом измеряемым выходом служит все состояние системы (то есть $y = x$). Отметим, что в данном примере (и ряде следующих) не выполнено одно из предположений теоремы 3.4.3: здесь размерность возмущений меньше числа состояний, однако алгоритм оптимизации применим и работает.

В результате имеем

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad B = D = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad C_1 = I, \quad K = (k_1 \ k_2).$$

Для

$$C_2 = (1 \ 0), \quad \rho = 1,$$

решая на двумерной сетке по k_1 и k_2 уравнение Ляпунова и вычисляя $\min_{\alpha} \text{tr } C_2 P(\alpha) C_2^T$, численным образом находим функцию

$$f(k_1, k_2) = C_2 P(K) C_2^T + \|K\|^2.$$

Ее линии уровня показаны на рис. 3.4.4. В данном случае линии уровня гладкие и выпуклые.

Для начального регулятора

$$K_0 = (-3 \ -3)$$

вычисления в соответствии с алгоритмом 8 доставляют на 119-й итерации регулятор

$$K_* = (-0,7224 \ -1,2205),$$

при котором $f(K_*) = 2,9377$ (линия 1 на рис. 3.4.4). Модифицированный алгоритм 8 требует всего лишь 7 итераций и приводит к регулятору

$$K_* = (-0,6477 \ -1,2038),$$

для которого $f(K_*) = 2,8670$ (линия 2 на рис. 3.4.4). ▼

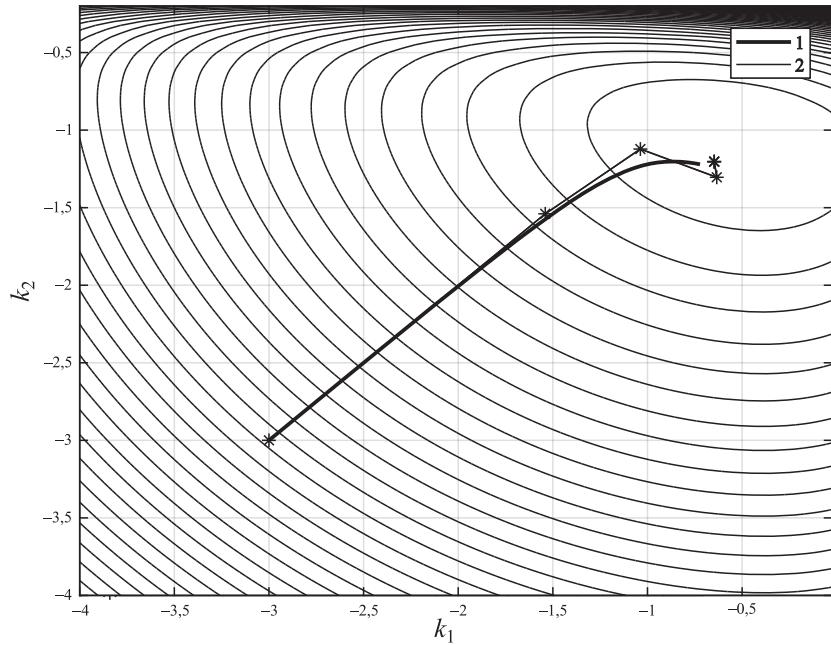


Рис. 3.4.4. Линии уровня $f(K)$ и оптимизационная процедура в примере 3.4.2.

Пример 3.4.3. Рассмотрим систему с матрицами

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & a \end{pmatrix}, \quad B = D = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

$$C_1 = \begin{pmatrix} 5 & 2 & 1 \end{pmatrix}, \quad C_2 = I, \quad \rho = 1.$$

При $a = -1,4$ функция

$$f(k) = \text{tr } C_2 P(k) C_2^T + k^2$$

имеет два локальных минимума (см. рис. 3.3.5), а при $a = -1$ ее область определения несвязна, см. рис. 3.3.4. Ясно, что отыскивать глобальный минимум в такой задаче очень трудно.

Заметим, что в данном примере используется управление по выходу, тогда как в предыдущем — управление по состоянию. ▼

Пример 3.4.4. Для системы с матрицами

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad B = D = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

$$C_1 = C_2 = I, \quad \rho = 1,$$

сечение области определения функции

$$f(K) = \text{tr } C_2 P(K) C_2^T + \|K\|_F^2$$

плоскостью $k_2 = k_3$ невыпукло, см. рис. 3.4.5. Более того, сама область \mathcal{S} негладкая, а области уровня \mathcal{S}_0 — гладкие.

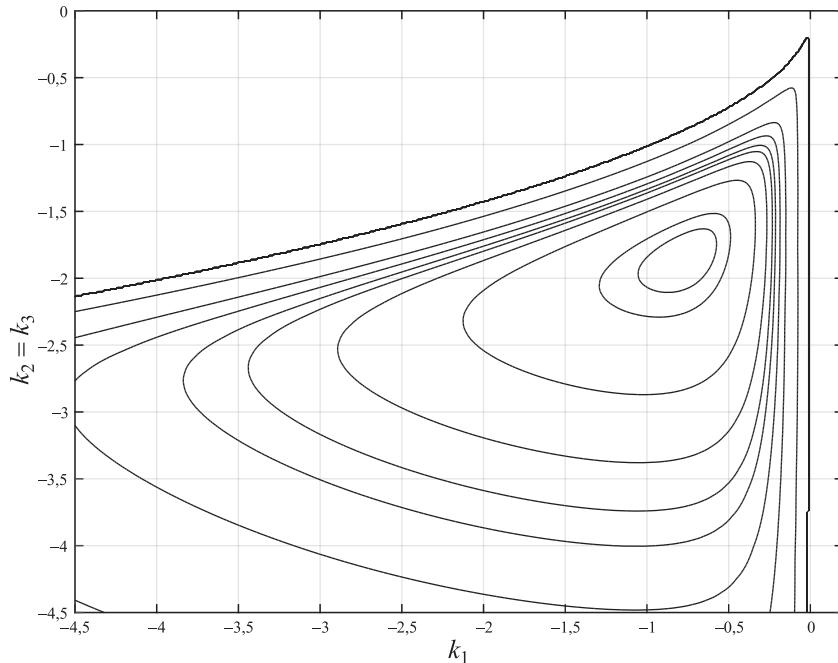


Рис. 3.4.5. Линии уровня в примере 3.4.4.

Пример 3.4.5. Вновь обратимся к задаче управления двухмассовой системой из примера 3.1.2. Пусть к левому телу приложено управление u , а к правому — внешнее возмущение w , $|w| \leq 1$, см. рис. 3.4.6.

Как и ранее, x_1, v_1 — координата и скорость левого тела, а x_2, v_2 — правого. Непрерывная модель колебаний системы описывается уравнениями

$$\begin{aligned}\dot{x}_1 &= v_1, \\ \dot{x}_2 &= v_2, \\ \dot{v}_1 &= -\frac{k}{m_1}x_1 + \frac{k}{m_2}x_2 + \frac{1}{m_1}u, \\ \dot{v}_2 &= \frac{k}{m_2}x_1 - \frac{k}{m_2}x_2 + \frac{1}{m_2}w.\end{aligned}$$

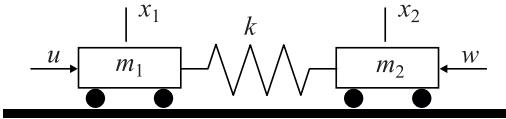


Рис. 3.4.6. Двухмассовая система из примера 3.4.5.

При единичных параметрах системы ($k_1 = k_2 = m_1 = m_2 = 1$) приходим к системе с матрицами

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad D = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Пусть измерению доступно все состояние системы ($C_1 = I$), а в качестве минимизируемого выхода возьмем координаты левого и правого тела:

$$C_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Пусть также $\rho = 1$.

Сначала в качестве начальной точки выберем некоторый начальный стабилизирующий регулятор

$$K_0 = (-1 \ 0 \ -1 \ 0)$$

и соответствующее ему начальное значение

$$\alpha_0 = \sigma(A + BK_0C_1) = 0,1484.$$

Критерием остановки процесса будет служить уменьшение значения функции $f(K)$ на некотором шаге менее чем на 0,001.

На рис. 3.4.7 линия 1 соответствует динамике изменения критерия $f(K) = \text{tr } C_2 P C_2^T + \|K\|_F^2$ в соответствии с алгоритмом 8. Процесс завершился нахождением регулятора

$$K_* = (-1,3800 \quad -0,0836 \quad -1,5482 \quad -0,8047);$$

при этом $f(K_*) = 17,5974$. Линия 2 на рис. 3.4.7 соответствует вычислениям по модифицированному алгоритму 8. Получен регулятор

$$K_* = (-1,0921 \quad -0,1720 \quad -1,4263 \quad -0,5916),$$

при этом значение функции $f(K_*) = 17,3148$ изменилось менее чем на 2 %, однако число итераций сократилось.

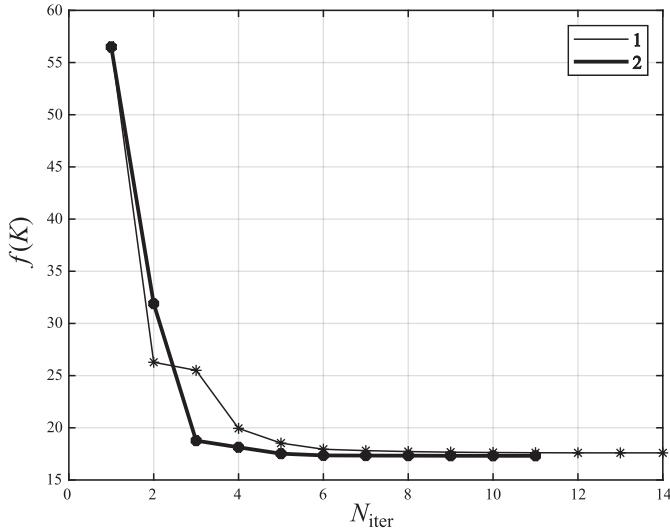


Рис. 3.4.7. Результат использования алгоритма 8 в примере 3.4.5.

Далее, для начального регулятора

$$K'_0 = (-2 \quad 0 \quad -3 \quad 1)$$

и соответствующего начального значения

$$\alpha'_0 = \sigma(A + BK_0C_1) = 0,0266$$

итерационный процесс на порядок более длителен; его динамика показана на рис. 3.4.8. Линией 1 показан расчет в соответствии с алгоритмом 8; на 125-й итерации процесс завершился нахождением регулятора

$$K'_* = (-1,6245 \quad -0,0794 \quad -1,5570 \quad -0,9869);$$

при этом $f(K'_*) = 18,0417$.

Линия 2 на рис. 3.4.8 соответствует модифицированному алгоритму 8. Процесс завершается быстрее (61 итерация) и доставляет регулятор

$$K'_* = (-1,6234 \quad -0,0793 \quad -1,5565 \quad -0,9843);$$

при этом значение минимизируемой функции практически не изменилось:

$$f(K'_*) = 18,0367.$$

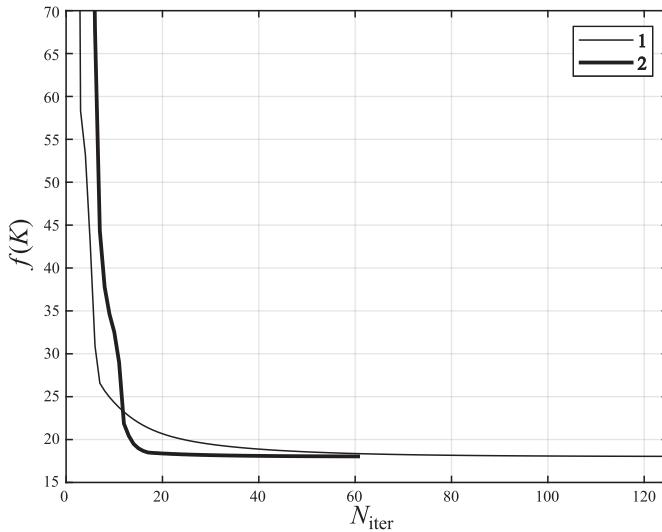


Рис. 3.4.8. Модифицированный алгоритм 8 в примере 3.4.5.:

Таким образом, из разных начальных точек получаем очень близкие регуляторы и близкие значения критерия качества. ▼

Пример 3.4.6. Рассмотрим двойной математический маятник, состоящий из двух невесомых стержней длины l_1 и l_2 , на концах которых укреплены грузики массами m_1 и m_2 .

Система движется в вязкой среде с коэффициентом сопротивления γ , в вертикальной плоскости xy , и положение маятника определяется углами φ_1 и φ_2 отклонения стержней от вертикали, см. рис. 3.4.9.

На «нижнее» тело действует ограниченное внешнее возмущение w , $|w| \leq 1$, для компенсации которого к «верхнему» телу приложено управляющее воздействие u .

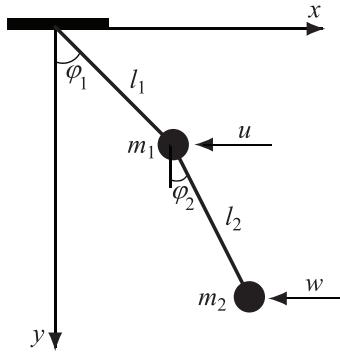


Рис. 3.4.9. Двойной математический маятник из примера 3.4.6.

Вводя переменные

$$\varphi_3 = \dot{\varphi}_1,$$

$$\varphi_4 = \dot{\varphi}_2,$$

приходим к линеаризованной системе

$$\dot{\varphi}_1 = \varphi_3,$$

$$\dot{\varphi}_2 = \varphi_4,$$

$$\dot{\varphi}_3 = -(1 + \frac{m_2}{m_1}) \frac{g}{l_1} \varphi_1 + \frac{m_2}{m_1} \frac{g}{l_1} \varphi_2 - \frac{\gamma}{m_1} \varphi_3 + \frac{1}{m_1} u,$$

$$\dot{\varphi}_4 = (1 + \frac{m_2}{m_1}) \frac{g}{l_2} \varphi_1 - (1 + \frac{m_2}{m_1}) \frac{g}{l_2} \varphi_2 - \frac{\gamma}{m_2} \varphi_4 + \frac{1}{m_2} w.$$

При

$$m_1 = m_2 = 1, \quad l_1 = l_2 = g, \quad \gamma = 0,2$$

матрицы системы имеют вид

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -2 & 1 & -0,2 & 0 \\ 2 & -2 & 0 & -0,2 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad D = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

В качестве наблюдаемого выхода выберем

$$y = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix},$$

то есть

$$C_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

а в качестве регулируемого выхода — вектор

$$z = \begin{pmatrix} \dot{\varphi}_1 \\ \dot{\varphi}_2 \end{pmatrix},$$

то есть

$$C_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Положим также $\rho = 1$.

Поскольку разомкнутая система устойчива, в качестве начального приближения для регулятора выберем

$$K_0 = (0 \quad 0).$$

При расчете по алгоритму 8 динамика изменения критерия

$$f(K) = \text{tr } C_2 P C_2^T + \|K\|_F^2$$

показана линией 1 на рис. 3.4.10. Процесс завершился на 10-м шаге нахождением стабилизирующего регулятора по выходу

$$K_* = (0,0088 \quad -0,8657),$$

при этом $f(K_*) = 29,0021$.

При выборе шага по модифицированному алгоритму 8 (линия 2 на рис. 3.4.10) процесс завершается на 7-м шаге и доставляет регулятор

$$K_* = (0,0083 \quad -0,8680),$$

при этом значение минимизируемой функции практически не изменилось:

$$f(K_*) = 29,0029.$$

Теперь в качестве начального выберем регулятор

$$K'_0 = (-1 \quad 1).$$

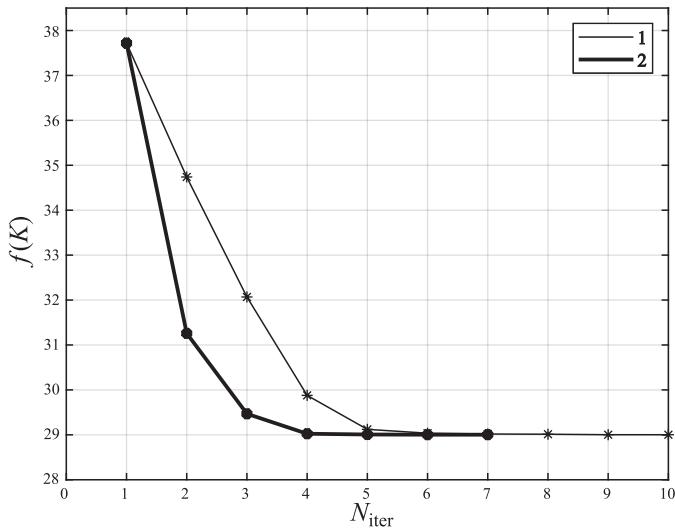


Рис. 3.4.10. Оптимизационная процедура в примере 3.4.6.

При выборе шага в соответствии с алгоритмом 8 процесс (линия 1 на рис. 3.4.11) завершается на 16-й итерации и доставляет стабилизирующий регулятор по выходу

$$K'_* = (0,0086 \quad -0,8697),$$

при этом $f(K'_*) = 29,0040$.

При выборе шага по модифицированному алгоритму 8 (линия 2 на рис. 3.4.11) процесс вновь завершается раньше — на 8-й итерации — и дает регулятор

$$K'_* = (0,0165 \quad -0,8711),$$

при этом значение минимизируемой функции вновь практически не изменилось:

$$f(K'_*) = 29,0071. \quad \blacktriangledown$$

3.5. Гарантирующая фильтрация

Классическая постановка задачи фильтрации (то есть оценки состояния динамической системы по измерениям) при случайных возмущениях допускает практически исчерпывающее решение с помощью

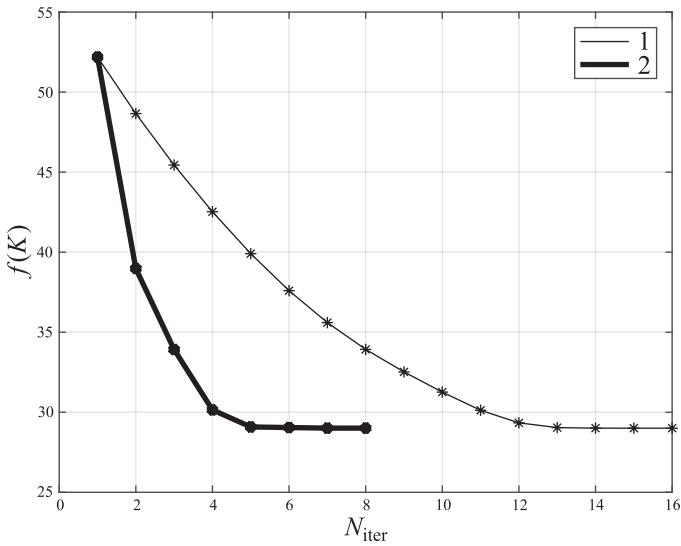


Рис. 3.4.11. Оптимизационная процедура в примере 3.4.6.

фильтра Калмана [96]; см. подробнее монографию [39], а также [12]. Однако часто известно лишь, что все возмущения являются ограниченными, а в остальном произвольными; в этом случае можно строить гарантирующие (а не вероятностные) оценки состояний.

В начале 1970-х гг. в работах Ф. Швеппе [44], А.Б. Куржанского [11] и Ф.Л. Черноуско [27] была развита эллипсоидальная техника фильтрации. Позже в [64] рассматривалась проблема фильтрации с ограниченными неслучайными возмущениями для стационарных задач: искалась оценка состояния такая, что ее ошибка гарантированно заключена в единый эллипсоид для всех моментов времени, то есть оценка является равномерной, а сам фильтр искался в классе линейных стационарных фильтров. В этом классе задач и оценок проблема оказалась полностью разрешимой: удалось построить оптимальный фильтр и оценку состояния. С технической точки зрения в [64] был применен аппарат линейных матричных неравенств [37], а исходная задача сведена к параметрической задаче полуопределенного программирования. Систематическое изложение этой техники можно найти в монографии [22].

В настоящем разделе рассматривается оптимизационный алгоритм

решения задачи фильтрации при неслучайных ограниченных внешних возмущениях. Он обладает невысокой вычислительной сложностью, предполагая на каждом итерационном шаге лишь решение уравнений Ляпунова. При этом весьма существенным с инженерно-практической точки зрения преимуществом (по сравнению с ЛМН-подходом) является возможность ограничения величины матрицы фильтра. Как показывают примеры, предлагаемая рекуррентная процедура является эффективной и приводящей ко вполне удовлетворительным результатам.

Важно отметить, что в отличие от фильтра Калмана в рамках обсуждаемого подхода оказывается возможным строить оптимальные матрицы фильтра по каждой из координат вектора состояния системы в отдельности.

3.5.1. Постановка задачи

Рассмотрим линейную динамическую систему в дискретном времени:

$$\begin{aligned} x_{k+1} &= Ax_k + B_1 u_k + D_1 w_k, \\ y_k &= Cx_k + B_2 u_k + D_2 w_k, \\ z_k &= C_1 x_k, \end{aligned} \tag{3.5.1}$$

где $A \in \mathbb{R}^{n \times n}$, $B_1 \in \mathbb{R}^{n \times p}$, $B_2 \in \mathbb{R}^{\ell \times p}$, $C \in \mathbb{R}^{\ell \times n}$, $C_1 \in \mathbb{R}^r \times n$, $D_1 \in \mathbb{R}^{n \times m}$, $D_2 \in \mathbb{R}^{\ell \times m}$, с состоянием $x_k \in \mathbb{R}^n$, начальным условием x_0 , входом $u_k \in \mathbb{R}^p$, наблюдаемым выходом $y_k \in \mathbb{R}^\ell$, оцениваемым выходом $z_k \in \mathbb{R}^r$ и внешним возмущением (шумом) $w_k \in \mathbb{R}^m$, ограниченным в каждый момент времени:

$$\|w_k\| \leq 1 \quad \text{для всех } k = 0, 1, \dots;$$

пара (A, D_1) управляема, пара (A, C) наблюдаема.

Пусть состояние x_k системы недоступно измерению и информация о системе предоставляется ее выходом y_k . Для оценивания выхода z_k будем использовать фильтр, описываемый линейным разностным уравнением относительно оценки состояния \hat{x}_k :

$$\hat{x}_{k+1} = A\hat{x}_k + B_1 u_k + L(y_k - C\hat{x}_k - B_2 u_k), \quad \hat{x}_0 = 0, \tag{3.5.2}$$

где $L \in \mathbb{R}^{n \times \ell}$. Подчеркнем, что структура фильтра задается заранее: он является линейным стационарным, выбору подлежит лишь постоянная матрица L . Эта структура фильтра такая же, как и в известном наблюдателе Люенбергера [109, 110]. По сути, можно рассматривать

этот фильтр как обобщение наблюдателя Люенбергера на задачи с помехами.

Задачей является минимизация ошибки оценки:

$$z_k - \hat{z}_k = C_1(x_k - \hat{x}_k) = C_1 e_k,$$

где $e_k = x_k - \hat{x}_k$ — невязка, удовлетворяющая согласно (3.5.1), (3.5.2) разностному уравнению

$$e_{k+1} = (A - LC)e_k + (D_1 - LD_2)w_k, \quad e_0 = x_0. \quad (3.5.3)$$

Обратим внимание, что *допустимая* матрица фильтра L стабилизирует систему (3.5.3) (то есть матрица $A - LC$ штурвальная); ее существование вытекает из свойства наблюдаемости исходной системы.

Для случайного, гауссовского шума вполне естественно применять калмановскую фильтрацию, однако здесь рассматриваются случай неслучайных ограниченных помех и соответствующая задача фильтрации с ограниченными шумами. На демонстрационных примерах будет рассмотрено, как работает фильтр Калмана с ограниченными помехами, и наоборот — к чему приведет применение к случайным помехам рассматриваемой модели с ограниченным шумом.

Заметим также, что рассматривается более общая постановка задачи, чем в [64]: в описании системы присутствует вход u_k .

Наконец, учет ограничения на внешнее возмущение вида

$$\|w_k\| \leq \gamma \quad \text{для всех } k = 0, 1, \dots$$

производится очевидным путем масштабирования матриц:

$$D_1 \doteq \gamma D_1, \quad D_2 \doteq \gamma D_2.$$

3.5.2. Гарантирующий фильтр

Мы рассмотрим и обсудим гарантирующий подход к решению задачи фильтрации при ограниченных шумах, для которого можно явным образом выписать прямые формулы, основанные на градиентном спуске. Он основывается на концепции инвариантных эллипсоидов, см. раздел 3.4.2 и подробнее в [22, 65].

Один из основных результатов раздела представлен следующим утверждением.

Теорема 3.5.1. *Пусть L^*, P^* — решение оптимизационной задачи*

$$\min f(L, \alpha), \quad f(L, \alpha) = \operatorname{tr} C_1 P C_1^T + \rho \|L\|_F^2, \quad (3.5.4)$$

при ограничении

$$\frac{1}{\alpha}(A - LC)P(A - LC)^T - P + \frac{1}{1-\alpha}(D_1 - LD_2)(D_1 - LD_2)^T = 0 \quad (3.5.5)$$

относительно матричных переменных $0 \prec P \in \mathbb{S}^n$, $L \in \mathbb{R}^{n \times \ell}$ и скалярного параметра $0 < \alpha < 1$.

Тогда ошибка оценки $z_k - \hat{z}_k$ выхода системы (3.5.1) с нулевым начальным условием при помощи наблюдателя (3.5.2) с матрицей L^* заключена в минимальный ограничивающий эллипсоид с матрицей

$$C_1 P^* C_1^T.$$

Переходя к обоснованию теоремы, напомним следующий результат [58], представляющий собой дискретную версию леммы 3.4.1.

Лемма 3.5.1. Пусть матрица A шуровская, $\rho = \max_i |\lambda_i(A)| < 1$, пара (A, D) управляема, а матрица $P(\alpha) \succ 0$, $\rho^2 < \alpha < 1$, удовлетворяет дискретному уравнению Ляпунова

$$\frac{1}{\alpha} A P A^T - P + \frac{1}{1-\alpha} D D^T = 0.$$

Тогда:

1) задача об оптимальном ограничивающем эллипсоиде для выхода системы

$$\begin{aligned} x_{k+1} &= Ax_k + Dw_k, \\ z_k &= Cx_k, \end{aligned}$$

с начальным условием x_0 и неслучайными ограниченными внешними возмущениями w_k : $\|w_k\| \leq 1$ сводится к минимизации одномерной функции $f(\alpha) = \text{tr } CP(\alpha)C^T$ на интервале $\rho^2 < \alpha < 1$;

2) если α_* — точка минимума и x_0 удовлетворяет условию

$$x_0^T P^{-1}(\alpha_*) x_0 \leq 1,$$

то гарантируется оценка

$$\|z_k\| \leq \sqrt{f(\alpha_*)}, \quad k = 0, 1, \dots$$

Итак, рассмотрим величину $C_1 e_k$ в качестве линейного выхода системы (3.5.3). Тогда если заключить невязку e_k в инвариантный эллипсоид

$$\mathcal{E} = \{e \in \mathbb{R}^n : e^T P^{-1} e \leq 1\}, \quad P \succ 0,$$

то $C_1 e_k$ будет содержаться в ограничивающем эллипсоиде

$$\mathcal{E}_z = \{e_z \in \mathbb{R}^r : e_z^T (C_1 P C_1^T)^{-1} e_z \leq 1\}, \quad (3.5.6)$$

размер которого и будем минимизировать. Таким образом, оценивается асимптотическая (а при малых уклонениях и равномерная по k) точность фильтрации.

В соответствии с леммой 3.5.1 исходная задача свелась к оптимационной задаче (3.5.4)–(3.5.5). Обратим внимание, что в минимизируемую функцию $f(L, \alpha)$ помимо компоненты, определяющей размер ограничивающего эллипса (3.5.6) по критерию следа, введен штраф за величину матрицы фильтра (при этом коэффициент $\rho > 0$ регулирует его важность). В то же время его наличие гарантирует коэрцитивность минимизируемой функции по L . Запись $f(L, \alpha)$ подчеркивает, что при заданных L и α матрица P находится из уравнения Ляпунова (3.5.5); тем самым независимыми переменными являются L и α .

Сделаем важное

Замечание 3.5.1. Рассматриваемый гарантирующий подход позволяет строить оптимальные матрицы фильтра по каждой из координат вектора состояния системы в отдельности (этой возможности нет в фильтре Калмана). Действительно, положив в задаче из теоремы 3.5.1 в качестве матрицы C_1 транспонированный i -й координатный вектор, приходим к матрице фильтра, минимизирующей невязку $x_k^{(i)} - \hat{x}_k^{(i)}$. ▀

3.5.3. Вычисление оптимальной матрицы фильтра

В [64] был предложен вариант гарантирующего подхода к решению задачи фильтрации при ограниченных шумах, который основан на технике линейных матричных неравенств и предполагает решение параметрической задачи полуопределенного программирования. В рамках рассматриваемой оптимационной задачи (3.5.4)–(3.5.5) нет необходимости применять этот технически сравнительно сложный аппарат (несмотря на то, что и минимизируемая функция, и ограничение невыпуклы по совокупности переменных P , L и α). В этом разделе будет рассмотрен регулярный итеративный подход к ее решению, в основе которого лежит применение градиентного метода по переменной L и минимизации по α по методу Ньютона.

Прежде всего приведем принципиальную схему алгоритма.

Алгоритм 9.

1. Задаемся параметрами $\varepsilon > 0$, $\gamma > 0$, $0 < \tau < 1$ и начальным допустимым приближением L_0 . Вычисляем величину

$$\alpha_0 = \frac{1 + \rho^2(A - L_0 C)}{2}.$$

2. На j -й итерации, имея величины L_j и α_j , находим градиент

$$H_j = f'_L(L_j, \alpha_j).$$

Если $\|H_j\| \leq \varepsilon$, то L_j принимаем за приближенное решение и завершаем работу алгоритма.

3. Делаем шаг градиентного метода

$$L_{j+1} = L_j - \gamma_j H_j,$$

при этом длину шага $\gamma_j > 0$ подбираем дроблением γ до выполнения условий:

- а. матрица $(A - L_{j+1} C)/\sqrt{\alpha_j}$ шуровская;
 - б. $f(L_{j+1}) \leq f(L_j) - \tau \gamma_j \|H_j\|^2$.
4. Для полученного приближения L_{j+1} решаем задачу минимизации $f(L_{j+1}, \alpha)$ по α по методу Ньютона:

$$\alpha_{j+1} = \alpha_j - \frac{f'(\alpha_j)}{f''(\alpha_j)}$$

и получаем α_{j+1} . Переходим к п. 2.

В алгоритме 9 величины $f'_L(L, \alpha)$, $f'(\alpha)$ и $f''(\alpha)$ определяются следующим образом:

$$f'_L(L, \alpha) = 2 \left(\rho L - \frac{1}{\alpha} Y (A - LC) PC^T - \frac{1}{1-\alpha} Y (D_1 - LD_2) D_2^T \right),$$

$$f'(\alpha) = \text{tr } Y \left(\frac{1}{(1-\alpha)^2} (D_1 - LD_2) (D_1 - LD_2)^T - \frac{1}{\alpha^2} (A - LC) P (A - LC)^T \right),$$

$$f''(\alpha) = 2 \operatorname{tr} Y \left(\frac{1}{(1-\alpha)^3} (D_1 - LD_2)(D_1 - LD_2)^T + \right. \\ \left. + \frac{1}{\alpha^3} (A - LC)(P - X)(A - LC)^T \right),$$

где матрицы P , Y и X являются решениями дискретных уравнений Ляпунова: (3.5.5),

$$\frac{1}{\alpha} (A - LC)^T Y (A - LC) - Y + C_1^T C_1 = 0 \quad (3.5.7)$$

и

$$\frac{1}{\alpha} (A - LC) X (A - LC)^T - X + \frac{1}{(1-\alpha)^2} (D_1 - LD_2)(D_1 - LD_2)^T - \\ - \frac{1}{\alpha^2} (A - LC) P (A - LC)^T = 0$$

соответственно.

Заметим, что на шаге 4 алгоритма реально требуется произвести не более трех–четырех итераций для получения решения с большой точностью, если начальная точка не слишком близка к границам интервала $(\rho^2(A - L_{j+1}C), 1)$.

Переходя к обоснованию алгоритма, начнем с оптимизации функции $f(L, \alpha)$ по переменной α . А именно, рассмотрим оптимизационную задачу

$$\min f(\alpha), \quad f(\alpha) = \operatorname{tr} CP(\alpha) C^T,$$

при ограничении

$$\frac{1}{\alpha} APA^T - P + \frac{1}{1-\alpha} DD^T = 0 \quad (3.5.8)$$

относительно матричных переменных $0 \prec P \in \mathbb{S}^n$ и скалярного параметра $0 < \alpha < 1$.

Далее наложим более жесткие требования к постановке задачи, чем обычно: будем предполагать, что матрица C выхода системы — квадратная невырожденная. Это предположение можно было бы ослабить, но мы хотим представить наиболее простые и наглядные результаты.

Как и ранее, зависимость P от α будем иногда опускать.

Лемма 3.5.2. *Пусть матрица A шурковская, $\rho = \rho(A)$ — спектральный радиус матрицы A , пара (A, D) управляема, а матрица C такова, что $C^T C \succ 0$. Тогда функция $f(\alpha) = \operatorname{tr} CPC^T$, где $\rho^2 < \alpha < 1$,*

а P — решение дискретного уравнения Ляпунова (3.5.8), обладает следующими свойствами:

а) функция $f(\alpha)$ определена, положительна и сильно выпукла на интервале $\rho^2 < \alpha < 1$, а ее значения стремятся к бесконечности на концах интервала, причем существует $c > 0$ такое, что

$$f(\alpha) \geq \frac{\alpha}{(1-\alpha)(\alpha-\rho^2)}c, \quad \rho^2 < \alpha < 1; \quad (3.5.9)$$

б) производная функции $f(\alpha)$ имеет вид

$$f'(\alpha) = \text{tr } Y \left(\frac{1}{(1-\alpha)^2} DD^T - \frac{1}{\alpha^2} APA^T \right),$$

где P и Y — решения дискретных уравнений Ляпунова

$$\frac{1}{\alpha} APA^T - P + \frac{1}{1-\alpha} DD^T = 0 \quad (3.5.10)$$

и

$$\frac{1}{\alpha} A^T YA - Y + C^T C = 0. \quad (3.5.11)$$

в) вторая производная функции $f(\alpha)$ определяется формулой

$$f''(\alpha) = 2 \text{tr } Y \left(\frac{1}{(1-\alpha)^3} DD^T + \frac{1}{\alpha^3} A(P-X)A^T \right),$$

где P , Y и X — решения дискретных уравнений Ляпунова (3.5.10) и (3.5.11) и

$$\frac{1}{\alpha} AXA^T - X + \frac{1}{(1-\alpha)^2} DD^T - \frac{1}{\alpha^2} APA^T = 0, \quad (3.5.12)$$

причем $f''(\alpha^*) > 0$ и $f''(\alpha)$ монотонно возрастает слева и справа от α^* .

Доказательство. а. Уравнение (3.5.8) представимо в виде

$$\left(\frac{1}{\sqrt{\alpha}} A \right) P \left(\frac{1}{\sqrt{\alpha}} A \right)^T - P = - \frac{1}{1-\alpha} DD^T$$

и согласно [22, лемма 1.2.6] имеет единственное решение тогда и только тогда, когда матрица $\frac{1}{\sqrt{\alpha}} A$ шуровская: $|\lambda_i(\frac{1}{\sqrt{\alpha}} A)| < 1$, то есть при $\rho^2 < \alpha < 1$.

Оценим величину $f(\alpha) = \operatorname{tr} CP(\alpha)C^T$, используя лемму [Б.11](#) с очевидными заменами:

$$\begin{aligned} f(\alpha) &= \operatorname{tr} CP(\alpha)C^T \geq \lambda_{\min}(C^TC)\lambda_{\max}(P(\alpha)) \geq \\ &\geq \frac{\|u^*D\|^2\lambda_{\min}(C^TC)}{(1-\alpha)(1-\rho^2(A/\sqrt{\alpha}))} = \frac{\alpha}{(1-\alpha)(\alpha-\rho^2)}\|u^*D\|^2\lambda_{\min}(C^TC), \end{aligned}$$

где u имеет тот же смысл, что и в лемме [Б.11](#), а величина $\|u^*D\|^2$ положительна в силу предположения об управляемости пары (A, D) , а тем самым и пары $(A/\sqrt{\alpha}, D)$.

Покажем теперь, что функция $f(\alpha) = \operatorname{tr} CP(\alpha)C^T$ строго выпукла на интервале $(\rho^2, 1)$. В соответствии с леммой [Б.6](#) решение уравнения [\(3.5.10\)](#) представимо в явном виде как

$$\begin{aligned} P(\alpha) &= \sum_{k=0}^{\infty} \left(\frac{1}{\sqrt{\alpha}} A \right)^k \frac{1}{1-\alpha} DD^T \left(\frac{1}{\sqrt{\alpha}} A^T \right)^k = \\ &= \sum_{k=0}^{\infty} \underbrace{\frac{1}{(1-\alpha)\alpha^k}}_{g(\alpha,k)} \underbrace{A^k DD^T (A^T)^k}_{H_k}. \end{aligned}$$

Но $H_k \succ 0$ и $g(\alpha, k) > 0$ при $0 < \alpha < 1$, поэтому на интервале $(\rho^2, 1)$ имеем

$$P(\alpha) = \sum_{k=0}^{\infty} g(\alpha, k) H_k \succ 0$$

и

$$f(\alpha) = \operatorname{tr} P(\alpha)C^TC > 0.$$

Прямым вычислением получаем

$$g'(\alpha, k) = \left(\frac{1}{1-\alpha} - \frac{k}{\alpha} \right) g(\alpha, k),$$

$$g''(\alpha, k) = \left(\left(\frac{1}{1-\alpha} - \frac{k}{\alpha} \right)^2 + \frac{1}{(1-\alpha)^2} + \frac{k}{\alpha^2} \right) g(\alpha, k) \geq \frac{1}{(1-\alpha)^2} g(\alpha, k)$$

(здесь дифференцирование производится по α), так что

$$f''(\alpha) = \sum_{k=0}^{\infty} g''(\alpha, k) \operatorname{tr} CH_k C^T \geq \frac{1}{(1-\alpha)^2} f(\alpha) \geq \frac{1}{(1-\rho^2)^2} f(\alpha^*) > 0.$$

Таким образом, вторая производная функции $f(\alpha)$ положительна и стремится к бесконечности на концах интервала $(\rho^2, 1)$.

Далее, прямым вычислением четвертой производной получаем

$$g^{(IV)}(\alpha, k) = \sum_{k=0}^{\infty} \frac{k(k+1)(k+2)(k+3)}{\alpha^{k+4}} + \frac{24}{(1-\alpha)^4} \geq \frac{24}{(1-\alpha)^4},$$

так что

$$\begin{aligned} f^{(IV)}(\alpha) &= \sum_{k=0}^{\infty} g^{(IV)}(\alpha, k) \operatorname{tr} CH_k C^T \geq \\ &\geq \frac{24}{(1-\alpha)^4} \sum_{k=0}^{\infty} \operatorname{tr} CH_k C^T > \frac{24}{(1-\rho^2)^4} \sum_{k=0}^{\infty} \operatorname{tr} CH_k C^T > 0, \end{aligned}$$

то есть вторая производная $f''(\alpha)$ сама является выпуклой и растет на границах интервала.

б. Выведем теперь формулу для производной функции $f(\alpha)$. В уравнении (3.5.10) решение P является функцией от α . Продифференцируем это уравнение; под P' будем понимать производную по α :

$$\frac{1}{\alpha} AP' A^T - P' + \frac{1}{(1-\alpha)^2} DD^T - \frac{1}{\alpha^2} APA^T = 0. \quad (3.5.13)$$

Применяя лемму Б.9 к уравнениям (3.5.13) и (3.5.11), получаем желаемую формулу:

$$f'(\alpha) = \operatorname{tr} CP' C^T = \operatorname{tr} P' C^T C = \operatorname{tr} Y \left(\frac{1}{(1-\alpha)^2} DD^T - \frac{1}{\alpha^2} APA^T \right).$$

в. Аналогичным образом получим выражение для второй производной $f(\alpha)$. Дифференцируя уравнение (3.5.13) по α , получаем

$$\frac{1}{\alpha} AP'' A^T - P'' + \frac{2}{(1-\alpha)^3} DD^T + \frac{2}{\alpha^3} APA^T - \frac{2}{\alpha^3} AP' A^T = 0.$$

Вновь применяя лемму Б.9 к этому уравнению и уравнению (3.5.11) (и имея в виду, что $X = P'$), получаем

$$\begin{aligned} f''(\alpha) &= \operatorname{tr} CP'' C^T = \operatorname{tr} P'' C^T C = \\ &= 2 \operatorname{tr} Y \left(\frac{1}{(1-\alpha)^3} DD^T + \frac{1}{\alpha^3} A(P - X)A^T \right). \end{aligned}$$

Лемма доказана. ■

Заметим, что для вычисления значения функции f и двух ее производных в точке α достаточно решить три дискретных уравнения Ляпунова: (3.5.10), (3.5.11) и (3.5.12).

Для минимизации функции $f(\alpha)$ воспользуемся методом Ньютона (раздел 1.4), задавшись начальным приближением $\rho^2(A) < \alpha_0 < 1$, например $\alpha_0 = (1 + \rho^2(A))/2$, и применим итерационный процесс

$$\alpha_{j+1} = \alpha_j - \frac{f'(\alpha_j)}{f''(\alpha_j)}.$$

Глобальную сходимость этого алгоритма гарантирует результат, полностью аналогичный сформулированному в теореме 3.4.2. Ее справедливость устанавливается аналогично сходному результату в разделе 3.3.

Вернемся к оптимизационной задаче (3.5.4)–(3.5.5) и займемся минимизацией функции

$$f(L) = \min_{\alpha} f(L, \alpha),$$

предварительно исследовав ее свойства.

Лемма 3.5.3. *Функция $f(L)$ определена и положительна на множестве \mathcal{S} допустимых матриц фильтра.*

Действительно, если матрица $A - LC$ шуровская, то $\rho(A - LC) < 1$ и для $\rho^2(A - LC) < \alpha < 1$ существует решение $P \succcurlyeq 0$ дискретного уравнения Ляпунова (3.5.5). Тем самым определена строго положительная функция $f(L, \alpha)$, при этом $f(L) > 0$ в силу (3.5.9). Как и в непрерывном случае, множество ее определения \mathcal{S} может быть невыпуклым и несвязным, а его границы — негладкими.

Лемма 3.5.4. *На множестве \mathcal{S} функция $f(L)$ коэрцитивна, причем справедливы следующие оценки:*

$$f(L) \geq \frac{1}{1 - \rho^2(A - LC)} \frac{\lambda_{\min}(C_1 C_1^T)}{1 - \sigma_{\min}^2(A - LC)} \|D_1 - LD_2\|_F^2, \quad (3.5.14)$$

$$f(L) \geq \rho \|L\|^2.$$

Доказательство. Рассмотрим последовательность $\{L_j\} \in \mathcal{S}$ допустимых матриц такую, что $L_j \rightarrow L \in \partial\mathcal{S}$, то есть $\rho(A - LC) = 1$. Это означает, что для любого $\varepsilon > 0$ найдется число $N = N(\varepsilon)$ такое, что неравенство

$$|\rho(A - L_j C) - \rho(A - LC)| = 1 - \rho(A - L_j C) < \varepsilon$$

справедливо для всех $j \geq N(\varepsilon)$.

Пусть P_j — решение уравнения (3.5.5), ассоциированного с матрицей фильтра L_j :

$$\frac{1}{\alpha_j}(A - L_j C)P_j(A - L_j C)^T - P_j + \frac{1}{1 - \alpha_j}(D_1 - L_j D_2)(D_1 - L_j D_2)^T = 0,$$

а Y_j — решение двойственного к нему дискретного уравнения Ляпунова

$$\frac{1}{\alpha_j}(A - L_j C)^T Y_j(A - L_j C) - Y_j + C_1 C_1^T = 0.$$

Тогда с учетом леммы Б.11 имеем:

$$\begin{aligned} f(L_j) &= \text{tr}(C_1 P_j C_1^T) + \rho \|L_j\|_F^2 \geq \text{tr}(P_j C_1 C_1^T) = \\ &= \text{tr}\left(Y_j \frac{1}{1 - \alpha_j}(D_1 - L_j D_2)(D_1 - L_j D_2)^T\right) \geq \\ &\geq \frac{1}{1 - \alpha_j} \lambda_{\min}(Y_j) \|D_1 - L_j D_2\|_F^2 \geq \\ &\geq \frac{1}{1 - \alpha_j} \frac{\lambda_{\min}(C_1 C_1^T)}{1 - \sigma_{\min}^2(A - L_j C)} \|D_1 - L_j D_2\|_F^2 \geq \\ &\geq \frac{1}{1 - \rho^2(A - L_j C)} \frac{\lambda_{\min}(C_1 C_1^T)}{1 - \sigma_{\min}^2(A - L_j C)} \|D_1 - L_j D_2\|_F^2 \geq \\ &\geq \frac{1}{\varepsilon} \frac{\lambda_{\min}(C_1 C_1^T)}{1 - \sigma_{\min}^2(A - L_j C)} \|D_1 - L_j D_2\|_F^2 \xrightarrow[\varepsilon \rightarrow 0]{} +\infty, \end{aligned}$$

поскольку $\rho^2(A - L_j C) < \alpha_j < 1$.

С другой стороны,

$$f(L_j) = \text{tr}(C_1 P_j C_1^T) + \rho \|L_j\|_F^2 \geq \rho \|L_j\|_F^2 \xrightarrow[\|L_j\| \rightarrow +\infty]{} +\infty,$$

что и завершает доказательство. ■

Введем в рассмотрение множество уровня

$$\mathcal{S}_0 = \{L \in \mathcal{S}: f(L) \leq f(L_0)\}.$$

Из леммы 3.5.4 вытекает очевидное

Следствие 3.5.1. Для любых $L_0 \in \mathcal{S}$ множество \mathcal{S}_0 ограничено.

С другой стороны, у непрерывной (в силу свойств решения дискретного уравнения Ляпунова) функции $f(L)$ на компактном множестве \mathcal{S}_0 существует точка минимума. При этом множество \mathcal{S}_0 не имеет общих точек с границей \mathcal{S} в силу (3.5.14). Далее будет показано, что $f(L)$ дифференцируема на \mathcal{S}_0 . Следовательно, справедливо

Следствие 3.5.2. *Существует точка минимума L_* на множестве \mathcal{S} , и в ней градиент функции $f(L)$ обращается в нуль.*

Перейдем к свойствам градиента функции $f(L, \alpha)$.

Лемма 3.5.5. *Функция $f(L, \alpha)$ определена на множестве стабилизирующих L и для $\rho^2(A - LC) < \alpha < 1$. На этом допустимом множестве она дифференцируема, причем градиент дается выражениями*

$$\begin{aligned} f'_\alpha(L, \alpha) = \text{tr } Y \left(\frac{1}{(1-\alpha)^2} (D_1 - LD_2)(D_1 - LD_2)^T - \right. \\ \left. - \frac{1}{\alpha^2} (A - LC)P(A - LC)^T \right), \quad (3.5.15) \end{aligned}$$

$$f'_L(L, \alpha) = 2 \left(\rho L - \frac{1}{\alpha} Y(A - LC)PC^T - \frac{1}{1-\alpha} Y(D_1 - LD_2)D_2^T \right), \quad (3.5.16)$$

где матрицы P и Y являются решениями дискретных уравнений Ляпунова (3.5.5) и (3.5.7).

Минимум $f(L, \alpha)$ достигается во внутренней точке допустимого множества и определяется условиями

$$f'_L(L, \alpha) = f'_\alpha(L, \alpha) = 0.$$

При этом $f(L, \alpha)$ как функция от α строго выпукла на интервале $\rho^2(A - LC) < \alpha < 1$ и достигает минимума в его внутренней точке.

Доказательство. Имеем задачу:

$$\min f(L, \alpha), \quad f(L, \alpha) = \text{tr } C_1 PC_1^T + \rho \|L\|_F^2,$$

при ограничении в виде дискретного уравнения Ляпунова (3.5.5) относительно матрицы P инвариантного эллипсоида.

Следуя лемме 3.5.2, дифференцирование по α производится в соответствии с соотношениями (3.5.15), (3.5.5) и (3.5.7). Для дифференцирования по L дадим ему приращение ΔL и обозначим соответствующее

приращение P через ΔP ; в результате (3.5.5) примет вид

$$\begin{aligned} \frac{1}{\alpha}(A - (L + \Delta L)C)(P + \Delta P)(A - (L + \Delta L)C)^T - (P + \Delta P) + \\ + \frac{1}{1-\alpha}(D_1 - (L + \Delta L)D_2)(D_1 - (L + \Delta L)D_2)^T = 0. \end{aligned}$$

Оставляя обозначение ΔP для главной части приращения, получаем

$$\begin{aligned} \frac{1}{\alpha}((A - LC)P(A - LC)^T - \Delta LCP(A - LC)^T - (A - LC)P(\Delta LC)^T + \\ + (A - LC)\Delta P(A - LC)^T) - (P + \Delta P) + \\ + \frac{1}{1-\alpha}((D_1 - LD_2)(D_1 - LD_2)^T - \\ - \Delta LD_2(D_1 - LD_2)^T - (D_1 - LD_2)(\Delta LD_2)^T) = 0. \end{aligned}$$

После вычитания уравнения (3.5.21) из этого уравнения имеем

$$\begin{aligned} \frac{1}{\alpha}(A - LC)\Delta P(A - LC)^T - \Delta P - \\ - \frac{1}{\alpha}(\Delta LCP(A - LC)^T + (A - LC)P(\Delta LC)^T) - \\ - \frac{1}{1-\alpha}(\Delta LD_2(D_1 - LD_2)^T + (D_1 - LD_2)(\Delta LD_2)^T) = 0. \quad (3.5.17) \end{aligned}$$

Теперь вычислим приращение функционала $f(L)$, линеаризуя соответствующие величины:

$$\begin{aligned} \Delta f(L) = \text{tr } C_1 \Delta PC_1^T + \rho \text{tr } L^T \Delta L + \rho \text{tr } (\Delta L)^T L = \\ = \text{tr } \Delta PC_1^T C_1 + 2\rho \text{tr } L^T \Delta L. \end{aligned}$$

По лемме Б.9 из двойственных уравнений (3.5.17) и (3.5.7) имеем

$$\begin{aligned} \Delta f(L) = -2 \text{tr } Y \left(\frac{1}{\alpha} \Delta LCP(A - LC)^T + \right. \\ \left. + \frac{1}{1-\alpha} \Delta LD_2(D_1 - LD_2)^T \right) + 2\rho \text{tr } L^T \Delta L = \\ = 2 \text{tr} \left(\rho L^T \Delta L - \frac{1}{\alpha} CP(A - LC)^T Y - \frac{1}{1-\alpha} D_2(D_1 - LD_2)^T Y \right) \Delta L = \\ = \left\langle 2 \left(\rho L - \frac{1}{\alpha} Y(A - LC)PC^T - \frac{1}{1-\alpha} Y(D_1 - LD_2)D_2^T \right), \Delta L \right\rangle. \end{aligned}$$

Таким образом, приходим к соотношению (3.5.16). ■

Градиент функции $f(L)$ не является липшицевым на множестве \mathcal{S} , однако можно показать, что он обладает этим свойством на его подмножестве \mathcal{S}_0 (аналогично тому, как это было показано в разделе 3.3).

Установленные свойства минимизируемой функции и ее производных обосновывают метод минимизации, реализованный в виде алгоритма 9.

Важным моментом является выбор пробного шага градиентного метода. Весьма перспективным является его выбор из следующих соображений. Найдем для некоторого допустимого L решение P дисcreteного уравнения Ляпунова

$$(A - LC)P(A - LC)^T - P = -I.$$

Рассмотрим приращение по L :

$$L \rightarrow L - \gamma H, \quad H = f'_L(L, \alpha),$$

и найдем, для каких γ матрица P останется матрицей квадратичной функции Ляпунова для $A - (L - \gamma H)C$, то есть

$$(A - (L - \gamma H)C)P(A - (L - \gamma H)C)^T - P \prec 0,$$

или по лемме Шура

$$\begin{pmatrix} P & A - (L - \gamma H)C \\ (A - (L - \gamma H)C)^T & P^{-1} \end{pmatrix} \succ 0.$$

Полученное матричное неравенство, которому можно придать вид

$$\begin{pmatrix} P & A - LC \\ (A - LC)^T & P^{-1} \end{pmatrix} + \gamma \begin{pmatrix} 0 & HC \\ (HC)^T & 0 \end{pmatrix} \succ 0,$$

выполняется при

$$\gamma < \lambda_{\max}^{-1} \left(\begin{pmatrix} 0 & HC \\ (HC)^T & 0 \end{pmatrix}, \begin{pmatrix} P & A - LC \\ (A - LC)^T & P^{-1} \end{pmatrix} \right).$$

3.5.4. Непрерывный случай

Рассмотрим непрерывную систему

$$\begin{aligned} \dot{x} &= Ax + B_1 u + D_1 w, & x(0) &= x_0, \\ y &= Cx + B_2 u + D_2 w, & & \\ z &= C_1 x, & & \end{aligned} \tag{3.5.18}$$

где $A \in \mathbb{R}^{n \times n}$, $B_1 \in \mathbb{R}^{n \times p}$, $B_2 \in \mathbb{R}^{\ell \times p}$, $C \in \mathbb{R}^{\ell \times n}$, $C_1 \in \mathbb{R}^{r \times n}$, $D_1 \in \mathbb{R}^{n \times m}$, $D_2 \in \mathbb{R}^{\ell \times m}$, с состоянием $x(t) \in \mathbb{R}^n$, входом $u(t) \in \mathbb{R}^p$, наблюдаемым выходом $y(t) \in \mathbb{R}^\ell$, оцениваемым выходом $z(t) \in \mathbb{R}^r$ и внешним возмущением (шумом) $w(t) \in \mathbb{R}^m$, ограниченным в каждый момент времени:

$$\|w(t)\| \leq 1 \quad \text{для всех } t \geq 0;$$

пара (A, D_1) управляема, пара (A, C) наблюдаема.

Для оценивания выхода z используется фильтр, описываемый линейным дифференциальным уравнением относительно оценки состояния \hat{x} :

$$\dot{\hat{x}} = A\hat{x} + B_1 u + L(y - C\hat{x} - B_2 u), \quad \hat{x}(0) = 0, \quad (3.5.19)$$

где $L \in \mathbb{R}^{n \times \ell}$.

Как и в дискретном случае, задачей является минимизация ошибки оценки

$$z - \hat{z} = C_1(x - \hat{x}) = C_1 e,$$

где $e(t) = x(t) - \hat{x}(t)$ — невязка, которая согласно (3.5.18), (3.5.19) удовлетворяет дифференциальному уравнению

$$\dot{e} = (A - LC)e + (D_1 - LD_2)w, \quad e(0) = x_0. \quad (3.5.20)$$

При этом *допустимая* матрица фильтра L стабилизирует систему (3.5.20) (то есть матрица $A - LC$ гурвицева); ее существование вытекает из свойства наблюдаемости исходной системы.

Рассуждая аналогично дискретному случаю и воспользовавшись леммой 3.4.1, приходим к следующему результату.

Теорема 3.5.2. *Пусть L^*, P^* — решение оптимизационной задачи*

$$\min f(L, \alpha), \quad f(L, \alpha) = \operatorname{tr} C_1 P C_1^T + \rho \|L\|_F^2,$$

при ограничении

$$\begin{aligned} & \left(A - LC + \frac{\alpha}{2} I \right) P + P \left(A - LC + \frac{\alpha}{2} I \right)^T + \\ & + \frac{1}{\alpha} (D_1 - LD_2)(D_1 - LD_2)^T = 0 \end{aligned} \quad (3.5.21)$$

относительно матричных переменных $0 \prec P \in \mathbb{S}^n$, $L \in \mathbb{R}^{n \times \ell}$ и скалярного параметра $\alpha > 0$.

Тогда ошибка оценки $z - \hat{z}$ выхода системы (3.5.18) с нулевым начальным условием при помощи наблюдателя (3.5.19) с матрицей L^* заключена в минимальный ограничивающий эллипсоид с матрицей

$$C_1 P^* C_1^T.$$

Приведем схему соответствующего алгоритма.

Алгоритм 10.

1. Задаемся параметрами $\varepsilon > 0$, $\gamma > 0$, $0 < \tau < 1$ и начальным допустимым приближением L_0 . Вычисляем величину

$$\alpha_0 = \sigma(A - L_0 C).$$

2. На j -й итерации, имея величины L_j и α_j , находим градиент

$$H_j = f'_L(L_j, \alpha_j).$$

Если $\|H_j\| \leq \varepsilon$, то L_j принимаем за приближенное решение и завершаем работу алгоритма.

3. Делаем шаг градиентного метода

$$L_{j+1} = L_j - \gamma_j H_j,$$

при этом длину шага $\gamma_j > 0$ подбираем дроблением γ до выполнения условий:

- a. матрица $A - L_{j+1} C + \frac{\alpha_j}{2} I$ гурвицева;
- b. $f(L_{j+1}) \leq f(L_j) - \tau \gamma_j \|H_j\|^2$.
4. Для полученного приближения L_{j+1} решаем задачу минимизации $f(L_{j+1}, \alpha)$ по α и получаем α_{j+1} . Переходим к п. 2.

В алгоритме 10 величины $f_L(L, \alpha)$, $f'(\alpha)$ и $f''(\alpha)$ определяются следующим образом:

$$f'_L(L, \alpha) = 2 \left(\rho L - Y P C^T - \frac{1}{\alpha} Y (D_1 - L D_2) D_2^T \right),$$

$$f'(\alpha) = \text{tr } Y \left(P - \frac{1}{\alpha^2} (D_1 - L D_2) (D_1 - L D_2)^T \right),$$

$$f''(\alpha) = 2 \text{tr } Y \left(X + \frac{1}{\alpha^3} (D_1 - L D_2) (D_1 - L D_2)^T \right),$$

где матрицы P , Y и X являются решениями дискретных уравнений Ляпунова: (3.5.21),

$$\left(A - LC + \frac{\alpha}{2} I \right)^T Y + Y \left(A - LC + \frac{\alpha}{2} I \right) + C_1^T C_1 = 0$$

и

$$\begin{aligned} \left(A - LC + \frac{\alpha}{2} I \right) X + X \left(A - LC + \frac{\alpha}{2} I \right)^T + P - \\ - \frac{1}{\alpha^2} (D_1 - LD_2)(D_1 - LD_2)^T = 0. \end{aligned}$$

Упражнение 3.5.1. Дайте обоснование алгоритму 10, установив и доказав свойства минимизируемой функции и ее производных по аналогии с дискретным случаем (см. раздел 3.5.3).

Выбор пробного шага градиентного метода предлагается проводить из следующих соображений. Найдем для некоторого допустимого L решение P уравнения Ляпунова

$$(A - LC)P + P(A - LC)^T = -I.$$

Рассмотрим приращение по L :

$$L \rightarrow L - \gamma H, \quad H = f'_L(L, \alpha),$$

и выясним, для каких γ матрица P останется матрицей квадратичной функции Ляпунова для $A - (L - \gamma H)C$.

А именно, для этого должно выполняться условие

$$(A - (L - \gamma H)C)P + P(A - (L - \gamma H)C)^T \prec 0.$$

С учетом исходного уравнения имеем

$$\gamma(HCP + P(HC)^T) \prec I,$$

откуда

$$\gamma < \lambda_{\max}^{-1}(HCP + P(HC)^T).$$

3.5.5. Примеры и обсуждение

Рассмотрим линейную дискретную систему

$$\begin{aligned} x_{k+1} &= Ax_k + B_1 u_k + Gw_k, \\ y_k &= Cx_k + B_2 u_k + v_k, \end{aligned} \tag{3.5.22}$$

с состоянием $x_k \in \mathbb{R}^n$, начальным условием x_0 , входом $u_k \in \mathbb{R}^p$, наблюдаемым выходом $y_k \in \mathbb{R}^\ell$, шумом $w_k \in \mathbb{R}^m$ и ошибкой измерений $v_k \in \mathbb{R}^\ell$; здесь A, B_1, B_2, C и G — известные матрицы соответствующих размерностей; величины w_k и v_k предполагаются независимыми.

Применительно к системе (3.5.22) фильтр Калмана имеет следующий вид (в предположении, что величины w_k и v_k распределены нормально с нулевым математическим ожиданием и ковариационными матрицами Q и R соответственно).

Этап экстраполяции:

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} + B_1 u_k,$$

$$P_{k+1|k} = AP_{k|k}A^T + GQG^T.$$

Этап коррекции:

$$K_k = P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1},$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1} - B_2 u_k),$$

$$P_{k|k} = (I - K_k C)P_{k|k-1}.$$

Будем рассматривать три следующие постановки задачи.

1. Модель \mathcal{M}_1 со случайными помехами: шум w_k распределен по нормальному закону с нулевым математическим ожиданием и ковариационной матрицей Q , а погрешность измерений v_k имеет нормальное распределение с нулевым математическим ожиданием и ковариационной матрицей R :

$$w_k \sim \mathcal{N}(0, Q), \quad v_k \sim \mathcal{N}(0, R).$$

2. Модель \mathcal{M}_2 со случайными ограниченными помехами: величина w_k равномерно распределена на кубе $[-w, w]^m$, а величина v_k равномерно распределена на кубе $[-v, v]^\ell$:

$$w_k \sim \mathcal{U}([-w, w]^m), \quad v_k \sim \mathcal{U}([-v, v]^\ell).$$

3. Модель \mathcal{M}_3 с неслучайными ограниченными помехами: величины w_k и v_k могут принимать произвольные значения на кубах $[-w, w]^m$ и $[-v, v]^\ell$ соответственно:

$$|w_k|_\infty \leq w, \quad |v_k|_\infty \leq v.$$

В рамках этих моделей сравним работоспособность фильтра Калмана и гарантирующего подхода на следующих примерах (несмотря на то, что гарантирующие оценки не обоснованы для модели \mathcal{M}_1 , а фильтр Калмана — для модели \mathcal{M}_3).

На приводимых ниже рисунках показаны истинная траектория системы, ее наблюдение (если оно имеется) и оценки, предоставляемые калмановским и гарантирующим фильтром; в последнем случае также показана трубка, в которой заведомо содержится оценка при всех допустимых помехах.

Пример 3.5.1. Рассмотрим вагонетку, которая может двигаться без трения по бесконечным горизонтально расположенным рельсам. В начальный момент вагонетка находится в нулевом положении и на нее воздействуют внешние возмущения. Положение вагонетки измеряется каждые Δt секунд, при этом измерения неточны. Задача состоит в отслеживании положения s и скорости $\dot{s} = v$ вагонетки.

Соответствующая система представима в формате (3.5.22) при

$$\begin{aligned} x_k &= \begin{pmatrix} s \\ \dot{s} \end{pmatrix}, \quad x_0 = 0, \\ A &= \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix}, \quad B_1 = B_2 = 0, \quad C = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad G = \begin{pmatrix} (\Delta t)^2/2 \\ \Delta t \end{pmatrix}. \end{aligned}$$

1. В рамках модели \mathcal{M}_1 предполагается, что на k -м такте вагонетка движется с постоянным ускорением, распределенным по нормальному закону с нулевым математическим ожиданием и среднеквадратическим отклонением σ_x , а погрешность измерений имеет нормальное распределение с нулевым математическим ожиданием и среднеквадратическим отклонением σ_y :

$$w_k \sim \mathcal{N}(0, \sigma_x^2), \quad v_k \sim \mathcal{N}(0, \sigma_y^2).$$

Построим фильтр Калмана, а также воспользуемся гарантированным подходом, объединив помехи w_k и v_k в общий вектор возмущений

$$\tilde{w}_k = \begin{pmatrix} w_k \\ v_k \end{pmatrix}. \quad (3.5.23)$$

При этом матрицы D_1 и D_2 в системе (3.5.1) примут вид

$$D_1 = 3\sigma_x \sqrt{2} \begin{pmatrix} G & 0 \end{pmatrix}, \quad D_2 = 3\sigma_y \sqrt{2} \begin{pmatrix} 0 & 1 \end{pmatrix},$$

что позволит варьировать величины w_k и v_k независимо друг от друга на интервалах

$$|w_k| \leq 3\sigma_x, \quad |v_k| \leq 3\sigma_y.$$

Гарантирующий подход привел к оптимальным матрицам фильтра

$$L_*^{(1)} = \begin{pmatrix} 0,2359 \\ 0,1412 \end{pmatrix}$$

для координаты $x^{(1)} = s$ и

$$L_*^{(2)} = \begin{pmatrix} 0,1122 \\ 0,0386 \end{pmatrix}$$

для координаты $x^{(2)} = \dot{s}$.

Результаты сравнения с фильтром Калмана при

$$\Delta t = 0,1, \quad \sigma_x = 0,1, \quad \sigma_y = 0,5,$$

показаны на рис. 3.5.1 (с. 315).

2. В рамках модели \mathcal{M}_2 будем предполагать, что на k -м такте вагонетка движется с постоянным ускорением, равномерно распределенным на отрезке $[-a, a]$, а погрешность измерений равномерно распределена на отрезке $[-v, v]$:

$$w_k \sim \mathcal{U}(-a, a), \quad v_k \sim \mathcal{U}(-v, v).$$

Построим фильтр Калмана (при $\sigma_x = a/3$, $\sigma_y = v/3$), а также воспользуемся гарантирующим подходом для возмущения (3.5.23) и матриц

$$D_1 = a\sqrt{2} \begin{pmatrix} G & 0 \end{pmatrix}, \quad D_2 = v\sqrt{2} \begin{pmatrix} 0 & 1 \end{pmatrix}.$$

Гарантирующий подход привел к оптимальным матрицам фильтра

$$L_*^{(1)} = \begin{pmatrix} 0,1393 \\ 0,0489 \end{pmatrix}$$

для координаты $x^{(1)} = s$ и

$$L_*^{(2)} = \begin{pmatrix} 0,0574 \\ 0,0101 \end{pmatrix}$$

для координаты $x^{(2)} = \dot{s}$.

Результаты сравнения с фильтром Калмана при

$$\Delta t = 0,1, \quad a = 0,1, \quad v = 2,$$

показаны на рис. 3.5.2 (с. 316).

3. В рамках модели \mathcal{M}_3 будем предполагать, что w_k и v_k принимают произвольные значения на отрезках $[-a, a]$ и $[-v, v]$ соответственно:

$$|w_k| \leq a, \quad |v_k| \leq v.$$

Применим фильтр Калмана (для $\sigma_x = a$, $\sigma_y = v$), а также воспользуемся гарантирующим подходом для возмущения (3.5.23) и матриц

$$D_1 = a\sqrt{2} \begin{pmatrix} G & 0 \end{pmatrix}, \quad D_2 = v\sqrt{2} \begin{pmatrix} 0 & 1 \end{pmatrix}.$$

Гарантирующий подход привел к оптимальным матрицам фильтра

$$L_*^{(1)} = \begin{pmatrix} 0,1397 \\ 0,0492 \end{pmatrix}$$

для координаты $x^{(1)} = s$ и

$$L_*^{(2)} = \begin{pmatrix} 0,0574 \\ 0,0101 \end{pmatrix}$$

для координаты $x^{(2)} = \dot{s}$.

Результаты сравнения с фильтром Калмана при

$$\Delta t = 0,1, \quad a = 0,1, \quad v = 2$$

показаны на рис. 3.5.3 (с. 317). ▼

Пример 3.5.2. Следующий пример (см. подробнее [94]) связан с оцениванием движения снаряда по баллистической траектории при наличии внешних возмущений и доступных наблюдению (зашумленных) координатах. Соответствующая система имеет вид (3.5.22), где

$$x = \begin{pmatrix} s_x \\ s_y \\ v_x \\ v_y \end{pmatrix}$$

есть вектор состояния системы, состоящий из проекций координаты и скорости снаряда на горизонтальную и вертикальную оси, а

$$A = \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 - b & 0 \\ 0 & 0 & 0 & 1 - b \end{pmatrix}, \quad u = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -g\Delta t \end{pmatrix},$$

$$B_1 = G = I, \quad B_2 = 0, \quad C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Здесь Δt — интервал между измерениями; $0 < b \ll 1$ — коэффициент сопротивления воздуха; g — гравитационная постоянная; шум w_k распределен нормально с нулевым математическим ожиданием и ковариационной матрицей $Q_k \succ 0$, а помехи измерения v_k распределены нормально с нулевым математическим ожиданием и ковариационной матрицей $R_k \succ 0$. При этом

$$\Delta t = 0,1, \quad b = 10^{-4}, \quad g = 9,8, \quad Q = 0,1I, \quad R = 500I.$$

В качестве начальной возьмем точку

$$x_0 = \begin{pmatrix} 0 \\ 0 \\ 300 \\ 600 \end{pmatrix}.$$

1. В рамках «стандартной» модели \mathcal{M}_1 предполагается, что

$$w_k \sim \mathcal{N}(0, \sigma_x^2 I), \quad \sigma_x^2 = 0,1,$$

$$v_k \sim \mathcal{N}(0, \sigma_y^2 I), \quad \sigma_y^2 = 500.$$

Как и в предыдущем примере, для того чтобы воспользоваться гарантирующим подходом, объединим помехи w_k и v_k в общий вектор возмущений, а матрицы D_1 и D_2 в системе (3.5.1) примут вид

$$D_1 = 3\sigma_x \sqrt{2} (G \quad 0), \quad D_2 = 3\sigma_y \sqrt{2} (0 \quad I),$$

что позволит варьировать величины w_k и v_k независимо друг от друга на интервалах

$$|w_k| \leq 3\sigma_x, \quad |v_k| \leq 3\sigma_y.$$

Гарантирующий подход доставил следующие оптимальные матрицы фильтра для каждой из четырех координат (здесь и далее в матрицах фильтра обнулены элементы, по абсолютной величине не превосходящие 10^{-6}):

$$L_*^{(1)} = \begin{pmatrix} 0,5946 & 0 \\ 0 & 0,6822 \\ 0,8467 & 0 \\ 0 & 1,0590 \end{pmatrix}, \quad L_*^{(2)} = \begin{pmatrix} 0,7277 & 0 \\ 0 & 0,6340 \\ 1,2376 & 0 \\ 0 & 0,9879 \end{pmatrix},$$

$$L_*^{(3)} = \begin{pmatrix} 0,0971 & 0 \\ 0 & 0,1389 \\ 0,0284 & 0 \\ 0 & 0,0456 \end{pmatrix}, \quad L_*^{(4)} = \begin{pmatrix} 0,1393 & 0 \\ 0 & 0,0975 \\ 0,0459 & 0 \\ 0 & 0,0285 \end{pmatrix}.$$

Результаты сравнения с фильтром Калмана для координат s_y и v_y показаны на рис. 3.5.4 (с. 318).

Обратим внимание, что поведение траекторий оценок на начальном участке обусловлено тем, что фактическая начальная точка X_0 оказалась вне минимального инвариантного эллипсоида для невязки. Однако в силу того, что инвариантный эллипсоид обладает свойством притягиваемости, после нескольких шагов гарантированные оценки начинают охватывать истинную траекторию.

2. В рамках модели \mathcal{M}_2 будем предполагать, что

$$w_k \sim \mathcal{U}(-w, w), \quad v_k \sim \mathcal{U}(-v, v),$$

где $w = 3\sigma_x$, $v = 3\sigma_y$.

Поскольку в этом случае

$$D_1 = w\sqrt{2} \begin{pmatrix} G & 0 \end{pmatrix}, \quad D_2 = v\sqrt{2} \begin{pmatrix} 0 & I \end{pmatrix},$$

то гарантированный подход приводит к тем же матрицам фильтра, что и в модели \mathcal{M}_2 . Результаты его сравнения с фильтром Калмана для координат s_y и v_y показаны на рис. 3.5.5 (с. 319).

3. В рамках модели \mathcal{M}_3 будем предполагать, что

$$|w_k| \leq \sigma_x, \quad |v_k| \leq \sigma_y.$$

Гарантирующий подход при

$$D_1 = \sigma_x \sqrt{2} \begin{pmatrix} G & 0 \end{pmatrix}, \quad D_2 = \sigma_y \sqrt{2} \begin{pmatrix} 0 & I \end{pmatrix}$$

привел к оптимальным матрицам фильтра

$$L_*^{(1)} = \begin{pmatrix} 0,6362 & 0 \\ 0 & 0,7256 \\ 0,9779 & 0 \\ 0 & 1,2126 \end{pmatrix}, \quad L_*^{(2)} = \begin{pmatrix} 0,6697 & 0 \\ 0 & 0,5747 \\ 1,0517 & 0 \\ 0 & 0,8206 \end{pmatrix},$$

$$L_*^{(3)} = \begin{pmatrix} 0,0971 & 0 \\ 0 & 0,1389 \\ 0,0284 & 0 \\ 0 & 0,0456 \end{pmatrix}, \quad L_*^{(4)} = \begin{pmatrix} 0,1393 & 0 \\ 0 & 0,0975 \\ 0,0459 & 0 \\ 0 & 0,0285 \end{pmatrix}.$$

Результаты его сравнения с фильтром Калмана для координат s_y и v_y показаны на рис. 3.5.6 (с. 320). ▼

Как видно из рассмотренных примеров, и фильтр Калмана, и гарантирующий фильтр, с одной стороны, не слишком сильно отличаются по своим результатам, а с другой — во всех трех моделях они вполне работоспособны. При этом, как ожидалось, для гауссовых помех фильтр Калмана дает немного (но не кардинально) лучшее поведение оценки по сравнению с гарантирующим фильтром, тогда как для неслучайных ограниченных помех преимущество остается за гарантирующим фильтром.

Отметим также, что ширина трубки, в которую гарантированно заключена соответствующая оценка, довольно сильно завышена, что хорошо видно на рисунках, — это характерное поведение методов гарантированного оценивания, направленных на противодействие «наихудшей» из возможных реализаций неопределенности.

Наконец, гарантирующий фильтр позволяет получить равномерную оценку точности фильтрации; здесь уместно обратить внимание на поведение гарантирующей и калмановской оценок на начальном участке траектории (при этом последняя также имеет явно выраженный всплеск).

3.6. Синтез ПИД-регуляторов

Теория ПИД-регуляторов имеет 80-летнюю историю, восходя к работе [135] Дж. Циглера и Н. Никольса 1942 г. С тех пор появилось множество работ, посвященных теории и практике их настройки, были предложены различные принципы их настройки (см. например, [45]). Здесь можно упомянуть монографии [30, 31, 35, 46] и многие другие. Однако до настоящего времени есть не так много работ, где формулируются явные критерии оптимальности ПИД-регуляторов, например такие, как H_∞ -оптимальность [78, 93].

В целом регуляторы низкого порядка настраивают по самым разным критериям, используя при этом, как правило, подбор, прямой поиск, перебор по сетке и т. п., см., например, [52, 55]. В этом разделе рассмотрим новый подход к настройке ПИД-регуляторов, включающий постановку задачи, решение которой определяет ПИД-регулятор, и алгоритм ее решения путем явного выписывания градиента для квадратичного функционала и применения градиентного метода.

В отличие от упомянутых выше критериев, будем рассматривать постановку, в некотором смысле близкую к линейно-квадратичной (см. раздел 3.3). А именно, предполагается, что неопределенность содержится не во входах системы, а в начальных условиях, при этом кри-

терий качества близок по своей структуре к линейно-квадратичной задаче: качество оценивается по квадратичному критерию от выхода системы. Таким образом, искомый ПИД-регулятор настраивается против неопределенности в начальных условиях так, чтобы выход системы был равномерно малым (в квадратичном смысле).

Помимо малости выхода, естественно стремиться к тому, чтобы синтезированный ПИД-регулятор удовлетворял и инженерным критериям качества. Так, если замкнутая система окажется близкой к границе устойчивости, это приведет к ее неудовлетворительной реакции на внешнее возмущение. Поэтому вполне естественно требовать, чтобы ПИД-регулятор дополнительно гарантировал замкнутой системе некоторую (заданную) степень устойчивости.

Рассматриваемый подход позволяет, с одной стороны, конструктивно решать задачи настройки и оптимизации параметров регулятора, а с другой — он предоставляет «хорошие» по обычным инженерным показателям регуляторы. Как показывают примеры, предлагаемая рекуррентная процедура является весьма эффективной и приводящей ко вполне удовлетворительным ПИД-регуляторам.

3.6.1. Постановка задачи и ее сведение к параметрической линейно-квадратичной задаче

Рассмотрим линейную непрерывную систему управления

$$\begin{aligned}\dot{x} &= Ax + bu, \quad x(0) = x_0, \\ y &= c^T x,\end{aligned}\tag{3.6.1}$$

где $A \in \mathbb{R}^{n \times n}$, $b, c \in \mathbb{R}^n$, с состоянием $x(t) \in \mathbb{R}^n$, скалярным выходом $y(t) \in \mathbb{R}$ и скалярным управлением $u(t) \in \mathbb{R}$ в виде *ПИД-регулятора* (см. подробнее [B.8](#))

$$u(t) = -k_P y(t) - k_I \int_0^t y(\tau) d\tau - k_D \dot{y}(t)\tag{3.6.2}$$

с некоторыми числовыми параметрами k_P , k_I и k_D .

Целью является определение вектора

$$K = \begin{pmatrix} k_P \\ k_I \\ k_D \end{pmatrix}$$

параметров обратной связи (3.6.2), которая

- а) доставляет замкнутой системе степень устойчивости $\sigma > 0$ и
- б) минимизирует квадратичный функционал

$$f(K) = \mathbb{E}_{\sim x(0)} \int_0^{\infty} y^2(t) dt + \rho \|K\|^2, \quad \rho > 0. \quad (3.6.3)$$

Вторая компонента в (3.6.3) представляет собой штраф за величину управления (при этом коэффициент $\rho > 0$ регулирует его важность). Ее наличие позволяет избежать появления больших значений коэффициентов ПИД-регулятора.

Далее будем предполагать, что начальные условия $x(0)$ распределены с нулевым средним и ковариационной матрицей Σ .

Рассмотрим вспомогательную скалярную переменную z :

$$\dot{z} = y, \quad z(0) = 0.$$

Тогда, вводя расширенный вектор состояния

$$g = \begin{pmatrix} x \\ z \end{pmatrix} \in \mathbb{R}^{n+1},$$

системе (3.6.1) можно придать эквивалентный вид

$$\begin{aligned} \dot{g} &= \begin{pmatrix} A & 0 \\ c^T & 0 \end{pmatrix} g + \begin{pmatrix} b \\ 0 \end{pmatrix} u, \quad g(0) = \begin{pmatrix} x_0 \\ 0 \end{pmatrix}, \\ y &= \begin{pmatrix} c^T & 0 \end{pmatrix} g. \end{aligned} \quad (3.6.4)$$

При этом согласно (3.6.1), (3.6.2) имеем

$$\begin{aligned} u &= -k_P y(t) - k_I \int_0^t y(\tau) d\tau - k_D \dot{y}(t) = \\ &= -k_P c^T x - k_I z - k_D c^T \dot{x} = -k_P c^T x - k_I z - k_D c^T (Ax + bu) = \\ &= -k_P (c^T \ 0) g - k_I (0 \ 1) g - k_D (c^T A \ 0) g - k_D c^T bu, \end{aligned}$$

откуда

$$(1 + k_D c^T b)u = -k_P (c^T \ 0) g - k_I (0 \ 1) g - k_D (c^T A \ 0) g,$$

или

$$u = -\frac{k_P}{1+k_D c^T b} (c^T \quad 0) g - \frac{k_I}{1+k_D c^T b} (0 \quad 1) g - \frac{k_D}{1+k_D c^T b} (c^T A \quad 0) g. \quad (3.6.5)$$

Если ввести новые переменные

$$k_1 = \frac{k_P}{1+k_D c^T b}, \quad k_2 = \frac{k_I}{1+k_D c^T b}, \quad k_3 = \frac{k_D}{1+k_D c^T b},$$

то (3.6.5) примет вид

$$u = -(k_1 c^T + k_3 c^T A \quad k_2) g. \quad (3.6.6)$$

Замыкая систему (3.6.4) обратной связью (3.6.6), приходим к замкнутой системе

$$\dot{g} = \begin{pmatrix} A - k_1 b c^T - k_3 b c^T A & -k_2 b \\ c^T & 0 \end{pmatrix} g, \quad g(0) = \begin{pmatrix} x_0 \\ 0 \end{pmatrix}, \quad (3.6.7)$$

которой можно придать вид

$$\dot{g} = (A_0 + k_1 A_1 + k_2 A_2 + k_3 A_3) g, \quad g(0) = \begin{pmatrix} x_0 \\ 0 \end{pmatrix},$$

где

$$A_0 = \begin{pmatrix} A & 0 \\ c^T & 0 \end{pmatrix}, \quad A_1 = \begin{pmatrix} -b c^T & 0 \\ 0 & 0 \end{pmatrix},$$

$$A_2 = \begin{pmatrix} 0 & -b \\ 0 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} -b c^T A & 0 \\ 0 & 0 \end{pmatrix}.$$

При этом исходные параметры ПИД-регулятора восстанавливаются единственным образом:

$$k_P = \frac{k_1}{1-k_3 c^T b}, \quad k_I = \frac{k_2}{1-k_3 c^T b}, \quad k_D = \frac{k_3}{1-k_3 c^T b}.$$

Замечание 3.6.1. Обратим внимание, что если векторы b и c ортогональны, то

$$k_1 = k_P, \quad k_2 = k_I, \quad k_3 = k_D.$$

Как показывают многочисленные примеры, это вполне типичная ситуация, так что — для упрощения записи — все дальнейшие выкладки относятся именно к этому случаю. \blacktriangledown

Для удобства введем обозначение:

$$\{A, K\} \doteq k_1 A_1 + k_2 A_2 + k_3 A_3.$$

Для того чтобы гарантировать желаемую степень устойчивости $\sigma > 0$ замкнутой системы, введем в ее матрицу компоненту σI :

$$\dot{g} = (A_0 + \{A, K\} + \sigma I)g. \quad (3.6.8)$$

В самом деле, для стабилизирующей систему (3.6.8) регулятора K матрица $A_0 + \{A, K\} + \sigma I$ является гурвицевой. Отсюда непосредственно вытекает, что

$$\max_i \operatorname{Re} \lambda_i(A_0 + \{A, K\} + \sigma I) = \max_i \operatorname{Re} \lambda_i(A_0 + \{A, K\}) + \sigma < 0,$$

то есть степень устойчивости исходной системы не меньше σ :

$$\max_i \operatorname{Re} \lambda_i(A_0 + \{A, K\}) < -\sigma.$$

Продолжим и перейдем к преобразованию функционала (3.6.3). Поскольку

$$x = (I \ 0) g,$$

то начальные условия $g(0) = \begin{pmatrix} x_0 \\ 0 \end{pmatrix}$ будут распределены с нулевым средним и ковариационной матрицей

$$\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}.$$

По лемме Беллмана 3.1.3 для системы (3.6.7) имеем

$$\begin{aligned} f(K) &= \mathbb{E}_{\sim x(0)} \int_0^\infty y^2(t) dt + \rho \|K\|^2 = \\ &= \mathbb{E}_{\sim g(0)} \int_0^\infty g^T(t) \begin{pmatrix} cc^T & 0 \\ 0 & 0 \end{pmatrix} g(t) dt + \rho \|K\|^2 = \\ &= \mathbb{E}_{\sim g(0)} g^T(0) Q g(0) + \rho \|K\|^2 = \operatorname{tr} Q \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} + \rho \|K\|^2. \end{aligned}$$

Здесь $Q \in \mathbb{S}^{n+1}$ — решение уравнения Ляпунова

$$A_K^T Q + Q A_K = - \begin{pmatrix} cc^T & 0 \\ 0 & 0 \end{pmatrix}, \quad (3.6.9)$$

где

$$A_K = A_0 + \{A, K\} + \sigma I.$$

Итак, возникает следующая задача матричной оптимизации

$$\min f(K), \quad f(K) = \operatorname{tr} Q \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} + \rho \|K\|^2,$$

при ограничении (3.6.9), где оптимизация проводится по матричной переменной $0 \prec Q \in \mathbb{S}^{n+1}$ и векторной переменной $K \in \mathbb{R}^3$.

Сделаем следующее

Предположение 3.6.1. Пусть известен регулятор

$$K_0 = \begin{pmatrix} k_1^0 \\ k_2^0 \\ k_3^0 \end{pmatrix},$$

стабилизирующий систему с заданной степенью устойчивости σ , то есть такой, что матрица $A_{K_0} = A_0 + \{A, K_0\} + \sigma I$ гурвицева.

Перейдем к описанию свойств функции $f(K)$.

Лемма 3.6.1. *Функция $f(K)$ определена и положительна на множестве \mathcal{S} стабилизирующих регуляторов.*

Действительно, если матрица A_K гурвицева, то решение $Q \succcurlyeq 0$ уравнения Ляпунова (3.6.9) существует; тем самым определена функция $f(K) > 0$. Множество ее определения \mathcal{S} может быть невыпуклым и несвязным, причем его границы могут быть негладкими.

Перейдем к вычислению градиента функции $f(K)$.

Лемма 3.6.2. *Функция $f(K)$ определена на множестве стабилизирующих обратных связей K . На этом допустимом множестве она дифференцируема, причем градиент дается выражениями*

$$\frac{\partial f(K)}{\partial k_i} = 2 \operatorname{tr} Y Q A_i + 2\rho k_i, \quad i = 1, 2, 3, \quad (3.6.10)$$

где матрица Y является решением уравнения Ляпунова

$$A_K Y + Y A_K^\top + \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} = 0. \quad (3.6.11)$$

Доказательство. В уравнении (3.6.9) придадим величине K приращение ΔK и обозначим соответствующее приращение Q через ΔQ :

$$A_{K+\Delta K}^T(Q + \Delta Q) + (Q + \Delta Q)A_{K+\Delta K} = - \begin{pmatrix} cc^T & 0 \\ 0 & 0 \end{pmatrix},$$

или

$$(A_K + \{A, \Delta K\})^T(Q + \Delta Q) + (Q + \Delta Q)(A_K + \{A, \Delta K\}) = - \begin{pmatrix} cc^T & 0 \\ 0 & 0 \end{pmatrix},$$

откуда после линеаризации имеем

$$A_K^T \Delta Q + \Delta Q A_K + Q \{A, \Delta K\} + \{A, \Delta K\}^T Q = 0. \quad (3.6.12)$$

Теперь вычислим приращение функционала $f(K)$, линеаризуя соответствующие величины:

$$\begin{aligned} \Delta f(K) &= \text{tr}(Q + \Delta Q) \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} + \rho \|K + \Delta K\|^2 - \text{tr} Q \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} - \rho \|K\|^2 = \\ &= \text{tr} \Delta Q \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} + 2\rho(K, \Delta K). \end{aligned}$$

Рассмотрим уравнение Ляпунова (3.6.11), двойственное к (3.6.12). По лемме Б.8 из уравнений (3.6.12) и (3.6.11) имеем

$$\Delta f(K) = \text{tr} \Delta Q \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} + 2\rho(K, \Delta K) = 2 \text{tr} Y Q \{A, \Delta K\} + 2\rho(K, \Delta K),$$

так что

$$df(K) = 2 \text{tr} Y Q \sum_{i=1}^3 A_i dk_i + 2\rho(K, dK),$$

откуда имеем (3.6.10). ■

Подытоживая сказанное, можно воспользоваться следующим итеративным подходом к решению задачи; в его основе лежит применение градиентного метода по матричной переменной K . Приведем принципиальную схему алгоритма.

Алгоритм 11.

1. Задаемся параметрами $\varepsilon > 0$, $\gamma > 0$, $0 < \tau < 1$ и начальным стабилизирующим приближением K_0 .

2. На j -й итерации имеем регулятор K_j . Вычисляем матрицу

$$A_{K_j} = A_0 + \{A, K_j\},$$

решаем уравнения (3.6.9), (3.6.11) и находим матрицы Q и Y ; вычисляем градиент

$$H_j = f'(K_j)$$

из уравнения (3.6.10). Если $\|H_j\| \leq \varepsilon$, то K_j принимаем за приближенное решение и завершаем работу алгоритма.

3. Делаем шаг градиентного метода

$$K_{j+1} = K_j - \gamma_j H_j.$$

Длину шага $\gamma_j > 0$ подбираем дроблением γ до выполнения условий:

- а. матрица $A_0 + \{A, K_{j+1}\} + \sigma I$ гурвицева;
- б. $f(K_{j+1}) \leq f(K_j) - \tau \gamma_j \|H_j\|^2$.

4. Переходим к п. 2.

Сделаем несколько замечаний. Прежде всего, нетривиальным моментом является выбор начального стабилизирующего регулятора K_0 . Здесь можно прибегнуть к помощи D -разбиения [53] или воспользоваться методами перебора, см. также [55]. В некоторых случаях может оказаться полезным следующий прием: для некоторой положительно определенной матрицы Q попытаться разрешить неравенство Ляпунова $A_{K_0}^T Q + Q A_{K_0} \prec 0$ относительно K_0 . Также часто в практических задачах требуется улучшить качество уже имеющегося ПИД-регулятора путем настройки его параметров. Наконец, если исходная система управления является устойчивой (как часто и бывает на практике), вопрос выбора начального регулятора решается очевидным образом.

Еще одним важным моментом является выбор пробного шага градиентного метода. Весьма перспективным является его выбор из следующих соображений. Найдем для некоторого стабилизирующего регулятора K_j решение Q уравнения Ляпунова

$$(A_0 + \sigma I + \{A, K_j\})^T Q + Q(A_0 + \sigma I + \{A, K_j\}) = -I.$$

Рассмотрим приращение по K :

$$K_j \rightarrow K_j - \gamma H_j, \quad H_j = f'(K_j),$$

и найдем, для каких γ матрица Q останется матрицей квадратичной функции Ляпунова для $A_{K_j - \gamma H_j} = A_0 + \sigma I + \{A, K_j - \gamma H_j\}$, то есть

$$(A_0 + \sigma I + \{A, K_j - \gamma H_j\})^T Q + Q(A_0 + \sigma I + \{A, K_j - \gamma H_j\}) \prec 0.$$

С учетом исходного уравнения имеем

$$\gamma(-\{A, H_j\}^T Q - Q\{A, H_j\}) \prec I,$$

откуда

$$\gamma < \lambda_{\max}^{-1}(-\{A, H_j\}^T Q - Q\{A, H_j\}).$$

3.6.2. Примеры

Рассматриваемые далее примеры взяты из статьи [73]. Всюду в этом разделе будем полагать $\Sigma = I$.

Пример 3.6.1. Рассмотрим передаточную функцию

$$G(s) = \frac{1}{(1+s)(1+\alpha s)(1+\alpha^2 s)(1+\alpha^3 s)}, \quad \alpha = 0,5.$$

MATLAB-процедура `tf2ss` доставляет матрицы системы в пространстве состояний:

$$A = \begin{pmatrix} -15 & -70 & -120 & -64 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad c = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 64 \end{pmatrix}.$$

Пусть

$$\sigma = 0,25, \quad \rho = 10.$$

При выборе в качестве начального стабилизирующего регулятора

$$K_0 = \begin{pmatrix} 9,5717 \\ 4,8538 \\ 8,0028 \end{pmatrix},$$

оптимизационная процедура приводит к регулятору

$$K_* = \begin{pmatrix} 7,6296 \\ 3,4331 \\ 3,1795 \end{pmatrix}, \quad \|K_*\| = 8,9502,$$

который доставляет интегральной части функционала $f(K)$ значение $3,5327 \cdot 10^3$.

Возьмем теперь начальное приближение

$$K'_0 = \begin{pmatrix} 4,3141 \\ 9,1065 \\ 1,8185 \end{pmatrix};$$

в результате получим регулятор

$$K'_* = \begin{pmatrix} 7,6265 \\ 3,4327 \\ 3,1782 \end{pmatrix}, \quad \|K'_*\| = 8,9469,$$

и значение функционала, равное $f(K'_*) = 3,5333 \cdot 10^3$.

Как видно, значения функционала и нормы получившихся регуляторов отличаются на доли процента.

Далее, передаточная функция ПИД-регулятора с коэффициентами K_* имеет вид

$$G_{\text{PID}}(s) = 7,6296 + \frac{3,4331}{s} + 3,1795s.$$

Система, замкнутая ПИД-регулятором K_* , является устойчивой по критерию Найквиста; ее минимальный запас устойчивости по модулю составляет 7,41 дБ, а по фазе — $27,5^\circ$, см. рис. 3.6.1.

Сравним полученный ПИД-регулятор с регуляторами, полученными по методу Циглера–Никольса (ZN) [135], гармонического поиска (Harmony Search, HS) [90], улучшенной версии пчелиного алгоритма (Improved Bees' Algorithm, IBA) [123] и алгоритма пчелиной колонии (Artificial Bee Colony, ABC) [98]. Соответствующие результаты представлены в таблице 3.1.

На рис. 3.6.2 показана динамика выхода $y(t)$ рассматриваемой системы при некотором начальном условии

$$x_0 = \begin{pmatrix} -0,5715 \\ -0,1249 \\ 0,6635 \\ 0,4664 \end{pmatrix}$$

из единичного шара: при замыкании найденным ПИД-регулятором K_* (жирная линия) и регуляторами из таблицы 3.1.

На рис. 3.6.3 показана динамика выхода $y(t)$ рассматриваемой системы, замкнутой найденным ПИД-регулятором K_* (жирная линия)

Таблица 3.1. Сравнение ПИД-регуляторов из примера 3.6.1 по интегральной части функционала и запасам устойчивости.

	k_P	k_I	k_D
Алгоритм 11	7,6296	3,4331	3,1795
ZN	3,9706	3,5749	1,1026
HS	2,8206	1,8022	1,4330
IBA	2,7852	1,7873	1,4157
ABC	2,8458	1,8278	1,4535
	$\mathbb{E}_{\sim x(0)} \int_0^{\infty} y^2(t) dt$	G_m (дБ)	P_m (град.)
Алгоритм 11	$2,7303 \cdot 10^3$	7,41	27,5
ZN	$4,3752 \cdot 10^3$	12,4	35,0
HS	$3,6654 \cdot 10^3$	15,6	68,0
IBA	$3,6872 \cdot 10^3$	15,7	68,5
ABC	$3,6545 \cdot 10^3$	15,5	67,8

и регуляторами из таблицы 3.1 при единичном ступенчатом возмущении (и нулевом начальном условии).

Как видно, синтезированный ПИД-регулятор вполне удовлетворителен по своим характеристикам. ▼

Пример 3.6.2. Рассмотрим передаточную функцию

$$G(s) = \frac{1 - \alpha s}{(s + 1)^2}, \quad \alpha = 0,1.$$

MATLAB-процедура `tf2ss` доставляет матрицы системы в пространстве состояний:

$$A = \begin{pmatrix} -3 & -3 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad c = \begin{pmatrix} 0 \\ -0,1 \\ 1 \end{pmatrix}.$$

Полагая $\rho = 10$, $\sigma = 0,1$ и выбрав

$$K_0 = \begin{pmatrix} 3,8710 \\ 6,1308 \\ 8,9234 \end{pmatrix}$$

в качестве начального стабилизирующего регулятора, при завершении

оптимизационной процедуры получаем регулятор

$$K_* = \begin{pmatrix} 0,4010 \\ 0,1870 \\ 0,0382 \end{pmatrix},$$

который доставляет интегральной части функционала $f(K)$ значение 8,5440, см. рис. 3.6.4.

При этом передаточная функция ПИД-регулятора с коэффициентами K_* имеет вид

$$G_{\text{PID}}(s) = 0,4010 + \frac{0,1870}{s} + 0,0382s.$$

Замкнутая система с ПИД-регулятором K_* является устойчивой по критерию Найквиста; ее минимальный запас устойчивости по модулю составляет 21,1 дБ, а по фазе — 78,9°, см. рис. 3.6.5.

▼

Пример 3.6.3. Рассмотрим передаточную функцию

$$G(s) = \frac{(s+6)^2}{s(s+1)^2(s+36)}.$$

MATLAB-процедура `tf2ss` доставляет матрицы системы в пространстве состояний:

$$A = \begin{pmatrix} -39 & -111 & -109 & -36 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad c = \begin{pmatrix} 0 \\ 1 \\ 12 \\ 36 \end{pmatrix}.$$

При

$$\rho = 10, \quad \sigma = 0,1$$

и выборе в качестве начального стабилизирующего регулятора

$$K_0 = \begin{pmatrix} 10,1955 \\ 9,8265 \\ 2,4392 \end{pmatrix}$$

оптимизационная процедура приводит к регулятору

$$K_* = \begin{pmatrix} 6,7538 \\ 1,2114 \\ 2,2965 \end{pmatrix},$$

доставляющему интегральной части минимизируемого функционала значение $1,4774 \cdot 10^3$.

Замкнутая система с ПИД-регулятором K_* является устойчивой по критерию Найквиста; ее минимальный запас устойчивости по модулю бесконечен, а по фазе равен $55,2^\circ$, см. рис. 3.6.6.



Пример 3.6.4. Вернемся к примеру 3.6.1 и положим в нем $\alpha = 1$. Оптимизационная процедура (при $\rho = 0,5$, $\sigma = 0,07$) завершается нахождением регулятора

$$G_{\text{PID}}(s) = 1,7408 + \frac{0,2849}{s} + 1,8615s,$$

см. рис. 3.6.7.

Он доставляет интегральной части минимизируемого функционала значение 8,9878 и обладает минимальным запасом устойчивости по модулю 13,2 дБ, а по фазе — $76,2^\circ$, см. рис. 3.6.8.

Сравнение найденного регулятора с тремя ПИД/ПИ-регуляторами, предложенными для этого же примера в работах [103, 106, 121], представлено в таблице 3.2.

Таблица 3.2. Сравнение ПИД-регуляторов из примера 3.6.4 по интегральной части функционала и запасам устойчивости.

	k_P	k_I	k_D
Алгоритм 11	1,7408	0,2849	1,8615
Из работы [121]	0,925	0,9	2,86
Из работы [106]	0,83	0,318	0,3569
Из работы [103]	1,031	0,3529	0
	$\mathbb{E}_{\sim x(0)} \int_0^\infty y^2(t)dt$	G_m (дБ)	P_m (град.)
Алгоритм 11	8,9878	13,2	76,2
Из работы [121]	9,6715	14,87	43,0
Из работы [106]	10,7752	14,34	62,5
Из работы [103]	12,5559	17,25	96,2

На рис. 3.6.9 показана динамика изменения выхода $y(t)$ рассматри-

ваемой системы при некотором начальном условии

$$x_0 = \begin{pmatrix} -0,2456 \\ -0,6435 \\ -0,6921 \\ -0,2161 \end{pmatrix}$$

из единичного шара: при замыкании найденным ПИД-регулятором K_* (жирная линия) и регуляторами из таблицы 3.2.

На рис. 3.6.10 показана динамика выхода $y(t)$ рассматриваемой системы, замкнутой найденным ПИД-регулятором K_* (жирная линия) и регуляторами из таблицы 3.2 при единичном ступенчатом возмущении (и нулевом начальном условии).



Пример 3.6.5. Рассмотрим передаточную функцию

$$G(s) = \frac{10s^3 + 9s^2 + 362,4s + 36,16}{2s^5 + 2,7255s^4 + 138,4292s^3 + 156,471s^2 + 637,6472s + 360,1779}$$

из статьи [78].

Алгоритм 11 (при $\rho = 0,001$, $\sigma = 0,01$) дает ПИД-регулятор

$$G_{\text{PID}}(s) = 201,1057 + \frac{75,9364}{s} + 6,2735s,$$

который доставляет интегральной части минимизируемого функционала значение $1,0614 \cdot 10^3$ и обладает бесконечным запасом устойчивости по модулю и запасом по фазе $52,2^\circ$, см. рис. 3.6.11.

Его сравнение с тремя ПИД-регуляторами, предложенными для рассматриваемой системы в работе [93], представлено в таблице 3.3.

Синтезированный ПИД-регулятор обладает весьма удовлетворительными характеристиками. На рис. 3.6.12 показана динамика изменения выхода $y(t)$ рассматриваемой системы при некотором начальном условии

$$x_0 = \begin{pmatrix} -0,5228 \\ -0,4387 \\ -0,3609 \\ -0,1148 \\ -0,6251 \end{pmatrix}$$

из единичного шара при замыкании найденным ПИД-регулятором K_* (жирная линия) и регуляторами из работы [93].

Таблица 3.3. Сравнение ПИД-регуляторов из примера 3.6.5 по интегральной части функционала и запасам устойчивости.

	k_P	k_I	k_D
Алгоритм 11	201,1057	75,9364	6,2735
ПИД-регулятор 1 из [93]	185	2986	9
ПИД-регулятор 2 из [93]	20	800	9
ПИД-регулятор 3 из [93]	19	200	9
	$\mathbb{E}_{\sim x(0)} \int_0^{\infty} y^2(t) dt$	G_m (дБ)	P_m (град.)
Алгоритм 11	$1,0614 \cdot 10^3$	∞	52,2
ПИД-регулятор 1 из [93]	$1,7338 \cdot 10^3$	-10,2	60,8
ПИД-регулятор 2 из [93]	$1,3503 \cdot 10^4$	-7,79	87,5
ПИД-регулятор 3 из [93]	$1,3769 \cdot 10^4$	-10	78,6

На рис. 3.6.13 показана динамика выхода $y(t)$ рассматриваемой системы, замкнутой найденным ПИД-регулятором K_* (жирная линия) и регуляторами из таблицы 3.3 при единичном ступенчатом возмущении (и нулевом начальном условии).

Как и в предыдущих примерах, синтезированный ПИД-регулятор обладает вполне удовлетворительными характеристиками с точки зрения инженерных требований. ▼

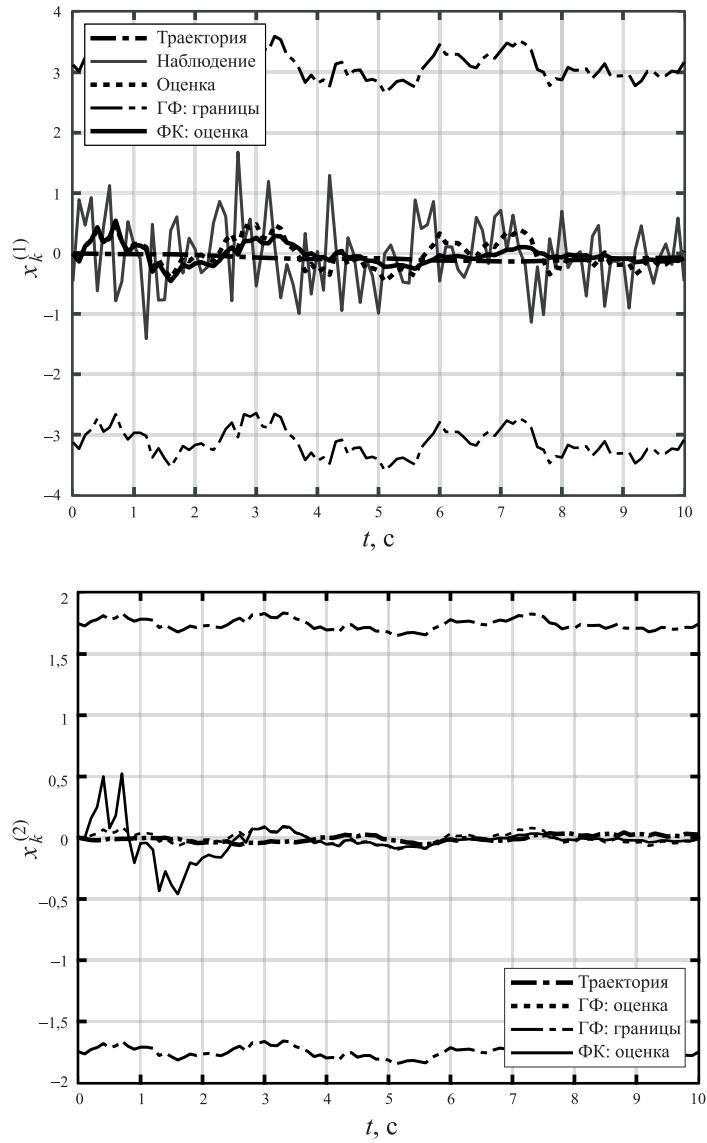


Рис. 3.5.1. Динамика координат и их оценок в примере 3.5.1 для модели \mathcal{M}_1 .

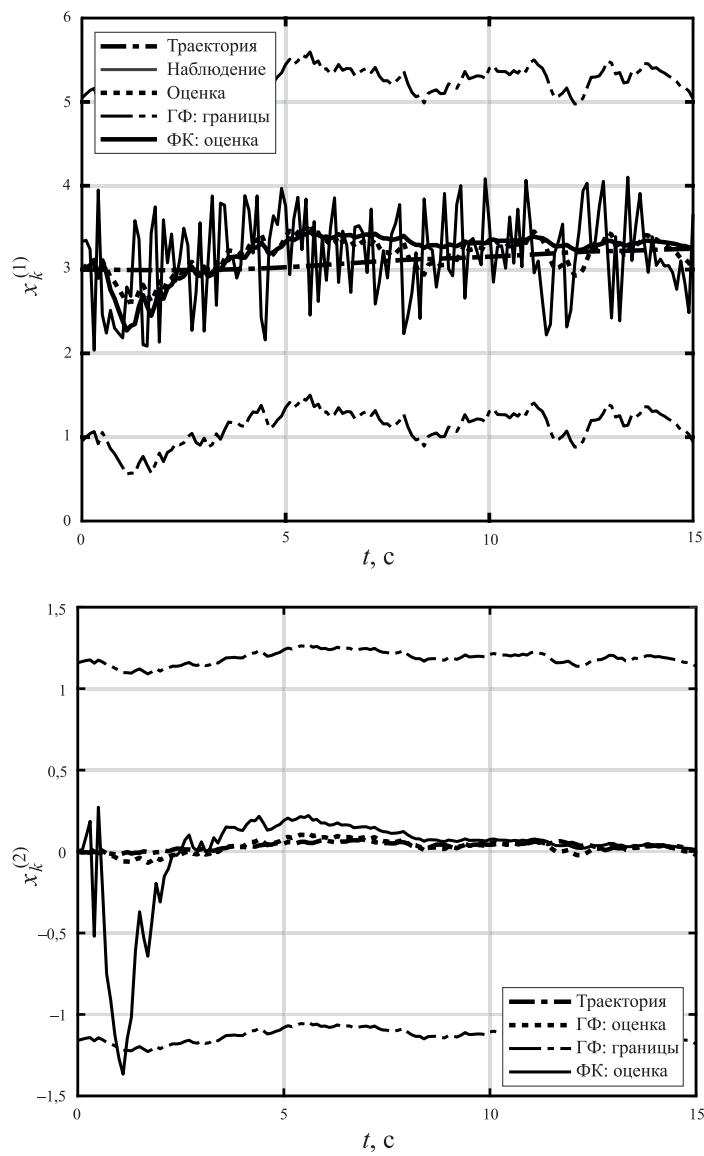


Рис. 3.5.2. Динамика координат и их оценок в примере 3.5.1 для модели \mathcal{M}_2 .

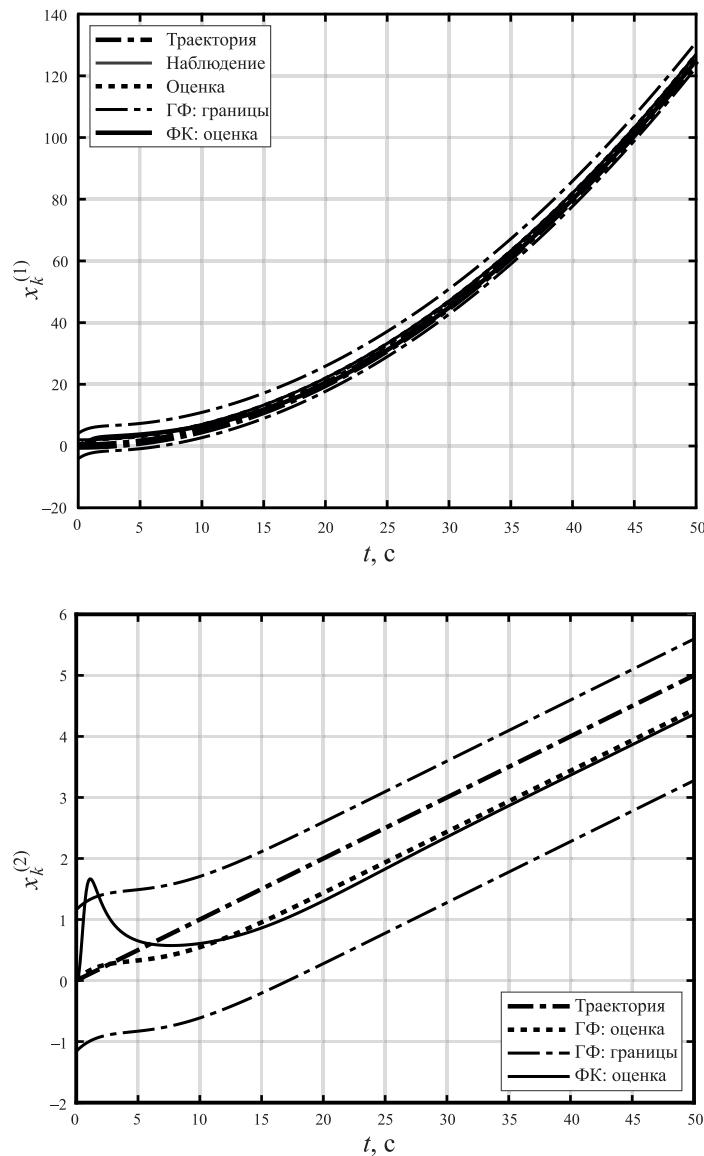


Рис. 3.5.3. Динамика координат и их оценок в примере 3.5.1 для модели \mathcal{M}_3 .

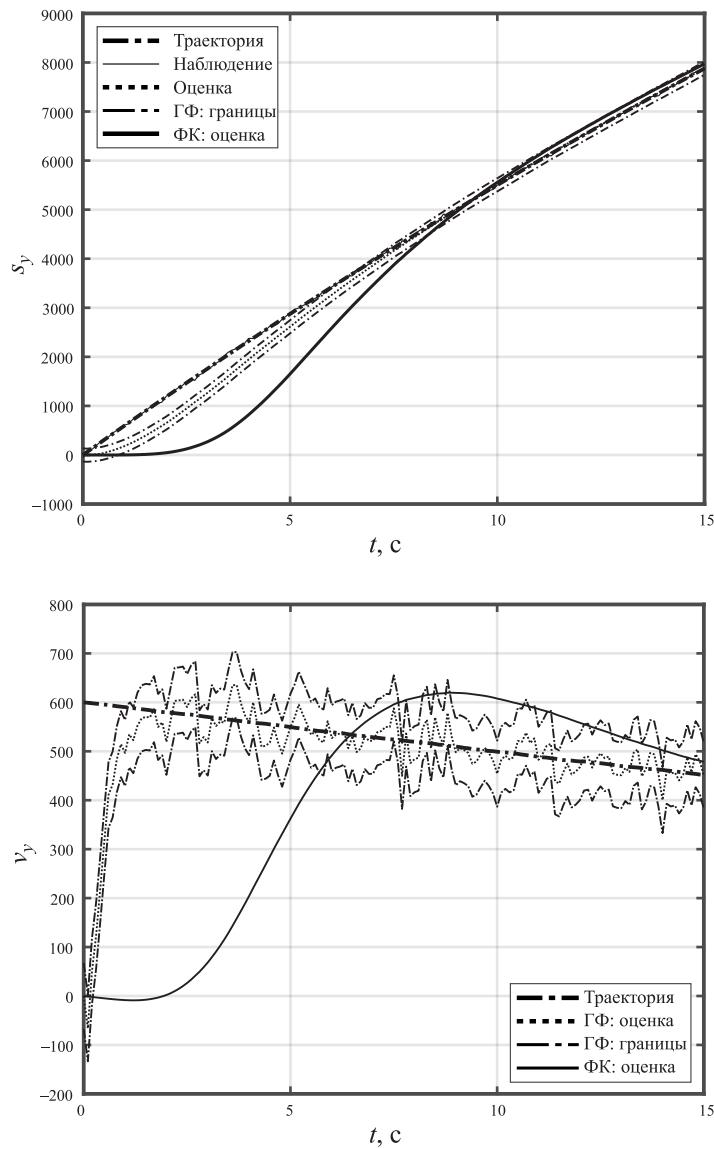


Рис. 3.5.4. Динамика координат и их оценок в примере 3.5.2 для модели \mathcal{M}_1 .

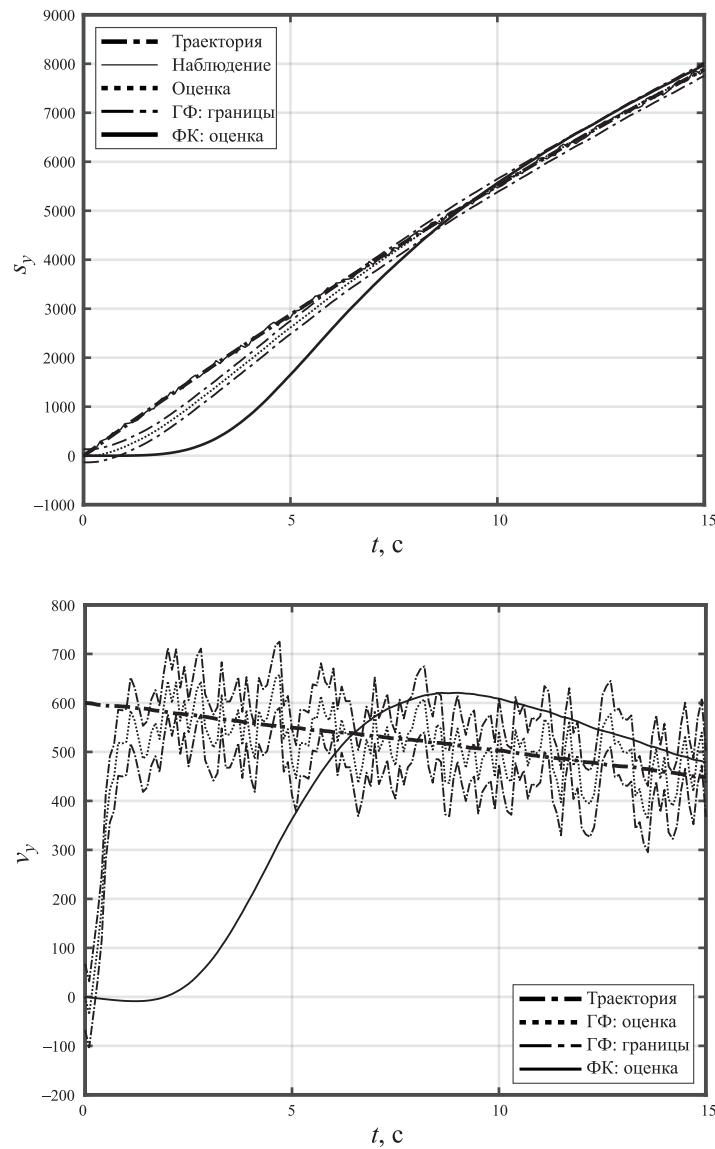


Рис. 3.5.5. Динамика координат и их оценок в примере 3.5.2 для модели \mathcal{M}_2 .

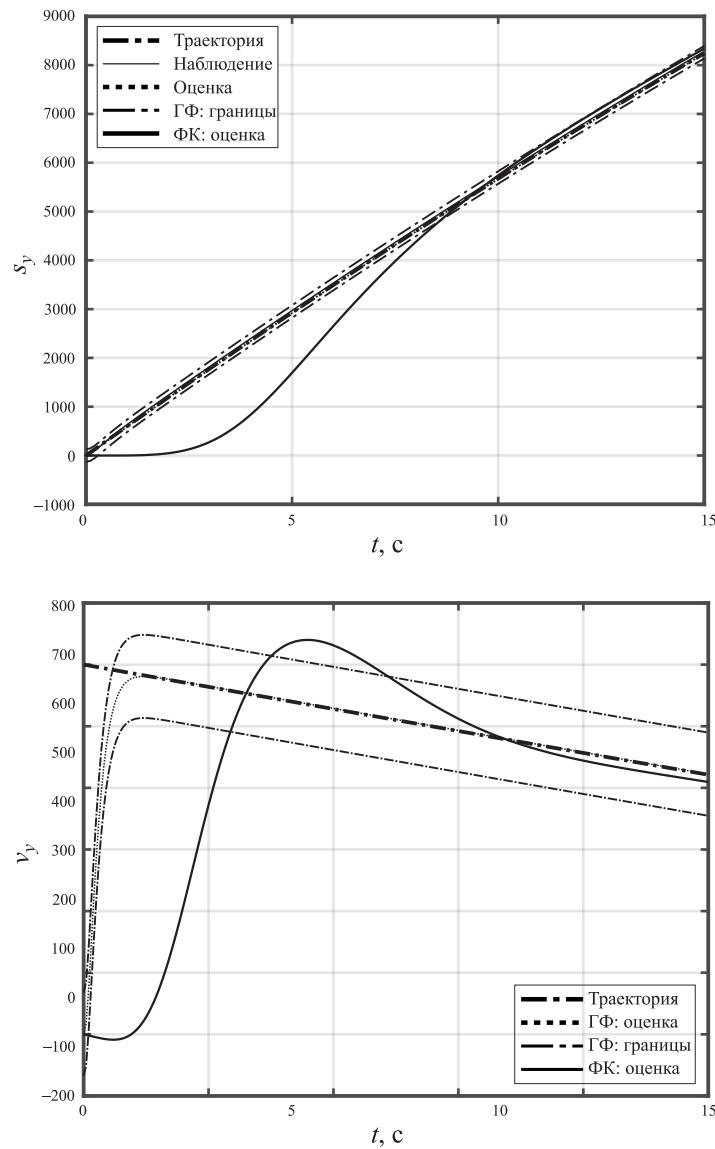


Рис. 3.5.6. Динамика координат и их оценок в примере 3.5.2 для модели \mathcal{M}_3 .

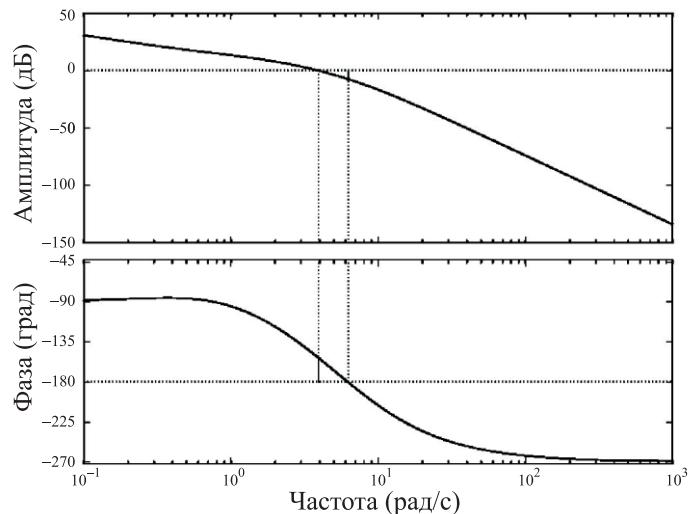


Рис. 3.6.1. ЛАФЧХ замкнутой системы из примера 3.6.1.

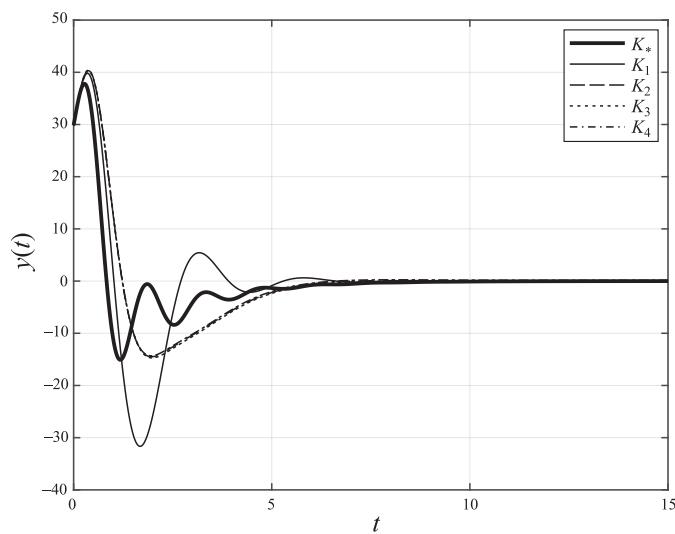


Рис. 3.6.2. Траектории выхода системы из примера 3.6.1.

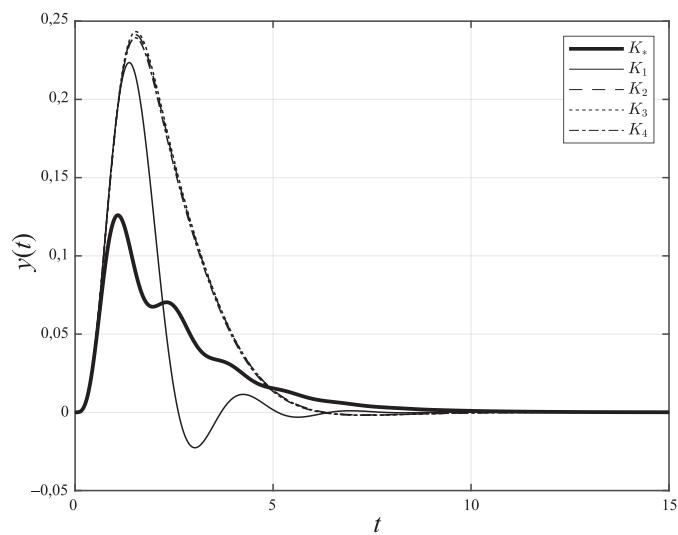


Рис. 3.6.3. Траектории выхода системы из примера 3.6.1 при единичном ступенчатом возмущении.

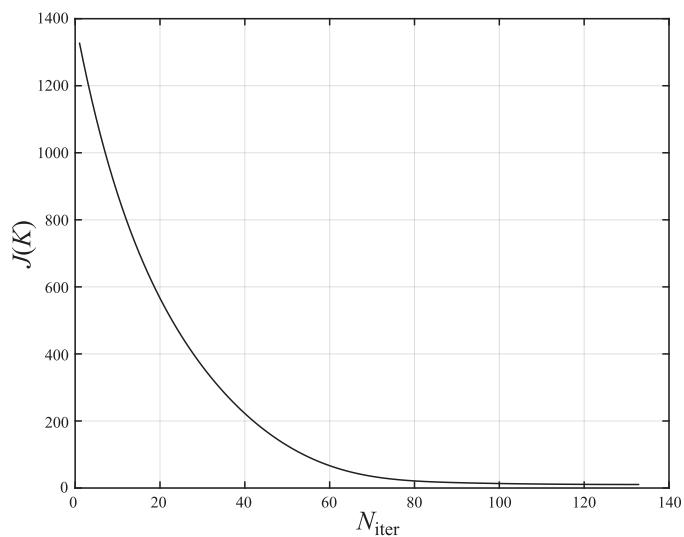


Рис. 3.6.4. Оптимизационная процедура в примере 3.6.2.

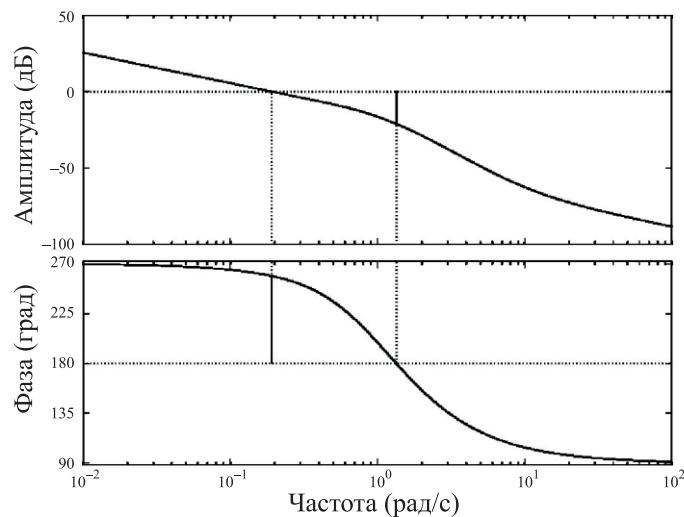


Рис. 3.6.5. ЛАФЧХ замкнутой системы из примера 3.6.2.

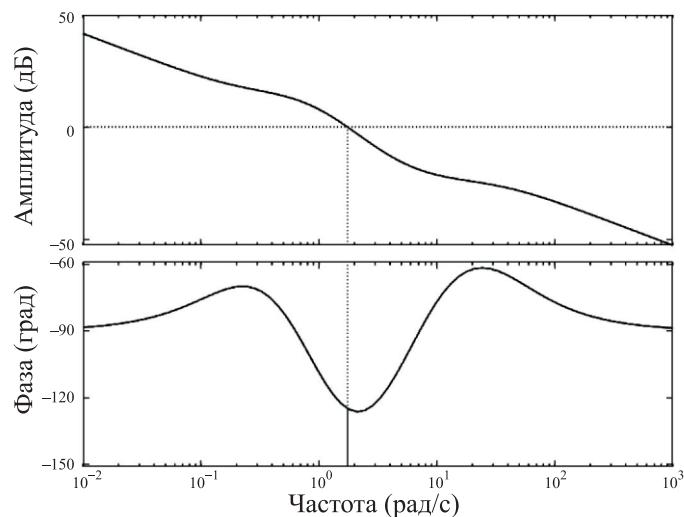


Рис. 3.6.6. ЛАФЧХ замкнутой системы из примера 3.6.3.

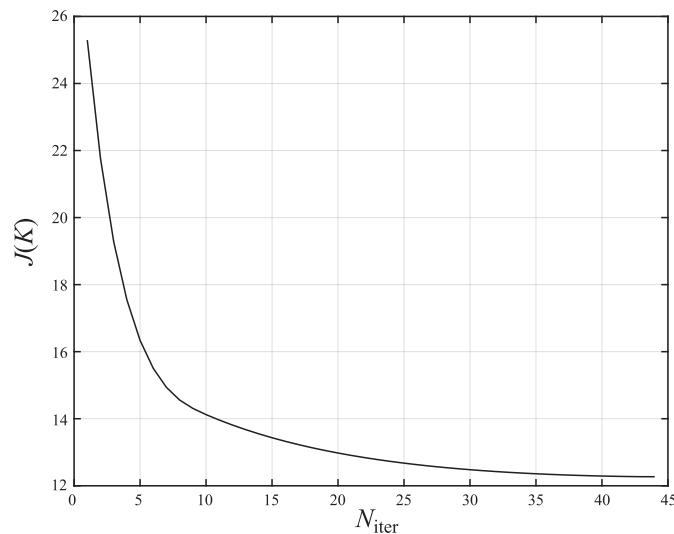


Рис. 3.6.7. Оптимизационная процедура в примере 3.6.4.

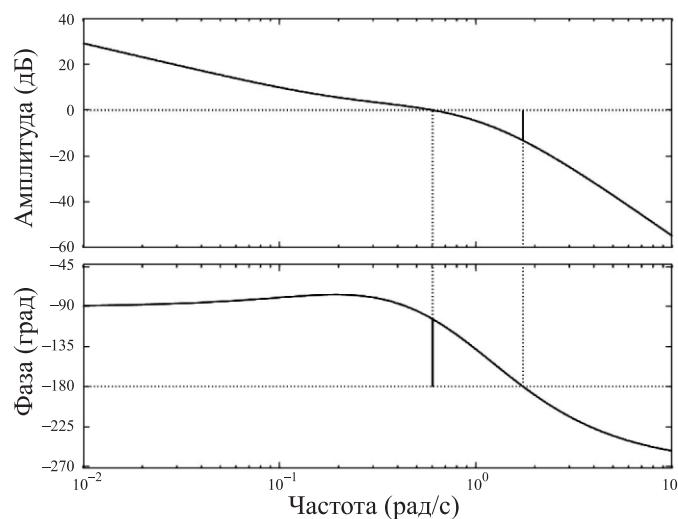


Рис. 3.6.8. ЛАФЧХ замкнутой системы из примера 3.6.4.

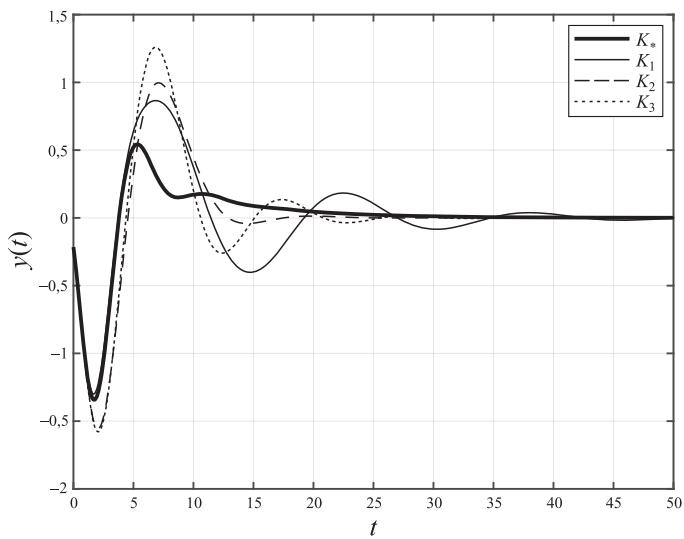


Рис. 3.6.9. Траектории выхода системы из примера 3.6.4.

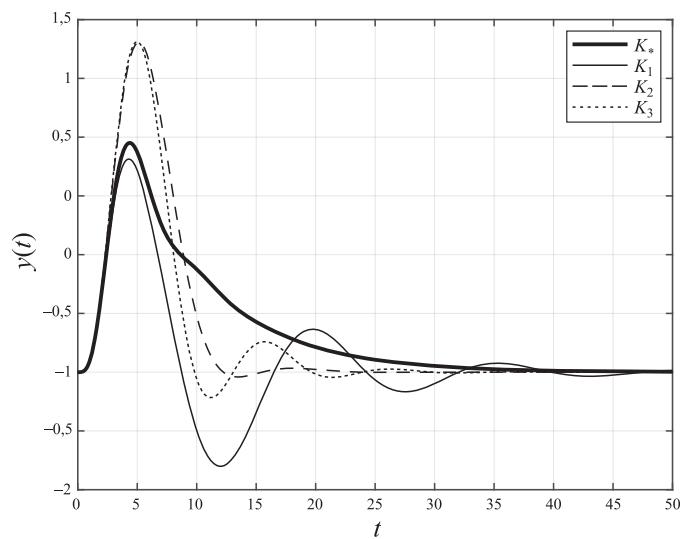


Рис. 3.6.10. Траектории выхода системы из примера 3.6.4 при единичном ступенчатом возмущении.

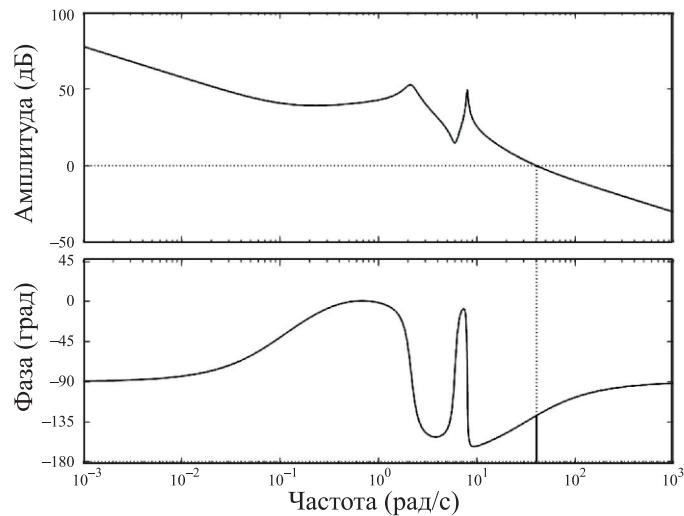


Рис. 3.6.11. ЛАФЧХ замкнутой системы из примера 3.6.5.

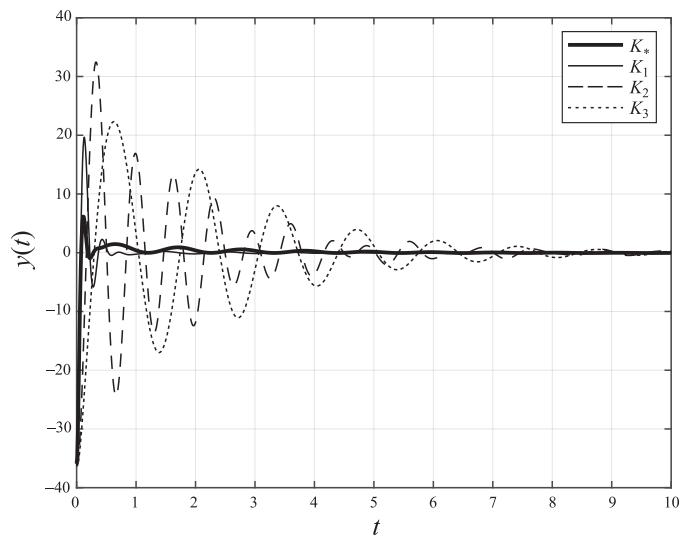


Рис. 3.6.12. Траектории выхода системы из примера 3.6.5.

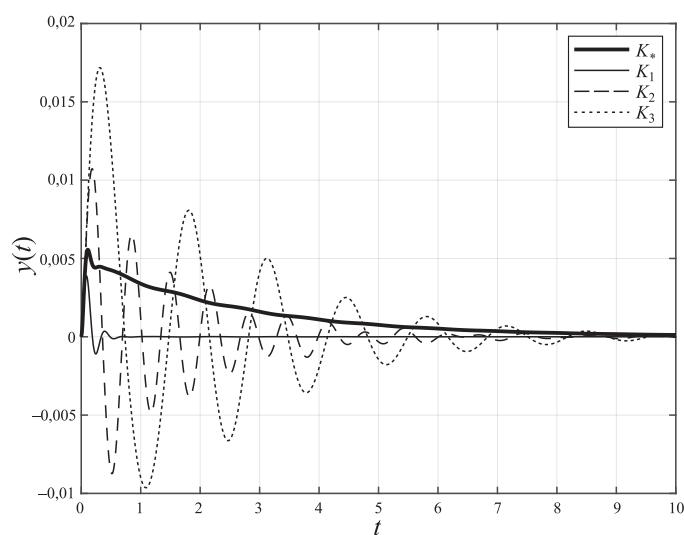


Рис. 3.6.13. Траектории выхода системы из примера 3.6.5 при единичном ступенчатом возмущении.

Приложение

A. Сведения из математического анализа

Основные результаты, изложенные в этом разделе, можно найти в стандартных монографиях по математическому анализу. Мы рекомендуем монографию С.М. Никольского [14] для начинающих и книгу У. Рудина [25] — для более искушенных читателей. На более тонкие результаты данного раздела даны ссылки.

A.1. Компактные множества

Множество $\mathcal{Q} \subset \mathbb{R}^n$ называется *компактным*, если из любого его покрытия открытыми множествами можно выделить конечное подпокрытие. Иными словами, если $\mathcal{Q} \subset \bigcup_{\alpha \in \mathcal{A}} \mathcal{U}_\alpha$, где \mathcal{U}_α открыты и \mathcal{A} — произвольное множество индексов, то $\mathcal{Q} \subset \bigcup_{1 \leq i \leq m} \mathcal{U}_{\alpha_i}$ для некоторого конечного множества $\{\alpha_i\}_{i=1}^m \subset \mathcal{A}$.

Множество $\mathcal{Q} \subset \mathbb{R}^n$ компактно тогда и только тогда, когда \mathcal{Q} замкнуто и ограничено. Компактность множества $\mathcal{Q} \subset \mathbb{R}^n$ эквивалентна тому, что из любой последовательности $\{x_k\}_{k=1}^\infty \subset \mathcal{Q}$ можно выделить подпоследовательность, сходящуюся к элементу из \mathcal{Q} .

Важное значение имеет следующая теорема о топологической оделимости компактного множества $\mathcal{Q} \subset \mathbb{R}^n$ и замкнутого множества $\mathcal{P} \subset \mathbb{R}^n$ [19, теорема 1.1.8]. Если $\mathcal{Q} \cap \mathcal{P} = \emptyset$, то найдется число $\varepsilon > 0$ такое, что

$$(\mathcal{Q} + \mathcal{B}_\varepsilon(0)) \cap \mathcal{P} = \left(\bigcup_{x \in \mathcal{Q}} \mathcal{B}_\varepsilon(x) \right) \cap \mathcal{P} = \emptyset.$$

А.2. Полунепрерывные функции. Теорема Вейерштрасса

Пусть $\text{int } \mathcal{B}_r(x_0)$ — открытый евклидов шар радиуса $r > 0$ с центром в точке $x_0 \in \mathbb{R}^n$. Напомним, что функция $f: \text{int } \mathcal{B}_r(x_0) \rightarrow \mathbb{R}$ *непрерывна* в точке x_0 , если $\lim_{x \rightarrow x_0} f(x) = f(x_0)$. Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — подмножество и $x_0 \in \mathcal{Q}$. Будем говорить, что сужение f на \mathcal{Q} непрерывно в точке x_0 , если $\lim_{\substack{x \rightarrow x_0, x \in \mathcal{Q}}} f(x) = f(x_0)$. Слово «сужение», как правило, будем опускать. Функцию $f: \mathcal{Q} \rightarrow \mathbb{R}$ будем называть *непрерывной*, если f непрерывна в каждой точке $x_0 \in \mathcal{Q}$.

Если точка $x_0 \in \mathcal{Q}$ изолирована, то будем считать функцию в ней непрерывной по определению.

Важным свойством непрерывных функций, заданных на компактном множестве, является утверждение следующей теоремы о достижимости точной верхней и нижней граней.

Теорема А.1 (Вейерштрасс). *Пусть множество $\mathcal{Q} \subset \mathbb{R}^n$ компактно и функция $f: \mathcal{Q} \rightarrow \mathbb{R}$ непрерывна. Тогда найдутся точки $x_{\min} \in \mathcal{Q}$ и $x_{\max} \in \mathcal{Q}$ такие, что*

$$f(x_{\min}) = \inf_{x \in \mathcal{Q}} f(x), \quad f(x_{\max}) = \sup_{x \in \mathcal{Q}} f(x).$$

Важными теоремами являются теорема Больцано–Коши о промежуточных значениях непрерывной на отрезке функции и теорема о непрерывности обратной функции к монотонной непрерывной функции.

Отметим, что непрерывность не является необходимой для выполнения одного из утверждений теоремы А.1. Так, функция

$$f(x) = \begin{cases} |x| + 1, & x \neq 0, \\ 0, & x = 0, \end{cases}$$

является разрывной в точке $x = 0$, но достигает минимума на отрезке $[-1, 1]$.

Напомним, что *нижний предел* функции $f: \mathcal{Q} \rightarrow \mathbb{R}$ в точке $x_0 \in \mathcal{Q}$ (\mathcal{Q} — подмножество \mathbb{R}^n) есть инфимум частичных пределов, то есть инфимум пределов $\lim_{k \rightarrow \infty} f(x_k)$ по всем последовательностям $\{x_k\} \subset \mathcal{Q}$ таким, что $x_k \rightarrow x_0$, $x_k \neq x_0$, и последовательность $\{f(x_k)\}$ имеет предел. Нижний предел функции всегда существует и либо конечен, либо равен $-\infty$. Будем обозначать его через $\liminf_{x \rightarrow x_0, x \in \mathcal{Q}} f(x)$. Аналогично *верхний предел* функции $f: \mathcal{Q} \rightarrow \mathbb{R}$ в точке $x_0 \in \mathcal{Q}$ есть супремум

частичных пределов f в точке x_0 . Верхний предел также всегда существует и либо конечен, либо равен $+\infty$; будем обозначать его через $\limsup_{x \rightarrow x_0, x \in Q} f(x)$.

Функция $f: Q \rightarrow \mathbb{R}$ полуценерывна снизу (сверху) в точке $x_0 \in Q$, если

$$f(x_0) \leq \liminf_{x \rightarrow x_0, x \in Q} f(x) \quad \left(\limsup_{x \rightarrow x_0, x \in Q} f(x) \leq f(x_0) \right).$$

Будем говорить, что функция $f: Q \rightarrow \mathbb{R}$ полуценерывна снизу (сверху), если она полуценерывна снизу (сверху) в каждой точке $x_0 \in Q$.

Теорема Вейерштрасса может быть доказана для полуценерывных функций.

Теорема А.2 (Вейерштрасс). Пусть множество $Q \subset \mathbb{R}^n$ компактно и функция $f: Q \rightarrow \mathbb{R}$ полуценерывна снизу (сверху). Тогда найдется точка $x_{\min} \in Q$ ($x_{\max} \in Q$) такая, что

$$f(x_{\min}) = \inf_{x \in Q} f(x) \quad \left(f(x_{\max}) = \sup_{x \in Q} f(x) \right).$$

Таким образом, понятие непрерывности в контексте экстремальных задач эффективно расщепляется на два понятия: полуценерывность снизу и сверху.

Теорема А.3 дает связь полуценерывности функции со свойствами надграфика функции и ее лебеговых множеств. Напомним, что функция $f: \mathbb{R}^n \supset Q \rightarrow \mathbb{R} \cup \{+\infty\}$ — собственная, если существует точка $x \in Q$ такая, что $f(x) < +\infty$.

Теорема А.3 ([19, теорема 1.5.1]). Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ — собственная функция. Тогда следующие условия эквивалентны:

- 1) f полуценерывна снизу;
- 2) лебеговы множества $\mathcal{L}_f(\mu) = \{x \in \mathbb{R}^n : f(x) \leq \mu\}$ замкнуты;
- 3) надграфик ері $f = \{(x, \mu) \in \mathbb{R}^n \times \mathbb{R} : f(x) < +\infty, f(x) \leq \mu\}$ замкнут.

A.3. Производные, градиенты, разложения

Напомним, что функция одной переменной $f: \mathbb{R} \supset \text{int } \mathcal{B}_r(x_0) \rightarrow \mathbb{R}$ называется дифференцируемой с производной $f'(x_0) \in \mathbb{R}$ в точке x_0 , если

$$f(x_0 + h) - f(x_0) = f'(x_0) \cdot h + o(h), \quad h \rightarrow 0.$$

Символ o -малое означает такую функцию, что $\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$. Очевидным следствием определения дифференцируемости функции f одной переменной в точке x_0 является непрерывность f в этой точке.

Для векторного аргумента имеется много разных понятий дифференцируемости.

Пусть $\{e_k\}_{k=1}^n$ — стандартный ортонормированный базис в \mathbb{R}^n . Градиенты Гато или Фреше, если они существуют, в декартовых координатах можно найти по формуле

$$f'(x_0) = \begin{pmatrix} \frac{\partial f(x_0)}{\partial \mathbf{x}_1} \\ \vdots \\ \frac{\partial f(x_0)}{\partial \mathbf{x}_n} \end{pmatrix},$$

где для $1 \leq k \leq n$ частные производные определены по формуле

$$\frac{\partial f(x_0)}{\partial \mathbf{x}_k} = \lim_{t \rightarrow 0} \frac{f(x_0 + te_k) - f(x_0)}{t}.$$

Для функции $f: \mathbb{R}^n \supset \text{int } \mathcal{B}_r(x_0) \rightarrow \mathbb{R}$ и вектора $e \in \mathbb{R}^n$ производной по направлению e в точке x_0 называется величина

$$f'(x_0)[e] = \frac{\partial f(x_0)}{\partial e} = \lim_{t \rightarrow 0} \frac{f(x_0 + te) - f(x_0)}{t}.$$

Легко видеть, что $\frac{\partial f(x_0)}{\partial \mathbf{x}_k} = f'(x_0)[e_k]$.

Будем говорить, что функция $f: \mathbb{R}^n \supset \text{int } \mathcal{B}_r(x_0) \rightarrow \mathbb{R}$ дифференцируема в точке x_0 по Гато (слабо дифференцируема) если существует вектор $f'(x_0) \in \mathbb{R}^n$ такой, что для любого вектора $h \in \mathbb{R}^n$ имеет место равенство

$$f(x_0 + th) - f(x_0) = t(f'(x_0), h) + o_h(t), \quad t \rightarrow 0. \quad (\text{A.1})$$

Заметим, что функция o -малое в формуле (A.1) зависит от вектора h . Вектор $f'(x_0)$ в (A.1) называется градиентом Гато функции f в точке x_0 .

Если функция $f: \mathbb{R}^n \supset \text{int } \mathcal{B}_r(x_0) \rightarrow \mathbb{R}$ дифференцируема по Гато в точке x_0 , то градиент Гато $f'(x_0)$ выражается через частные производные по формуле

$$f'(x_0) = \left(\frac{\partial f(x_0)}{\partial \mathbf{x}_1}, \dots, \frac{\partial f(x_0)}{\partial \mathbf{x}_n} \right)^T.$$

Будем говорить, что функция $f: \mathbb{R}^n \supset \text{int } \mathcal{B}_r(x_0) \rightarrow \mathbb{R}$ *дифференцируема в точке x_0 по Фреше (сильно дифференцируема)*, если существует вектор $f'(x_0) \in \mathbb{R}^n$ такой, что для любого вектора $h \in \mathbb{R}^n$ имеет место равенство

$$f(x_0 + h) - f(x_0) = (f'(x_0), h) + o(\|h\|), \quad h \rightarrow 0. \quad (\text{A.2})$$

Функция o -малое в формуле (A.2) не зависит от вектора h . Вектор $f'(x_0)$ в (A.2) называется *градиентом Фреше* функции f в точке x_0 .

Легко видеть, что дифференцируемость по Фреше влечет непрерывность f в точке x_0 . Например, функция

$$f(x_1, x_2) = \begin{cases} \frac{x_2}{x_1}(x_1^2 + x_2^2), & x_1 \neq 0, \\ 0, & x_1 = 0, \end{cases}$$

имеет в точке $(0, 0)$ градиент Гато $f'(0, 0) = (0, 0)$, но не является непрерывной в $(0, 0)$.

При изучении субдифференциала в разделе Г.5 будет использоваться понятие *односторонней производной* функции f по направлению e в точке x_0 , то есть

$$f'(x_0, e) = \lim_{t \rightarrow +0} \frac{f(x_0 + te) - f(x_0)}{t}.$$

Отметим, что если выпуклая функция дифференцируема по Гато в точке x_0 , то она в этой точке дифференцируема и по Фреше [24, теорема 25.2].

Будем говорить, что вектор-функция $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ дифференцируема по Фреше, если каждый компонент F дифференцируем по Фреше.

Если функция $f: \mathbb{R}^n \supset \text{int } \mathcal{B}_r(x_0) \rightarrow \mathbb{R}$ является $m + 1$ раз непрерывно дифференцируемой в окрестности точки x_0 (то есть $f \in C^{m+1}$), то выполнено следующее локальное представление функции f по формуле Тейлора: если $h = (h_1, \dots, h_n)$, то

$$\begin{aligned} f(x_0 + h) - f(x_0) &= \\ &= \sum_{k=1}^m \frac{1}{k!} \sum_{i_1, \dots, i_k} \frac{\partial^k f(x_0)}{\partial x_{i_1} \dots \partial x_{i_k}} h_{i_1} \dots h_{i_k} + R(x_0, h, m), \end{aligned} \quad (\text{A.3})$$

где второе суммирование ведется по всем наборам целых индексов i_j и $1 \leq i_j \leq n$ для всех $1 \leq j \leq k$. Остаток $R(x_0, h, m)$ есть $o(\|h\|^m)$

при $\|h\| \rightarrow 0$. Его также можно представить в форме Лагранжа: для некоторого $\theta \in [0, 1]$ выполнено равенство

$$R(x_0, h, m) = \frac{1}{(m+1)!} \sum_{i_1, \dots, i_{m+1}} \frac{\partial^k f(x_0 + \theta \cdot h)}{\partial x_{i_1} \dots \partial x_{i_{m+1}}} h_{i_1} \dots h_{i_{m+1}}.$$

Отметим, что θ зависит от x_0 и h .

Особо важным для рассматриваемых задач является случай $m = 2$. В этом случае из формулы (A.3) следует, что локальное представление функции f в окрестности точки x_0 задается равенством

$$\begin{aligned} f(x_0 + h) - f(x_0) &= \\ &= (f'(x_0), h) + \frac{1}{2} (h, f''(x_0)h) + o(\|h\|^2), \quad \|h\| \rightarrow 0, \end{aligned} \quad (\text{A.4})$$

где матрица вторых производных

$$f''(x_0) = \left(\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} \right)_{1 \leq i, j \leq n}$$

называется *матрицей Гессе*.

А.4. ЛИПШИЦЕВЫ ФУНКЦИИ И ФУНКЦИИ С ЛИПШИЦЕВЫМ ГРАДИЕНТОМ

Функция $f: \mathbb{R}^n \supset Q \rightarrow \mathbb{R}$ называется *непрерывной по Липшицу* (или просто *липшицевой*) с константой L на множестве Q , если для любых точек $x_1, x_2 \in Q$ выполнено неравенство

$$|f(x_1) - f(x_2)| \leq L \|x_1 - x_2\|,$$

называемое *условием Липшица*.

Класс липшицевых функций занимает промежуточное положение между классом непрерывных функций и классом дифференцируемых по Фреше функций с ограниченным градиентом. Действительно, легко видеть, что всякая липшицева на множестве Q функция непрерывна, а всякая функция с ограниченным сильным градиентом липшицева в силу теоремы Лагранжа о среднем.

Более того, справедлива *теорема Радемахера*, которая утверждает, что любая липшицева функция на открытом множестве $Q \subset \mathbb{R}^n$ дифференцируема по Фреше на Q почти всюду в смысле меры Лебега. Из теоремы Г.2 отсюда следует в частности, что любая выпуклая

функция, заданная на открытом множестве, дифференцируема почти всюду. Для незнакомых с теорией меры отметим, что если случайно выбрать точку из множества \mathcal{Q} , то с априорной вероятностью 1 функция f будет в ней дифференцируема.

Дифференцируемость почти всюду может вызвать иллюзию, что для липшицевой функции должны успешно работать основные алгоритмы, справедливые для дифференцируемой функции. Однако это не так. Напомним простейший пример из раздела 1.3.6. Для функции $f(x) = |x|$ (которая дифференцируема всюду, кроме точки $x = 0$, и при $x \neq 0$ имеем $f'(x) = 1$ ($x > 0$) или $f'(x) = -1$ ($x < 0$)) рассмотрим градиентный спуск с фиксированным шагом $\alpha > 0$, то есть

$$x_1 > 0, \quad x_{k+1} = x_k - \alpha f'(x_k).$$

Отсюда $\|x_{k+1} - x_k\| = \alpha \|f'(x_k)\|$, и если будет игнорироваться *единственная* точка негладкости $x = 0$, то последовательность $\{x_k\}$ будет расходящейся, поскольку для любого k будет выполняться равенство $\|x_{k+1} - x_k\| = \alpha > 0$. Таким образом, даже негладкость функции в одной точке может существенно испортить многие алгоритмы.

Особый интерес в приложениях представляют функции с непрерывным по Липшицу градиентом. Такие функции обладают свойствами верхней и нижней квадратичных аппроксимаций. Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R}$ имеет непрерывный по Липшицу градиент с константой Липшица L_1 (иногда для краткости будем называть такие функции функциями с L_1 -липшицевым градиентом). Тогда для любых точек $x, x_0 \in \mathbb{R}^n$ имеют место следующие оценки, см. [20, гл. 1, §1, п. 2, формула (15)]:

$$\begin{aligned} f(x_0) + (f'(x_0), x - x_0) - \frac{L_1}{2} \|x - x_0\|^2 &\leq f(x) \leq \\ &\leq f(x_0) + (f'(x_0), x - x_0) + \frac{L_1}{2} \|x - x_0\|^2. \end{aligned} \quad (\text{A.5})$$

Для вектор-функции $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ формула (A.5) доказывается аналогично, она приобретает вид

$$\|F(x) - F(x_0) - F'(x_0)(x - x_0)\| \leq \frac{L_1}{2} \|x - x_0\|^2. \quad (\text{A.6})$$

Отметим, что здесь $F'(x_0) = \left(\frac{\partial F_j(x_0)}{\partial x_i} \right)_{1 \leq i \leq n, 1 \leq j \leq m}$ — матрица $m \times n$.

Если функция f имеет L_1 -липшицев градиент и дополнительно является выпуклой, то эквивалентным свойством является выполнение

неравенства

$$(x - x_0, f'(x) - f'(x_0)) \geq \frac{1}{L_1} \|f'(x) - f'(x_0)\|^2 \quad (\text{A.7})$$

для любых x_0, x [19, лемма 1.19.5].

А.5. Теорема о неявной функции и обратном отображении

Теорема А.4 (о неявной функции). Пусть отображение

$$F: \mathbb{R}^n \times \mathbb{R}^m \supset \mathcal{U}(x_0) \times \mathcal{V}(y_0) \rightarrow \mathbb{R}^m,$$

где $\mathcal{U}(x_0), \mathcal{V}(y_0)$ — окрестности точек x_0 и y_0 в соответствующих пространствах, удовлетворяют условиям:

- 1) $F(x_0, y_0) = 0$;
- 2) ранг матрицы Якоби

$$F'_y(x_0, y_0) = \left(\frac{\partial F_i(x_0, y_0)}{\partial y_j} \right)_{1 \leq i, j \leq m}$$

равен m ;

- 3) $F \in C^1(\mathcal{U}(x_0) \times \mathcal{V}(y_0))$.

Тогда существуют окрестность $\mathcal{U}_0(x_0) \subset \mathcal{U}(x_0)$ и единственная функция $\varphi \in C^1(\mathcal{U}_0(x_0))$, $\varphi: \mathcal{U}_0(x_0) \rightarrow \mathcal{V}(y_0)$ такие, что $F(x, \varphi(x)) = 0$ для всех $x \in \mathcal{U}_0(x_0)$ и $\varphi(x_0) = y_0$.

Если требовать $F \in C^k(\mathcal{U}(x_0) \times \mathcal{V}(y_0))$ для $k \geq 1$, то и функция $\varphi \in C^k(\mathcal{U}_0(x_0))$.

Теорема А.4 позволяет доказать следующую теорему об обратном отображении [14].

Теорема А.5 (об обратном отображении). Пусть заданы открытое множество $\mathcal{U} \subset \mathbb{R}^n$ и непрерывно дифференцируемое отображение $F: \mathcal{U} \rightarrow \mathbb{R}^n$ с невырожденной матрицей Якоби, то есть

$$\det F'(x) = \det \left(\frac{\partial F_i(x)}{\partial x_j} \right)_{1 \leq i, j \leq n} \neq 0 \quad \forall x \in \mathcal{U}.$$

Тогда отображение F локально обратимо: для любой точки $x_0 \in \mathcal{U}$ существуют открытые множества $x_0 \in \mathcal{U}_0 \subset \mathcal{U}$ и \mathcal{V}_0 такие, что $F: \mathcal{U}_0 \rightarrow \mathcal{V}_0$ является взаимно однозначным отображением \mathcal{U}_0 на \mathcal{V}_0 . При этом обратное к отображению $F: \mathcal{U}_0 \rightarrow \mathcal{V}_0$ является непрерывно дифференцируемым с неравным нулю якобианом.

Следствием теоремы A.5 является то, что образ открытого множества при действии непрерывно дифференцируемого отображения с неравным нулю якобианом является открытым множеством.

Простейший вариант теоремы об обратном отображении для матриц состоит в следующем. Пусть $A \in \mathbb{R}^{n \times n}$ — обратимая матрица и возмущение $\Delta A \in \mathbb{R}^{n \times n}$ — матрица такая, что $\|\Delta A\| < 1/\|A^{-1}\|$ (нормы везде операторные). Тогда матрица $A + \Delta A$ обратима и

$$\|(A + \Delta A)^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \cdot \|\Delta A\|}. \quad (\text{A.8})$$

Результат (A.8) вытекает, например, из [10, глава IV, §5, теоремы 4 и 5].

В примере 1.5.6 было показано, что итерационный процесс

$$x_{k+1} = F(x_k)$$

может не сходиться даже при наличии неподвижной точки и гладкости F . Приведем достаточные условия сходимости этого процесса.

Теорема А.6 ([20, гл. 2, §1, теорема 1]). *Пусть $x_0 \in \mathbb{R}^n$ — неподвижная точка отображения $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$, F дифференцируемо по Фреше в точке x_0 и спектральный радиус матрицы Якоби $F'(x_0)$ удовлетворяет условию $\varrho = \varrho(F'(x_0)) < 1$ (то есть спектр $F'(x_0)$ находится в круге радиуса ϱ из \mathbb{C} с центром в нуле).*

Тогда процесс $x_{k+1} = F(x_k)$ локально линейно сходится к x_0 , а именно, для всякого $\varepsilon \in (0, 1 - \varrho)$ найдутся $\delta > 0$ и $c > 0$ такие, что при условии $\|x_1 - x_0\| < \delta$ справедлива оценка

$$\|x_k - x_0\| \leq c(\varepsilon + \varrho)^k \quad \text{для всех } k \geq 1.$$

Б. Сведения из линейной алгебры

Систематическое и подробное изложение материала разделов Б.1 и Б.2 читатель найдет в монографиях [3, 26].

Б.1. Элементарные свойства матриц

Пусть A прямоугольная матрица. *Рангом матрицы A* называется число ее линейно независимых строк или, что то же самое, число ее линейно независимых столбцов. Обозначение: $\text{rank } A$.

Полином $f(s)$ скалярной переменной s называется *аннулирующим полиномом* матрицы A , если

$$f(A) = 0.$$

Характеристическим полиномом матрицы $A \in \mathbb{R}^{n \times n}$ называется полином

$$p(\lambda) = \det(\lambda I - A) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0,$$

корни λ_i которого называются *собственными значениями* матрицы A .

Важным свойством характеристического полинома матрицы A является то, что он является аннулирующим для A , то есть имеет место

Теорема Б.1 (Гамильтон–Кэли). *Матрица $A \in \mathbb{R}^{n \times n}$ удовлетворяет своему характеристическому уравнению: если*

$$p(\lambda) = \det(\lambda I - A) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0,$$

то

$$p(A) = A^n + a_{n-1}A^{n-1} + \cdots + a_0I = 0.$$

Следствие Б.1. *Матрица A^n (а следовательно и все матрицы A^m при $m \geq n$) представима в виде линейной комбинации матриц $I, A, A^2, \dots, A^{n-1}$.*

Из других полезных свойств характеристического полинома отмечим следующие:

$$a_{n-1} = -\operatorname{tr} A,$$

$$a_0 = \det A.$$

Аннулирующий полином $\psi(s)$ наименьшей степени μ с единичным старшим коэффициентом называется *минимальным полиномом* матрицы A . Соответственно, любая степень матрицы A представима в виде линейной комбинации матриц $I, A, \dots, A^{\mu-1}$, $\mu \leq n$. Корнями минимального полинома служат все различные собственные значения матрицы A и только они.

Отметим следующие полезные свойства следа матриц (A и B — матрицы соответствующей размерности):

$$\operatorname{tr} A = \sum_{i=1}^n \lambda_i,$$

$$\begin{aligned}\operatorname{tr}(AB) &= \operatorname{tr}(BA), \\ \operatorname{tr}(A+B) &= \operatorname{tr} A + \operatorname{tr} B.\end{aligned}$$

Напомним, что симметрическая матрица A называется *положительно определенной* и обозначается через $A \succ 0$, если

$$x^T Ax > 0 \quad \forall x \in \mathbb{R}^n, \quad x \neq 0;$$

матрица называется *неотрицательно определенной* и обозначается через $A \geq 0$, если

$$x^T Ax \geq 0 \quad \forall x \in \mathbb{R}^n.$$

При этом запись $A \succ B$ означает, что $A - B \succ 0$.

Симметрическая матрица A называется *отрицательно (неположительно)* определенной и обозначается через $A \prec 0$ ($A \preccurlyeq 0$), если матрица $-A$ положительно (неотрицательно) определена.

У положительно (неотрицательно) определенных матриц все собственные значения вещественны и положительны (неотрицательны); у отрицательно (неположительно) определенных матриц все собственные значения вещественны и отрицательны (неположительны).

Отметим некоторые важные свойства знакоопределенных матриц.

1. Если A и B — матрицы одинаковой размерности, причем $A \preccurlyeq 0$ и $B \preccurlyeq 0$, то

$$A + B \preccurlyeq 0.$$

2. Если A и B — матрицы одинаковой размерности, причем $A \prec 0$ и $B \prec 0$, то

$$A + B \prec 0.$$

3. Если A, B, C, D — матрицы одинаковой размерности, причем $A \preccurlyeq B$ и $C \preccurlyeq D$, то

$$A + C \preccurlyeq B + D.$$

4. Если A, B, C, D — матрицы одинаковой размерности, причем $A \prec B$ и $C \prec D$, то

$$A + C \prec B + D.$$

5. Для любой матрицы $A \in \mathbb{R}^{n \times m}$ верно $A^T A \geq 0$ и $AA^T \geq 0$. При этом, если матрица A квадратная невырожденная, то $A^T A \succ 0$ и $AA^T \succ 0$.

6. Если $A \succ 0$ и матрица B полного строчного ранга, то $BAB^T \succ 0$, то есть строгое матричное неравенство можно одновременно умножать слева на матрицу полного строчного ранга B и справа — на B^T .

7. Если $A \succcurlyeq 0$, то для произвольной матрицы B верно $BAB^T \succcurlyeq 0$, то есть нестрогое матричное неравенство можно одновременно умножать слева на произвольную матрицу B и справа — на B^T .
8. Если $A \succcurlyeq B \succ 0$, то обратные матрицы A^{-1} и B^{-1} существуют и при этом

$$0 \prec A^{-1} \preccurlyeq B^{-1}.$$

9. Для матрицы $A \succ 0$ существует единственная матрица $B \succ 0$ такая, что $BB = A$; она называется *квадратным корнем* из матрицы A и обозначается через $B = A^{1/2}$.
10. Для симметрической матрицы A верно

$$\lambda_{\min}(A)I \preccurlyeq A \preccurlyeq \lambda_{\max}(A)I.$$

11. Если $0 \preccurlyeq A \preccurlyeq B$, то

$$\lambda_{\max}(A) \leq \lambda_{\max}(B).$$

12. Если $A \succcurlyeq 0$ и $\text{tr } A = 0$, то $A = 0$.

Б.2. Матричные нормы

Определение Б.1. Функция

$$\|\cdot\|: \mathbb{R}^{n \times m} \longrightarrow \mathbb{R},$$

определенная на множестве матриц размера $n \times m$, называется *матричной нормой*, если для любых $A, B \in \mathbb{R}^{n \times m}$ выполнены следующие аксиомы:

1. $\|A\| \geq 0$.
2. $\|A\| = 0$ тогда и только тогда, когда $A = 0$.
3. $\|\alpha A\| = |\alpha| \|A\|$ для любого $\alpha \in \mathbb{R}$.
4. $\|A + B\| \leq \|A\| + \|B\|$.

Отметим два полезных свойства, непосредственно вытекающие из определения нормы:

$$\|I\| \geq 1,$$

$$\|A\| \cdot \|A^{-1}\| \geq 1 \quad \text{для обратимой матрицы } A.$$

Наиболее употребимы следующие явно задаваемые нормы:

— *спектральная норма*:

$$\|A\|_2 \doteq \max_{1 \leq i \leq n} \lambda_i^{1/2}(A^T A);$$

— строчная норма:

$$\|A\|_1 \doteq \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|;$$

— столбцовая норма:

$$\|A\|_\infty \doteq \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|;$$

— фробениусова норма:

$$\|A\|_F \doteq \left(\sum_{i,j=1}^n |a_{ij}|^2 \right)^{1/2}.$$

Все эти нормы обладают свойством *субмультипликативности*:

$$\|AB\| \leq \|A\| \cdot \|B\|.$$

Между матричными нормами имеются следующие соотношения:

$$\begin{aligned} \frac{1}{\sqrt{n}} \|A\|_1 &\leq \|A\|_2 \leq \sqrt{n} \|A\|_1; \\ \frac{1}{\sqrt{n}} \|A\|_\infty &\leq \|A\|_2 \leq \sqrt{n} \|A\|_\infty; \\ \frac{1}{\sqrt{n}} \|A\|_F &\leq \|A\|_2 \leq \|A\|_F; \\ \frac{1}{n} \|A\|_\infty &\leq \|A\|_1 \leq n \|A\|_\infty; \\ \frac{1}{\sqrt{n}} \|A\|_F &\leq \|A\|_1 \leq \sqrt{n} \|A\|_F; \\ \frac{1}{\sqrt{n}} \|A\|_F &\leq \|A\|_\infty \leq \sqrt{n} \|A\|_F, \end{aligned}$$

и эти оценки достижимы, то есть, например,

$$\max_{A \neq 0} \frac{\|A\|_F}{\|A\|_2} = \frac{\|I\|_F}{\|I\|_2} = \sqrt{n}.$$

Пусть задана некоторая векторная норма $\|\cdot\|$; функция

$$\|A\| \doteq \max_{\|x\| \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\| \leq 1} \|Ax\| = \max_{\|x\|=1} \|Ax\|$$

удовлетворяет всем вышеуказанным аксиомам и поэтому является матричной нормой. Она называется *подчиненной* данной векторной норме $\|\cdot\|$ (или *индукцированной* нормой). Иногда используют термин *операторная норма* (в частности, для таких норм $\|I\| = 1$).

Матричные нормы $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$ подчинены векторным нормами $\|\cdot\|_\infty, \|\cdot\|_2, \|\cdot\|_1$ соответственно, то есть

$$\begin{aligned}\|A\|_2 &= \max_{\|x\|_2=1} \|Ax\|_2, \\ \|A\|_1 &= \max_{\|x\|_\infty=1} \|Ax\|_\infty, \\ \|A\|_\infty &= \max_{\|x\|_1=1} \|Ax\|_1,\end{aligned}$$

а $\|\cdot\|_F$ не является подчиненной нормой.

Отметим важное свойство спектральной нормы:

$$\lambda_{\max}(A) = \|A\|_2$$

для любой матрицы $A \succcurlyeq 0$.

Лемма Б.1. Пусть A — прямоугольная вещественнозначная матрица. Неравенство

$$\|A\|_2 \leq \gamma$$

эквивалентно матричному неравенству

$$A^T A \preccurlyeq \gamma^2 I.$$

Лемма Б.2. Для матриц A и B соответствующих размерностей справедливы соотношения

$$\|AB\|_F \leq \|A\|_F \|B\|,$$

$$|\operatorname{tr} AB| \leq \|A\|_F \|B\|_F,$$

$$\|A\| \leq \|A\|_F \leq \sqrt{\operatorname{rank} A} \|A\|,$$

$$AB + B^T A^T \leq \varepsilon A A^T + \frac{1}{\varepsilon} B^T B \quad \text{для любого } \varepsilon > 0.$$

Лемма Б.3. Для неотрицательно определенных матриц A и B справедливы соотношения

$$0 \leq \lambda_{\min}(A) \lambda_{\max}(B) \leq \lambda_{\min}(A) \operatorname{tr} B \leq \operatorname{tr} AB \leq \lambda_{\max}(A) \operatorname{tr} B \leq \operatorname{tr} A \operatorname{tr} B.$$

Б.3. Уравнения и неравенства Ляпунова и их свойства

Определение Б.2. Матрица $A \in \mathbb{R}^{n \times n}$ называется *устойчивой по Гурвицу*, или *гурвицовой*, если все ее собственные значения имеют отрицательные вещественные части:

$$\operatorname{Re} \lambda_i(A) < 0, \quad i = 1, \dots, n.$$

Для гурвицовой матрицы величина

$$\sigma(A) = -\max_i \operatorname{Re} \lambda_i(A) > 0$$

называется ее *степенью устойчивости* — это минимальное расстояние от собственных значений матрицы до мнимой оси на комплексной плоскости.

Определение Б.3. Пусть $A \in \mathbb{R}^{n \times n}$, $R \in \mathbb{S}^n$. Матричное уравнение

$$AP + PA^T = -R \tag{Б.1}$$

называется *уравнением Ляпунова* относительно матрицы $P \in \mathbb{R}^{n \times n}$.

Матричные уравнения и неравенства Ляпунова составляют основу принятого в главе 3 подхода к решению задач синтеза управления. Рассмотрим свойства решений уравнений и неравенств Ляпунова.

Лемма Б.4 ([22]). Если для всех собственных значений λ_i матрицы A выполнено условие $\operatorname{Re}(\lambda_i + \lambda_j) \neq 0$, то существует единственное решение $P = P^T$ уравнения Ляпунова (Б.1).

При этом если $R \succ 0$ или $R = BB^T$ и матрица

$$U = (B \ AB \ \cdots \ A^{n-1}B)$$

имеет ранг n , то уравнение Ляпунова (Б.1) имеет положительно определенное решение

$$P = \int_0^\infty e^{At} R e^{A^T t} dt \succ 0$$

тогда и только тогда, когда матрица A гурвицева.

Наконец, если матрица A гурвицева и $R \succ 0$, то и $P \succ 0$.

Часто рассматривается уравнение Ляпунова вида

$$A^T Q + QA = -R.$$

Поскольку операция транспонирования сохраняет гурвицевость матрицы, лемма Б.4 (с соответствующими изменениями) сохраняет свою силу.

Из леммы Б.4 вытекают почти очевидные, но важные следствия.

Следствие Б.2. *Если матрица A гурвицева, то матричное неравенство*

$$AP + PA^T \prec 0 \quad (\text{Б.2})$$

разрешимо, причем все его решения положительно определены.

Если неравенство (Б.2) имеет положительно определенное решение, то матрица A гурвицева.

Если матрица A гурвицева, то матричное неравенство

$$AP + PA^T \preccurlyeq 0 \quad (\text{Б.3})$$

разрешимо и все его решения неотрицательно определены.

Следствие Б.3. *Решение матричного неравенства*

$$A^T Q + QA \prec 0, \quad (\text{Б.4})$$

положительно определено тогда и только тогда, когда матрица A гурвицева.

Если матрица A гурвицева, то решение матричного неравенства

$$A^T Q + QA \preccurlyeq 0 \quad (\text{Б.5})$$

является неотрицательно определенным.

Матричное неравенство (Б.4) будем называть *двойственным* к матричному неравенству (Б.2), а неравенство (Б.5) — *двойственным* к матричному неравенству (Б.3). Эти двойственные неравенства эквивалентны в том смысле, что если (Б.3) выполнено с некоторой матрицей $P \succ 0$, то (Б.5) выполнено с $Q = P^{-1}$, и наоборот.

Определение Б.4. Матричное неравенство

$$AP + PA^T \preccurlyeq -R \quad (\text{Б.6})$$

называется *неравенством Ляпунова* относительно матрицы P .

Лемма Б.5. Пусть A — гурвицева матрица. Если $R \succ 0$ или $R = BB^T$ и матрица

$$U = (B \ AB \ \cdots \ A^{n-1}B)$$

имеет ранг n , то неравенство Ляпунова (Б.6) разрешимо, причем для любого его решения P справедлива оценка

$$P \succcurlyeq P_-,$$

где $P_- \succ 0$ — решение соответствующего уравнения Ляпунова.

Таким образом, среди всех решений неравенства Ляпунова минимальным (по отношению порядка \preceq) является то, которое доставляет решение соответствующему уравнению Ляпунова. Принимая во внимание свойство 12 знакопределенных матриц (с. 339) получаем следующий результат, который позволяет находить решение матричных уравнений через решение соответствующих неравенств.

Следствие Б.4. Пусть матрица A гурвицева, матрица

$$U = (B \ AB \ \cdots \ A^{n-1}B)$$

имеет ранг n и пусть C — матрица максимального ранга. Тогда решение задачи SDP

$$\min \operatorname{tr} CPC^T \quad \text{при ограничении } AP + PA^T + BB^T \preceq 0$$

достигается на решении уравнения Ляпунова

$$AP + PA^T + BB^T = 0.$$

Перейдем к дискретному случаю.

Определение Б.5. Величина

$$\rho(A) = \max_i |\lambda_i(A)|$$

называется *спектральным радиусом* матрицы $A \in \mathbb{R}^{n \times n}$.

Матрица $A \in \mathbb{R}^{n \times n}$ называется *устойчивой по Шуру*, или *шуровской*, если все ее собственные значения лежат внутри единичного круга:

$$|\lambda_i(A)| < 1, \quad i = 1, \dots, n$$

(иными словами, ее спектральный радиус меньше 1).

Для шуровской матрицы величина

$$\sigma(A) = 1 - \rho(A) > 0$$

называется ее *степенью устойчивости* — это минимальное расстояние от собственных значений матрицы до единичной окружности.

Определение Б.6. Пусть $A \in \mathbb{R}^{n \times n}$, $R \in \mathbb{S}^n$. Матричное уравнение

$$APA^T - P = -R, \quad R = R^T, \quad (\text{Б.7})$$

называется *дискретным уравнением Ляпунова* относительно матрицы $P \in \mathbb{R}^{n \times n}$.

Следующая лемма является дискретным аналогом леммы Б.4.

Лемма Б.6 ([22]). Пусть для всех собственных значений λ_i матрицы A выполнено условие $\lambda_i \lambda_j \neq 1$. Тогда существует единственное решение $P = P^T$ дискретного уравнения Ляпунова (Б.7).

При этом если $R \succ 0$ или $R = BB^T$ и матрица

$$U = (B \ AB \ \cdots \ A^{n-1}B)$$

имеет ранг n , то дискретное уравнение Ляпунова (Б.7) имеет положительно определенное решение

$$P = \sum_{k=0}^{\infty} A^k R (A^T)^k \succ 0$$

тогда и только тогда, когда A — шуровская матрица.

Наконец, если матрица A шуровская и $R \succcurlyeq 0$, то $P \succcurlyeq 0$.

С другой стороны, уравнение (Б.7) представимо в виде

$$P = APA^T + R$$

и, в предположениях леммы Б.6, для нахождения его решения можно воспользоваться простым итерационным методом

$$P_{k+1} = AP_k A^T + R,$$

который сходится со скоростью геометрической прогрессии при любом начальном приближении.

Отметим важные следствия из леммы Б.6.

Следствие Б.5. *Если матрица A шуровская, то матричное неравенство*

$$APA^T - P \prec 0 \quad (\text{Б.8})$$

разрешимо, причем все его решения положительно определены.

Если неравенство (Б.8) имеет положительно определенное решение, то матрица A шуровская.

Если матрица A шуровская, то матричное неравенство

$$APA^T - P \preccurlyeq 0 \quad (\text{Б.9})$$

разрешимо и все его решения неотрицательно определены.

Следствие Б.6. *Решение матричного неравенства*

$$A^T Q A - Q \prec 0 \quad (\text{Б.10})$$

положительно определено тогда и только тогда, когда матрица A шуровская.

Если матрица A шуровская, то решение матричного неравенства

$$A^T Q A - Q \preccurlyeq 0 \quad (\text{Б.11})$$

является неотрицательно определенным.

Матричное неравенство (Б.10) будем называть *двойственным* к матричному неравенству (Б.8), а неравенство (Б.11) — *двойственным* к матричному неравенству (Б.9).

Часто рассматривается дискретное уравнение Ляпунова вида

$$A^T Q A - Q = -R.$$

Поскольку операция транспонирования сохраняет спектр матрицы, то лемма Б.6 (с соответствующими очевидными изменениями) сохраняет свою силу.

Определение Б.7. Матричное неравенство

$$APA^T - P \preccurlyeq -R \quad (\text{Б.12})$$

называется *дискретным неравенством Ляпунова* относительно матрицы P .

Лемма Б.7. Пусть A — шуровская матрица. Если $R \succ 0$ или $R = BB^T$ и матрица

$$U = (B \ AB \ \cdots \ A^{n-1}B)$$

имеет ранг n , то неравенство Ляпунова (Б.12) разрешимо, причем для любого его решения P справедлива оценка

$$P \succcurlyeq P_-,$$

где $P_- \succ 0$ — решение соответствующего уравнения Ляпунова (то есть P_- — минимальное решение неравенства (Б.12)).

Использование свойства 12 знакопределенных матриц (с. 339) приводит к дискретному аналогу следствия Б.4.

Следствие Б.7. Пусть матрица A шуровская, матрица

$$U = (B \ AB \ \cdots \ A^{n-1}B)$$

имеет ранг n и пусть C — матрица максимального ранга. Тогда решение задачи SDP

$$\min \operatorname{tr} CPC^T \quad \text{при ограничении } APA^T - P + BB^T \preceq 0$$

достигается на решении дискретного уравнения Ляпунова

$$APA^T - P + BB^T = 0.$$

Перейдем к свойствам решений матричных уравнений Ляпунова.

Лемма Б.8. Пусть X и Y — решения двойственных уравнений Ляпунова с гурвицевой матрицей A :

$$A^T X + X A + W = 0$$

и

$$AY + Y A^T + V = 0.$$

Тогда $\operatorname{tr}(XV) = \operatorname{tr}(YW)$.

Доказательство. Прямым вычислением имеем

$$\begin{aligned} \operatorname{tr}(XV) &= \operatorname{tr}(X(-AY - YA^T)) = -\operatorname{tr}(XAY) - \operatorname{tr}(XYA^T) = \\ &= -\operatorname{tr}(XAY)^T - \operatorname{tr}(A^TXY)^T = \operatorname{tr}(Y(-A^TX - XA)) = \operatorname{tr}(YW). \end{aligned}$$

Лемма доказана. ■

Лемма Б.9. Пусть X и Y — решения двойственных дискретных уравнений Ляпунова с шуровской матрицей A :

$$A^T X A - X + W = 0 \quad u \quad A Y A^T - Y + V = 0.$$

Тогда $\text{tr}(XV) = \text{tr}(YW)$.

Лемма Б.10. Для решения P уравнения Ляпунова

$$AP + PA^T + Q = 0$$

с гурвицевой матрицей A и $Q \succ 0$ справедливы оценки:

$$\lambda_{\max}(P) \geq \frac{\lambda_{\min}(Q)}{2\sigma}, \quad \lambda_{\min}(P) \geq \frac{\lambda_{\min}(Q)}{2\|A\|}, \quad (\text{Б.13})$$

где $\sigma = -\max_i \operatorname{Re} \lambda_i(A)$.

Если же $Q = DD^T$ и пара (A, D) управляема, то

$$\lambda_{\max}(P) \geq \frac{\|u^* D\|^2}{2\sigma} > 0, \quad (\text{Б.14})$$

где

$$u^* A = \lambda u^*, \quad \operatorname{Re} \lambda = -\sigma, \quad \|u\| = 1,$$

то есть u — левый собственный вектор матрицы A , отвечающий собственному значению λ матрицы A с наибольшей вещественной частью. Вектор u и число λ могут быть комплексными; здесь u^* означает комплексное сопряжение и транспонирование.

Доказательство. Оценки (Б.13) хорошо известны [102]. Докажем справедливость оценки (Б.14). Явное решение уравнения Ляпунова для гурвицевой матрицы имеет вид

$$P = \int_0^{+\infty} e^{At} D D^T e^{A^T t} dt.$$

Умножая это равенство справа на u и слева на u^* и учитывая, что $u^* e^{At} = e^{\lambda t} u^*$, $e^{A^T t} u = e^{\lambda^* t} u$, получаем, что

$$\begin{aligned} \lambda_{\max}(P) &\geq u^* P u = \int_0^{+\infty} u^* e^{At} D D^T e^{A^T t} u dt = \\ &= \int_0^{+\infty} e^{(\lambda + \lambda^*)t} u^* D D^T u dt = \frac{\|u^* D\|^2}{2\sigma}, \end{aligned}$$

причем $\|u^*D\| > 0$ в силу управляемости пары (A, D) , см., например, [22, теорема Д.1.5]. \blacksquare

Лемма Б.11. Для решения P дискретного уравнения Ляпунова

$$APA^T - P + Q = 0$$

с шурковской матрицей A и $Q \succ 0$ справедливы оценки:

$$\lambda_{\max}(P) \geq \frac{\lambda_{\min}(Q)}{1 - \rho^2}, \quad \lambda_{\min}(P) \geq \frac{\lambda_{\min}(Q)}{1 - \sigma_{\min}^2(A)}, \quad (\text{Б.15})$$

где $\rho = \max_i |\lambda_i(A)|$, а $\sigma_{\min}(A)$ — наименьшее сингулярное число матрицы A .

Если же $Q = DD^T$ и пара (A, D) управляема, то

$$\lambda_{\max}(P) \geq \frac{\|u^*D\|^2}{1 - \rho^2} > 0, \quad (\text{Б.16})$$

где

$$u^*A = \lambda u^*, \quad |\lambda| = \rho, \quad \|u\| = 1,$$

то есть u — левый собственный вектор матрицы A , отвечающий собственному значению λ матрицы A с наибольшим модулем. Вектор u и число λ могут быть комплексными; здесь u^* означает комплексное сопряжение и транспонирование.

Доказательство. Как и в предыдущей лемме, оценки (Б.15) хорошо известны, см., например, [101]. Докажем справедливость оценки (Б.16). Явное решение дискретного уравнения Ляпунова для шурковской матрицы A имеет вид

$$P = \sum_{k=0}^{\infty} A^k D D^T (A^T)^k.$$

Умножая это равенство справа на u и слева на u^* и учитывая, что $u^* A^k = \lambda^k u^*$, $(A^T)^k u = (\lambda^*)^k u$, получаем, что

$$\begin{aligned} \lambda_{\max}(P) &\geq u^* P u = \sum_{k=0}^{\infty} u^* A^k D D^T (A^T)^k u = \\ &= \sum_{k=0}^{\infty} (\lambda \lambda^*)^k u^* D D^T u = \frac{\|u^* D\|^2}{1 - \rho^2}, \end{aligned}$$

причем $\|u^* D\| > 0$ в силу управляемости пары (A, D) , см., например, [22, теорема Д.1.5]. \blacksquare

Б.4. Матричные многообразия

В пространстве матриц $\mathbb{R}^{n \times k}$ для матриц $X, Y \in \mathbb{R}^{n \times k}$ напомним скалярное произведение

$$\langle X, Y \rangle = \text{tr } X^T Y.$$

Соответствующая ему норма называется фробениусовой:

$$\|X\|_F = \sqrt{\langle X, X \rangle}.$$

В настоящем разделе все расстояния измеряются в смысле фробениусовой нормы $\|\cdot\|_F$.

Напомним, что *многообразием Штифеля* $\mathcal{S}_{n,k}$, $k \leq n$, называется множество таких матриц $X \in \mathbb{R}^{n \times k}$, что $X^T X = I_k$, где I_k — единичная матрица размера $k \times k$. Таким образом, столбцы матрицы X есть k векторов ортонормированного базиса из \mathbb{R}^n . Пусть $X = U\Sigma V^T$ — сингулярное разложение матрицы X , где $U \in O(n)$, $V \in O(k)$, $\Sigma = (\Lambda \ 0)^T \in \mathbb{R}^{n \times k}$, $\Lambda = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_k\}$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k \geq 0$. Числа $\sigma_i > 0$ называются *сингулярными числами* матрицы X .

Для любой матрицы $X \in \mathbb{R}^{n \times k}$ выполнено включение

$$\sum_{i=1}^k U_i V_i^T = U I_{n,k} V^T \in P_{\mathcal{S}_{n,k}} X, \quad (\text{Б.17})$$

вытекающее из [49, предложение 2]. Таким образом, указанная в левой части формулы (Б.17) матрица является в общем случае одним из элементов метрической проекции X на $\mathcal{S}_{n,k}$, здесь $I_{n,k} = (I_k \ 0)^T \in \mathbb{R}^{n \times k}$.

Многообразие Грассмана $\mathcal{G}_{n,k}$, $k \leq n$, есть множество всех k -мерных подпространств в \mathbb{R}^n . Известно [84], что многообразие Грассмана $\mathcal{G}_{n,k}$ может быть вложено в пространство симметричных $n \times n$ матриц \mathbb{S}^n по формуле $\mathcal{G}_{n,k} = \{XX^T : X \in \mathcal{S}_{n,k}\} \subset \mathbb{S}^n$. Заметим, что при $X \in \mathcal{S}_{n,k}$ матрица XX^T задает метрический проектор на подпространство из \mathbb{R}^n , базисом в котором являются столбцы матрицы X .

Пусть $X \in \mathcal{S}_{n,k}$; тогда XX^T задает элемент многообразия Грассмана и при этом из сингулярного разложения матрицы X следует, что любой элемент многообразия Грассмана можно задать по формуле

$$XX^T = UZU^T, \quad U \in O(n), \quad Z = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}.$$

Пусть $X \in \mathbb{S}^n$; его диагональное представление есть $X = W^T \Lambda W$, где $W \in O(n)$, а матрица $\Lambda = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_n\}$ со свойством

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k \geq \sigma_{k+1} \geq \dots \geq \sigma_n.$$

Тогда [49, теорема 3]

$$W^T Z W \in P_{\mathcal{G}_{n,k}} X, \quad \text{где } Z = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}. \quad (\text{Б.18})$$

В. Элементы теории автоматического управления

Систематическое изложение основ математической теории автоматического управления читатель найдет в монографиях [21] и [22].

В.1. Пространство состояний

Линейная непрерывная система управления описывается векторным линейным обыкновенным дифференциальным уравнением первого порядка

$$\dot{x} = Ax + Bu + D_1w, \quad (\text{Б.1})$$

где t — (непрерывное) время, $x(t) \in \mathbb{R}^n$ — вектор *состояния* системы, $u(t) \in \mathbb{R}^p$ — вектор *управления* (*управляющие воздействия*), а $w(t) \in \mathbb{R}^m$ — вектор *внешних возмущений*, или *шумов*. И управление, и возмущение называются *входами* системы.

Обычно на практике само состояние системы неизвестно, а известен ее линейный *выход* $y(t) \in \mathbb{R}^\ell$, $1 \leq \ell \leq n$:

$$y = Cx + D_2w, \quad (\text{Б.2})$$

который также может быть подвержен воздействию внешних возмущений. Хотя природа возмущений в состоянии и выходе системы, вообще говоря, различна, удобно считать их одними и теми же, полагая что матрицы D_1 и D_2 «вырезают» из вектора w разные «куски».

Система управления, в которой матрицы $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $D_1 \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{\ell \times n}$, $D_2 \in \mathbb{R}^{\ell \times m}$ являются постоянными, называется *стационарной*. Если же элементы этих матриц изменяются во времени, соответствующая система называется *нестационарной*.

В случае когда матрицы A , B , C , D_1 , D_2 заданы, система управления является *полностью определенной*; если же эти матрицы известны не полностью, говорят о *системах с (параметрической) неопределенностью*.

Уравнение (Б.1) называется *уравнением состояния* системы управления, а уравнение (Б.2) — *уравнением выхода*. Пространство \mathbb{R}^n на-

зывается *пространством состояний*, а \mathbb{R}^p — *пространством управлений*.

Форма записи системы управления в виде (3.4.17)–(B.2) называется *описанием в пространстве состояний*. При этом матрица A называется *матрицей состояния* системы, а C — *матрицей выхода*.

Помимо наблюдаемого выхода, иногда у системы управления имеются и дополнительные выходы (регулируемый и т. п.). При этом будем обозначать через y *наблюдаемый выход*, а через z — *регулируемый*.

Относительно характера внешних возмущений $w(t)$ могут делаться самые разнообразные предположения: они могут отсутствовать или быть полностью известными; могут быть произвольными детерминированными, ограниченными в некоторой норме или же случайными. Более подробно некоторые классы внешних возмущений будут описаны ниже.

Целью управления является выбор такого управления $u = u(t)$ (или $u = u(x, t)$), которое придает системе управления (B.1) требуемые свойства (например, устойчивость, оптимальность по некоторому показателю качества и т. п.).

Если в системе (B.1) выбрано управление в виде

$$u = Kx, \quad K \in \mathbb{R}^{p \times n},$$

(*линейная обратная связь по состоянию*, см. подробнее раздел B.8), то приходим к линейной системе вида

$$\begin{aligned} \dot{x} &= Ax + D_1w, \\ y &= Cx + D_2w, \end{aligned} \tag{B.3}$$

в которой отсутствует управление (и с иной матрицей состояния A); при этом говорят о *системе, замкнутой линейной обратной связью*.

Решение ее уравнения состояния может быть записано в явном виде:

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}D_1w(\tau)d\tau, \tag{B.4}$$

где x_0 — *начальное условие*, то есть значение $x(t)$ в начальный момент $t = 0$. Таким образом, зная x_0 и $w(t)$, можно найти состояние системы $x(t)$ для всех моментов времени t .

Обратим внимание на структуру решения (B.4). Оно состоит из двух независимых друг от друга слагаемых: одно определяется только начальными условиями, другое — лишь внешними возмущениями, поэтому их поведение можно анализировать отдельно друг от друга.

Наряду с линейными непрерывными системами (B.1) рассматриваются линейные *дискретные системы*, описываемые не дифференциальными, а разностными уравнениями:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + D_1w_k, \\ y_k &= Cx_k + D_2w_k, \end{aligned} \quad (\text{B.5})$$

где $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, $D_1 \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{\ell \times n}$, $D_2 \in \mathbb{R}^{\ell \times m}$, индекс $k = 0, 1, 2, \dots$ играет роль времени (дискретное время), $x_k \in \mathbb{R}^n$ — *состояние* системы, $u_k \in \mathbb{R}^p$ — *управление (управляющие воздействия)*, $w_k \in \mathbb{R}^m$ — *внешнее возмущение (шум)*, $y_k \in \mathbb{R}^\ell$ — *выход* системы.

Дискретные системы могут возникать как при дискретной аппроксимации непрерывных систем, так и в других случаях. Например, k может быть номером итерации в итерационном процессе или время может быть дискретным в процессах, связанных с цифровым управлением.

Если в дискретной системе выбрано управление в виде линейной обратной связи

$$u_k = Kx_k, \quad K \in \mathbb{R}^{p \times n},$$

она также называется *системой, замкнутой линейной обратной связью*, и имеет вид

$$\begin{aligned} x_{k+1} &= Ax_k + D_1w_k, \\ y_k &= Cx_k + D_2w_k, \end{aligned}$$

а ее решение также выписывается в явной форме:

$$x_k = A^k x_0 + \sum_{i=0}^{k-1} A^{k-i-1} D_1 w_i. \quad (\text{B.6})$$

B.2. Передаточная функция

Существует альтернативный способ описания линейных систем — с помощью передаточных функций.

Пусть имеется описание непрерывной системы в пространстве состояний; проведем некоторые формальные преобразования ее уравнений. Для этого введем в рассмотрение символ дифференцирования

$$s = \frac{d}{dt},$$

так что для гладкой функции $x(t)$ имеем

$$sx(t) = \dot{x}(t).$$

и если

$$p(s) = a_0 + a_1s + \cdots + a_k s^k,$$

то

$$p(s)x(t) = a_0x(t) + a_1\dot{x}(t) + \cdots + a_kx^{(k)}(t),$$

где $x^{(k)}(t)$ — k -я производная по времени от $x(t)$.

При этом будем рассматривать s как формальную (комплексную) переменную. Тогда, записав уравнение системы (B.1) при $x(0) = 0$ в виде

$$sx = Ax + Bu + D_1w$$

и разрешив получившееся соотношение относительно x , получаем

$$x = (sI - A)^{-1}(Bu + D_1w)$$

и имеем следующее выражение для выхода (B.2) системы:

$$y = \underbrace{C(sI - A)^{-1}B u}_{G_{yu}(s)} + \underbrace{(C(sI - A)^{-1}D_1 + D_2)w}_{G_{yw}(s)}. \quad (\text{B.7})$$

Матричная функция

$$G_{yu}(s) \doteq C(sI - A)^{-1}B$$

комплексной переменной s называется *передаточной функцией от управления и к выходу* y , а функция

$$G_{yw}(s) \doteq C(sI - A)^{-1}D_1 + D_2$$

называется *передаточной функцией от возмущения* w к выходу y .

Элементы g_{ki} матриц $G_{yu}(s)$ и $G_{yw}(s)$ являются дробно-рациональными функциями комплексной переменной s , которые имеют общий знаменатель

$$P(s) \doteq \det(sI - A) \quad (\text{B.8})$$

— характеристический полином матрицы A , см. приложение B.1. Этот полином от переменной s называется *характеристическим полиномом системы*; от расположения его корней зависят такие важные свойства системы, как устойчивость.

Поскольку матричная функция $(sI - A)^{-1}$ определена при всех $s \in \mathbb{C}$, не являющихся собственными значениями матрицы A , то функции $G_{yu}(s)$ и $G_{yw}(s)$ аналитичны на всей комплексной плоскости за исключением конечного числа точек (называемых *полюсами передаточных функций*).

С учетом обозначения (B.8) функция $G_{yu}(s)$ (и аналогично функция $G_{yw}(s)$) представима в виде

$$G_{yu}(s) \doteq H(s) = \frac{1}{P(s)} W(s), \quad (\text{B.9})$$

где $W(s)$ — матрица, элементы которой являются полиномами от s . Действительно, для любой невырожденной матрицы $M \in \mathbb{R}^{n \times n}$

$$M^{-1} = \frac{1}{\det M} \widehat{M},$$

где $\widehat{M} = (\widehat{m}_{ij})$ — присоединенная матрица:

$$\widehat{m}_{ij} = (-1)^{i+j} \det M_{ji};$$

здесь M_{ji} — матрицы, получающиеся из M вычеркиванием j -й строки и i -го столбца. Поэтому если $M = sI - A$, то \widehat{m}_{ij} — полиномы от s , откуда и следует выражение (B.9).

Нули s_i характеристического полинома $P(s)$ называются *полюсами передаточной функции* $G(s)$. Таким образом, полюса $G(s)$ совпадают с собственными числами матрицы A ; для всех остальных s матрица $G(s)$ определена. В частности, если полином $P(s)$ устойчив, то есть все его корни лежат в открытой левой полуплоскости, то $G(s)$ — матричная функция, аналитичная в правой полуплоскости; такие передаточные функции называются *устойчивыми*.

Возвращаясь к соотношению (B.7), отметим, что выход системы можно записать как функцию от управления и внешних входов на языке передаточных функций:

$$y = G_{yu}(s)u + G_{yw}(s)w. \quad (\text{B.10})$$

Строгое обоснование перехода от записи (B.1) системы в пространстве состояний к форме (B.10) может быть сделано с помощью преобразования Лапласа.

Разумеется, описание системы в форме (B.10) может быть и исходным; возможен обратный переход к записи (3.4.17), при этом говорят,

что получена *реализация системы* в пространстве состояний. Рассмотрим для простоты ситуацию, когда внешние возмущения и ошибки измерения выхода отсутствуют:

$$y = G(s)u, \quad u \in \mathbb{R}^p, \quad y \in \mathbb{R}^\ell. \quad (\text{B.11})$$

В этой записи под *передаточной функцией* $G(s)$ будем понимать матрицу размера $\ell \times p$, элементы которой есть дробно-рациональные функции от s , то есть $G(s)$ представима в виде

$$G(s) = \frac{1}{P(s)}W(s), \quad (\text{B.12})$$

где элементы матрицы $W(s)$ размера $\ell \times p$ являются полиномами от s . Полином $P(s)$ — общий знаменатель элементов матрицы $G(s)$ — называется *характеристическим полиномом системы*, а его корни — *полюсами передаточной функции* (системы).

Такое определение характеристического полинома и полюсов системы не вполне точно, и более строгое определение дается следующим образом. Формально умножив обе части (B.11) на $P(s)$, с учетом (B.12) получим

$$P(s)y = W(s)u \quad (\text{B.13})$$

и, рассматривая теперь s как оператор дифференцирования, приходим к системе дифференциальных уравнений высокого порядка относительно $y(t) \in \mathbb{R}^\ell$ и $u(t) \in \mathbb{R}^p$. На элементы матрицы $G(s)$ естественно накладывать дополнительное *условие реализуемости*: степень полинома в числителе не превосходит степени полинома в знаменателе; такие передаточные функции называются *правильными*, или *реализуемыми*. Тогда, вводя «искусственные» переменные — состояния, — можно привести уравнение (B.13) к виду, аналогичному (3.4.17).

Иными словами, от записи системы с помощью реализуемой передаточной функции можно перейти к эквивалентному описанию в пространстве состояний, которое принято называть *реализацией передаточной функции в пространстве состояний*. При этом используют записи

$$G(s) = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

или

$$G(s) = (A, B, C, D),$$

которые означают, что система $y = G(s)u$ эквивалентна системе

$$\begin{aligned} \dot{x} &= Ax + Bu, \quad x(0) = 0, \\ y &= Cx + Du, \end{aligned}$$

и при этом

$$G(s) = C(sI - A)^{-1}B + D.$$

Переход от $G(s)$ к (A, B, C, D) -реализации может быть осуществлен различными способами, и таких реализаций много. Среди них существуют такие, в которых размерность матрицы A (то есть размерность вектора состояний x) минимальна; они называются *минимальными реализациями*.

Запись с помощью передаточных функций чрезвычайно удобна; проиллюстрируем это на простом примере. Пусть имеется несколько объектов соответствующих размерностей, соединенных последовательно, так что выход y_k каждого служит входом u_{k+1} следующего, см. рис. В.1. Пусть при этом каждый объект имеет свою передаточную функцию $G_k(s)$:

$$y_k = G_k(s)u_k, \quad k = 1, \dots, p;$$

здесь для простоты полагаем, что имеется единственный входной сигнал u , а ошибки измерения v отсутствуют.



Рис. В.1. Последовательное соединение объектов.

Отсюда получаем для связи общего входа $u = u_1$ и выхода $y = y_p$ выражение

$$y = y_p = G_p(s) \cdot \dots \cdot G_1(s)u_1 = G(s)u,$$

то есть передаточная функция последовательного соединения равна произведению передаточных функций объектов:

$$G(s) = G_p(s) \cdot \dots \cdot G_1(s). \quad (\text{B.14})$$

Выразить такое соотношение на языке пространства состояний было бы гораздо труднее. Поэтому в инженерной практике, где нередко рассматриваются сложные соединения простых звеньев (*блок-схемы* системы), язык передаточных функций является общепринятым. При этом существуют простые правила, позволяющие рассчитать итоговую передаточную функцию блок-схемы по передаточным функциям звеньев, подобные правилу (B.14) для последовательного соединения.

Наконец, рис. В.2 иллюстрирует систему, замкнутую отрицательной единичной обратной связью.

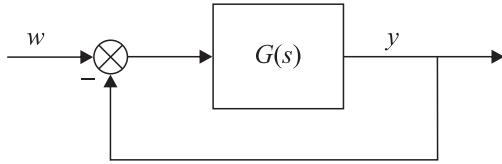


Рис. В.2. Отрицательная единичная обратная связь.

Обсудим еще одно важное свойство передаточных функций, поясняющее удобство такого способа описания систем. Предположим, что система (B.1) имеет вид

$$\begin{aligned}\dot{x} &= Ax + Bu, \\ y &= Cx,\end{aligned}$$

а входное воздействие $u(t)$ — комплексный гармонический сигнал

$$u(t) = ae^{j\omega t},$$

где a — некоторый постоянный вектор, а ω — частота колебаний (напомним, что в теории управления мнимую единицу принято обозначать j , а не i , как обычно в математике). Из формулы (B.4) для решения системы $x(t)$ получим

$$\begin{aligned}x(t) &= e^{At}x(0) + e^{At} \int_0^t e^{(j\omega I - A)\tau} B a d\tau = \\ &= e^{At}x(0) + (j\omega I - A)^{-1} B a e^{j\omega t} - (j\omega I - A)^{-1} e^{At} B a,\end{aligned}\quad (\text{B.15})$$

и через $\bar{x}(t)$ обозначим установившееся значение вектора состояния

$$\bar{x}(t) = (j\omega I - A)^{-1} B u(t).\quad (\text{B.16})$$

Предположим, что матрица A гурвицева, то есть все ее собственные значения λ_i лежат в левой полуплоскости: $\operatorname{Re} \lambda_i < 0$, $i = 1, \dots, n$. Можно показать, что для гурвицевых матриц $e^{At} \rightarrow 0$ при $t \rightarrow \infty$. Тогда из (B.15)–(B.16) следует

$$\|x(t) - \bar{x}(t)\| \rightarrow 0 \quad \text{при } t \rightarrow \infty.$$

Таким образом, для установившегося значения выхода имеем

$$\bar{y}(t) = C\bar{x}(t) = C(j\omega I - A)^{-1} B u(t),$$

$$\|y(t) - \bar{y}(t)\| \rightarrow 0 \text{ при } t \rightarrow \infty,$$

или, иначе говоря,

$$\bar{y}(t) = H(j\omega)u(t), \quad (\text{B.17})$$

где матричная функция $H(j\omega)$ называется *частотной характеристической* системы.

Поясним смысл полученного соотношения (B.17). Пусть все компоненты входного вектора $u(t)$ равны нулю, кроме i -й, которую представим в виде $u_i(t) = a \cos \omega t + ja \sin \omega t$ (где a — число). Тогда k -я компонента установившегося значения выходного сигнала равна

$$\bar{y}_k(t) = |h_{ki}(j\omega)|a \cos(\omega t + \varphi) + j|h_{ki}(j\omega)|a \sin(\omega t + \varphi),$$

где $h_{ki}(j\omega)$ — (k, i) -й элемент матрицы $H(j\omega)$, а $\varphi = \arg h_{ki}(j\omega)$. В силу линейности $H(\cdot)$, отклик системы на сумму вещественной и мнимой составляющих $u(t)$ равен сумме откликов на каждую из них, то есть если в качестве $u_i(t)$ взять вещественную гармонику $a \cos \omega t$, то установившееся значение на k -м выходе будет

$$\bar{y}_k(t) = |h_{ki}(j\omega)|a \cos(\omega t + \varphi).$$

Приходим к важному выводу: если на i -й вход системы с устойчивой (гуровицовой) матрицей A подается гармонический сигнал с частотой ω , то на k -м выходе в пределе получается также гармонический сигнал с той же частотой. Его амплитуда в $|h_{ki}(j\omega)|$ раз отличается от амплитуды входного сигнала (то есть $|h_{ki}(j\omega)|$ имеет смысл *коэффициента усиления* входного гармонического сигнала), а фаза изменяется на $\arg h_{ki}(j\omega)$. Это свойство используется на практике для определения частотной характеристики системы экспериментальным путем.

Введем теперь передаточные функции для линейной дискретной системы (B.5). Для этого определим оператор сдвига назад z :

$$zx_k = x_{k-1}$$

и аналогично непрерывному случаю будем рассматривать его как формальную переменную. Тогда при $x_0 = 0$ уравнение состояния системы (B.5) примет вид

$$x_k = zAx_k + zBu_k + zD_1w_k,$$

то есть

$$\begin{aligned} x_k &= z(I - zA)^{-1}Bu_k + z(I - zA)^{-1}D_1w_k, \\ y_k &= zC(I - zA)^{-1}Bu_k + (zC(I - zA)^{-1}D_1 + D_2)w_k. \end{aligned}$$

Передаточные функции выражаются через переменную z по формулам

$$\begin{aligned} G_{yu}(z) &= zC(I - zA)^{-1}B, \\ G_{yw}(z) &= zC(I - zA)^{-1}D_1 + D_2, \end{aligned} \quad (\text{B.18})$$

а *характеристическим полиномом системы* так же, как и раньше, называется общий знаменатель элементов матричных передаточных функций, то есть полином

$$P(z) \doteq \det(I - zA)$$

от переменной z . Соответственно, передаточные функции (B.18), как и в непрерывном случае, имеют вид

$$G(z) = \frac{1}{P(z)}W(z),$$

где $W(z)$ — матрица, элементы которой являются полиномами от z .

В терминах передаточных функций удобно описывать трансформацию гармонических сигналов, имеющих фиксированную частоту, при прохождении через линейную систему. Поэтому методы, основанные на таком подходе, обычно называют *частотными*.

B.3. Нормы сигналов

Посмотрим на внешнее возмущение в системе (B.1) (или (B.3)) с другой точки зрения — как на сигнал, подаваемый на вход системы. Рассмотрим основные функциональные пространства сигналов и соответствующие нормы. Будем считать, что все сигналы определены при $t \geq 0$.

1. *L_2 — пространство интегрируемых с квадратом функций.* Оно состоит из измеримых функций $w(t)$, заданных на $t \geq 0$ и имеющих ограниченную L_2 -норму:

$$\|w\|_2 \doteq \left(\int_0^\infty |w(t)|_2^2 dt \right)^{1/2} < \infty, \quad (\text{B.19})$$

где $|w(t)|_2$ — евклидова норма вектора $w(t) \in \mathbb{R}^m$.

Бо многих физических приложениях $\|w\|_2^2$ означает энергию сигнала, поэтому L_2 называется *пространством сигналов ограниченной энергии*.

Из конечности интеграла в (B.19) следует, что функции $w \in L_2$, по существу, убывают с течением времени.

Пример В.1. Легко видеть, что функция $w(t) \equiv 1$ не принадлежит классу L_2 , так же как и функция $w(t) = \sin \omega t$ при любом значении $\omega \neq 0$. \blacktriangledown

2. L_∞ — пространство существенно ограниченных функций. В него входят измеримые функции $w(t)$ с ограниченной ∞ -нормой:

$$\|w\|_\infty \doteq \sup_{0 \leq t < \infty} |w(t)| < \infty,$$

где $|\cdot|$ — какая-либо векторная норма в \mathbb{R}^m (как правило, ∞ - или 2-норма). Заметим, что более правильно было бы писать

$$\|w\|_\infty = \text{ess sup}_{0 \leq t < \infty} |w(t)|,$$

где ess sup — *существенная верхняя грань функции*, полученная при пренебрежении множествами нулевой меры. Мы, однако, будем считать, что функция $w(t)$ уже изменена на множестве нулевой меры так, что величина $\sup_{0 \leq t < \infty} |w(t)|$ оказалась минимальной. Например, вместо

$$w(t) = \begin{cases} 0, & t \neq 1, \\ 1, & t = 1, \end{cases}$$

будем рассматривать эквивалентную ей функцию $w(t) \equiv 0$. Нетрудно видеть, что для обеих функций $\|w\|_\infty = 0$.

Обычно величина $\|w\|_\infty$ измеряет интенсивность сигнала, так что L_∞ называется *пространством сигналов ограниченной интенсивности*.

Пример В.2. Функции

$$w(t) \equiv 1 \quad \text{и} \quad w(t) = \sin \omega t$$

принадлежат классу L_∞ , а функция

$$w(t) = \begin{cases} t^{-\alpha}, & 0 < t \leq 1, \\ 0, & t > 1, \end{cases}$$

при $0 < \alpha < 1/2$ принадлежит классу L_2 , так как

$$\|w\|_2 = \frac{1}{\sqrt{1 - 2\alpha}},$$

но не принадлежит классу L_∞ , поскольку не ограничена в нуле. \blacktriangledown

3. L_1 — пространство абсолютно интегрируемых функций, содержащее измеримые функции с ограниченной 1-нормой:

$$\|w\|_1 \doteq \int_0^\infty |w(t)| dt < \infty,$$

где $|w(t)|$ — некоторая норма вектора $w(t) \in \mathbb{R}^m$ (как правило, 1-норма).

Пример B.3. Функция

$$w(t) = \begin{cases} t^{-\alpha}, & 0 < t \leq 1, \\ 0, & t > 1, \end{cases}$$

при $1/2 \leq \alpha < 1$ принадлежит классу L_1 , так как

$$\|w\|_1 = \frac{1}{1-\alpha},$$

но не принадлежит ни классу L_2 (поскольку интеграл $\int_0^\infty t^{-2\alpha} dt$ расходится в нуле), ни классу L_∞ . \blacktriangledown

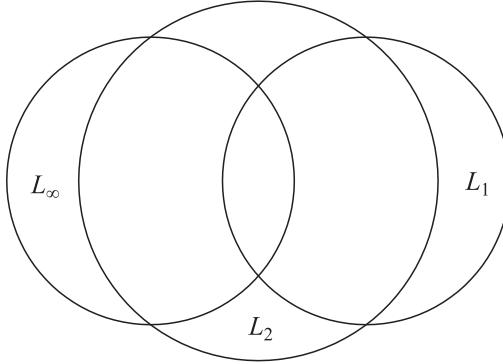
Соотношение между рассмотренными пространствами можно условно изобразить в виде диаграммы, представленной на рис. B.3. В частности, из ограниченности функции в L_1 -норме и L_∞ -норме следует ограниченность в L_2 -норме. Действительно,

$$\|w\|_2^2 = \int_0^\infty |w(t)| |w(t)| dt \leq \|w\|_\infty \int_0^\infty |w(t)| dt = \|w\|_\infty \|w\|_1.$$

Аналогичные пространства и нормы вводятся для дискретного случая, при этом рассматриваются не вектор-функции $w(t)$, а последовательности векторов $\{w_k\}$.

1. l_2 — пространство суммируемых с квадратом последовательностей. Оно состоит из последовательностей $\{w_k\}$, имеющих ограниченную 2-норму:

$$\|w\|_2 \doteq \left(\sum_{k=0}^{\infty} |w_k|_2^2 \right)^{1/2} < \infty.$$

Рис. В.3. Пространства L_1 , L_2 и L_∞ .

2. l_∞ — пространство существенно ограниченных последовательностей. Оно состоит из последовательностей $\{w_k\}$ с ограниченной ∞ -нормой:

$$\|w\|_\infty \doteq \sup_{0 \leq k < \infty} |w_k|_\infty < \infty.$$

3. l_1 — пространство абсолютно суммируемых последовательностей. Оно состоит из последовательностей $\{w_k\}$, имеющих ограниченную 1-норму:

$$\|w\|_1 \doteq \sum_{k=0}^{\infty} |w_k|_1 < \infty.$$

Пример В.4. Если

$$w_k \equiv 1 \quad \text{или} \quad w_k = (-1)^k,$$

то $\{w_k\} \in l_\infty$, но $\{w_k\} \notin l_2$, $\{w_k\} \notin l_1$.

Если же

$$w_k = (k+1)^{-\alpha},$$

то $\{w_k\} \in l_\infty$ при любом $\alpha > 0$, тогда как $\{w_k\} \in l_2$ при $\alpha > 1/2$ и $\{w_k\} \in l_1$ при $\alpha > 1$. \blacktriangledown

B.4. Коэффициенты усиления

Рассмотрим линейную непрерывную систему

$$\begin{aligned} \dot{x} &= Ax + Bu, & x(0) &= 0, \\ y &= Cx, \end{aligned} \tag{B.20}$$

с состоянием $x(t)$, управлением $u(t)$ и выходом $y(t)$. Сигнал на ее выходе y линейно зависит от сигнала на входе u . Для ответа на вопрос, насколько «усиливаются» входные сигналы, используется понятие *коэффициента усиления* (англоязычный аналог — *gain*). Для него используется обозначение $\mathfrak{g}_{p,q}$, смысл которого состоит в том, что входные сигналы измеряются в норме L_p , а выходные — в норме L_q ; обычно рассматриваются $p, q = 1, 2$ или ∞ .

Для различных классов входных и выходных сигналов коэффициент усиления системы имеет прозрачный физический смысл, и многие задачи теории управления могут быть сведены к его вычислению или оптимизации. Например, обычно желательно выбирать управление так, чтобы коэффициент усиления был по возможности мал; это означает, что выход системы будет мал при любых *возмущениях*, ограниченных в соответствующей норме.

В разделе B.2 было введено понятие передаточной функции, которая для системы (B.20) имеет вид

$$G(s) = C(sI - A)^{-1}B.$$

Выражение

$$\|G(s)\|_\infty \doteq \sup_{\omega} \|C(j\omega I - A)^{-1}B\|_2,$$

где $\|M\|_2$ — *спектральная норма* фиксированной матрицы M с комплексными элементами, определяет так называемую *H_∞ -норму передаточной функции* $G(s)$. Величина $\|G(s)\|_\infty$ равна коэффициенту усиления $\mathfrak{g}_{2,2}$, соответствующему входным и выходным сигналам, измеряемым в норме L_2 . Она показывает, во сколько раз может измениться *энергия* сигнала при прохождении через систему.

Коэффициент $\mathfrak{g}_{2,\infty}$ соответствует входным сигналам из L_2 и выходным сигналам, измеряемым в норме L_∞ ; методы его вычисления также хорошо разработаны. По определению, L_∞ -норма выходного сигнала $y(t)$ равна радиусу минимального *шара* \mathcal{B} в евклидовой норме, которому принадлежит $y(t)$ при всех $t \geq 0$ и всех входных сигналах $\|w(t)\|_2 \leq 1$.

Наибольший интерес представляет случай, когда в системе (B.20) рассматривается входной сигнал из пространства L_∞ и нужно получить оценку на L_∞ -норму выхода. В частности, при $C = I$ выход системы совпадает с ее состоянием; в западной литературе величину $\mathfrak{g}_{\infty,\infty}$ называют «peak-to-peak gain» — коэффициент усиления от возмущений, ограниченных единицей в норме L_∞ , к состоянию системы, также измеряемому в норме L_∞ . Как и в предыдущих случаях, коэффициент $\mathfrak{g}_{\infty,\infty}$ может вычисляться в терминах передаточной функции.

B.5. Нормы передаточных функций

Рассмотрим еще один вид норм — не для функций от времени, а для передаточных функций. Пусть $M(s)$ — матрица размера $n \times n$, элементы которой являются аналитическими функциями комплексной переменной s в правой полуплоскости.

Определим H_∞ -норму этой функции как

$$\|M\|_\infty = \sup_{\operatorname{Re} s \geq 0} \|M(s)\|_2 = \sup_{-\infty < \omega < \infty} \|M(j\omega)\|_2, \quad (\text{B.21})$$

где $\|M(s)\|_2$ означает спектральную норму фиксированной матрицы $M(s)$:

$$\|M(s)\|_2 = \left(\lambda_{\max}(M^*(s)M(s)) \right)^{1/2}.$$

В (B.21) супремум достаточно искать не по всей правой полуплоскости комплексной плоскости, а лишь по ее границе — мнимой оси.

В частном случае, когда $M(s)$ — скалярная функция ($n = 1$), определение H_∞ -нормы приобретает вид

$$\|M\|_\infty = \sup_{-\infty < \omega < \infty} |M(j\omega)|.$$

Пример B.5. Функция

$$M(s) = \frac{1}{s+1}$$

имеет единственный полюс в $s = -1$ и потому аналитична в правой полуплоскости. Для нее получаем

$$\|M\|_\infty = \sup_{\omega} \left| \frac{1}{j\omega + 1} \right| = \sup_{\omega} \frac{1}{\sqrt{1 + \omega^2}} = 1. \quad \blacktriangledown$$

Оказывается, что если система (B.20) записана с помощью передаточных функций в форме

$$y = H(s)u, \quad H(s) = C(sI - A)^{-1}B, \quad (\text{B.22})$$

то при условии, что A гурвицева матрица, имеем

$$\|y\|_2 \leq \|H(s)\|_\infty \|u\|_2,$$

где нижний индекс 2 для функций означает 2-норму, а $\|H(s)\|_\infty$ означает H_∞ -норму, которая (в предположении гурвицевости матрицы A) является конечной. Более того, как вытекает из следующей леммы, H_∞ -норма является точной верхней границей отношения энергий выход/вход.

Лемма B.1. Для системы (B.22) с устойчивой матрицей A справедливо соотношение

$$\|H(s)\|_\infty = \sup_{\|u\|_2 \neq 0} \frac{\|y\|_2}{\|u\|_2}.$$

Как отмечалось выше, H_∞ -норма передаточной функции имеет простой физический смысл: она показывает, во сколько раз может измениться энергия сигнала при прохождении через данную линейную систему. Если вспомнить, что другая трактовка передаточной функции характеризовала $H(j\omega)$ как *коэффициент усиления* системы для гармонического входа заданной частоты ω , то получаем, что H_∞ -норма является также равномерно-частотной характеристикой системы. Эти два результата относительно двух типов входных сигналов (ограниченные в 2-норме и гармонические с произвольной частотой) различны, так как гармонические сигналы не принадлежат L_2 .

Другая норма, которую можно ввести для тех же матричных функций $M(s)$, — это *H_2 -норма*:

$$\|M\|_2 = \left(\int_{-\infty}^{\infty} \|M(j\omega)\|_2^2 d\omega \right)^{1/2} = \left(\int_{-\infty}^{\infty} \lambda_{\max}(M^*(j\omega)M(j\omega)) d\omega \right)^{1/2}.$$

В частности, для скалярных функций определение принимает вид

$$\|M\|_2 = \left(\int_{-\infty}^{\infty} |M(j\omega)|^2 d\omega \right)^{1/2}.$$

В примере B.5 имеем

$$\|M\|_2 = \left(\int_{-\infty}^{\infty} \frac{1}{|j\omega + 1|^2} d\omega \right)^{1/2} = \left(\int_{-\infty}^{\infty} \frac{d\omega}{1 + \omega^2} \right)^{1/2} = \sqrt{\pi}.$$

Все вышеприведенные понятия и результаты имеют дискретные аналоги. В этом случае вместо матриц $M(s)$, аналитичных в правой полуплоскости, фигурируют матрицы $M(z)$, аналитические внутри единичного круга. H_∞ -норма таких матричных функций равна

$$\begin{aligned} \|M\|_\infty &= \sup_{|z| \leq 1} \|M(z)\|_2 = \sup_{0 \leq \omega \leq 2\pi} \|M(e^{j\omega})\|_2 = \\ &= \sup_{0 \leq \omega \leq 2\pi} \left(\lambda_{\max}(M^*(e^{j\omega})M(e^{j\omega})) \right)^{1/2}, \end{aligned}$$

а для скалярных функций —

$$\|M\|_\infty = \sup_{0 \leq \omega \leq 2\pi} |M(e^{j\omega})|.$$

При этом H_∞ -норма передаточной функции дискретных систем имеет тот же физический смысл, что и для непрерывных систем.

B.6. Одномерные системы управления

Остановимся теперь на простом, но важном специальном классе систем управления, имеющих скалярный вход и скалярный выход. Будем называть их *одномерными*, в отличие от общего *многомерного* случая (раньше использовался термин «односвязные» и «многосвязные» системы). В англоязычной научной литературе общеупотребительны термины SISO (Single-Input Single-Output) и MIMO (Multiple-Input Multiple-Output), то есть системы с одним входом — одним выходом и со многими входами — многими выходами соответственно.

В инженерных приложениях часто используется способ описания систем не в пространстве состояний, а в виде дифференциального уравнения высокого порядка, связывающего вход $u(t)$ с выходом $y(t)$ напрямую, без использования понятия состояния. При этом говорят о *входо-выходном описании систем во временной области*. Так, уравнения, описывающие поведение механических систем, содержат координаты, скорости и ускорения взаимодействующих тел.

В этом разделе на примере одномерных систем рассмотрим способы описания, приведенные в предыдущих разделах (а также обсудим входо-выходное описание во временной области) и их соотношения и приведем общепринятую терминологию, используемую в инженерной практике.

Рассмотрим сначала одномерную линейную систему управления в пространстве состояний

$$\begin{aligned}\dot{x} &= Ax + Bu, \\ y &= Cx\end{aligned}\tag{B.23}$$

с состоянием $x(t) \in \mathbb{R}^n$, выходом $y(t) \in \mathbb{R}$ и управлением $u(t) \in \mathbb{R}$. Соответственно, $A \in \mathbb{R}^{n \times n}$, а B и C^\top — векторы из \mathbb{R}^n . В этом случае передаточная функция

$$H_{yu}(s) \doteq H(s) = C(sI - A)^{-1}B$$

является скалярной функцией от s .

Для удобства перепишем систему (B.23) в виде

$$\begin{aligned}\dot{x} &= Ax + bu, \\ y &= c^T x,\end{aligned}\tag{B.24}$$

где $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}$, $y(t) \in \mathbb{R}$, $A \in \mathbb{R}^{n \times n}$, $b, c \in \mathbb{R}^n$.

Перейдем от записи в пространстве состояний к описанию в виде одного соотношения, связывающего вход u с выходом y напрямую, без использования понятия состояния. С этой целью продифференцируем n раз второе уравнение в (B.24), подставляя выражение для \dot{x} из первого уравнения. В результате получим

$$\begin{aligned}y &= c^T x, \\ \dot{y} &= c^T Ax + c^T bu, \\ \ddot{y} &= c^T A^2 x + c^T Abu + c^T b\dot{u}, \\ &\vdots \\ y^{(k)} &= c^T A^k x + c^T A^{k-1} bu + \cdots + c^T bu^{(k-1)}, \\ &\vdots \\ y^{(n)} &= c^T A^n x + c^T A^{n-1} bu + \cdots + c^T A^{n-k} bu^{(k-1)} + \cdots + c^T bu^{(n-1)}.\end{aligned}$$

Сложим уравнения, предварительно умножив первое на a_0 , второе — на a_1 , и т. д., последнее — на единицу, где a_i — коэффициенты характеристического полинома матрицы A . По теореме Гамильтона–Кэли (теорема B.1) каждая матрица удовлетворяет своему характеристическому уравнению, поэтому члены с x в правой части исчезнут; в результате приходим к обыкновенному дифференциальному уравнению n -го порядка относительно входного и выходного сигналов, то есть к эквивалентной записи на языке входо-выходных соотношений:

$$\begin{aligned}y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1\dot{y} + a_0y &= \\ = \beta_{n-1}u^{(n-1)} + \cdots + \beta_1\dot{u} + \beta_0u, &\tag{B.25}\end{aligned}$$

где

$$\begin{aligned}\beta_k &= c^T (A^{n-k-1} + a_{n-1}A^{n-k-2} + \cdots \\ &\quad \cdots + a_{k+2}A + a_{k+1}I)b, \quad k = 0, \dots, n-1.\end{aligned}$$

С помощью оператора дифференцирования s соотношение (B.25) записывается в компактной форме

$$P(s)y = Q(s)u,$$

$$P(s) = \sum_{k=0}^n a_k s^k, \quad a_n = 1, \quad Q(s) = \sum_{k=0}^{n-1} \beta_k s^k,$$

или

$$y = H(s)u, \quad H(s) = \frac{Q(s)}{P(s)},$$

где передаточная функция $H(s)$ — скалярная дробно-рациональная (в отличие от общего многомерного случая), а ее знаменатель — полином $P(s) = \det(sI - A)$ — является характеристическим полиномом рассматриваемой системы.

В разделе B.2 обсуждалось условие реализуемости: степень m полинома в числителе передаточной функции не превосходит степени n полинома в знаменателе. В (B.25) это условие выполнено: $m \leq n - 1$.

Итак, система в пространстве состояний сведена к эквивалентной записи на языке передаточных функций.

Обратно, пусть теперь доступно описание системы в форме одного дифференциального уравнения n -го порядка

$$\begin{aligned} y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1\dot{y} + a_0y &= \\ &= \beta_m u^{(m)} + \cdots + \beta_1\dot{u} + \beta_0u, \quad n \geq m. \end{aligned} \quad (\text{B.26})$$

Тогда, действуя в обратном порядке, это уравнение можно привести к эквивалентной записи в пространстве состояний. Например, пусть в уравнение (B.26) входит лишь входной сигнал, но не его производные:

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1\dot{y} + a_0y = u, \quad (\text{B.27})$$

или в операторной записи

$$P(s)y = u, \quad P(s) = s^n + a_{n-1}s^{n-1} + \cdots + a_1s + a_0.$$

Вводя новые переменные x_i («состояния») по правилам

$$\begin{aligned} x_1 &= y, \\ x_2 &= \dot{y}, \\ &\vdots \\ x_n &= y^{(n-1)}, \end{aligned}$$

с учетом (B.27) получим

$$\begin{aligned}\dot{x}_1 &= x_2, \\ \dot{x}_2 &= x_3, \\ &\vdots \\ \dot{x}_n &= -a_0x_1 - a_1x_2 - \cdots - a_{n-1}x_n + u,\end{aligned}$$

или

$$\begin{aligned}\dot{x} &= Ax + bu, \\ y &= c^T x.\end{aligned}$$

Таким образом, получена *форма записи системы* (B.27) в пространстве состояний с

$$\begin{aligned}A &= \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ -a_0 & -a_1 & \dots & -a_{n-2} & -a_{n-1} \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}, \quad (\text{B.28}) \\ c &= \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.\end{aligned}$$

Нетрудно убедиться в том, что для матрицы A вида (B.28) полином $\det(sI - A) = P(s)$ является характеристическим полиномом системы (B.27).

Система с матрицами вида (B.28) (и вектором c произвольного вида) является так называемой *канонической управляемой формой* записи системы в пространстве состояний (при этом говорят, что матрица A приведена к *фробениусовой форме*). Оказывается, что при не слишком обременительных предположениях ее можно привести к такому виду линейными заменами переменных.

Закончим этот раздел кратким обзором классических частотных методов исследования одномерных систем. В инженерных приложениях различным физическим устройствам — звеньям — соответствуют так называемые типовые передаточные функции; они имеют специальные названия. Звено с передаточной функцией

$$H(s) = \frac{k}{Ts + 1}$$

называется *апериодическим (инерционным)*, при этом $k = H(0) > 0$ называется *коэффициентом усиления*, а $T > 0$ — *постоянной времени*.

Передаточная функция

$$H(s) = \frac{k}{T_2^2 s^2 + T_1 s + 1}, \quad k > 0,$$

отвечает *апериодическому звену второго порядка* при $T_1 \geq 2T_2 > 0$ (тогда корни характеристического полинома $1 + T_1 s + T_2^2 s^2$ вещественны) и *колебательному звену* при $0 < T_1 < 2T_2$ (случай комплексных корней).

Простейший элемент с

$$H(s) = \frac{k}{s}, \quad k > 0,$$

называется *идеальным интегратором*, а с

$$H(s) = k$$

— *идеальным усилителем*.

Для приведенных выше простых звеньев легко изобразить их частотные характеристики $H(j\omega)$. На рис. B.4 на комплексной плоскости показаны $H(j\omega)$ для апериодического звена первого (I) и второго (II) порядка, для колебательного звена (III) и интегратора (IV).

Часто удобно отдельно изображать амплитуду и фазу частотной характеристики (вспомним их физический смысл как реакции системы на гармоническое воздействие). Функция

$$A(\omega) \doteq |H(j\omega)|$$

называется *амплитудной частотной характеристикой* ($A \chi X$), она дает коэффициент усиления на входной сигнал частоты ω .

Функция

$$\varphi(\omega) \doteq \arg H(j\omega)$$

называется *фазовой частотной характеристикой* ($\Phi \chi X$), она характеризует сдвиг по фазе входа и выхода.

Наконец, во многих случаях удобно применять логарифмические частотные характеристики (*диаграммы Боде*), когда величины $A(\omega)$ и $\varphi(\omega)$ откладываются в логарифмической шкале. Именно, величина

$$\text{Lm}(\omega) \doteq 20 \log A(\omega)$$

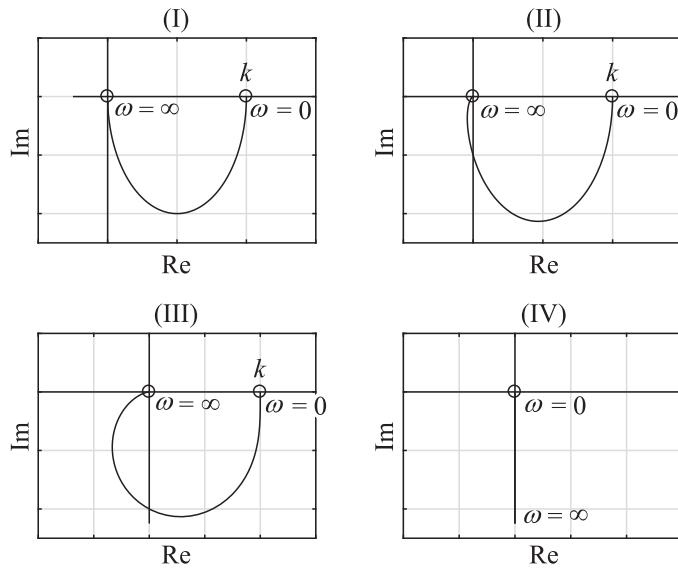


Рис. В.4. Частотные характеристики типовых звеньев.

называется *логарифмической амплитудной характеристикой (ЛАХ)*; она измеряется в децибелях. Как было показано, при последовательном соединении элементов их передаточные функции перемножаются. Поэтому

$$H(s) = H_1(s) \cdot \dots \cdot H_m(s),$$

$$A(\omega) = A_1(\omega) \cdot \dots \cdot A_m(\omega),$$

и

$$|H(j\omega)| = |H_1(j\omega)| \cdot \dots \cdot |H_m(j\omega)|,$$

$$\text{Lm}(\omega) = \text{Lm}_1(\omega) + \dots + \text{Lm}_m(\omega),$$

то есть логарифмические амплитудные характеристики при таком соединении складываются. При построении ЛАХ по оси абсцисс откладывается частота ω в логарифмической шкале; единицей измерения при этом является декада — отрезок, на котором частота увеличивается в 10 раз. По оси ординат откладывается величина $\text{Lm}(\omega)$, единицей измерения которой является децибел. Точка пересечения ЛАХ с осью абсцисс называется *частотой среза* ω_c , то есть

$$20 \log |H(j\omega_c)| = 0, \quad |H(j\omega_c)| = 1.$$

При построении *логарифмической фазовой характеристики* (*ЛФХ*) по оси абсцисс откладывается частота ω в декадах (то есть в логарифмической шкале), а по оси ординат — углы φ в градусах в равномерной шкале.

Пример В.6. На рис. В.5 показаны ЛАХ и ЛФХ для системы с передаточной функцией

$$\begin{aligned} H(s) &= \frac{s^2 - 0,25}{s^4 + 2s^3 - 0,79s^2 - 1,79s - 0,42} = \\ &= \frac{(s - 0,5)(s + 0,5)}{(s - 1)(s + 0,3)(s + 0,7)(s + 2)}. \end{aligned}$$

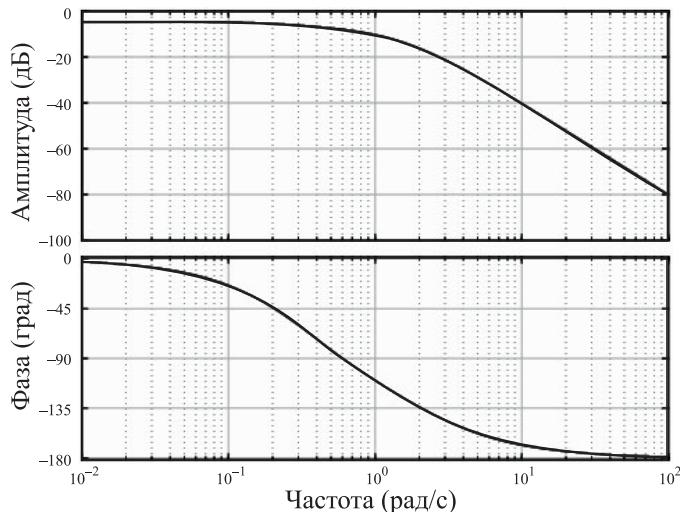


Рис. В.5. Диаграммы Боде.



B.7. Программное управление. Управляемость

Рассмотрим линейную непрерывную систему управления

$$\begin{aligned} \dot{x} &= Ax + Bu + D_1w, \\ y &= Cx + D_2w \end{aligned} \tag{B.29}$$

с состоянием $x(t) \in \mathbb{R}^n$, управлением $u(t) \in \mathbb{R}^p$, выходом $y(t) \in \mathbb{R}^\ell$ и внешним возмущением $w(t) \in \mathbb{R}^m$. Обсудим способ выбора управления как функции от времени, то есть в виде $u = u(t)$. Такой способ выбора называется *программным управлением*; на первый взгляд он представляется наиболее естественным.

Здесь прежде всего следует уточнить, в каком классе ищутся функции $u(t)$ и какие на них накладываются ограничения. Так, управление обычно предполагаются либо гладкими, либо непрерывными, либо произвольными измеримыми функциями времени. Чаще всего задача ставится именно в последней формулировке, то есть допускаются и разрывные управлений. Однако нередко оказывается, что для некоторых задач оптимальное управление достигается на дифференцируемых функциях от времени. Кроме того, на класс управлений обычно накладываются дополнительные условия типа ограниченности управлений. Типичным является условие

$$u(t) \in \mathcal{U} \quad \text{для всех } t \in [0, T],$$

где T — заданная длительность процесса управления, а \mathcal{U} — заданное замкнутое ограниченное множество в \mathbb{R}^p .

Другим типом ограничений являются интегральные, например,

$$J(u) = \int_0^T \|u(t)\|_2^2 dt \leq c^2 \quad (\text{B.30})$$

(то есть управления предполагаются L_2 -ограниченными на отрезке $[0, T]$). Функционал (B.30) называется *функционалом энергии*.

После выбора программного управления и подстановки его в систему (B.29) приходим к системе обыкновенных дифференциальных уравнений, в которых фигурируют лишь внешние воздействия. Например, уравнение состояния приобретает вид

$$\dot{x} = Ax + \underbrace{Bu(t) + D_1w(t)}_{w_1(t)},$$

и если величины $x(0) = x_0$ и $w(t)$ известны, то $x(t)$ может быть определено по формуле

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}w_1(\tau) d\tau$$

для всех значений $t \geq 0$.

Разумеется, программное управление может применяться и в дискретных системах (в этом случае управление u_k выбирается заранее как функция от k); никаких принципиально новых моментов при этом не возникает.

Используем программное управление для анализа важного понятия *управляемости*.

Определение В.1 (Определение управляемости на физическом уровне). Линейная система управления

$$\dot{x} = Ax + Bu \quad (\text{B.31})$$

с состоянием $x(t) \in \mathbb{R}^n$ и управлением $u(t) \in \mathbb{R}^p$ называется *управляемой*, если для любых точек $x_0, x_1 \in \mathbb{R}^n$ и любого момента времени $T > 0$ найдется ограниченное кусочно-непрерывное управление $u(t)$, переводящее систему из $x(0) = x_0$ в $x(T) = x_1$.

Свойство управляемости системы — естественное требование; оно означает, что имеются достаточные возможности воздействия на систему.

Замечание В.1. Важно отметить, что из свойств решений линейных дифференциальных уравнений вытекает, что управляемость системы эквивалентна выполнению следующего свойства: для любого начального $x(0)$ и любого $0 < T < \infty$ найдется ограниченное кусочно-непрерывное управление $u(t)$ на отрезке $[0, T]$ такое, что решение системы (B.31) переходит в начало координат в момент T . ▼

Иначе говоря, в управляемой системе начальное отклонение может быть устранено за (любое) конечное время. Очевидно, что изменив направление времени, получим, что управляемая система может быть переведена из $x(0) = 0$ в любую заданную точку $x(T)$; это свойство иногда называют *достижимостью*.

Трудность задачи проверки управляемости системы состоит в том, что обычно количество управляющих воздействий (размерность p пространства управлений) меньше количества управляемых величин (размерности n пространства состояний), однако она разрешается следующим образом.

Определение В.2. Назовем *матрицей управляемости* для системы (B.31) матрицу

$$U = (B \ AB \ \dots \ A^{n-r}B),$$

где r — ранг матрицы B .

Заметим, что матрица U составлена из $n - r + 1$ блоков, каждый из которых — матрица размера $n \times p$, то есть $U \in \mathbb{R}^{n \times (n-r+1)p}$.

Определение B.3. Если матрица управляемости имеет полный ранг, то есть

$$\operatorname{rank} U = n,$$

то пару матриц (A, B) называют *управляемой*.

В частности, для линейной системы управления

$$\dot{x} = Ax + bu$$

с состоянием $x(t) \in \mathbb{R}^n$ и скалярным входом $u(t) \in \mathbb{R}$ (при этом $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$) матрица управляемости будет квадратной (n -го порядка) со столбцами

$$b, Ab, \dots, A^{n-1}b.$$

Таким образом, пара (A, b) будет управляемой, если эти векторы линейно независимы. В этом случае систему можно с помощью преобразования подобия привести к *канонической управляемой форме*.

Следующая теорема дает конструктивное необходимое и достаточное условие управляемости.

Теорема B.1 (Ранговый критерий управляемости). *Линейная система (B.31) управляема тогда и только тогда, когда матрица управляемости имеет полный ранг.*

Доказательство. *Необходимость.* Пусть $\operatorname{rank} U < n$. Тогда найдется вектор $v \in \mathbb{R}^n$, $v \neq 0$, такой, что

$$v^T B = v^T AB = \dots = v^T A^{n-1}B = 0.$$

По теореме Гамильтона–Кэли матрица A удовлетворяет своему характеристическому уравнению:

$$A^n + a_{n-1}A^{n-1} + \dots + a_0I = 0, \quad (\text{B.32})$$

откуда

$$v^T A^n B = -a_{n-1}v^T A^{n-1}B - \dots - a_0v^T B = 0.$$

Аналогично, умножая равенство (B.32) слева на $v^T A^k$, $k = 1, 2, \dots$, и справа на B , получим

$$v^T A^k B = 0$$

для всех $k \geq n + 1$, откуда

$$v^T e^{-A\tau} B = v^T \left(I - A\tau + \frac{1}{2!} A^2 \tau^2 - \dots \right) B \equiv 0.$$

С другой стороны, согласно (B.4) решение системы (B.31) имеет вид

$$x(T) = e^{AT} x_0 + \int_0^T e^{A(T-\tau)} B u(\tau) d\tau = e^{AT} \left(x_0 + \int_0^T e^{-A\tau} B u(\tau) d\tau \right),$$

поэтому для $x(T) = 0$ имеем

$$0 = x_0 + \int_0^T e^{-A\tau} B u(\tau) d\tau,$$

или, после домножения слева на v^T ,

$$0 = v^T x_0 + \int_0^T \underbrace{v^T e^{-A\tau} B}_{=0} u(\tau) d\tau = v^T x_0$$

для любого управления $u(\tau)$. Однако равенство $v^T x_0 = 0$ не может выполняться при любых x_0 (например, при $x_0 = v$).

Достаточность. Укажем конкретный вид управления, которое переводит систему (B.31) из $x(0) = x_0$ в $x(T) = 0$. Именно, возьмем

$$\hat{u}(t) \doteq B^T e^{A^T(T-t)} v,$$

где вектор v подлежит выбору из условий

$$x(T) = 0, \quad x(0) = x_0,$$

то есть

$$\begin{aligned} 0 &= e^{AT} x_0 + \int_0^T e^{A(T-\tau)} B \hat{u}(\tau) v d\tau = \\ &= e^{AT} x_0 + \left(\int_0^T e^{A(T-\tau)} B B^T e^{A^T(T-\tau)} d\tau \right) v. \end{aligned}$$

Покажем, что матрица

$$W(T) \doteq \int_0^T e^{A(T-\tau)} BB^T e^{A^T(T-\tau)} d\tau = \int_0^T e^{A\tau} BB^T e^{A^T\tau} d\tau,$$

называемая *грацианом управляемости*, невырождена.

Прежде всего,

$$d^T W(T)d = \int_0^T |d^T e^{A\tau} B|^2 d\tau$$

для любого $d \in \mathbb{R}^n$, то есть эта матрица неотрицательно определена.
Если $d^T W(T)d = 0$ для некоторого $d \neq 0$, то

$$\varphi(\tau) = d^T e^{A\tau} B \equiv 0 \quad \text{для всех } \tau \in [0, T].$$

Но тогда и все производные функции φ равны нулю, в частности,

$$\varphi^{(k)}(0) = d^T A^k B = 0, \quad k = 0, 1, \dots, n - 1.$$

Следовательно, $d^T U = 0$, что противоречит условию $\text{rank } U = n$.

Таким образом, $W(T) \succ 0$, поэтому уравнение

$$e^{AT} x_0 + W(T)v = 0$$

имеет решение

$$v = -W^{-1}(T)e^{AT}x_0$$

при любом x_0 . Тем самым найдено управление

$$\hat{u}(t) = -B^T e^{A^T(T-t)} W^{-1}(T)e^{AT}x_0,$$

которое переводит систему (B.31) из состояния $x(0) = x_0$ в состояние $x(T) = 0$. ■

В случае когда

$$\text{rank } U = q < n,$$

рассмотрим подпространство \mathbb{R}^q , натянутое на столбцы матрицы управляемости. Нетрудно видеть, что на этом подпространстве система (B.31) обладает свойством управляемости, то есть для любого начального $x(0) \in \mathbb{R}^q$ и любого $0 < T < \infty$ найдется ограниченное

кусочно-непрерывное управление $u(t)$ на отрезке $[0, T]$ такое, что решение системы (B.31) обращается в нуль в момент T . Назовем \mathbb{R}^q управляемым подпространством рассматриваемой системы; в нем замечание B.1 сохраняет свою силу.

Далее, выше не просто доказано, что найдется управление, которое переводит систему из произвольного начального состояния в начало координат, но и явно указана одна такая функция: $u = \hat{u}(t)$. Более того, предложенное управление оказалось гладкой функцией, в то время как в определении управляемости требуется лишь кусочная непрерывность.

Замечание B.2. Рассмотрим линейную дискретную систему управления

$$x_{k+1} = Ax_k + Bu_k \quad (\text{B.33})$$

с состоянием $x_k \in \mathbb{R}^n$ и входом $u_k \in \mathbb{R}^p$, $\text{rank } B = r$. Согласно (B.6), решение системы (B.33) представимо в виде

$$x_k = A^k x_0 + \sum_{i=0}^{k-1} A^{k-i-1} Bu_i,$$

поэтому условию управляемости можно придать следующую форму: для всякого $a \in \mathbb{R}^n$ найдутся число $k > 0$ и векторы u_0, \dots, u_{k-1} такие, что

$$\sum_{i=0}^{k-1} A^{k-i-1} Bu_i = a. \quad (\text{B.34})$$

Отсюда вытекает критерий управляемости для линейных дискретных систем: система (B.33) управляема тогда и только тогда, когда пара (A, B) управляема. Действительно, взяв $k = n - r + 1$, перепишем уравнение (B.34) в виде

$$Uu = a,$$

где

$$U = (B \quad AB \quad \dots \quad A^{n-r}B) \in \mathbb{R}^{n \times (n-r+1)p},$$

$$u = \begin{pmatrix} u_{n-r} \\ u_{n-r-1} \\ \vdots \\ u_0 \end{pmatrix} \in \mathbb{R}^{(n-r+1)p}.$$

Остается заметить, что уравнение имеет решение при любом a тогда и только тогда, когда ранг матрицы U равен n . \blacktriangledown

В отличие от непрерывных систем дискретное время не может выбираться произвольно: требуется конечное число шагов, чтобы «накопить» ранг матрицы U . Если ранг матрицы B больше единицы, то систему можно привести в начало координат за число шагов, меньшее n ; в частности, если матрица B квадратная невырожденная, то можно достичь цели всего за один шаг.

Замечание B.3. Подход с точки зрения программного управления допустим в случае, когда внешние воздействия на систему отсутствуют, матрицы A , B и C известны и задан критерий оптимальности вида

$$J(u) = \int_0^T \|u(t)\|_2^2 dt.$$

В этом случае можно найти оптимальное управление

$$u = \hat{u}(t), \quad 0 \leq t \leq T.$$

Однако в более общих ситуациях — при наличии внешних возмущений или неопределенности в описании системы — применение программного управления может привести к резкому ухудшению качества процесса либо к полной катастрофе¹. Это же относится и к подавляющему большинству иных ситуаций, связанных с управлением производственными процессами, транспортом, системами связи, финансами и т. д. Лишь в очень небольшом числе случаев (при расчете оптимального режима космического полета или модели, в которых t не играет роль времени, например, при расчете оптимальной трассы дороги) решение в виде программного управления $u = u(t)$ является удовлетворительным. ▼

B.8. Управление по обратной связи

Рассмотрим линейную непрерывную систему управления

$$\begin{aligned} \dot{x} &= Ax + Bu + D_1w, \\ y &= Cx + D_2w \end{aligned} \tag{B.35}$$

¹ Представим себе, например, процесс управления самолетом, рассчитанный заранее, до начала полета, и не предусматривающий использования текущей информации о скорости ветра, высоте и т. п. Вряд ли кому-нибудь придет в голову управлять самолетом подобным образом.

с состоянием $x(t) \in \mathbb{R}^n$, управлением $u(t) \in \mathbb{R}^p$, выходом $y(t) \in \mathbb{R}^\ell$ и внешним возмущением $w(t) \in \mathbb{R}^m$.

В разделе B.7 в качестве управлений рассматривались заранее выбранные функции от времени; такое управление было названо *программным*. Другой подход к построению управления связан с идеей *обратной связи*, в рамках которого управление не выбирается заранее, а корректируется в каждый текущий момент на основании информации о состоянии (или о выходе) системы.

Выбор управления в форме функции от состояния и момента времени называется *комбинированным управлением*:

$$u = u(x, t). \quad (\text{B.36})$$

Функция $u(x, t)$ может быть нелинейной по x , и в общем случае при выборе управления в форме (B.36) уравнение состояния становится нелинейным и нестационарным. Существуют различные подходы (например, *динамическое программирование*), позволяющие решать задачи синтеза управления в некоторых постановках.

Рассмотрим управление в форме *статической линейной обратной связи по состоянию*

$$u = Kx, \quad (\text{B.37})$$

где матрица $K \in \mathbb{R}^{p \times n}$, называемая *матрицей регулятора* (или *матрицей усиления*), не зависит от t . Оказывается, что в некоторых задачах (например, в задаче о линейно-квадратичном регуляторе) управление такого типа обеспечивает наилучшее значение критерия оптимальности в классе любых управлений, то есть переход к нелинейным нестационарным обратным связям не улучшает критерий качества.

Подставив управление (B.37) в уравнение состояния (B.35), придем к замкнутой системе

$$\begin{aligned} \dot{x} &= (A + BK)x + D_1w, \\ y &= Cx + D_2w. \end{aligned}$$

Таким образом, после замыкания системы линейной обратной связью по состоянию (B.37) вновь получили линейное уравнение состояния — с матрицей состояния

$$A_c = A + BK,$$

которая называется *матрицей замкнутой системы*. За счет выбора матрицы усиления K можно оптимизировать те или иные показатели качества системы управления.

Управления в виде (B.37) называются *линейными (статическими) регуляторами*. Большинство используемых на практике регуляторов являются именно линейными; они легко реализуются технически.

Отметим, что возможности статических линейных регуляторов в некоторых отношениях ограничены. Так, регулятор в линейной системе $\dot{x} = Ax + Bu$ не может устраниć начальное отклонение за конечное время, то есть не может перевести систему из состояния $x_0 \neq 0$ в начало координат за конечное время T . Действительно, из уравнения замкнутой системы $\dot{x} = A_c x$ и условия $x(T) = 0$ очевидно следует, что $x(t) \equiv 0$ при всех t .

В дискретных системах линейная обратная связь имеет вид

$$u_k = Kx_k,$$

поэтому уравнения замкнутой системы приобретают вид

$$\begin{aligned} x_{k+1} &= A_c x_k + D_1 w_k, \\ y_k &= C x_k + D_2 w_k, \end{aligned}$$

где $A_c = A + BK$. Таким образом, матрица состояния A_c замкнутой дискретной системы вычисляется по той же формуле, что и в непрерывном случае.

Как отмечалось выше, состояние системы не всегда доступно измерению; часто единственная информация о системе предоставляется ее выходом y . Попытка построить регулятор в форме *статической линейной обратной связи по выходу*

$$u = Ky,$$

где $K \in \mathbb{R}^{p \times l}$, как правило, бывает неудовлетворительной; обычно с помощью управлений такого вида систему даже не удается сделать устойчивой.

Вместе с тем в реальных системах автоматического управления часто используется обратная связь в виде *пропорционально-интегрально-дифференцирующего регулятора*, или *ПИД-регулятора*. Для одномерной системы он имеет структуру

$$u = k_p y + k_i \int_0^t y + k_d \dot{y}, \quad (\text{B.38})$$

где k_p , k_i и k_d — некоторые числа, называемые *коэффициентами усиления* пропорциональной, интегрирующей и дифференцирующей составляющих регулятора соответственно.

Как видно из (B.38), ПИД-регулятор формирует управление, являющееся суммой трех слагаемых: первое из них пропорционально выходу y , второе — интегралу от выхода, а третье — его производной. При этом *настройка* ПИД-регулятора сводится к нахождению его коэффициентов усиления так, чтобы обеспечить наилучшее значение заданного критерия оптимальности.

B.9. Наблюдаемость

Управление может эффективно строиться исходя из текущего состояния системы, однако состояние не всегда доступно наблюдению; часто в нашем распоряжении имеются лишь измерения выходной переменной. В этом разделе обсудим принципиальную возможность восстановления состояния по выходу.

Определение B.4. Линейная динамическая система

$$\begin{aligned}\dot{x} &= Ax, \\ y &= Cx\end{aligned}\tag{B.39}$$

с состоянием $x(t) \in \mathbb{R}^n$ и выходом $y(t) \in \mathbb{R}^\ell$ называется *ненаблюдаемой*, если ее разным траекториям могут отвечать одинаковые выходы, то есть найдутся такие $x_0 \neq x'_0$, что для соответствующих траекторий $x(t), x'(t)$ и выходов $y(t), y'(t)$ будет выполнено $y(t) \equiv y'(t)$. В противном случае система называется *наблюдаемой*.

Покажем, что если линейная система наблюдаема, то можно восстановить значение $x(t)$ по значениям

$$y(t), \dot{y}(t), \dots, y^{(n-1)}(t),$$

то есть достаточно знать значения выхода и $n - 1$ его последовательных производных в тот момент времени, когда производится оценка состояния.

Определение B.5. Назовем *матрицей наблюдаемости* для линейной системы (B.39) матрицу

$$V = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-q} \end{pmatrix} \in \mathbb{R}^{\ell(n-q+1) \times n},$$

где q — ранг матрицы C .

Определение B.6. Если матрица наблюдаемости имеет полный ранг, то есть

$$\text{rank } V = n,$$

то пару матриц (C, A) называют *наблюдаемой*.

В следующей теореме устанавливается необходимое и достаточное условие наблюдаемости.

Теорема B.2 (Ранговый критерий наблюдаемости). *Линейная система (B.39) наблюдаема тогда и только тогда, когда матрица наблюдаемости имеет полный ранг.*

Доказательство. *Необходимость.* Пусть $\text{rank } V < n$. Тогда найдется вектор $v \in \mathbb{R}^n$, $v \neq 0$, такой, что

$$Cv = CAv = \dots = CA^{n-1}v = 0.$$

Как и при доказательстве необходимости в теореме B.1 об управляемости, отсюда следует, что

$$CA^k v = 0 \quad \text{для всех } k = 0, 1, 2, \dots$$

и, следовательно, $Ce^{At}v \equiv 0$. Поэтому при $x(0) = v$ имеем

$$y(t) = Cx(t) = Ce^{At}v \equiv 0.$$

С другой стороны, при $x(0) = 0$ также имеем $y(t) \equiv 0$. Таким образом, разным начальным условиям отвечают одинаковые выходы, поэтому система ненаблюдаема.

Достаточность. Поскольку выход $y(t)$ порожден линейной системой (B.39), имеем

$$\begin{aligned} y(t) &= Cx(t), \\ \dot{y}(t) &= CAx(t), \end{aligned}$$

⋮

$$y^{(n-1)}(t) = CA^{n-1}x(t),$$

то есть

$$Vx = Y,$$

где

$$x = x(t), \quad Y = \begin{pmatrix} y(t) \\ \dot{y}(t) \\ \vdots \\ y^{(n-1)}(t) \end{pmatrix} \in \mathbb{R}^{n\ell}.$$

Полученная линейная система имеет единственное решение, поскольку $\text{rank } V = n$. \blacksquare

Замечание В.4. Матрица

$$V^T \doteq (C^T \quad A^T C^T \quad \dots \quad (A^T)^{n-1} C^T)$$

является матрицей управляемости для линейной системы управления

$$\begin{aligned}\dot{x} &= A^T x + C^T u, \\ y &= B^T x,\end{aligned}$$

которая называется *двойственной* к линейной системе управления

$$\begin{aligned}\dot{x} &= Ax + Bu, \\ y &= Cx,\end{aligned}$$

поскольку управляемость одной из них эквивалентна наблюдаемости другой. \blacktriangledown

Г. Выпуклость множеств и функций.

Элементы негладкого анализа

Г.1. Выпуклость. Теорема Каратеодори. Непрерывность выпуклых функций

Определение Г.1. Множество $Q \subset \mathbb{R}^n$ называется *выпуклым*, если для любых точек $x_0, x_1 \in Q$ и для любого числа $\lambda \in [0, 1]$ выполнено включение $(1 - \lambda)x_0 + \lambda x_1 \in Q$ (см. рис. Г.1).

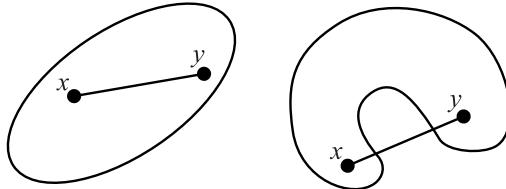


Рис. Г.1. Выпуклое (слева) и невыпуклое (справа) множество.

Выпуклое множество Q называется *строго выпуклым*, если его граница ∂Q не содержит отрезков.

Определение Г.2. Функция $f: \mathcal{Q} \ni x \rightarrow \mathbb{R}$, где $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое множество, называется *выпуклой*, если для любых точек $x_0, x_1 \in \mathcal{Q}$ и для любого числа $\lambda \in [0, 1]$ выполнено неравенство

$$f((1 - \lambda)x_0 + \lambda x_1) \leq (1 - \lambda)f(x_0) + \lambda f(x_1).$$

Функция $f: \mathcal{Q} \ni x \rightarrow \mathbb{R}$ называется *строго выпуклой*, если неравенство в определении Г.2 строгое:

$$f((1 - \lambda)x_0 + \lambda x_1) < (1 - \lambda)f(x_0) + \lambda f(x_1).$$

Геометрически строгая выпуклость функции f означает, что график f не содержит отрезков.

Легко видеть, что функция f выпукла, если ее *надграфик*

$$\text{epi } f = \{(x, \mu) : x \in \mathcal{Q}, \mu \geq f(x)\}$$

есть выпуклое множество в \mathbb{R}^{n+1} .

Функция $f: \mathcal{Q} \ni x \rightarrow \mathbb{R}$ называется *вогнутой*, если функция $-f$ выпуклая.

Часто выпуклую функцию $f: \mathcal{Q} \ni x \rightarrow \mathbb{R}$, где $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое множество, доопределяют на всем пространстве \mathbb{R}^n , полагая, что $f(x) = +\infty$ при всяком $x \in \mathbb{R}^n \setminus \mathcal{Q}$. Таким образом, мы будем частно рассматривать функции $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$. Легко видеть, что такое доопределение сохраняет справедливость определения Г.2.

Определение Г.3. Выпуклую функцию $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, которая принимает в некоторых точках конечные значения, будем называть *собственной*. Эффективным множеством функции f называется множество

$$\text{dom } f = \{x \in \mathbb{R}^n : f(x) < +\infty\}.$$

Определение Г.4. Пусть $\{x_k\}_{k=1}^m \subset \mathbb{R}^n$. Будем называть выражение $\sum_{k=1}^m \lambda_k x_k$, $\lambda_k \in \mathbb{R}$,

- *линейной комбинацией* точек $\{x_k\}_{k=1}^m$, если числа λ_k произвольны;
- *аффинной комбинацией*, если $\sum_{k=1}^m \lambda_k = 1$;
- *конической комбинацией*, если $\lambda_k \geq 0$ для всех k ;
- *выпуклой комбинацией*, если $\sum_{k=1}^m \lambda_k = 1$ и $\lambda_k \geq 0$ для всех k .

Лемма Г.1. *Множество $\mathcal{Q} \subset \mathbb{R}^n$ выпукло тогда и только тогда, когда оно содержит любую выпуклую комбинацию любого конечного числа своих точек.*

Аналогично лемме Г.1 легко доказать, что функция $f: \mathcal{Q} \rightarrow \mathbb{R}$ выпукла тогда и только тогда, когда для любой выпуклой комбинации точек $\sum_{k=1}^m \lambda_k x_k$ справедливо неравенство

$$f\left(\sum_{k=1}^m \lambda_k x_k\right) \leq \sum_{k=1}^m \lambda_k f(x_k). \quad (\text{Г.1})$$

Неравенство (Г.1) называется *неравенством Йенсена*. Частный случай этого неравенства (при $m = 2$) есть неравенство из определения Г.2; иногда именно оно называется неравенством Йенсена.

Пример Г.1. Рассмотрим $f(x) = -\ln x$ при $x > 0$. Поскольку $f''(x) = \frac{1}{x^2} > 0$, то функция f выпуклая [14]. Поэтому для любой выпуклой комбинации $\sum_{k=1}^m \lambda_k x_k$ имеем

$$-\ln\left(\sum_{k=1}^m \lambda_k x_k\right) \leq -\sum_{k=1}^m \lambda_k \ln x_k$$

или, потенцируя, получаем неравенство о средних следующего вида:

$$\prod_{k=1}^m x_k^{\lambda_k} \leq \sum_{k=1}^m \lambda_k x_k \quad \forall m \in \mathbb{N}, \quad x_k > 0, \quad \lambda_k \geq 0, \quad \sum_{k=1}^m \lambda_k = 1. \quad \blacktriangledown$$

Множество \mathcal{Q} будем называть *конусом*, если из условия $x \in \mathcal{Q}$ для любого $\lambda > 0$ следует $\lambda x \in \mathcal{Q}$.

Определение Г.5. Пусть $\mathcal{Q} \subset \mathbb{R}^n$. Будем называть соответственно *линейной, аффинной и выпуклой оболочкой* множества \mathcal{Q} минимальное по включению соответственно линейное множество, аффинное множество или выпуклое множество, содержащее \mathcal{Q} . Будем обозначать их через $\text{lin } \mathcal{Q}$, $\text{aff } \mathcal{Q}$ и $\text{co } \mathcal{Q}$ соответственно.

Лемма Г.2. *Для произвольного множества $\mathcal{Q} \subset \mathbb{R}^n$ его выпуклая оболочка со \mathcal{Q} совпадает со множеством*

$$\mathcal{X} = \left\{ \sum_{k=1}^m \lambda_k x_k : \quad m \in \mathbb{N}, \quad \lambda_k \geq 0, \quad \sum_{k=1}^m \lambda_k = 1, \quad x_k \in \mathcal{Q} \right\}$$

всевозможных выпуклых комбинаций точек из \mathcal{Q}

Лемма легко вытекает из определений, см. [19, теорема 1.2.2].

Недостатком леммы Г.2 является достаточно сложное описание выпуклой оболочки множества \mathcal{Q} , поскольку выпуклые комбинации могут содержать любое число слагаемых m . В таком виде структура множества со \mathcal{Q} неясна даже для случая компактного множества \mathcal{Q} .

Теорема Г.1 (Каратеодори). Для любого множества $\mathcal{Q} \subset \mathbb{R}^n$

$$\text{co } \mathcal{Q} = \left\{ \sum_{k=1}^{n+1} \lambda_k x_k : \quad \lambda_k \geq 0, \quad \sum_{k=1}^{n+1} \lambda_k = 1, \quad x_k \in \mathcal{Q} \right\}.$$

Следствие Г.1. Пусть множество $\mathcal{Q} \subset \mathbb{R}^n$ компактно. Тогда $\text{co } \mathcal{Q}$ — компактное множество.

Доказательство. Утверждение следует из компактности множества $[0, 1]^{n+1} \times \mathcal{Q}^{n+1}$, непрерывности функции

$$\varphi: [0, 1]^{n+1} \times \mathcal{Q}^{n+1} \rightarrow \mathbb{R}^n, \quad \varphi(\lambda_1, \dots, \lambda_{n+1}, x_1, \dots, x_{n+1}) = \sum_{k=1}^{n+1} \lambda_k x_k$$

и равенства $\varphi([0, 1]^{n+1} \times \mathcal{Q}^{n+1}) = \text{co } \mathcal{Q}$, которое вытекает из теоремы Каратеодори. ■

Аналогично теореме Каратеодори доказываются два следующие утверждения.

Утверждение Г.1. Пусть множество $\mathcal{Q} \subset \mathbb{R}^n$ есть невыпуклый конус. Тогда

$$\text{co } \mathcal{Q} = \left\{ \sum_{k=1}^n \lambda_k x_k : \quad \lambda_k \geq 0, \quad \sum_{k=1}^n \lambda_k = 1, \quad x_k \in \mathcal{Q} \right\}.$$

Утверждение Г.2. Пусть множество $\mathcal{Q} \subset \mathbb{R}^n$ — линейно связное, то есть любые его две точки можно соединить непрерывной кривой. Тогда

$$\text{co } \mathcal{Q} = \left\{ \sum_{k=1}^n \lambda_k x_k : \quad \lambda_k \geq 0, \quad \sum_{k=1}^n \lambda_k = 1, \quad x_k \in \mathcal{Q} \right\}.$$

Определение Г.6. Суммой Минковского (или суммой Минковского–Понтрягина) двух множеств $\mathcal{P}, \mathcal{Q} \subset \mathbb{R}^n$ называется множество

$$\mathcal{X} = \mathcal{P} + \mathcal{Q} = \{x + y \in \mathbb{R}^n : x \in \mathcal{P}, y \in \mathcal{Q}\}.$$

Отметим, что операция $\mathcal{X} + (-\mathcal{P})$ не является обратной к сумме Минковского. Легко видеть, что если множества \mathcal{P} и \mathcal{Q} не одноточечные, то $\mathcal{P} + \mathcal{Q} + (-\mathcal{P}) \supset \mathcal{Q}$, но равенства здесь нет.

Мы будем рассматривать выпуклые функции, принимающие не только конечные значения, но и значение $+\infty$. Это удобно, например, при работе со множествами. Для произвольного множества $\mathcal{Q} \subset \mathbb{R}^n$ вводится его индикаторная функция $\delta(x, \mathcal{Q})$, которая есть 0 при $x \in \mathcal{Q}$ и $+\infty$ при $x \notin \mathcal{Q}$. Определяется также *опорная функция* множества $\mathcal{Q} \subset \mathbb{R}^n$: для всякого $p \in \mathbb{R}^n$ ее значение равно $s(p, \mathcal{Q}) = \sup_{x \in \mathcal{Q}} \langle p, x \rangle$. Опорная функция положительно однородна: $s(\lambda p, \mathcal{Q}) = \lambda s(p, \mathcal{Q})$ для любого $\lambda \geq 0$ и $p \in \mathbb{R}^n$. Поскольку $\text{epi } s(\cdot, \mathcal{Q}) = \bigcap_{x \in \mathcal{Q}} \text{epi}(\cdot, x)$, то надграфик $\text{epi } s(\cdot, \mathcal{Q})$ есть выпуклое замкнутое множество. Поэтому функция $s(p, \mathcal{Q})$ выпукла и полуунепрерывна снизу, см. [19, лемма 1.6.2]. Если множество \mathcal{Q} компактно, то функция $s(p, \mathcal{Q})$ удовлетворяет по p условию Липшица с константой $L = \max_{x \in \mathcal{Q}} \|x\|$. При неограниченном множестве \mathcal{Q} для некоторых ненулевых векторов p может быть $s(p, \mathcal{Q}) = +\infty$.

Выпуклая функция может быть разрывна.

Пример Г.2. Определим в \mathbb{R}^2 множество

$$\mathcal{Q} = \{(x_1, x_2) : x_1^2 + x_2^2 \leq 1\}$$

и функцию $f: \mathcal{Q} \rightarrow \mathbb{R}$ такую, что $f(x_1, x_2) = 0$ при $x_1^2 + x_2^2 < 1$. Если же $x_1^2 + x_2^2 = 1$ и $(x_1, x_2) \neq (1, 0)$, то положим $f(x_1, x_2) = \varphi$, где φ — угол, отсчитанный против часовой стрелки от положительного направления оси Ox_1 к вектору (x_1, x_2) . Наконец, пусть $f(1, 0) = \pi$.

Построенная функция f выпукла на \mathcal{Q} (поскольку каждая хорда, соединяющая точки графика f , лежит выше графика) и непрерывна на внутренности \mathcal{Q} . Однако f разрывна во всех точках границы \mathcal{Q} , причем в точке $(1, 0)$ функция f не полуунепрерывна ни сверху, ни снизу. ▼

Непрерывность f на внутренности \mathcal{Q} в предыдущем примере не случайна. Приведем соответствующий результат.

Теорема Г.2 ([19]). Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое открытое подмножество и $f: \mathcal{Q} \rightarrow \mathbb{R}$ — выпуклая функция. Тогда f локально липшицева на \mathcal{Q} : для любой точки $x_0 \in \mathcal{Q}$ существуют числа $\delta(x_0) > 0$ и $L(x_0) > 0$ такие, что для всех $x_1, x_2 \in \text{int } \mathcal{B}_{\delta(x_0)}(x_0) \cap \mathcal{Q}$ выполнено неравенство

$$|f(x_1) - f(x_2)| \leq L(x_0) \|x_1 - x_2\|.$$

Г.2. Метрическая проекция точки на выпуклое замкнутое множество. Теоремы об отделимости

Расстоянием от точки $x_0 \in \mathbb{R}^n$ до множества $\mathcal{Q} \subset \mathbb{R}^n$ называется величина $\varrho(x_0, \mathcal{Q}) = \inf_{x \in \mathcal{Q}} \|x - x_0\|$.

Из условия Липшица с константой 1 для нормы для любых точек $x_1, x_2 \in \mathbb{R}^n$ вытекает неравенство $|\varrho(x_1, \mathcal{Q}) - \varrho(x_2, \mathcal{Q})| \leq \|x_1 - x_2\|$. Множество \mathcal{Q} при этом произвольно.

Определение Г.7. *Метрической проекцией* точки $x_0 \in \mathbb{R}^n$ на множество $\mathcal{Q} \subset \mathbb{R}^n$ называется множество

$$P_{\mathcal{Q}}x_0 = \{x \in \mathcal{Q}: \|x - x_0\| = \varrho(x_0, \mathcal{Q})\}.$$

Оператор метрического проектирования является важным объектом исследования в анализе и его приложениях и его можно определять с помощью различных норм и даже метрик. Здесь будет изучаться только случай евклидовой нормы $\|x\| = \sqrt{(x, x)}$ для любого $x \in \mathbb{R}^n$.

Заметим, что множество $P_{\mathcal{Q}}x_0$ может быть пустым, например, когда $x_0 \notin \mathcal{Q}$ и \mathcal{Q} — открытое множество. Также множество $P_{\mathcal{Q}}x_0$ может состоять из более чем одной точки, например, когда $x_0 = 0$ и $\mathcal{Q} = \{x \in \mathbb{R}^n: \|x\| = 1\}$. В последнем случае $P_{\mathcal{Q}}x_0 = \mathcal{Q}$.

Лемма Г.3. Для выпуклого замкнутого подмножества $\mathcal{Q} \subset \mathbb{R}^n$ множество $P_{\mathcal{Q}}x_0$ одноточечно.

Доказательство. Достаточно рассмотреть случай $x_0 \notin \mathcal{Q}$. Положим $r = \varrho(x_0, \mathcal{Q}) > 0$. Тогда $P_{\mathcal{Q}}x_0 = P_{\mathcal{Q}_r}x_0$, где $\mathcal{Q}_r = B_{r+1}(x_0) \cap \mathcal{Q}$.

Множество \mathcal{Q}_r компактно, поэтому задача $\min_{x \in \mathcal{Q}_r} \|x - x_0\|$ имеет решение $x \in \mathcal{Q}_r$ по теореме Вейерштрасса А.1. Значит, $P_{\mathcal{Q}}x_0 \neq \emptyset$.

Предположим, что $\{x_1, x_2\} \subset P_{\mathcal{Q}}x_0$, $x_1 \neq x_2$. В силу выпуклости \mathcal{Q} отрезок $[x_1, x_2]$ целиком содержится в \mathcal{Q} , значит и $z = \frac{x_1+x_2}{2} \in \mathcal{Q}$. Поскольку в треугольнике x_0x_1z гипotenуза x_0x_1 строго больше катета x_0z , получаем противоречие равенству $\|x_0 - x_1\| = \min_{x \in \mathcal{Q}} \|x_0 - x\|$. ■

Теорема Г.3. Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое множество и $x_0 \notin \mathcal{Q}$. Тогда существует единичный вектор $p_0 \in \mathbb{R}^n$ и число $\varepsilon > 0$ такие, что $(p_0, x_0) - \varepsilon \geq (p_0, x)$ для всех $x \in \mathcal{Q}$.

Доказательство. В силу леммы Г.3 множество $P_{\mathcal{Q}}x_0 = \{y\}$ одноточечно. Определим $p_0 = \frac{x_0 - y}{\|x_0 - y\|}$ и $\varepsilon = \|x_0 - y\|$.

Допустим, что для точки $x \in \mathcal{Q}$ верно $(p_0, x) > (p_0, x_0) - \varepsilon = (p_0, y)$; см. рис. Г.2. Тогда, положив $x(\lambda) = \lambda x + (1 - \lambda)y \in \mathcal{Q}$ для $\lambda \in [0, 1]$, получаем

$$\|x(\lambda) - x_0\|^2 = \|x_0 - y\|^2 - 2\lambda(x_0 - y, x - y) + \lambda^2\|x - y\|^2.$$

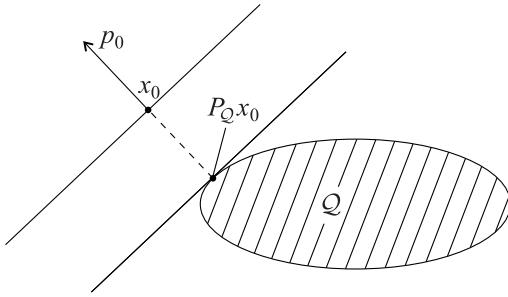


Рис. Г.2. Иллюстрация к теореме Г.3.

Поскольку $(x_0 - y, x - y) > 0$, то при малых $\lambda > 0$ для точки $x(\lambda) \in \mathcal{Q}$ имеем $\|x(\lambda) - x_0\| < \|x_0 - y\| = \varrho(x_0, \mathcal{Q})$. В результате приходим к противоречию. ■

Теорема Г.4 ([19, §1.9]). Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое множество, а $\mathcal{P} \subset \mathbb{R}^n$ — выпуклое компактное множество и $\mathcal{Q} \cap \mathcal{P} = \emptyset$. Тогда существуют единичный вектор $p_0 \in \mathbb{R}^n$ и число $\varepsilon > 0$ такие, что $(p_0, x) - \varepsilon \geq (p_0, y)$ для всех $x \in \mathcal{Q}$, $y \in \mathcal{P}$.

Пусть в \mathbb{R}^2

$$\mathcal{Q} = \{(x_1, 0)^T : x_1 \in \mathbb{R}\}, \quad \mathcal{P} = \left\{ (x_1, x_2)^T : x_1 > 0, x_2 \geq \frac{1}{x_1} \right\}.$$

Оба множества выпуклы, замкнуты и не компактны, $\mathcal{Q} \cap \mathcal{P} = \emptyset$. При этом множество $\mathcal{P} - \mathcal{Q} = \{(x_1, x_2)^T : x_2 > 0\}$ открыто.

Результат теоремы Г.4 называется сильной отделимостью: имеется «запас» $\varepsilon > 0$.

Приведем другие варианты теоремы об отделимости.

Теорема Г.5 ([19, §1.9]). 1) Пусть $\mathcal{Q}, \mathcal{P} \subset \mathbb{R}^n$ — выпуклые множества и $\inf_{x \in \mathcal{Q}, y \in \mathcal{P}} \|x - y\| > 0$. Тогда существует единичный вектор p_0 и $\varepsilon > 0$ такие, что $(p_0, x) - \varepsilon \geq (p_0, y)$ для всех $x \in \mathcal{Q}$, $y \in \mathcal{P}$.

2) Пусть $\mathcal{Q}, \mathcal{P} \subset \mathbb{R}^n$ — выпуклые замкнутые множества такие, что $\mathcal{Q} \cap \mathcal{P} = \emptyset$. Тогда существует такой единичный вектор $p_0 \in \mathbb{R}^n$, что для всех $x \in \mathcal{Q}$, $y \in \mathcal{P}$ выполнено неравенство $(p_0, x) \geq (p_0, y)$.

3) Пусть $\mathcal{Q}, \mathcal{P} \subset \mathbb{R}^n$ — выпуклые множества, $\mathcal{Q} \cap \mathcal{P} = \emptyset$, причем \mathcal{Q} открыто. Тогда существует единичный вектор $p_0 \in \mathbb{R}^n$ такой, что $(p_0, x) > (p_0, y)$ для всех $x \in \mathcal{Q}$, $y \in \mathcal{P}$.

Теорема Г.6 ([19, теорема 1.9.2]). Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое подмножество и $x \in \partial\mathcal{Q}$. Тогда существует такой единичный вектор $p \in \mathbb{R}^n$, что

$$(p, x) = s(p, \mathcal{Q}).$$

Теорема Г.7 ([19, следствие 1.9.5]). Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое подмножество. Тогда

$$\mathcal{Q} = \{x \in \mathbb{R}^n : (p, x) \leq s(p, \mathcal{Q}) \quad \forall \|p\| = 1\}.$$

Теорема Г.7 означает, что если взять пересечение всех замкнутых полупространств, каждое из которых содержит выпуклое замкнутое множество \mathcal{Q} , то получится в точности множество \mathcal{Q} .

Отметим, что метрическая проекция точки на выпуклое замкнутое множество в \mathbb{R}^n удовлетворяет по точке условию Липшица с константой Липшица 1, то есть для всякого выпуклого замкнутого подмножества $\mathcal{Q} \subset \mathbb{R}^n$ и любых точек $x_0, x \in \mathbb{R}^n$ выполнена оценка

$$\|P_{\mathcal{Q}}x - P_{\mathcal{Q}}x_0\| \leq 1 \cdot \|x - x_0\|.$$

Это прямое следствие теоремы Г.3. Оно крайне важно в проксимальных методах оптимизации.

Прямым следствием теорем об отделимости являются свойства полярного множества. Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — произвольное множество. *Полярой* \mathcal{Q} называется множество

$$\mathcal{Q}^0 = \{p \in \mathbb{R}^n : (p, x) \leq 1 \quad \forall x \in \mathcal{Q}\}.$$

Очевидно, что полярное множество \mathcal{Q}^0 выпукло, замкнуто и содержит ноль. Множество

$$\mathcal{Q}^{00} = (\mathcal{Q}^0)^0 = \{x \in \mathbb{R}^n : (p, x) \leq 1 \quad \forall p \in \mathcal{Q}^0\}$$

называется второй полярой к \mathcal{Q} . Очевидно, что $\mathcal{Q} \subset \mathcal{Q}^{00}$.

Теорема Г.8 ([19, теорема 1.12.1]). Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое подмножество, $0 \in \mathcal{Q}$. Тогда

$$\mathcal{Q} = \mathcal{Q}^{00}.$$

Г.3. Выпуклые конусы¹

Множество $\mathcal{K} \subset \mathbb{R}^n$ называется *конусом*, если для любого $x \in \mathcal{K}$ имеет место

$$\lambda x \in \mathcal{K} \quad \forall \lambda > 0.$$

Если $0 \in \mathcal{K}$, то конус \mathcal{K} часто называют *заостренным*. Множество \mathcal{K} называется выпуклым конусом, если оно выпукло и в то же время является конусом. Очевидно, для того чтобы конус \mathcal{K} являлся выпуклым конусом, необходимо и достаточно, чтобы выполнялось

$$x_1 + x_2 \in \mathcal{K} \quad \forall x_1, x_2 \in \mathcal{K}.$$

В настоящей книге рассматриваются главным образом замкнутые конусы. Очевидно, что замкнутый конус является заостренным, так как нуль принадлежит замыканию любого непустого конуса.

Приведем простейшие примеры конусов. Выпуклым конусом являются: любое линейное подпространство; неотрицательный ортант в \mathbb{R}^n , который состоит из таких векторов $x = (x_1, \dots, x_n)$, что все $x_i \geq 0$; множество квадратных неотрицательно определенных матриц и т. д.

Пусть \mathcal{K} — конус. Конус

$$\mathcal{K}^* = \{p \in \mathbb{R}^n : (p, x) \geq 0 \quad \forall x \in \mathcal{K}\}$$

называется *сопряженным конусом* к \mathcal{K} .

Очевидно, сопряженный конус \mathcal{K}^* всегда является выпуклым и замкнутым конусом независимо от того, является конус \mathcal{K} выпуклым и замкнутым или нет. Наконец, если верно включение $\mathcal{K}_1 \subset \mathcal{K}_2$, то $\mathcal{K}_2^* \subset \mathcal{K}_1^*$.

Отметим, что наряду с сопряженным часто рассматривают *нормальный конус* к конусу \mathcal{K} , который обозначается через \mathcal{K}^0 и определяется соотношением

$$\mathcal{K}^0 = \{p \in \mathbb{R}^n : (p, x) \leq 0 \quad \forall x \in \mathcal{K}\}.$$

Очевидно, $\mathcal{K}^0 = -\mathcal{K}^*$. Наряду с термином «нормальный» используется также термин «полярный конус». Ограничимся рассмотрением сопряженных конусов.

Будем использовать обозначение $\mathcal{K}^{**} = (\mathcal{K}^*)^*$. Конус \mathcal{K}^{**} называют вторым сопряженным к конусу \mathcal{K} .

¹При написании раздела были использованы материалы, предоставленные А.В. Арутюновым.

Утверждение Г.3. Имеет место включение $\mathcal{K} \subset \mathcal{K}^{**}$.

Если конус \mathcal{K} выпуклый и замкнутый, то, как и всякое выпуклое замкнутое множество, содержащее 0, он удовлетворяет условию $\mathcal{K}^{**} = \mathcal{K}$, см. теорему Г.8.

Конус, представимый в виде

$$\mathcal{K} = \left\{ x : x = \sum_{i=1}^m \lambda_i a_i, \lambda_i \geq 0 \right\},$$

где m натуральное число, а $a_i \in \mathbb{R}^n$, $1 \leq i \leq m$, — заданные векторы, называется *конечно порожденным*. Отметим, что конечно порожденный конус еще называют *многогранным*.

Лемма Г.4. Конечно порожденный конус замкнут.

Наряду с конечно порожденными конусами, в приложениях часто возникают так называемые *полиэдральные конусы*. Полиэдральным называется конус, представимый в виде

$$\mathcal{K} = \{x \in \mathbb{R}^n : (a_i, x) = 0, (b_j, x) \leq 0, i = 1, \dots, m_1, j = 1, \dots, m_2\},$$

где m_1, m_2 — заданные натуральные числа, а a_i и b_i — заданные векторы. Оказывается, что в конечномерных пространствах между конечно порожденными и полиэдральными конусами существует тесная взаимосвязь. А именно, справедливо следующее общее утверждение, восходящее к Г. Минковскому и Г. Вейлю.

Теорема Г.9. Любой конечно порожденный конус является полиэдральным, и наоборот: любой полиэдральный конус является конечно порожденным.

Доказательства этих известных результатов можно найти, например, в [24, с. 188–190] и [1]. Отметим, что из теоремы Г.9 вытекает справедливость леммы Г.4 о конечно порожденном конусе, поскольку любой полиэдральный конус замкнут, что несложно проверить.

Вычисление сопряженных конусов весьма важно для приложений. Поэтому приведем одну формулу, помогающую вычислять сопряженные конусы. Начнем с утверждения, дающего формулу для вычисления сопряженного конуса в следующем важном случае. Для вектора $x \in \mathbb{R}^n$ будем писать $x \geq 0$, если все его координаты неотрицательны, и $x \leq 0$, если они неположительны.

Лемма Г.5 (Фаркаш [1]). Пусть для конуса $\mathcal{K} \subset \mathbb{R}^n$ имеет место представление

$$\mathcal{K} = \{x \in \mathbb{R}^n : Ax = 0, Bx \geq 0\},$$

где A, B — заданные матрицы размерностей $m \times n$ и $k \times n$ соответственно. Тогда для \mathcal{K}^* справедлива формула

$$\mathcal{K}^* = \{y \in \mathbb{R}^n : y = A^T a + B^T b, a \in \mathbb{R}^m, b \in \mathbb{R}^k, b \geq 0\}.$$

Г.4. Теоремы о крайних точках и лучах

Определение Г.8. Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое подмножество. Точка $x \in \mathcal{Q}$ называется *крайней* для множества \mathcal{Q} , если множество $\mathcal{Q} \setminus \{x\}$ выпукло. Множество крайних точек \mathcal{Q} будем обозначать через $\text{extr } \mathcal{Q}$.

Пусть $x \in \mathcal{Q}$ и $y \in \mathbb{R}^n \setminus \{0\}$. Луч $l(x, y) = \{x + \lambda y : \lambda \geq 0\}$ называется *крайним* для множества \mathcal{Q} , если $l(x, y) \subset \mathcal{Q}$, множество $\mathcal{Q} \setminus l(x, y)$ выпукло и для любого $\lambda < 0$ имеем $x + \lambda y \notin \mathcal{Q}$. Множество крайних лучей \mathcal{Q} будем обозначать через $\text{rect } \mathcal{Q}$, см. рис. Г.3.

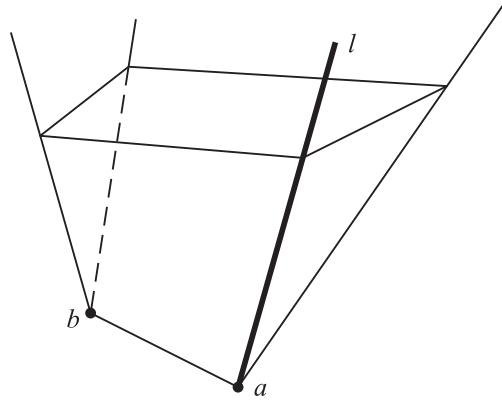


Рис. Г.3. Точки a и b — крайние, луч l — крайний.

Другими словами, точка $x \in \mathcal{Q}$ крайняя, если для любых $y, z \in \mathcal{Q}$, $y \neq x$, $z \neq x$ имеем $x \notin [y, z]$. Также легко видеть, что для крайнего луча $l(x, y) \subset \mathcal{Q}$ точка $x \in \text{extr } \mathcal{Q}$.

Заметим, что любой цилиндр в \mathbb{R}^n , то есть множество вида $\mathcal{Q} + l$, где l — прямая (то есть $l = \{(1 - \lambda)x + \lambda y: \lambda \in \mathbb{R}\}$, $x \neq y$ из \mathbb{R}^n), а \mathcal{Q} — выпуклое замкнутое подмножество, не имеет ни крайних точек, ни крайних лучей. Следующая теорема показывает, что для множества, которое не содержит прямых, набор крайних точек и лучей его вполне характеризует.

Теорема Г.10 (Кли [19, теорема 1.18.4]). *Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое множество, которое не содержит прямых. Тогда $\text{extr } \mathcal{Q} \neq \emptyset$ и справедливо равенство $\mathcal{Q} = \text{co}(\text{extr } \mathcal{Q} \cup \text{rect } \mathcal{Q})$.*

Если $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое подмножество, $x_0 \in \partial \mathcal{Q}$ и $H = \{x: (p, x - x_0) = 0\}$ — опорная гиперплоскость к \mathcal{Q} (то есть $p \neq 0$ и $(p, x) \leqslant (p, x_0)$ для всех $x \in \mathcal{Q}$), то $\text{extr}(\mathcal{Q} \cap H) = (\text{extr } \mathcal{Q}) \cap H$ и $\text{rect}(\mathcal{Q} \cap H) = (\text{rect } \mathcal{Q}) \cap H$ [19, лемма 1.18.1].

Более известен следующий вариант теоремы Г.10, который был доказан Г. Минковским для \mathbb{R}^n в начале XX века, а затем М. Крейном и Д. Мильманом для бесконечномерных пространств [19, теорема 1.18.3].

Теорема Г.11 (Минковский). *Пусть множество $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклый компакт. Тогда $\mathcal{Q} = \text{co}(\text{extr } \mathcal{Q})$.*

Для выпуклого замкнутого множества $\mathcal{Q} \subset \mathbb{R}^n$ можно определить асимптотический (или рецессивный) конус

$$O^+ \mathcal{Q} = \{p \in \mathbb{R}^n: x + \lambda p \in \mathcal{Q} \quad \forall \lambda \geqslant 0 \quad \forall x \in \mathcal{Q}\}. \quad (\Gamma.2)$$

Заметим, что $O^+ \mathcal{Q}$ есть выпуклый замкнутый конус. Легко видеть, что множество \mathcal{Q} ограничено тогда и только тогда, когда $O^+ \mathcal{Q} = \{0\}$.

В случае произвольного замкнутого выпуклого множества \mathcal{Q} , которое не содержит прямых, имеет место равенство

$$\mathcal{Q} = \text{co}(\text{extr } \mathcal{Q}) + O^+ \mathcal{Q}. \quad (\Gamma.3)$$

Докажем его. Из включения $\text{extr } \mathcal{Q} \cup \text{rect } \mathcal{Q} \subset \text{extr } \mathcal{Q} + O^+ \mathcal{Q}$ и теоремы Г.10 вытекает включение

$$\mathcal{Q} = \text{co}(\text{extr } \mathcal{Q} \cup \text{rect } \mathcal{Q}) \subset \text{co}(\text{extr } \mathcal{Q} + O^+ \mathcal{Q}) = \text{co}(\text{extr } \mathcal{Q}) + O^+ \mathcal{Q}.$$

Включение в обратную сторону следует из импликации

$$\text{extr } \mathcal{Q} + O^+ \mathcal{Q} \subset \mathcal{Q} \Rightarrow \text{co}(\text{extr } \mathcal{Q} + O^+ \mathcal{Q}) \subset \mathcal{Q}.$$

Еще одно важное следствие теоремы Г.10 — условия достижимости максимума выпуклой функции в крайней точке выпуклого замкнутого множества.

Теорема Г.12. Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое множество, которое не содержит прямых. Пусть для любого $v \in O^+ \mathcal{Q}$ и для любого $x \in \text{co extr } \mathcal{Q}$ выполнено условие $f(x + v) \leq f(x)$. Тогда

$$\sup_{x \in \mathcal{Q}} f(x) = \sup_{x \in \text{extr } \mathcal{Q}} f(x).$$

Доказательство. Пусть $\{x_k\}$ — минимизирующая последовательность, $f(x_k) \rightarrow \sup_{x \in \mathcal{Q}} f(x)$. Тогда $x_k = z_k + v_k$, где $z_k \in \text{co extr } \mathcal{Q}$. Тогда по

теореме Каратеодори $z_k = \sum_{i=1}^{n+1} \lambda_i^k u_i^k$, $u_i^k \in \text{extr } \mathcal{Q}$, а $\lambda_i^k \geq 0$, $\sum_{i=1}^{n+1} \lambda_i^k = 1$.

При этом $v_k \in O^+ \mathcal{Q}$. Из условия имеем $f(x_k) = f(z_k + v_k) \leq f(z_k)$ и

$$f(x_k) \leq f(z_k) \leq \sum_{i=1}^{n+1} \lambda_i^k f(u_i^k) \leq \max_{1 \leq i \leq n+1} f(u_i^k) \leq \sup_{x \in \text{extr } \mathcal{Q}} f(x),$$

то есть $\sup_{x \in \mathcal{Q}} f(x) \leq \sup_{x \in \text{extr } \mathcal{Q}} f(x)$. Обратное неравенство очевидно. ■

Г.5. Субдифференциал.

Теорема Моро – Рокафеллара

Определение Г.9. Пусть $f: \mathbb{R}^n \supset \mathcal{Q} \rightarrow \mathbb{R} \cup \{+\infty\}$ — собственная выпуклая функция. Субградиентом функции f в точке x будем называть такой вектор $p \in \mathbb{R}^n$, что для всех $z \in \mathbb{R}^n$ выполнено неравенство $f(z) \geq f(x) + (p, z - x)$.

Множество всех субградиентов f в точке x будем называть *субдифференциалом* функции f в точке x и обозначать через $\partial f(x)$.

На рис. Г.4 схематично изображен субдифференциал функции f в точке x_0 . Нормали $(p_i, -1)$, $i = 1, 2$, перпендикулярны опорным гиперплоскостям к надграфику $\text{epi } f$ в точке $(x_0, f(x_0))$, которые обозначены на рисунке пунктирными линиями.

Если функция f вогнута, то для нее вводится понятие *супердифференциала*

$$\partial^+ f(x) = \{p \in \mathbb{R}^n : f(z) \leq f(x) + (p, z - x) \quad \forall z \in \mathbb{R}^n\}.$$

Субдифференциал обладает свойством монотонности. Пусть выпуклая функция $f: \mathbb{R}^n \rightarrow \mathbb{R}$ имеет субградиенты $p_i \in \partial f(x_i)$, $i = 1, 2$. Тогда $f(x_2) \geq f(x_1) + (p_1, x_2 - x_1)$ и $f(x_1) \geq f(x_2) + (p_2, x_1 - x_2)$. Складывая два последних неравенства и сокращая $f(x_1) + f(x_2)$, получаем

$$(p_1 - p_2, x_1 - x_2) \geq 0. \tag{Г.4}$$

Субдифференциал выпуклой функции на границе области определения может быть пуст, например для функции $f(x) = -\sqrt{1-x^2}$ в точках $x = \pm 1$. Однако во внутренности области определения выпуклой функции субдифференциал всегда не пуст.

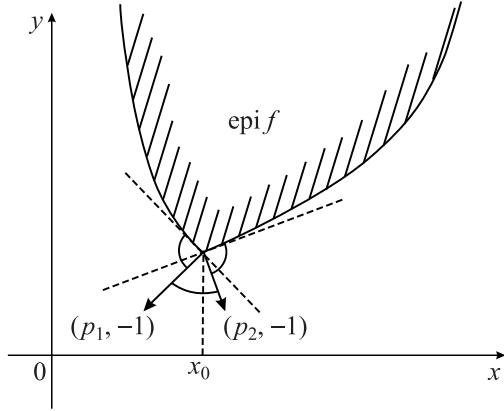


Рис. Г.4. Пример субдифференциала, $[p_1, p_2] \subset \partial f(x_0)$.

Следующую теорему можно найти в [19, следствие 1.16.1].

Теорема Г.13. Пусть $r > 0$ и $f: \text{int } \mathcal{B}_r(x_0) \rightarrow \mathbb{R}$ — выпуклая функция. Тогда $\partial f(x_0)$ — выпуклое компактное множество. При этом его опорная функция $s(p, \partial f(x_0))$ для любого единичного вектора $p \in \mathbb{R}^n$ есть

$$s(p, \partial f(x_0)) = f'(x_0, p),$$

где $f'(x_0, p) = \lim_{\alpha \rightarrow +0} \frac{f(x_0 + \alpha p) - f(x_0)}{\alpha}$ — односторонняя производная по направлению p функции f в точке x_0 .

Можно показать, что при дифференцируемости выпуклой функции f в точке x_0 выполнено равенство $\partial f(x_0) = \{f'(x_0)\}$.

Заметим, что равенство $s(p, \partial f(x_0)) = f'(x_0, p)$ справедливо для любой точки x_0 . Отсюда вытекает, что функция $p \rightarrow f'(x_0, p)$ положительно однородная, выпуклая и полуунпрерывная снизу. Если x_0 — внутренняя точка области определения f , то $p \rightarrow f'(x_0, p)$ липшицева.

Из определения субдифференциала вытекает, что необходимым и достаточным условием глобального минимума выпуклой функции f в точке x_0 является включение

$$0 \in \partial f(x_0). \quad (\Gamma.5)$$

Односторонняя производная по направлению и субдифференциал не являются непрерывными по переменной x . Например, для функции $f(x) = |x|$ имеем $\partial f(x) = \{-1\}$ при $x < 0$, $\partial f(0) = [-1, 1]$, $\partial f(x) = \{1\}$ при $x > 0$. Тем не менее выполняется некоторое условие, которое называется полуунпрерывностью сверху.

Следующий результат вытекает из теоремы 24.5 и следствия 24.5.1 из [24].

Теорема Г.14. *Пусть $r > 0$ и $f: \text{int } \mathcal{B}_r(x_0) \rightarrow \mathbb{R}$ — выпуклая функция. Тогда для любой последовательности $\{x_k\} \rightarrow x_0$ выполнены следующие условия полуунпрерывности сверху:*

$$\forall \varepsilon > 0 \quad \forall \|p\| = 1 \quad \exists k_{p,\varepsilon} \in \mathbb{N} \quad \forall k > k_{p,\varepsilon} \quad f'(x_k, p) \leq f'(x_0, p) + \varepsilon; \quad (\Gamma.6)$$

$$\forall \varepsilon > 0 \quad \exists k_\varepsilon \in \mathbb{N} \quad \forall k > k_\varepsilon \quad \partial f(x_k) \subset \partial f(x_0) + \mathcal{B}_\varepsilon(0). \quad (\Gamma.7)$$

Легко видеть, что из (Г.7) вытекает (Г.6): надо взять опорные функции от обеих частей включения (Г.7) для вектора p . Менее тривиальным оказывается тот факт, что в \mathbb{R}^n эти условия эквивалентны.

Теорема Г.15 (Моро – Рокафеллар [24, теорема 23.8]). *Пусть заданы собственные выпуклые функции $f_1, f_2: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$. Тогда для любой точки $x \in \text{dom } f_1 \cap \text{dom } f_2$*

$$\partial(f_1 + f_2)(x) \supset \partial f_1(x) + \partial f_2(x). \quad (\Gamma.8)$$

Если $\text{int } \text{dom } f_1 \cap \text{dom } f_2 \neq \emptyset$, то

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x). \quad (\Gamma.9)$$

Неравенство (Г.8) вытекает из определения субдифференциала. Нетривиальная часть теоремы — равенство (Г.9). Для его справедливости нужно условие $\text{int } \text{dom } f_1 \cap \text{dom } f_2 \neq \emptyset$.

Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое множество. Рассмотрим задачу

$$\min_{x \in \mathcal{Q}} f(x) \quad (\Gamma.10)$$

для выпуклой функции $f: \mathcal{U} \rightarrow \mathbb{R}$, где \mathcal{U} — некоторая выпуклая окрестность \mathcal{Q} . На языке индикаторной функции задачу (Г.10) можно переписать как задачу безусловной минимизации

$$\min_{x \in \mathbb{R}^n} (f(x) + \delta(x, \mathcal{Q})). \quad (\Gamma.11)$$

Пусть $x \in \mathcal{Q}$. Субдифференциал $\partial\delta(x, \mathcal{Q})$ есть множество таких векторов $p \in \mathbb{R}^n$, что для любого $z \in \mathbb{R}^n$

$$\delta(z, \mathcal{Q}) \geq \delta(x, \mathcal{Q}) + (p, z - x) = (p, z - x).$$

Поскольку при $z \in \mathcal{Q}$ выполнено равенство $\delta(z, \mathcal{Q}) = 0$, то

$$\partial\delta(x, \mathcal{Q}) = \{p \in \mathbb{R}^n : (p, x - z) \geq 0 \quad \forall z \in \mathcal{Q}\}. \quad (\Gamma.12)$$

Последнее множество есть выпуклый конус, его называют *нормальным конусом* ко множеству \mathcal{Q} в точке $x \in \mathcal{Q}$ и обозначают через $N(\mathcal{Q}, x)$. Если $x \in \text{int } \mathcal{Q}$, то $N(\mathcal{Q}, x) = \{0\}$. Если $x \in \partial\mathcal{Q}$, то в силу теоремы об отделимости [Г.6](#) $N(\mathcal{Q}, x)$ имеет размерность не менее 1.

Из необходимого условия безусловного экстремума [\(Г.5\)](#) для задачи [\(Г.11\)](#) получаем, что любое решение x_0 задачи [\(Г.10\)](#) удовлетворяет условию

$$0 \in \partial(f(x_0) + N(\mathcal{Q}, x_0)).$$

Если $\text{int dom } f \cap \mathcal{Q} \neq \emptyset$, то по теореме Моро–Рокафеллара критерий оптимальности в задаче [\(Г.10\)](#) есть

$$0 \in \partial f(x_0) + N(\mathcal{Q}, x_0),$$

или в эквивалентном виде

$$(-\partial f(x_0)) \cap N(\mathcal{Q}, x_0) \neq \emptyset.$$

Если функция f дифференцируема, то $-f''(x_0) \in N(\mathcal{Q}, x_0)$.

Рассмотрим в \mathbb{R}^2 множества

$$\mathcal{Q}_1 = \mathcal{B}_1((0; 1)), \quad \mathcal{Q}_2 = -\mathcal{Q}_1;$$

очевидно, что $\mathcal{Q}_1 \cap \mathcal{Q}_2 = \{(0; 0)\}$. Имеем

$$\partial(\delta((0; 0), \mathcal{Q}_1) + \delta((0; 0), \mathcal{Q}_2)) = \partial\delta((0; 0), \{(0; 0)\}) = \mathbb{R}^2,$$

$$\partial\delta((0; 0), \mathcal{Q}_1) = N(\mathcal{Q}_1, (0; 0)) = \{\lambda e_2 : \lambda \leq 0\},$$

$$\partial\delta((0; 0), \mathcal{Q}_2) = N(\mathcal{Q}_2, (0; 0)) = \{\lambda e_2 : \lambda \geq 0\}$$

и $\partial\delta((0; 0), \mathcal{Q}_1) + \partial\delta((0; 0), \mathcal{Q}_2) = \{\lambda e_2 : \lambda \in \mathbb{R}\}$. Таким образом, равенство [\(Г.9\)](#) в теореме Моро–Рокафеллара может не выполняться.

Приведем теорему об очистке [[19](#), теоремы 1.17.3, 1.17.4].

Теорема Г.16 (о субдифференциале максимума). Предположим, что $\mathcal{S} \subset \mathbb{R}^m$ — компактное подмножество, и пусть функция $f: \mathcal{S} \times \mathbb{R}^n \rightarrow \mathbb{R}$ такова, что отображение $f(s, \cdot): \mathbb{R}^n \rightarrow \mathbb{R}$ выпукло при всяком $s \in \mathcal{S}$, а отображение $f(\cdot, x): \mathcal{S} \rightarrow \mathbb{R}$ полуинкрементно сверху при всяком $x \in \mathbb{R}^n$. Определим $f(x) = \max_{s \in \mathcal{S}} f(s, x)$ и $\mathcal{S}_0(x) = \{s \in \mathcal{S}: f(s, x) = f(x)\}$.

Если для каждого $s \in \mathcal{S}$ в точке $x_0 \in \text{dom } f$ функция $f(s, \cdot)$ непрерывна, то справедливо равенство

$$\partial f(x_0) = \text{co} \left(\bigcup_{s \in \mathcal{S}_0(x_0)} \partial_x f(s, x_0) \right), \quad (\Gamma.13)$$

где ∂_x — субдифференциал по переменной x .

Теорема Г.17 (об очистке). При выполнении условий теоремы Г.16 каждый элемент $p \in \partial f(x_0)$ можно представить в виде

$$p = \sum_{i=1}^k \lambda_i p_i, \quad (\Gamma.14)$$

где $k \leq n+1$, $\lambda_i > 0$, $\sum_{i=1}^k \lambda_i = 1$, $p_i \in \partial_x f(s_i, x_0)$, $s_i \in \mathcal{S}_0(x_0)$, $1 \leq i \leq k$.

Отметим, что формула (Г.14) вытекает из (Г.13) и теоремы Кареодори. Одно из приложений теоремы Г.17 можно найти в примере 2.8.4.

Г.6. Сильно выпуклые функции

Определение Г.10. Функцию $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ будем называть сильно выпуклой с константой $\varkappa \geq 0$, если функция $f(x) - \frac{\varkappa}{2} \|x\|^2$ выпуклая. При $\varkappa = 0$ функция f является выпуклой.

Первое серьезное внимание классу сильно выпуклых функций в задачах оптимизации было уделено в работах Б.Т. Поляка, см. [57].

Легко видеть, что определение Г.10 эквивалентно следующему условию: для любого вектора $x_0 \in \mathbb{R}^n$ функция $f(x) - \frac{\varkappa}{2} \|x - x_0\|^2$ выпуклая. Действительно,

$$f(x) - \frac{\varkappa}{2} \|x - x_0\|^2 = \left(f(x) - \frac{\varkappa}{2} \|x\|^2 \right) + \varkappa(x, x_0) - \frac{\varkappa}{2} \|x_0\|^2$$

и функция в скобках выпуклая. Добавка $\kappa(x, x_0) - \frac{\kappa}{2}\|x_0\|^2$, линейная по x , не нарушает выпуклость функции.

Таким образом, чтобы получить сильно выпуклую с константой $\kappa > 0$ функцию, надо к любой выпуклой функции $g: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ прибавить $\frac{\kappa}{2}\|x\|^2$.

Пусть функция $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ сильно выпуклая с константой $\kappa > 0$. Для выпуклой функции $g(x) = f(x) - \frac{\kappa}{2}\|x\|^2$, точек $x, x_0 \in \mathbb{R}^n$ и $\lambda \in [0, 1]$ запишем неравенство выпуклости:

$$g((1-\lambda)x_0 + \lambda x) \leq (1-\lambda)g(x_0) + \lambda g(x).$$

Если переписать последнее неравенство в терминах функции f , то получим следующее эквивалентное условие сильной выпуклости: для любого $\lambda \in [0, 1]$ и любых $x, x_0 \in \mathbb{R}^n$ выполнено неравенство

$$f((1-\lambda)x_0 + \lambda x) \leq (1-\lambda)f(x_0) + \lambda f(x) - \lambda(1-\lambda)\frac{\kappa}{2}\|x_0 - x\|^2. \quad (\Gamma.15)$$

Из формулы (Г.15), верной для любых $\lambda \in [0, 1]$, $x, x_0 \in \mathbb{R}^n$, вытекает определение Г.10.

Перепишем неравенство (Г.15) в виде

$$\frac{1}{\lambda} \left(f(x_0 + \lambda(x - x_0)) - f(x_0) \right) \leq f(x) - f(x_0) - (1-\lambda)\frac{\kappa}{2}\|x_0 - x\|^2$$

и перейдем к пределу при $\lambda \rightarrow +0$. В силу дифференцируемости по направлению выпуклых функций получим

$$s(x - x_0, \partial f(x_0)) = f'(x_0, x - x_0) \leq f(x) - f(x_0) - \frac{\kappa}{2}\|x_0 - x\|^2.$$

Если $p_0 \in \partial f(x_0) \neq \emptyset$, то последнее условие можно переписать в виде

$$f(x) \geq f(x_0) + (p_0, x - x_0) + \frac{\kappa}{2}\|x_0 - x\|^2. \quad (\Gamma.16)$$

Формула (Г.16) уточняет субградиентное неравенство для сильно выпуклых функций. Из нее также вытекает, что надграфик сильно выпуклой функции можно представить в виде пересечения надграфиков параболоидов вращения вида $\varphi_{x_0}(x) = f(x_0) + (p_0, x - x_0) + \frac{\kappa}{2}\|x_0 - x\|^2$ по всем точкам x_0 эффективного множества f , в которых $\partial f(x_0) \neq \emptyset$, и по всем $p_0 \in \partial f(x_0)$. Отсюда в частности следует сверхлинейный рост сильно выпуклой функции $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$:

$$\lim_{\|x\| \rightarrow +\infty} \frac{f(x)}{\|x\|} = +\infty.$$

Из формулы (Г.16) вытекает несколько полезных следствий. Поменяв в (Г.16) местами x и x_0 и сложив полученное неравенство с (Г.16), мы получим

$$\varkappa \|x_0 - x\|^2 \leq (p_0 - p, x_0 - x) \quad \forall p_0 \in \partial f(x_0) \quad \forall p \in \partial f(x). \quad (\Gamma.17)$$

Применяя к правой части (Г.17) неравенство Коши–Буняковского, получаем

$$\varkappa \|x_0 - x\| \leq \|p_0 - p\| \quad \forall p_0 \in \partial f(x_0) \quad \forall p \in \partial f(x), \quad (\Gamma.18)$$

см. рис. Г.5.

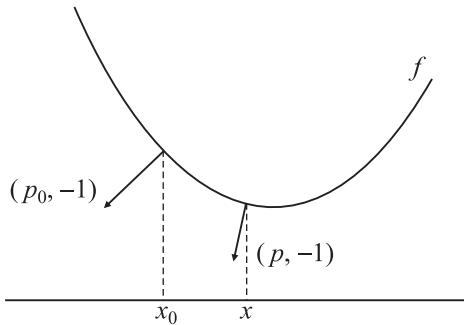


Рис. Г.5. Близость p и p_0 влечет близость x и x_0 .

Отметим, что каждое из условий (Г.16), (Г.17) и (Г.18) для выпуклой функции f эквивалентно определению Г.10, см. [19, §1.19]

Получим еще одно полезное неравенство. Пусть $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ — сильно выпуклая с константой $\varkappa > 0$ собственная функция, сублифференцируемая в точке $x \in \mathbb{R}^n$. Зафиксируем $p \in \partial f(x)$. Пусть $x_0 \in \mathbb{R}^n$ — глобальный минимум f на \mathbb{R}^n .

Для сильно выпуклой (определение Г.10) собственной функции ее минимум на \mathbb{R}^n достигается в единственной точке $x_0 \in \mathbb{R}^n$. Это следует из строгой выпуклости и сверхлинейного роста сильно выпуклой функции. Из субградиентного неравенства для выпуклой функции f получаем $f(x_0) \geq f(x) + (p, x_0 - x)$, то есть

$$\|p\| \cdot \|x - x_0\| \geq (p, x - x_0) \geq f(x) - f(x_0). \quad (\Gamma.19)$$

Из неравенства (Г.18) для точки минимума x_0 , где $p_0 = 0 \in \partial f(x_0)$, следует

$$\|x_0 - x\| \leq \frac{1}{\varkappa} \|p\|. \quad (\Gamma.20)$$

Из условий (Г.19) и (Г.20) находим, что для любого $x \in \text{dom } f$ выполнено неравенство

$$\|p\|^2 \geq \varkappa(f(x) - f(x_0)) \quad \forall p \in \partial f(x), \quad f(x_0) = \min_{z \in \mathbb{R}^n} f(z). \quad (\Gamma.21)$$

Оно играет важную роль в методах градиентной минимизации.

Г.7. Сильно выпуклые множества и проксимально гладкие множества. Слабо выпуклые и слабо вогнутые функции

Определение Г.11. Выпуклый компакт $\mathcal{Q} \subset \mathbb{R}^n$ называется *сильно выпуклым* с радиусом $R > 0$, если его можно представить как пересечение замкнутых евклидовых шаров радиуса $R > 0$, то есть

$$\mathcal{Q} = \bigcap_{x \in \mathcal{X}} \mathcal{B}_R(x)$$

для некоторого множества центров $\mathcal{X} \subset \mathbb{R}^n$.

Сильно выпуклые множества и их обобщения изучались в работах [19, 59, 89, 133].

Сильная выпуклость выпуклого компакта \mathcal{Q} в \mathbb{R}^n с радиусом $R > 0$ эквивалентна следующим свойствам:

- 1) *опорный принцип*: для любого $\|p\| = 1$ выполнено включение $\mathcal{Q} \subset \mathcal{B}_R(\mathcal{Q}(p) - Rp)$, см. рис. Г.6;
- 2) для любых единичных векторов $\|p\| = \|q\| = 1$ выполнено неравенство $\|\mathcal{Q}(p) - \mathcal{Q}(q)\| \leq R\|p - q\|$;
- 3) для любых точек $x_1, x_2 \in \mathbb{R}^n$, $\varrho(x_i, \mathcal{Q}) \geq r > 0$ и проекций $a_i = P_{\mathcal{Q}}x_i$, $i = 1, 2$, выполнена оценка

$$\|a_1 - a_2\| \leq \frac{R}{R+r} \|x_1 - x_2\|;$$

- 4) множество \mathcal{Q} равномерно выпукло с модулем $\delta_{\mathcal{Q}}(t) = \frac{t^2}{8R}$. Последнее означает (см. подробнее [75]), что для любых точек $x_1, x_2 \in \mathcal{Q}$ выполнено включение

$$\frac{x_1 + x_2}{2} + \mathcal{B}_{\delta_{\mathcal{Q}}(\|x_1 - x_2\|)}(0) \subset \mathcal{Q}.$$

Сумма Минковского сильно выпуклых множеств \mathcal{Q}_1 с радиусом $R_1 > 0$ и \mathcal{Q}_2 с радиусом $R_2 > 0$ сильно выпукла с радиусом $R_1 + R_2$. Эллипсоид является сильно выпуклым множеством с радиусом $R = \frac{\lambda_{\max}^2}{\lambda_{\min}}$,

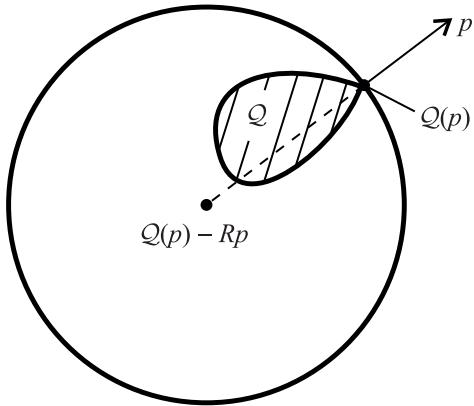


Рис. Г.6. Опорный принцип для сильно выпуклых множеств.

где λ_{\max} — максимальная полуось эллипсоида, а λ_{\min} — минимальная [19, глава 3].

Определение Г.12. Замкнутое множество $Q \subset \mathbb{R}^n$ будем называть *проксимально гладким* (или *прокс-регулярным, слабо выпуклым*) с константой $R > 0$, если функция расстояния $\varrho(x, Q)$ непрерывно дифференцируема на $U_Q(R) = \{x \in \mathbb{R}^n : 0 < \varrho(x, Q) < R\}$ — R -окрестности множества Q .

Заметим, что выпуклое замкнутое подмножество $Q \subset \mathbb{R}^n$ можно считать проксимально гладким с любой константой $R > 0$.

Напомним, что градиент Фреше $\varrho(x, Q)$ в точке $x \notin Q$ есть

$$\varrho'(x, Q) = \frac{x - P_Q x}{\varrho(x, Q)}.$$

Проксимально гладкие множества в гильбертовом пространстве изучались в работах [82, 133]. О проксимальных градиентах, построенных на основании свойств проксимально гладких множеств, и их применении в теории оптимального управления можно найти в книге [38].

Пусть $Q \subset \mathbb{R}^n$ — замкнутое подмножество и $x \in \partial Q$. Будем говорить, что вектор $p \in \mathbb{R}^n$ есть *проксимальная нормаль* ко множеству Q в точке x , если существует такое $t_x > 0$, что для всех $t \in (0, t_x)$ выполнено равенство $P_Q(x + tp) = \{x\}$. Легко видеть, что множество всех проксимальных нормалей ко множеству Q в точке $x \in \partial Q$ образует *конус проксимальных нормалей*, который будем обозначать через $N_P(Q, x)$.

Проксимальная гладкость замкнутого множества \mathcal{Q} в \mathbb{R}^n с константой $R > 0$ эквивалентна следующим свойствам:

- 1) *опорный принцип*: для любых $x \in \partial\mathcal{Q}$ и $p \in N_P(\mathcal{Q}, x)$, $\|p\| = 1$, выполнено условие $\mathcal{Q} \cap \text{int } \mathcal{B}_R(x + Rp) = \emptyset$, см. рис. Г.7;

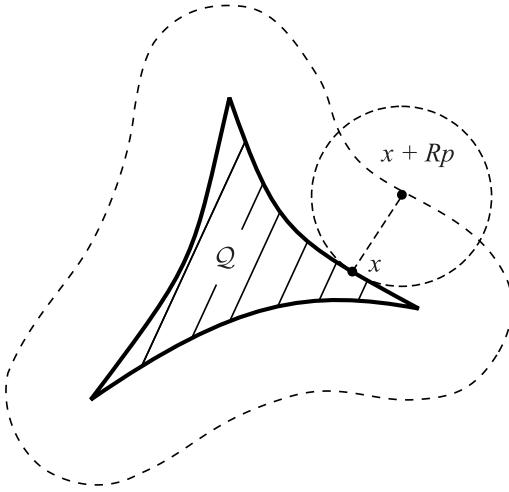


Рис. Г.7. Опорный принцип для проксимально гладких множеств.

- 2) для любой точки $x \in \mathcal{U}_{\mathcal{Q}}(R)$ метрическая проекция $P_{\mathcal{Q}}x$ одноточечна и непрерывно зависит от x ;
- 3) для $r \in (0, R)$, любых точек $x_1, x_2 \in \mathcal{U}_{\mathcal{Q}}(r)$ и $a_i = P_{\mathcal{Q}}x_i$, $i = 1, 2$, выполнено неравенство $\|a_1 - a_2\| \leq \frac{R}{R-r} \|x_1 - x_2\|$.

Отметим, что любое выпуклое замкнутое множество $\mathcal{Q} \subset \mathbb{R}^n$ является проксимально гладким с любой константой $R > 0$, также можно считать такое множество проксимально гладким с $R = +\infty$.

Приведем примеры проксимально гладких множеств.

Пусть $g: \mathbb{R}^n \rightarrow \mathbb{R}$ — C^1 -функция и $\mathcal{S} = \{x \in \mathbb{R}^n : g(x) = 0\}$. Предположим, что единичная ориентированная нормаль к \mathcal{S} липшицево зависит от точки: существует константа $L > 0$ такая, что для любых $x_1, x_2 \in \mathcal{S}$ выполнено неравенство

$$\left\| \frac{g'(x_1)}{\|g'(x_1)\|} - \frac{g'(x_2)}{\|g'(x_2)\|} \right\| \leq L \|x_1 - x_2\|.$$

Тогда множество \mathcal{S} проксимально гладкое с константой $R = 1/L$ [7, теорема 1.10.2]. Проксимальная гладкость пересечения множеств указанного выше вида рассматривается в работе [70].

Многие матричные многообразия проксимально гладкие [49]. Расстояние в *примерах* ниже измеряется в смысле нормы Фробениуса $\|X\|_F = \sqrt{\text{tr } X^T X}$.

Так, многообразие Штифеля $\mathcal{S}_{n,k}$ для любых $k \leq n$ является проксимально гладким множеством с неулучшаемой для любых k, n константой $R = 1$, причем метрическая проекция на $\mathcal{S}_{n,k}$ вычисляется по формуле (Б.17). Многообразие Грассмана $\mathcal{G}_{n,k}$ для любых $k \leq n$, реализованное как вложение в пространство симметричных матриц \mathbb{S}^n , также проксимально гладкое с неулучшаемой для любых k, n константой $R = 1/\sqrt{2}$. Метрическая проекция на $\mathcal{G}_{n,k}$ вычисляется по формуле (Б.18).

Константы проксимальной гладкости и формулы для метрической проекции для некоторых других матричных множеств можно найти в [51].

Определение Г.13. Пусть $f: \mathbb{R}^n \supset \mathcal{U} \rightarrow \mathbb{R}$, где \mathcal{U} — выпуклое открытое подмножество. Функция f называется *слабо выпуклой* с константой $L_1 \geq 0$, если функция $f(x) + \frac{L_1 \|x\|^2}{2}$ выпуклая.

Функция f называется *слабо вогнутой* с константой $L_1 \geq 0$, если функция $h(x) = f(x) - \frac{L_1 \|x\|^2}{2}$ вогнутая, то есть $-h(x)$ выпуклая.

Концепция слабой вогнутости и ее применение в задачах оптимизации возникла, вероятно, в работах Е.А. Нурминского [120]. Современное состояние теории слабо выпуклых и слабо вогнутых функций может быть найдено в монографии [7].

Аналогично формуле (Г.15) доказываются следующие утверждения.

Если функция f из определения Г.13 слабо выпуклая с константой $L_1 > 0$, то для любых $x_0, x \in \mathcal{U}, \lambda \in [0, 1]$, это эквивалентно выполнению неравенства

$$\lambda f(x) + (1 - \lambda)f(x_0) \geq f((1 - \lambda)x_0 + \lambda x) - \frac{\lambda(1 - \lambda)}{2} L_1 \|x_0 - x\|^2.$$

Если функция f из определения Г.13 слабо вогнутая с константой $L_1 > 0$, то для любых $x_0, x \in \mathcal{U}, \lambda \in [0, 1]$ это эквивалентно неравенству

$$f((1 - \lambda)x_0 + \lambda x) \geq \lambda f(x) + (1 - \lambda)f(x_0) - \frac{\lambda(1 - \lambda)}{2} L_1 \|x_0 - x\|^2.$$

Известно, что слабая выпуклость и слабая вогнутость функции f с константой L_1 эквивалентны липшицевости градиента f' с константой Липшица L_1 [7, теорема 2.1.2]. Однако можно рассматривать слабую

выпуклость или вогнутость отдельно. Напомним, что *субдифференциалом Фреше* функции f в точке $x_0 \in \mathcal{U}$ называется множество

$$\partial_F^- f(x_0) = \left\{ p \in \mathbb{R}^n : \liminf_{x \rightarrow x_0} \frac{f(x) - f(x_0) - (p, x - x_0)}{\|x - x_0\|} \geq 0 \right\},$$

а *супердифференциалом Фреше* функции f в точке $x_0 \in \mathcal{U}$ есть

$$\partial_F^+ f(x_0) = \left\{ p \in \mathbb{R}^n : \limsup_{x \rightarrow x_0} \frac{f(x) - f(x_0) - (p, x - x_0)}{\|x - x_0\|} \leq 0 \right\}.$$

Пусть функция f слабо вогнута с константой L_1 . Тогда функция

$$h(x) = f(x) - \frac{L_1}{2} \|x\|^2$$

вогнута и из [41, утверждение 1.107(i)] вытекает равенство

$$\partial_F^+ f(x_0) = \partial^+ h(x_0) + L_1 x_0,$$

где $\partial^+ h(x_0)$ — супердифференциал вогнутой функции h в точке x_0 .

Обозначим через $f'(x_0) \in \partial_F^+ f(x_0)$ произвольный суперградиент Фреше. Тогда

$$\begin{aligned} h(x) &\leq h(x_0) + (f'(x_0) - L_1 x_0, x - x_0) = \\ &= h(x_0) + (f'(x_0), x - x_0) - L_1(x_0, x - x_0). \end{aligned}$$

Переходя к функции f , получаем

$$f(x) - \frac{L_1}{2} \|x\|^2 \leq f(x_0) - \frac{L_1}{2} \|x_0\|^2 + (f'(x_0), x - x_0) - L_1(x_0, x - x_0).$$

После преобразований для любого суперградиента $f'(x_0) \in \partial_F^+ f(x_0)$

$$f(x) \leq f(x_0) + (f'(x_0), x - x_0) + \frac{L_1}{2} \|x - x_0\|^2. \quad (\Gamma.22)$$

Из приведенных выкладок вытекает следующая теорема.

Теорема Г.18. *Слабая вогнутость функции f в определении Г.13 эквивалентна выполнению неравенства (Г.22) для любых $x, x_0 \in \mathcal{U}$ и любого суперградиента Фреше $f'(x_0) \in \partial_F^+ f(x_0)$. При этом имеет место равенство*

$$\partial_F^+ f(x_0) = \partial^+ h(x_0) + L_1 x_0,$$

где $\partial^+ h(x_0)$ — супердифференциал вогнутой функции

$$h(x) = f(x) - \frac{L_1}{2} \|x\|^2$$

в точке x_0 .

Формула (Г.22) показывает, что для слабо вогнутой (не липшицевой!) с константой $L_1 > 0$ функции f выполнена верхняя квадратичная аппроксимация, где $f'(x_0) \in \partial_F^+ f(x_0)$ — произвольный элемент. Этого бывает достаточно для доказательства сходимости некоторых оптимизационных алгоритмов. В настоящей монографии обычно требуется условие Липшица градиента для выполнения верхней квадратичной аппроксимации, хотя, в силу результата теоремы Г.18, в некоторых случаях можно работать с более общим классом слабо вогнутых функций. Пример использования слабой вогнутости приведен в лемме 2.3.1.

Более подробно об обобщенных градиентах можно прочитать в монографиях [41, 43].

Г.8. Метрика Хаусдорфа. Погрешности многогранных аппроксимаций

Определение Г.14. Пусть \mathcal{P} и \mathcal{Q} — компактные подмножества \mathbb{R}^n . Расстоянием в метрике Хаусдорфа между \mathcal{P} и \mathcal{Q} называется величина

$$h(\mathcal{P}, \mathcal{Q}) = \max \left\{ \max_{x \in \mathcal{Q}} \varrho(x, \mathcal{P}), \max_{x \in \mathcal{P}} \varrho(x, \mathcal{Q}) \right\}$$

или, эквивалентно,

$$h(\mathcal{P}, \mathcal{Q}) = \inf \left\{ \varepsilon > 0 : \quad \mathcal{P} \subset \mathcal{Q} + \mathcal{B}_\varepsilon(0), \quad \mathcal{Q} \subset \mathcal{P} + \mathcal{B}_\varepsilon(0) \right\}. \quad (\text{Г.23})$$

Для компактных множеств \mathcal{P} , \mathcal{Q} инфимум в (Г.23) достигается.

В случае выпуклости компактных множеств \mathcal{P} и \mathcal{Q} расстояние $h(\mathcal{P}, \mathcal{Q})$ связано с опорными функциями \mathcal{P} и \mathcal{Q} по формуле [19, лемма 1.11.4]

$$h(\mathcal{P}, \mathcal{Q}) = \max_{\|p\|=1} |s(p, \mathcal{P}) - s(p, \mathcal{Q})|.$$

В настоящем разделе обсудим оценки погрешности выпуклого компакта многогранником в метрике Хаусдорфа. Будут рассмотрены внешние и внутренние аппроксимации компакта многогранником.

Далее через $\widehat{\mathcal{Q}}$ будем обозначать внешнюю аппроксимацию многогранником множества \mathcal{Q} , то есть $\mathcal{Q} \subset \widehat{\mathcal{Q}}$, а через $\widehat{\mathcal{Q}}$ — внутреннюю, то

есть $\tilde{\mathcal{Q}} \subset \mathcal{Q}$. Обычно выпуклый компакт $\mathcal{Q} \subset \mathbb{R}^n$ приближают внешней многогранной аппроксимацией. Для этого берут сетку $\mathbb{G} = \{p_i\}_{i=1}^I$ единичных векторов мелкости $\Delta \in (0, \frac{1}{2})$, под мелкостью Δ понимается такое число, что каждый вектор p можно представить (возможно, не единственным образом) в виде $p = \sum_{i \in I_p} \alpha_i p_i$, $\alpha_i > 0$, $p_i \in \mathbb{G}$, I_p — некоторое подмножество $\{1, \dots, I\}$ и $\|p_i - p_j\| \leq \Delta$ для всех $i, j \in I_p$. В качестве внешней аппроксимации обычно рассматривается множество

$$\hat{\mathcal{Q}} = \{x \in \mathbb{R}^n : (p_i, x) \leq s(p_i, \mathcal{Q}) \quad \forall p_i \in \mathbb{G}\}. \quad (\Gamma.24)$$

В случае если выпуклый компакт произвольный, погрешность аппроксимации есть

$$h(\mathcal{Q}, \hat{\mathcal{Q}}) \leq \Delta \frac{\text{diam } \mathcal{Q}}{1 - \frac{\Delta^2}{2}}.$$

Если множество \mathcal{Q} сильно выпукло с радиусом $R > 0$, то [19, §4.9]

$$h(\mathcal{Q}, \hat{\mathcal{Q}}) \leq \Delta^2 \frac{R}{1 - \frac{\Delta^2}{2}}.$$

Внутренняя аппроксимация может быть построена на основе внешней (Г.24). Предположим, что $\mathcal{B}_r(0) \subset \mathcal{Q}$ для некоторого $r > 0$. Тогда для $h = h(\mathcal{Q}, \tilde{\mathcal{Q}})$ имеем

$$\tilde{\mathcal{Q}} \subset \mathcal{Q} + \frac{h}{r} \mathcal{B}_r(0) \subset \left(1 + \frac{h}{r}\right) \mathcal{Q},$$

то есть внутренняя аппроксимация $\tilde{\mathcal{Q}}$ есть

$$\tilde{\mathcal{Q}} = \frac{r}{r+h} \hat{\mathcal{Q}} \subset \mathcal{Q}. \quad (\Gamma.25)$$

При этом имеем цепочку включений

$$\mathcal{Q} \subset \frac{r}{r+h} \hat{\mathcal{Q}} + \frac{h}{r+h} \hat{\mathcal{Q}} \subset \tilde{\mathcal{Q}} + \frac{h}{r+h} (\mathcal{Q} + h \mathcal{B}_1(0)),$$

и погрешность $h(\tilde{\mathcal{Q}}, \mathcal{Q})$ оценивается как

$$h(\tilde{\mathcal{Q}}, \mathcal{Q}) \leq \frac{h}{r+h} (\text{diam } \mathcal{Q} + h), \quad h = h(\mathcal{Q}, \hat{\mathcal{Q}}).$$

Итак, для внешней (Г.24) и внутренней (Г.25) аппроксимаций выпуклого компакта \mathcal{Q} получены оценки погрешности приближения в метрике Хаусдорфа.

Г.9. Выживание решений*

В разделе 3.1.1 было введено понятие инвариантного эллипсоида, то есть такого эллипсоида, из которого не выходит никакое решение устойчивой линейной системы с начальным условием также из этого эллипсоида. Оказывается, что аналогичные инвариантному эллипсоиду конструкции можно рассмотреть и в гораздо более общей ситуации.

Пусть $\mathcal{M} \subset \mathbb{R}^n$ — замкнутое подмножество, а решение системы

$$\dot{x} = f(t, x), \quad x(t_0) = x_0 \in \mathcal{M} \subset \mathbb{R}^n, \quad t_0 \geq 0, \quad (\Gamma.26)$$

существует на луче $t \in [t_0, +\infty)$ при любом начальном условии $x_0 \in \mathcal{M}$. Будем предполагать измеримость по Лебегу функции $f(\cdot, x)$ при каждом x и непрерывность $f(t, \cdot)$ при каждом t .

Рассмотрим вопрос: при каких условиях решение задачи Коши для (Г.26) «выживает» в \mathcal{M} , то есть $x(t) \in \mathcal{M}$ для всех $t \in [t_0, +\infty)$? Выживание решений — важный раздел негладкого и многозначного анализа. Классические монографии [33] и [34] посвящены этому вопросу.

Напомним определение *верхнего касательного конуса* (или *конуса Булигана*) для замкнутого множества $\mathcal{M} \subset \mathbb{R}^n$ в точке $x \in \mathcal{M}$:

$$T(\mathcal{M}, x) = \left\{ v \in \mathbb{R}^n : \forall \varepsilon > 0 \exists \{t_k\}, t_k \rightarrow +0, \mathcal{B}_\varepsilon(v) \bigcap \frac{\mathcal{M} - x}{t_k} \neq \emptyset \right\}.$$

Эквивалентные формы конуса Булигана:

$$T(\mathcal{M}, x) = \left\{ v \in \mathbb{R}^n : \exists \{t_k\}, t_k \rightarrow +0, x + t_k v + o(t_k) \in \mathcal{M} \right\},$$

$$T(\mathcal{M}, x) = \left\{ v \in \mathbb{R}^n : \liminf_{t \rightarrow +0} \varrho \left(v, \frac{\mathcal{M} - x}{t} \right) = 0 \right\},$$

или

$$T(\mathcal{M}, x) = \bigcap_{\varepsilon > 0} \bigcap_{\delta > 0} \bigcup_{0 < t < \delta} \left(\frac{\mathcal{M} - x}{t} + \mathcal{B}_\varepsilon(0) \right).$$

Вектор $v \in \mathbb{R}^n$ содержится в конусе $T(\mathcal{M}, x)$ тогда и только тогда, когда найдутся последовательность $\{t_k\}$ положительных чисел, стремящаяся к нулю, и последовательность векторов $\{v_k\}$, сходящаяся к v , такие, что $x + t_k v_k \in \mathcal{M}$. Если множество \mathcal{M} выпукло и замкнуто, то для каждой точки $x \in \mathcal{M}$ верно равенство

$$T(\mathcal{M}, x) = \text{cl} \bigcup_{t > 0} \frac{\mathcal{M} - x}{t}.$$

Все указанные свойства конуса $T(\mathcal{M}, x)$ можно найти в [19, §1.4] и в [34].

Будем говорить, что для системы (Г.26) выполнено *условие выживания*, если для почти всех $t \geq 0$ выполнено включение

$$f(t, x) \in T(\mathcal{M}, x) \quad \forall x \in \mathcal{M}. \quad (\text{Г.27})$$

Поскольку для внутренних точек $x \in \text{int } \mathcal{M}$ конус $T(\mathcal{M}, x)$ очевидно равен \mathbb{R}^n , то условие выживания (Г.27) достаточно проверять в граничных точках множества \mathcal{M} .

Пусть $L_1([0, +\infty), [0, +\infty))$ есть пространство интегрируемых по Лебегу функций, которые определены на луче $[0, +\infty)$ со значениями в $[0, +\infty)$. Приведенная ниже теорема является частным случаем теоремы П. Таллоса о выживании для однозначной функции $f(t, x)$, см. [33, теорема 11.7.1].

Теорема Г.19. *Пусть в системе (Г.26) $\mathcal{M} \subset \mathbb{R}^n$ — замкнутое подмножество и функция $f(t, x)$ удовлетворяет условиям:*

- 1) *для любого $x \in \mathcal{M}$ отображение $t \mapsto f(t, x)$ измеримо;*
- 2) *для любого $t \geq 0$ отображение $x \mapsto f(t, x)$ непрерывно;*
- 3) *существует функция $c(\cdot) \in L_1([0, +\infty), [0, +\infty))$ такая, что*

$$\|f(t, x)\| \leq c(t)(1 + \|x\|).$$

Если выполнено условие выживания (Г.27), то для любого $t_0 \geq 0$ и любого начального условия $x_0 \in \mathcal{M}$ существует решение (Г.26), которое удовлетворяет условию $x(t) \in \mathcal{M}$ для всех $t \geq 0$.

Для простоты рассмотрим далее случай автономной системы (Г.26) с $f(t, x) = f(x)$. Пусть функция $f(x)$ для некоторой константы $c > 0$ удовлетворяет для всех x условию линейного роста $\|f(x)\| \leq c(1 + \|x\|)$, которое гарантирует продолжимость любого решения (Г.26) вправо. Поскольку $c \notin L_1([0, +\infty), [0, +\infty))$, то невозможно применять теорему Г.19 к такой функции. Однако положение спасает следующее рассуждение. Зафиксируем $T > 0$. Рассмотрим функцию

$$f_T(t, x) = \begin{cases} f(x), & t \in [0, T], \\ (T + 1 - t)f(x), & t \in (T, T + 1], \\ 0, & t > T + 1. \end{cases}$$

Легко видеть, что f_T удовлетворяет условию теоремы Г.19 с

$$c(t) = \begin{cases} c, & t \in [0, T + 1], \\ 0, & t > T + 1, \end{cases}$$

и последняя функция принадлежит $L_1([0, +\infty), [0, +\infty))$. Таким образом, некоторое решение задачи Коши $\dot{x} = f_T(t, x)$, $x(t_0) = x_0 \in \mathcal{M}$ по теореме Г.19 выживает при $t \geq t_0$, а значит, существует решение задачи Коши (Г.26), которое выживает на любом отрезке $[t_0, T]$.

Слова «существует решение» в формулировке теоремы Г.19 возникли не случайно, поскольку для непрерывной по x правой части в задаче (Г.26) помимо выживающих могут быть и не выживающие решения. Рассмотрим простейший пример. Пусть $f(x) = 3x^{\frac{2}{3}}$ и $\mathcal{M} = \{0\}$. Тогда условие выживания (Г.27) выполнено, так как $T(\mathcal{M}, 0) = \{0\}$ и для точки $x = 0$ имеем $f(0) = 0$. Но у задачи Коши

$$\dot{x} = 3x^{\frac{2}{3}}, \quad x(0) = 0,$$

есть как выживающее в \mathcal{M} решение $x_0(t) = 0$, так и не выживающее решение $x_1(t) = t^3$.

Для того чтобы исключить эту неоднозначность, будем далее дополнительно предполагать, что $f(t, x)$ удовлетворяет условию Липшица по x при каждом фиксированном t . Как известно, это гарантирует единственность решения задачи Коши.

Пусть $v = v(t, x)$ — возмущение в системе

$$\dot{x} = f(x) + v, \quad x(t_0) = x_0 \in \mathcal{M}, \quad t_0 \geq 0.$$

Будем считать, что функция v измерима по t и липшицева по x . Если значения v подчинены геометрическому ограничению

$$v(t, x) \in \mathcal{V} = \bigcap_{x \in \partial \mathcal{M}} (T(\mathcal{M}, x) - f(x)) \quad \forall t \geq t_0 \quad \forall x \in \mathcal{M}, \quad (\text{Г.28})$$

то в силу теоремы Г.19 решение указанной выше возмущенной задачи Коши выживает во множестве \mathcal{M} . Таким образом, в случае $\mathcal{V} \neq \emptyset$ формула (Г.28) задает множество допустимых возмущений \mathcal{V} , к которым нечувствительно решение: эти возмущения не могут изменить включение $x(t) \in \mathcal{M}$ при $t \geq t_0$.

Рассмотрим систему

$$\dot{x} = f(x), \quad x(t_0) = x_0 \in \mathcal{M} \subset \mathbb{R}^n, \quad t_0 \geq 0. \quad (\text{Г.29})$$

Будем считать, что правая часть уравнения липшицева и удовлетворяет условию линейного роста. Рассмотрим для простоты случай выпуклого множества \mathcal{M} , содержащего нуль в своей внутренности и имеющего липшицеву нормаль: существует константа $L > 0$ такая, что

для всех $x, y \in \partial\mathcal{M}$ и единичных нормальных векторов $p(x)$ и $p(y)$ в точках x и y ко множеству \mathcal{M} соответственно выполнено условие $\|p(x) - p(y)\| \leq L\|x - y\|$.

Пусть $[\lambda]_+ = \lambda$ при $\lambda \geq 0$ и $[\lambda]_+ = 0$ при $\lambda < 0$. Рассмотрим обратную связь вида

$$u(\lambda x) = \begin{cases} -p(x) [(p(x), f(\lambda x))]_+, & \lambda \in (\frac{1}{2}, 1], \\ 0, & \lambda \in [0, \frac{1}{2}], \end{cases} \quad \forall x \in \partial\mathcal{M}. \quad (\Gamma.30)$$

Отметим, что для любого $x \in \partial\mathcal{M}$ и $\lambda \in (0, 1]$ выполнено равенство $p(x) \in N(\lambda\mathcal{M}, \lambda x)$, то есть единичная нормаль $p(\lambda x)$ ко множеству $\lambda\mathcal{M}$ в точке λx равна $p(x)$. Заметим, что функция $u(z)$ кусочно-липшицева по аргументу. По построению $u(\cdot)$ для любой точки $z = \lambda x \in \mathcal{M} \setminus (\frac{1}{2}\mathcal{M})$, $x \in \partial\mathcal{M}$, $\lambda \in (\frac{1}{2}, 1]$, выполнено условие

$$u(z) \in T(\lambda\mathcal{M}, z) - f(z),$$

поскольку величина $[(p(x), f(\lambda x))]_+$ есть в точности расстояние от точки $f(z)$ до

$$T(\lambda\mathcal{M}, z) = \{y \in \mathbb{R}^n : (p(x), y) \leq 0\}.$$

Неформально говоря, управление u «стремится» не выпустить вектор состояния системы из $\lambda\mathcal{M}$, если вектор состояния оказался во множестве $\lambda\mathcal{M}$.

Таким образом, возникает следующая идея кусочно-липшицевой обратной связи, которая заставляет решение (Г.29) выживать во множестве \mathcal{M} . Сначала выбираем управление $w(t, x) = u$ по формуле (Г.30). Если вектор состояния системы $x(t)$ в момент $t = t_1 \geq t_0$ попал во множество $\frac{1}{2}\mathcal{M}$, то положим $w(t, x) = 0$ при $t > t_1$ до тех пор, пока вектор состояния не выйдет на границу $\partial\mathcal{M}$ в момент времени $t = t_2$. Если $x(t_2) \in \partial\mathcal{M}$, то определяем $w(t, x) = u$ при $t > t_2$ по формуле (Г.30).

Если вектор состояния системы попадает во множество $\frac{1}{2}\mathcal{M}$ в момент времени $t_3 > t_2$, то снова задаем $w(t, x) = 0$ при $t > t_3$ до выхода (если таковой случится) вектора состояния на границу $\partial\mathcal{M}$ и т. д. Получаем систему

$$\dot{x} = f(x) + w(t, x), \quad x(t_0) = x_0 \in \mathcal{M},$$

в которой правая часть кусочно-липшицева по x . По построению функции w решение этой задачи Коши выживает в \mathcal{M} .

Г.10. Неподвижные точки и точки совпадения отображений^{1*}

Пусть даны множество $\mathcal{X} \subset \mathbb{R}^n$ и отображение $F: \mathcal{X} \rightarrow \mathcal{X}$. Точка $x_0 \in \mathcal{X}$ называется *неподвижной точкой* отображения F , если $F(x_0) = x_0$. Теоремами о неподвижных точках называются различные утверждения, содержащие достаточные условия, основанные на свойствах множества \mathcal{X} и отображения F , которые гарантируют существование у него неподвижной точки, см. пример 1.5.6. Соответствующие результаты составляют содержание теории неподвижных точек, имеющей большое значение как для теории, так и для приложений. К вопросам существования неподвижных точек сводятся теоремы существования многих типов уравнений, в том числе дифференциальных, задачи теории управляемых систем, динамических систем, теории игр и т. д.

Напомним, что метрическое пространство полно, если любая фундаментальная последовательность из этого пространства сходится к элементу этого пространства.

Начнем с *принципа сжимающих отображений*, который формулируется следующим образом.

Теорема Г.20 (Банах – Качионполи). *Пусть заданы $\mathcal{X} \subset \mathbb{R}^n$ – замкнутое подмножество и $\rho_{\mathcal{X}}$ – метрика на \mathcal{X} , такая, что пространство $(\mathcal{X}, \rho_{\mathcal{X}})$ полное, а $F: \mathcal{X} \rightarrow \mathcal{X}$ – сжимающее отображение, то есть существует такое число $k \in (0, 1)$, что*

$$\rho_{\mathcal{X}}(F(x_1), F(x_2)) \leq k \rho_{\mathcal{X}}(x_1, x_2) \quad \forall x_1, x_2 \in \mathcal{X}.$$

Тогда у отображения F существует неподвижная точка x_0 , причем она единственна. Для произвольной точки $\xi \in \mathcal{X}$ последовательность итераций $F^{(n)}(\xi)$ отображения F сходится при $n \rightarrow \infty$ к точке x_0 .

В приложениях часто в качестве $\rho_{\mathcal{X}}(x_1, x_2)$ рассматривается евклидово расстояние $\|x_1 - x_2\|$, и в этом случае полнота пространства $(\mathcal{X}, \rho_{\mathcal{X}})$ следует из замкнутости \mathcal{X} . Однако метрика $\rho_{\mathcal{X}}$ может быть произвольной.

Доказательство теоремы Г.20 есть во многих учебниках по анализу (см., например, [10]). Оно вытекает из более общего утверждения, которое будет приведено ниже.

¹При написании раздела были использованы материалы, предоставленные А.В. Арутюновым.

Приведем теперь понятие точки совпадения двух отображений, которое является естественным развитием понятия неподвижной точки.

Итак, пусть $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{Y} \subset \mathbb{R}^m$ и $\Psi, \Phi: \mathcal{X} \rightarrow \mathcal{Y}$ – заданные непрерывные отображения.

Определение Г.15. Точка $x \in \mathcal{X}$ называется *точкой совпадения* отображений Ψ и Φ , если $\Psi(x) = \Phi(x)$.

Пусть $\rho_{\mathcal{X}}$ – метрика на \mathcal{X} , а $\rho_{\mathcal{Y}}$ – метрика на \mathcal{Y} . Определим для точки $x_0 \in \mathcal{X}$ метрический шар

$$\mathcal{B}^{\mathcal{X}}(x_0, r) = \{x \in \mathcal{X}: \rho_{\mathcal{X}}(x, x_0) < r\}.$$

Аналогично для точки $y_0 \in \mathcal{Y}$ определим метрический шар

$$\mathcal{B}^{\mathcal{Y}}(y_0, r) = \{y \in \mathcal{Y}: \rho_{\mathcal{Y}}(y, y_0) < r\}.$$

Ниже также будем предполагать полноту пространства $(\mathcal{X}, \rho_{\mathcal{X}})$.

Определение Г.16. Пусть задано $\alpha > 0$. Отображение $\Psi: \mathcal{X} \rightarrow \mathcal{Y}$ называется *α -накрывающим*, если

$$\Psi(\mathcal{B}^{\mathcal{X}}(x, r)) \supset \mathcal{B}^{\mathcal{Y}}(\Psi(x), \alpha r) \quad \forall r \geq 0 \quad \forall x \in \mathcal{X}. \quad (\text{Г.31})$$

Отображение называется накрывающим, если оно является α -накрывающим для некоторого $\alpha > 0$.

Обсудим понятие накрывающего отображения. Очевидно, накрывающее отображение сюръективно. Также очевидно, что если для некоторого $\alpha > 0$ отображение Ψ является α -накрывающим, то оно является γ -накрывающим и для произвольного $0 < \gamma \leq \alpha$. Простейшим примером 1-накрывающего отображения $\Psi: \mathcal{X} \rightarrow \mathcal{X}$ является тождественное отображение $\Psi(x) \equiv x$.

Если $F = A$, где $A: \mathcal{X} \rightarrow \mathcal{Y}$ – линейный непрерывный оператор, то он является накрывающим для некоторого $\alpha > 0$ тогда и только тогда, когда его образ $\{F(x): x \in \mathcal{X}\}$ совпадает со всем множеством \mathcal{Y} . Это утверждение составляет теорему Банаха об открытом отображении и известно [10].

Определение Г.17. Пусть $\beta \geq 0$. Будем говорить, что отображение $\Phi: \mathcal{X} \rightarrow \mathcal{Y}$ является *β -липшицевым*, если оно удовлетворяет условию Липшица с константой Липшица β , то есть

$$\rho_{\mathcal{Y}}(\Phi(x_1), \Phi(x_2)) \leq \beta \rho_{\mathcal{X}}(x_1, x_2) \quad \forall x_1, x_2 \in \mathcal{X}.$$

Напомним, что любое липшицево отображение непрерывно. Кроме того, если $\mathcal{X} = \mathcal{Y}$, то отображение Φ является сжимающим тогда, когда оно является β -липшицевым при некотором $\beta < 1$.

В следующей теореме даны достаточные условия существования точек совпадения.

Теорема Г.21 ([47]). *Пусть даны числа $\alpha > \beta > 0$. Предположим, что отображение Ψ является α -накрывающим, отображение Φ является β -липшицевым и пространство $(\mathcal{X}, \rho_{\mathcal{X}})$ полное.*

Тогда для произвольного $x \in \mathcal{X}$ существует такое $\xi = \xi(x) \in \mathcal{X}$, что

$$\Psi(\xi) = \Phi(\xi), \quad \rho_{\mathcal{X}}(\xi, x) \leq \frac{\rho_{\mathcal{Y}}(\Psi(x), \Phi(x))}{\alpha - \beta}. \quad (\Gamma.32)$$

Доказательство. Деля метрику $\rho_{\mathcal{Y}}$ на α , не теряя общности будем считать, что $\alpha = 1$ и, значит, $\beta < 1$. Доказательство этой теоремы, как и доказательство приведенной выше теоремы Банаха–Качиоппали, основано на методе простой итерации.

Зафиксируем произвольное $x \in \mathcal{X}$ и положим $x_0 = x$. В силу условия накрывания (Г.31) существует точка $x_1 \in \mathcal{X}$ такая, что

$$\Psi(x_1) = \Phi(x_0), \quad \rho_{\mathcal{X}}(x_0, x_1) \leq \rho_{\mathcal{Y}}(\Psi(x_0), \Phi(x_0)). \quad (\Gamma.33)$$

Построим по индукции такую последовательность $\{x_i\}$, что при $i \geq 2$ выполняется

$$\Psi(x_i) = \Phi(x_{i-1}), \quad \rho_{\mathcal{X}}(x_i, x_{i-1}) \leq \beta \rho_{\mathcal{X}}(x_{i-1}, x_{i-2}). \quad (\Gamma.34)$$

Действительно, пусть искомые x_0, x_1, \dots, x_i уже построены. Тогда в силу (Г.31) существует точка $x_{i+1} \in \mathcal{X}$ такая, что

$$\Psi(x_{i+1}) = \Phi(x_i), \quad \rho_{\mathcal{X}}(x_{i+1}, x_i) \leq \rho_{\mathcal{Y}}(\Psi(x_i), \Phi(x_i)).$$

Но тогда

$$\rho_{\mathcal{X}}(x_{i+1}, x_i) \leq \rho_{\mathcal{Y}}(\Psi(x_i), \Phi(x_i)) = \rho_{\mathcal{Y}}(\Phi(x_{i-1}), \Phi(x_i)) \leq \beta \rho_{\mathcal{X}}(x_i, x_{i-1}),$$

где последнее неравенство вытекает из условия Липшица для отображения Φ . Построение x_{i+1} , а вместе с этим и всей искомой последовательности, завершено.

Покажем, что последовательность $\{x_i\}$ является фундаментальной. Действительно, возьмем произвольные номера $j > l$. В силу неравенства треугольника имеем

$$\rho_{\mathcal{X}}(x_j, x_l) \leq \rho_{\mathcal{X}}(x_j, x_{j-1}) + \dots + \rho_{\mathcal{X}}(x_{l+1}, x_l).$$

Но в силу неравенства из (Г.34) имеет место

$$\rho_{\mathcal{X}}(x_{s+1}, x_s) \leq \beta^s \rho_{\mathcal{X}}(x_1, x_0)$$

для любого s . Поэтому, используя формулу для суммы геометрической прогрессии, имеем

$$\rho_{\mathcal{X}}(x_j, x_l) \leq \sum_{s=l}^{j-1} \beta^s \rho_{\mathcal{X}}(x_1, x_0) \leq \beta^l (1-\beta)^{-1} \rho_{\mathcal{X}}(x_1, x_0)$$

и, значит, последовательность $\{x_i\}$ является фундаментальной, так как в силу полученного неравенства $\rho_{\mathcal{X}}(x_j, x_l) \rightarrow 0$ при $j > l \rightarrow \infty$. Кроме того, из него в силу (Г.33) имеем, что для любого i

$$\rho_{\mathcal{X}}(x_i, x_0) \leq (1-\beta)^{-1} \rho_{\mathcal{Y}}(\Psi(x_0), \Phi(x_0)). \quad (\text{Г.35})$$

Поскольку пространство $(\mathcal{X}, \rho_{\mathcal{X}})$ полное, то фундаментальная последовательность $\{x_i\}$ сходится к некоторой точке $\xi \in \mathcal{X}$. Поэтому в силу непрерывности отображения Φ последовательность $\{\Phi(x_i)\}$ также сходится. Следовательно, в силу равенства из (Г.34) последовательность $\{\Psi(x_i)\}$ также сходится, причем в силу непрерывности отображения Ψ она сходится к $\Psi(\xi)$. В силу указанного равенства из (Г.34) $\Psi(\xi) = \Phi(\xi)$. Переходя в (Г.35) к пределу при $i \rightarrow \infty$, окончательно получаем (Г.32). ■

Если в доказанной теореме взять $\mathcal{X} = \mathcal{Y}$, а в качестве Ψ — тождественное отображение $\Psi(x) \equiv x$, то получим принцип Банаха–Качиополи сжимающих отображений (точнее, существование неподвижной точки; ее единственность легко проверяется).

Д. Сведения из теории экстремальных задач

В этом разделе будут рассмотрены гладкие и выпуклые задачи. Мы отсылаем читателя по поводу гладких задач к стандартным курсам по анализу, например [14], а выпуклых — к монографиям по выпуклому анализу [6, 19, 24]. Мы также обращаем внимание на классические монографии, где освещены и гладкие, и выпуклые задачи, например [8].

Напомним, что точка $x_0 \in \mathcal{Q} \subset \mathbb{R}^n$ есть локальный минимум (максимум) $f: \mathbb{R}^n \rightarrow \mathbb{R}$ на \mathcal{Q} , если найдется такое $\delta > 0$, что для всех $x \in \mathcal{B}_\delta(x_0) \cap \mathcal{Q}$ выполнено неравенство $f(x) \geq f(x_0)$ ($f(x) \leq f(x_0)$).

Экстремум x_0 называется строгим, если предыдущие неравенства строгие для всех $x \in (\mathcal{B}_\delta(x_0) \setminus \{x_0\}) \cap \mathcal{Q}$.

Д.1. Безусловная оптимизация

Рассмотрим задачу $\min_{x \in \mathbb{R}^n} f(x)$ для функции $f: \mathbb{R}^n \rightarrow \mathbb{R}$; можно также рассмотреть задачу $\max_{x \in \mathbb{R}^n} f(x)$.

Пусть функция дифференцируема по Гато в точке x_0 . Тогда необходимое условие экстремума функции f в точке $x_0 \in \mathbb{R}^n$ задается *правилом Ферма*:

$$f'(x_0) = 0, \quad (\text{Д.1})$$

где $f'(x_0)$ — градиент Гато. Для дифференцируемой по Фреше функции f необходимое условие экстремума в точке x_0 также задается формулой (Д.1), где $f'(x_0)$ — градиент Фреше, см. (А.1), (А.2).

Если $f \in C^2$ в окрестности $x_0 \in \mathbb{R}^n$, то необходимое условие минимума (максимума) есть $f'(x_0) = 0$ и $(h, f''(x_0)h) \geq 0$ для всех $h \in \mathbb{R}^n$ ($f'(x_0) = 0$ и $(h, f''(x_0)h) \leq 0$ для всех $h \in \mathbb{R}^n$). Если $f'(x_0) = 0$ и найдутся два таких вектора $h_i \in \mathbb{R}^n$, $i = 1, 2$, что выполнены неравенства $(h_1, f''(x_0)h_1) < 0$ и $(h_2, f''(x_0)h_2) > 0$, то в точке x_0 экстремума нет.

Достаточное условие локального минимума $f \in C^2$ в точке x_0 дает условие $f'(x_0) = 0$ и $(h, f''(x_0)h) > 0$ для всех $h \in \mathbb{R}^n \setminus \{0\}$. Достаточным условием локального максимума является условие $f'(x_0) = 0$ и $(h, f''(x_0)h) < 0$ для всех $h \in \mathbb{R}^n \setminus \{0\}$.

Все приведенные выше условия вытекают из определений экстремума и разложения (А.4).

Если функция f выпуклая и дифференцируемая в точке x_0 , то условие $f'(x_0) = 0$ является необходимым и достаточным условием глобального минимума f на \mathbb{R}^n .

Для негладкой выпуклой функции f необходимым и достаточным условием минимума в точке $x_0 \in \mathbb{R}^n$ является включение $0 \in \partial f(x_0)$, см. раздел Г.5.

Довольно редко в экстремальных задачах бывают известны разложения вида (А.3). Тогда при четном m при условии $\frac{\partial^k f(x_0)}{\partial \mathbf{x}_{i_1} \dots \partial \mathbf{x}_{i_k}} = 0$ для всех $k < m$ и при $\sum_{i_1, \dots, i_m} \frac{\partial^m f(x_0)}{\partial \mathbf{x}_{i_1} \dots \partial \mathbf{x}_{i_m}} h_{i_1} \dots h_{i_m} > 0$ при всех $h \in \mathbb{R}^n \setminus \{0\}$ точка x_0 есть точка локального минимума, а при обратном неравенстве (< 0) — точка локального максимума. Если m нечетно, $\frac{\partial^k f(x_0)}{\partial \mathbf{x}_{i_1} \dots \partial \mathbf{x}_{i_k}} = 0$ при всех $k < m$ и $\sum_{i_1, \dots, i_m} \frac{\partial^m f(x_0)}{\partial \mathbf{x}_{i_1} \dots \partial \mathbf{x}_{i_m}} \neq 0$ для некоторых наборов i_1, \dots, i_m , то экстремума нет.

**Д.2. Гладкая задача условного экстремума.
Метод множителей Лагранжа**

Рассмотрим задачу в \mathbb{R}^n

$$\min(\max) f_0(x) \quad \text{при условии } f_i(x) = 0, \quad 1 \leq i \leq m. \quad (\text{Д.2})$$

Как и в разделе 2.8.1, введем в рассмотрение *функцию Лагранжа*

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

где $\mathbb{R}^m \ni \lambda = (\lambda_1, \dots, \lambda_m)^T$ — вектор множителей Лагранжа и

$$f(x) = (f_1(x), \dots, f_m(x))^T.$$

Аналогом правила Ферма (Д.1) в рассматриваемом случае является следующая теорема.

Теорема Д.1. Пусть функции $f_i \in C^1$, $0 \leq i \leq m$, точка x_0 удовлетворяет системе $f(x_0) = 0$ и является локальным экстремумом в задаче (Д.2). Если в точке x_0 выполнено условие полного ранга $\text{rank}(f'_1(x_0), \dots, f'_m(x_0)) = m$, то найдется вектор множителей Лагранжа $\lambda \in \mathbb{R}^m$ такой, что

$$L'_x(x_0, \lambda) = f'_0(x_0) + \sum_{i=1}^m \lambda_i f'_i(x_0) = 0.$$

Точка x_0 , удовлетворяющая условиям $f(x_0) = 0$ и $L'_x(x_0, \lambda) = 0$ для некоторого вектора λ , называется *стационарной точкой функции Лагранжа*, или просто *стационарной точкой*. Таким образом, чтобы найти все подозрительные на экстремум точки в задаче (Д.2), надо решить систему $f(x) = 0$ и $L'_x(x, \lambda) = 0$. В этой системе имеется $n+m$ переменных $(x, \lambda) \in \mathbb{R}^{n+m}$ и столько же уравнений. Отметим, что без выполнения условия полного ранга, то есть при

$$\text{rank}(f'_1(x_0), \dots, f'_m(x_0)) < m,$$

можно утверждать, что существует вектор $(\lambda_0, \lambda) \in \mathbb{R}^{m+1}$, $\lambda_0 = 0$, такой, что

$$\sum_{i=0}^m \lambda_i f'_i(x_0) = \sum_{i=1}^m \lambda_i f'_i(x_0) = 0,$$

то есть приходим к тавтологии. Именно условие полного ранга гарантирует, что $\lambda_0 = 1$.

Приведем простейший пример. В пространстве \mathbb{R}^2 рассмотрим задачу (Д.2) с функцией $f_0(x_1, x_2) = x_1^2 + (x_2 - 1)^2$ и $f_1(x_1, x_2) = x_2^2 - \frac{x_1^4}{4} = 0$ (одно ограничение). Отметим, что ограничение $f_1(x_1, x_2) = 0$ задает две параболы $x_2 = \pm \frac{x_1^2}{2}$. Решением (минимумом) является точка $(0; 0)$, так как круг радиуса 1 с центром $(0; 1)$ является касательным к параболе $x_2 = \frac{x_1^2}{2}$ в точке $(0; 0)$. С другой стороны,

$$f'_1(0, 0) = (-x_1^3, 2x_2)|_{(0, 0)} = (0, 0).$$

Условие полного ранга не выполнено, теорема Д.1 не работает.

Для системы $f(x) = 0$, $f \in C^1$, в точке x_0 ($f(x_0) = 0$) определим касательное подпространство

$$T(x_0) = \{h \in \mathbb{R}^n : f'(x_0)h = 0\}. \quad (\text{Д.3})$$

Линейное подпространство $T(x_0)$ есть пересечение касательных подпространств — решений систем $f'_i(x_0)h = 0$, $1 \leq i \leq m$.

Следующая теорема дает достаточные условия для минимума или максимума в точке x_0 в задаче (Д.2) при условии, что все функции $f_i \in C^2$, $0 \leq i \leq m$.

Теорема Д.2. Пусть функции $f_i \in C^2$, $0 \leq i \leq m$, и точка x_0 является стационарной точкой функции Лагранжа, $\lambda \in \mathbb{R}^m$ есть соответствующий вектор множителей Лагранжа. Определим квадратичную форму

$$k(h) = (h, L''_{xx}(x_0, \lambda)h), \quad h \in \mathbb{R}^n.$$

Тогда для задачи (Д.2) выполнены следующие утверждения:

- если квадратичная форма $k(h) > 0$ для всех $h \in T(x_0) \setminus \{0\}$, то точка x_0 — локальный минимум;
- если квадратичная форма $k(h) < 0$ для всех $h \in T(x_0) \setminus \{0\}$, то точка x_0 — локальный максимум;
- если квадратичная форма $k(h)$ принимает разные знаки на векторах $h_1, h_2 \in T(x_0) \setminus \{0\}$, то x_0 не является точкой экстремума.

Заметим, что теорема Д.2 не дает ответа на вопрос об экстремуме при условии $k(h) \geq 0$ или $k(h) \leq 0$ для всех $h \in T(x_0) \setminus \{0\}$.

Отметим, что необходимое условие экстремума в теореме Д.1 геометрически означает, что антиградиент $-f'(x_0)$ находится в нормальном конусе к касательному конусу (и в данном случае подпространстве) $T(x_0)$ (Д.3).

Д.3. Выпуклая задача условного экстремума

Пусть $f_i: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, $i \in \{0\} \cup I_1$, $I_1 = \{1, \dots, m\}$, — выпуклые функции, а $f_i(x) = (p_i, x) + \nu_i = 0$, $i \in I_2 = \{m+1, \dots, N\}$, — линейные функции. Далее будем считать, что векторы $\{p_i\}_{i=m+1}^N$ линейно независимы. Пусть $\mathcal{Q} \subset \mathbb{R}^n$ — выпуклое замкнутое множество. Рассмотрим задачу

$$\min f_0(x) \quad \text{при } f_i(x) \leq 0, \quad i \in I_1, \quad f_i(x) = 0, \quad i \in I_2, \quad x \in \mathcal{Q}. \quad (\text{Д.4})$$

Определим конус

$$\mathcal{K} = \{\lambda = (\lambda_0, \dots, \lambda_{N+1}) \in \mathbb{R}^{N+1}: \lambda_i \geq 0, 0 \leq i \leq m\}.$$

Функцией Лагранжа для задачи (Д.4) будем называть

$$L(x, \lambda) = \sum_{i=0}^N \lambda_i f_i(x), \quad \lambda \in \mathcal{K}, \quad x \in \mathcal{Q}.$$

Седловой точкой функции Лагранжа будем называть такой вектор $(x_0, \bar{\lambda}) \in \mathcal{Q} \times \mathcal{K}$, что

$$L(x_0, \lambda) \leq L(x_0, \bar{\lambda}) \leq L(x, \bar{\lambda}) \quad \forall x \in \mathcal{Q} \quad \forall \lambda \in \mathcal{K}.$$

Очевидным следствием приведенного определения седловой точки является равенство

$$\min_{x \in \mathcal{Q}} L(x, \bar{\lambda}) = \max_{\lambda \in \mathcal{K}} L(x_0, \lambda) = L(x_0, \bar{\lambda}).$$

Следующая теорема доказана Г. Куном и А. Таккером; она дает необходимые, а также достаточные условия для решения задачи (Д.4).

Теорема Д.3 (Кун–Таккер [19, теорема 2.4.1]). 1. Если точка x_0 — решение задачи (Д.4), то найдется ненулевой вектор множителей Лагранжа $\bar{\lambda} \in \mathcal{K}$ такой, что выполняются условия:

a) принцип минимума для функции Лагранжа

$$\min_{x \in \mathcal{Q}} L(x, \bar{\lambda}) = L(x_0, \bar{\lambda}); \quad (\text{Д.5})$$

б) условия дополняющей несвесткости

$$\bar{\lambda}_i f_i(x_0) = 0, \quad i \in I_1. \quad (\text{Д.6})$$

2. Если для допустимой точки x_0 в задаче (Д.4) существует ненулевой вектор $\bar{\lambda} \in \mathcal{K}$ такой, что для него выполнены условия а), б) и $\bar{\lambda}_0 > 0$, то точка x_0 — решение задачи (Д.4), а $(x_0, \bar{\lambda})$ — седловая точка функции Лагранжа.

3. Если в задаче (Д.4) выполнено условие Слейтера, то есть существует такая точка $\bar{x} \in \mathbb{R}^n$, что $\bar{x} \in \text{int } \mathcal{Q}$ и $f_i(\bar{x}) < 0$ для всех $i \in I_1$, а для допустимой точки x_0 существует ненулевой вектор $\bar{\lambda} \in \mathcal{K}$ такой, что выполнены условия а), б), то точка x_0 — решение задачи (Д.4), а $(x_0, \bar{\lambda})$ — седловая точка функции Лагранжа.

Таким образом, при выполнении условия Слейтера теорема дает необходимые и достаточные условия решения задачи (Д.4). Условие Слейтера не требуется, если множество \mathcal{Q} есть выпуклый полиэдр, а функции f_i , $i \in I_1$, — аффинные [24, §19, §20].

Если преобразовать (Д.5) к субградиентной форме, то получим с учетом теоремы Моро–Рокафеллара Г.15 (если она применима):

$$0 \in \sum_{i=0}^N \bar{\lambda}_i \partial f_i(x_0) + N(\mathcal{Q}, x_0),$$

где $N(\mathcal{Q}, x_0)$ — нормальный конус к \mathcal{Q} в точке x_0 (Г.12). Совместно с условиями дополняющей нежесткости (Д.6) получаем систему для нахождения точки $(x_0, \bar{\lambda})$.

Д.4. Замечание о необходимых условиях минимума функции на проксимально гладком множестве

Если множество $\mathcal{Q} \subset \mathbb{R}^n$ проксимально гладкое (см. определение Г.12), то в каждой точке $x_0 \in \mathcal{Q}$ оно регулярно. Последнее означает, что основные касательные конусы ко множеству \mathcal{Q} в точке x_0 , включая конус Кларка, нижний и верхний касательный конус (конус Булигана) (см. определения указанных конусов в [19, §1.4]) совпадают [80, раздел 6], [124, следствие 2.2]. Это же доказано в [7, теорема 1.17.1].

Пусть

$$T_H(\mathcal{Q}, x_0) = \{q \in \mathbb{R}^n : x_0 + \lambda q + o(\lambda) \in \mathcal{Q}, \lambda \rightarrow +0\}$$

есть нижний касательный конус к \mathcal{Q} в точке $x_0 \in \mathcal{Q}$. Для любой точки $x_0 \in \mathcal{Q}$ с учетом равенства касательных конусов и теоремы [7, теорема 1.17.3] для проксимально гладкого множества \mathcal{Q} выполнено равенство для конуса проксимальных нормалей $N_P(\mathcal{Q}, x_0)$:

$$N_P(\mathcal{Q}, x_0) = T_H^-(\mathcal{Q}, x_0) = \{p \in \mathbb{R}^n : (p, q) \leq 0 \quad \forall q \in T_H(\mathcal{Q}, x_0)\}.$$

Покажем, что необходимое условие минимума в точке $x_0 \in \mathcal{Q}$ дифференцируемой по Фреше функции f есть включение

$$-f'(x_0) \in N_P(\mathcal{Q}, x_0). \quad (\text{Д.7})$$

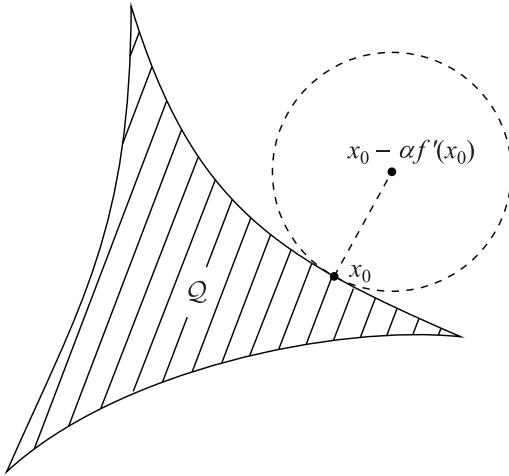


Рис. Д.1. Иллюстрация к формуле (Д.8).

От противного: пусть включение неверно. Тогда для некоторого вектора $q \in T_H(\mathcal{Q}, x_0)$ выполнено неравенство $(f'(x_0), q) < 0$. Отсюда в силу определения нижнего касательного конуса для всех малых $\lambda > 0$ выполнено включение $x_0 + \lambda q + o(\lambda) \in \mathcal{Q}$ и

$$f(x_0) \leq f(x_0 + \lambda q + o(\lambda)) = f(x_0) + \lambda(f'(x_0), q) + o(\lambda) < f(x_0)$$

при достаточно малом $\lambda > 0$ в силу условия $(f'(x_0), q) < 0$. Итак, условие (Д.7) доказано.

В силу опорного условия 1) для проксимально гладких множеств включение (Д.7) эквивалентно условию

$$P_{\mathcal{Q}}(x_0 - \alpha f'(x_0)) = \{x_0\} \quad \forall \alpha \in \left(0, \frac{R}{\|f'(x_0)\|}\right) \quad (\text{Д.8})$$

при $f'(x_0) \neq 0$, см. рис. Д.1.

Литература

Учебники и монографии

- [1] Арутюнов А.В. Лекции по выпуклому и многозначному анализу. М.: Физматлит, 2014.
- [2] Баландин Д.В., Коган М.М. Синтез законов управления на основе линейных матричных неравенств. М.: Физматлит, 2007.
- [3] Голуб Дж., Ван Лоун Ч. Матричные вычисления. М.: Мир, 1999.
- [4] Грюнбаум Б. Этюды по комбинаторной геометрии и теории выпуклых тел. М.: Наука, 1971.
- [5] Дистель Дж. Геометрия банаховых пространств. Киев: Вища школа, 1980.
- [6] Дмитрук А.В. Выпуклый анализ. Элементарный курс. 2-е изд., перераб. и доп. М.: Макс-Пресс, 2023.
- [7] Иванов Г.Е. Слабо выпуклые функции и множества. М.: Физматлит, 2006.
- [8] Иоффе А.Д., Тихомиров В.М. Теория экстремальных задач. М.: Наука, 1974.
- [9] Квакернаак Х., Сиван Р. Линейные оптимальные системы управления. М.: Мир, 1977.
- [10] Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. М.: Наука, 1972.
- [11] Куржанский А.Б. Управление и наблюдение в условиях неопределенности. М.: Наука, 1977.

- [12] МАТАСОВ А.И. Основы теории фильтра Калмана. М.: Изд-во МГУ, 2021.
- [13] НЕСТЕРОВ Ю.Е. Введение в выпуклую оптимизацию. М.: МЦНМО, 2010.
- [14] НИКОЛЬСКИЙ С.М. Курс математического анализа. Т. 1. М.: Наука, 1983.
- [15] НОВИКОВ Д.А. (ред.) Теория управления. Дополнительные главы. М.: ЛЕНАНД, 2019.
- [16] НОВИКОВ Д.А. Теория управления организационными системами. 3-е изд. М.: Физматлит, 2012.
- [17] НОВИКОВ Д.А. (ред.) Теория управления: словарь системы основных понятий. М.: ИПУ РАН, 2024.
- [18] ОБЕН Ж.-П., ЭКЛАНД И. Прикладной нелинейный анализ. М.: Мир, 1988.
- [19] ПОЛОВИНКИН Е.С., БАЛАШОВ М.В. Элементы выпуклого и сильно выпуклого анализа. 2-е изд., испр. и доп. М.: Физматлит, 2007.
- [20] ПОЛЯК Б.Т. Введение в оптимизацию. 2-е изд. М.: ЛЕНАНД, 2014.
- [21] ПОЛЯК Б.Т., ХЛЕБНИКОВ М.В., РАПОПОРТ Л.Б. Математическая теория автоматического управления. М.: ЛЕНАНД, 2019.
- [22] ПОЛЯК Б.Т., ХЛЕБНИКОВ М.В., ЩЕРБАКОВ П.С. Управление линейными системами при внешних возмущениях: Техника линейных матричных неравенств. М.: ЛЕНАНД, 2014.
- [23] ПОНТРЯГИН Л.С. Обыкновенные дифференциальные уравнения. М.: Наука, 1974.
- [24] РОКАФЕЛЛАР Р.Т. Выпуклый анализ. М.: Мир, 1973.
- [25] РУДИН У. Основы математического анализа. М.: Мир, 1976.
- [26] ХОРН Р., ДЖОНСОН Ч. Матричный анализ. М.: Мир, 1989.
- [27] ЧЕРНОУСЬКО Ф.Л. Оценивание фазового состояния динамических систем. М.: Наука, 1988.

- [28] ABSIL P.-A., MAHONY R., SEPULCHRE R. Optimization Algorithms on Matrix Manifolds. Princeton: Princeton University Press, 2008.
- [29] ANDERSON B.D.O., MOORE J.B. Linear Optimal Control. N.J.: Prentice Hall, 1971.
- [30] ÅSTRÖM K.J., HÄGGLUND T. PID Controllers: Theory, Design, and Tuning. Research Triangle Park: Instrument Society of America, 1995.
- [31] ÅSTRÖM K.J., HÄGGLUND T. Advanced PID Control. Research Triangle Park: The Instrumentation, Systems, and Automation Society, 2006.
- [32] ATHANS M., FALB P.L. Optimal Control: An Introduction to the Theory and Its Applications. N.Y.: McGraw-Hill, 1966.
- [33] AUBIN J.-P. Viability Theory. N.Y.: Springer Science & Business Media, 2009.
- [34] AUBIN J.-P., CELLINA A. Differential Inclusions. Set-Valued Maps and Viability Theory. Berlin: Springer-Verlag, 1984.
- [35] BHATTACHARYYA S.P., KEEL L.H. Linear Multivariable Control Systems. N.Y.: Cambridge University Press, 2022.
- [36] BONNANS J.F., SHAPIRO A. Perturbation Analysis of Optimization Problems. N.Y.: Springer, 2000.
- [37] BOYD S., EL GHAOUI L., FERON E., BALAKRISHNAN V. Linear Matrix Inequalities in System and Control Theory. Philadelphia: SIAM, 1994.
- [38] CLARKE F.H., LEDYAEV YU.S., STERN R.J., WOLENSKI R.R. Nonsmooth Analysis and Control Theory. N.Y.: Springer, 1998.
- [39] KAILATH T., SAYED A.H., HASSIBI B. Linear Estimation. N.J.: Prentice Hall, 2000.
- [40] KWAKERNAAK H., SIVAN R. Linear Optimal Control Systems. N.Y.: Wiley, 1972.
- [41] MORDUKHOVICH B.S. Variational Analysis and Generalized Differentiation I: Basic Theory. Berlin: Springer, 2005.

- [42] NESTEROV Y., NEMIROVSKII A. Interior-Point Polynomial Algorithms in Convex Programming. 3rd printing. Philadelphia: SIAM, 2001.
- [43] ROCKAFELLAR R.T., WETS R.J.-B. Variational Analysis. Berlin: Springer, 2009.
- [44] SCHWEPPE F.C. Uncertain Dynamic Systems. N.J.: Prentice Hall, 1973.
- [45] VISIOLI A. Practical PID Control. London: Springer, 2006.
- [46] WANG Q.-G., YE Z., CAI W.-J., HANG C.-C. PID Control for Multivariable Processes. Berlin: Springer, 2008.

Статьи

- [47] Арутюнов А.В. Накрывающие отображения в метрических пространствах и неподвижные точки // Доклады Академии наук. 2007. Т. 416. № 2. С. 151–155.
- [48] Балашов М.В. Условие Липшица метрической проекции и сходимость градиентных методов // Математический сборник. 2024. <https://www.mathnet.ru/rus/sm9982>.
- [49] Балашов М.В. Метод проекции градиента на матричных многообразиях // Журнал вычислительной математики и математической физики. 2020. Т. 60. № 9. С. 1453–1461.
- [50] Балашов М.В., Биглов К.З. Опорное условие сильной выпуклости и условие Липшица для метрической проекции // Математические заметки. 2024. Т. 115. Вып. 2. С. 197–207.
- [51] Балашов М.В., Камалов Р.А. Метод проекции градиента с шагом Армихо на многообразиях // Журнал вычислительной математики и математической физики. 2021. Т. 61. № 11. С. 1814–1824.
- [52] Грязина Е.Н., Поляк Б.Т., ТРЕМБА А.А. Синтез регуляторов низкого порядка по критерию H_∞ : параметрический подход // Автоматика и телемеханика. 2007. № 3. С. 94–105.
- [53] Грязина Е.Н., Поляк Б.Т., ТРЕМБА А.А. Современное состояние метода D-разбиения // Автоматика и телемеханика. 2008. № 12. С. 3–40.

- [54] КАНТОРОВИЧ Л.В. Функциональный анализ и прикладная математика // Успехи математических наук. 1948. Т. 3. Вып. 6(28). С. 89–185.
- [55] КИСЕЛЕВ О.Н., Поляк Б.Т. Синтез регуляторов низкого порядка по критерию H^∞ и по критерию максимальной робастности // Автоматика и телемеханика. 1999. № 3. С. 119–130.
- [56] ЛЕВИН А.Ю. Об одном алгоритме минимизации выпуклых функций // Доклады АН СССР. 1965. Т. 160. № 6. С. 1244–1247.
- [57] ЛЕВИТИН Е.С., Поляк Б.Т. Методы минимизации при наличии ограничений // Журнал вычислительной математики и математической физики. 1966. Т. 6. № 5. С. 787–823.
- [58] НАЗИН С.А., Поляк Б.Т., Топунов М.В. Подавление ограниченных внешних возмущений с помощью метода инвариантных эллипсоидов // Автоматика и телемеханика. 2007. № 3. С. 106–125.
- [59] ПОЛОВИНКИН Е.С. Сильно выпуклый анализ // Математический сборник. 1996. Т. 187. № 2. С. 103–130.
- [60] Поляк Б.Т. Градиентные методы минимизации функционалов // Журнал вычислительной математики и математической физики. 1963. Т. 3. № 4. С. 643–653.
- [61] Поляк Б.Т. Градиентные методы решения уравнений и неравенств // Журнал вычислительной математики и математической физики. 1964. Т. 4. № 6. С. 995–1005.
- [62] Поляк Б.Т. Теоремы существования и сходимость минимизирующих последовательностей для задач на экстремум при наличии ограничений // Доклады АН СССР. 1966. Т. 166. № 2. С. 287–290.
- [63] Поляк Б.Т. Минимизация негладких функционалов // Журнал вычислительной математики и математической физики. 1969. Т. 9. № 3. С. 509–521.
- [64] Поляк Б.Т., Топунов М.В. Фильтрация при неслучайных возмущениях: метод инвариантных эллипсоидов // Доклады Академии наук. 2008. Т. 418. № 6. С. 749–753.

- [65] Поляк Б.Т., ХЛЕБНИКОВ М.В., ЩЕРБАКОВ П.С. Линейные матричные неравенства в системах управления с неопределенностью // Автоматика и телемеханика. 2021. № 1. С. 3–54.
- [66] ХЛЕБНИКОВ М.В., Поляк Б.Т., Кунцевич В.М. Оптимизация линейных систем при ограниченных внешних возмущениях (техника инвариантных эллипсоидов) // Автоматика и телемеханика. 2011. № 11. С. 9–59.
- [67] ХЛЕБНИКОВ М.В., ЩЕРБАКОВ П.С., ЧЕСТНОВ В.Н. Линейно-квадратичное управление: I. Новое решение // Автоматика и телемеханика. 2015. № 12. С. 78–92.
- [68] ABADIE J., CARPENTIER J. Generalization of the Wolfe reduced gradient method to the case of nonlinear constraints // Optimization. R. Fletcher (Ed.). N.Y.: Academic Press, 1969. P. 37–47.
- [69] ABEDOR J., NAGPAL K., POOLLA K. A linear matrix inequality approach to peak-to-peak gain minimization // International Journal on Robust and Nonlinear Control. 1996. Vol. 6. No. 9–10. P. 899–927.
- [70] ADLY S., NACRY F., THIBAULT L. Preservation of prox-regularity of sets with applications to constrained optimization // SIAM Journal on Optimization. 2016. Vol. 26. Iss. 1. P. 448–473.
- [71] ACKERMANN J. Parameter space design of robust control systems // IEEE Transactions on Automatic Control. 1980. Vol. 25. No. 6. P. 1058–1072.
- [72] AMMAN H.M., NEUDECKER H. Numerical solutions of the algebraic matrix Riccati equation // Journal of Economic Dynamics and Control. 1997. Vol. 21. Iss. 2–3. P. 363–369.
- [73] ÅSTRÖM K.J., HAGGLUND T. Benchmark systems for PID control // IFAC Proceedings Volumes. 2000. Vol. 33. Iss. 4. P. 165–166.
- [74] BALASHOV M.V. About the gradient projection algorithm for a strongly convex function and a proximally smooth set // Journal of Convex Analysis. 2017. Vol. 24. Iss. 2. P. 493–500.
- [75] BALASHOV M.V., REPOVŠ D. Uniformly convex subsets of the Hilbert space with modulus of convexity of the second order //

Journal of Mathematical Analysis and Applications. 2011. Vol. 377. Iss. 2. P. 754–761.

- [76] BELLMAN R. Notes on matrix theory – X. A problem in control // Quarterly of Applied Mathematics. 1957. Vol. 14. No. 4. P. 417–419.
- [77] BENNER P., LI J.-R., PENZL T. Numerical solution of large-scale Lyapunov equations, Riccati equations, and linear-quadratic optimal control problems // Numerical Linear Algebra with Applications. 2008. Vol. 15. Iss. 9. P. 755–777.
- [78] BLANCHINI F., LEPSCHY A., MIANI S., VIARO U. Characterization of PID and lead/lag compensators satisfying given H_∞ specifications // IEEE Transactions on Automatic Control. 2004. Vol. 49. No. 5. P. 736–740.
- [79] BLONDEL V., TSITSIKLIS J.N. NP-hardness of some linear control design problems // SIAM Journal on Control and Optimization. 1997. Vol. 35. No. 6. P. 2118–2127.
- [80] BOUNKHEL M., THIBAULT L. On various notions of regularity of sets in nonsmooth analysis // Nonlinear Analysis. 2002. Vol. 48. No. 2. P. 223–246.
- [81] BRAUN G., CARDERERA A., COMBETTES C.W., HASSANI H., KARBASI A., MOKHTARI A., POKUTTA S. Conditional gradient methods // arXiv:2211.14103v1. 2022.
- [82] CLARKE F.H., STERN R.J., WOLENSKI P.R. Proximal smoothness and the lower- C^2 property // Journal of Convex Analysis. 1995. Vol. 2. No. 1-2. P. 117–144.
- [83] COHEN A.I. Rate of Convergence of Several Conjugate Gradient Algorithms // SIAM Journal on Numerical Analysis. 1972. Vol. 9. No. 2. P. 248–259.
- [84] CONWAY J.H., HARDIN R.H., SLOANE N.J.A. Packing lines, planes, etc.: packings in grassmannian spaces // Experimental Mathematics. 1996. Vol. 5. No. 2. P. 139–159.
- [85] EILBRECHT J., JILG M., STURSBERG O. Distributed H_2 -optimized output feedback controller design using the ADMM // IFAC-PapersOnLine. 2017. Vol. 50. No. 1. P. 10389–10394.

- [86] FATKHULLIN I., POLYAK B. Optimizing static linear feedback: Gradient method // SIAM Journal on Control and Optimization. 2021. Vol. 59. No. 5. P. 3887–3911.
- [87] FAZEL M., GE R., KAKADE S., MESBAHI M. Global convergence of policy gradient methods for the linear quadratic regulator // Proc. 35th International Conference on Machine Learning. 2018. P. 1467–1476.
- [88] FENG H., LAVAEI J. On the exponential number of connected components for the feasible set of optimal decentralized control problems // Proc. 2019 American Control Conference. 2019. P. 1430–1437.
- [89] FRANKOWSKA H., OLECH C. R-convexity of the integral of the set-valued functions // Contributions to Analysis and Geometry. Baltimore: Johns Hopkins Univ. Press, 1981. P. 117–129.
- [90] GEEM Z.W., KIM J.H., LOGANATHAN G.V. A new heuristic optimization algorithm: Harmony search // Simulation. 2002. Vol. 76. No. 2. P. 60–68.
- [91] GRANT M., BOYD S. CVX: Matlab software for disciplined convex programming, version 2.1. URL: <http://cvxr.com/cvx>
- [92] GREENBERG H. Klee–Minty polytope shows exponential time complexity of simplex method. URL: <https://glossary.informs.org/notes/Klee-Minty.pdf>
- [93] HAN S., KEEL L.H., BHATTACHARYYA S.P. PID controller design with an H^∞ criterion // IFAC-PapersOnLine. 2018. Vol. 51. No. 4. P. 400–405.
- [94] HUMPHERRYS J., REDD P., WEST J. A fresh look at the Kalman filter // SIAM Review. 2012. Vol. 54. Iss. 4. P. 801–823.
- [95] IWASAKI T., SKELTON R.E., GEROMEL J.C. Linear quadratic suboptimal control with static output feedback // Systems & Control Letters. 1994. Vol. 23. Iss. 6. P. 421–430.
- [96] KALMAN R.E. Contributions to the theory of optimal control // Boletin de la Sociedad Matematica Mexicana. 1960. Vol. 5. No. 1. P. 102–119.

- [97] KALMAN R.E. On the general theory of control systems // Proc. 1st IFAC Congress. 1960. P. 481–491.
- [98] KARABOGA D., AKAY B. Proportional-integral-derivative controller design by using artificial bee colony, harmony search, and the bees algorithms // Journal of Systems and Control Engineering. 2010. Vol. 224. No. 7. P. 869–883.
- [99] KARIMI H., NUTINI J., SCHMIDT M. Linear convergence of gradient and proximal-gradient methods under the Polyak–Łojasiewicz condition // Lecture Notes in Computer Science. 2016. Vol. 9851. P. 795–811.
- [100] KLEE V., MINTY G.J. How good is the simplex algorithm // Inequalities. 1972. Vol. 3. No. 3. P. 159–175.
- [101] KWON W.H., MOON Y.S., AHN S.C. Bounds in algebraic Riccati and Lyapunov equations: A survey and some new results // International Journal of Control. 1996. Vol. 64. No. 3. P. 377–389.
- [102] LEE C.-H. New results for the bounds of the solution for the continuous Riccati and Lyapunov equations // IEEE Transactions on Automatic Control. 1997. Vol. 42. No. 1. P. 118–123.
- [103] LEVA A., PAPADOPOULOS A.V. Tuning of event-based industrial controllers with simple stability guarantees // Journal of Process Control. 2013. Vol. 23. P. 1251–1260.
- [104] LEVINE W., ATHANS M. On the determination of the optimal constant output feedback gains for linear multivariable systems // IEEE Transactions on Automatic Control. 1970. Vol. 15. Iss. 1. P. 44–48.
- [105] LEZANSKI T. Über das Minimumproblem für Funktionale in Banachschen Räumen // Mathematische Annalen. 1963. Vol. 152. No. 4. P. 271–274.
- [106] LI Y., ANG K.H., CHONG G.C.Y. PID control system analysis and design // IEEE Control Systems Magazine. 2006. Vol. 26. No. 1. P. 32–41.
- [107] LÖFBERG J. YALMIP: A toolbox for modeling and optimization in MATLAB // Proc. 2004 IEEE International Symposium on Computer Aided Control Systems Design (CACSD). Taipei, Taiwan, September 2–4, 2004. URL: <https://yalmip.github.io/>

- [108] LOJASIEWICZ S. Une propriété topologique des sous-ensembles analytiques réels // Les Équations aux Dérivées Partielles. 1963. P. 87–89.
- [109] LUENBERGER D.G. Observing the state of a linear system // IEEE Transactions on Military Electronics. 1964. Vol. 8. No. 2. P. 74–80.
- [110] LUENBERGER D.G. An introduction to observers // IEEE Transactions on Automatic Control. 1971. Vol. AC-16. No. 6. P. 596–602.
- [111] MAKILA P.M., TOIVONEN H.T. Computational methods for parametric LQ problems – A survey // IEEE Transactions on Automatic Control. 1987. Vol. 32. No. 8. P. 658–671.
- [112] MOERDER D., CALISE A. Convergence of a numerical algorithm for calculating optimal output feedback gains // IEEE Transactions on Automatic Control. 1985. Vol. 30. No. 9. P. 900–903.
- [113] MOHAMMADI H., ZARE A., SOLTANOLKOTABI M., JOVANOVIC M.R. Global exponential convergence of gradient methods over the nonconvex landscape of the linear quadratic regulator // Proc. 2019 IEEE 58th Conference on Decision and Control. 2019. P. 7474–7479.
- [114] MOHAMMADI H., ZARE A., SOLTANOLKOTABI M., JOVANOVIC M.R. Convergence and sample complexity of gradient methods for the model-free linear quadratic regulator problem // IEEE Transactions on Automatic Control. 2022. Vol. 67. Iss. 5. P. 2435–2450.
- [115] MOSTAFA E.-S.M.E. A conjugate gradient method for discrete-time output feedback control design // Journal of Computational Mathematics. 2012. Vol. 30. No. 3. P. 279–297.
- [116] NECOARA I., NESTEROV YU., GLINEUR F. Linear convergence of first order methods for non-strongly convex optimization // Mathematical Programming. Series A. 2019. Vol. 175. No. 1–2. P. 69–107.
- [117] NESTEROV YU. Efficiency of Coordinate Descent Methods on Huge-Scale Optimization Problems // SIAM Journal on Optimization. 2012. Vol. 22. No. 2. P. 341–362.

- [118] NESTEROV YU., PROTASOV V.YU. Computing closest stable nonnegative matrices // SIAM Journal on Matrix Analysis and Applications. 2020. Vol. 41. Iss. 1. P. 1–28.
- [119] NEWMAN D.J. Location of the maximum on unimodal surfaces // Journal of the ACM (JACM). 1965. Vol. 12. No 3. P. 395–398.
- [120] NURMINSKII E.A. The quasigradient method for the solving of the nonlinear programming problems // Cybernetics. 1973. Vol. 9. No. 1. P. 145–150.
- [121] PANAGOPOULOS H., ÅSTRÖM K.J., HÄGGLUND T. Design of PID controllers based on constrained optimization // Proc. 1999 American Control Conference. San Diego, USA, June 2–4, 1999. P. 3858–3862.
- [122] PETERSEN I. A stabilization algorithm for a class of uncertain systems // Systems & Control Letters. 1987. Vol. 8. Iss. 4. P. 351–357.
- [123] PHAM D.T., SHOLEDOLU M. The bees algorithm with attraction to global best solutions // Proc. 5th IPROMS International Virtual Conference on Innovative Production Machines and Systems (IPROMS 2009). Cardiff, UK, July 6–17, 2009.
- [124] POLIQUIN R.A., ROCKAFELLAR R.T., THIBAULT L. Local differentiability of distance functions // Transactions of the American Mathematical Society. 2000. Vol. 353. No. 11. P. 5231–5249.
- [125] POLYAK B., TREMBA A. New versions of Newton method: step-size choice, convergence domain and under-determined equations // Optimization Methods and Software. 2020. Vol. 35. No. 6. P. 1272–1303.
- [126] RAUTERT T., SACHS E.W. Computational design of optimal output feedback controllers // SIAM Journal on Optimization. 1997. Vol. 7. Iss. 3. P. 837–852.
- [127] SCHUR J. Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind // Journal für die reine und angewandte Mathematik. 1918. Vol. 148. P. 122–145.
- [128] STURM J.F. Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones // Optimization Methods and Software. 1999. No. 11–12. P. 625–653. URL: <https://sedumi.ie.lehigh.edu>

- [129] SYRMONS V.L., ABDALLAH C.T., DORATO P., GRIGORIADIS K. Static output feedback: A survey // Automatica. 1997. Vol. 33. No. 2. P. 125–137.
- [130] TAPIA R.A. The Kantorovich theorem for Newton's method // The American Mathematical Monthly. 1971. Vol. 78. No. 4. P. 389–392.
- [131] TOIVONEN H.T. A globally convergent algorithm for the optimal constant output feedback problem // International Journal of Control. 1985. Vol. 41. Iss. 6. P. 1589–1599.
- [132] VELIOV V.M. On the convexity of integrals of multivalued mappings: application in control theory // Journal of Optimization Theory and Applications. 1987. Vol. 54. No. 3. P. 541–563.
- [133] VIAL J.-P. Strong and weak convexity of sets and functions // Mathematics of Operations Research. 1983. Vol. 8. No. 2. P. 231–259.
- [134] WOLFE P. Methods of nonlinear programming // Nonlinear Programming. 1967. P. 97–131.
- [135] ZIEGLER J.B., NICHOLS N.B. Optimum settings for automatic controllers // Transactions of the ASME. 1942. Vol. 64. P. 759–768.

Предметный указатель

Аппроксимация
Моро–Иосиды, 95

Барьер, 152

Вариационный принцип
Экланда, 99

Возмущение внешнее, 247, 351, 353
ограниченное, 179

Время
дискретное, 353, 380
непрерывное, 351

Гармонический сигнал, 358, 359

Гессиан, 72

Градиент
Гато, 331
Фреше, 332

Градиентный спуск, 47

Грамиан управляемости, 378

Двухмассовая система, 187, 198

Диаграммы Боде, 371

Дифференцируемость, 330
по Гато, 331
по Фреше, 332

Дополнение по Шуру, 160

Допустимые направления, 39

Задача
 H_∞ -оптимизации, 199, 203
робастный вариант, 217
безусловной оптимизации, 24
квадратичного
программирования, 119
линейно-квадратичная, 219, 222
линейного
программирования, 109
каноническая, 110
о линейно-квадратичном
регуляторе, 194
подавления возмущений, 246
полуопределенного
программирования (SDP), 157
разрешимости ЛМН, 157
условной оптимизации, 24
фильтрации, 275

Инфимальная конволюция, 99

Комбинация точек
аффинная, 386
выпуклая, 386
коническая, 386
линейная, 386

Конус, 387, 393
асимптотический, 396
Булигана, 411

- верхний касательный, 411
конечно порожденный, 394
многогранный, 394
нормальный, 393
полиэдральный, 394
проксимальных нормалей, 405
рецессивный, 396
Коэффициент усиления, 364, 366, 371, 382
Критерий
качества квадратичный, 193
наблюдаемости, 384
управляемости
дискретной системы, 379
непрерывной системы, 376
устойчивости
дискретной системы, 181
непрерывной системы, 179
Лемма
Беллмана, 194
Питтерсена, 208, 209, 211, 215, 218
для строгого неравенства, 165
Шура, 160
Линейное матричное
неравенство (ЛМН), 153
допустимая область, 155
каноническая форма, 153
матричная форма, 154
Луч крайний, 395
MATLAB, 187
cvx, 187
SeDuMi, 187
YALMIP, 187
Матрица
выхода, 352
Гессе, 333
гурвицева, 342
замкнутой системы, 381
знакоопределенная, 338
наблюдаемости, 383
ранг, 336, 342, 345, 378
регулятора, 381
сингулярное число, 350
след, 337
собственные значения, 337, 342, 345
состояния, 352
степень устойчивости, 342, 345
управляемости, 375
усиления, 381
фробениусова форма, 370
характеристический
полином, 354
шуровская, 344
Якоби, 72
Матричная
замена переменных, 183, 192, 196
неопределенность, 207
пара
наблюдаемая, 384
управляемая, 376
Мелкость, 28
Метод
градиентного спуска, 47
золотого сечения, 33
имитации отжига, 43
Нелдера–Мида, 45
Ньютона, 73
демпифицированный, 79
комбинированный, 80
нулевого порядка, 28
отсекающей гиперплоскости, 143
покоординатного спуска, 39

- случайного, 41
 полной релаксации, 39
 проекции
 градиента, 122
 субградиента, 134
 симплекс-метод, 116
 случайного блуждания, 39,
 41
 сопряженных градиентов, 60
 тяжелого шарика, 56
 условного градиента, 135
 штрафных функций, 150
 эллипсоидов, 143
 инвариантных, 247
 Метрика Хаусдорфа, 409
 Минимизирующая
 последовательность, 26
 Минимум
 глобальный, 24
 локальный, 24, 42
 строгий, 24
 Многообразие
 Грассмана, 350, 407
 Штифеля, 350, 407
 Множество
 выпуклое, 32, 385
 компактное, 328
 крайних лучей, 395
 крайних точек, 395
 Лебега, 32
 линейно связное, 388
 прокс-регулярное, 405
 проксимально гладкое, 405
 равномерно выпуклое, 105,
 141
 связное, 221
 сильно выпуклое, 404
 слабо выпуклое, 405
 строго выпуклое, 385
 уровня, 223
 эффективное, 386
 Невырожденная вершина, 117
 Неравенство
 Йенсена, 387
 Ляпунова, 154, 178, 343
 двойственное, 343, 346
 дискретное, 181, 346
 Риккати, 161
 Юнга, 175
 Норма
 матричная, 339
 индуцированная, 341
 подчиненная, 341
 спектральная (2-норма),
 339, 364
 столбцовая (∞ -норма),
 340
 строчная (1-норма), 340
 фробениусова, 340, 350
 передаточной функции
 H_2 -норма, 366
 H_∞ -норма, 200, 364–366
 сигнала
 L_1 , 362
 L_2 , 360
 L_∞ , 361
 l_1 , 363
 l_2 , 362
 l_∞ , 363
 Нормаль проксимальная, 405
 Оболочка множества
 аффинная, 387
 выпуклая, 387
 линейная, 387
 Обратная связь, 381
 по выходу, 222, 254, 382
 по оценке состояния, 383
 по состоянию, 352, 381
 Одномерная минимизация, 28

- Описание системы
в пространстве состояний,
352, 370
входо-выходное, **367**
с помощью передаточной
функции, **355**
- Опорный принцип
для проксимально гладких
множеств, **406**
для сильно выпуклых
множеств, **404**
- Отображение
 α -накрывающее, **416**
 β -липшицевое, **416**
- ПИД-регулятор, **300, 301, 382**
настройка, **383**
- Полином
аннулирующий, **337**
минимальный, **337**
характеристический, **337**
- Постоянная времени, **371**
- Правило
Армихо, **58**
Ферма, **419**
- Принцип скимающих
отображений, **415**
- Проекция метрическая, **390**
- Производная, **330**
односторонняя, **332**
по направлению, **331**
- Пространство
последовательностей
абсолютно суммируемых,
363
суммируемых с
квадратом, **362**
существенно
ограниченных, **363**
сигналов
- ограниченной
интенсивности, **361**
ограниченной энергии, **360**
- сигналов дискретных
 l_1 , **363**
 l_2 , **362**
 l_∞ , **363**
- сигналов непрерывных
 L_1 , **362**
 L_2 , **360**
 L_∞ , **361**
- состояний, **352**
управлений, **352**
- функций
абсолютно
интегрируемых, **362**
интегрируемых с
квадратом, **360**
существенно
ограниченных, **361**
- Радиус
квадратичной
стабилизируемости, **212**
устойчивости, **209**
спектральный, **344**
- Распределение
нормальное, **40**
равномерное, **40**
- Расстояние от точки до
множества, **390**
- Регулятор
линейно-квадратичный, **193**
весовые коэффициенты,
194
робастный, **214**
линейный статический, **382**
- Рекордное значение, **28**
- Свойство Поляка–Лоясевича,
51

- Сетка равномерная, 28
Система линейная
 блок-схема, 357
 выход, 351
 наблюдаемый, 352
 регулируемый, 352
 двойственная, 385
 дискретная, 353
 замкнутая, 382
 устойчивая, 180
 характеристический полином, 360
 каноническая управляемая форма, 370, 376
 наблюдаемая, 383
 ненаблюдаемая, 383
 непрерывная, 351
 асимптотически устойчивая, 177
 возмущенная, 179, 181
 замкнутая, 352, 381
 управления, 351
 устойчивая, 177
 характеристический полином, 354, 356, 369
 нестационарная, 351
 полностью определенная, 351
 с неопределенностью, 351
 состояние, 351, 353
 стационарная, 351
 управляемая, 375
 частотная характеристика, 359
Скорость сходимости
 квадратичная, 27
 линейная, 27
 сублинейная, 27
Сопряженные векторы, 63
Стабилизация, 177
квадратичная, 182, 193
робастная, 210
Степень устойчивости
 матрицы, 342
 системы, 179
Субградиент, 397
Субдифференциал, 397
 Фреше, 408
Сумма Минковского, 388
Супердифференциал, 397
 Фреше, 408
Сходимость
 по точке, 27
 по функции, 27
 со скоростью
 геометрической прогрессии, 27
Теорема
 Вейерштрасса, 329, 330
 Гамильтона–Кэли, 337, 368
 Грюнбаума–Хаммера, 144
 Канторовича, 75
 Каратеодори, 388
 Кли, 396
 Куна–Таккера, 422
 Люка, 92
 Моро–Рокафеллара, 399
 о неявной функции, 335
 о субдифференциале максимума, 401
 об обратном отображении, 335
 об отделимости, 390
 об очистке, 401
 Радемахера, 333
 Хофмана, 115
Точка
 крайняя, 395
 неподвижная, 415

- совпадения отображений, 416
стационарная, 133, 420
- Управление, 351, 353
комбинированное, 381
программное, 374
робастное, 207
- Управляемость, 375
- Уравнение
выхода, 351
Ляпунова, 342
дискретное, 345
- Риккати, 194
- состояния, 351
- Условие
выживания, 412
Лежанского–Поляка–
Лоясевича (LPL), 51
Липшица, 333
начальное, 352
ограничения ошибки, 91
опорное сильной
выпуклости, 129
острого минимума, 70
Поляка–Лоясевича, 51
реализуемости, 356, 369
- Установившееся значение
выхода, 358
состояния, 358
- Устойчивость, 176
по входу, 179
по начальному
приближению, 179
- Фильтрация
дискретное время, 277
непрерывное время, 291
- Функционал энергии, 374
- Функция
 L -гладкая, 235
- барьерная, 152
вогнутая, 386
выпуклая, 386
квазивыпуклая, 31, 89
коэрцитивная, 37, 230
Лагранжа, 420
липшицевая, 29, 333
модельная, 26
надграфик, 386
непрерывная, 329
в точке, 329
по Липшицу, 333
передаточная, 354
дискретной системы, 360
матричная, 356
непрерывной системы, 354
от возмущения к выходу,
354
от управления к выходу,
354
полюс, 355, 356
скалярная, 369
устойчивая, 355
положительной срезки, 87
полунепрерывная, 330
предел
верхний, 329
нижний, 329
сильно выпуклая, 401
сильно квазивыпуклая, 89
слабо вогнутая, 407
слабо выпуклая, 407
собственная, 386
строго выпуклая, 386
штрафная, 150
с квадратичным
штрафом, 150
- Функция Ляпунова, 177, 180
квадратичная, 178, 181, 182
метод, 177

- общая квадратичная, 209
- Характеристика
логарифмическая
амплитудная, 372
фазовая, 373
- частотная
амплитудная, 371
фазовая, 371
- Частота среза, 372
- Частотные методы, 360
- Штраф
внешний, 152
внутренний, 152
- Штрафной член, 150
- Шум, 351, 353
- Эллипсоид
инвариантный, 178, 248
ограничивающий, 249
- Эффект всплеска, 190