А. С. НЕМИРОВСКИЙ
Д. Б. ЮДИН

СЛОЖНОСТЬ ЗАДАЧ И ЭФФЕКТИВНОСТЬ МЕТОДОВ ОПТИМИЗАЦИИ

ТЕОРИЯ И МЕТОДЫ СИСТЕМНОГО АНАЛИЗА

МОСКВА «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1979

А. С. НЕМИРОВСКИЙ, Д. Б. ЮДИН

СЛОЖНОСТЬ ЗАДАЧ И ЭФФЕКТИВНОСТЬ МЕТОДОВ ОПТИМИЗАЦИИ

МОСКВА «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1979

32.81
Н 50
УДК 62.50

Сложность задач и эффективность методов оптимизации. Немировский А. С., Юдин Д. Б. — М.: Наука. Главная редакция физико-математической литературы, 1979, 384 с.

Монография посвящена исследованию круга вопросов, относящихся к сложности задач и трудоемкости методов математического программирования. В книге рассматриваются теоретические потенциальные нижние границы трудоемкости численных методов решения экстремальных задач стандартных классов (гладких, негладких выпуклых, сильно выпуклых и гладких выпуклых, выпуклых стохастических) при различных предположениях о типе и качестве информации о задаче, доступной методу на каждом шаге. Предложены методы, в существенном реализующие эти потенциальные границы.

Монография рассчитана на специалистов, занимающихся теорией и приложениями численных методов оптимизации, в том числе на разработчиков алгоритмов для АСУ, и на студентов и аспирантов — математиков и вычислителей.

Илл. 4, библ. 34.

Н $\frac{30501-174}{053(02)-79}$ 185-80. 1502000000

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Численным методам решения экстремальных задач посвящена обширная литература. Центральное место в ней занимают исследования, связанные с конкретными алгоритмами оптимизации — их описанием, областями применимости, скоростью сходимости. Имеется также ряд работ, посвященных эмпирическому — на примерах — сравнению алгоритмов по эффективности с целью выбора «наилучшего» метода среди наличествующего их арсенала.

В гораздо меньшей степени изучен вопрос о том, чего вообще можно ожидать от численных методов решения задач данного типа, вопрос о потенциальных границах их возможностей, о том, насколько сложны задачи того или иного сорта не по отношению к данному конкретному, а ко «всем вообще» методам решения. Именно эта естественная проблема «потенциально достижимой эффективности численных методов на задачах данного типа» является объектом изучения в предлагаемой вниманию читателя монографии.

Типичный вопрос, который мы будем рассматривать, выглядит следующим образом. Дано семейство оптимизационных задач вместе с доступным для методов источником информации о каждой решаемой задаче из этого семейства. Каковы потенциальные нижние границы трудоемкости всевозможных методов, решающих все задачи семейства с заданной точностью? Какой метод реализует эту потенциальную границу и тем самым является наилучшим? Ясно, что точная постановка такого вопроса требует формализации понятий «метод», «трудоемкость метода» и т. п. Мы фиксируем определенную формализацию такого рода (по нашему мнению, наиболее удобную при изучении «непрерывных» задач математического программирования, которыми мы и будем заниматься) и исследуем затем поставленный выше вопрос применительно к серии стандартных семейств задач нелинейного программирования: гладких многоэкстремальных, «всех выпуклых», сильно выпуклых и т. д. В большинстве случаев на интересующий нас вопрос удается получить в достаточной мере законченный ответ.

Мы не претендуем на то, что принятая нами формальная постановка проблемы выбора «наилучшего» численного метода оптимизации вполне адекватна исходной содержательной постановке. В данный момент нет необходимости останавливаться на этой стороне дела более подробно; содержательная мотивировка принятого нами подхода, обсуждение его достоинств и вместе с тем ограниченности проводится в § 1 гл. I. Отметим лишь, что, по нашему мнению, используемая нами формализация позволяет получить достаточно полезную, хотя и грубую — «порядковую» — информацию о потенциальных возможностях численных методов решения экстремальных задач стандартных классов. Дело читателя — разделить или отвергнуть это мнение.

К вопросу об отыскании потенциальной нижней границы трудоемкости методов, решающих задачи данного класса с данной погрешностью, т. е. в определенном смысле — к задаче вычисления «сложности» класса, тесно примыкает проблема оценки эффективности тех или иных методов решения задач этого класса. Эту эффективность естественно определить как обратное отношение трудоемкости рассматриваемого метода к «эталонной трудоемкости» — к сложности класса. Эффективность метода показывает, в какой мере он может быть улучшен по трудоемкости, насколько он неоптимален.

Задача оценки эффективности традиционных численных методов оптимизации нами затрагивается лишь в небольшой своей части. Это и понятно — для ее решения недостаточно располагать эталоном (на получении которого в основном и сосредоточены наши усилия); надо еще иметь оценки трудоемкости и погрешности интересующего нас метода на рассматриваемом классе задач. Обилие стандартных численных методов исключает, по-видимому, возможность усилиями авторов оценить эти характеристики для сколь-нибудь представительной группы методов.

В книге оценивается эффективность нескольких наиболее естественных и простых методов. По причинам, которые станут ясны ниже, все эти методы — методы выпуклого программирования. Среди методов негладкой выпуклой оптимизации оцениваются градиентный метод и метод Келли (пожалуй, ими и исчерпывается список традиционных алгоритмов негладкой выпуклой оптимизации). Обширное поле алгоритмов минимизации гладких и (сильно) выпуклых задач исследовано на эффективность несравненно хуже — здесь мы ограничились градиентным методом с минимизацией в направлении антиградиента и несколькими простыми версиями методов сопряженных направлений. Оказывается, что рассмотренные методы сильно выпуклого программирования не являются эффективными — они излишне чувствительны к степени обусловленности решаемой задачи, и их эффективность стремится к нулю при ухудшении обусловленности. Отметим, что такого рода негативные результаты позволяют сделать некоторые выводы и в отношении эффективности ряда традиционных методов, в явном виде в работе не рассматриваемых.

Приведем пример. Хорошо известно обширное семейство методов возможных направлений решения условных выпуклых задач. Скорость сходимости этих методов обычно оценивается в предположении сильной выпуклости решаемой задачи. Таким образом, эти методы естественно рассматривать на классе сильно выпуклых задач. В главе VII показано, что сложность этого класса по существу определяется *лишь* требуемой точностью, обусловленностью задачи и ее размерностью, но не числом ограничений. С другой стороны, если ограничения отсутствуют, то большая часть версий метода возможных направлений превращается в градиентный спуск с минимизацией в направлении спуска. Стало быть, эффективность методов возможных направлений не может быть существенно более высокой, чем у указанного градиентного метода, и потому также стремится к нулю с ухудшением обусловленности задачи.

Ограниченный объем книги не позволил нам останавливаться на такого рода следствиях; они, безусловно, будут читателю самоочевидны. Отметим еще, что высказанное выше суждение о неэффективности метода возможных направлений (как и другие высказывания такого рода в основном тексте) есть суждения «в рамках принятого определения трудоемкости», как указывалось, не вполне адекватного интуитивно понимаемой вычислительной сложности метода. Поэтому его не следует интерпретировать как призыв к безоговорочной дискриминации соответствующих методов; категорические приговоры такого рода вряд ли допустимы вообще.

Отметим некоторые отличия принятых в монографии подходов и полученных результатов от традиционных установок в теории оптимизации.

Традиционный подход к оценкам скорости сходимости численных методов оптимизации носит большей частью асимптотический характер — выясняется тип асимптотики трудоемкости метода по требуемой точности. Вопрос о том, когда происходит «выход» на эту асимптотику, изучается сравнительно редко, как, впрочем, и важный вопрос о влиянии на трудоемкость других, отличных от точности, параметров класса задач (например, размерности решаемых задач).

В то же время все приводимые в монографии оценки трудоемкости численных методов оптимизации (как, впрочем, и большая часть оценок сложности) имеют неасимптотический характер. В книге явно выписываются верхние оценки трудоемкости рассматриваемых методов в функции от требуемой точности (измеряемой разумным образом) и от параметров, выделяющих класс подлежащих решению задач (таких как размерность задачи, число ограничений, характеристики геометрии области определения задачи и т. п.).

В литературе по методам оптимизации скорость сходимости (именно скорость, а не сам факт сходимости) устанавливается обычно лишь применительно к классам достаточно «хороших» задач. В предлагаемой монографии методы оптимизации изучаются, как правило, на более широких классах задач. В большей части книги от задач, *подлежащих решению, ничего не требуется сверх* выпуклости, разве что ограниченности области определения задачи и непрерывности (иногда — липшицевости) фигурирующих в ней функционалов. Даже рассматриваемые в главе VII классы сильно выпуклых задач все еще несколько шире классов «хороших» задач, на которых традиционно принято оценивать методы оптимизации. Разумеется, оценка методов по их поведению на более широких, чем обычно, семействах задач приводит к ухудшению потенциально достижимых гарантий их работы. Оказывается, однако, что во многих случаях ухудшение это не слишком значительно и представляется приемлемой платой за расширение запаса задач, на которые распространяются новые гарантии.

Обрисовав в общих чертах цели монографии, охарактеризуем вкратце содержание работы по разделам (более подробная сводка результатов дана в § 1 гл. I). Первая ее глава носит вводный характер: здесь формируется язык, на котором далее ставятся и решаются проблемы потенциальной эффективности численных методов. Разделение последующего материала на главы диктуется необходимостью по отдельности рассматривать различные классы экстремальных задач. В последнем параграфе гл. I рассматриваются гладкие (не обязательно выпуклые) задачи. Помещенные здесь частью известные, частью новые результаты носят негативный характер (оказывается, что этот класс в общем случае безнадежно сложен для решения). Выпуклое программирование, которому в книге уделено основное внимание, являет собой гораздо более оптимистическую картину. В главах II, III, IV рассматриваются классы «всех вообще», в том числе и негладких, выпуклых задач, решаемых методами первого порядка при точном вычислении значений и производных компонент задачи. В главах V и VI изучаются выпуклые задачи, значения и производные компонент которых наблюдаются в смеси с шумом — задачи выпуклого стохастического программирования. В главе VII рассматриваются классы гладких выпуклых и сильно выпуклых задач. В главе IX речь идет о выпуклых задачах, решаемых методами нулевого порядка (оперирующими лишь со значениями, но не с производными компонент задачи). Несколько особняком стоит глава VIII, посвященная оценке трудоемкости нескольких популярных методов решения строго выпуклых задач. Приложение к работе содержит сводку ряда классических математических понятий и теорем, используемых в книге и не всегда знакомых прикладникам. Вообще следует заметить, что по характеру изложения книга доступна читателю, имеющему стандартную для занимающихся численными методами оптимизации математическую подготовку. Более сложный аппарат привлекается лишь во второстепенных «формальных тонкостях» отдельных доказательств.

Изложение материала, связанного с конкретным классом экстремальных задач, включает обычно следующие основные этапы: (1) описание рассматриваемого класса задач; (2) описание некоторых методов решения задач данного класса и оценку сверху их трудоемкости и погрешности; (3) оценку снизу потенциально (по всем мыслимым методам решения задач данного класса) достижимой трудоемкости этих методов при заданной их погрешности. Как правило, выбор конкретных методов в (2) приводит к оценкам трудоемкости, в существенном совпадающим с потенциальными нижними оценками из (3). В результате, с одной стороны, мы получаем достаточно полное представление об «объективной сложности» рассматриваемого класса задач, а с другой — основания для теоретических рекомендаций по применению методов из (2), как существенно неулучшаемых по трудоемкости. Отметим, что приводимые нами «субоптимальные» методы в ряде случаев являются в существенной мере новыми.

Укажем, что при желании читатель может независимо знакомиться с разделами, посвященными тем или иным конкретным классам задач.

По поводу характера изложения стоит заметить следующее. Мы старались по возможности ясно выделять идеи, лежащие в основе предлагаемых конструкций, и четко описывать изучаемые численные методы. Формальные доказательства локализованы; часть их вынесена в специальные параграфы. При первом чтении доказательства при желании можно опускать, что не помешает, скажем, применять описываемые методы, но затруднит детальное понимание их механизма.

Существенная часть результатов оформлена в виде упражнений, как правило, предлагающих читателю доказать некоторое тут же формулируемое утверждение. Если речь идет о не вполне тривиальном факте, то часто вслед за формулировкой в угловых скобках $\langle\ldots\rangle$ приводится и его доказательство. Подчеркнем, что от читателя требуется ознакомление с формулировками излагаемых в упражнениях предложений независимо от того, будет ли он выполнять сами упражнения.

О терминологии. Наряду со стандартными терминами, употребляемыми без специальных пояснений, нам приходится широко пользоваться рядом специфических понятий и обозначений. При первом столкновении с нестандартными и некомментируемыми обозначениями или терминами читателю следует обратиться к списку обозначений в конце книги или предметному указателю, где он найдет ссылку на раздел, в котором описывается соответствующий объект. Исключение составляют некоторые второстепенные понятия и обозначения, «область действий» которых не превышает одной главы. Соответственно при ознакомлении с каким-нибудь разделом данной главы, вообще говоря, требуется знакомство со всеми предыдущими ее разделами.

Несколько *слов о списке литературы и ссылках*. В список литературы вынесены лишь те работы, на которые имеются прямые ссылки из основного текста; этот список невелик и ни в коей мере не претендует на сколько-нибудь полное перечисление работ по теме монографии. Используя чужие результаты, авторы старались упоминать об этом, указывая источник информации (но не ставя себе целью указать непременно первоисточник); говоря о результатах, так сказать, ставших хрестоматийными, мы заменяли прямые ссылки фразами типа «хорошо известно, что...». Не исключено, что те или иные результаты, рассматриваемые авторами как оригинальные, на самом деле являются переоткрытиями уже известных фактов (с этой возможностью связана оговорка «старались» в предыдущей фразе). В этом случае мы заранее приносим первооткрывателям свои извинения.

В заключение считаем своим приятным долгом поблагодарить Е. Г. Гольштейна и Б. Т. Поляка за стимулирующие обсуждения результатов работы.

А. С. Немировский, Д. Б. Юдин

Июнь 1978 г.

Глава I

ВВЕДЕНИЕ

§ 1. Постановка проблемы оценки сложности задач оптимизации и основные результаты работы: неформальное описание

В этом параграфе неформально описан круг вопросов, которыми мы займемся, и направление наших исследований. Цель описания — мотивировка естественности принятого в работе подхода к оценке сложности задач и эффективности методов оптимизации.

1.1. Мы будем изучать возможности численных методов решения задач математического программирования. Едва ли надо говорить о роли таких методов в приложениях математики к практическим задачам. Расширение приложений и совершенствование вычислительных средств приводят к резкому усложнению задач оптимизации, которые приходится «доводить до ответа», и к непрерывному пополнению арсенала используемых для этого методов. В этой ситуации закономерно возникает тенденция к более пристальному теоретическому исследованию самих методов. По современным стандартам одно только обоснование сходимости еще не дает методу права на существование. «Хороший тон» требует вдобавок оценить его трудоемкость.

На следующем шаге возникает желание найти потенциально достижимые границы трудоемкости решения задач данного типа и построить методы, реализующие эти границы, методы, в некотором смысле оптимальные — обеспечивающие требуемое качество решения всех рассматриваемых задач при минимально возможной трудоемкости. Именно этим проблемам и посвящена настоящая работа. Наша цель — теоретический анализ потенциальных возможностей численных методов.

Строгие формулировки возникающих в этой связи проблем требуют формализации таких понятий, как «класс задач данного типа», «метод решения задач данного класса», «трудоемкость метода» и т. д. Имеется много естественных в тех или иных отношениях формализаций такого рода.

1.2. Наиболее общий подход к анализу вычислений развивается в рамках общей теории алгоритмов и связан с работами [10], [12]. Для ознакомления с духом подхода можно рекомендовать небольшую, но чрезвычайно глубокую работу Левина [16]. Суть этого направления, в неформальном и самом общем виде, такова. Принимается, что вычислительный метод есть алгоритм (в любом из формальных определений, принятых в математической логике, например машина Тьюринга). Каждый такой алгоритм, как известно, перерабатывает слова в некотором алфавите (например, двоичном) в некоторые другие слова. Чтобы применить алгоритм к какой-нибудь задаче, сформулированной, возможно, на неформализованном (в смысле математической логики) языке, скажем на языке классической математики, надо сопоставить ей слово во входном алфавите алгоритма — код этой задачи. Алгоритм перерабатывает код задачи в выходное слово, интерпретируемое как код ее решения. Этот код по определенным правилам транслируется на исходный язык описания задачи. Трудоемкость решения задачи измеряется числом тактов работы алгоритма на ее коде.

Могут быть рассмотрены и другие меры трудоемкости, связанные с расходом памяти и т. п. На таком пути получен ряд чрезвычайно общих и глубоких результатов. Нет сомнения, что разрабатываемый в общей теории алгоритмов аппарат наиболее приспособлен для изучения общей проблемы вычислений. Вместе с тем ясно, что при указанном подходе проблема перевода задачи с языка (как выражаются логики — жаргона) классической математики на формализованный язык — проблема кодировки задачи — остается вне теории. Для задач с четкой алгебраической структурой (например, для задач целочисленного программирования, к которым, в частности, сводятся комбинаторные задачи) эта проблема легко решается: способы естественной кодировки здесь очевидны. Хотя естественных кодировок много, но все они перерабатываются друг в друга простыми алгоритмами, так что результаты по существу не зависят от способа кодировки.

1.3. Ситуация меняется, если мы переходим от дискретных задач (или задач простой алгебраической структуры) к непрерывным задачам типа интегрирования или минимизации «любой» (скажем, гладкой) функции, определенной, к примеру, на отрезке действительной оси. Проблема кодировки здесь уже не столь проста. Конечно, при практическом решении любая задача с необходимостью кодируется (в наших примерах — алгоритмом, вычисляющим значения функции в «любой» точке) и дискретизуется, так что формально речь всегда идет о решении дискретных задач, а метод решения есть некоторый алгоритм переработки слов. Однако при таком формализованном (в смысле математической логики) подходе с трудом просматриваются «непрерывные», «классические» свойства задачи (в наших примерах — свойства гладкости, выпуклости и т. п.), а описание (и тем более оптимизация) методов становится чрезвычайно, если не безнадежно, сложным делом.

Естественно описывать методы решения «непрерывных» задач на обычном — классическом языке и не задумываться при этом над тем, что умножение действительных чисел или вычисление синуса нельзя, строго говоря, реализовать алгоритмом. Но даже и описание методов решения неалгебраических задач в терминах идеализированной вычислительной машины, выполняющей (точно) арифметические действия, сравнения, вычисления стандартных функций и т. п., все еще «слишком детально». Представляется, что изучение в этих терминах, скажем, метода минимизации Ньютона — Рафсона функции многих переменных «потонет» в выяснении вопроса о минимальном числе действий, необходимых для обращения матрицы (этот вопрос, кстати, до сих пор не решен). Построенная в таком стиле теория численных методов решения задач определенного (но сколько-нибудь широкого) класса (скажем, экстремальных задач) обязана была бы, грубо говоря, включать в себя теорию «всех вообще» численных методов. Но тогда она не была бы приспособлена к специфике именно данного класса задач.

Таким образом, наряду с наиболее общей, универсальной теорией вычислений должны существовать и специальные исследования, связанные со специальными классами вычислительных задач. Язык этих исследований (т. е. в первую очередь формализация понятий «задача» и «метод»), не претендуя на универсальность, должен быть приспособлен к специфике рассматриваемого класса. Исследования такого типа позволяют получить существенную (хотя и специальную) информацию, которую вряд ли можно извлечь средствами универсальной теории.

В разнообразии подходов нет ничего удивительного — вряд ли можно рассчитывать на одну-единственную удовлетворительную во всех без исключения отношениях формализацию понятий, связанных с анализом численных методов.

1.4. Подход, принятый в этой книге (по нашему мнению — наиболее удобный для изучения методов «непрерывной оптимизации»), восходит к работам Н. С. Бахвалова [2]. В соответствии с этим подходом численный метод есть набор правил накопления информации о решаемой задаче для последующего формирования ее решения. Источник этой информации фиксируется заранее. Процесс получения информации управляется методом в следующем смысле: заранее фиксировано множество «вопросов» X, которые можно задавать источнику информации — оракулу. Работа метода происходит по шагам. На i-м шаге можно задать один вопрос $x_i \in X$ оракулу и получить на него ответ — точку некоторого множества I (информационного пространства — множества возможных ответов оракула). Очередные вопросы можно свободно выбирать в пределах X. После некоторого числа таких вопросов на основе накопленной информации (набора ответов оракула) формируется результат применения метода к задаче. Метод решения представляет собой, таким образом, набор правил формирования очередных вопросов, момента остановки и результата в функции от накопленной к данному моменту информации. На эти правила не налагается никаких ограничений типа вычислимости. Они могут быть любыми *) функциями соответствующих аргументов.

Трудоемкость решения данной задачи данным методом определяется как число шагов его работы на ней. Помимо трудоемкости, метод на данной задаче характеризуется своей погрешностью (мерой неточности результата его применения к задаче в качестве приближенного решения последней).

Заметим, что при описанном подходе ограничения на возможности методов носят чисто информационный характер: метод «не знает» заранее, какую именно задачу ему приходится решать. По существу его цель как раз и состоит в «опознании» задачи до такой степени, при которой уже возможно формирование решения требуемой точности.

Мы видим, что если в общей теории вычислений полная информация о задаче (ее код) является входом метода и его возможности ограничены лишь средствами, которые разрешается применять для ее переработки в код ответа (метод должен быть алгоритмом в смысле математической логики), то при описываемом подходе картина обратная: применяемые средства никак не ограничены, зато исходная информация о задаче неполна, и за ее накопление приходится платить.

Ясно теперь, что бессмысленно ставить вопрос о выборе наилучшего метода решения данной вполне конкретной задачи. Этот метод состоит просто в выписывании ее решения и имеет трудоемкость 0. Не надо думать, что это замечание доказывает бессмысленность намеченного подхода. Указанное явление не содержит никакой патологии. Если надо решить одну-единственную задачу, то не возникает и вопроса о методе. По самому смыслу слова понятие «метод» связано не с индивидуальной, а с массовой задачей — классом задач данного типа. Таким образом, речь должна идти не о «методе решения конкретной задачи», а о «методе решения задач данного класса» (т. е. задач из некоторого заранее фиксированного множества).

*) С точностью до практически совершенно необременительных формальных оговорок.

Заметим, кстати, что вообще вряд ли можно разумно поставить вопрос о «наилучшем по трудоемкости» методе решения индивидуальной задачи, как бы ни уточнять понятие «метод». Так, в рамках общей теории алгоритмов такая «минимальная» трудоемкость есть по существу длина слова, кодирующего ответ. Смысла в таком результате не многим больше, чем в утверждении, что эта трудоемкость равна 0.

1.5. Итак, изучение численных методов в рамках описываемого подхода предполагает фиксацию следующих объектов:

— класса задач, подлежащих решению рассматриваемыми методами (для нас это всегда будут классы экстремальных задач);

— источника информации о решаемой задаче (оракула);

— способа определения погрешности результата в качестве приближенного решения рассматриваемой задачи.

Зафиксировав эти объекты, мы четко выделяем класс отвечающих им методов и для каждого метода определяем характеристики — трудоемкость и погрешность — на каждой задаче класса. Этого еще недостаточно для того, чтобы ставить задачу оптимизации методов. Действительно, из двух методов, вообще говоря, первый на одних задачах будет хуже, на других — лучше второго, так что надо еще указать способ сравнения методов друг с другом по их свойствам на всем классе задач, а не на отдельных его представителях Для этой цели мы определяем характеристики метода на классе (из минимаксных соображений).

Обсуждение правомерности минимаксного подхода уместно отложить до п. 4.5.

Уточнение понятия «характеристики метода на классе» и позволяет четко поставить задачу изучения потенциальных возможностей методов решения задач данного класса.

1.6. Обсудим принятый нами подход. Он универсален в том смысле, что надлежащая фиксация понятий «класс», «оракул» и т. п. позволяет применить его к любому классу задач. Однако эта универсальность — лишь кажущаяся. Дело в том, что определение трудоемкости, которое ему отвечает, вовсе не адекватно интуитивно понимаемому понятию «сложности решения». Наше определение — чисто информационное, никак не учитывающее затрат вычислительных средств на обработку полученной информации.

Если, скажем, можно за небольшое число k вопросов однозначно выделить задачу из класса (что типично для «алгебраических» задач), то оптимальная трудоемкость решения задач этого класса мала — она не больше k. В такой ситуации наш подход может давать бессодержательные результаты. Тем не менее ясно, что трудоемкость метода в принятом определении уже заведомо оценивает снизу его «реальную» трудоемкость. В ряде случаев эти оценки полезны. Да и «оптимальные» в нашем смысле методы достаточно часто оказываются, по-видимому, вполне приемлемыми и с точки зрения практических соображений. Короче говоря, принятый подход никоим образом не является панацеей от всех трудностей и не дает исчерпывающего анализа всей совокупности характеристик вычислительных методов (например, важные вопросы вычислительной устойчивости вообще выходят за его рамки). Однако мы считаем (и надеемся, что содержание книги подтверждает это мнение *)), что результаты, которые доставляет его применение, дают достаточно много полезной информации о численных методах оптимизации. Как и всякое другое средство, этот подход эффективен — если, конечно, его разумно употреблять и не требовать от него слишком многого.

Другой «ограничивающий момент» состоит в том, что движение по намеченному пути требует привлечения «новой сущности» — оракула (эта проблема в общей теории вычислений возникает только в специальных случаях). Соответствующее ограничение не очень важно с точки зрения возможных применений. Оракул для решения «непрерывных» задач оптимизации — задач математического программирования,— по счастью, традиционен (он вычисляет значения (или значения и производные) функционалов задачи в указанной ему точке). Тем не менее авторы не понимают, по каким причинам надо ограничиваться именно этим оракулом. Мы можем мотивировать употребление этого оракула во всех конкретных результатах книги лишь ссылкой на упомянутые традиции. Однако распространенность порока не может его оправдать.

1.7. Приведем теперь краткую сводку результатов работы. Как уже говорилось, мы рассматриваем конкретные классы задач (вместе с фиксированными оракулом и способом измерения погрешности приближенных решений) и отвечающий этим объектам класс методов. Каждый из этих методов на рассматриваемом классе задач характеризуется своими трудоемкостью и погрешностью — верхними гранями (по задачам класса) числа шагов, соответственно, погрешности результата, своей работы на задаче. Мы определяем, далее, сложность $N(\nu)$ данного класса задач как функцию погрешности ν, равную минимально возможной трудоемкости метода, решающего всякую задачу класса с погрешностью, не превосходящей ν. Наша цель, во-первых, оценить «потенциальную нижнюю границу трудоемкости» $N(\nu)$ и, во-вторых, построить методы, «в существенном» реализующие эту границу. Прежде чем перейти к изложению полученных в этом направлении результатов, отметим, что погрешность, фигурирующая далее, есть должным образом определяемая «относительная» погрешность. Нетривиален только случай $\nu < 1$, о котором далее и ведется речь.

*) Впрочем, что же нам еще остается.

Первые (пессимистические) результаты связаны с классами гладких (многоэкстремальных) задач. Оказывается (§ 6 гл. I), что для сложности $N(\nu)$ класса всех экстремальных задач с k раз непрерывно-дифференцируемыми функционалами на компактном теле G в E^n справедлива нижняя оценка $N(\nu) \geqslant c(k, G)(1/\nu)^{n/k}$, как для обычных (детерминированных) методов решения, так и для методов случайного поиска. Катастрофический рост $N(\nu)$ при $\nu \to 0$ и особенно при $n \to \infty$ показывает, что бессмысленно ставить вопрос о построении универсальных методов решения «всех вообще» гладких задач сколько-нибудь заметной размерности. Интересно отметить, что практически те же оценки имеют место даже в случае безусловных задач, порожденных одноэкстремальными (но не выпуклыми) функциями.

Неприличное поведение сложности классов невыпуклых задач заставляет нас в дальнейшем сосредоточить внимание на гораздо более благодарном объекте — выпуклом программировании.

си Мы начинаем с классов «всех вообще» выпуклых экстремальных задач на выпуклом компакте G размерности n; оракул поставляет значения и субградиенты компонент задачи. Оказывается, что сложность этого класса допускает оценку

$$c_1 \leqslant \frac{N(\nu)}{1 + n \ln \frac{1}{\nu}} \leqslant c_2$$

(здесь и далее c_i — положительные абсолютные константы). Правое неравенство верно при всех $\nu < 1$, а левое — в асимптотике по $\nu \to 0$ (главы II, IV). «Момент установления» асимптотики зависит от аффинных свойств G. Эта зависимость является предметом специального исследования, о результатах которого речь пойдет ниже. Метод погрешности ν и трудоемкости $O(n \ln(1/\nu))$, «в существенном реализующий» сложность класса при $\nu \to 0$ (идея его высказана А. Ю. Левиным [15]), к сожалению, практически не может быть использован при $n > 3$ из-за чрезвычайной сложности шага. Мы описываем другой, уже реализуемый практически метод трудоемкости $O(n^2 \ln(n/\nu))$ с числом элементарных операций на организацию $O(n^4 \ln(n/\nu))$ и памятью $O(n^2)$ (§ 5 гл. II)). Подчеркнем, что применимость и гарантии точности и трудоемкости указанных методов не связаны ни с наличием (и числом) ограничений задачи, ни с ее гладкостью и степенью обусловленности, а предполагают лишь ее выпуклость.

Итак, асимптотика сложности рассматриваемого класса по $\nu \to 0$ есть $O(n \ln(1/\nu))$. Теперь настала очередь выяснить, как ведет себя сложность при росте размерности задачи n. Оказывается, все дело в том, каковы аффинные свойства G. Если G — параллелепипед, то $N(\nu) \sim n \ln(1/v)$ при всех $\nu < 1/4$ и всех n, так что сложность линейно растет с ростом n. Если же G — эллипсоид, то при увеличении n $N(\nu)$ стабилизируется на уровне $\sim 1/\nu^2$ — сложность «общих» выпуклых задач на эллипсоидах ограничена сверху величиной $O(1/\nu^2)$ независимо от размерности, и при данном ν есть $O(1/\nu^2)$ для всех достаточно больших n. Верхняя оценка сложности $O(1/\nu^2)$ может быть получена стандартным градиентным методом, который, таким образом, «существенно неулучшаем» для задач высокой размерности. Чем же определяется принципиальное различие в интересующем нас отношении между параллелепипедом и эллипсоидом? Оказывается (глава III), тем, что эллипсоид «равномерно выпуклый», а параллелепипед — нет.

Изучение вопроса приводит в главе III к новому взгляду на природу градиентного метода. Известно, что применительно к негладким выпуклым задачам работоспособность градиентного метода связана не с тем, что он работает с локальным линейным приближением (точка зрения, унаследованная от случая гладких задач), а с тем, что при сдвиге вдоль антиградиента мы приближаемся (с точностью до квадрата величины сдвига) к точке минимума. Сам этот факт казался (по крайней мере, авторам) неким чудом, специально изготовленным господом богом на благо выпуклого программирования. Оказывается, что никакого чуда нет: имеется простая и естественная конструкция, сопоставляющая функциям обширного класса методы решения негладких выпуклых задач (мы назвали их методами зеркального спуска). Эта конструкция позволяет поставить производство «чудес» на конвейер. Стандартный градиентный метод является простейшим случаем ее применения.

Мы строим методы зеркального спуска для решения выпуклых задач на выпуклых телах G типа L_p-шаров ($1 \leqslant p < \infty$). Если G — такой шар, то оценка трудоемкости соответствующего метода есть $c(p)(1/\nu)^{\max(2,p)}$ при $p > 1$ (для $p = 1$ имеем $c(1)(1/\nu)^2 \ln n$; случай $p = 1$, по-видимому, небезынтересен для приложений). Оказывается, что в асимптотике по размерности построенные методы «существенно» неулучшаемы.

Исследовав вопрос о сложности класса «всех вообще» выпуклых задач при точном оракуле первого порядка, мы рассматриваем затем случай, когда оракул стохастичен — наблюдения значений компонент задачи и их опорных функционалов искажены случайной помехой (главы V, VI). И в этой ситуации нам удается применить методы зеркального спуска. Оказывается, что их трудоемкость при обеспечении средней (по реализациям помех) погрешности $\nu < 1$ в случае, когда G — выпуклое тело типа L_p-шара, есть

$$c(p)(1/\nu)^{\max(2,p)} \ln^2 \frac{1}{\nu} \ln(m+2), \quad 1 < p < \infty;$$

(m — число ограничений задачи; гипотезы о помехе — стандартные для проблем стохастической аппроксимации; при $m = 0$ множитель $O\left(\ln^2 \frac{1}{\nu}\right)$ можно убрать). Эта оценка, оказывается, в широких предположениях совпадает в существенном со сложностью соответствующего класса и потому в принципе «почти» $\left(\text{с точностью до логарифмического множителя } O\left(\ln^2 \frac{1}{\nu}\right)\right)$ не улучшается. «По дороге» к описанным результатам мы строим имеющие самостоятельный интерес методы зеркального спуска решения игр. Следует отметить, что даже в простейшей ситуации ($p = 2$, $m = 0$) предложенный метод не вполне совпадает с традиционными методами стохастической аппроксимации.

Изучив в первом приближении сложность класса «всех вообще» выпуклых задач, мы изучаем затем (глава VII) важный его подкласс, образованный задачами с гладкими сильно выпуклыми компонентами: минимальные собственные числа матриц вторых производных компонент отделены от 0, а максимальные ограничены сверху; отношение Q соответствующих верхней и нижней границ собственных чисел — модуль сильной выпуклости задачи — характеризует степень ее обусловленности. Простейший градиентный метод решения задач этого класса имеет трудоемкость $O(Q \ln (1/\nu))$. Оказывается, он «чересчур чувствителен» к степени обусловленности. Мы строим другой, менее чувствительный метод с оценкой трудоемкости $O(\sqrt{Q} \ln^2 Q \ln(1/\nu))$ (не зависящей от числа ограничений задачи). Эта последняя оценка в существенном неулучшаема, ибо, как удается доказать, сложность рассматриваемого класса (по крайней мере, в асимптотике по размерности) допускает нижнюю оценку

$$N(\nu) \geqslant c_3 \sqrt{Q} \ln \frac{1}{\nu}.$$

В главе VIII предпринята попытка установить, какие из стандартных методов решения сильно выпуклых задач, считающихся наиболее эффективными (методы сопряженных градиентов), реализуют эту потенциальную границу. Рассмотрены градиентный метод, метод Полака — Рибьера, метод Флетчера — Ривса, метод Зойтендейка. Для всех этих методов получены негативные результаты: в рамках принятых критериев оценки они оказались не лучше градиентного метода.

В главе IX рассматриваются методы нулевого порядка решения выпуклых задач (наблюдаются — точно или с помехами — лишь значения, но не производные функционалов задачи). Сколько-нибудь конечных результатов здесь не получено, построены методы с оценками трудоемкости $O(P(n) \ln(n/\nu))$ ($P(n)$ — полином, информация точна) и $O(P(n) \ln^2(1/\nu) \ln(m+2) \cdot 1/\nu^2)$

(информация искажена помехой). Характер зависимости этих оценок от ν, ввиду сказанного выше, не допускает существенного улучшения даже и в классе методов первого порядка.

§ 2. Семейства задач математического программирования. Приближенные решения и их погрешность

В этом и следующем параграфах описываются основные понятия, которыми мы будем оперировать: «задача математического программирования», «класс таких задач», «метод решения задач данного класса» и т. п. Наша цель — дать формальные определения этим понятиям. Без формализации такого рода мы не можем четко ставить вопрос об оценке потенциальных возможностей численных методов оптимизации.

2.1. Мы будем заниматься в основном методами численного решения задач математического программирования. Каждая такая задача записывается в виде

$$f_0(x) \to \min \mid x \in G, \quad f_j(x) \leqslant 0, \quad 1 \leqslant j \leqslant m, \tag{2.1}$$

где G — подмножество (всегда замкнутое) некоторого вещественного банахова пространства E, $f_j(x)$, $0 \leqslant j \leqslant m$,— вещественнозначные определенные, по крайней мере, на G функции (всегда непрерывные). По техническим причинам пространство E удобно считать сепарабельным (что с практической точки зрения вряд ли является серьезным ограничением). В дальнейшем, если явно не оговорено противное, E считается сепарабельным. Функция f_0 называется *целевым функционалом* задачи, функции f_j, $1 \leqslant j \leqslant m$,— *функционалами-ограничениями* (или просто *ограничениями*), G — *областью определения* задачи.

Точка $x \in E$ называется *планом* задачи (2.1), если $x \in G$ и $f_j(x) \leqslant 0$, $1 \leqslant j \leqslant m$ (т. е. x лежит в области определения задачи и удовлетворяет ограничениям). Сама задача состоит в отыскании плана с наименьшим возможным значением целевого функционала f_0. Соответственно *решением* задачи называется всякий ее план $\bar{x}$, такой, что для любого другого плана x' имеем $f_0(\bar{x}) \leqslant f_0(x')$.

Задача, имеющая планы, называется *совместной* (в противном случае — *несовместной*). Задача, имеющая решение, называется *разрешимой*. Разрешимая задача с необходимостью совместна.

Из сказанного ясно, что задача математического программирования с m ограничениями с формальной точки зрения есть набор объектов $(f = (f_0, \ldots, f_m); G; E)$, где G — область определения задачи, E — соответствующее банахово пространство, $f = (f_0, \ldots, f_m)$ — $(m+1)$-мерная непрерывная вектор-функция на G.

Замечание. Вообще говоря, при формализации понятия «задача математического программирования» нет необходимости наделять G топологией и тем более фиксировать «банахово погружение» $G \subset E$. Однако «непрерывные» задачи, которыми только мы и будем заниматься, удобно определять именно так.

Как правило, мы всегда будем иметь дело с семействами задач, имеющими общие m, G и E. В такой ситуации удобно отождествлять сами задачи с отвечающими им функциями f, что далее всегда и делается. Таким образом, вместо фразы «задача (2.1), определяемая функцией f» мы будем писать «задача f»; из контекста всегда будет ясно, какие G и E имеются в виду.

Обозначим через f_* оптимальное значение целевого функционала задачи f:

$$f_* = \begin{cases} +\infty, & f \text{ несовместна}, \\ \inf\{f_0(x) \mid x \text{ — план } f\}, & f \text{ совместна}. \end{cases}$$

2.2. *Полем задач* математического программирования называется любое семейство задач (2.1) с общими m, G и E. Каждое такое семейство можно отождествить с некоторым множеством $\mathcal{F}$ $(m+1)$-мерных вектор-функций на G. Поле задач, порождаемое объектами G, E, m, $\mathcal{F}$, обозначается $\mathfrak{A}(\mathcal{F}, G, m, E)$.

Пусть, к примеру, G — выпуклое замкнутое ограниченное множество в $R^n = E$ и $\mathcal{F}$ образовано всеми непрерывными покомпонентно выпуклыми функциями f на G. Соответствующее поле — это множество всех выпуклых задач с m ограничениями и областью определения $G \subset R^n$.

2.3. Приближенные решения и их погрешность.

2.3.1. Численные методы, как правило, не могут обеспечить точное решение любой задачи из сколько-нибудь широкого поля. Результаты их применения к задачам представляют собой приближенные решения последних. Для определения характеристик методов необходимо научиться измерять погрешность результата работы метода в качестве приближенного решения рассматриваемой задачи. Этот результат может быть точкой множества G либо констатацией несовместности задачи (такой результат будем обозначать символом $*$). Таким образом, множество возможных результатов применения метода к задаче $f \in \mathfrak{A}(\mathcal{F}, G, m, E)$ есть $G_* = G \cup \{*\}$. Каждой точке $x \in G_*$ и каждой точке $f \in \mathfrak{A}$ мы должны сопоставить число — погрешность точки x в качестве приближенного решения f.

2.3.2. Имеются два основных способа определения погрешности. Один связан с анализом близости кандидата в решение x к истинному решению $\bar{x}$ (в метрике G). Другой связан с изучением отклонений значений $f_j(x)$ от номинальных — требуемых от $\bar{x}$. В этой книге используется только второй из этих двух способов.

Коротко скажем о причинах такого выбора. Прежде всего, если содержательно задача и в самом деле экстремальна, то аппроксимация решения «по функционалу» вполне отвечает сути дела, чего нельзя сказать об аппроксимации «по решению». Это соображение еще не решает вопроса — к экстремальным задачам часто сводят задачи, содержательно неэкстремальные (систему линейных уравнений, скажем, решают редукцией к квадратичной задаче оптимизации). В таких случаях хорошая аппроксимация решения по функционалу сама по себе не обязана обеспечить удовлетворительного решения в исходной постановке. Но есть еще одна — главная — причина нашего выбора. Дело в том, что во всех стандартных (по крайней мере, для этой книги) ситуациях задача аппроксимации решения по функционалу корректна (при правильной постановке), тогда как задача аппроксимации по решению в ряде важных случаев некорректна и не может быть «с гарантией точности» решена ни за какое время (см. упражнение 5 § 3 гл. IV). Таким образом, если требуется аппроксимация «по решению», то необходимо сделать специальные допущения о задаче, делающие эту цель достижимой. Стандартные же допущения такого рода (типа строгой выпуклости) таковы, что при них из аппроксимации по функционалу следует аппроксимация по решению, причем соответствующие погрешности эффективно пересчитываются друг в друга. Итак, и в этом случае можно ограничиться аппроксимацией решения «по функционалу».

2.3.3. В соответствии со сказанным введем две меры погрешности точки $x \in G_*$ в качестве приближенного решения задачи f: векторную $\vec{\varepsilon}\,(x, f)$ и скалярную $\varepsilon_*\,(x, f)$.

Векторная мера погрешности определяется как

$$\vec{\varepsilon}\,(x, f) = (\varepsilon_0(x, f), \ldots, \varepsilon_m(x, f)) = \\ = \begin{cases} (\underbrace{+\infty, \ldots, +\infty}_{m+1 \text{ раз}}), & \text{если } x = * \text{ и } f \text{ совместна}, \\ (\underbrace{0, \ldots, 0}_{m+1 \text{ раз}}), & \text{если } x = * \text{ и } f \text{ несовместна}, \\ \{[f_0(x) - f_*]_+, [f_1(x)]_+ \ldots [f_m(x)]_+\} \\ \text{в остальных случаях (т. е. при } x \in G). \end{cases} \tag{2.2}$$

Здесь и далее $[t]_+$ (где t — скаляр) означает $\max\{0, t\}$. Содержательный смысл определения (2.2) ясен: $\vec{\varepsilon}\,(x, t)$ показывает, насколько значения функционалов задачи f в точке x превышают свои номинальные (требуемые от точного решения) значения. Номинальное значение целевого функционала задачи есть f_*, а номинальные значения функционалов-ограничений равны 0. При этом объявление совместной задачи несовместной карается штрафом $+\infty$. Заметим, что малость $\vec{\varepsilon}\,(x, t)$ еще не означает, что x есть план задачи: при оценке погрешности решения вектором $\vec{\varepsilon}\,(x, f)$ мы не настаиваем на точном выполнении ограничений, но фиксируем соответствующие невязки. Малость $\vec{\varepsilon}\,(x, f)$ означает, в частности, что эти невязки малы: x есть «почти оптимальный почти-план» задачи f.

Скалярная мера погрешности определяется как

$$\varepsilon_*(x, f) = \\ = \begin{cases} +\infty, & x = * \text{ и } f \text{ совместна}, \\ +\infty, & x \in G \text{ не есть план } f, \\ 0, & x = * \text{ и } f \text{ несовместна}, \\ f_0(x) - f_* & \text{в остальных случаях} \\ & \text{(т. е. когда } x \in G \text{ есть план } f). \end{cases} \tag{2.3}$$

Погрешность $\varepsilon_*\,(x, f)$ фактически исключает из рассмотрения приближенные решения $x \in G$, не являющиеся планами f (погрешность таких решений равна $+\infty$). Эту меру стоит использовать в тех случаях, когда по содержательным соображениям допустимы лишь точно удовлетворяющие ограничениям приближенные решения.

Мы будем иметь дело в основном с первым определением (причина снова в том, что проблема отыскания плана совместной задачи поставлена некорректно, если не сделать дополнительных допущений о самой задаче, см. (1.17) гл. IV).

Меру $\vec{\varepsilon}\,(x, f)$ мы будем называть *абсолютной погрешностью* точки x в качестве приближенного решения задачи f, а меру $\varepsilon_*\,(x, f)$ *$*$-абсолютной погрешностью.*

2.3.4. Пользоваться абсолютными погрешностями не всегда удобно, особенно в случаях, когда речь идет о потенциальных границах трудоемкости методов, обеспечивающих заданную погрешность в решении всех задач данного поля. Например, бессмысленно ставить вопрос о трудоемкости метода, решающего все, скажем, выпуклые задачи

$$f(x) \to \min \mid x \in \mathbb{R}, \quad |x| \leqslant 1 \tag{2.4}$$

с абсолютной погрешностью 1. В самом деле, метод такого рода фактически обязан позволить решать все указанные задачи с *любой*, сколь угодно малой, погрешностью. Действительно, если мы желаем решить задачу f с точностью ε, нам достаточно применить исследуемый метод к задаче $\bar{f}(x) = (1/\varepsilon) f(x)$. Ясно, что решение $\bar{f}$ с абсолютной погрешностью 1 будет решением f абсолютной погрешности ε. Соответственно бессмысленно ставить вопрос о трудоемкости такого метода на всем рассматриваемом поле задач; она, разумеется, будет бесконечна.

Выход из положения состоит в том, чтобы перейти от измерения невязок в «абсолютных» единицах к измерению их в подходящей относительной шкале. Иными словами, вместо вектора абсолютных погрешностей $\vec{\varepsilon}(x, t)$ будем рассматривать вектор *относительных погрешностей*

$$\vec{\nu}(x, f) = (\nu_0(x, f), \ldots, \nu_m(x, f)) = \left(\frac{\varepsilon_0(x, f)}{r_0(f)}, \ldots, \frac{\varepsilon_m(x, f)}{r_m(f)}\right).$$

Здесь $r_j(f)$ — единицы измерения погрешностей по j-му функционалу задачи. Выбор этих единиц диктуется соображениями математического удобства и стремлением к наиболее прозрачным результатам. Ясно, впрочем, что этот выбор не играет решающей роли. В качестве примера рассмотрим задачу (2.4). Здесь в качестве $r(f)$ естественнее всего выбрать изменение f на отрезке $\Delta = \{|x| \leqslant 1\}$, т. е. величину $\sup\limits_{|x| \leqslant 1} f(x) - \inf\limits_{|x| \leqslant 1} f(x)$. При таком выборе $r(f)$ представляет собой по существу максимально возможную абсолютную погрешность точки $x \in \Delta$ в качестве приближенного решения задачи f, так что фраза «$x \in \Delta$ суть приближенное решение задачи f с относительной погрешностью ν» имеет четкий содержательный смысл. Она утверждает, что x в $1/\nu$ раз лучшее приближенное решение f, чем наихудшее решение, которое можно получить «методом тыка».

Выбор величин $r_j(f)$ (мы будем называть их *нормирующими множителями*) будет далее явно описываться для каждого из рассматриваемых ниже классов задач. Мы всегда будем выбирать их неотрицательными. Как правило, нормирующие множители будут выбираться так, что относительные погрешности любой (или некоторой априори известной) точки $\hat{x} \in G$ в качестве решения любой задачи $f \in \mathfrak{A}$ не будут превышать 1. *Поле задач* $\mathfrak{A}$ вместе с правилом формирования нормирующих множителей — *нормирующим отображением* $f \mapsto r(f) = (r_0(f), \ldots, r_m(f))$, $f \in \mathfrak{A}$, будем называть *нагруженным полем* задачи.

2.3.5. Введенный выше вектор относительных погрешностей — все еще не слишком удобная для анализа методов мера погрешности. Погрешность желательно измерять скалярной, а не векторной величиной (без этого трудно сравнивать методы по точности). Можно указать различные «естественные» приемы получения скалярной меры погрешности из векторной. Проще всего взять максимум компонент вектора относительной погрешности. Именно так мы и поступим. На первый взгляд при таком определении все компоненты задачи оказываются равноценными с точки зрения влияния их «невязок» на погрешность x в качестве решения задачи f, что содержательно не всегда оправдано. Однако этот недостаток — кажущийся: априорную «неравноценность» невязок по разным компонентам задачи можно учесть надлежащим выбором нормирующих множителей.

Окончательно, мы принимаем два определения *относительной погрешности* точки $x \in G_*$ в качестве приближенного решения задачи $f \in \mathfrak{A}$:

$$\blacksquare \quad \nu(x, f) = \begin{cases} 1, & f \text{ совместна}, \ x = *, \\ 0, & f \text{ несовместна}, \ x = *, \\ \max\left\{\dfrac{f_0(x) - f_*}{r_0(f)}, \dfrac{f_1(x)}{r_1(f)}, \ldots, \dfrac{f_m(x)}{r_m(f)}\right\}, & x \neq *; \end{cases}$$

$$\nu_*(x, f) = \begin{cases} 1, & (f \text{ совместна}, \ x = *) \text{ или } (x \in G \text{ не есть план } f), \\ 0, & f \text{ несовместна}, \ x = *, \\ \dfrac{f_0(x) - f_*}{r_0(f)} & \text{— в остальных случаях.} \end{cases} \qquad \square \ (2.5)^{*)}$$

Здесь $r_j(f) \geqslant 0$, $0 \leqslant j \leqslant m$,— нормирующие множители, выбор которых в каждом из рассматриваемых далее случаев будет указан.

Чтобы не делать тривиальных оговорок, приходится допускать возможность равенства нулю знаменателей в (2.5). В связи с этим примем раз и навсегда, что

$$\frac{a}{0} = \begin{cases} +\infty, & a > 0, \\ 0, & a = 0, \\ -\infty, & a < 0. \end{cases}$$

Заметим, что штрафы $+\infty$ в (2.2), (2.3) заменены в (2.5) штрафами 1. О причинах такой замены будет сказано ниже, в п. 4.1.

Мы надеемся, что приведенная выше содержательная мотивировка введенных в (2.2), (2.3), (2.5) мер погрешности достаточно убедительна. Что же касается формальной стороны дела, то принятые определения удовлетворяют естественным требованиям к мерам погрешности: величины $\vec{\varepsilon}(x, f)$, $\varepsilon_*(x, f)$, $\nu(x, f)$, $\nu_*(x, f)$ — неотрицательные числа (векторы), равные нулю тогда и только тогда, когда x есть либо точное решение f, либо правильный ответ о несовместимости f.

§ 3. Численные методы решения задач математического программирования

В этом параграфе приводится формальное определение основного объекта нашего исследования — численного метода оптимизации. Начнем с описания информационной базы методов — *оракулов*.

---

*) Значок $\square$ перед номером формулы означает, что данный номер относится к группе формул, перед первой из которых стоит значок $\blacksquare$. (*Прим. ред.*)

3.1. Оракул. Как уже говорилось, метод решения задач некоторого семейства в начале работы «не знает» в точности, какую именно задачу ему приходится решать. Все его «априорное» знание состоит в факте принадлежности решаемой задачи некоторому заранее известному полю задач $\mathfrak{A}$. Дальнейшую информацию о задаче, необходимую для решения, метод накапливает, обращаясь к источнику информации — задавая вопросы оракулу. Ответы оракула могут, вообще говоря, искажаться шумом.

Формализуем понятие «оракул» следующим достаточно общим образом. *Оракул* $\mathcal{O}$ для поля задач $\mathfrak{A}(\mathcal{F}, G, m, E)$ представляет собой совокупность нескольких объектов, а именно

— *пространства* Ω *«шумов оракула»* с заданным на нем распределением вероятностей F_ω (Ω предполагается польским пространством, а F_ω — регулярной борелевой полной по Лебегу мерой) (см. § 1 Приложения);

— *функции наблюдения* $\psi(x, f, \omega)$: $G \times \mathcal{F} \times \Omega \to I$, принимающей значения в некотором множестве I (*информационном пространстве*).

В определение включается требование борелевости ψ по ω, x; при этом I предполагается польским пространством.

Если функция наблюдения не зависит от ω (т. е. ответ оракула однозначно определен парой (x, f)), то оракул называется *детерминированным*. В этом случае Ω, разумеется, можно считать одноточечным множеством. Недетерминированные оракулы называются *стохастическими*. Оракул, задаваемый объектами Ω, F_ω, I, $\psi(x, f, \omega)$, будет обозначаться $\mathcal{O}((\Omega, \mathsf{F}_\omega), I, \psi(x, f, \omega))$.

Пусть ω_i — независимые случайные величины со значениями в Ω и функцией распределения F_ω. Будем считать, что на i-м шаге метод может задать оракулу вопрос о решаемой задаче f в любой точке $x_i \in G$. Ответом оракула служит точка $\psi(x_i, f, \omega_i) \in I$. Случайная величина ω_i — это реализовавшийся на i-м такте шум оракула. Вся информация, накопленная методом на первых i шагах его работы на задаче f, представляет собой последовательность $(x_1, \psi(x_1, f, \omega_1); x_2, \psi(x_2, f, \omega_2); \ldots; x_i, \psi(x_i, f, \omega_i))$ заданных вопросов и ответов на них.

3.1.1. Приведем простой пример. Пусть Ω — одноточечное множество (так что фактически оракул свободен от помех), $I = R^{m+1}$ и $\psi(x, f) = f(x)$. Определенный таким образом оракул способен сообщить в каждой точке G значения всех функционалов задачи.

Более «информативные» оракулы могут сообщить не только значения функционалов задачи, но и значения их производных до некоторого порядка. Оракулы такого типа обладают некоторым свойством локальности, которое в общем случае можно определить следующим образом. Оракул $\mathcal{O} = \mathcal{O}((\Omega, \mathsf{F}_\omega), I, \psi(x, f, \omega))$ для поля задач $\mathfrak{A}(\mathcal{F}, G, m, E)$ называется *локальным*, если для всяких x, f, f' ($x \in G$, $f, f' \in \mathcal{F}$), таких, что $f \equiv f'$ в некоторой окрестности x в G, имеем $\psi(x, f, \omega) = \psi(x, f', \omega)$ (для всех $\omega \in \Omega$).

Подчеркнем, что фигурирующая в определении окрестность может зависеть от x, f, f'. Локальный оракул «чувствует» только локальное устройство функционалов задачи в «опрашиваемой» точке. Все «обычные» оракулы математического программирования локальны, и именно с такими оракулами мы и будем работать дальше.

3.2. Теперь мы в состоянии определить основной объект нашего исследования — метод решения задач данного класса. Пусть $\mathfrak{A}(\mathcal{F}, G, m, E)$ — нагруженное поле задач, $\mathcal{O}((\Omega, \mathsf{F}_\omega), I, \psi(x, f, \omega))$ — отвечающий ему оракул. Содержательно ясно, что такое «метод решения задач из $\mathfrak{A}$, использующий оракул $\mathcal{O}$». Действительно, процесс решения состоит в последовательном задании вопросов оракулу и формировании на основе полученных ответов очередных вопросов (в некоторый момент времени — результата). Метод же решения должен быть набором правил, регламентирующих процесс решения. Таким образом, *метод решения задач из* $\mathfrak{A}$, *использующий оракул* $\mathcal{O}$ (короче, $(\mathfrak{A}, \mathcal{O})$-*метод*), *есть набор правил формирования очередных вопросов, момента остановки и выдачи результата, а также самого этого результата в функции от накопленной к данному шагу информации о решаемой задаче.* Эти правила могут быть как детерминированными, так и рандомизированными, что приводит соответственно к детерминированным и стохастическим методам математического программирования.

Итак, работа метода на задаче f состоит в формировании некоторой последовательности $x_1, \ldots, x_{N_f}$ точек G (последовательности вопросов, задаваемых оракулу при решении этой задачи), пополненной точкой $\bar{x} \in G \cup \{*\}$ — результатом применения метода к f. Число N_f шагов метода на задаче f в общем случае может формироваться самим методом (и, стало быть, зависеть от задачи). Последовательность $x_1, \ldots, x_{N_f}, \bar{x}$ *) можно было бы назвать *траекторией метода* на данной задаче. Формально удобнее, однако, присоединить к G, помимо элемента $*$, еще один элемент ϕ («символ бездействия») и считать траекторией метода на f бесконечную последовательность $x_1, \ldots, x_{N_f}, \bar{x}, \phi, \phi, \ldots, \phi, \ldots$

Формальное определение понятия «метод» разумно начать с формального определения траекторий.

3.2.1. **Определение.** Пусть $\mathfrak{A}(\mathcal{F}, G, m, E)$, $\mathcal{O} = \mathcal{O}((\Omega, \mathsf{F}_\omega), I, \psi)$ — нагруженное поле задач и отвечающий ему оракул, и пусть $\bar{G} = G \cup \{*\} \cup \{\phi\}$.

1. *G-траекторией* называется всякая последовательность $x^\infty = (x_1, x_2, \ldots)$, $x_i \in \bar{G}$.

*) Нижний индекс f обычно будет опускаться.

2. *Трудоемкость* траектории x^∞ есть число $l(x^\infty)$ отличных от ϕ членов последовательности x^∞. (Трудоемкость траектории может принимать неотрицательные целые значения и значение $+\infty$.)

3. Траектория x^∞ называется *результативной*, если множество $\{i \mid x_i \neq \phi\}$ непусто и конечно (т. е. если $0 < l(x^\infty) < +\infty$). *Результатом* $\bar{x}(x^\infty)$ результативной траектории называется последний отличный от ϕ элемент последовательности x^∞.

4. *Погрешностью* траектории x^∞ на задаче $f \in \mathfrak{A}$ называется число

$$\nu(x^\infty, f) = \begin{cases} +\infty, \ x^\infty & \text{нерезультативна,} \\ \nu(\bar{x}(x^\infty), f) & \text{в противном случае.} \end{cases}$$

Таким образом, погрешность результативной траектории есть погрешность ее результата в качестве приближенного решения задачи f. Погрешность «нерезультативной» траектории по определению есть $+\infty$.

Аналогично определим $*$-погрешность траектории на задаче f как

$$\nu_*(x^\infty, f) = \begin{cases} +\infty, \ x^\infty & \text{нерезультативна,} \\ \nu_*(\bar{x}(x^\infty), f) & \text{в противном случае.} \end{cases}$$

3.3. Ясно теперь, что такое «$(\mathfrak{A}, \mathcal{O})$-метод». Каждый такой метод есть набор рекуррентных правил формирования траектории в функции от имеющейся к данному моменту информации.

Приведем точные определения. Пусть $J = \bar{G} \times I$, $J^i = \underbrace{J \times \ldots \times J}_{i \text{ раз}}$ (J^0 — одноточечное множество). Точки J будем обозначать $\eta = (x, \xi)$, $x \in \bar{G}$, $\xi \in I$, а точки J^i — $\eta^i = (\eta_1, \ldots, \eta_i)$, $\eta_s \in J$.

3.3.1. Определим вначале детерминированный $(\mathfrak{A}, \mathcal{O})$-метод.

Определение. *Детерминированным* $(\mathfrak{A}, \mathcal{O})$*-методом* называется набор $\mathcal{B} = \{\bar{x}_i(\eta^{i-1})\}_{i=1}^\infty$ борелевых *) функций $\bar{x}_i(\cdot)$: $J^{i-1} \to \bar{G}$.

Пусть $\omega^\infty = (\omega_1, \ldots, \omega_k, \ldots) \in \Omega \times \Omega \times \ldots \times \Omega \times \ldots \ldots = \Omega^\infty$. *Траекторией детерминированного метода* $\mathcal{B} = \{\bar{x}_i(\eta^{i-1})\}$ *на задаче* $f \in \mathfrak{A}$ *при шуме* ω^∞ называется последовательность $x^\infty(\omega^\infty, f, \mathcal{B})$, определяемая рекуррентными правилами:

$$x_i = \bar{x}_i(x_1, \psi(x_1, f, \omega_1); \ldots; x_{i-1}, \psi(x_{i-1}, f, \omega_{i-1})) \ {}^{**)}.$$

Можно доказать (см. п. 1.2 Приложения), что $x^\infty(\omega^\infty, f, \mathcal{B})$ как функция ω^∞ есть борелево отображение Ω^∞ в $\bar{G}^\infty = \bar{G} \times \ldots$

*) Топология на $\bar{G}$ совпадает на G с исходной топологией G; $*$ и ϕ изолированы.

**) Функция $\psi(x, f, \omega)$ не была определена для $x = *$ и $x = \phi$. Будем считать, что $\psi(*, f, \omega) = \psi(\phi, f, \omega) = a$, a — фиксированная точка I.

$\ldots \times \bar{G} \times \ldots$ (топология в прямых произведениях всегда тихоновская). Поэтому можно говорить о распределении значений $x^\infty(\omega^\infty, f, \mathcal{B})$, индуцированном мерой $\mathsf{F}_{\omega^\infty} = \mathsf{F}_{\omega_1} \times \ldots \times \mathsf{F}_{\omega_k} \times \ldots$ на Ω^∞. Это распределение — назовем его $\Phi_{x^\infty}(f, \mathcal{B})$ — называется *распределением траекторий метода* $\mathcal{B}$ *на задаче* f. Можно также рассмотреть совместное распределение $(x^\infty, \omega^\infty)$ — обозначим его $\Phi_{x^\infty, \omega^\infty}(f, \mathcal{B})$ и назовем *совместным распределением траекторий и шума метода* $\mathcal{B}$ *на задаче* f.

Замечание. Требование борелевости поисковых правил $\bar{x}_i(\eta^{i-1})$ и функций наблюдения (равно как и требование, чтобы E было сепарабельным, а I — польским пространством) существенно лишь при рассмотрении стохастических оракулов. Эти предположения обеспечивают измеримость траекторий методов и нужны лишь для того, чтобы иметь возможность применять к анализу методов вероятностный язык. Если же оракул детерминирован и мы занимаемся изучением детерминированных методов, то траектория метода на задаче, результат работы метода, а также вводимые ниже характеристики метода на задаче и на классе допускают естественное определение и без предположений типа борелевости функции наблюдения оракула и поисковых правил рассматриваемых методов, сепарабельности E и J и т. п. Соответственно в дальнейшем, при описании тех или иных детерминированных методов решения оптимизационных задач при детерминированном оракуле, мы не будем заботиться о сепарабельности E и борелевости поисковых правил описываемых методов. Конечно, фактически интерес представляют лишь сепарабельные E, и в этом случае описываемые далее методы автоматически будут иметь борелевы поисковые правила, но мы не будем специально следить за этим.

3.3.2. Определим теперь стохастический $(\mathfrak{A}, \mathcal{O})$-метод. Начнем с наиболее естественного, но не самого удобного определения.

Определение. *Стохастическим* $(\mathfrak{A}, \mathcal{O})$*-методом* называется набор $\{\Phi_{x_i|\eta^{i-1}}\}_{i=1}^\infty = \tilde{\mathcal{B}}$ распределений вероятностей на $\bar{G}$, зависящих, притом борелевым образом *), от параметра $\eta^{i-1} \in J^{i-1}$. Содержательно $\Phi_{x_i|\eta^{i-1}}$ есть распределение, по которому разыгрывается очередной (i-й) элемент траектории метода $\tilde{\mathcal{B}}$ при условии, что имеющаяся к данному шагу информация есть η^{i-1}.

Определим теперь распределение траекторий метода $\tilde{\mathcal{B}}$ на задаче f следующим естественным образом. Вначале рекуррентно определяются распределения $\Phi_{x^i|\omega^\infty}(\tilde{\mathcal{B}}, f)$ фрагментов $x^i = (x_1, \ldots, x_i)$ траекторий при фиксированном шуме $\omega^\infty =$

*) По поводу борелевых семейств мер и используемых операций над мерами см. § 1 Приложения.

$= (\omega_1, \omega_2, \ldots)$ в оракуле;

$$\Phi_{x^i \mid \omega^\infty}(\widetilde{\mathcal{B}}, f) = \Phi_{x_i \mid \psi^{i-1}(x^{i-1}, f, \omega^{i-1})} \otimes \Phi_{x^{i-1} \mid \omega^\infty}(f, \widetilde{\mathcal{B}}).$$

Здесь

$$\psi^i(x^i, f, \omega^i) = (x_1, \psi(x_1, f, \omega_1); \ldots; x_i, \psi(x_i, f, \omega_i)).$$

Очевидно, семейство мер $\{\Phi_{x^i|\omega^\infty}(f, \widetilde{\mathcal{B}})\}_{i=1}^\infty$ согласовано и, стало быть, определяет меру $\Phi_{x^\infty|\omega^\infty}(f, \widetilde{\mathcal{B}})$ на $\overline{G}^\infty$, которая вместе со всеми мерами $\Phi_{x^i|\omega^\infty}(f, \widetilde{\mathcal{B}})$ борелева по ω^∞. Мера $\Phi_{x^\infty,\omega^\infty}(f, \widetilde{\mathcal{B}}) = \mathsf{F}_{\omega^\infty} \otimes \Phi_{x^\infty|\omega^\infty}(f, \widetilde{\mathcal{B}})$ называется *совместным с шумом распределением траекторий* $\widetilde{\mathcal{B}}$ *на задаче* f, а ее проекция $\Phi_{x^\infty}(f, \widetilde{\mathcal{B}})$ на $\overline{G}^\infty$ называется *распределением траекторий* $\widetilde{\mathcal{B}}$ *на* f.

3.4. Смеси и строение стохастических методов. В предыдущем пункте определение стохастического $(\mathfrak{A}, \mathcal{O})$-метода представляется вполне естественным и «самым общим образом» охватывающим возможности рандомизации поиска. Существует и другой способ рандомизации, на первый взгляд, гораздо более тривиальный — образование смесей детерминированных методов. Пусть T — пространство с мерой F_t и $\overline{\mathcal{B}} = \{\mathcal{B}^t = \{\bar{x}_i\ (\eta^{i-1}, t)\}_{i=1}^\infty\}_{t\in T}$ — семейство детерминированных $(\mathfrak{A}, \mathcal{O})$-методов, занумерованных элементами t множества T. Предположим, что это семейство борелево по t (т. е. все $\bar{x}_i\ (\eta^{i-1}, t)$ борелевы по совокупности η^{i-1}, t).

Представим теперь себе следующий процесс решения задач из $\mathfrak{A}$. Перед началом решения разыгрывается наудачу (в соответствии с распределением F_t) «номер» $t \in T$, а затем в ходе решения применяется все время метод $\mathcal{B}^t$. Эта фраза (неформально) определяет некоторый рандомизированный метод решения задач класса $\mathfrak{A}$, который мы назовем *смесью методов* $\mathcal{B}^t$ и обозначим $\overline{\mathcal{B}}$:

$$\overline{\mathcal{B}} = \int_T \mathcal{B}^t d\mathsf{F}_t.$$

В некотором смысле смесь есть смешанная стратегия решения (чистые стратегии естественно отождествить с детерминированными методами).

Формальное определение смеси дать легко. $(\mathfrak{A}, \mathcal{O})$-*смесь* — это пара, составленная из пространства с мерой (T, F_t) и борелева по $t \in T$ семейства детерминированных $(\mathfrak{A}, \mathcal{O})$-методов $\{\mathcal{B}^t\}$. Без труда определяется и основная характеристика смеси $\overline{\mathcal{B}} = \int_T \mathcal{B}^t d\mathsf{F}_t$ на задаче $f \in \mathfrak{A}$ — совместное с шумом распределение $\Phi_{x^\infty,\omega^\infty}(f, \overline{\mathcal{B}})$ ее траекторий на задаче f. Именно, пусть $x^\infty(\omega^\infty, t, f)$ — траектория $\mathcal{B}^t$ на задаче f при шуме оракула ω^∞.

Из борелевости семейства $\mathcal{B}^t$ по t легко извлечь борелевость по (ω^∞, t) отображения $(\omega^\infty, x^\infty(\omega^\infty, t, f))$: $\Omega^\infty \times T \to \Omega^\infty \times \overline{G}^\infty$ (f — параметр). Распределение значений этого отображения, индуцированное мерой $\mathsf{F}_{\omega^\infty} \times \mathsf{F}_t$ на $\Omega^\infty \times T$, и есть по определению $\Phi_{x^\infty,\omega^\infty}(f, \overline{\mathcal{B}})$. Это определение полностью соответствует исходному неформальному описанию понятия «смесь». По $\Phi_{x^\infty,\omega^\infty}(f, \overline{\mathcal{B}})$ способом из п. 3.3.2 определяется распределение $\Phi_{x^\infty}(f, \overline{\mathcal{B}})$ траекторий $\overline{\mathcal{B}}$ на задаче f.

Оказывается, что возможности смесей в некотором точном смысле ничуть не у́же, чем возможности определенных выше стохастических методов.

Теорема. *Для каждого стохастического* $(\mathfrak{A}, \mathcal{O})$-*метода* $\widetilde{\mathcal{B}}$ *существует смесь* $\overline{\mathcal{B}} = \int_0^1 \mathcal{B}^t dt$ *детерминированных* $(\mathfrak{A}, \mathcal{O})$-*методов, эквивалентная* $\widetilde{\mathcal{B}}$ *в следующем точном смысле: совместные с шумом распределения траекторий* $\widetilde{\mathcal{B}}$ *и* $\overline{\mathcal{B}}$ *на каждой задаче из* $\mathfrak{A}$ *совпадают друг с другом.*

Доказательство теоремы вынесено в § 7. Обсудим сам результат. На первый взгляд он может показаться удивительным — настолько определение смесей тривиальнее исходного определения стохастического метода. Мы видим, что для реализации любого такого метода достаточно обратиться к датчику случайных чисел — притом стандартному — всего один раз. Впрочем, легко указать интуитивно убедительное обоснование этого утверждения. Действительно, как мог бы быть реализован стохастический метод, скажем, на ЭВМ? Очевидно, в виде программы, на каждом такте своей работы обращающейся к некоему датчику случайных чисел, работающему совершенно независимо от нее. Но тогда можно заранее до начала работы обратиться к датчику нужное число раз и весь набор его ответов записать в память. В ходе же работы можно обращаться в память, а не к датчику. Получившаяся теперь программа работы есть программа детерминированного метода. Смесью таких методов и является исходный стохастический метод.

Конечно, сформулированное утверждение не означает, что на практике все методы случайного поиска следует организовать именно как смеси. Эквивалентность, устанавливаемая теоремой, есть эквивалентность поведения на задачах, а вовсе не эквивалентность внутренней организации. Однако в рамках нашего подхода, игнорирующего внутреннее устройство методов, эквивалентные в смысле этой теоремы методы неразличимы.

Значение теоремы не практическое, а формальное. Теорема позволяет интерпретировать стохастические методы как некоторые простые конгломераты детерминированных. Как техническое средство этот результат — основа всего анализа рандомизированных методов.

Заметим, что мы не утверждали пока еще, что всякая смесь эквивалентна некоторому стохастическому методу. Это, разумеется, так и есть (при весьма общих предположениях относительно свойств участвующих в определениях объектов), но мы не будем уточнять эти свойства и доказывать соответствующую теорему. Вместо этого мы примем определение смеси за единственное, действующее в книге, определение стохастического $(\mathfrak{A}, \mathcal{O})$-метода и отменим за ненадобностью определение п. 3.3.2, сыгравшее свою роль до конца. Множество всех стохастических $(\mathfrak{A}, \mathcal{O})$-методов обозначим $\widetilde{\mathfrak{B}}(\mathfrak{A}, \mathcal{O})$. Очевидно, детерминированные $(\mathfrak{A}, \mathcal{O})$-методы являются частными случаями *) смесей, так что множество этих методов отождествляется с некоторым подмножеством $\mathfrak{B}(\mathfrak{A}, \mathcal{O})$ в $\widetilde{\mathfrak{B}}(\mathfrak{A}, \mathcal{O})$. Учитывая включение, достаточно давать дальнейшие определения лишь для стохастических методов.

§ 4. Характеристики методов на задаче и на классе. Сложность

Продолжим формирование языка, на котором мы будем описывать свойства методов математического программирования. Пусть фиксировано нагруженное поле задач $\mathfrak{A}(\mathcal{F}, G, m, E)$ и отвечающий ему оракул $\mathcal{O}((\Omega, \mathsf{F}_\omega), I, \psi)$. Наша конечная цель — изучение наилучших $(\mathfrak{A}, \mathcal{O})$-методов. Чтобы продвинуться в этом направлении, надо прежде всего научиться сравнивать методы друг с другом: вначале на каждой задаче из $\mathfrak{A}$, а потом и на $\mathfrak{A}$ в целом. Начнем с характеризации методов на задаче.

4.1. Как уже отмечалось, принятый подход игнорирует свойства методов, связанные с их внутренней организацией, и ограничивается описанием методов в терминах «качество (погрешность) решения — трудоемкость».

С этой точки зрения поведение метода $\widetilde{\mathcal{B}}$ на задаче исчерпывающим образом описывается распределением его траекторий $\Phi_{x^\infty}(f, \widetilde{\mathcal{B}})$ и даже производными от него распределениями «трудоемкость — погрешность», т. е. индуцированным Φ_{x^∞} совместным распределением $\Phi_{\nu, l}(f, \widetilde{\mathcal{B}})$ величины $(\nu(x^\infty, f), \; l(x^\infty))$.

В некоторых случаях мы будем рассматривать и распределение «трудоемкость — *-погрешность» $\Phi^*_{\nu, l}(f, \widetilde{\mathcal{B}})$ — совместное распределение пары $(\nu_*(x^\infty, f), l(x^\infty))$. Оба этих распределения представляют собой меры на $[0, \infty] \times [0, \infty]$.

*) Говоря более строго, детерминированные методы эквивалентны надлежащим смесям в смысле теоремы 3.4.

Описание методов в терминах их распределений «трудоемкость — погрешность» на задачах не позволяет непосредственно сравнивать методы. Чтобы сделать такое сравнение возможным, приходится «загрубить» описание, перейти к характеризации методов на задачах числами, а не распределениями. Существует много способов «естественного» сопоставления распределениям чисел. Самый простой из них — осреднение. Он и будет для нас основным. Именно, определим (средние) трудоемкость, погрешность и *-погрешность метода $\widetilde{\mathcal{B}}$ на задаче $f \in \mathfrak{A}$ соотношениями

$$l(\widetilde{\mathcal{B}}, f) = \int l\, d\,\Phi_l(f, \widetilde{\mathcal{B}}), \quad \nu(\widetilde{\mathcal{B}}, f) = \int \nu\, d\,\Phi_\nu(f, \widetilde{\mathcal{B}}),$$
$$\nu_*(\widetilde{\mathcal{B}}, f) = \int \nu\, d\Phi^*_\nu(f, \widetilde{\mathcal{B}}).$$

Здесь Φ_l, Φ_ν, Φ^*_ν — проекции мер $\Phi^*_{\nu, l}$, $\Phi^*_{\nu, l}$ на соответствующие прямые сомножители $[0, \infty] \times [0, \infty]$.

Более «жесткие» характеристики — это *сильная трудоемкость* и *сильная погрешность*:

$$\bar{l}(\widetilde{\mathcal{B}}, f) = \sup_{\Phi_{l(f, \widetilde{\mathcal{B}})}} \operatorname{vrai} l, \; \bar{\nu}(\widetilde{\mathcal{B}}, f) = \sup_{\Phi_\nu(f, \widetilde{\mathcal{B}})} \operatorname{vrai} \nu,$$
$$\bar{\nu}_*(\widetilde{\mathcal{B}}, f) = \sup_{\Phi^*_\nu(f, \widetilde{\mathcal{B}})} \operatorname{vrai} \nu_*,$$

при которых метод характеризуется свойствами «наихудших» траекторий.

Заметим еще, что в важном случае свободного от шума оракула и детерминированных методов все намеченные способы характеризации совпадают.

З а м е ч а н и е. Теперь мы можем мотивировать проделанную в (2.5) замену штрафов $+\infty$, фигурирующих в (2.2), (2.3), штрафами 1. Дело в том, что при наличии помех нельзя рассчитывать на достоверное решение вопроса о совместности задачи (и гарантировать точное удовлетворение ограничений задачи результатом работы метода). В этих условиях бесконечные штрафы за невыполнение указанных требований привели бы к бесконечным средним (и сильным) погрешностям методов на задачах.

4.2. Теперь уместно приостановить поток определений и проанализировать взаимосвязь введенных понятий. Нашей целью была формализация понятий «численный метод решения оптимизационных задач данного типа» и «характеристики данного метода». Посмотрим, каким путем мы достигли этой цели.

Отправной точкой было стандартное понятие:

(1) *задача математического программирования*

(и непосредственно примыкающее к нему понятие поля таких задач — т. е. содержательно — множества всех задач данного типа).

Чтобы формализовать понятие «метод решения задач данного типа», нам пришлось ввести в рассмотрение еще одно понятие:

(2) *оракул* для данного поля задач.

Фиксация объектов «поле задач» и «оракул» позволили определить понятие

(3) *метод*,

формализующее содержательное понятие «метода решения задач данного типа, использующего данный источник информации о задаче».

Подчеркнем, что на самом деле запас $(\mathfrak{A}, \mathcal{O})$-методов, отвечающий приведенному определению, зависит не от самих объектов $\mathfrak{A}$ и $\mathcal{O}$, а от их компонент: области G определения задач поля $\mathfrak{A}$ и информационного пространства I оракула $\mathcal{O}$. Пара (G, I) однозначно определяет множество $(\mathfrak{A}, \mathcal{O})$-методов. Функционирование же данного метода $\tilde{\mathcal{B}}$ на данной задаче $f \in \mathfrak{A}$ (т. е. отвечающее $\tilde{\mathcal{B}}$, f совместное с шумом распределение траекторий $\tilde{\mathcal{B}}$ на f) определяется уже не только объектами f, G и I, но еще и устройством оракула $\mathcal{O}$. Однако это распределение все еще не зависит от определения мер погрешности точек G_* в качестве приближенных решений f, т. е. от

(4) *нормирующего отображения*,

которым снабжено поле $\mathfrak{A}$. Нам приходится фиксировать это отображение для того, чтобы определить характеристики (трудоемкость и погрешность) метода на задаче. Имея в своем распоряжении не только понятие метода, но и способ его характеризации на задачах из $\mathfrak{A}$, мы уже в состоянии ставить вопрос о выборе наилучшего метода решения задач $\mathfrak{A}$.

Таким образом, отправной точкой для обсуждаемого подхода к анализу возможностей численных методов оптимизации является фиксация набора следующих трех объектов:

(I) поле задач $\mathfrak{A}$;

(II) нормирующее отображение $f \mapsto r(f)$ $(f \in \mathfrak{A})$ и

(III) оракул $\mathcal{O}$ для поля задач $\mathfrak{A}$.

Объекты (I) и (III) определяют запас подлежащих рассмотрению методов, а (II) позволяет охарактеризовать их работу на задачах из $\mathfrak{A}$.

Совокупность объектов (I)—(III) будем называть *классом задач* математического программирования (класс будет обозначаться $\bar{\mathfrak{A}}$). Это понятие, как уже говорилось, служит исходным пунктом для анализа методов, цель которого — оценка потенциальных возможностей методов решения задач данного класса и отыскание в некотором смысле наилучших среди таких методов.

4.3. Итак, мы охарактеризовали метод $\tilde{\mathcal{B}}$ на каждой задаче $f \in \mathfrak{A}$ парой чисел — трудоемкостью и погрешностью. Этого опять-таки недостаточно для того, чтобы сравнивать методы друг с другом. Из двух методов первый будет, вообще говоря, лучше второго на одних и хуже — на других задачах из $\mathfrak{A}$. Мы снова вынуждены загрубить характеризацию методов — перейти от характеристики-функции задачи класса к характеристике-числу. С этой целью определим трудоемкость, погрешность и $*$-погрешность $(\mathfrak{A}, \mathcal{O})$-метода $\tilde{\mathcal{B}}$ на классе $\bar{\mathfrak{A}}$ как

$$l(\tilde{\mathcal{B}}, \bar{\mathfrak{A}}) = \sup_{f \in \mathfrak{A}} l(\tilde{\mathcal{B}}, f), \quad \nu(\tilde{\mathcal{B}}, \bar{\mathfrak{A}}) = \sup_{f \in \mathfrak{A}} \nu(\tilde{\mathcal{B}}, f),$$
$$\nu_*(\tilde{\mathcal{B}}, \bar{\mathfrak{A}}) = \sup_{f \in \mathfrak{A}} \nu_*(\tilde{\mathcal{B}}, f).$$

Аналогично — взятием верхней грани $\bar{l}(\tilde{\mathcal{B}}, f)$, $\bar{\nu}(\tilde{\mathcal{B}}, f)$ и $\bar{\nu}_*(\tilde{\mathcal{B}}, f)$ по задачам класса — определяются *сильные характеристики метода на классе* $\bar{l}(\tilde{\mathcal{B}}, \bar{\mathfrak{A}})$, $\bar{\nu}(\tilde{\mathcal{B}}, \bar{\mathfrak{A}})$, $\bar{\nu}^*(\tilde{\mathcal{B}}, \bar{\mathfrak{A}})$.

Если погрешность метода $\tilde{\mathcal{B}}$ на классе $\bar{\mathfrak{A}}$ (обычная, сильная, $*$-погрешность) не превосходит ν, то мы будем говорить также, что *метод* $\tilde{\mathcal{B}}$ *имеет на* $\bar{\mathfrak{A}}$ *точность* ν (соответственно обычную, сильную, $*$-точность).

Упражнение 1. Пусть $\mathfrak{A}$ не более чем счетно и $\tilde{\mathcal{B}}$ — $(\mathfrak{A}, \mathcal{O})$-метод. Докажите, что существует детерминированный $(\mathfrak{A}, \mathcal{O})$-метод $\bar{\mathcal{B}}$ такой, что $\bar{l}(\bar{\mathcal{B}}, \bar{\mathfrak{A}}) \leqslant \bar{l}(\tilde{\mathcal{B}}, \mathfrak{A})$, $\bar{\nu}(\bar{\mathcal{B}}, \bar{\mathfrak{A}}) \leqslant \bar{\nu}(\tilde{\mathcal{B}}, \bar{\mathfrak{A}})$ и $\bar{\nu}_*(\bar{\mathcal{B}}, \bar{\mathfrak{A}}) \leqslant \bar{\nu}_*(\tilde{\mathcal{B}}, \bar{\mathfrak{A}})$. Таким образом, при «сильной» характеризации методов рандомизация поиска становится излишней.

Указание. См. теорему 3.4.

4.4. Теперь мы в состоянии определить потенциальную границу трудоемкости решения задач данного класса — сложность класса задач $\bar{\mathfrak{A}}$. Грубо говоря, сложность определяется как функция точности, равная минимально возможной трудоемкости метода, обеспечивающего еще решение с нужной точностью всех задач класса. В зависимости от того, как определяются точность и трудоемкость (а мы дали несколько определений этих понятий) и какие ограничения налагаются на применяемые методы, получается не одно определение сложности, а много. Мы ограничимся двумя — «самым слабым» и «самым сильным». Именно, пусть

$$\tilde{N}(\nu) = \inf\{l \mid \exists \tilde{\mathcal{B}} \in \tilde{\mathfrak{B}}(\mathfrak{A}, \mathcal{O}) : l(\tilde{\mathcal{B}}, \bar{\mathfrak{A}}) \leqslant l, \nu(\tilde{\mathcal{B}}, \bar{\mathfrak{A}}) \leqslant \nu\},\ ^{*)}$$
$$N(\nu) = \inf\{l \mid \exists \mathcal{B} \in \mathfrak{B}(\mathfrak{A}, \mathcal{O}) : \bar{l}(\tilde{\mathcal{B}}, \bar{\mathfrak{A}}) \leqslant l, \nu(\mathcal{B}, \bar{\mathfrak{A}}) \leqslant \nu\}.$$

Функция $\tilde{N}(\nu)$ называется *стохастической*, а $N(\nu)$ — *сильной детерминированной сложностью класса* $\bar{\mathfrak{A}} = (\mathfrak{A}, \mathcal{O}, r(\cdot))$.

*) Всегда $\inf\{t \mid t \in \phi\} = +\infty$.

Очевидно, $\tilde{N}(\nu) \leqslant N(\nu)$. Если в определениях сложностных функций погрешность заменить $*$-погрешностью, то получится определение $*$-сложностей $\tilde{N}_*(\nu)$ и $N_*(\nu)$.

Основное содержание книги состоит в вычислении оценок сложности стандартных классов задач математического программирования и в построении реализующих эти оценки методов решения задач соответствующих классов.

4.5. Избранный нами минимаксный способ характеризации методов на классах «по худшему случаю» несомненно нуждается в содержательной мотивировке. Распространено мнение, что «минимаксный подход» чересчур пессимистичен: более разумным считается усреднение характеристик методов на отдельных задачах класса по некоторому априорному распределению. Содержательно такой «байесов» подход постулирует, что «в жизни» задачи данного типа определенным образом распределены. Заметим, однако, что для сколько-нибудь широких классов задач (например, для всех рассматриваемых далее) нет никакого хоть в какой-то мере оправданного в содержательном отношении способа задать такое априорное распределение на классе задач. Надежды на «экспериментальное определение» такого распределения беспочвенны: если класс задач параметрический с, скажем, 50 параметрами, то сколько-нибудь достоверное «прямое» построение их совместного распределения требует выборки фантастического, безусловно нереального объема. Поэтому даже в простейших линейных задачах эмпирический путь построения априорных распределений безнадежен.

Итак, «байесов подход» к изучению методов решения сколько-нибудь широких классов задач практически бесперспективен: рекомендуемые методы обязаны были бы хорошо работать *при любом* априорном распределении, но тогда они были хорошими и в минимаксном смысле.

4.6. Остановимся еще на одном отличии данного выше определения понятия «метод» от традиционных представлений о численных методах. С принятой здесь точки зрения метод есть конечношаговый процесс, формирующий за конечное время результат требуемой точности и после этого останавливающийся. Традиционное представление состоит в том, что метод работает бесконечное время *) и формирует последовательность приближений, сходящуюся в том или ином смысле к точному решению. Чем выше скорость сходимости, тем лучше такой «бесконечный» метод. Принятая нами схема применения метода такова: задается требуемая точность решения ν, по которой и выбирается метод решения $\mathscr{B}_\nu$. Трудоемкость наилучшего такого метода и есть сложность класса.

*) «Конечных» методов (типа методов линейного программирования) для общих классов задач не существует.

Эта схема, по-видимому, не имеет изъянов в теоретическом плане. Однако на практике часто поступают иначе — «запускают» в принципе бесконечношаговый метод и заставляют его работать столько времени, сколько позволяют имеющиеся вычислительные ресурсы. Какая при этом получится точность — такая и получится. Никаких явных гарантий в этом отношении обычно не дается. При всем том, что теоретически такой подход не вполне удовлетворителен, на практике он более прост.

Говоря более формально, пусть $\mathscr{B}^\infty$ — бесконечношаговый метод решения задач класса $\overline{\mathfrak{A}}$, использующий оракул $\mathcal{O}$, и $\nu(N)$ — максимальная (по задачам класса) погрешность даваемого им на N-м шаге приближения *). Пусть $N_{\mathscr{B}^\infty}(\nu) = \min\{N \mid \nu(M) \leqslant \nu,\ M \geqslant N\}$. Максимальные гарантии, связанные с применением $\mathscr{B}^\infty$, состоят в утверждении, что при *всяком* ν метод $\mathscr{B}^\infty$, проработав время $N_{\mathscr{B}^\infty}(\nu)$, достигнет точности ν, какова бы ни была задача $f \in \mathfrak{A}$. (Как уже говорилось, традиционно методы сравниваются по «скорости сходимости», причем последняя оценивается, как правило, достаточно грубо — порядком зависимости точности, достигнутой на данной задаче за N шагов, от N. Ясно, что такие оценки не дают особых гарантий.)

Пусть теперь $\widetilde{\mathfrak{B}}^l(\overline{\mathfrak{A}})$ — класс конечношаговых $\overline{\mathfrak{A}}$-методов с «остановкой на шаге l» (решение независимо ни от чего принимается на l-м шаге; такие методы будем называть *l-шаговыми*), и пусть $N_{\text{п}}(\nu)$ — сложность $\overline{\mathfrak{A}}$ по отношению к семейству классов методов $\{\widetilde{\mathscr{B}}^l\}_{l=1}^\infty$:

$$N_{\text{п}}(\nu) = \min\{l \mid \exists\ \widetilde{\mathscr{B}} \in \widetilde{\mathfrak{B}}^l(\overline{\mathfrak{A}}):\ \nu(\widetilde{\mathscr{B}}, \overline{\mathfrak{A}}) \leqslant \nu\}.$$

Для каждого *заранее заданного* ν можно подобрать в классе $\widetilde{\mathfrak{B}}_{\text{п}} = \bigcup_l \widetilde{\mathfrak{B}}^l$ метод, который именно эту точность обеспечит быстрее, чем $\mathscr{B}^\infty$. Из определения ясно, что $N_{\text{п}}(\nu) - 1 \leqslant N_{\mathscr{B}^\infty}(\nu)$ **). Однако такой «хороший» метод — свой для каждого ν, и неясно, можно ли из них «построить» бесконечношаговый, столь же хороший уже для *всех* ν, метод. Априори возможно поэтому, что функция $N_{\text{п}}(\nu)$ (которую мы фактически будем изучать) «существенно меньше» (скажем, в асимптотике по $\nu \to 0$) любой из функций $N_{\mathscr{B}^\infty}(\nu)$ и не является тем самым эффективной нижней границей трудоемкости традиционных бесконечношаговых методов.

*) Пусть для определенности погрешность этого приближения на задаче f будет $\nu(\mathscr{B}^N, f)$, где $\mathscr{B}^N$ — очевидное N-шаговое «усечение» $\mathscr{B}^\infty$. Аналогичным образом можно было бы рассмотреть и другие способы измерения погрешности.

**) Напомним, что метод из $\widetilde{\mathfrak{B}}^l$ формирует результат на основе $l-1$ наблюдений.

Представляется, таким образом, что принятый здесь подход не позволяет оценить потенциальную эффективность таких методов и, стало быть, не охватывает широко распространенной на практике схемы применения численных методов.

К счастью, все обстоит вполне благополучно: $N_П(\nu)$ — эффективная граница трудоемкости «сходящихся» методов. Чтобы доказать это, укажем способ получения «хорошего» метода такого рода из «хороших» конечношаговых методов. Пусть $N_П(\nu)$, $0 \leqslant \nu \leqslant \bar{\nu}$, — определенная выше функция сложности класса $\bar{\mathfrak{A}}$ (снабженного оракулом $\mathcal{O}$), а $\varphi(\nu) \geqslant 1$ — любая *непрерывная* невозрастающая функция со значениями в $[0, \infty]$, мажорирующая $N_П(\nu) - 1$ (такую функцию можно выбрать совпадающей с $N_П(\nu) - 1$ вне произвольно малой окрестности счетного множества точек скачка $N_П(\nu)$ и множества, на котором $N_П(\nu) = 1$).

Теорема. *В сделанных предположениях существует бесконечношаговый* $(\mathfrak{A}, \mathcal{O})$*-метод* $\mathscr{B}^\infty$ *), *для которого* $N_{\mathscr{B}^\infty}(\nu) \leqslant 6\varphi(\nu)$, $0 \leqslant \nu \leqslant \bar{\nu}$.

Доказательство. Случай $\varphi \equiv +\infty$ тривиален. Пусть $\varphi \not\equiv +\infty$ и $\underline{\nu} = \inf\{\nu \mid \varphi(\nu) < \infty\}$. Положим $\bar{\nu} = \nu_1$, и пусть уже определены $\nu_1, \ldots, \nu_s$. Определим ν_{s+1} как $\max\{\nu \mid \varphi(\nu) \geqslant 2\varphi(\nu_s)\}$.

Оборвем процесс на первом шаге, для которого $\varphi(\nu) < 2\varphi(\nu_s)$ при всех $\nu < \nu_s$ (это может произойти лишь при $\varphi(0) < +\infty$). В результате для некоторого $\bar{s} \leqslant +\infty$ и всех натуральных $i < \bar{s}$, $\bar{s} > 1$ будут определены числа $\nu_i \in [\underline{\nu}, \bar{\nu}]$, такие, что $\nu_1 > \nu_2 > \ldots > \nu_j > \ldots$, $j \leqslant \bar{s} - 1$ и $+\infty > \varphi(\nu_j) = 2\varphi(\nu_{j-1})$. Если $\bar{s} < \infty$, то $\varphi(0) < 2\varphi(\nu_{\bar{s}-1})$. Так как $\varphi(\cdot)$ — мажоранта $N_П(\cdot) - 1$, то для каждого $i \leqslant \bar{s} - 1$ можно найти $N_i = N_П(\nu_i) \leqslant (\varphi(\nu_i) + 1)$-шаговый $(\mathfrak{A}, \mathcal{O})$-метод $\mathscr{B}^i$ (элемент $\tilde{\mathfrak{B}}^{N_i}(\mathfrak{A}, \mathcal{O})$), имеющий погрешность ν_i на классе $\bar{\mathfrak{A}}$. Если $\bar{s} < \infty$, то пусть $\mathscr{B}^{\bar{s}}$ — $N_{\bar{s}} = N_П(0)$-шаговый $(\mathfrak{A}, \mathcal{O})$-метод с погрешностью 0 на $\bar{\mathfrak{A}}$.

Сконструируем метод $\mathscr{B}^\infty$ следующим образом. Метод состоит из этапов, i-й из которых ($i \leqslant \bar{s}$) содержит $N_i - 1$ шагов и сводится к применению к решаемой задаче метода $\mathscr{B}^i$. Очередные приближения, определяемые $\mathscr{B}^\infty$ на i-м этапе, совпадают с результатом применения к решаемой задаче метода $\mathscr{B}^{i-1}$ (при $i = 1$ это приближение — любая точка G). Если $\bar{s} < \infty$, то на заключительном ($(\bar{s} + 1)$-м) этапе никакой оптимизации не проводится, и очередные приближения — результат применения $\mathscr{B}^{\bar{s}}$.

Убедимся, что $\mathscr{B}^\infty$ — искомый метод. Действительно, пусть $\nu \in [0, \bar{\nu}]$. Возможно, что $\varphi(\nu) = +\infty$; тогда утверждение тривиально. Пусть теперь $\varphi(\nu) < +\infty$ и, значит, $\nu \geqslant \underline{\nu}$. Тогда при некотором $i_0 < \bar{s}$ имеем $\nu_{i_0} \geqslant \nu \geqslant \nu_{i_0+1}$ (для $\bar{s} < \infty$ принято

*) Читателю предлагается самому определить это понятие.

$\nu_{\bar{s}} = 0$). Отсюда $\varphi(\nu_{i_0}) \leqslant \varphi(\nu) \leqslant \varphi(\nu_{i_0+1})$. Так как $\nu_{i_0+1} \leqslant \nu$, то из строения $\mathscr{B}^\infty$ ясно, что, начиная с $(i_0 + 2)$-го этапа, погрешность $\mathscr{B}^\infty$ на любой задаче не выше ν. Поэтому

$$N_{\mathscr{B}^\infty}(\nu) \leqslant N_1 + \ldots + N_{i_0+1} - (i_0 + 1) \leqslant \varphi(\nu_1) + \ldots + \varphi(\nu_{i_0+1}) \leqslant 3\varphi(\nu_{i_0+1})$$

(мы учли, что $\varphi(\nu_{i+1}) = 2\varphi(\nu_i)$ при $i + 1 < \bar{s}$ и *всегда* $\varphi(\nu_{i+1}) \geqslant \varphi(\nu_i)$). Но вместе с тем $\varphi(\nu_{i_0+1}) \leqslant 2\varphi(\nu_{i_0})$, т. е. $N_{\mathscr{B}^\infty}(\nu) \leqslant 6\varphi(\nu_{i_0}) \leqslant 6\varphi(\nu)$. Теорема доказана.

Замечание. Утверждение теоремы, очевидно, справедливо и в случае, когда мы ограничиваемся рассмотрением детерминированных конечно- и бесконечношаговых методов. Разумеется, при этом необходимо очевидным образом изменить определение функции сложности $N_П(\nu)$.

§ 5. Некоторые утверждения о методах математического программирования

В этом параграфе собраны некоторые общие и широко используемые в дальнейшем утверждения о методах решения экстремальных задач и сложностных функциях. Пусть $\bar{\mathfrak{A}} = (\mathfrak{A}(\mathcal{F}, G, m, E), \mathcal{O}((\Omega, \mathsf{F}_\omega), I, \psi), r(\cdot))$ — класс задач.

5.1. Канонические формы методов. В соответствии с определением $\bar{\mathfrak{A}}$-метод может формировать в качестве траекторий *любые* последовательности элементов $\bar{G}$, в том числе и те, у которых «символ бездействия» встречается и до результата. Содержательно такого рода траектории мало естественны, и некоторые рассуждения с методами удобнее проводить, считая, что таких траекторий метод не формирует. Напротив, при доказательстве общих теорем о методах (скажем, теоремы о строении случайного поиска) удобно именно принятое выше определение траекторий (поэтому-то мы его и выбрали). Оказывается — и это интуитивно очевидно, — что всякий метод можно «привести к каноническому виду» — несущественно изменить и добиться этим, чтобы он формировал только «естественные» траектории.

Определение. Детерминированный $\bar{\mathfrak{A}}$-метод $\mathscr{B}$ называется *правильным*, если $\mathscr{B} = \{\bar{x}_i(\xi^{i-1})\}$ (т. е. аргументами в правилах являются только наборы ответов оракула, но не наборы вопросов, заданных на предыдущих шагах), и при этом

(1) из $\bar{x}_i(\xi^{i-1}) = \phi$ следует $\bar{x}_{i+1}(\xi^{i-1}, \xi_i) = \phi$ для всех $\xi_i \in I$;
(2) из $\bar{x}_i(\xi^{i-1}) = *$ следует $\bar{x}_{i+1}(\xi^{i-1}, \xi_i) = \phi$ для всех $\xi_i \in I$.

Из определения ясно, что правильный метод формирует траектории одного из трех видов: $\phi, \phi, \phi, \ldots$;

$$x_1, \ldots, x_n, \quad \phi, \phi, \ldots, x_1, \ldots, x_{n-1} \in G, \quad x_n \in G_*;$$
$$x_1, \ldots, x_n, \ldots, x_i \in G.$$

Правильной $\overline{\mathfrak{A}}$*-смесью называется* смесь правильных $\overline{\mathfrak{A}}$-методов. Оказывается, что изучение «всех» методов можно свести к изучению правильных.

Определение. Два $\overline{\mathfrak{A}}$-метода $\tilde{\mathscr{B}}$ и $\tilde{\mathscr{B}}'$ называются *сильно эквивалентными* (*эквивалентными*), если совместные с шумом распределения их траекторий (соответственно их распределения «трудоемкость — погрешность») совпадают друг с другом на любой задаче класса.

Мы интересуемся только такими свойствами методов, которые однозначно определяются их распределением «трудоемкость — погрешность». Поэтому эквивалентные (и тем более сильно эквивалентные) методы в рамках принятого подхода неразличимы.

Теорема. *Всякий детерминированный* $\overline{\mathfrak{A}}$*-метод* $\mathscr{B}$ *эквивалентен правильному детерминированному* $\overline{\mathfrak{A}}$*-методу. Всякая* $\overline{\mathfrak{A}}$*-смесь эквивалентна правильной* $\overline{\mathfrak{A}}$*-смеси.*

Доказательство этого (интуитивно очевидного) результата мы опустим.

Пример применения теоремы. В § 4 была введена функция $N_{\text{п}}(\nu)$ — некоторая характеристика сложности класса $\overline{\mathfrak{A}}$ — и была продемонстрирована роль этой характеристики в качестве нижней границы эффективности «бесконечношаговых» $\overline{\mathfrak{A}}$-методов. Докажем, что $N_{\text{п}}(\nu)$ удовлетворяет неравенствам

$$N(\nu) \geqslant N_{\text{п}}(\nu) \geqslant \tilde{N}(\nu).$$

Эти неравенства показывают, что информация об основных изучаемых здесь объектах $N(\nu)$ и $\tilde{N}(\nu)$ есть в то же время и информация о $N_{\text{п}}(\nu)$.

Правое неравенство очевидно. Чтобы доказать левое, заметим, что

$$N(\nu) = \inf\{l \mid \exists \mathscr{B} \in \mathfrak{B}(\overline{\mathfrak{A}}):\ \tilde{l}(\mathscr{B}, \overline{\mathfrak{A}}) \leqslant l, \nu(\mathscr{B}, \overline{\mathfrak{A}}) \leqslant \nu\}.$$

По теореме 5.1 методы $\mathscr{B}$, фигурирующие в определении $N(\cdot)$, можно считать правильными. Но всякий детерминированный *правильный* метод $\mathscr{B}$ с *сильной трудоемкостью* l на классе $\overline{\mathfrak{A}}$ выдает результат на любой задаче не позднее $[l]$-го шага *) с вероятностью 1 (а тогда можно считать, что наверное). Стало быть, любой такой метод есть детерминированный метод из $\tilde{\mathfrak{B}}^{[l]}(\overline{\mathfrak{A}})$ (см. п. 4.6). Но тогда неравенство $N(\nu) \geqslant N_{\text{п}}(\nu)$ сразу следует из определения $N_{\text{п}}(\nu)$.

Приведенная теорема позволяет в дальнейшем работать только с правильными детерминированными методами и их смесями (и мы, начиная с этого места, пользуемся этой возможностью, раз и

*) Как всегда $[\cdot]$ означает целую часть числа.

навсегда считая все рассматриваемые методы правильными и не оговаривая этого специально). Заметим еще, что если детерминированный $\overline{\mathfrak{A}}$-метод $\mathscr{B}$ способен сформировать траекторию $\phi, \phi, \phi, \ldots$, то он формирует эту траекторию на любой задаче и при любом шуме (почему?). Поэтому погрешность такого метода тождественно равна $+\infty$, и рассматривать его нет смысла. Соответственно далее «методы» всегда предполагаются либо детерминированными, не формирующими траекторий $\phi, \phi, \ldots$, либо смесями таких детерминированных методов.

5.2. Смеси задач. Пусть $\mathfrak{A}_0$ — некоторое не более чем счетное подмножество в $\mathfrak{A}$, а P_f — распределение вероятностей на $\mathfrak{A}_0$. Пару $\overline{\overline{\mathfrak{A}}}_0 = (\mathfrak{A}_0, \mathsf{P}_f)$ будем называть $\overline{\mathfrak{A}}$*-смесью* задач. Будем характеризовать $\overline{\mathfrak{A}}$-методы $\tilde{\mathscr{B}}$ на смеси $\overline{\overline{\mathfrak{A}}}_0$ их *средними* (по распределению P_f) трудоемкостью и погрешностью (*-погрешностью), соответственно, величинами

$$\tilde{l}(\tilde{\mathscr{B}}, \overline{\overline{\mathfrak{A}}}_0) = \int\limits_{\mathfrak{A}_0} l(\tilde{\mathscr{B}}, f)\, d\mathsf{P}_f,$$

$$\tilde{\nu}(\tilde{\mathscr{B}}, \overline{\overline{\mathfrak{A}}}_0) = \int\limits_{\mathfrak{A}_0} \nu(\tilde{\mathscr{B}}, f)\, d\mathsf{P}_f,$$

$$\tilde{\nu}_*(\tilde{\mathscr{B}}, \overline{\overline{\mathfrak{A}}}_0) = \int\limits_{\mathfrak{A}_0} \nu_*(\tilde{\mathscr{B}}, f)\, d\mathsf{P}_f.$$

Оказывается, что при такой «байесовой» характеризации возможности рандомизованных методов примерно такие же, что и у детерминированных.

Теорема. *Пусть* $\tilde{\mathscr{B}}$ *— любой* $\overline{\mathfrak{A}}$*-метод, а* $\overline{\overline{\mathfrak{A}}}_0$ *—* $\overline{\mathfrak{A}}$*-смесь задач. Найдется детерминированный* $\overline{\mathfrak{A}}$*-метод* $\mathscr{B}$*, для которого* $\tilde{l}(\mathscr{B}, \overline{\overline{\mathfrak{A}}}_0) \leqslant 2\tilde{l}(\tilde{\mathscr{B}}, \overline{\overline{\mathfrak{A}}}_0)$ *и* $\tilde{\nu}(\mathscr{B}, \overline{\overline{\mathfrak{A}}}_0) \leqslant 2\tilde{\nu}(\tilde{\mathscr{B}}, \overline{\overline{\mathfrak{A}}}_0)$ *(а также и такой детерминированный метод* $\mathscr{B}$*, для которого* $\tilde{l}(\mathscr{B}, \overline{\overline{\mathfrak{A}}}_0) \leqslant 2\tilde{l}(\tilde{\mathscr{B}}, \overline{\overline{\mathfrak{A}}}_0)$ *и* $\tilde{\nu}_*(\mathscr{B}, \overline{\overline{\mathfrak{A}}}_0) \leqslant 2\tilde{\nu}_*(\tilde{\mathscr{B}}, \overline{\overline{\mathfrak{A}}}_0)$*).*

Упражнение 1. Докажите теорему.

⟨Достаточно доказать первое утверждение. Второе доказывается дословно так же. Можно считать $\tilde{\mathscr{B}}$ смесью $\int\limits_0^1 \mathscr{B}^t\, dt$. Тогда, очевидно,

$$\tilde{l}(\tilde{\mathscr{B}}, \overline{\overline{\mathfrak{A}}}_0) = \int\limits_{\mathfrak{A}_0 \times [0,1]} l(\mathscr{B}^t, f)\, d\mathsf{P}_f\, dt = \int\limits_0^1 \tilde{l}(\mathscr{B}^t, \overline{\overline{\mathfrak{A}}}_0)\, dt$$

и, аналогично,

$$\tilde{\nu}(\tilde{\mathscr{B}}, \overline{\overline{\mathfrak{A}}}_0) = \int\limits_0^1 \tilde{\nu}(\mathscr{B}^t, \overline{\overline{\mathfrak{A}}}_0)\, dt.$$

Так как $\widetilde{l}(\cdot)$ неотрицательно, то найдется борелево $T \subset \Delta = [0, 1]$ с mes $T \geqslant 1/2$, такое, что $\widetilde{l}(\mathscr{B}^t, \overline{\mathfrak{A}}_0) \leqslant 2\widetilde{l}(\widetilde{\mathscr{B}}, \overline{\mathfrak{A}}_0)$ при $t \in T$. Имеем, далее, $\int\limits_T \widetilde{\nu}(\mathscr{B}^t, \overline{\mathfrak{A}}_0)\,dt \leqslant \widetilde{\nu}(\widetilde{B}, \overline{\mathfrak{A}}_0)$, и так как mes $T \geqslant 1/2$, то из этого неравенства следует, что при некотором $t_0 \in T$ $\widetilde{\nu}(\mathscr{B}^{t_0}, \overline{\mathfrak{A}}_0) \leqslant 2\widetilde{\nu}(\widetilde{\mathscr{B}}, \overline{\mathfrak{A}}_0)$. Ясно, что метод $\mathscr{B}^{t_0}$ искомый.〉

Эта простая теорема распространяет на случай «векторной» характеризации решающих правил хорошо известное в теории оценивания утверждение: при байесовом подходе рандомизация правил не нужна. Этот факт будет весьма полезен при оценке «стохастической сложности» $\widetilde{N}(\nu)$ снизу: действительно, если определить естественным образом «среднюю» сложность $\widetilde{N}_{\overline{\mathfrak{A}}_0}(\nu)$ $\overline{\mathfrak{A}}$-смеси $\overline{\mathfrak{A}}_0$ (исходя из средних по $\mathfrak{A}_0$ характеристик методов), то будет выполняться очевидное неравенство $\widetilde{N}_{\overline{\mathfrak{A}}_0}(\nu) \geqslant \widetilde{N}_{\overline{\mathfrak{A}}_0}(\nu)$. Полученный результат показывает, что для оценки $\widetilde{N}_{\overline{\mathfrak{A}}_0}(\nu)$ фактически достаточно ограничиться рассмотрением детерминированных методов, а это, как правило, провести проще, чем рассмотрение всех стохастических методов.

5.3. Сложности, связанные с локальными оракулами. Нижние оценки сложности, вычисляемые в последующих главах, как правило, будут инвариантны по отношению к выбору оракула, которым снабжен рассматриваемый класс задач. Существенно лишь, чтобы этот оракул был локален (см. п. 3.1.1). Наметим способ получения таких оценок. Как правило, при получении нижних оценок сложности мы сводим дело к изучению поведения детерминированных методов, использующих детерминированный оракул. В основе последнего лежит следующий очевидный, но чрезвычайно полезный факт.

Лемма о неразличении. *Пусть* $\overline{\mathfrak{A}} = (\mathfrak{A}\ (\mathscr{F}, G, m, E);\ \mathscr{O}(\Omega, \mathsf{F}_\omega, I, \psi);\ r(\cdot))$ — *класс задач, снабженный детерминированным оракулом* $\mathscr{O}$, *а* $\mathfrak{B}$ — *детерминированный* $(\mathfrak{A}, \mathscr{O})$*-метод. Справедливы следующие утверждения*:

(*i*) *если* $f, f' \in \mathfrak{A}$, $x^\infty = (x_1, \ldots, x_i, \ldots) \in \overline{G}^\infty$, *таковы, что* x^∞ *есть траектория* $\mathfrak{B}$ *на* f *и* $\psi(x_i, f) = \psi(x_i, f')$ *при всех* i, *для которых* $x_i \in G$, *то* x^∞ *есть траектория* $\mathfrak{B}$ *на* f';

(*ii*) *пусть* N — *натуральное число (или* $+\infty$), *и пусть* $f_i \in \mathfrak{A}$, $0 \leqslant i < N$. *Пусть, далее, при* $0 \leqslant i < N$ x_{i+1} *есть* $(i+1)$*-я точка траектории* $\mathfrak{B}$ *на* f_i. *Предположим, что для всех* $i < N$, *для которых* $x_{i+1} \in G$, *имеем* $\psi(x_{i+1}, f_{i+1}) = \psi(x_{i+1}, f_j)$ *при всех* j *таких, что* $N > j \geqslant i+1$. *Тогда при всех* $i < N$ *точки* $x_1, \ldots, x_{i+1}$ *являются первыми* $i+1$ *точками траектории* $\mathfrak{B}$ *на* f_i.

Пусть, в частности, оракул $\mathscr{O}$ не только детерминирован, но и локален. Тогда для справедливости утверждения (i) достаточно, чтобы было $f = f'$ в окрестности каждой из лежащих в G точек траектории $\mathfrak{B}$ на f. Для справедливости утверждения (ii) достаточно, чтобы было $f_{i+1} = f_j$ в окрестности x_{i+1} для всех i и j, таких, что $x_{i+1} \in G$ и $N > i+1 \geqslant 1$, $j \geqslant i+1$ (в обоих последних утверждениях имеются в виду окрестности в G).

Лемма о неразличении, при всей своей очевидности, весьма полезна при получении нижних оценок сложности, и это легко понять — утверждения леммы отражают то *единственное* ограничение, которое наложено на возможности методов — ограничение «информационной реализуемости», состоящее в том, что траектория метода рекуррентно определяется накапливаемой в ходе движения по ней информацией.

Упражнение 1. Докажите лемму.

〈Ограничимся доказательством утверждения (ii). Проведем индукцию по i.

База $i = 0$ тривиальна, так как начальная точка траектории $\mathscr{B}$ формируется им в отсутствие информации о решаемой задаче, и, следовательно, эта точка одна и та же для всех задач. Для $i = 1$ доказываемое утверждение также тривиально — по определению x_2 есть вторая точка траектории $\mathscr{B}$ на f_1 (и мы уже видели, что x_1 — первая точка этой траектории).

Шаг индукции. Пусть мы уже знаем, что при всех s, $1 \leqslant s \leqslant k$, справедливо утверждение:

$$I(s):\ x_1, \ldots, x_s \text{ есть траектория } \mathscr{B} \text{ на } f_{s-1}.$$

Мы извлечем отсюда, что $I(k+1)$ тоже верно, если только $k+1 \leqslant N$. Это завершит доказательство. Мы уже знаем, что $I(2)$ справедливо, так что будем считать $k+1 > 2$.

Так как $\mathscr{B}$ — правильный метод и $x_1, \ldots, x_k$ — первые k точек траектории $\mathscr{B}$ на f_{k-1}, то при некотором $\overline{k}$, $\overline{k} \leqslant k$, имеем $x_l \in G$, $l \leqslant \overline{k}$, тогда как $x_l \in \{\phi, *\}$ при $\overline{k} < l \leqslant k$. Пусть уже известно, что

$$\psi(x_l, f_{k-1}) = \psi(x_l, f_k), \quad 1 \leqslant l \leqslant k-1.$$

Ввиду детерминированности метода $\mathscr{B}$ и рекуррентного характера правила формирования траекторий отсюда, очевидно, следует, что $x_1, \ldots, x_k$ — первые k точек траектории $\mathscr{B}$ на f_k. По определению x_{k+1} есть $(k+1)$-я точка этой траектории, так что $I(k+1)$ верно. Итак, осталось проверить, что и в самом деле

$$\psi(x_l, f_{k-1}) = \psi(x_l, f_k), \quad 1 \leqslant l \leqslant k-1.$$

При $x_l \in \{\phi, *\}$ $\psi(x_l, f)$ не зависит от f, так что достаточно проверить сформулированное утверждение для $l \leqslant \min\{\overline{k}, k-1\}$. Пусть $l \leqslant k-1$ таково, что $1 \leqslant l \leqslant \overline{k}$. Тогда по условию $\psi(x_l, f_j) = \psi(x_l, f_l)$ при $j \geqslant l$, т. е., в частности, $\psi(x_l, f_{k-1}) = \psi(x_l, f_l) = \psi(x_l, f_k)$, что и требуется. Лемма доказана.〉

Упражнение 2. Используя лемму, докажите следующий (интуитивно очевидный) результат. Пусть G состоит из N точек, и класс задач образован всеми задачами вида $f_0(x) \to \min \mid x \in G$, оракул сообщает значение f_0 в «опрашиваемой» точке x, $r_0(f) \equiv 1$. Докажите, что при всех $\nu < 1$ сильная детерминированная сложность класса $\overline{\mathfrak{A}}$ равна N.

5.4. В заключение сделаем следующее замечание. До сих пор мы считали, что вектор-функция f, порождающая задачу (2.1), определена только на G. Соответственно и оракул считался поставляющим информацию только в точках G. Между тем иногда априори известно, что функция f определена в некоторой большей, чем G, области G_I (например, во всем E), причем оракул способен поставлять информацию о f во всей этой бо́льшей области. Ясно, что с информационной точки зрения такой «выход за пределы G» может оказаться полезным, так что в рассматриваемой ситуации целесообразно разрешить методам задавать вопросы о решаемой задаче и вне G (в пределах G_I). Данная выше схема описания классов задач, методов их решения и характеризации последних очевидным образом распространяется и на этот, более общий, случай с сохранением (при естественной переформулировке) всех приведенных выше утверждений.

Мы, разумеется, не будем приводить новые версии данных выше определений и предложений. Отметим лишь, что в рассматриваемой теперь ситуации G_I, так же, как и G, должно входить в список объектов, задающих класс задач. Далее, погрешность точки x в качестве приближенного решения задачи f надо теперь уметь определять и для $x \in G_I \setminus G$; будем считать ее равной $+\infty$ (соответственно вектором с координатами $+\infty$, если речь идет о векторных мерах погрешности). Имея в виду эти замечания, читатель без труда распространит описанную выше схему на случай $G_I \supset G$ (по этой причине мы ранее позволили себе не загромождать изложения введением еще одного объекта — G_I).

§ 6. О сложности классов многоэкстремальных задач

Наиболее просто проверяемое и часто встречающееся свойство задачи математического программирования — свойство гладкости, состоящее в непрерывной дифференцируемости некоторое число раз функционалов задачи. Соответственно методы решения классов гладких задач могли бы рассчитывать на наиболее широкие приложения. К сожалению, «универсальные» методы такого рода не могут не иметь катастрофически большой, безусловно неприемлемой трудоемкости. Чтобы убедиться в этом, приведем соответствующие нижние оценки сложности. Опишем сначала рассматриваемые классы задач.

6.1. Пусть $G \subset E^n$ — ограниченное замкнутое множество, k — натуральное число. Мы будем рассматривать задачи вида

$$f_0(x) \to \min \mid f_j(x) \leqslant 0, \quad 1 \leqslant j \leqslant m, \quad x \in G, \tag{6.1}$$

порожденные k-гладкими функциями $f = (f_0, \ldots, f_m)$, т. е. определенными во всем E^n и k раз непрерывно-дифференцируемыми функциями f. Одной только гладкости для получения конструктивных результатов, очевидно, мало. Надо еще ограничить скорость изменения функций f_j и их производных. Удобнее всего сделать это следующим образом. Пусть $\mathsf{L}_0, \ldots, \mathsf{L}_m$ — заданные положительные числа. Обозначим через $\boldsymbol{S}^k(\mathsf{L}_0, \ldots, \mathsf{L}_m)$ множество всех k раз непрерывно-дифференцируемых вектор-функций $f = (f_0, \ldots, f_m)$, у которых k-я производная f_j по любому направлению не превосходит величины L_j:

$$\left|\frac{d^k}{dt^k} f_j(x+th)\right| \leqslant \mathsf{L}_j \|h\|^k, \quad x \in E^n, \; h \in E^n, \; 0 \leqslant j \leqslant m. \tag{6.2}$$

Пусть, далее, $\mathcal{O}$ — произвольный локальный детерминированный оракул для множества $\boldsymbol{S}^k(\mathsf{L}_0, \ldots, \mathsf{L}_m)$ (вопросы можно задавать во всем E^n). Обозначим через $\boldsymbol{S}^k_{\mathfrak{G}}(G; \mathsf{L}_0, \ldots, \mathsf{L}_m)$ класс всех задач вида (6.1), порожденных функциями $f \in \boldsymbol{S}^k(\mathsf{L}_0, \ldots, \mathsf{L}_m)$. Нормирующие множители для класса удобно выбрать в виде

$$r_j(f) \equiv r_j = \frac{\rho^k(G)\,\mathsf{L}_j}{k!},$$

где $\rho(G)$ — радиус G. Содержательно r_j есть естественная оценка максимального (в G) отклонения функции f_j от ее полинома Тейлора $(k-1)$-го порядка, построенного для «центра» множества G.

6.2. Следующая теорема определяет нижнюю оценку сложности класса $\boldsymbol{S}^k_{\mathfrak{G}}(G; \mathsf{L}_0, \ldots, \mathsf{L}_m)$.

Теорема. *Пусть внутренность G непуста, так что асферичность $\alpha_{\|\cdot\|}(G)$ тела G конечна. Тогда стохастическая сложность $\tilde{N}(\nu)$ класса задач $\boldsymbol{S}^k_{\mathfrak{G}}(G; \mathsf{L}_0, \ldots, \mathsf{L}_m)$ допускает нижнюю оценку*

$$\tilde{N}(\nu) \geqslant d^n(k)\,\alpha^{-n}_{\|\cdot\|}(G)\left(\frac{1}{\nu}\right)^{n/k} \equiv \Phi_{G,k}(\nu). \tag{6.3}$$

Здесь $d(k) > 0$ *).

Обсудим оценку (6.3). Прежде всего заметим, что аналогичная оценка для детерминированной сложности класса $\boldsymbol{S}^k_{\mathfrak{G}}(G; \mathsf{L}_0, \ldots, \mathsf{L}_m)$ хорошо известна. По сведениям авторов, она впервые получена в [11]. Обсудим следствия из (6.3). Ясно, что с практической точки зрения характер зависимости $\Phi_{G,k}(\nu)$ от ν и n приводит к катастрофическому росту сложности класса при $\nu \to 0$ и при $n \to \infty$, если только k фиксировано и $\alpha_{\|\cdot\|}(G)$ остается постоянным (т. е. если не рассматривать задач высокой размерности на «сильно уплощенных» телах). Увеличением гладкости можно несколько снизить $\Phi_{G,k}(\nu)$; однако на практике k обычно невелико ($k = 1 \div 2$). Дело здесь, конечно, не в том, что реальные задачи недостаточно гладки; просто мы обычно не в состоянии оценить старшие производные компонент задачи и понять, в какой именно класс типа $\boldsymbol{S}^k_{\mathfrak{G}}(G; \cdot, \ldots, \cdot)$ они попадают. Кроме того, производ-

*) По поводу доказательства см. работу авторов [34].

ные «обычно» быстро растут с ростом их порядка k, и эффект от повышения гладкости «съедается» тем, что для обеспечения требуемой абсолютной точности приходится применять методы, хотя и ориентированные на бо́льшую гладкость, но зато настраиваемые на все более и более высокие относительные точности.

Конечно, с практической точки зрения дело могло бы обстоять не так уж плохо, если бы величина $d(k)$ была достаточно мала. Однако это не так. Даже весьма грубые (заниженные) нижние оценки $\tilde{N}(\nu)$, определяемые доказательством теоремы 6.2 (см. [34]), показывают, что при $\alpha_{\|\cdot\|}(G) = 1$ (G — шар) и $k = 1$ или 2 гарантированное решение всех соответствующих многоэкстремальных задач с относительной погрешностью, скажем, 10^{-3} — абсолютно безнадежное дело уже в размерности $n \sim 20$.

Заметим далее, что нижняя оценка сложности (6.3) точна (в асимптотике по $\nu \to 0$) в следующем точном смысле: при надлежащих $\mathcal{O}$ существуют методы решения задач класса $\boldsymbol{S}^k_{\mathfrak{G}}(G;\ \mathsf{L}_0, \ldots, \mathsf{L}_m)$ с погрешностью $\leqslant \nu$ и верхней оценкой трудоемкости вида $c(G, k)(1/\nu)^{n/k}$ (см. [11]).

Подчеркнем, что оценка (6.3) получена для стохастической сложности класса. Таким образом, применение методов случайного поиска для решения многоэкстремальных задач немалой размерности столь же безнадежно, как и применение детерминированных методов (речь идет, разумеется, о методах, дающих некоторые гарантии; на отдельных задачах может «повезти» любому методу). В этой связи неясны источники оптимизма, проявляемого поклонниками случайного поиска, в отношении его применимости к многоэкстремальным задачам.

Грубо говоря, не только много, но и трехэкстремальные гладкие задачи недоступны решению методами с гарантированными (и приемлемыми в заметных размерностях) оценками трудоемкости: именно, можно показать, что оценка (6.3) «набегает» на задачах с $m = 0$ и не очень большим (всего 3) числом критических точек во всем E^n (точек x, в которых $f_0'(x) = 0$).

Можно утверждать и большее. Рассмотрим подкласс $\overline{\boldsymbol{S}}^k_{\mathfrak{G}}(G, 1)$ класса $\boldsymbol{S}^k_{\mathfrak{G}}(G, 1)$ (отвечающего единичному шару G пространства E^n), составленный всевозможными *одноэкстремальными* задачами $f_0 \in \boldsymbol{S}^k(1)$, такими, что $f(x) = p_f x^2$ при $\|x\| \geqslant 1$, $p_f > 0$, $\inf\limits_G f_0 = \inf\limits_{E^n} f_0$. Подчеркнем, что при $f \in \overline{\boldsymbol{S}}^k_{\mathfrak{G}}(G, 1)$ уравнение $\nabla f(x) = 0$ *по определению класса* $\overline{\boldsymbol{S}}^k_{\mathfrak{G}}(G, 1)$ во всем E^n имеет точно одно решение. Можно доказать, что детерминированная сложность $N(\nu)$ этого класса удовлетворяет оценке вида

$$N(\nu) \geqslant c(n, k)\left(\frac{1}{\nu}\right)^{\frac{n-1}{k}}. \tag{6.4}$$

Оценки (6.4) немногим лучше, чем (6.3). Таким образом, даже одноэкстремальные (невыпуклые) задачи не допускают «гарантированного» метода решения с приемлемой оценкой трудоемкости.

6.3. Обсудим некоторые выводы из полученных результатов. Мы видели, что не только все гладкие, но даже и одноэкстремальные гладкие задачи — класс, слишком сложный для того, чтобы допускать приемлемые по трудоемкости методы решения. Этот вывод основывается на нижних — информационных — оценках реальной вычислительной сложности методов, так что с точки зрения последней дело обстоит еще хуже. С другой стороны, часто именно такого рода классы задач наиболее естественны с точки зрения приложений. Конечно, на практике обычно применяют методы, не слишком задумываясь о том, какие гарантии они способны доставить, и удовлетворяются полученными при этом результатами, которые, по крайней мере, не хуже (а иногда значительно лучше) исходных — нулевых — приближений к решению. Какие бы успехи ни относились на счет этого, чисто эмпирического подхода к проблеме, нельзя считать, что он способен заменить теоретическое обоснование рекомендаций по применению тех или иных методов. Мы видим, что достаточно общая позитивная теория такого рода не может быть связана с классом всех (хотя бы и гладких) задач оптимизации. Выходом из положения было бы выделение таких классов оптимизационных задач, которые, с одной стороны, были бы достаточно общими и естественными (т. е., в конечном счете, охватывающими достаточно многие практические задачи) и, с другой стороны, имели бы приемлемую по величине сложность (хотя бы информационную). Насколько известно авторам, в настоящее время выделен лишь один класс нелинейных задач, удовлетворяющий этим требованиям — это класс задач выпуклого программирования. Выпуклое программирование, по-видимому, имеет достаточно широкое поле приложений, в то же время этот класс приемлем и по сложности. По этой причине мы переходим к изучению классов выпуклых экстремальных задач.

§ 7. Доказательство теоремы 3.4

По поводу используемых в доказательстве понятий см. § 1 Приложения. Результат получен в работе авторов [33].

Лемма. *Пусть $\Phi_{x|y}$ — семейство распределений вероятностей (регулярных, борелевых и полных по Лебегу) на польском пространстве X, зависящее борелевым образом от параметра $y \in Y$, Y — польское пространство. Существует борелева по s, y функция $\varphi(s, y)$: $[0, 1] \times Y \to X$, такая, что распределение значений φ как функции $s \in \Delta = [0, 1]$ (Δ снабжено мерой Лебега) при всяком $y \in Y$ есть $\Phi_{x|y}$.*

Вывод теоремы 3.4 из леммы. Пусть $\{\Phi_{x_i \mid \eta^{i-1}}\}_{i=1}^{\infty}$ есть набор распределений, отвечающий рассматриваемому стохастическому методу $\widetilde{\mathscr{B}}$. Пусть $\Delta^\infty = \prod\limits_{i=1}^{\infty} \Delta = \{t^\infty = (t_1, \ldots, t_k, \ldots),\ t_k \in \Delta\}$ (произведение снабжено тихоновской топологией и мерой F_{t^∞}, являющейся произведением мер Лебега на сомножителях). По лемме существуют борелевы функции $\bar{x}_i\,(\eta^{i-1}, t_i)\colon \Delta^\infty \times J^{i-1} \to \overline{G}$, такие, что распределение их значений как функций t^∞ на Δ^∞ (индуцированное мерой F_{t^∞}) при каждом η^{i-1} есть $\Phi_{x_i \mid \eta^{i-1}}$.

Рассмотрим смесь методов $\mathscr{B}^{t^\infty} = \{\bar{x}_i\,(\eta^{i-1}, t_i)\}_{i=1}^{\infty}$ — метод $\overline{\overline{\mathscr{B}}} = \int\limits_{\Delta^\infty} \mathscr{B}^{t^\infty}\, d\mathsf{F}_{t^\infty}$. Из определения распределений $\Phi_{x^i \mid \omega^\infty}(\widetilde{\mathscr{B}}, t)$ и определения $\bar{x}_i\,(\eta^{i-1}, t_i)$ индукцией по i непосредственно выводится (мы не будем здесь этого делать), что при всяком $f \in \mathfrak{A}$ и всяком ω^∞, i распределение начального фрагмента длины i траектории $\overline{\overline{\mathscr{B}}}$ на f при шуме оракула ω^∞ есть в точности $\Phi_{x^i \mid \omega^\infty}(\widetilde{\mathscr{B}}, f)$. Стало быть, распределение траекторий $\overline{\overline{\mathscr{B}}}$ на f при шуме оракула ω^∞ есть $\Phi_{x^\infty \mid \omega^\infty}(\widetilde{\mathscr{B}}, f)$, так что $\overline{\overline{\mathscr{B}}}$ эквивалентно $\widetilde{\mathscr{B}}$ в требуемом теоремой 3.4 смысле.

Пусть теперь $t^\infty(t)$ — борелева функция на Δ со значениями в Δ^∞, распределение значений которой, индуцированное мерой Лебега на Δ, есть F_{t^∞} (такая функция есть в силу леммы). Ясно, что смесь $\overline{\overline{\mathscr{B}}}$ эквивалентна смеси $\overline{\mathscr{B}} = \int\limits_0^1 \mathscr{B}^t\, dt$ методов $\mathscr{B}^t = \{\bar{x}_i\,(\eta^{i-1}, (t^\infty(t))_i)\}_{i=1}^{\infty}$, так что и $\overline{\mathscr{B}}$ эквивалентно $\widetilde{\mathscr{B}}$. Теорема 3.4 доказана.

Доказательство леммы. Пусть $X = \bigcup\limits_{i=1}^{\infty} X_i^1$ — разбиение X на борелевы попарно непересекающиеся множества диаметра $\leqslant {}^1/_2$, $X^k_{i_1, \ldots, i_k} = \bigcup\limits_{i=1}^{\infty} X^{k+1}_{i_1, \ldots, i_k, i}$ — разбиение $X^k_{i_1, \ldots, i_k}$ на попарно не пересекающиеся борелевы множества диаметра $\leqslant {}^1/_{2^{(k+1)}}$, $k = 1, 2, \ldots$ Пусть при $X^k_{i_1, \ldots, i_k}$ непустом $\xi^k_{i_1, \ldots, i_k}$ — какая-нибудь точка $X^k_{i_1, \ldots, i_k}$. Для пустого $X^k_{i_1, \ldots, i_k}$ точка $\xi^k_{i_1, \ldots, i_k} = \bar{\xi}$, $\bar{\xi}$ — фиксированная точка X. Обозначим

$$t^k_{i_1, \ldots, i_k}(y) = \Phi_{x \mid y}\{X^k_{i_1, \ldots, i_k}\}.$$

Тогда $t^k_{i_1, \ldots, i_k}(y)$ — неотрицательная борелева функция y.

Определим функции $T^k_{i_1, \ldots, i_k}(y)$, $i_j \geqslant 1$, $j < k$, $i_k \geqslant 0$, следующим образом: при $k = 1$

$$T^1_0(y) = 0,\ T^1_{i_1}(y) = \sum_{j=1}^{i_1} t^1_j(y).$$

Пусть, далее, уже определены $T^k_{i_1, \ldots, i_k}(y)$. Тогда

$$T^{k+1}_{i_1, \ldots, i_k, 0}(y) \equiv T^k_{i_1, \ldots, i_k - 1}(y), \quad T^{k+1}_{i_1, \ldots, i_k, i}(y) = \sum_{j=1}^{i} t^{k+1}_{i_1, \ldots, i_k, j}(y) + T^k_{i_1, \ldots, i_k - 1}(y).$$

Очевидно, что $T^k_{i_1, \ldots, i_k}(y)$ — борелевы функции y и $T^k_{i_1, \ldots, i_{k-1}, i}$ растет с ростом i,

$$\lim_{i \to \infty} T^k_{i_1, \ldots, i_{k-1}, i}(y) = T^{k-1}_{i_1, \ldots, i_{k-1}}(y), \quad \lim_{i \to \infty} T^1_i(y) = 1.$$

Все эти соотношения следуют из очевидного равенства

$$T^{k+1}_{i_1, \ldots, i_k, i}(y) = \Phi_{x \mid y}\Big\{\Big(\bigcup_{1 \leqslant j \leqslant i} X^{k+1}_{i_1, \ldots, i_k, j}\Big) \cup \Big(\bigcup_{1 \leqslant j \leqslant i_k - 1} X^k_{i_1, \ldots, i_{k-1}, j}\Big) \cup \ldots \cup \Big(\bigcup_{1 \leqslant j \leqslant i_1 - 1} X^1_j\Big)\Big\}.$$

Определим функции

$$\varphi^k(s, y) = \begin{cases} \bar{\xi}, & \text{если } s = 0, s = 1 \text{ или } s = T^k_{i_1, \ldots, i_k}(y) \\ & \text{при некоторых } i_1, \ldots, i_k; \\ \xi^k_{i_1, \ldots, i_k}, & \text{если } T^k_{i_1, \ldots, i_k - 1}(y) < s < T^k_{i_1 \ldots i_k}(\ ) \\ & \text{при некоторых } i_1, \ldots, i_k\,^{*)}. \end{cases}$$

Убедимся, что $\varphi^k(s, y)$ борелева по s, y. Действительно, пусть s_1, s_2 — вещественные переменные и

$$\theta_1(s_1, s_2, y) = \begin{cases} 0, & s_1 \neq s_2, \\ 1, & s_1 = s_2, \end{cases} \text{ и } \theta_2(s_1, s_2, y) = \begin{cases} 0, & s_1 < s_2, \\ 1, & s_1 \geqslant s_2. \end{cases}$$

Очевидно, θ_1 и θ_2 борелевы по s, y. Но тогда для борелевой скалярной функции $T(y)$ борелевы *график — множество* $\{(s, y) \mid s = T(y)\}$ *и надграфик — множество* $\{(s, y) \mid s \geqslant T(y)\}$, ибо характеристические функции этих множеств суть $\theta_1(T(y), s, y)$ и $\theta_2(T(y), s, y)$, а эти функции борелевы по s, y в силу модельного рассуждения п. 1.2.2 Приложения. Ясно теперь, что $\varphi^k(s, y)$ — борелева функция.

---

*) Если такой набор $i_1, \ldots, i_k$ для данного s есть, то он единствен.

Проверим, что при каждом фиксированном $y = \bar{y}$ при $k \to \infty$ функции $\varphi^k(s, y)$ сходятся (даже равномерно вне некоторого счетного множества) к функции $\varphi(s, y)$. Действительно, пусть $L(\bar{y})$ есть счетное множество, состоящее из концов отрезка Δ и значений всех функций $T^k_{i_1, \ldots, i_k}$ в точке $\bar{y}$, и пусть $\Delta^k_{i_1, \ldots, i_k}(\bar{y})$, $i_1, \ldots, i_k \geqslant 1$, есть множество $\{s \mid T^k_{i_1, \ldots, i_{k-1}, i_k - 1}(\bar{y}) < s < T^k_{i_1, \ldots, i_k}(\bar{y})\}$. Очевидно,

$$\Delta^k_{i_1, \ldots, i_k}(\bar{y}) \supset \bigcup_{i \geqslant 1} \Delta^{k+1}_{i_1, \ldots, i_k, i}(\bar{y});$$

при этом

$$\varphi^k(\Delta^k_{i_1, \ldots, i_k}(\bar{y}), \bar{y}) \subset X^k_{i_1, \ldots, i_k}.$$

Стало быть, на

$$\Delta \setminus L(\bar{y}) \subset \bigcup_{i_1, \ldots, i_k \geqslant 1} \Delta^k_{i_1, \ldots, i_k}(\bar{y})$$

(при всяком $k \geqslant 1$) $\varphi^k(s, y)$ сходятся равномерно на s, причем

$$\rho_X(\varphi^k(s, \bar{y}), \varphi(s, \bar{y})) \leqslant 2^{-k+1}$$

(так как диаметры $X^k_{i_1, \ldots, i_k}$ не превосходят 2^{-k}). На $L(\bar{y})$ последовательность $\{\varphi^k(s, \bar{y})\}_{k=1}^{\infty}$ в каждой точке стабилизируется, так что указанная сходимость $\varphi^k(s, \bar{y})$ действительно имеет место.

Функция $\varphi(s, y)$ борелева (как предел борелевых). Докажем, что распределение $\widetilde{\Phi}_{x|y}$ ее значений как функции $s \in \Delta$ есть $\Phi_{x|y}$. Ввиду регулярности мер $\Phi_{x|y}$, $\widetilde{\Phi}_{x|y}$, достаточно показать, что если $K \subset X$ — компакт и U — содержащее K открытое множество, то

$$\alpha \equiv \Phi_{x \mid y}(K) \leqslant \widetilde{\Phi}_{x \mid y}(U)$$

(так как обе меры вероятностные, то отсюда следует совпадение Φ с $\widetilde{\Phi}$). Фиксируем указанные K, U и $\bar{y} \in Y$, и пусть $\delta > 0$ таково, что замкнутая 3δ-окрестность K лежит в U. Выберем m так, чтобы было $2^{-m} < \delta$, и рассмотрим все множества $X^m_{i_1, \ldots, i_m}$, пересекающие K (и, значит, лежащие в δ-окрестности K). Сумма мер $\Phi_{\cdot|\bar{y}}$ этих множеств не меньше α. Но каждое из них, скажем $X^m_{\bar{i}_1, \ldots, \bar{i}_m}$, имеющее ненулевую меру, отвечает непустому интервалу $\Delta^m_{\bar{i}_1, \ldots, \bar{i}_m}(\bar{y})$. На $\Delta^m_{\bar{i}_1, \ldots, \bar{i}_m}(\bar{y})$ функция $\varphi^m(s, \bar{y})$ принимает значение из $X^m_{\bar{i}_1, \ldots, \bar{i}_m}$, и при этом

$$\operatorname{mes} \Delta^m_{\bar{i}_1, \ldots, \bar{i}_m}(\bar{y}) = \Phi_{x \mid \bar{y}}(X^m_{\bar{i}_1, \ldots, \bar{i}_m}).$$

Вне счетного множества точек $\Delta^m_{\bar{i}_1, \ldots, \bar{i}_m}(\bar{y})$ имеем

$$\rho_X(\varphi(s, \bar{y}), \varphi^m(s, \bar{y})) \leqslant 2^{-m+1} < 2\delta.$$

Итак, почти везде на $\Delta^m_{\bar{i}_1, \ldots, \bar{i}_m}(\bar{y})$ имеем $\varphi(s, \bar{y}) \in U$. Стало быть, обозначая

$$\mathfrak{J} = \{i_1, \ldots, i_m \mid X^m_{i_1, \ldots, i_m} \cap K \neq \phi\},$$

получим

$$\widetilde{\Phi}_{x|\bar{y}}(U) \geqslant \sum_{(i_1, \ldots, i_m) \in \mathfrak{J}} \operatorname{mes} \Delta^m_{i_1, \ldots, i_m}(y) = \\ = \sum_{(i_1, \ldots, i_m) \in \mathfrak{J}} \Phi_{x \mid \bar{y}}(X^m_{i_1, \ldots, i_m}) \geqslant \Phi_{x \mid \bar{y}}(K),$$

что и требуется. Лемма доказана.

Глава II

ВЫПУКЛОЕ ПРОГРАММИРОВАНИЕ. МЕТОДЫ С ЛИНЕЙНОЙ СХОДИМОСТЬЮ ДЛЯ КЛАССОВ ОБЩИХ ВЫПУКЛЫХ ЗАДАЧ

В этой главе описывается основной объект нашего изучения — задача выпуклого программирования. Намечена классификация таких задач. Строятся методы решения общих (наиболее широких) классов задач такого рода, обладающие линейной сходимостью (т. е. сходящиеся со скоростью геометрической прогрессии). Скорость сходимости этих методов определяется размерностью пространства. Методы реализуют сложность решения соответствующих классов задач в асимптотике по точности (это будет доказано в главе IV). В следующей главе будет построено другое семейство методов, скорость сходимости которых хотя и не является линейной, но зато не зависит явно от размерности задачи. Грубо говоря, методы второго типа реализуют сложность решения общих классов выпуклых задач в асимптотике по размерности. Комбинируя обе группы методов, мы научимся в главе IV реализовывать (точнее, «почти реализовывать») сложность решения общих выпуклых задач на большом спектре стандартных выпуклых тел любой размерности при любой точности.

§ 1. Выпуклые множества и выпуклые задачи

Введем основные, изучаемые в главе, классы выпуклых задач и соответствующие им информационные отображения.

Начнем с элементарных сведений из выпуклого анализа, которые нам потребуются. Мы стремимся не только привести некоторое количество справок, но и сформировать полезный для дальнейшего геометрический взгляд на вводимые объекты. Везде далее E означает вещественное банахово пространство.

1.1. Определение. Пусть $G \subset E$. Множество G называется *выпуклым*, если из $x, y \in G$ и $t \in [0, 1]$ следует $tx + (1-t)y \in G$.

Геометрическая интерпретация: множество выпукло, если вместе с любыми двумя своими точками оно содержит и соединяющий их отрезок.

Упражнение 1 (к определению). Если G выпукло и $x_1, \ldots, x_s \in G$, то всякая *выпуклая комбинация* $x = \sum_{i=1}^{s} a_i x_i$, $a_i \geqslant 0$, $\sum_{i=1}^{s} a_i = 1$, точек x_i лежит в G.

Ясно, что пересечение любого семейства выпуклых множеств в E выпукло.

Простейший пример выпуклого множества — это полупространство $\{x \in E \mid \langle e \mid x - x_0 \rangle \leqslant 0\}$, где $e \in E^*$, $e \neq 0$ — фиксированный элемент E^*, x_0 — какая-нибудь точка E. Это множество не только выпукло, но и замкнуто. Пересекая полупространства из какого-либо семейства, мы снова получаем выпуклое замкнутое множество. Оказывается, что таким путем можно получить любое (отличное от E) выпуклое замкнутое подмножество E.

1.2. Теорема. *Пусть $G \subset E$ — выпуклое замкнутое множество, $x \in E \setminus G$. Тогда существует полупространство, содержащее G и не содержащее x. В частности, если $G \neq E$, то G есть пересечение всех содержащих G полупространств.*

Граница фигурирующего в теореме полупространства называется *гиперплоскостью, разделяющей x и G.*

Доказательство теоремы (для случая гильбертова пространства E) проводится в следующей серии полезных для читателя упражнений.

Упражнение 2. Пусть E — гильбертово пространство, $G \subset E$ — непустое выпуклое замкнутое множество, $\bar{x} \in E \setminus G$. Пусть $x_i \in G$ — последовательность, минимизирующая $\|\bar{x} - x\|$ по $x \in G$: $\|\bar{x} - x_i\| \xrightarrow[i\to\infty]{} \inf_{x \in G} \|x - \bar{x}\| \equiv \rho(\bar{x}, G)$. Докажите, что последовательность x_i сходится к точке $x^* \in G$: $\|\bar{x} - x^*\| = \rho(\bar{x}, G)$] и что точка x^*, удовлетворяющая последнему соотношению (т. е. ближайшая к $\bar{x}$ среди точек G), единственна.

Упражнение 3. В условиях предыдущего упражнения рассмотрим полупространство $\{x \in E \mid \langle x - x^* \mid \bar{x} - x^* \rangle \leqslant 0\}$. Докажите, что это полупространство содержит G и не содержит $\bar{x}$. Геометрически: G и $\bar{x}$ лежат по разные стороны от гиперплоскости, проходящей через x^* ортогонально $\bar{x} - x^*$.

Замечание. В общем случае банахова пространства E «евклидова» схема из упражнений 2, 3 уже не годится для доказательства теоремы 1.2. Эта теорема эквивалентна фундаментальной теореме Хана — Банаха.

1.3. Определим теперь понятие выпуклой функции.

Определение. Пусть $G \subset E$ — непустое выпуклое множество и $f\colon G \to \mathbb{R}$ — вещественная функция на G. Функция f называется *выпуклой на G*, если при всех $x, y \in G$ и $t \in [0, 1]$ справедливо неравенство

$$f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y). \tag{1.1}$$

Эквивалентное определение: f выпукла на G, если сужение f на любой отрезок в G — выпуклая функция на этом отрезке.

Упражнение 4 (к определению). *Неравенство Иенсена.* Пусть G выпукло, f выпукла на G и $x = \sum_{i=1}^{s} \alpha_i x_i, \alpha_i \geqslant 0, \sum_{i=1}^{s} \alpha_i = 1$ — выпуклая комбинация точек $x_i \in G$. Докажите, что тогда

$$f(x) \leqslant \sum_{i=1}^{s} \alpha_i f(x_i).$$

Воспользовавшись выпуклостью функции $(-\ln t)$, $t > 0$, и неравенством Иенсена, докажите неравенство Коши

$$(t_1 \dots t_n)^{\frac{1}{n}} \leqslant \frac{t_1 + \dots + t_n}{n}, \quad t_i > 0.$$

Упражнение 5. *Оценка модуля непрерывности выпуклой функции через ее колебание.* Пусть f определена и выпукла на выпуклом множестве G, $x, z \in G$ и y — точка на отрезке $[x, z]$. Докажите, что тогда

$$f(y) - f(x) \leqslant \frac{\|x - y\|}{\|x - z\|}(f(z) - f(x)) \leqslant \frac{\|x - y\|}{\|x - z\|} \mathbf{V}_G(f),$$

где

$$\mathbf{V}_G(f) = \sup_G f - \inf_G f$$

— изменение f на G. В частности, докажите, что если f выпукла в ρ-окрестности G_ρ множества G и $\mathbf{V}_{G_\rho}(f) < \infty$, то f липшицева на G с константой $\frac{1}{\rho} \mathbf{V}_{G_\rho}(f)$.

Удобно интерпретировать свойство выпуклости функции f в терминах ее надграфика. Определим пространство E_+ как $E \times R^1$. *Надграфиком функции* $f\colon G \to \mathbb{R}$ называется подмножество в E_+, определенное как $\{(x, t) \mid x \in G,\ t \geqslant f(x)\}$. Функция f выпукла тогда и только тогда, когда ее надграфик — выпуклое множество в E_+ (полезно представить себе ситуацию при $E = R^1$ или $E = R^2$).

Простейшим примером выпуклой (на всем E) функции является аффинный функционал $f(x) = \langle x \mid e\rangle + c$, $e \in E^*$. Для него неравенство (1.1) превращается в равенство. Аффинные функционалы — это единственно возможные непрерывные функции f, такие, что и f, и $-f$ выпуклы на E.

Подобно тому как все выпуклые замкнутые множества могут быть сконструированы из простейших (полупространств) посредством операции пересечения, так и все непрерывные выпуклые функции могут быть сконструированы из аффинных с помощью операции взятия верхней грани.

Теорема. *Пусть* f_ω, $\omega \in \Omega$,— *выпуклые функции на выпуклом множестве* $G \subset E$. *Тогда таковы же и функции* $\sup_\omega f_\omega(x)$ (*если эта функция определена на* G), *и, при* $\alpha_\omega \geqslant 0$, $\sum_{\omega \in \Omega} \alpha_\omega f_\omega$ (*в этой формуле* $\alpha_\omega \neq 0$ *лишь для конечного числа индексов* ω).

Доказательство предоставляется читателю.

1.4. Теорема. *Пусть* $G \subset E$ — *непустое выпуклое замкнутое множество, а* $f\colon G \to \mathbb{R}$ — *непрерывная выпуклая на* G *функция. Пусть еще* $\operatorname{int} G \neq \phi$ (*или* E *конечномерно*). *Обозначим через* A *множество всех непрерывных аффинных функций* g, *таких, что* $g \leqslant f$ *на* G. *Тогда*

$$f(x) = \sup\{g(x) \mid g \in A\} \quad \text{при} \quad x \in G. \tag{1.2}$$

Эта теорема тесно связана с теоремой 1.2.

Доказательство теоремы проводится в следующей серии упражнений.

Упражнение 6 (редукция к случаю $\operatorname{int} G \neq \phi$). Пусть в условиях теоремы E конечномерно, а E^G — минимальное содержащее G аффинное подпространство E. Докажите, что внутренность $\operatorname{int}_0 G$ по отношению к E^G непуста. Верно ли это и без допущения о конечномерности E?

Упражнение 7. Пусть в условиях теоремы $x \in \operatorname{int} G$, $x_+ = (x, t) \in E_+$, $t < f(x)$. Докажите, что надграфик f — выпуклое и замкнутое множество, а разделяющая его и x_+ гиперплоскость представляет собой график некоторой аффинной функции из A. Выведите отсюда, что (1.2) верно на $\operatorname{int} G$.

Упражнение 8. Покажите, что в условиях теоремы при $\operatorname{int} G \neq 0$ множество $\operatorname{int} G$ плотно в G. Выведите отсюда и из результата предыдущего упражнения утверждение теоремы. (Указание. Полезно иметь в виду, что выпуклая функция на отрезке полунепрерывна сверху.)

1.5. Определим теперь фундаментальное понятие опорного функционала к выпуклой функции.

Определение. Пусть $G \subset E$ выпукло, а $f\colon G \to \mathbb{R}$ — функция. *Опорным к* f *на* G *функционалом в точке* $x_0 \in G$ называется всякий функционал $g \in E^*$, такой, что

$$f(x) - f(x_0) \geqslant \langle g \mid x - x_0\rangle \quad \text{при} \quad x \in G. \tag{1.3}$$

Геометрическая интерпретация: график аффинной функции $y = f(x_0) + \langle g \mid x - x_0\rangle$ лежит (при $x \in G$) под графиком функции f, а над точкой x_0 соприкасается с графиком f.

Упражнение 9 (к определению). Пусть $x_0 \in \operatorname{int} G$ и f дифференцируема в x_0. Докажите, что тогда опорный функционал к f на G в x_0 (если он есть) единствен и совпадает с производной f в x_0.

Упражнение 10 (к определению). Пусть f выпукла на выпуклом множестве G и $x_0 \in G$. Пусть функционал g опорен к f в x_0 *на некоторой окрестности* x_0 *в* G. Докажите, что g опорен к f в x_0 и на всем G.

Множество опорных к f на G в точке x функционалов обозначается $\partial_G f(x)$. В случае гильбертова пространства E элементы E^* канонически отождествляются с элементами E, так что опорные функционалы можно считать векторами E. В этой ситуации их называют и *субградиентами*.

В силу (1.3) опорный функционал несет глобальную информацию о поведении f на G, информацию, как мы увидим, чрезвычайно ценную для решения экстремальных задач. Следует подчеркнуть, что при выпуклых f и G опорный к f на G в точке x функционал определяется поведением f в как угодно малой окрестности x в G, или, как говорят, *ростком* f на G в точке x (утверждение упражнения 10). Поэтому в выпуклом случае знание *локального* устройства f дает чрезвычайно ценную информацию о ее *глобальных* свойствах. Это, по-видимому, и определяет объективную «легкость» решения выпуклых экстремальных задач по сравнению с многоэкстремальными задачами.

Выясним, достаточно ли много опорных функционалов у выпуклой функции. Этот вопрос решается в следующей теореме.

Теорема. *Пусть G выпукло и f: $G \to \mathbb{R}$ непрерывна в точке $x_0 \in \operatorname{int} G$ и выпукла на G. Тогда множество $\partial_G f(x_0)$ непусто.*

Доказательство. В силу утверждения упражнения 10 и условия $x_0 \in \operatorname{int} G$ можно G заменить в утверждении на шар V достаточно малого радиуса с центром в x_0. Так как f непрерывна в x_0, то можно считать f ограниченной в ρ-окрестности V ($\rho > 0$ достаточно мало, равно как и радиус V). В силу утверждения упражнения 5 f непрерывна на V. Используя теорему 1.4, найдем последовательность аффинных функционалов $g_i(x) = c_i + \langle g_i \mid x - x_0\rangle$, $g_i \in E^*$, такую, что $f(x) \geqslant g_i(x)$ на V и $g_i(x_0) \equiv c_i \xrightarrow[i\to\infty]{} f(x_0)$. Линейные функционалы $\langle g_i \mid x\rangle$ ограничены сверху в шаре $V - x_0$ с центром в 0 одной и той же константой $\sup\limits_G f - \min\limits_i c_i$. Стало быть, их нормы равномерно ограничены. Если E конечномерно, то $\{g_i\}$ имеет предельную точку g (в общем случае надо говорить о предельной в смысле слабой топологии сопряженного пространства точке). Ясно, что g и есть опорный к f на G в точке x_0 функционал.

Упражнение 11 (к теореме). Сохранится ли утверждение теоремы, если отбросить в ее формулировке условие выпуклости f? Включение $x_0 \in \operatorname{int} G$?

Упражнение 12 (к теореме). *Обращение теоремы.* Пусть $G \subset E$ выпукло и $\operatorname{int} G \neq \phi$, а f: $G \to \mathbb{R}$ — такая непрерывная функция, что $\partial_G f(x) \neq \phi$ при всех $x \in \operatorname{int} G$. Докажите, что f выпукла на G.

Упражнение 13 (к теореме). Пусть G выпукло в $x_0 \in \operatorname{int} G$, а f выпукла на G и непрерывна в x_0. Определим функцию от $h \in E$ $f'(x_0, h)$ соотношением

$$f'(x_0, h) = \lim_{t \to +0} \frac{f(x_0 + th) - f(x_0)}{t}.$$

Докажите, что $f'(x_0, h)$ определена, непрерывна и выпукла по h, $f'(x_0 + \lambda h) = \lambda f'(x_0, h)$, $\lambda \geqslant 0$. Докажите, что множество опорных к $f'(x_0, h)$ как функции h функционалов в точке $h = 0$ совпадает с $\partial_G f(x_0)$. Покажите, далее, что если $g(h) = a + \langle g \mid h\rangle$ — аффинный функционал, такой, что $g \in E^*$ и $f'(x_0, h) \geqslant g(h)$, $h \in E$, то g опорно к $f'(x_0, h)$ в точке $h = 0$. Выведите отсюда и из теоремы 1.4 непустоту $\partial_G f(x_0)$ (тем самым будет получено новое доказательство теоремы 1.5).

Замечание. Из теоремы Хана — Банаха можно извлечь и более сильный результат, чем утверждение теоремы 1.5. Именно, в условиях упражнения 13 для каждого $h \in E$ существует такое $g \in \partial_G f(x_0)$, что $\langle g \mid h\rangle = f'(x_0, h)$. Таким образом, $\partial_G f(x_0)$ достаточно богато: для каждого направления h существует опорный к f в x_0 функционал с той же, что и у f, производной по направлению h.

Упражнение 14 (к теореме). Пусть G выпукло, f непрерывна и выпукла на G, $\operatorname{int} G \neq \phi$. Докажите, что f липшицева на G с константой L в том и только в том случае, когда для всякого $x_0 \in \operatorname{int} G$ и $g \in \partial_G f(x_0)$ имеем $\| g \|_* \leqslant L$.

Упражнение 15 (к теореме). Пусть G выпукло, f выпуклая и липшицева с константой L функция на G. Докажите, что $\partial_G f(x) \neq \varnothing$ при всех $x \in G$. Более того, для каждого $x \in G$ можно найти $g \in \partial_G f(x_0)$: $\| g \|_* \leqslant L$ (при доказательстве этого факта читатель, не знакомый с такими результатами функционального анализа, как теорема Хана — Банаха и компактность шара в E^* в надлежащей топологии, может ограничиться случаем евклидова конечномерного пространства E).

1.6. Определим теперь понятие строгой выпуклости. Для наших целей достаточно наиболее сильное определение, да и то лишь в гильбертовом случае.

Определение. Пусть E — гильбертово пространство, $G \subset E$ — *непустое выпуклое замкнутое множество и* $\alpha \geqslant 0$. Функция f: $G \to \mathbb{R}$ называется *α-строго выпуклой на* G, если при $x, y \in G$

$$2f\left(\frac{x+y}{2}\right) \leqslant -\frac{\alpha}{4}\| x - y \|^2 + f(x) + f(y)$$

и f непрерывна на G.

Замечание. 0-строго выпуклые функции на G — это в точности непрерывные выпуклые на G функции.

Упражнение 16 *). Пусть f_ω, $\omega \in \Omega$, α-строго выпуклы на G, а $f(x) = \sup\limits_\omega f_\omega(x)$ определена и непрерывна на G. Докажите, что f α-строго выпукла на G.

Упражнение 17. Пусть f, g соответственно α- и β-строго выпуклы на G. Докажите, что $f + g$ $(\alpha + \beta)$-строго выпукла на G.

Упражнение 18. *Пусть f α-строго выпукла на G, $0 \leqslant \beta \leqslant \alpha$.* Докажите, что функция $\beta x^2/2$ β-строго выпукла на E, а $f - \beta\frac{x^2}{2}$ -выпукла на G.

Упражнение 19. Пусть f α-строго выпукла на G, $x_0 \in G$ и $g \in \partial_G f(x_0)$. Докажите, что

$$f(y) \geqslant f(x_0) + \langle g \mid y - x_0\rangle + \frac{\alpha}{2}(y - x_0)^2$$

при $y \in G$.

Упражнение 20. Пусть f α-строго выпукла на G, $\alpha > 0$. Докажите, что $f(x) \to +\infty$ при $\| x \| \to \infty$, $x \in G$. Докажите, что точка x^* минимума f на G существует и единственна и справедливо неравенство

$$\| x - x^* \|^2 \leqslant \frac{4}{\alpha}(f(x) - f(x^*)).$$

*) В упражнениях 16—21 E гильбертово, а $G \subset E$ выпукло, замкнуто и непусто.

Упражнение 21. Пусть $f\colon G \to \mathbb{R}$ непрерывна и $\operatorname{int} G \neq \phi$. Пусть также f имеет вторую производную $\langle f''(x)h \mid h\rangle$ на $\operatorname{int} G$. Докажите, что следующие три утверждения эквивалентны:

1. f α-строго выпукла на G.
2. $\forall x,\ y \in \operatorname{int} G$: $\langle f'(x) - f'(y) \mid x - y\rangle \geqslant \alpha (x-y)^2$.
3. Квадратичная форма $\langle f''(x)h \mid h\rangle - \alpha \langle h \mid h\rangle$ от h неотрицательно определена при $x \in \operatorname{int} G$.

Утверждение упражнения 20 показывает важную роль, которую предположение строгой выпуклости играет в экстремальных задачах: всякая минимизирующая f на G последовательность x_i *сходится* к точке минимума x^*, причем $\| x_i - x^* \|$ эффективно оценивается через $f(x_i) - f(x^*)$.

1.7. Выпуклые множества можно задавать с помощью выпуклых функций. Пусть f — непрерывная выпуклая функция на E. Тогда при любом вещественном a множество $\{x \in E \mid f(x) \leqslant a\}$ выпукло и замкнуто (проверить!). Обратно, если $G \subset E$ непусто, выпукло и замкнуто, то его можно представить в указанном виде — именно, как $\{x \in E \mid \rho_{\|\cdot\|}(x, G) \leqslant 0\}$, где $\rho_{\|\cdot\|}(x, G) = \inf \{\| x - y \| \mid y \in G\}$ — расстояние от x до G.

Упражнение 22. Проверьте, что в описанной выше ситуации $g(x) \equiv \rho_{\|\cdot\|}(x, G)$ — непрерывная выпуклая функция с константой Липшица 1.

Упражнение 23. Покажите (хотя бы в случае конечномерного E, хотя утверждение верно всегда), что $\partial_G g(x)$ при $x \notin G$ состоит из функционалов нормы 1. Если $\operatorname{int} G \neq \phi$, то и при $x \in G \setminus \operatorname{int} G$ $\partial_G g(x)$ содержит функционал нормы 1.

Другой способ задавать выпуклые множества состоит в использовании *функции Минковского* $p(x)$. Пусть G выпукло и замкнуто, а $x_0 \in \operatorname{int} G$. Определим функцию $p(x)$ как $\inf\{\lambda \mid x_0 + \lambda^{-1}(x - x_0) \in G\}$. Геометрически $p(x)$ есть отношение отрезка от точки x_0 до точки x к отрезку от x_0 до пересечения исходящего из x_0 в направлении $x - x_0$ луча с границей G.

Упражнение 24. Докажите, что в описанной выше ситуации $p(x)$ — непрерывная выпуклая функция, однородная первой степени относительно точки x_0 (т. е. $p(\lambda(x - x_0) + x_0) = \lambda p(x)$, $\lambda \geqslant 0$). Докажите, что G описывается как множество $\{x \mid p(x) \leqslant 1\}$.

1.8. Теперь мы в состоянии определить основной объект нашего изучения — выпуклую экстремальную задачу.

Определение. Задача

$$\{f_0(x) \to \min \mid x \in G \subset E,\ f_i(x) \leqslant 0, \quad i = 1, \ldots, m, \tag{1.4}$$

называется *выпуклой*, если G — выпуклое замкнутое непустое подмножество E, а f_i, $i = 0, 1, \ldots, m$, — непрерывные выпуклые на G функции. Совместные выпуклые задачи, как правило, разрешимы. ■

Теорема. *Пусть* (1.4) — *выпуклая задача,* E *рефлексивно, задача* (1.4) *совместна и множество*

$$G_a = \{x \in G \mid f_0(x) \leqslant a,\ f_i(x) \leqslant 0, \quad 1 \leqslant i \leqslant m\}$$

непусто и ограничено при каком-нибудь $a = a_0$. *Тогда задача* (1.4) *разрешима.*

Доказательство. Пусть $f_{0*} = \inf\{f_0(x) \mid x \in G, f_i(x) \leqslant 0,\ i = 1, \ldots, m\} \subset R \cup \{-\infty\}$ — оптимальное значение целевого функционала в (1.4). Рассмотрим множества $G_a \neq \phi$, $f_{0*} < a \leqslant a_0$. Ясно, что они выпуклы и замкнуты и $G_a \subset G_{a'}$ при $a \leqslant a'$. Множество G_{a_0} вдобавок ограничено по условию. В силу рефлексивности E G_{a_0} — компакт в слабой топологии E, а G_a, $f_{0*} < a \leqslant a_0$ — центрированное семейство его замкнутых (в этой топологии) подмножеств. Стало быть, $\bigcap_{a > f_{0*}} G_{a'} \equiv G_* \neq \phi$. Очевидно, G_* есть в точности множество решений задачи (1.4).

Упражнение 25 (к теореме). Сохранится ли утверждение, если снять условие ограниченности G_{a_0}? А если заменить его условием α-строгой выпуклости ($\alpha > 0$) (при гильбертовом E) функции $\sum_{i=1}^{m} f_i$ на G? (В последнем случае следует воспользоваться утверждением упражнения 19 п. 1.6).

В большинстве приложений E рефлексивно (даже гильбертово) и либо G ограничено, либо $\sum_{i=1}^{m} f_i$ α-строго выпукла на G с $\alpha > 0$, так что из совместности рассматриваемых задач сразу следует их разрешимость.

§ 2. Классы выпуклых экстремальных задач

2.1. В соответствии с идеологией, описанной в главе I, класс экстремальных задач задается фиксацией трех объектов: поля задач, оракула и нормирующего отображения. Фактически из этих трех объектов существенны первые два. Нормирующее отображение выбирается по ним из соображений математической естественности и удобства. Таким образом, классификация основных объектов нашего исследования — классов выпуклых задач — предполагает классификацию как множеств таких задач (или, если угодно, самих задач), так и применяемых оракулов.

Что касается оракулов, то здесь мы следуем традиционному их делению на *оракулы нулевого порядка* (сообщающие только значения функционалов задачи), *первого* (сообщающие и значения, и первые производные) и оракулы *более высоких порядков*. Кроме того, оракулы этих типов разделяют еще на *детерминированные* и *стохастические*.

Традиционно считается, что методы, использующие оракулы порядка выше первого, не могут найти широких приложений (по крайней мере, при решении задач заметной размерности), в основном из-за трудностей, связанных с построением оракула. В теоретическом отношении исследование такого рода методов также не слишком интересно. По этим причинам мы не будем сколько-нибудь основательно заниматься этими методами. Несколько замечаний по их поводу имеется в п. 1.1 гл. VIII. *Методы нулевого порядка* (использующие оракул нулевого порядка) представляют несомненный интерес — именно такие оракулы легче всего строить на практике. Некоторые результаты о таких методах содержит глава IX. Из них следует, что возможности таких методов, к сожалению, довольно узки.

В соответствии со сказанным наиболее интересны методы первого порядка, которыми в основном мы и будем заниматься. Отметим, что именно они занимают центральное место в работах по численной оптимизации.

2.2. Наметим теперь классификацию выпуклых экстремальных задач, которой мы будем придерживаться. Всякая классификация основана на выделении и последующей фиксации некоторых характеристик (параметров) задачи. Такими характеристиками являются, в частности,

— *размерность* задачи n, т. е. линейная размерность пространства E, $n = 1, 2, \ldots, \infty$;

— *характер ограничений* задачи, по которому они делятся на *задачи безусловные* ($m = 0$, $G = E$); *задачи без ограничений* (G любое, $m = 0$); *условные задачи* (m любое, G любое);

— *свойства гладкости* задачи (т. е. свойства гладкости функций f_j).

В большинстве случаев (и везде в этой главе) в основе классификации будут лежать свойства гладкости задачи. Прочие характеристики будут определять «более тонкое» деление задач.

2.3. Проведем разделение выпуклых задач вида (1.4) в соответствии со степенью гладкости фигурирующих в них функций. Пусть фиксированы E, $G \subset E$ и $m \geqslant 0$, причем G выпукло и замкнуто. Тогда задачу (1.4) можно отождествить с непрерывной $(m + 1)$-мерной вектор-функцией $f = (f_0, \ldots, f_m)$ на G, компоненты которой выпуклы.

Минимальное естественное (по крайней мере, для ограниченных G) ограничение гладкости f — это требование ее непрерывности (которое мы раз и навсегда предъявляем ко всем рассматриваемым задачам) и ограниченности на G. Последнее эквивалентно требованию конечности всех величин

$$\mathsf{V}_G(f_i) \equiv \mathsf{V}_i(f) = \sup_{x \in G} f_i - \inf_{x \in G} f_i.$$

Соответственно обозначим через $\boldsymbol{C}(G, E, m)$ поле всех выпуклых задач f вида (1.4) с данными G, E, m и конечными $\mathsf{V}_i(f)$, $0 \leqslant i \leqslant m$. Снабдим это множество нормирующим отображением

$$r_i(f) = \begin{cases} \mathsf{V}_0(f), & i = 0, \\ \max\{0, \sup_G f_i\}, & i > 0. \end{cases}$$

Класс задач, получающийся оснащением поля $\boldsymbol{C}(G, E, m)$ оракулом $\mathcal{O}$, будем обозначать $\boldsymbol{C}^{\mathcal{O}}(G, E, m)$ и называть *классом общих выпуклых задач.*

Нормировка погрешности, избранная нами, имеет простой содержательный смысл. Пусть $f \in \boldsymbol{C}^{\mathcal{O}}(G, E, m)$ — совместная задача и мы выдаем ее решение априори — «не глядя». Тогда максимальная возможная абсолютная погрешность такого тривиального (трудоемкости 1) метода на f по i-й компоненте есть $r_i(f)$. Стало быть, фраза «$\boldsymbol{C}^{\mathcal{O}}(G, E, m)$-метод имеет точность ν на f» означает, что этот метод в $1/\nu$ раз точнее указанного тривиального «метода тыка» (ср. с рассуждениями п. 2.3.4 гл. I).

Можно наложить и дополнительные ограничения на модули непрерывности компонент f_i задачи f. Наиболее естественное из них — требование липшицевости f_i на G. Иными словами, мы требуем конечности величин

$$\mathsf{L}_{\|\cdot\|}(f_i) = \sup_{x \neq y,\, x, y \in G} \frac{|f_i(x) - f_i(y)|}{\|x - y\|},$$

где $\|\cdot\|$ — рассматриваемая норма в E. Соответственно обозначим через $\boldsymbol{C}_{\mathrm{Lip}} = \boldsymbol{C}_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$ поле всех выпуклых задач f вида (1.4) с данными G*), E, m и конечными $\mathsf{L}_{\|\cdot\|}(f)$, $0 \leqslant i \leqslant m$. Снабдим это множество задач нормирующим отображением

$$r_j(f) = \begin{cases} 2\rho_{\|\cdot\|}(G)\,\mathsf{L}_{\|\cdot\|}(f_0), & j = 0, \\ \max\{0, \min_x f_j(x)\} + 2\rho_{\|\cdot\|}(G)\,\mathsf{L}_{\|\cdot\|}(f_j), & j > 0, \end{cases}$$

где $\rho_{\|\cdot\|}(G)$ — $\|\cdot\|$-радиус G (для совместных задач, очевидно, $r_j(f) = 2\rho_{\|\cdot\|}(G)\mathsf{L}_{\|\cdot\|}(f_j)$ при всех j). Поле задач $\boldsymbol{C}_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$, снабженное оракулом $\mathcal{O}$ и указанным нормирующим отображением, будем обозначать $\boldsymbol{C}^{\mathcal{O}}_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$ и называть *классом липшицевых выпуклых задач.*

Очевидно, $r_j(f)$ — естественные оценки сверху величин $\sup_G f_0 - \inf_G f_0$ (при $j = 0$) и $[\max_x f_j(x)]_+$, $j > 0$, т. е. нормирующих множителей задачи f, рассматриваемой как элемент $\boldsymbol{C}(G, E, m)$. Поэтому содержательная интерпретация нормировки — та же, что и выше.

*) При рассмотрении липшицевых выпуклых задач мы всегда будем считать G ограниченным.

2.4. Указанные ограничения гладкости относились лишь к модулям непрерывности f_i. Можно налагать требования типа непрерывной дифференцируемости заданное число раз с ограничениями на модуль непрерывности старшей производной, а также комбинировать их с ограничениями типа строгой выпуклости. Возникающие на этом пути классы гладких выпуклых задач изучаются далее в главе VII. В ближайших трех главах мы сосредоточим внимание на классах общих (и липшицевых) выпуклых задач, решаемых с помощью точного оракула первого порядка.

Сделаем еще одно замечание, существенное для всего дальнейшего. Намеченная классификация задач «по гладкости» содержательно оправдывается не тем, что на практике часто встречаются выпуклые, но не липшицевы, или выпуклые липшицевы, но мало раз дифференцируемые задачи — такое оправдание было бы чистой воды демагогией, ибо сам факт высокой гладкости практической задачи, как правило, не вызывает сомнений. Но надо иметь в виду, что сама по себе гладкость немногого стоит — важны значения числовых параметров, характеризующих эту гладкость (величины соответствующих производных и т. п.).

При «естественной» для данного класса нормировке погрешностей эти параметры явно учитываются (см. приведенные выше описания нормирующих множителей для классов общих и липшицевых выпуклых задач). Кроме того, иногда они непосредственно влияют на сложность класса или даже должны быть оценены априори — без этого решить соответствующие задачи трудно или вовсе невозможно (см., например, ситуацию в главе VII). Таким образом, численные значения параметров, характеризующих гладкость задачи, так или иначе влияют на сложность ее решения. В связи с этим полезно иметь в своем распоряжении методы, чувствительные лишь к наиболее грубым и легко оцениваемым параметрам гладкости задачи. Наша классификация задач как раз и имеет целью выделить наиболее естественные и по возможности «не слишком тонкие» свойства ситуации и построить методы, нечувствительные к численным характеристикам прочих ее свойств. Выбор «существенных» черт ситуации, разумеется, неоднозначен, так что приведенные далее рассмотрения далеко не исчерпывают всего круга возникающих здесь проблем.

Мы надеемся, однако, что дальнейшее убедит читателя в разумности (но, повторяем, не в полноте) избранного пути классификации выпуклых экстремальных задач.

2.5. Опишем теперь оракулы, которые мы будем рассматривать при изучении общих (липшицевых) выпуклых задач. Ими будут детерминированные оракулы первого порядка, сообщающие значения и опорные функционалы к компонентам задачи в рассматриваемой точке. Мы не будем требовать абсолютно точных ответов, разрешая оракулу ошибаться (но требуя, чтобы эта ошибка не превосходила заданной точности решения).

Чтобы продемонстрировать естественность приведенных ниже формальных определений, начнем со случая точного оракула первого порядка. В ответ на вопрос о задаче $f \in \boldsymbol{C}(G, E, m)$ в точке x он должен сообщить значения f_i в точке x и опорные к ним в x (на G) функционалы $\varphi_{i,x} \in \partial_G f_i(\boldsymbol{x})$. Если x не есть внутренняя точка G, то $\partial_G f_i(x)$ может быть пусто. Поэтому при рассмотрении классов общих выпуклых задач естественно ограничиться случаем $\operatorname{int} G \neq \phi$ (в конечномерном случае это ограничение несущественно, так как можно заменить E наименьшим содержащим G аффинным подпространством E) и считать, что область вопросов оракулу $\mathcal{O}$ есть $\operatorname{int} G$, а не все G *). Далее, ответ оракула о f_i в точке x можно рассмотреть как аффинный функционал

$$\bar{g}_{i,x}(y) = f_i(x) + \langle \varphi_{i,x} | y - x \rangle, \ \varphi_{i,x} \in \partial_G f_i(x). \tag{2.1}$$

При этом $\bar{g}_{i,x}(x) = f_i(x)$ и $\bar{g}_{i,x}(y) \leqslant f_i(y)$ при всех $y \in G$. Ясно теперь, что значит «оракул ошибается на величину $\leqslant \nu_0$»: это значит, что он сообщает не $\bar{g}_{i,x}$, а близкий «с точностью ν_0» к $\bar{g}_{i,x}$ аффинный функционал $g_{i,x}$. Оказывается, что «близость» разумно измерять максимальным расхождением $\bar{g}_{i,x}(y)$ и $g_{i,x}(y)$ по $y \in G$ (которую, разумеется, надо нормировать делением на $r_i(f)$). Если вычисленное таким образом расхождение $\bar{g}_{i,x}$ и $g_{i,x}$ не больше ν_0, то переход от $g_{i,x}$ к функционалу $\tilde{g}_{i,x}(y) = g_{i,x}(y) - \nu_0 r_i(f)$ гарантирует выполнение неравенства

$$f_i(y) \geqslant \tilde{g}_{i,x}(y), \tag{2.2}$$

тогда как $\tilde{g}_{i,x}(x)$ по-прежнему близко в $f_i(x)$. По техническим соображениям удобно считать, что ответы оракула удовлетворяют (2.2) с самого начала.

Мы приходим к следующему определению, в котором $\nu_0 \geqslant 0$ — параметр (точность оракула).

Определение. *Оракулом типа* $(\operatorname{int} G, \nu_0)$ *для поля задач* $\boldsymbol{C}(G, E, m)$ называется любой детерминированный оракул с областью вопросов $\operatorname{int} G$ и ответами о задаче f в точках $x \in \operatorname{int} G$ — наборами $m+1$ аффинных функционалов $g^f_{j,x}(y)$, $0 \leqslant j \leqslant m$, обладающий при всех $f \in \boldsymbol{C}(G, E, m)$ и всех $x \in \operatorname{int} G$ следующими свойствами:

$$(1)\qquad \begin{cases} g^f_{i,x}(y) \leqslant f_i(y), & 0 \leqslant i \leqslant m, \ y \in G, \\ g^f_{i,x}(x) \geqslant f_i(x) - \nu_0 r_i(f), & 0 \leqslant i \leqslant m. \end{cases}$$

*) Очевидно, это соглашение не запрещает пользоваться общей схемой определений главы I, где областью вопросов оракулу всегда была вся область G. Действительно, можно считать, что вопросы можно задавать и на границе G, а ответ оракула на такой вопрос есть раз навсегда фиксированная точка информационного пространства (такой ответ на самом деле не содержит никакой информации).

Аналогичные соображения приводят к определению оракула для поля задач $C_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$.

Определение. *Оракулом типа* (G, ν_0) *для поля задач* $C_{\mathrm{Lip}} = C_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$ называется всякий детерминированный оракул с областью вопросов G и ответами о задаче f в точках $x \in G$ — наборами аффинных функционалов $g^f_{i,x}(y)$, обладающий — при всех $f \in C_{\mathrm{Lip}} = C_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$ и всех $x \in G$ — следующими свойствами:

$$(1) \begin{cases} g^f_{i,x}(y) \leqslant f_i(y), & 0 \leqslant i \leqslant m, \ y \in G, \\ g^f_{i,x}(x) \geqslant f_i(x) - 2\nu_0 \rho_{\|\cdot\|}(G) L_{\|\cdot\|}(f_i), & 0 \leqslant i \leqslant m; \end{cases}$$

(2) $g^f_{i,x}(y)$ липшицево по y с константой $L_{\|\cdot\|}(f_i)$.

Второе определение отличается от первого, во-первых, возможностью задавать вопросы во всем G, а не только на $\mathrm{int}\, G$. Кроме того, мы требуем, чтобы норма опорного функционала к f_i, сообщаемого оракулом, не превосходила $L_{\|\cdot\|}(f_i)$ (такой опорный функционал есть в каждой точке G, см. упражнение 15 § 1).

Классы $C^{\mathcal{O}}(G, E, m)$, отвечающие оснащению $C(G, E, m)$ $(\mathrm{int}\, G, \nu_0)$-оракулами, будем называть *классами типа* $C^{\nu_0}(G, E, m)$. Аналогично определяются классы типа $C^{\nu_0}_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$.

В этой и следующей главах мы будем заниматься детерминированными методами решения задач таких классов. Существенно отметить, что предлагаемые методы не будут нуждаться в априорном знании оракула $\mathcal{O}$. Им достаточно знать тип рассматриваемого класса (т. е. список объектов ν_0, G, E, m — для первого типа и ν_0, G, E, $\|\cdot\|$, m — для второго). Фраза «класс C имеет тип $C^{\nu_0}(G, E, m)$ ($C^{\nu_0}_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$)» будет кратко записываться $C \in C^{\nu_0}(G, E, m)$ ($C \in C^{\nu_0}_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$).

Упражнение 1. Пусть G — ограниченное выпуклое замкнутое тело в E. Докажите, что всякий метод решения задач класса типа $C^0(G, E, m)$ естественно индуцирует метод решения задач класса типа $C^0_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$ с теми же (или лучшими) характеристиками трудоемкости и погрешности на классе.

2.6. В заключение обсудим вопрос о целесообразности отнесения имеющейся экстремальной задачи к тому или иному классу таких задач. Заметим, что содержательно одна и та же задача может быть записана в виде (1.4) многими эквивалентными способами. Пусть, например, имеется записанная в виде (1.4) задача, причем f_i непрерывны и выпуклы на всем E. Вводя $f_{m+1}(x) = \rho_{\|\cdot\|}(x, G)$, можем записать эквивалентную задачу, но уже без жестких ограничений:

$$f_0(x) \to \min \mid f_i(x) \leqslant 0, \ 1 \leqslant i \leqslant m+1. \qquad (1.4.1)$$

В новой записи $G = E$.

Другие примеры эквивалентных записей

$$f_0(x) \to \min \mid x \in G, f^1(x) \leqslant 0, (f^1(x) = \max_{1 \leqslant j \leqslant m} f_j(x)), \qquad (1.4.2)$$

$$f_0(x) \to \min \mid \tilde{f}(x) \leqslant 0, (\tilde{f}(x) = \max_{1 \leqslant j \leqslant m+1} f_j(x)), \qquad (1.4.3)$$

$$u \to \min \mid f_0(x) - u \leqslant 0, f_j(x) \leqslant 0, 1 \leqslant j \leqslant m, \quad (x, u) \in E_+, \quad x \in G. \qquad (1.4.4)$$

Заметим, что эти записи эквивалентны лишь в том смысле, что полученные задачи имеют общее решение (точнее, их решения канонически пересчитываются друг в друга). Однако по отношению к намеченной классификации, да и к возможностям численного решения эти задачи вовсе не эквивалентны. Скажем, если исходная задача лежала в классе типа $C^0(G, E, m)$ с ограниченным G, то задачи (1.4.1), (1.4.3), (1.4.4) вообще не лежат в классе такого типа. Далее, если исходная задача была гладкой и мы имели метод ее решения, использующий это свойство, то, скажем, задачи (1.4.2), (1.4.3) могут потерять гладкость, и к такой записи имеющийся метод может быть неприменим.

Указанное обстоятельство не является ни дефектом классификации, ни дефектом численных методов. Последние оперируют не с исходными содержательными задачами и не с классами взаимоэквивалентных их формальных постановок, но с вполне определенными постановками (записями) этих задач. Выбор такой записи является делом математика. Соответственно и приведенная классификация выпуклых задач, предназначенная для анализа численных методов их решения, классифицирует не «содержательные задачи», а их формальные записи.

Подчеркнем, что наиболее важный этап в решении всякой содержательной задачи точными методами — этап формализации, теоретического изучения формальной задачи и редукции ее к задаче, решаемой численными методами,— есть этап сугубо человеческой, творческой деятельности, не поддающейся (и, по-видимому, долго не собирающейся поддаваться) формальному изучению и автоматизации. На этом этапе действует специалист, учитывающий по мере своих способностей специфику конкретной задачи и продвигающийся как можно дальше по пути аналитического ее решения. Численные же методы по самому своему смыслу должны быть универсальными, приспособленными не для решения конкретных задач, но для решения всех задач достаточно широких классов. Цель последующего изучения сложности традиционных классов экстремальных задач — это, если угодно, разработка объективной «шкалы ценностей» этих классов. Руководствуясь ею, можно понять, к чему следует стремиться на первом — «человеческом» — этапе решения содержательных задач.

Приведем пример. В большом числе содержательных задач типа (1.4) жесткие ограничения $x \in G$ либо вовсе отсутствуют ($G = E$), либо «слабы» — множество G неограниченное. Мы видели выше, что при слабых предположениях относительно f всегда можно снять эти жесткие ограничения ((1.4.1) вместо (1.4)). Оказывается, однако, что сложность класса, скажем, $C(\{t \geqslant 0\}, \mathbb{R}, 0)$ бесконечна, так что задачи этого класса не могут «решаться эффективно». Поэтому, если речь идет об общих выпуклых задачах, то не только не следует снимать жесткие ограничения, но, напротив, следует налагать их (если вначале их не было), добиваясь притом ограниченности G. Для этого необходимо иметь априорную оценку нормы решения (т. е. увеличить путем некоторого предварительного исследования априорную информацию). Мы увидим дальше, что не только свойства ограниченности, но даже и геометрия G существенно влияет на сложность соответствующего класса выпуклых задач.

§ 3. Метод центров тяжести для решения общих выпуклых задач

3.1. В этом и следующем параграфах описывается некоторый набор методов решения общих выпуклых задач и оценивается их трудоемкость. Все они — методы первого порядка. Идея этих методов весьма проста. Пусть имеется выпуклая задача f вида (1.4) и задается вопрос о ней в точке $x_0 \in G$. Пусть, кроме того, ответ на вопрос содержит информацию о значениях составляющих вектора $f(x_0)$ и наборе опорных функционалов g_i, $i = 0, 1, \ldots, m$, к функциям f_i в точке x_0. Тем самым получены аффинные функции

$$g^f_{i,x_0} \equiv f_i(x_0) + \langle g_i \mid x - x_0 \rangle$$

такие, что $g^f_{i,x_0}(x) \leqslant f_i(x)$ на G, а $g^f_{i,x_0}(x_0) = f_i(x_0)$. Обозначим через ν требуемую относительную точность решения задачи f (принадлежащей одному из описанных в предыдущем параграфе классов выпуклых задач) и через $r_i(f)$ — нормирующие множители этого класса. Возможно, что $g^f_{i_0,x_0}(x_0) > \nu r_{i_0}(f)$ при некотором $i_0 \geqslant 1$. Проведем через точку x_0 гиперплоскость уровня функции $g^f_{i_0,x_0}(y)$ — обозначим ее Π_{i_0}. Гиперплоскость Π_{i_0} делит G на две части — G_+ и G_-, на одной из которых,— скажем, на G_+ — имеем

$$g^f_{i_0,x_0}(x) \geqslant g^f_{i_0,x_0}(x_0).$$

Стало быть, G_+ заведомо не содержит точек x, могущих быть приближенными решениями f с погрешностью $\nu(x, f) \leqslant \nu$. В самом деле,

при $x \in G_+$ имеем

$$f_{i_0}(x) \geqslant g^f_{i_0,x_0}(x) \geqslant g^f_{i_0,x_0}(x_0) > \nu r_{i_0}(f).$$

Таким образом, в рассматриваемом случае ответ на вопрос позволил исключить из дальнейшего рассмотрения часть G_+ области G: точки этой части заведомо не являются решениями f с требуемой точностью.

Пусть теперь

$$f_i(x_0) \leqslant \nu r_i(f), \quad i = 1, \ldots, m.$$

Тогда сама точка x_0 удовлетворяет (с требуемой точностью) ограничениям задачи. Рассмотрим проходящую через x_0 гиперплоскость Π_0 уровня аффинной функции $g^f_{0,x_0}(x)$. Обозначим, как и выше, через G_+ и G_- части, на которые Π_0 делит G, причем

$$g^f_{0,x_0}(x) \geqslant g^f_{0,x_0}(x_0) \text{ в } G_+.$$

По тем же причинам, что и выше, на G_+ выполняется неравенство $f_0(x) \geqslant f_0(x_0)$. Стало быть, если в G_+ и содержится приближенное решение x' задачи f с точностью $\nu(x', f) \leqslant \nu$, то

$$f_0(x_0) \leqslant f_0(x') \leqslant f_* + \nu r_0(f)$$

(как всегда, f_* — оптимальное значение целевого функционала задачи f). Напомним, что условия

$$f_i(x_0) \leqslant \nu r_i(f), \quad 1 \leqslant i \leqslant m,$$

выполнены по предположению. Итак, если G_+ и содержит приближенные решения f нужной точности, то уже x_0 есть такое решение. Если же G_+ таких решений не содержит, то тем более все они лежат в G_-. Во всех случаях область G_+ можно исключить из дальнейших рассмотрений.

Приведенные рассуждения показывают, что, задав вопрос о f в точке x, можно всегда выделить *) часть G_- области G, обладающую тем свойством, что *если f имеет приближенное решение погрешности $\leqslant \nu$ в G, то f имеет такое решение и в G_-*. При этом G_- есть часть G, отсекаемая некоторой проходящей через x_0 гиперплоскостью Π. Эта гиперплоскость (а, значит, и область G_-) определяется локальной информацией о f в точке x_0.

Таким образом, ответ на вопрос позволяет локализовать решение в области G_- — части области G. Теперь можно принять G_- за новую область определения входящих в задачу функций, заменить x_0 на подходящую точку G_- и повторить описанную выше процедуру, и т. д. При переходе от итерации к итерации очередная область локализации решения становится все меньше и меньше. Ин-

*) Точнее, можно это сделать, если известны $r_i(f)$. Далее придется принять специальные меры, чтобы избавиться от необходимости в этой информации.

туитивно ясно, что локализация решения в «достаточно малой» области позволяет получить результат требуемой точности.

Простая геометрическая идея, описанная выше, и лежит в основе предлагаемых ниже методов решения общих выпуклых задач.

3.2. Метод центров тяжести (МЦТ). Опишем метод решения задач класса типа $C^{\nu_0}(G, E, m)$. Идея метода восходит к [15]. Последующее изложение основано на работе авторов [30]. Метод применим в случае, когда E конечномерно, а G ограничено, выпукло и замкнуто. Эти условия ниже предполагаются выполненными. Вместо E будем писать R^n. Кроме того, ниже считается, что $\operatorname{int} G \neq \phi$. В принятых предположениях последнее не ограничивает общности, так как всегда можно заменить E наименьшим содержащим G аффинным подпространством E (ср. упражнение 6 § 1).

Основанные на описанной выше геометрической идее методы различаются по существу лишь правилом выбора очередной точки (в которой задается вопрос о задаче) в имеющейся к данному моменту области локализации решения. При выборе этого правила следует обеспечить «как можно более быстрое убывание» области локализации решения от шага к шагу. Следует, разумеется, выбрать подходящую меру «величины» областей. Полезной в интересующем нас круге вопросов мерой оказывается лебегов объем области.

Пусть мы имеем очередную область локализации решения $\widetilde{G}$ и выбрали точку $x \in \widetilde{G}$ в качестве «опрашиваемой» на данном шаге. Заранее мы не знаем, какая именно из проходящих через x гиперплоскостей отсечет новую область локализации решения. При неудачно выбранном x (скажем, при $x \in \partial\widetilde{G}$) и «плохой» плоскости сечения (например, касающейся $\partial\widetilde{G}$ при $x \in \partial\widetilde{G}$) новая область локализации может быть лишь немногим меньше имеющейся (в указанном крайнем случае — вообще совпадающей с нею). Интуитивно ясно, что чем больше x похожа на «центр» $\widetilde{G}$, тем больше будет убывание области локализации на шаге и при «самой неблагоприятной» среди проходящих через x гиперплоскостей отсекающей гиперплоскости.

Естественный кандидат на роль «центра» выпуклой области $\widetilde{G}$ — это ее центр тяжести (центр тяжести равномерно распределенной по объему $\widetilde{G}$ массы), т. е. точка

$$x = \int_{\widetilde{G}} y\,dy \Big/ \int_{\widetilde{G}} dy.$$

Метод выпуклого программирования, реализующий приведенную выше геометрическую идею и использующий в качестве очередной опрашиваемой точки центр тяжести полученной к этому моменту области локализации решения, назовем *методом центров тяжести* (МЦТ).

Упражнение. Пусть $\widetilde{G}$ — выпуклое ограниченное тело (последнее означает, что $\operatorname{int} \widetilde{G} \neq \phi$) в R^n, x — центр тяжести $\widetilde{G}$. Покажите, что

1. Определение x не зависит от выбора евклидовой структуры R^n (иными словами, центр тяжести — аффинный инвариант G: если $A: R^n \to R^n$ — обратимое аффинное преобразование, $\tilde{x}$ — центр тяжести $A\widetilde{G}$, то $\tilde{x} = Ax$). Вывести отсюда, что если Δ — симплекс в R^n с вершинами $x_0, \ldots, x_n$, то центр тяжести Δ есть

$$x = \frac{x_0 + \ldots + x_n}{n+1}.$$

2. Если f — аффинный функционал на R^n, то

$$f(x) = \int_{\widetilde{G}} f(y)\,dy \Big/ \int_{\widetilde{G}} dy.$$

3. x есть внутренняя точка $\widetilde{G}$ (использовать п. 2 и утверждения упражнения 23 § 1).

Полезность центра тяжести в интересующем нас вопросе обосновывается следующей геометрической леммой (здесь и далее $|\cdot|_n$ обозначает n-мерный лебегов объем).

3.2.1. **Лемма** [18]. *Пусть G — выпуклое замкнутое ограниченное тело в R^n с центром тяжести x, G_+, G_- — части, на которые делит G проходящая через x гиперплоскость* П. *Тогда*

$$|G_+|_n,\ |G_-|_n \leqslant \left(1 - \left(\frac{n}{n+1}\right)^n\right)|G_n| \leqslant \frac{e-1}{e}\,|G|_n.$$

Геометрический комментарий. Плоскостью, проходящей через центр тяжести выпуклого тела объема 1, нельзя отсечь от него кусок, бо́льший отсекаемого от симплекса того же объема плоскостью, проходящей через центр тяжести параллельно грани (в последнем случае речь идет о куске, прилегающем к этой грани).

Лемма показывает, что выбор в качестве очередной «опрашиваемой» точки центра тяжести построенной области локализации решения действительно гарантирует значительное сокращение этой области — не менее чем в $(e-1)/e = a(\infty)$ раз по объему, где

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

Приведенных соображений достаточно для построения МЦТ. Рассмотрим ограниченное выпуклое замкнутое тело $G \subset R^n$ и класс C типа $C^{\nu_0}(G, R^n, m)$. Пусть $\nu > \nu_0$ — требуемая точность решения задач класса C. Опишем метод МЦТ, настроенный на эту точность. Будем считать $\nu < 1$: всякая точка G является решением любой задачи $f \in C$ с точностью 1, так что принятое допущение не ограничивает общности.

3.2.2. Работа МЦТ, настроенного на точность ν, на задаче $f \in C$ состоит в построении последовательностей $\{x_i \in \text{int}\, G\}_{i=1}^{M_f+1}$ и $\{G_i \subset G\}_{i=0}^{M_f}$. При этом $M_f + 1$ есть число шагов метода на задаче f (оно формируется самим методом), а G_i — выпуклое замкнутое ограниченное тело, лежащее в G при всех $i < M_f$. Множество G_{M_f} выпукло и замкнуто, лежит в G, но не обязано быть телом. При всяком $i \leqslant M_f$ x_i есть центр тяжести G_{i-1}; $G_0 = G$.

Помимо указанных последовательностей, МЦТ строит еще вспомогательную числовую последовательность $\{a_i\}_{i=0}^{M_f}$. Содержательно a_i есть, грубо говоря, рекордное значение целевого функционала решаемой задачи, достигнутое после i шагов; $a_0 = +\infty$.

Ниже будет доказано, что число $M_f + 1$ шагов МЦТ на любой задаче $f \in C$ не превосходит величины $\Phi_{n,\infty}(\nu - \nu_0)$, где *)

$$\Phi_{n,\infty}(\nu) \equiv \left] \frac{n \ln \frac{1}{\nu}}{\ln \frac{1}{a(n)}} \right[+ 2 \leqslant \left] \frac{n}{\ln \frac{e}{e-1}} \ln \frac{1}{\nu} \right[+ 2. \quad (3.1)$$

Работа МЦТ, настроенного на точность ν, $1 > \nu > \nu_0$, на задаче $f \in C$ описывается следующими правилами.

МЦТ0. Начальная настройка.

Положить $G_0 = G$, $a_0 = +\infty$. Перейти к МЦТ1.

МЦТ1. i-й шаг. К i-му ($1 \leqslant i \leqslant M_f$) шагу имеется выпуклое замкнутое ограниченное тело $G_{i-1} \subset G$. На i-м шаге последовательно производятся следующие операции:

МЦТ1. 1. Строится центр тяжести x_i области G_{i-1}.

МЦТ1. 2. В точке x_i задается вопрос **) оракулу о решаемой задаче.

Пусть $g_i^j(y)$ — сообщенные оракулом аффинные функционалы. Положим $r_i^j = \max\{0, \max_{y \in G} g_i^j(y)\}$. Если при всех $j \geqslant 1$ справедливо неравенство $(\max_{s \leqslant i} r_s^j)(\nu - \nu_0) \geqslant g_i^j(x_i)$, положим $j(i) = 0$. В противном случае положим $j(i)$ равным какому-нибудь $j \geqslant 1$, для которого указанное неравенство неверно.

МЦТ1. 3. Положим

$$a_i = \begin{cases} a_{i-1}, & j(i) \neq 0, \\ \min\{g_i^0(x_i), a_{i-1}\}, & j(i) = 0, \end{cases}$$

так что

$$a_i = \min\{g_s^0(x_s) \mid 1 \leqslant s \leqslant i,\ j(s) = 0\}.$$

*) Здесь и далее $]t[$ наименьшее целое, $\geqslant t$.

**) Это возможно, ибо $x_i \in \text{int}\, G_{i-1} \subset \text{int}\, G$.

МЦТ1. 4. Пусть

$$\tilde{G}_i = \{x \in G_{i-1} \mid g_i^0(x) < a_i,\quad g_i^j(x) \leqslant (\nu - \nu_0)(\max_{s \leqslant i} r_s^j),\quad 1 \leqslant j \leqslant m\}.$$

В качестве G_i возьмем замыкание множества $\tilde{G}_i$. Если

$$|G_i|_n < (\nu - \nu_0)^n\, |G|_n, \quad (3.2)$$

перейдем к МЦТ2. В противном случае закончим шаг (т. е. увеличим i на 1 и вернемся к МЦТ1).

МЦТ2. Правило остановки и выдачи результата.

Пусть обращение к МЦТ2 произошло в момент $i = M_f$. Работа метода на этом прекращается выдачей результата

$$\begin{matrix} *, & \text{если } a_{M_f} = \infty \text{ (т. е. для всех } s \leqslant M_f \text{ было } j(s) > 0), \\ x_{i_0}, & \text{если } a_{M_f} < \infty,\quad j(i_0) = 0 \text{ и } g_{i_0}^0(x_{i_0}) = a_{M_f}. \end{matrix}$$

К о м м е н т а р и й. В качестве результата выдается наилучшая (по оценкам $g_i^0(x_i)$ чисел $f_0(x_i)$) среди всех просмотренных точек, удовлетворяющих (опять-таки по оценкам $g_i^j(x_i)$) с нужной точностью ограничениям задачи.

3.2.3. **Теорема.** *При $1 > \nu > \nu_0$ описанный метод, настроенный на точность ν, обеспечивает эту точность на классе C при трудоемкости, не превосходящей $\Phi_{n,\infty}(\nu - \nu_0)$, где $\Phi_{n,\infty}(\nu)$ — функция из* (3.1).

Д о к а з а т е л ь с т в о (в нем, как в первом в последовательности рассуждений одного типа, рекомендуется хорошо разобраться).

1°. Начнем с оценки трудоемкости метода. Из правила формирования $j(i)$ и a_i следует, что при всех $i \leqslant M_f$

$$g_i^{j(i)}(x_i) \begin{cases} \geqslant a_i, & j(i) = 0, \\ > (\nu - \nu_0) \max_{s \leqslant i} r_s^{j(i)}, & j(i) > 0, \end{cases}$$

т. е. $\tilde{G}_i$ есть подмножество части G_{i-1}, отсекаемой от G_{i-1} гиперплоскостью, проходящей через центр тяжести G_{i-1}. В силу сформулированной выше леммы

$$|\tilde{G}_i|_n \leqslant a(n)\, |G_{i-1}|_n, \quad \text{т. е.} \quad |G_i|_n \leqslant a^i(n)\, |G|_n.$$

Стало быть, (2.2) выполнится не позднее, чем при

$$i = \left] n \frac{\ln \frac{1}{\nu - \nu_0}}{\ln \frac{1}{a(n)}} \right[+ 1,$$

что и доставляет требуемую оценку трудоемкости

$$M_f + 1 \leqslant \Phi_{n,\infty}(\nu - \nu_0).$$

2°. Теперь оценим погрешность метода. Фиксируем $f \in C$. Пусть вначале f несовместна. По правилу МЦТ2 результат $\bar{x}$ применения метода к f есть либо (правильный) ответ о несовместности условий задачи, либо точка $x_{i_0^*}$, такая, что для всех $j \geqslant 1$

$$g^j_{i_0}(x_{i_0}) \leqslant (\nu - \nu_0) \max_{s \leqslant i_0} r^j_s \leqslant (\nu - \nu_0) \max \{0, \max_{x \in G} f_j(x)\} \qquad (3.3)$$

(мы учли, что в силу свойств оракула

$$r^j_s = \max\{0, \max_{x \in G} g^j_s(x)\} \leqslant \max\{0, \max_{x \in G} f_j(x)\}).$$

Из (3.3) и свойств оракула

$$f_j(\bar{x}) = f_j(x_{i_0}) \leqslant (\nu - \nu_0)\, r_j(f) + \nu_0 r_j(f) \leqslant \nu r_j(f),$$
$$1 \leqslant j \leqslant m. \qquad (3.4)$$

Итак, для несовместных f имеем $\nu(\bar{x}, f) \leqslant \nu$, что и требуется.

3°. Пусть теперь f совместна и x^* — ее решение. Возможно, что справедливо неравенство

$$a_{M_f} \leqslant f_0(x^*) + (\nu - \nu_0)\, r_0(f). \qquad (3.5)$$

В этом случае метод обеспечивает нужную точность. Действительно, из (3.5) следует, что $a_{M_f} < \infty$, а из МЦТ2 тогда следует, что $\bar{x} \neq *$. Но тогда $\bar{x} = x_{i_0}$, где $j(i_0) = 0$ и $a_{M_f} = g^{\bullet}_{i_0}(x_{i_0})$. Как мы видели, из $j(i_0) = 0$ следует (3.4). Остается проверить, что $f_0(x_{i_0}) \leqslant f_0(x^*) + \nu r_0(f)$. Но это сразу следует из неравенств, вытекающих из (2.5) и определения i_0:

$$f_0(x_{i_0}) \leqslant g^0_{i_0}(x_{i_0}) + \nu_0 r_0(f) \leqslant f_0(x^*) + \nu r_0(f).$$

4°. Докажем теперь, что (3.5) действительно имеет место. Приведем к противоречию противоположное допущение. Пусть

$$a_{M_f} > f_0(x^*) + (\nu - \nu_0)\, r_0(f). \qquad (3.6)$$

Возможно, что $x^* \notin \tilde{G}_{M_f}$. Это значит, что для некоторого i_0 и $j_0 \in \overline{0, m}$ имеем

$$f_{j_0}(x^*) \geqslant g^{j_0}_{i_0}(x^*) \begin{cases} \geqslant a_{i_0}, & j_0 = 0, \\ > (\nu - \nu_0) \max\limits_{s \leqslant i_0} r^{j_0}_s, & j_0 > 0. \end{cases} \qquad (3.7)$$

Из (3.7) следует, что $j_0 = 0$. Действительно, при $j_0 > 0$ $g^{j_0}_{i_0}(x^*) > 0$ в силу (3.7), что невозможно, ибо $f_{j_0}(x^*) \leqslant 0$. Итак, в (3.7) $j_0 = 0$. Но тогда (3.7) дает

$$f_* \geqslant g^0_{i_0}(x^*) \geqslant a_{i_0},$$

что невозможно из-за (3.6). Это значит, что $x^* \in \tilde{G}_{M_f}$, и тем самым $\tilde{G}_{M_f} \neq \phi$.

Рассмотрим тело $\bar{G}$, получающееся подобным сжатием G в $1/(\nu - \nu_0)$ раз к точке x^*:

$$\bar{G} = (\nu - \nu_0)(G - x^*) + x^*.$$

В силу того, что

$$|\tilde{G}_{M_f}|_n < (\nu - \nu_0)^n |G|_n,$$

а $\bar{G}$ имеет объем $(\nu - \nu_0)^n |G|_n$, включение $\tilde{G}_{M_f} \supset \operatorname{int} \bar{G}$ невозможно. Итак, $\operatorname{int} \bar{G} \not\subset \tilde{G}_{M_f}$. Это значит, что существуют ν', $\nu_0 < \nu' < \nu$, и начинающийся в x^* отрезок с концом $z \in G$, такой, что точка $y = (\nu' - \nu_0)(z - x^*) + x^*$ не лежит в $\tilde{G}_{M_f}$. По определению $\tilde{G}_{M_f}$ это означает, что при некоторых $i_0 \leqslant M_f$ и $j_0 \in \overline{0, m}$

$$g^{j_0}_{i_0}(y) \begin{cases} \geqslant a_{i_0}, & j_0 = 0 \\ > (\nu - \nu_0) \max\limits_{s \leqslant i_0} r^{j_0}_s, & j_0 > 0. \end{cases} \qquad (3.8)$$

В то же время

$$g^{j_0}_{i_0}(y) = (1 - (\nu' - \nu_0))\, g^{j_0}_{i_0}(x^*) + (\nu' - \nu_0)\, g^{j_0}_{i_0}(z). \qquad (3.9)$$

При $j_0 > 0$ правая часть (3.9) не превосходит

$$(\nu' - \nu_0)\, r^{j_0}_{i_0} \leqslant (\nu' - \nu_0) \max_{s \leqslant i_0} r^{j_0}_s.$$

Итак, (3.8) в принятом предположении $j_0 > 0$ давало бы

$$(\nu' - \nu_0) \max_{s \leqslant i_0} r^{j_0}_s > (\nu - \nu_0) \max_{s \leqslant i_0} r^{j_0}_s,$$

что невозможно, ибо $\nu' < \nu$.

Значит, $j_0 = 0$. Но тогда (3.8) дает $g^0_{i_0}(y) \geqslant a_{i_0}$ и в силу (3.9)

$$g^0_{i_0}(y) \leqslant (1 - (\nu' - \nu_0))\, f_0(x^*) + (\nu' - \nu_0) \max_{x \in G} f_0(x) \leqslant$$
$$\leqslant f_0(x^*) + (\nu' - \nu_0)\, r_0(f) \leqslant f_0(x^*) + (\nu - \nu_0)\, r_0(f),$$

и так как $g^0_{i_0}(y) \geqslant a_{i_0}$, то

$$a_{M_f} \leqslant a_{i_0} \leqslant f_0(x^*) + (\nu - \nu_0)\, r_0(f),$$

что противоречит (3.6). Полученное противоречие доказывает утверждение.

§ 4. Специальные версии МЦТ

Этот параграф посвящен изложению версий МЦТ для специальных ситуаций: для случая, когда заранее задана не относительная, а абсолютная погрешность решения, и для случая, когда приближенное решение обязано точно удовлетворять ограничениям (т. е. погрешность измеряется в шкале $\nu_*(x, f)$, а не $\nu(x, f)$).

4.1. Начнем с первой из этих двух ситуаций.

МЦТ описан в предположении, что метод должен решить задачи некоторого класса $\boldsymbol{C}$ типа $\boldsymbol{C}^{\nu_0}(G, R^n, m)$ и априори задана требуемая относительная погрешность ν. Иными словами, применение к f МЦТ, настроенного на относительную точность $\nu > \nu_0$, гарантирует получение результата с абсолютными погрешностями $\leqslant \nu r_j(f)$. Если заранее задана не относительная погрешность, требуемая от решения, а абсолютные погрешности ε_j, тогда как величины $r_j(f)$ априори неизвестны, то мы не можем указать ту относительную точность (именно, $\nu(f, \vec{\varepsilon}) = \min\limits_{0\leqslant j\leqslant m} \varepsilon_j / r_j(f)$), на которую требуется настроить метод, чтобы обеспечить заданные абсолютные погрешности.

Таким образом, описанный вариант МЦТ не приспособлен для обеспечения заданных абсолютных погрешностей. Оказывается, его можно модифицировать так, чтобы обойти эту трудность. Опишем соответствующую модификацию. Она получается из исходной следующими заменами ($\varepsilon_j > 0$ означают заранее заданные абсолютные погрешности).

1°. Правило определения $j(i)$ в МЦТ1. 2 заменяется на

$$j(i) = \begin{cases} 0, & \text{если } g_i^j(x_i) \leqslant \varepsilon_j, \quad 1 \leqslant j \leqslant m, \\ \text{любое } j \geqslant 1, & \text{такое, что } g_i^j(x_i) > \varepsilon_j, \\ \text{если такое } j \text{ есть.} & \end{cases}$$

2°. Правило МЦТ1. 4 заменяется на правило

МЦТ1. 4'. Положим

$$r_i = \begin{cases} r_i^{j(i)}, & j(i) > 0, \\ \max\limits_{y\in G} g_i^0(y) - g_i^0(x_i), & j(i) = 0. \end{cases}$$

Пусть

$$\nu_i = \begin{cases} \dfrac{\varepsilon_{j(i)}}{r_i}, & \text{если } j(i) \neq 0, \\ \dfrac{\varepsilon_0}{\varepsilon_0 + r_i}, & \text{если } j(i) = 0. \end{cases}$$

Обозначим

$$\tilde{G}_i = \begin{cases} \{x \in G_{i-1} \mid g_i^{j(i)}(x) \leqslant \varepsilon_{j(i)}\} & \text{при } j(i) > 0, \\ \{x \in G_{i-1} \mid g_i^0(x) < g_i^0(x_i)\}, & j(i) = 0. \end{cases}$$

Пусть, кроме того, $\nu^i = \min\limits_{0\leqslant s\leqslant i} \nu_s$, $G_i = \overline{(\tilde{G}_i)}$ ($\overline{(\cdot)}$ — знак замыкания). Здесь принято $\nu_0 = 1$.

Если $|G_i|_n < (\nu^i)^n |G_0|_n$, перейдем к МЦТ2, в противном случае увеличим i на 1 и перейдем к МЦТ1.

Будет доказано, что устроенный таким образом метод обеспечивает требуемые абсолютные погрешности ε_j в решении задач $f \in \boldsymbol{C}$ «с точностью до неустранимой погрешности оракула». Именно, для результата $\bar{x}$ применения метода к $f \in \boldsymbol{C}$ справедливо неравенство

$$\vec{\varepsilon}(x, f) \leqslant \vec{\varepsilon} + \nu_0 \vec{r}(f), \tag{4.1}$$

где $\vec{\varepsilon} = (\varepsilon_0, \ldots, \varepsilon_m)$, а $\vec{r}(f) = (r_0(f), \ldots, r_m(f))$. Таким образом, погрешности решения с точностью до «неустранимых погрешностей оракула» $\nu_0 r_j(f)$ суть ε_j.

Пусть, скажем, априори известно, что требуемые погрешности превосходят «неустранимые шумы оракула»; для определенности $\varepsilon_j \geqslant 2\nu_0 r_j(f)$. В этом случае можно гарантировать точное выполнение заданных ограничений на погрешности (достаточно применить описанный метод, настроенный на погрешности $\varepsilon_j/2$). Мы теряем такого рода возможность, если погрешность оракула сравнивается с заданной точностью решения и тем более превосходит последнюю. Этот эффект не порочит метод: подобный оракул по очевидным объективным причинам мало приспособлен (или вовсе не приспособлен) для решения задачи с такой высокой точностью.

Трудоемкость описанного метода, как будет доказано, допускает следующую оценку. Пусть $\nu(f, \vec{\varepsilon}) = \min\limits_j \varepsilon_j / r_j(f)$ — максимальная относительная погрешность решения задачи f, обеспечивающая требуемые относительные погрешности, а

$$\bar{\nu}(f, \vec{\varepsilon}) = \frac{\nu(f, \vec{\varepsilon})}{1 + \nu(f, \vec{\varepsilon}) + \nu_0}.$$

Тогда оценка трудоемкости есть $\Phi_{n,\infty}(\bar{\nu}(f, \vec{\varepsilon}))$. Мы видим, что обеспечение заданных абсолютных погрешностей при априори неизвестных $r_i(f)$ «стоит» лишь немногим дороже, чем при известных («цена» в первом случае $\Phi_{n,\infty}(\bar{\nu}(f, \vec{\varepsilon}))$, во втором $\Phi_{n,\infty}(\nu(f, \vec{\varepsilon}))$. Хотя $\bar{\nu}(f, \vec{\varepsilon}) < \nu(f, \vec{\varepsilon})$, но отношение этих величин близко к $1 + \nu_0$ для малых $\nu(f, \vec{\varepsilon})$: соответственно при таких $\nu(f, \vec{\varepsilon})$ близко к 1 и отношение трудоемкостей, отвечающих обеим гипотезам об априорной информации о задаче).

4.1.1. Докажем, что описанный метод действительно удовлетворяет сформулированным утверждениям о погрешности и трудоемкости. Доказательство в основном повторяет приведенное выше. Пусть $f \in \boldsymbol{C}$ — решаемая задача. Начнем с оценки трудоемкости. Из определения ясно, что при тех i, для которых $j(i) > 0$, имеем

$$\nu_i = \frac{\varepsilon_{j(i)}}{r_i} \geqslant \frac{\varepsilon_{j(i)}}{r_{j(i)}(f)} \geqslant \nu(f, \vec{\varepsilon}) \geqslant \bar{\nu}(f, \vec{\varepsilon}).$$

Если же $j(i) = 0$, то

$$\nu_i = \frac{\varepsilon_0}{\varepsilon_0 + \max\limits_{y\in G} g_i^0(y) - g_i^0(x_i)} \geqslant \frac{\varepsilon_0}{\varepsilon_0 + r_0(f)(1+\nu_0)} \geqslant \bar\nu(f, \vec\varepsilon)$$

(мы учли, что

$$\max_{y\in G} g_i^0(y) - g_i^0(x_i) \leqslant \max_{y\in G} f_0(y) - f_0(x_i) + \nu_0 r_0(f) \leqslant (1+\nu_0) r_0(f)).$$

Итак, $\nu_i \geqslant \bar\nu(f, \vec\varepsilon)$, откуда, как и в п. 3.2.3, следует, что трудоемкость метода действительно не превосходит $\Phi_{n,\infty}(\bar\nu(f, \vec\varepsilon))$.

Теперь оценим точность метода. Как и выше, результат $\bar x$ применения метода к f есть либо $*$, либо такая точка G, для которой выполнены неравенства

$$f_j(\bar x) \leqslant \varepsilon_j + \nu_0 r_j(f), \quad 1 \leqslant j \leqslant m. \tag{4.2}$$

Отсюда следует, что при *несовместной* f утверждение о точности метода действительно имеет место. Далее, $\bar x = *$ тогда и только тогда, когда $a_{M_f} = +\infty$. Если $a_{M_f} < \infty$, то $\bar x = x_{i_0}$ при некотором i_0, причем

$$g_{i_0}^0(\bar x) \leqslant a_{M_f}. \tag{4.3}$$

Отсюда ясно, что для доказательства (4.1) достаточно доказать соотношение

$$a_{M_f} \leqslant f_* + \varepsilon_0. \tag{4.4}$$

Докажем (4.4). Для несовместных f оно очевидно. Пусть f совместна и x^* — ее решение. Для каждого $x \in G \setminus \bar G_{M_f}$ существует — по определению $\bar G_{M_f}$ — такое $i = i(x) \leqslant M_f$, что

$$g_i^{j(i)}(x) \begin{cases} \geqslant g_i^{j(i)}(x_i), & j(i) = 0, \\ > \varepsilon_{j(i)}, & j(i) > 0. \end{cases} \tag{4.5}$$

Если $x^* \notin \bar G_{M_f}$, то из (4.5), примененного при $x = x^*$, следует $j(i) = 0$ и $a_{M_f} \leqslant f_*$, что и требуется. Пусть теперь $x^* \in \bar G_{M_f}$. Тогда как и выше, существуют $\nu' < \nu^{M_f}$ и $z \in G$, такие, что точка

$$y = x^* + \nu'(z - x^*)$$

не лежит в $\bar G_{M_f}$. Тогда в силу (4.5) при некотором i

$$g_i^{j(i)}(y) \geqslant \begin{cases} g_i^{j(i)}(x_i), & j(i) = 0, \\ \varepsilon_{j(i)}, & j(i) \neq 0, \end{cases}$$

или

$$g_i^{j(i)}(x^*) + \nu'(g_i^{j(i)}(z) - g_i^{j(i)}(x^*)) \geqslant \begin{cases} g_i^0(x_i), & j(i) = 0, \\ \varepsilon_{j(i)}, & j(i) \neq 0. \end{cases} \tag{4.6}$$

При $j(i) > 0$ (4.6) дает

$$\varepsilon_{j(i)} \leqslant \nu' r_i < \nu_i r_i = \varepsilon_{j(i)},$$

что невозможно. Итак, в (4.6) $j(i) = 0$, и (4.6) дает

$$(1 - \nu')\, g_i^0(x^*) + \nu'\,(g_i^0(z) - g_i^0(x_i)) \geqslant (1 - \nu')\, g_i^0(x_i),$$

или

$$g_i^0(x_i) \leqslant f_* + \frac{\nu'}{1-\nu'} r_i \tag{4.7}$$

(мы учли, что $\nu_i \leqslant 1$ при $j(i) = 0$, так что $\nu' < \nu_i \leqslant 1$). Поскольку $\nu' < \nu_i \leqslant 1$, то $\nu'/(1 - \nu') \leqslant \varepsilon_0/r_i$. Итак,

$$g_i^0(x_i) \leqslant f_* + \varepsilon_0$$

и $a_{M_f} \leqslant f_* + \varepsilon_0$, что завершает доказательство (4.1), а вместе с тем и обоснование свойств рассматриваемого метода.

4.1.2. Замечание. Мы описывали версии МЦТ, приспособленные для решения общих выпуклых задач — задач классов типа $C^{\nu_0}(G, R^n, m)$ с заданной относительной (§ 3) и абсолютной (§ 4) погрешностями. Из обоснования методов легко извлечь, что эти же версии пригодны и для решения липшицевых выпуклых задач (классов типа $C_{\mathrm{Lip}}^{\nu_0}(G, R^n, \|\cdot\|, m)$) с теми же погрешностями (разумеется, в ситуации этого параграфа абсолютные погрешности во всех случаях обеспечиваются с точностью до неустранимых ошибок оракула).

Утверждение о том, что методы, решающие задачи более широкого класса (общие выпуклые), не теряют своих характеристик и на более узком классе липшицевых выпуклых задач, может показаться тавтологией. Здесь есть, однако, некоторая тонкость: класс C типа $C_{\mathrm{Lip}}^{\nu_0}(G, E, \|\cdot\|, m)$ не есть, вообще говоря, подкласс класса типа $C^{\nu_0}(G, E, m)$, если только $\nu_0 > 0$. Действительно, ошибка оракула для «липшицевых» классов измеряется в более грубой шкале, чем для общих выпуклых, так что (G, ν_0) — оракул для поля задач $C_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$, рассматриваемого как подмножество в поле задач $C(G, E, m)$, не всегда может быть расширен на все $C(G, E, m)$ до $(\operatorname{int} G, \nu_0)$-оракула.

4.2. Строго совместные задачи. Метод центров тяжести не обеспечивает, вообще говоря, точного удовлетворения ограничений задачи в найденной точке. Это обстоятельство не является дефектом метода. Как мы увидим в свое время в (1.17) гл. IV, $*$-сложность класса C типа $C^0(G, E, m)$, отвечающего любому детерминированному локальному оракулу, бесконечна при $m \geqslant 1$ и $\nu < 1$ (т. е. никакой конечношаговый метод не может гарантировать отыскания допустимого плана у всякой совместной задачи $f \in C$). Если же заранее известно, что задача f не только совместна, но и имеет «достаточно массивное» множество допустимых планов, то надлежащая модификация МЦТ гарантирует обеспечение требуемой $*$-погрешности. В качестве «меры массивности» множества допустимых планов f естественно принять лебегову n-мерную меру этого множества $(E = R^n)$. Определим в этой связи понятие α-строго совместной задачи.

Определение. $f \in C(G, R^n, m)$ называется *α-строго совместной*, если для множества G_f ее планов имеет место оценка

$$| G_f |_n \geqslant \alpha^n | G |_n.$$

Упражнение 1 (к определению). Пусть $f \in C(G, R^n, m)$ — выпуклая задача, удовлетворяющая условию Слейтера с параметром $\varepsilon > 0$, т. е. при некотором $x \in G$ имеем $f_j(x) \leqslant -\varepsilon r_j(f)$, $1 \leqslant j \leqslant m$. Докажите, что f α-строго совместна с $\alpha = \varepsilon/(1+\varepsilon)$.

Опишем версию МЦТ$_*$ метода МЦТ, пригодную для решения строго совместных задач с заданной $*$-погрешностью. Пусть C — класс задач типа $C^0(G, R^n, m)$, а C_α — подкласс C, содержащий в точности α-строго совместные задачи. Метод МЦТ$_*$ имеет активную остановку (т. е. сам определяет момент окончания работы и выдачи результата). Он применим к любой задаче из $C(G, R^n, m)$. Если задача лежит в C_α при каком-нибудь $\alpha > 0$, то метод выдает результат на ней не позднее, чем через

$$M(\alpha, \nu) = \Phi_{n,\infty}(\alpha\nu) \tag{4.8}$$

шагов (предполагается, что требуемая точность ν, $1 > \nu > 0$, задана заранее), и при этом задача будет решена с $*$-погрешностью, не большей ν. Подчеркнем, что заранее не требуется знать, в каком из классов C_α лежит решаемая задача. Далее считается $m \geqslant 1$.

МЦТ$_*$ состоит из шагов. На i-м шаге строятся области G_i и H_i, $H_i \supset G_i$, а также число a_i. При этом $G_0 = H_0 = G$, $a_0 = +\infty$. На i-м шаге производятся следующие операции.

МЦТ$_*$ 1. Вычисляется центр тяжести x_i области G_{i-1}. В точке x_i задается вопрос о задаче. Пусть $g_i^j(x)$ — сообщенные оракулом аффинные функционалы. Положим

$$j(i) = \begin{cases} 0, & \text{если } g_i^j(x_i) \leqslant 0, \quad 1 \leqslant j \leqslant m, \\ 1 & \text{в противном случае.} \end{cases}$$

Пусть, далее,

$$a_i = \begin{cases} a_{i-1}, & j(i) = 1, \\ \min\{a_i, g_i^0(x_i)\}, & j(i) = 0, \end{cases}$$

$$H_i = \{x \in H_{i-1} \mid g_i^j(x) \leqslant 0, \quad 1 \leqslant j \leqslant m\},$$

$$\tilde{G}_i = \{x \in G_{i-1} \mid g_i^j(x) \leqslant 0, \ 1 \leqslant j \leqslant m, \quad g_i^0(x) < a_i\},$$

$$\tilde{G}_i = \overline{(G_i)}$$

МЦТ$_*$ 2. Вычисляется отношение $| G_i |_n / | H_i |_n$. Если $| G_i |_n \geqslant \nu^n | H_i |_n$, то переходим к следующему шагу. В противном случае работа метода заканчивается и выдается результат x_{i_0}, где i_0 таково, что $j(i_0) = 0$ и $g_{i_0}^0(x_{i_0}) = \min\{g_s^0(x_s) \mid j(s) = 0, s \leqslant M_f\} = a_{M_f}$ (здесь M_f — номер шага остановки).

Видно, что МЦТ$_*$ отличается от исходного метода МЦТ, помимо правила остановки, лишь тем, что правило сокращения области G_{i-1} (правило построения G_i) «отсекает» все точки G_i, в которых $g_i^j(x) > 0$, при каком-нибудь j $1 \leqslant j \leqslant m$, т. е. ограничения заведомо не выполняются, тогда как МЦТ «отсекает» (при $\nu_0 = 0$) лишь те точки, в которых «ограничения не выполняются с заданной точностью».

4.2.1. **Теорема.** *Пусть $\nu < 1$ и $f \in C_\alpha$. Тогда МЦТ$_*$, настроенный на заданную точность, выдает результат $\bar{x}$ не позднее, чем через $M(\alpha, \nu)$ шагов (см. (4.8)), причем $\nu_*(\bar{x}, f) \leqslant \nu$.*

Доказательство. 1°. Пусть f такова, что метод остановит свою работу на f и выдаст результат $\bar{x}$. Убедимся, что тогда $\nu_*(\bar{x}, f) \leqslant \nu$. Пусть x^* — решение задачи f. Заметим прежде всего, что $x^* \in H_{M_f}$, где $M_f + 1$ — трудоемкость метода на f. Действительно, $H_0 = G$ и в $H_i \setminus H_{i-1}$ имеем $\max\limits_{1 \leqslant j \leqslant m} f_j(x) > 0$ по определению H_i. Стало быть, H_{M_f} содержит множество G_f допустимых планов f, а тогда и x^*.

Далее, если метод остановится на шаге M_f, то $| G_{M_f} |_n < | H_{M_f} |_n$. Последнее невозможно, если $a_{M_f} = +\infty$, так что $a_{M_f} < \infty$, и стало быть, $a_{M_f} = \min\{f_0(x_s) \mid x_s \in G_f\}$ (в силу правила построения a_i по a_{i-1} и правила выбора $j(i)$). Ясно, далее, что вне G_i в H_i либо $\max\limits_{1 \leqslant j \leqslant m} f_j(x) > 0$, либо $f_0(x) \geqslant a_i$. Отсюда следует, что вне G_{M_f} в G либо $x \notin G_f$, либо $f_0(x) \geqslant a_{M_f} = f_0(\bar{x})$. По построению $f_j(\bar{x}) \leqslant 0$, $1 \leqslant j \leqslant m$, и остается доказать только, что $f_0(x^*) \leqslant a_{M_f} + \nu r_0(f)$. Пусть это не так, и $f_0(x^*) > a_{M_f} + \nu r_0(f)$. Заметим, что тогда $x^* \in G_{M_f}$, ибо, как мы видели, на $G_f \setminus G_{M_f}$ везде $f_0(x) \geqslant a_{M_f}$. Далее, G_{M_f}, очевидно, допускает описание вида

$$G_{M_f} \supset \{x \in H_{M_f} \mid g_s^0(x) < a_s, \quad s \leqslant M_f\}.$$

Поскольку $| G_{M_f} |_n < \nu^n | H_{M_f} |_n$, то существуют $z \in H_{M_f}$ и $y = \nu z + (1 - \nu) x^* : y \notin G_{M_f}$. Стало быть, при некотором s имеем $g_0^s(y) \geqslant a_s$ и тем более $f_0(y) \geqslant a_s$. Но

$$f_0(y) - f_0(x^*) \leqslant \nu (f_0(z) - f_0(x^*)) \leqslant \nu r_0(f),$$

т. е.

$$a_{M_f} \leqslant a_s \leqslant f_0(y) \leqslant \nu r_0(f) + f_0(x^*).$$

Это противоречит сделанному выше допущению. Итак, $\nu_*(\bar{x}, f) \leqslant \nu$.

2°. Теперь оценим M_f, предполагая, что $f \in C_\alpha$. Выше было замечено, что $H_i \supset G_f$ при всех i, так что $|H_i|_n \geqslant \alpha^n |G|_n$ при всех i. Вместе с тем, как и в доказательстве теоремы 3.2.3, $|G_i|_n \leqslant a^i(n)|G|_n$. Стало быть,

$$|G_i|_n / |H_i|_n \leqslant \frac{a^i(n)}{\alpha^n}$$

и приведенное отношение становится меньше ν^n, как только $i \geqslant M(\alpha, \nu) - 1$. Это значит, что на f метод остановится, притом не позднее, чем через $M(\alpha, \nu)$ шагов. Теорема доказана.

Подобно МЦТ, метод МЦТ$_*$ можно приспособить для отыскания приближенного решения $f \in C$ заданной абсолютной (а не относительной, как выше) $*$-погрешности ε_0. Для этого в МЦТ$_*$ **1**, наряду со всеми описанными выше операциями, надо еще вычислять на каждом шаге i с $j(i) = 0$ величину

$$\nu_i = \frac{\varepsilon_0}{\varepsilon_0 + \sup\limits_{y \in G} g_i^0(y) - g_i^0(x_i)}$$

и $\nu^i = \min\{\nu_s \mid j(s) = 0,\ s \leqslant i\}$ (если при $s \leqslant i$ все время было $j(s) > 0$, то $\nu^i = 1$).

Далее, правило построения $\widetilde{G}_i$ в МЦТ$_*$ 1 заменяется на

$$\widetilde{G}_i = \begin{cases} G_{i-1} \cap H_i, & j(i) = 1, \\ \{x \in G_{i-1} \cap H_i \mid g_i^0(x) < g_i^0(x_i)\}, & j(i) = 0. \end{cases}$$

Кроме того, в МЦТ$_*$ 2 в неравенстве $|G_i|_n \geqslant \nu^n |H_i|_n$, определяющем условие продолжения процесса, следует заменить ν на ν^i.

Так модифицированный метод решает всякую строго совместную задачу из C с абсолютной $*$-погрешностью, $\leqslant \varepsilon$. При этом его трудоемкость на α-строго совместной задаче оценивается (сверху) величиной $M(\alpha, \bar{\nu}(f, \varepsilon_0))$, где

$$\bar{\nu}(f, \varepsilon_0) = \frac{\varepsilon_0}{\varepsilon_0 + r_0(f)} = \frac{\nu(f, \varepsilon)}{1 + \nu(f, \varepsilon)}$$

(здесь $\nu(f, \varepsilon) = \varepsilon_0 / r_0(f)$ — относительная погрешность, обеспечивающая заданную абсолютную).

Обоснование указанных свойств описанной модификации МЦТ предоставляется читателю.

§ 5. Реализуемый вариант МЦТ

5.1. Метод центров тяжести, как метод решения задач классов типа $C^{\nu_0}(G, R^n, m)$, существенно неулучшаем по трудоемкости (в некотором точном смысле, см. главу IV). Однако «вычислительная стоимость» каждого шага метода крайне велика: отыскание центра тяжести, вообще говоря, произвольного выпуклого тела

при сколько-нибудь значительной размерности — более чем сложная задача для самых совершенных вычислительных средств. Поэтому, если МЦТ и может быть использован, то лишь тогда, когда сложность получения информации о задаче весьма велика (например, эту информацию можно получить лишь в дорогостоящих натурных экспериментах), так что экономия в числе вопросов окупает огромную вычислительную сложность шага. В остальных случаях (которыми сегодня только и занимается вычислительная математика) чрезмерная сложность шага ставит неодолимое препятствие на пути использования МЦТ.

Возникает желание так видоизменить этот метод, чтобы упростить до разумного уровня шаг без чрезмерного увеличения оценки трудоемкости. Такое упрощение реализуется ниже. Соответствующий метод, предложенный в работе авторов [31], мы называем *ММЦТ* (*модифицированный метод центров тяжести*).

К идее ММЦТ легко прийти, проанализировав источник вычислительных трудностей в МЦТ — «произвольность форм» строящихся в нем последовательных областей локализации решения G_i. Предположим, что G есть самое простое с точки зрения отыскания центра тяжести тело — эллипсоид (т. е. шар в надлежащих координатах). Тогда первый шаг МЦТ реализуется легко, и G_1 есть полушарие (в указанных координатах). Отыскать центр тяжести полушария все еще довольно легко. Однако G_2 может иметь уже довольно сложную форму и т. д.— вычислительные трудности быстро растут с увеличением номера шага.

Можно, однако, сознательно увеличивать строящиеся области локализации, обеспечивая им достаточно удобную для вычислений форму. Проще всего иметь дело только с эллипсоидами. Чтобы этого добиться, на первом шаге, после построения G_1, заключим это полушарие в эллипсоид $\widetilde{G}_1$ минимально возможного объема и будем считать $\widetilde{G}_1$ новой областью локализации решения. Геометрически ситуация, возникающая к началу второго шага, точно такая же, как и исходная. Таким образом, сознательное игнорирование части полученной на шаге информации (увеличение G_1 до $\widetilde{G}_1$) обеспечивает геометрическую стабильность ситуации на всех шагах и позволяет построить простую рекуррентную схему вычислений.

5.2. Предпошлем описанию ММЦТ одну простую геометрическую лемму

Лемма. *Пусть $V = \{x \in E^n \mid \|x\| \leqslant R\}$ — шар радиуса R в E^n, а $e \in E^n$ — единичный вектор. Полушарие $V_e = \{x \in V \mid \langle x | e\rangle \leqslant 0\}$ можно заключить в эллипсоид W объема $\beta^n(n)|V|_n$, где*

$$\beta(n) = \begin{cases} \dfrac{n}{\sqrt{n^2-1}}\left(\dfrac{n-1}{n+1}\right)^{1/2n}, & n > 1, \\ \dfrac{1}{2}, & n = 1. \end{cases} \tag{5.1}$$

Отметим, что

$$\beta(n) = 1 - \frac{c_n}{n^2}, \quad c_n > 0, \quad c_n \xrightarrow[n\to\infty]{} \frac{1}{2}.$$

Обозначим через O_e систему координат (с тем же, что и исходная декартова система координат O_0 в R^n, элементом объема), в которой W есть шар с центром в 0. Радиус этого шара будет $\beta(n)R$. Преобразование координат из O_e в O_0 задается формулой

$$x = A_{e,n}x' - \frac{R}{n+1}e. \tag{5.2}$$

Здесь

$$A_{e,n} = \left(\varkappa^{n-1} - \frac{1}{\varkappa}\right)P_e + \frac{1}{\varkappa}I,$$

где P_e — матрица ортопроектора на ось вектора e, I — единичная $n \times n$-матрица,

$$\varkappa = \left(\frac{n+1}{n-1}\right)^{-\frac{1}{2n}}. \tag{5.3}$$

Геометрически: координаты O_e получаются из координат O_0 сдвигом начала в точку $(-(R/(n+1))e)$, изменением единицы масштаба вдоль оси e в $\varkappa^{n-1}$ раз и изменением единицы масштаба в ортогональной гиперплоскости в $1/\varkappa$ раз.

Упражнение 1. Доказать лемму. Показать, что построенный в ней эллипсоид есть эллипсоид наименьшего объема, содержащий V_e.

5.3. Модифицированный метод центров тяжести (ММЦТ) применим к задачам классов типа $C^{\nu_0}(G, E, m)$ в случае, когда E конечномерно, а G — ограниченное выпуклое замкнутое тело в $E = R^n$. Далее эти условия предполагаются выполненными. Предполагаются известными некоторый содержащий G эллипсоид W_0 и число β, такое, что

$$|W_0|_n \leqslant \beta^n |G|_n. \tag{5.4}$$

Обозначим через $f_{m+1}(x)$ функцию $\rho_{\|\cdot\|}(X, G)$. Здесь $\|\cdot\|$ — какая-нибудь норма на E. Как и МЦТ, ММЦТ применим к любому классу C типа $C^{\nu_0}(G, R^n, m)$. Если требуемая точность ν решения задач класса C удовлетворяет условию $1 > \nu > \nu_0$, то метод в состоянии обеспечить эту точность.

Метод, настроенный на точность $\nu > \nu_0$ (считаем $\nu < 1$), состоит не более чем из

$$\overline{\Phi}_{n,\infty}(\nu - \nu_0) = \left]\frac{\ln \frac{\beta}{\nu - \nu_0}}{\ln \frac{1}{\beta(n)}}\right[+ 2 \leqslant \left]d_n n^2 \ln \frac{\beta}{\nu - \nu_0}\right[+ 2 \tag{5.5}$$

шагов. Здесь $d_n \xrightarrow[n\to\infty]{} 2$. Число шагов $M_f + 1$ метода на задаче $f \in C$ зависит от f и от требуемой точности ν.

Приведем аналитическое описание ММЦТ, сопровождая его геометрическим комментарием. ММЦТ состоит из шагов, к i-му $(1 \leqslant i \leqslant M_f)$ из которых имеется система координат O_{i-1} на R^n с началом в точке x_i, шар W_{i-1} (в координатах относительно O_{i-1}) радиуса r_{i-1} с центром в x_{i-1} и число a_{i-1}. При этом O_0 есть система координат, в которой описанный в (5.4) эллипсоид W_0 есть шар с центром в 0 и элементы объема во всех системах координат O_{i-1} совпадают.

С аналитической точки зрения к i-му шагу имеется матрица B_{i-1} с определителем 1 (матрица пересчета координат из O_{i-1} в O_0), так что $B_0 = I$ — единичная матрица. Имеется также точка x_i (начало координат системы O_{i-1}) и число $r_{i-1} > 0$ (радиус W_{i-1} в координатах O_{i-1}). Точки R^n будем отождествлять с вектор-столбцами их координат в системе O_0.

Пусть ν, $1 > \nu > \nu_0$, — требуемая точность. Работа метода ММЦТ, настроенного на точность ν, на задачах класса C описывается следующими правилами.

ММЦТ0. Начальная настройка. Положить $a_0 = \infty$, $B_0 = I$, выбрать O_0, x_1, r_0 так, как описано выше. Перейти к ММЦТ1.

ММЦТ1. i-й шаг. К i-му шагу имеются: матрица B_{i-1}, $\det B_{i-1} = 1$, числа $a_{i-1} \leqslant +\infty$, $r_{i-1} \geqslant 0$ и система координат O_{i-1} с началом в точке x_i. При этом B_{i-1} есть матрица пересчета координат из системы O_{i-1} в систему O_0. На i-м шаге последовательно производятся следующие действия.

ММЦТ1.1. Если $x_i \notin \operatorname{int} G$, то строится опорный функционал $\eta_i^{m+1} \neq 0$ к функции $\rho_{\|\cdot\|}(x, G)$ в точке x_i (такой есть, ибо $\operatorname{int} G \neq \varnothing$, см. упражнение 23 § 1). Полагаем $j(i) = m + 1$, $g_i^{m+1}(y) = \rho_{\|\cdot\|}(x_i, G) + \langle \eta_i^{m+1} \mid y - x_i \rangle$ и переходим к ММЦТ1.3. Если же $x_i \in \operatorname{int} G$, то переходим к ММЦТ1.2.

ММЦТ1.2. Зададим в точке x_i вопрос оракулу о решаемой задаче f. Пусть $g_i^j(y)$ — сообщенные им аффинные функционалы, η_i^j — их производные. Положим

$$r_i^j = \max\{0, \max_{y\in G} g_i^j(y)\}, \quad 1 \leqslant j \leqslant m,$$

$$r_i^0 = \max_{z\in G} g_i^0(z) - g_i^0(x_i).$$

Если при всех $j \geqslant 1$ имеем $g_i^j(x_i) \leqslant (\nu - \nu_0) r_i^j$, положим $j(i) = 0$, в противном случае пусть $j(i)$ есть такое $j \geqslant 1$, для которого $g_i^j(x_i) > (\nu - \nu_0) r_i^j$.

ММЦТ1.3. Положим $\eta_i = \eta_i^{j(i)}$ (η_i рассматривается как вектор-столбец своих координат относительно базиса $(R^n)^*$, отвечающего O_0-реперу R^n). Вычислим вектор-столбец $\bar{p}_\iota = B_{\iota-1}^{\mathrm{T}} \eta_i$ («т» — знак транспонирования).

Если $\bar{p}_\iota = 0$, перейдем к ММЦТ1.4. При $\bar{p}_\iota \neq 0$ положим

$$p_\iota = \frac{\bar{p}_\iota}{\sqrt{\bar{p}_\iota^{\mathrm{T}} \bar{p}_\iota}}.$$

Комментарий. x_ι — «центр» очередной области локализации решения, вернее, объемлющего ее эллипсоида $W_{\iota-1}$ (являющегося шаром с центром в 0 в координатах $O_{\iota-1}$); p_ι есть единичный (в координатах $O_{\iota-1}$) вектор, ортогональный (в $O_{\iota-1}$-евклидовой структуре *)) гиперплоскости, отсекающей от $W_{\iota-1}$ очередную область локализации.

ММЦТ1.4. Положим

$$a_\iota = \begin{cases} a_{\iota-1}, & j(\iota) > 0, \\ \min\{a_{\iota-1}, \ g_\iota^0(x_\iota)\}, & j(\iota) = 0. \end{cases}$$

Если $\bar{p}_\iota = 0$, перейдем к ММЦТ2, иначе положим

$$d_\iota = \begin{cases} g_\iota^0(x_\iota) - a_\iota, & j(\iota) = 0, \\ g_\iota^{j(\iota)}(x_\iota) - (\nu - \nu_0) r_\iota^{j(\iota)}, & 1 \leqslant j(\iota) \leqslant m, \\ g_\iota^{m+1}(x_\iota), & j(\iota) = m+1. \end{cases}$$

Пусть

$$r_\iota = \beta(n) \sqrt{r_{\iota-1}^2 - \frac{d_\iota^2}{(\bar{p}_\iota^{\mathrm{T}} \bar{p}_\iota)}}$$

(если $d_\iota^2/(\bar{p}_\iota^{\mathrm{T}} \bar{p}_\iota) \geqslant r_{\iota-1}^2$, переходим к ММЦТ2). Положим

$$x_{\iota+1} = x_\iota - \left(\frac{d_\iota}{\sqrt{(\bar{p}_\iota^{\mathrm{T}} \bar{p}_\iota)}} + \frac{r_\iota}{\beta(n)(n+1)} \right) B_{\iota-1} p_\iota,$$

$$B_\iota = B_{\iota-1} A_{p_\iota, n},$$

где $A_{p_\iota, n}$ — матрица, описанная в лемме 5.2. Если $r_\iota < ((\nu - \nu_0)/\beta) r_0$, перейдем к ММЦТ2. В противном случае закончим шаг (т. е. увеличим ι на 1 и перейдем к ММЦТ1).

Комментарий. В качестве очередной области локализации можно взять область, описываемую — в координатах $O_{\iota-1}$ — как

$$U_\iota = \{x \in W_{\iota-1} \mid \bar{p}_\iota^{\mathrm{T}} x_\iota \leqslant -d_\iota\}$$

*) То есть евклидовой структуре, в которой координаты $O_{\iota-1}$ — декартовы.

(это сразу следует из определения d_ι). U_ι содержится в $O_{\iota-1}$-шаре $\overline{W}_\iota$ радиуса $\bar{r}_\iota = \sqrt{r_{\iota-1}^2 - d_\iota^2/(\bar{p}_\iota^{\mathrm{T}} \bar{p}_\iota)}$ с центром в точке $-d_\iota p_\iota / \sqrt{(\bar{p}_\iota^{\mathrm{T}} \bar{p}_\iota)}$ (все рассуждения — в координатах $O_{\iota-1}$). На самом деле U_ι содержится в полушарии $\bar{U}_\iota$ этого шара, отсекаемом проходящей через его центр $O_{\iota-1}$-ортогонально p_ι плоскостью.

В силу леммы координаты O_ι, матрица перехода от которых к $O_{\iota-1}$ есть $A_{p_\iota, n}$, а начало которых (в координатах $O_{\iota-1}$) есть

$$v_\iota = -\frac{d_\iota p_\iota}{\sqrt{\bar{p}_\iota^{\mathrm{T}} \bar{p}_\iota}} - \frac{r_\iota p_\iota}{n+1},$$

обладают тем свойством, что O_ι-шар W_ι радиуса r_ι с центром в v_ι содержит $\bar{U}_\iota$. Объекты O_ι, r_ι, W_ι и есть объекты, строящиеся на ι-м шаге. При этом $B_\iota = B_{\iota-1} A_{p_\iota, n}$ есть матрица пересчета координат из O_ι в O_0, x_ι — столбец O_0-координат центра W_ι.

ММЦТ2. Пусть обращение к ММЦТ2 произошло при $i = M_f$. Работа метода на этом прекращается выдачей результата:

$*$, если $j(s) > 0$ для всех $s \leqslant M_f$,

x_{ι_0}, где ι_0 таково, что $j(\iota_0) = 0$ и $g_{\iota_0}^0(x_{\iota_0}) = a_{M_f}$.

5.4. Теорема. *Пусть* $C \in C^{\nu_0}(G, R^n, m)$ *и* $1 > \nu > \nu_0$. *Тогда метод ММЦТ, настроенный на относительную точность* ν, *решает с этой точностью все задачи из* C *при трудоемкости, не превосходящей*

$$\overline{\Phi}_{n,\alpha}(\nu - \nu_0) = \left] \frac{\ln \dfrac{\beta}{\nu - \nu_0}}{\ln \dfrac{1}{\beta(n)}} \right[+ 2. \tag{5.5}$$

Замечание. Таким образом, ММЦТ по трудоемкости проигрывает МЦТ в $O(n)$ раз (в асимптотике по $\nu \to 0$). Это плата за «консерватизм» ММЦТ — обеспечение стабильности форм, строящихся методом областей локализации решения. Заметим еще, что оценка трудоемкости ММЦТ зависит от β (которое, очевидно, можно считать не большим $\alpha_{2,n}(G)$ — минимальной асферичности G относительно всевозможных евклидовых нормирований R^n). Известно, что $\alpha_{2,n}(G) \leqslant n$, так что при надлежащем выборе O_0, W_0 можно взять $\beta = n$. Правда, такой выбор может оказаться трудно реализуемым; впрочем, влияние β на трудоемкость — всего лишь логарифмическое. Лучше всего применять ММЦТ, когда G — эллипсоид (т. е. можно взять $W_0 = G$ и $\beta = 1$).

Заметим еще, что апостериори ММЦТ оказался конкретным вариантом общего метода Н. З. Шора градиентного спуска с растяжением пространства [28]. Геометрические соображения, лежащие в его основе, приводят к вполне определенным правилам выбора параметра растяжения и регулировки шага. Они позволяют доказать сходимость без каких-либо существенных дополнительных условий типа имеющихся в [28] (это, правда, относится к конкретному рассматриваемому варианту общего метода Н. З. Шора).

Доказательство теоремы в основном повторяет доказательство теоремы 3.2.3. Докажем вначале оценку трудоемкости. Из правила формирования r_i следует, что $r_i \leqslant \beta(n) r_{i-1}$, $i \geqslant 1$. Отсюда $r_i \leqslant \beta^i(n) r_0$. Но при $i < M_f$ в соответствии с ММЦТ1.4 $r_i \geqslant ((\nu - \nu_0)/\beta) r_0$, так что

$$M_f \leqslant \left] \frac{\ln(\beta/(\nu-\nu_0))}{\ln(1/\beta(n))} \right[+ 1,$$

что и требуется.

Теперь докажем утверждение о точности метода. Как всегда, результат $\bar{x}$ применения ММЦТ к $f \in C$ есть либо $*$ (что бывает тогда и только тогда, когда $a_{M_f} = +\infty$), либо точка $x_{i_0} \in G$, такая, что

$$f_j(x_{i_0}) \leqslant (\nu - \nu_0) r^j_{i_0} + \nu_0 r_j(f) \leqslant \nu r_j(f), \quad 1 \leqslant j \leqslant m, \tag{5.6}$$

и

$$f_0(x_{i_0}) \leqslant a_{M_f} + \nu_0 r_0(f). \tag{5.7}$$

Отсюда ясно, что если f несовместна, то $\nu(\bar{x}, f) \leqslant \nu$. Пусть теперь f совместна, x^* — ее решение. Чтобы доказать неравенство $\nu(\bar{x}, f) \leqslant \nu$, надо доказать, что

$$a_{M_f} \leqslant (\nu - \nu_0) r_0(f) + f_* \tag{5.8}$$

Пусть, напротив, (5.8) не имеет места:

$$a_{M_f} > (\nu - \nu_0) r_0(f) + f_* \tag{5.9}$$

Возможно, что $g_i^{j(i)}$ было постоянно при некотором $i \leqslant M_f$. По определению $j(i)$ это невозможно при $j(i) = m+1$. Если $1 \leqslant j(i) \leqslant m$, то по определению $j(i)$ $g_i^{j(i)}(x_i) > 0$ и $f_{j(i)}(x) > 0$ при $x \in G$, что невозможно, ибо задача f по предположению совместна. Итак, $j(i) = 0$, и если $g_i^0(y)$ постоянна, то $f_* \geqslant g_i^0(x_i) \geqslant a_i \geqslant a_{M_f}$, что противоречит (5.9).

Таким образом, можно считать, что $g_i^{j(i)}(y)$ непостоянно при всех i, $1 \leqslant i \leqslant M_f$. Пусть $G_0 = G$,

$$G_i = \begin{cases} \{x \in G_{i-1} \mid g_i^{j(i)}(x) < g_i^{j(i)}(x_i) - d_i\}, & j(i) \leqslant m, \\ \{x \in G_{i-1} \mid g_i^{j(i)}(x) \leqslant g_i^{j(i)}(x_i) - d_i\}, & j(i) = m+1. \end{cases}$$

Из правил ММЦТ 1.3 — ММЦТ1.4 и комментариев к ним ясно, что $G_i \subset W_i$. Из правила ММЦТ1.3 — ММЦТ1.4 очевидно, что если обращение к ММЦТ2 произошло ввиду

$$\frac{d^2_{M_f}}{\bar{p}^{\mathrm{T}}_{M_f}\bar{p}_{M_f}} \geqslant r^2_{M_f-1},$$

то $\operatorname{int} G_{M_f}$ пусто. Если же этого не произошло, то

$$|G_{M_f}|_n \leqslant |W_{M_f}|_n = \left(\frac{r_{M_f}}{r_0}\right)^n |W_0|_n < (\nu - \nu_0)^n |G|_n$$

(мы учли, что

$$\frac{r_{M_f}}{r_0} < \frac{\nu - \nu_0}{\beta},$$

элемент объема в O_i тот же, что и в O_0, и что $|W_0|_n \leqslant (\beta)^n |G|_n$). Итак, во всех случаях

$$|G_{M_f}|_n < (\nu - \nu_0)^n |G|_n. \tag{5.10}$$

Пусть $|G_{M_f}|_n \leqslant (\nu' - \nu_0)^n |G|_n$, $\nu_0 \leqslant \nu' < \nu$. Возможно, что $x^* \notin G_{M_f}$. Это значит, что при некотором i

$$\left.\begin{aligned} g_i^{j(i)}(x^*) &> g_i^{j(i)}(x_i) - d_i, \quad \text{если } j(i) = m+1, \\ g_i^{j(i)}(x^*) &\geqslant g_i^{j(i)}(x_i) - d_i, \qquad\quad j(i) \leqslant m. \end{aligned}\right\} \tag{5.11}$$

Это соотношение невозможно при $j(i) = m+1$, ибо $x^* \in G$ (см. определение $g_i^{m+1}(y)$). Итак, $j(i) \leqslant m$ и

$$f_{j(i)}(x^*) \geqslant g_i^{j(i)}(x^*) \geqslant g_i^{j(i)}(x_i) - d_i. \tag{5.12}$$

Если $1 \leqslant j(i) \leqslant m$, то правая часть (5.12) не меньше $(\nu - \nu_0) \times r_i^{j(i)} > 0$ (последнее — по определению $j(i)$), так что (5.12) невозможно. Следовательно, $j(i) = 0$, но тогда (5.11) дает $f_* \geqslant a_i \geqslant a_{M_f}$, что противоречит (5.9).

Итак, $x^* \in G_{M_f}$. Так как G_{M_f} открыто в G и $|G_{M_f}|_n \leqslant (\nu' - \nu_0)^n |G|_n$, то найдутся $z \in G$ и $y = (\nu' - \nu_0) z + (1 - (\nu' - \nu_0)) x^*$, такие, что $y \notin G_{M_f}$. Тогда при некотором i выполняется (5.12) с заменой x^* на y (как и выше, $j(i) \neq m+1$). Если $1 \leqslant j(i) \leqslant m$, то (5.12) дает

$$g_i^{j(i)}(y) \geqslant (\nu - \nu_0) r_i^{j(i)} > 0. \tag{5.13}$$

В то же время

$$g_i^{j(i)}(y) = (\nu' - \nu_0) g_i^{j(i)}(z) + (1 - (\nu' - \nu_0)) g_i^{j(i)}(x^*) \leqslant (\nu' - \nu_0) r_i^{j(i)},$$

что противоречит (5.13). Итак, $j(i) = 0$. Но тогда из (5.12) $g_i^0(y) \geqslant a_i$. В то же время

$$\begin{aligned} g_i^0(y) &= (\nu' - \nu_0) g_i^0(z) + (1 - (\nu' - \nu_0)) g_i^0(x^*) \leqslant \\ &\leqslant (\nu' - \nu_0) g_i^0(z) + (1 - (\nu' - \nu_0)) f_0(x^*) \leqslant \\ &\leqslant (\nu' - \nu_0) \left(\max_{z' \in G} g_i^0(z') - f_0(x^*)\right) + \\ &\qquad + f_* \leqslant (\nu' - \nu_0) r_0(f) + f_*. \end{aligned}$$

Следовательно, $a_{M_f} \leqslant a_i \leqslant (\nu' - \nu_0)\, r_0(f) + f_*$, что противоречит (5.9). Полученное противоречие доказывает теорему.

5.5. Как и МЦТ, метод ММЦТ можно приспособить для решения задач $f \in C \in C^{\nu_0}(G, R^n, m)$ с заранее заданными абсолютными погрешностями $\varepsilon_j > 0$, $0 \leqslant j \leqslant m$. Для этого метод модифицируется следующим образом.

Правило формирования $j(i)$ в ММЦТ1.2 заменяется на

$$j(i) = \begin{cases} 0, & \text{если } g_i^j(x_i) \leqslant \varepsilon_j, \quad 1 \leqslant j \leqslant m, \\ j \in \overline{1, m}, & \text{такое, что } g_i^j(x_i) > \varepsilon_j, \text{ если такое } j \text{ есть.} \end{cases}$$

Правило ММЦТ1.4 заменяется на

ММЦТ1.4'. Полагаем

$$a_i = \begin{cases} a_{i-1}, & j(i) > 0, \\ \min\{a_{i-1}, g_i^0(x_i)\}, & j(i) = 0, \end{cases}$$

$$d_i = \begin{cases} 0, & j(i) = 0, \\ g_i^{j(i)}(x_i) - \varepsilon_{j(i)}, & 0 < j(i) \leqslant m, \\ g_i^{m+1}(x_i), & j(i) = m+1. \end{cases}$$

При $\bar{p}_i = 0$ переходим к ММЦТ2. Иначе пусть

$$r_i = \beta(n) \sqrt{r_{i-1}^2 - \frac{d_i^2}{(\bar{p}_i^{\mathrm{T}} \bar{p}_i)}},$$

$$x_{i+1} = x_i - \left(\frac{d_i}{\sqrt{\bar{p}_i^{\mathrm{T}} \bar{p}_i}} + \frac{r_i}{\beta(n)(n+1)} \right) B_{i-1} p_i.$$

Определим

$$r_i^0 = \max_{z \in G} g_i^0(z) - g_i^0(x_i),$$

$$\nu_i = \begin{cases} 1, & j(i) = m+1, \\ \dfrac{\varepsilon_{j(i)}}{r_i^{j(i)}}, & 1 \leqslant j(i) \leqslant m, \\ \dfrac{\varepsilon_0}{\varepsilon_0 + r_i^0}, & j(i) = 0. \end{cases}$$

Пусть $\nu^i = \min\limits_{0 \leqslant s \leqslant i} \nu_s$ $(\nu_0 = 1)$. Если $r_i < (\nu^i r_0)/\beta$, то перейдем к ММЦТ2. В противном случае увеличим i на 1 и перейдем к ММЦТ1.

Читателю предоставляется доказать, что модифицированный таким образом метод ММЦТ обладает следующими свойствами.

(1) Результат его применения к любой задаче $f \in C \in C^{\nu_0}(G, R^n, m)$ в качестве решения f имеет абсолютные погрешности, не превосходящие заданных величин ε_j с точностью до неустранимых

погрешностей оракула (т. е. для результата $\bar{x}$ применения метода к f выполнены условия (4.1)).

(2) Трудоемкость описанного метода при этом не превзойдет $\overline{\Phi}_{n,\infty}(\bar{\nu}(f, \vec{\varepsilon}))$, где

$$\bar{\nu}(f, \vec{\varepsilon}) = \frac{\nu(f, \vec{\varepsilon})}{1 + \nu_0 + \nu(f, \vec{\varepsilon})},$$

а $\nu(f, \vec{\varepsilon}) = \min\limits_j \varepsilon_j / r_j(f)$ — максимальная относительная погрешность, обеспечивающая данные абсолютные погрешности.

Разумеется, обе изложенные версии ММЦТ можно применять к задачам классов типа $C_{\mathrm{Lip}}^{\nu_0}(G, R^n, \|\cdot\|, m)$ точно так же, как и к рассмотренным выше задачам классов типа $C^{\nu_0}(G, R^n, m)$.

5.6. Скажем несколько слов о практической реализации описанного метода. Она требует умения вычислять значения и производные функций $f_{m+1}(x) = \rho_{\|\cdot\|}(x, G)$. По поводу возникающих здесь вопросов отошлем читателя к п. 3.3.4 гл. III. Помимо этого, метод требует хранения в памяти и рекуррентного пересчета $n \times n$-матрицы B_i. В соответствии с леммой 5.2 этот пересчет производится по формуле

$$B_{i+1} = \lambda_n B_i + \mu_n B_i P_{e_i},$$

где λ_n и μ_n — зависящие лишь от n величины, выражения для которых даются (5.2) — (5.3), а P_{e_i} — матрица ортопроектора на единичный вектор-столбец $e_i = (e_i^1, \ldots, e_i^n)^{\mathrm{T}}$. Эта матрица имеет элементы $p_{kl} = e_i^k e_i^l$.

Отметим, что произведение $B_i P_{e_i}$ можно вычислить не за $O(n^3)$ арифметических действий, как обычно, а всего за $O(n^2)$. Действительно, его можно найти так: вычислить вектор $\bar{e}_i = B_i e_i$ и построить матрицу, j-й столбец которой есть $e_i^j \bar{e}_i$. Это и будет $B_i P_{e_i}$. Ясно, что при этом получение B_{i+1} требует $O(n^2)$ арифметических операций. Расход операций на организацию одного шага метода (без учета расхода операций оракулом и сравнений, необходимых для вычисления $j(i)$) будет также величиной $O(n^2)$, так что количество элементарных операций для организации метода точности ν будет $O\left(n^4 \left]\ln \frac{\beta}{\nu}\right[\right)$.

5.7. Мы видели, что методы с линейной сходимостью, описанные в этой главе, имеют при данной точности трудоемкость тем большую, чем больше размерность решаемой задачи. Между тем иногда есть основание надеяться, что функционалы задачи, априори зависящие от большого числа переменных, на самом деле не зависят (или почти не зависят) от подавляющего их числа; более точно, надлежащим (заранее неизвестным) преобразованием системы координат можно достичь того, что компоненты задачи будут зависеть лишь от небольшой доли всех переменных. Если бы мы заранее знали эти координаты, то можно было бы с самого начала редуцировать задачу к задаче меньшей размерности и, в случае применения методов этой главы, в конечном счете выиграть в трудоемкости.

Оказывается, что практически того же эффекта можно достичь, и не зная заранее «существенных» переменных. Именно, методам МЦТ и ММЦТ можно придать «адаптивную к истинной размерности решаемой задачи форму». Соответствующая конструкция описана в работе авторов [20], и мы отсылаем к ней интересующегося читателя. Здесь достаточно указать, что адаптивные версии упомянутых методов применимы к любой (липшицевой выпуклой конечномерной) задаче. Оценка трудоемкости этих версий на классах $C^{\mathscr{G}}_{\mathrm{Lip}}(G, E^n, |\cdot|, m)$ ($|\cdot|$ — евклидова норма) почти такая же, что и у соответствующих неадаптивных методов. Зато при решении задач с малым числом «существенных» переменных трудоемкость адаптивных версий оказывается примерно той же, что и при использовании неадаптивных методов в «истинной» размерности данной задачи.

Несущественно проигрывая описанным в этой главе методам на классе всех выпуклых задач данной размерности, адаптивные версии, таким образом, сильно выигрывают в трудоемкости на подклассах задач малой «существенной» размерности, что и определяет их привлекательность.

Глава III

МЕТОДЫ ЗЕРКАЛЬНОГО СПУСКА

В этой главе описывается общая конструкция, позволяющая построить обширное семейство методов выпуклой оптимизации. Особенность этих методов состоит в том, что их трудоемкость не зависит явно от размерности задачи. Соответственно их целесообразно применять при решении выпуклых задач большой размерности. В § 1 намечается идея конструкции; § 2 содержит некоторые предварительные сведения. В § 3 конструируются методы решения липшицевых, а в § 4 — общих выпуклых задач. Результаты этой главы в основном базируются на работе авторов [21].

§ 1. Идея методов

1.1. Методы решения общих выпуклых задач из предыдущей главы не могут нас вполне удовлетворить ни в практическом, ни в теоретическом отношении. Недостатки первого рода в основном присущи МЦТ и очевидны. Что же касается до теоретической стороны дела, то можно заметить следующее. МЦТ, как мы увидим, существенно неулучшаем по трудоемкости, если речь идет о решении общих выпуклых задач на *параллелепипедах* G. Для произвольных выпуклых тел G он существенно неулучшаем лишь в асимптотике по точности. При этом «точность, начиная с которой устанавливается асимптотика», зависит от аффинных свойств G.

Пусть, например, G — n-мерный эллипсоид. Тогда, как будет доказано в § 1 гл. IV, сложность класса выпуклых задач на G имеет вид $O(1/\nu^2)$, $\nu \geqslant 1/\sqrt{n}$, тогда как оценка трудоемкости МЦТ — $O(n \ln(1/\nu))$. Стало быть, если требуется решать выпуклые задачи на эллипсоидах с фиксированной точностью ν, а размерность эллипсоида достаточно высока ($n \gg 1/\nu^2$), то никаких теоретических оснований применять МЦТ нет. Таким образом, общие выпуклые задачи на эллипсоидах можно решать при трудоемкости, не зависящей вовсе от размерности задачи, тогда как трудоемкость (и, значит, «степень неоптимальности») МЦТ линейно растет с ростом размерности.

Эффект «равномерной по размерности ограниченности сложности» в практическом отношении выглядит весьма заманчиво, по крайней мере, применительно к экономическим задачам. Действительно, обычно требования к точности решения этих задач не слишком велики, тогда как размерность их может быть весьма значительной. Поэтому построение «нечувствительных к размерности» методов выпуклого программирования представляется весьма актуальным. Именно в этом направлении проводится последующий анализ.

1.2. Ясно, что нельзя предложить метод решения общих выпуклых задач с оценкой трудоемкости, не зависящей от размерности, если не сделать определенных гипотез об аффинных свойствах соответствующего тела G. Действительно, мы уже указывали, что сложность класса общих выпуклых задач на параллелепипеде линейно растет с ростом размерности последнего. Таким образом, строя «нечувствительный к размерности» метод выпуклой оптимизации, мы должны как-то выделить нужные аффинные свойства рассматриваемых областей G, понять, какие свойства эллипсоидов обеспечивают равномерную по размерности ограниченность сложности решения выпуклых задач на них. Забегая вперед, дадим ответ на этот вопрос. Этот ответ поначалу, возможно, покажется странным, но после приведенного ниже обсуждения он будет выглядеть вполне естественным.

Эллипсоид можно рассматривать как единичный шар евклидова пространства E с евклидовой нормой $\|\cdot\|_2$. Рассмотрим сопряженное к E, $\|\cdot\|_2$ пространство E^*, $\|\cdot\|_*$ (конечно, оно канонически отождествляется с E, $\|\cdot\|_2$, но мы увидим, что с точки зрения обобщений важно перейти к сопряженному пространству). На E^* есть функция $V(\xi)$ (именно, $1/2\,\|\xi_*^2\|$), обладающая следующим свойством: $V(\xi)$ равномерно дифференцируема *) и растет на бесконечности быстрее любой линейной функции. При этом скорость ее роста на бесконечности и модуль непрерывности производной не зависят от размерности E. Так вот, все дело — в существовании функции с такими свойствами на E^*, $\|\cdot\|_*$.

Уточним это утверждение. Пусть имеется рефлексивное пространство E с нормой $\|\cdot\|$, сопряженное к которому допускает существование функции V, обладающей указанными свойствами. Мы сумеем сопоставить функции V метод решения общих выпуклых задач на шарах E с оценкой трудоемкости, зависящей (помимо точности) лишь от модуля непрерывности V' и от скорости роста V на бесконечности. Заметим, что явно размерность в наших предположениях не фигурирует. Фактически мы налагаем ограничения лишь на аффинный тип G (поскольку именно он определяет возможность реализовать G как единичный шар в подходящей норме).

*) То есть функция $V'(\xi)$ равномерно непрерывна на ограниченных подмножествах E^*.

Последнее замечание крайне существенно для понимания идейной стороны всего последующего. Действительно, мы начинаем с общих выпуклых задач — класса, определенного в чисто аффинных терминах, безотносительно к каким-либо нормам, тогда как все дальнейшее изложение будет вестись на языке норм. В этих терминах будут описаны и предлагаемые ниже методы. Не следует удивляться такой несогласованности. Нормирование будет в некотором роде лишь инструментом. В своем месте будет объяснено, что возможность применения этого инструмента для решения общих выпуклых задач на теле $G \subset R^n$ связана с аффинными свойствами G.

1.3. Наметим план действий. Мы начнем с рассмотрения классов *липшицевых* выпуклых задач на ограниченных выпуклых G в нормированных пространствах E, $\|\cdot\|$. При определенных допущениях относительно E, $\|\cdot\|$ (фактически они сформулированы выше) мы научимся решать указанные задачи с любой точностью при трудоемкости, зависящей лишь от этой точности, констант Липшица решаемых задач, диаметра G в рассматриваемой норме и некоторых характеристик $(E, \|\cdot\|)$, в которые размерность E явным образом не входит (она, скажем, может быть и бесконечной). После этого мы научимся р[illegible]щие выпуклые задачи на телах G. В оценку трудоемкости при этом войдет еще и асферичность G в рассматриваемой норме E. Заметим, что в этой главе пространство E не предполагается непременно сепарабельным.

По поводу намеченной программы можно было бы высказать следующее возражение. Бесконечномерные задачи с вычислительной точки зрения — фикция. Численные методы реально работают лишь с конечномерными задачами. В конечномерном же случае все нормы эквивалентны друг другу и, стало быть, эквивалентны евклидовой норме. Казалось бы, достаточно рассмотреть евклидов случай, а более общие (и, значит, более громоздкие и менее ясные) рассмотрения — то все от лукавого!

Приведенное рассуждение, однако, обманчиво. Чтобы пояснить это, рассмотрим вначале случай липшицевых задач. Сам факт липшицевости выпуклой функции f на ограниченном выпуклом замкнутом теле $G \subset R^n$ не зависит от выбора нормы $\|\cdot\|$ на R^n, задающей метрику, но $\rho_{\|\cdot\|}(G)$ — радиус G, т. е. радиус минимального содержащего G шара с центром в G и константа Липшица $L_{\|\cdot\|}(f)$ функции f существенно зависят от выбора нормы. Трудоемкость ассоциированного с данной нормой $\|\cdot\|$ метода, доставляемого предлагаемым подходом и обеспечивающего *абсолютную погрешность* ε, имеет вид $\Phi(\|\cdot\|;\ L_{\|\cdot\|}(f)\,\rho_{\|\cdot\|}(G)/\varepsilon)$. При изменении $\|\cdot\|$ меняются оба аргумента Φ и «наилучшая» норма вовсе не обязана быть евклидовой *).

*) Хотя $\Phi(\|\cdot\|, t)$ при *фиксированном* t достигает минимума по $\|\cdot\|$ действительно для евклидовых норм.

Приведем довольно эффектный пример. Пусть решается задача

$$f(x) = \sum_{i=1}^{n} f_i(x_i) \to \min | x_i \geqslant 0, \ \sum_{i=1}^{n} x_i = 1,$$

$f_i(x)$ — липшицевы с константой 1 выпуклые функции на оси. Задачи такого рода, в некотором смысле, весьма типичны. Для решения можно было бы применить метод, ассоциированный с евклидовой нормой $\| x \|_2 = \sqrt{\Sigma x_i^2}$ (им, кстати, будет обычный градиентный метод). Константа Липшица f в этой норме будет $\sqrt{n}$ (и, вообще говоря, не будет меньшей), а диаметр области определения равен $\sqrt{2}$. При этом $\Phi(\|\cdot\|_2, t) = O(1/t^2)$, так что «евклидов» метод абсолютной точности ε будет иметь оценку трудоемкости $O(n/\varepsilon^2)$. Между тем, метод той же точности, ассоциированный с нормой $\| x \| = \sum_{i=1}^{n} | x^i |$, будет в указанных условиях иметь оценку трудоемкости $O(\ln n/\varepsilon^2)$ — выигрыш, как мы видим, заметный. С ситуацией такого рода мы встретимся в главе VI. Там станет особенно ясно, что *мы не зря намереваемся выйти за рамки евклидовой ситуации*.

В случае общих выпуклых задач ситуация примерно та же. Ассоциированный с нормой $\|\cdot\|$ метод решения общих выпуклых задач на теле G имеет оценку трудоемкости вида $\Phi(\|\cdot\|, \alpha_{\|\cdot\|}(G)/\nu)$, где ν — требуемая относительная точность, $\alpha_{\|\cdot\|}(G)$ — асферичность G в норме $\|\cdot\|$. Влияние на трудоемкость величины $\alpha_{\|\cdot\|}(G)$ может привести к тому, что лучшая оценка будет соответствовать норме, отличной от евклидовой. Например, если G — «гипероктаэдр» $\{x \in R^n \mid \sum_{i=1}^{n} | x^i | \leqslant 1\}$, то лучший «евклидов» метод решения общих выпуклых задач на G точности ν будет иметь трудоемкость $O(n/\nu^2)$, тогда как метод, отвечающий $\|\cdot\|_1$,— лишь $O(\ln n/\nu^2)$. Представляется, что эти аргументы должны убедить читателя в том, что общность последующих рассмотрений не является самоцелью.

1.4. *Теперь опишем идею предлагаемой конструкции методов.* Проще всего понять суть дела, рассматривая задачу минимизации выпуклой липшицевой функции f на банаховом пространстве $(E, \|\cdot\|)$. Предположим, что на сопряженном к E пространстве E^*, $\|\cdot\|_*$ определена функция $V(\xi)$, равномерно дифференцируемая и растущая на бесконечности быстрее любой линейной функции. Предположим, далее, что f имеет точку минимума x^*, а E рефлексивно. Мы намерены сопоставить V метод минимизации f. Чтобы упростить изложение идеи конструкции, опишем «непрерывный аналог» метода (сам метод получается отсюда естественной дискретизацией).

Пусть $\varphi \in E^*$. Тогда $V'(\varphi)$ — линейный функционал на E^*, т. е. элемент E (ибо E рефлексивно). Стало быть, можно рассмотреть траекторию $\varphi(t)$ дифференциального уравнения

$$\frac{d\varphi}{dt} = -f'(V'(\varphi(t))), \quad \varphi(0) = \varphi_0. \tag{1.1}$$

Отображение V' переводит E^* в E, а f' — E в E^*, так что правая часть есть элемент E^*.

Наряду с траекторией $\varphi(t)$ рассмотрим ее «тень» $x(t) = V'(\varphi(t))$. Введем функцию $V_*(\varphi) = V(\varphi) - \langle \varphi \mid x^* \rangle$ и убедимся, что она убывает вдоль траектории $\varphi(t)$. Действительно,

$$\frac{dV_*(\varphi(t))}{dt} = \langle \varphi'(t) \mid V'(\varphi(t)) - x^* \rangle = -\langle f'(V'(\varphi(t))) \mid V'(\varphi(t)) - x^* \rangle = \\ = \langle f'(x(t)) \mid x^* - x(t) \rangle \leqslant f(x^*) - f(x(t)) \leqslant 0.$$

Видно, далее, что если $\varliminf_{t\to\infty} f(x(t)) \neq f(x^*)$ (т. е. если траектория $x(t)$ все время плохо приближает x^* по функционалу), то $dV_*(\varphi(t))/dt \leqslant -\varepsilon < 0$, так что $V_*(\varphi(t))$ убывает неограниченно. Но это невозможно: ведь $V(\varphi)$ растет на бесконечности быстрее любой линейной функции, так что $V_*(\varphi)$ ограничена снизу на E^*. Итак, $\varliminf_{t\to\infty} f(x(t)) = f(x^*)$ — метод при надлежащем правиле формирования результата в момент t сходится по функционалу. Теперь легко «дискретизировать» это рассуждение и сделать его «конструктивным», т. е. явно указать, после какого числа шагов дискретизированного метода будет достигнута точность ν. Можно видеть, что это число шагов зависит от константы Липшица f, модуля непрерывности V' в подходящем шаре и ν (эти величины в совокупности определяют шаг дискретизации, не приводящий к «опасному» для наших целей расхождению траекторий «дискретного» и «непрерывного» методов), а также от $\| x^* \|$ и скорости роста V на бесконечности. (Эти величины определяют возможное минимальное значение V_*, т. е. время, в течение которого $V_*(\varphi(t))$ могла бы быстро убывать.)

Мы не будем сейчас проводить соответствующего анализа; он в более общем виде будет проведен дальше. При этом мы учтем и наличие ограничений. При построении методов простые соображения, которые мы изложили, будут, конечно, маскироваться рядом второстепенных деталей, так что сами описания могут показаться громоздкими. Авторы, однако, не имеют возможности подробно объяснять природу той или иной детали. При желании читатель сможет сам пройти путь от элементарных исходных посылок до результирующих схем.

Возвращаясь к идее конструкции, заметим, что главное движение происходит фактически в сопряженном пространстве. Движение в E является лишь «тенью» главного, что и определило выбранное название — *методы зеркального спуска*. Искушенный читатель сразу заметит, что $V(\varphi)$ (точнее, $V_*(\varphi)$) есть просто функция Ляпунова ассоциированного с V метода. Использование функций Ляпунова — традиционный способ обоснования сходимости итеративных процедур. Однако обычно процедура уже имеется, и для нее ищется функция Ляпунова. Мы действуем в обратном порядке — выбираем функцию Ляпунова (ею, оказывается, может быть почти любая функция на E^*), а уж по ней строим метод.

Посмотрим теперь, во что превращается предложенная конструкция, когда $E, \|\cdot\|$ гильбертово. В этом случае $E^*, \|\cdot\|_*$ канонически отождествляется с E, и можно взять $V(\varphi) = {}^1/_2\,\varphi^2$, т. е. $V'(\varphi) = \varphi$. Процесс (1.1) превращается в процесс

$$\frac{dx(t)}{dt} = -f'(x(t)), \quad x(0) = x_0,$$

т. е. в обычный градиентный спуск. Функция V_*, убывающая вдоль траектории спуска, есть — с точностью до константы — ${}^1/_2\,(x - x^*)^2$. Сходимость обычного градиентного метода как раз и доказывают, отслеживая поведение этой функции (т. е. расстояния до x^*) вдоль траекторий (имеется в виду минимизация выпуклых, не обязательно гладких функций).

Если попытаться непосредственно распространить такого рода схему обоснования градиентного метода на пространства $E, \|\cdot\|$, отличные от евклидовых, ничего не получится — на E ничего похожего на убывание расстояния до x^* не будет. Таким образом, «случайное» с общей точки зрения обстоятельство — то, что в гильбертовой ситуации стандартная функция Ляпунова градиентного метода определена на E, а не на E^* (ибо E отождествляется с E^*) и имеет простой геометрический смысл, только затемняет истинную природу градиентного метода.

В заключение заметим следующее. Приведенная схема имеет, на первый взгляд, колоссальную свободу — выбор функции V. Например, в евклидовой ситуации можно было бы наряду с обычным градиентным методом рассмотреть и метод, ассоциированный, скажем, с функцией ${}^1/_2\,x^2 + \cos^2|x|$. Ситуация, таким образом, чревата неким хаосом. Ничего страшного, однако, не произойдет. В тех случаях, когда мы доведем общую схему до конкретных методов, их право на существование будет основано на теоретической субоптимальности (доказываемой сравнением трудоемкости методов с нижними оценками сложности соответствующих классов). Правда, метод, ассоциированный с функцией ${}^1/_2\,x^2 + \cos^2|x|$ с точностью до абсолютной константы, столь же эффективен, что и обычный градиентный метод, отвечающий ${}^1/_2\,x^2$. В ситуациях такого рода разумно предпочитать более «естественные» методы, что мы и делаем.

Теперь мы кончаем несколько затянувшееся «идейное введение» и переходим к делу. Начнем со специального изучения устраивающих нас пар $(E, \|\cdot\|)$.

§ 2. Регулярные пространства

В этом параграфе мы подвергнем специальному изучению класс подходящих для наших целей нормированных пространств.

2.1. Пусть E — вещественное банахово пространство с нормой $\|\cdot\|$, а E^*, $\|\cdot\|_*$ — сопряженное пространство.

Определение. Пространство $E, \|\cdot\|$ называется *регулярным*, если

(1) E рефлексивно;

(2) существует равномерно дифференцируемая функция $V: E^* \to \mathbb{R}$, такая, что $V(\xi) \geqslant \theta(\|\xi\|_*)$, $\theta(t) \underset{t\to\infty}{\longrightarrow} \infty$.

Предостережение. Не следует думать, что требование (2) всегда автоматически выполнено. Это, разумеется, так в конечномерном случае (кроме того, всякое конечномерное пространство удовлетворяет и (1), так что оно регулярно). В бесконечномерных ситуациях (2) может быть не выполнено, даже если выполнено (1).

Для последующего изложения удобно определенным образом нормировать функции V, участвующие в определении регулярных пространств.

Определение. Пусть $E, \|\cdot\|$ — регулярное пространство. Функция $V: E^* \to \mathbb{R}$ называется *регулярной функцией*, отвечающей $E, \|\cdot\|$ (или $(E, \|\cdot\|)$-*регулярной*), если

(i) $V(\varphi) \geqslant 0$, $V(0) = 0$;

(ii) $\|\varphi\|_* \leqslant \frac{1}{2} + V(\varphi)$;

(iii) $V(\varphi)$ имеет равномерно непрерывную и ограниченную во всем E^* производную $V'(\varphi)$.

Можно доказать, что если $E, \|\cdot\|$ — регулярное пространство, то всегда существуют отвечающие $E, \|\cdot\|$ регулярные функции.

Приведем один (как мы вскоре увидим — универсальный) способ построения регулярных функций. Пусть $\|\cdot\|$ — такая норма на E, что функция $\|\varphi\|_*$ равномерно дифференцируема в окрестности единичной сферы E^* (т. е. существует функция $p(\xi): E^* \to E$, такая, что

$$\frac{1}{t}\,|\,\|\varphi + t\xi\|_* - \|\varphi\|_* - t\langle\xi\,|\,p(\varphi)\rangle\,| \underset{t\to 0}{\longrightarrow} 0$$

равномерно по φ, ξ с $\|\varphi\|_* = \|\xi\|_* = 1$). В этом случае в качестве $(E, \|\cdot\|)$-регулярной функции можно взять

$$V_{E,\|\cdot\|}(\varphi) = \begin{cases} \frac{1}{2}\|\varphi\|_*^2, & \|\varphi\|_* \leqslant 1, \\ \|\varphi\|_* - \frac{1}{2}, & \|\varphi\|_* \geqslant 1. \end{cases} \tag{2.1}$$

Упражнение 1. Докажите, что в сделанных выше предположениях (2.1) действительно определяет $(E, \|\cdot\|)$-регулярную функцию.

Отметим, что свойство $E, \|\cdot\|$ быть регулярным пространством есть, очевидно, свойство топологии E, а не нормы (т. е. оно сохраняется, если заменить $\|\cdot\|$ на эквивалентную ей $\|\cdot\|'$). Напротив, свойство функции V быть $(E, \|\cdot\|)$-регулярной зависит от конкретной нормы на E, а не от топологии E.

2.2. Цель этой главы состоит в описании конструкции, сопоставляющей $(E, \|\cdot\|)$-регулярным функциям V методы выпуклой оптимизации (мы их будем называть ЗС-*методы*, от названия «зеркальный спуск»). Характеристики ассоциированных с V ЗС-методов зависят от определенных свойств V. Эти свойства мы сейчас и опишем.

Пусть $(E, \|\cdot\|)$ — регулярное пространство, $V(\varphi)$ — соответствующая регулярная функция. Рассмотрим отображение $\varphi \to \to V'(\varphi)$, заданное на E^*. $V'(\varphi)$ есть непрерывный линейный функционал на E^*, т. е. элемент $(E^*)^*$. Так как E рефлексивно, то $(E^*)^*$ канонически отождествляется с E, так что мы будем всегда считать, что $V'(\varphi)$ принимает значения в E. Функция $V'(\varphi): E^* \to E$ по условию равномерно непрерывна, так что можно говорить о ее модуле непрерывности

$$\omega_V(t) = \sup\{\| V'(\varphi) - V'(\psi)\| \mid \varphi, \psi \in E^*, \|\varphi - \psi\|_* \leqslant t\}.$$

Функцию $\omega_V(t)$ будем называть *модулем гладкости* V. Она будет играть основную роль в оценке трудоемкости ассоциированного с V ЗС-метода. Из определения ясно, что $\omega_V(t)$ — неубывающая неотрицательная ограниченная на полуоси $t \geqslant 0$ функция, непрерывная и стремящаяся к 0 при $t \to +0$.

Существенную роль в дальнейшем будет играть «обратная» к $\omega_V(t)$ функция

$$\gamma_V(s) = \sup\{t \mid \omega_V(t) \leqslant s\}.$$

Очевидно, $\gamma_V(s) > 0$ при $s > 0$, $\gamma_V(s)$ не убывает с ростом s и $\omega_V(\gamma_V(s)) \leqslant s$.

Трудоемкость ассоциированного с V метода $O(1/\nu\gamma_V\cdot(\nu))$, т. е. она тем меньше, чем больше $\gamma_V(t)$, и, стало быть, тем меньше, чем меньше $\omega_V(t)$ (чем «глаже» V). Соответственно мы будем обращать внимание на оценку $\omega_V(t)$ снизу.

2.3. Исследуем теперь вопрос о запасе регулярных пространств. Оказывается, все они могут быть описаны в стандартных терминах функционального анализа. Именно, банахово пространство E регулярно тогда и только тогда, когда топология на E может быть задана с помощью *равномерно-выпуклой* нормы $\|\cdot\|$. Норма $\|\cdot\|$ называется *равномерно-выпуклой*, если стрелка не слишком малой хорды единичного шара сама не слишком мала. Более точно: для некоторой неубывающей положительной при $s > 0$ функции $\varkappa(s)$

должно выполняться соотношение

$$\left\|\frac{x+y}{2}\right\| \leqslant 1 - \varkappa(\|x-y\|)$$

при всех x, y с $\|x\| = \|y\| = 1$. Примером нормы такого типа может служить обычная гильбертова норма.

Известно (теорема Шмульяна, [8, гл. 7]), что норма $\|\cdot\|$ на рефлексивном пространстве E равномерно-выпукла тогда и только тогда, когда двойственная норма $\|\cdot\|_*$ на E^* равномерно-дифференцируема в окрестности единичной сферы E^* (см. п. 2.1). Таким образом, в качестве регулярной функции, отвечающей равномерно-выпукло нормируемому E с равномерно-выпуклой нормой $\|\cdot\|$, можно взять $V_{E,\|\cdot\|}(\cdot)$. Обратно, можно доказать, что если E регулярно, то E допускает введение равномерно-выпуклой нормы.

Приведем теперь конкретные примеры регулярных пространств. Прежде всего гильбертово пространство регулярно (как уже говорилось, его естественная норма равномерно-выпукла; впрочем, регулярную функцию для гильбертова пространства легко указать непосредственно). Следовательно, все конечномерные пространства регулярны (они допускают задание топологии с помощью евклидовой нормы). Приведем еще одно семейство примеров: стандартные пространства $L_p(T, \mu)$.

Описание этих пространств дано в п. 2.8 Приложения. Там объясняется, что при $1 < p < \infty$ $L_p(T, \mu)$ рефлексивно. Более того, обычная норма $\|\cdot\|_p$ на нем при этих p равномерно-выпукла, так что при $1 < p < \infty$ пространства $L_p(T, \mu)$ регулярны. В качестве отвечающих им регулярных функций можно взять ($q = = p/(p-1)$)

$$V_{L_p, \|\cdot\|_p}(\varphi) = \begin{cases} \dfrac{1}{2}\|\varphi\|_q^2, & \|\varphi\|_q \leqslant 1, \\ \|\varphi\|_q - \dfrac{1}{2}, & \|\varphi\|_q \geqslant 1. \end{cases} \tag{2.2}$$

Эти функции будем обозначать просто $V_p(\cdot)$.

Модуль гладкости $\omega_{V_p}(t) \equiv \omega_p(t)$ и обратная к нему функция $\gamma_p(t) \equiv \gamma_{V_p}(t)$ вычислены в § 5. Там показано, что функция $\gamma_{V_p}(t)$ допускает следующую оценку снизу:

$$\gamma_p(t) \geqslant \begin{cases} c(p)\,t, & 1 < p \leqslant 2, \\ c(p)\,t^{p-1}, & p > 2. \end{cases} \tag{2.3}$$

Здесь $c(p) > 0$ зависит лишь от p. Вид этой функции указан в § 5.

При $p = 1$ (а также и при $p = \infty$) пространство $L_p(T, \mu)$ нерефлексивно (и, стало быть, нерегулярно), если только его линейная размерность бесконечна. Если же она конечна и равна n, $1 \leqslant n < \infty$, то эти пространства, разумеется, регулярны. Нас будет специально интересовать конечномерное пространство L_1, т. е. пространство $l_1^{(n)}$, $1 \leqslant n < \infty$ (см. п. 2.8 Приложения).

Сопряженное к $l_1^{(n)}$ есть $l_\infty^{(n)}$. В качестве $l_1^{(n)}$-регулярной функции можно взять любую из функций

$$V^\tau(\varphi) = \begin{cases} \frac{1}{2}\|\varphi\|_\tau^2, & \|\varphi\|_\tau \leqslant 1, \\ \|\varphi\|_\tau - \frac{1}{2}, & \|\varphi\|_\tau \geqslant 1, \end{cases} \tag{2.4}$$
$$1 < \tau < \infty.$$

Разумно выбирать τ с тем, чтобы максимизировать гладкость V^τ. Соответствующий анализ проведен в § 5. Он приводит к выбору τ в виде $c \ln n$ (считаем $n > 1$, c — абсолютная константа). Для соответствующей функции $V \equiv V_{1,n}(\cdot)$ справедлива следующая оценка:

$$\gamma_{V_{1,n}}(s) \geqslant c(1)\frac{s}{\ln n} \tag{2.5}$$

с абсолютной константой $c(1) > 0$. Таким образом, характеристики гладкости $V_{1,n}$, как и следовало ожидать, портятся с ростом n, но достаточно медленно.

2.4. В заключение сформулируем элементарную оценку приращения функции V, весьма полезную в дальнейшем.

Предложение. Пусть E, $\|\cdot\|$ — регулярное пространство и V — отвечающая E, $\|\cdot\|$ регулярная функция. Введем функцию

$$V_*(\varphi) = V(\varphi) - \langle\varphi|x\rangle, \quad x \in E.$$

Пусть, далее, φ, $\xi \in E^*$. Тогда

$$V_*(\varphi + \xi) \leqslant V_*(\varphi) + \langle\xi|V_*'(\varphi)\rangle + \|\xi\|_*\omega_V(\|\xi\|_*). \tag{2.6}$$

Упражнение 2. Докажите (2.6).

§ 3. ЗС-методы на классах липшицевых выпуклых задач

В этом параграфе строятся методы зеркального спуска решения выпуклых липшицевых задач на регулярных пространствах.

Пусть E, $\|\cdot\|$ — регулярное пространство, G — ограниченное выпуклое замкнутое множество в E радиуса $\rho_{\|\cdot\|}(G)$, m — неотрицательное целое. Мы будем заниматься методами решения классов липшицевых выпуклых задач — *классов типа* $C_{\mathrm{Lip}}^{\circ}(G, E, \|\cdot\|, m)$. Соответствующие понятия были введены в § 2 предыдущей главы.

3.1. Обозначим через $V(\cdot)$ $(E, \|\cdot\|)$-регулярную функцию. Опишем ассоциированный с ней метод зеркального спуска (назовем его ЗС$_V$) решения задач класса $C_{\mathrm{Lip}} \in C_{\mathrm{Lip}}^{\circ}(G, E, \|\cdot\|, m)$. Будем описывать метод в предположении, что $\rho_{\|\cdot\|}(G) = 1$ и «центр» G есть 0. Очевидно, этого всегда можно добиться заменой переменных — растяжением и сдвигом G.

Трудоемкость ЗС$_V$ при точности решения $\nu > \nu_0$ есть $M_V(\nu - \nu_0)$, где

$$M_V(\nu) = \left]\frac{2}{\nu\gamma_V\left(\frac{\nu}{4}\right)}\right[+ 2. \tag{3.1}$$

Введем следующие обозначения. Пусть $\pi_G(x): E \to G$ — функция, относящая точке $x \in E$ ближайшую (в метрике, задаваемой нормой E) к x точку G. Пусть, далее, $\mu_G(x): E \to E^*$ сопоставляет точке $x \in E$ опорный в x функционал к функции $\rho_{\|\cdot\|}(\cdot, G)$.

Перейдем к описанию ЗС$_V$, настроенного на точность $\nu > \nu_0$. Случай $\nu \geqslant 1$ тривиален. При таких низких требованиях к точности решения задачу можно решить за один шаг. Любая точка G есть решение любой задачи $f \in C_{\mathrm{Lip}}$ погрешности 1.

В соответствии со сказанным далее считаем, что заданная точность ν удовлетворяет условию $1 > \nu > \nu_0$. Работа метода ЗС$_V$, настроенного на точность ν, $1 > \nu > \nu_0$, на задаче $f \in C_{\mathrm{Lip}}$ состоит в построении последовательности точек $\varphi_i \in E^*$, $0 \leqslant i \leqslant M_f$ (M_f — число вопросов метода о f) и ее «теней» $x_i \in E$, $\bar{x}_i \in G$. При этом строятся еще вспомогательные числовые последовательности $a_i \leqslant \infty$, $b_i \geqslant 0$. После M_f шагов процесса (M_f формируется автоматически) на основе последовательности $\{\bar{x}_i\}$ строится результат $\bar{x}$ работы метода на f.

3.2. Правила, описывающие ЗС$_V$, таковы.

ЗС0. Начальная настройка. Положить $\varphi_0 = 0$, $i = 1$, $a_0 = +\infty$, $b_0 = 0$ и перейти к ЗС1.

ЗС1. i-й шаг. К i-му шагу имеются точка $\varphi_{i-1} \in E^*$, числа $a_{i-1} \leqslant \infty$ и $b_{i-1} \geqslant 0$. На i-м шаге последовательно производятся следующие операции.

ЗС1.1. Строится точка $x_i = V'(\varphi_{i-1})$ и $\bar{x}_i = \pi_G(x_i)$. В точке $\bar{x}_i$ задается вопрос оракулу о решаемой задаче. Пусть $g_i^j(y)$ — сообщенные им аффинные функционалы, η_i^j — их производные по y.

ЗС1.2. Вычисляются числа $L_j(i) = \|\eta_i^j\|_*$. Определяется число $j(i)$:

$$j(i) = \begin{cases} 0, \text{ если } g_i^j(\bar{x}_i) \leqslant 2(\nu - \nu_0)L_j(i), & 1 \leqslant j \leqslant m, \\ \text{любое } j \geqslant 1, \text{ такое, что } g_i^j(\bar{x}_i) > 2(\nu - \nu_0)L_j(i), \\ \text{если такое } j \text{ есть.} \end{cases}$$

ЗС1.3. Определяется

$$a_i = \begin{cases} a_{i-1}, & j(i) > 0, \\ \min\{a_{i-1}, g_i^0(\bar{x}_i)\}, & j(i) = 0. \end{cases}$$

ЗС1.4. Если $\eta_i^{j(i)} = 0$, то переходим к ЗС2. В противном случае полагаем $\xi_i = \eta_i^{j(i)}/\|\eta_i^{j(i)}\|_*$ и

$$\hat{\delta}_i = \frac{1}{\|\eta_i^{j(i)}\|_*}\begin{cases} g_i^0(\bar{x}_i) - a_i + 2(\nu - \nu_0)L_0(i), & j(i) = 0,\\ g_i^{j(i)}(\bar{x}_i), & j(i) > 0.\end{cases}$$

ЗС1.5. Полагаем $\bar{\zeta}_i = \xi_i + \mu_G(x_i)$. Если $\bar{\zeta}_i = 0$, то переходим к ЗС2. В противном случае полагаем

$$\zeta_i = \bar{\zeta}_i/\|\bar{\zeta}_i\|_*, \quad \delta_i = \hat{\delta}_i/\|\bar{\zeta}_i\|_*.$$

ЗС1.6. Определяем $\rho_i = \gamma_V(\delta_i/2)$ и полагаем $\gamma_i = \delta_i\rho_i/2$.

ЗС1.7. Полагаем $\varphi_i = \varphi_{i-1} - \rho_i\zeta_i$, $b_i = b_{i-1} + \gamma_i$. Если $V(\varphi_i) - \|\varphi_i\|_* > -b_i$, то перейдем к ЗС2. В противном случае закончим шаг (т. е. увеличим i на 1 и перейдем к ЗС1).

ЗС2. Правило выдачи результата. Напомним, что обращение к ЗС2 происходит при $i = M_f$. Собственно, M_f, т. е. число вопросов о задаче f, задаваемых методом, как раз и определяется как номер шага, на котором происходит обращение к ЗС2. Работа метода на этом прекращается. Результат формируется в виде

$$\begin{array}{ll} *, & \text{если } j(i) > 0, \quad \forall i \leqslant M_f,\\ \bar{x}_{i_0}, & \text{если } j(i_0) = 0 \text{ и } g_{i_0}^0(x_{i_0}) = a_{M_f} \text{ в противном случае.}\end{array}$$

По определению a_i такое i_0 найдется, если $j(i) = 0$ хотя бы для одного $i \leqslant M_f$.

З а м е ч а н и е. Правило ЗС1.2, вообще говоря, неоднозначно определяет $j(i)$. Если в выборе $j(i)$ при данном i имеется свобода, то ее целесообразно использовать с тем, чтобы обеспечить максимально возможное значение γ_i. Такой прием в принципе способствует ускорению поиска. $\diamondsuit$

Со временем будет доказано, что описанный метод и его версии, приведенные ниже, гарантируют решение всякой задачи класса $\boldsymbol{C}_{\mathrm{Lip}}$ с относительной погрешностью, $\leqslant \nu$. Будет доказано также, что трудоемкость его не превосходит $M_V(\nu - \nu_0)$. Аналогичные оценки трудоемкости других версий ЗС-методов, формулируемые ниже, при описании этих версий, также будут доказаны в п. 3.5.

3.3. Теперь опишем некоторые модификации построенного метода.

3.3.1. Версия с «оптимальным» выбором смещения ρ_i. Правило ЗС 1.6 можно заменить следующим правилом.

ЗС 1.6′. Полагаем

$$r_i(t) = (\delta_i - \langle x_i \mid \zeta_i\rangle)t + V(\varphi_{i-1}) - V(\varphi_{i-1} - t\zeta_i),$$

и пусть t_i — точка максимума $r_i(t)$ по всем $t \geqslant 0$ (если этот максимум не достигается, то пусть, скажем, как и в ЗС1.6, $t_i = \gamma_V(\delta_i/2)$).

Полагаем $\rho_i = t_i$ и $\gamma_i = r_i(t_i)$.

Эта модификация, вообще говоря, ускоряет процесс поиска решения, но усложняет шаг. Трудоемкость указанной версии не превосходит $M_V(\nu - \nu_0)$.

3.3.2. Гильбертова версия метода. Пусть теперь E, $\|\cdot\|$ — гильбертово пространство. В этом случае возможны некоторые упрощения, приводящие ЗС_V к определенной версии стандартного градиентного метода. В предыдущем описании можно сделать следующие изменения:

— положить $V(\varphi) = {}^1/_2\,\varphi^2$ (в гильбертовом случае E и E^* канонически отождествляются, и мы их не различаем);

— заменить в ЗС1.7 правило формирования φ_i на $\varphi_i = \pi_G(\varphi_{i-1} - \rho_i\zeta_i)$ (при этом автоматически будет $\varphi_{i-1} = x_i = \bar{x}_i$, и можно считать $\mu_G(x_i) = 0$);

— заменить в ЗС1.6 правило выбора ρ_i, γ_i на $\rho_i = \delta_i$, $\gamma_i = \delta_i^2/2$. В этой ситуации ЗС1.6′ тождественно описанной модификации ЗС1.6, так что в гильбертовом случае мы ограничиваемся единственной версией метода — обозначим ее Г. Трудоемкость метода Г, настроенного на точность ν, $1 > \nu > \nu_0$, не превосходит величины $M(\nu - \nu_0)$, где

$$M(\nu) = \left]\frac{1}{4\nu^2}\right[+ 2. \tag{3.2}$$

3.3.3. Настройка на абсолютные погрешности. До сих пор мы предполагали, что задана относительная точность, с которой требуется решать задачи из $\boldsymbol{C}_{\mathrm{Lip}}$. Пусть теперь заранее задана не относительная точность, а абсолютные погрешности $\varepsilon_j > 0$, $j = 0, \ldots, m$, допустимые для приближенного решения. Оказывается, что ЗС-методы и в такой постановке способны обеспечить решение нужной точности. Точнее, описанные выше версии ЗС-методов нуждаются для этого в некоторой модификации. Прежде всего надо «совершить нулевой шаг» — задать вопрос о решаемой задаче в точке 0. Если окажется, что

$$g_{j,0}^f(0) \leqslant \varepsilon_j, \quad 1 \leqslant j \leqslant m, \quad \text{и} \quad \min_{y \in G} g_{0,0}^f(y) \geqslant g_{0,0}^f(0) - \varepsilon_0,$$

то в качестве приближенного решения задачи выдадим 0. Если же этого не произойдет, но окажется, что при каком-нибудь $j \geqslant 1$ $\min_G g_{j,0}^f(y) > 0$, то в качестве приближенного решения выдадим $*$. Если обе эти альтернативы не имеют места, применим для решения f любую допустимую (при данных E, $\|\cdot\|$) из описанных выше трех версий методов зеркального спуска решения задач из

$C_{\mathrm{Lip}} \subset C_{\mathrm{Lip}}^{\nu_0}(G, E, \|\cdot\|, m)$, модифицированную следующим образом.

Правило ЗС1.2 заменяется на ЗС1.2′:

$$j(i) = \begin{cases} 0, \text{ если } g_i^j(\bar{x}_i) \leqslant \varepsilon_j, \quad 1 \leqslant j \leqslant m, \\ \text{любое } j \geqslant 1, \text{ такое, что } g_i^j(\bar{x}_i) > \varepsilon_j, \text{ если такое } j \text{ есть.} \end{cases}$$

Кроме того, формулу для $\hat{\delta}_i$ в ЗС1.4 следует заменить на

$$\hat{\delta}_i = \frac{1}{\|\eta_i^{j(i)}\|} \begin{cases} g_i^0(\bar{x}_i) - a_i + \varepsilon_0, & j(i) = 0, \\ g_i^{j(i)}(\bar{x}_i), & j(i) > 0. \end{cases}$$

Полученные таким образом методы обладают следующим свойством: их применение к задаче $f \in C_{\mathrm{Lip}}$ доставляет результат $\bar{x}$, удовлетворяющий условиям (4.1) гл. II. При этом трудоемкость применения указанных методов не превосходит соответственно $M_V(\nu(f, \vec{\varepsilon}))$, $M(\nu(f, \vec{\varepsilon}))$, где

$$\nu(f, \vec{\varepsilon}) = \min_{0 \leqslant j \leqslant m} \frac{\varepsilon_j}{r_j(f)}$$

— максимально возможная относительная погрешность, гарантирующая требуемую абсолютную погрешность. Обратите внимание, что в случае неуспеха «нулевого шага» — когда он не приводит к результату — $\nu(f, \vec{\varepsilon}) \leqslant 1$. Таким образом, ЗС-методы приспосабливаются к обеспечению требуемой абсолютной погрешности с неменьшим успехом, чем методы, описанные в предыдущей главе.

3.3.4. Скажем несколько слов о практической реализации описанных методов. Конечно, речь может идти лишь о решении конечномерных задач. Ограничимся случаем, когда пространства E, $\|\cdot\|$ суть $l_p^{(n)}$ с нормой $\|\cdot\|_p$, $1 \leqslant n < \infty$. Алгоритмическая реализация ЗС-методов в этой ситуации требует умения решать три задачи.

(I) Находить производную $V'(\varphi)$ $(l_p^{(n)}, \|\cdot\|_p)$-регулярной функции V.

(II) Строить проекцию $\pi_G(x)$ точки $x \in l_p^{(n)}$ на выпуклое замкнутое ограниченное множество $G \subset l_p^{(n)}$.

(III) Строить опорный функционал $\mu_G(x)$ к функции $\rho_{\|\cdot\|_p}(\cdot, G)$.

Подчеркнем, что речь идет лишь о проблемах, связанных с организацией самих методов. Мы не касаемся вопросов алгоритмической реализации оракула, не имеющих прямого отношения собственно к теории методов оптимизации.

Разберем задачи (I) — (III). Первая из них весьма проста, по крайней мере, в случае, если рассматриваемый метод ассоциирован со стандартной $(l_p^{(n)}, \|\cdot\|_p)$-регулярной функцией из п. 2.3. Действительно, такие функции допускают явное представление, из которого легко извлечь аналитические выражения для координат $V'(\varphi)$. Вычисление по этим выражениям $V'(\varphi)$ требует $O(n)$ элементарных операций (в число которых мы включаем и вычисление стандартных функций — степеней, экспонент и т. п.).

Задача (II), вообще говоря, может быть далеко не простой. Действительно, отыскать ближайшую к x по метрике $\|\cdot\|_p$ точку G — это все равно, что решить специальную выпуклую задачу оптимизации, вообще говоря, не менее сложную, чем исходная. Однако тут приходит на помощь следующее соображение. Если область G имеет простую форму, то отыскивать $\pi_G(x)$ довольно легко (примеры будут даны ниже). В то же время выбор этой области во многом в нашей власти.

Действительно, область «сложной формы» обычно задается некоторым числом неравенств, для выпуклых задач — выпуклых. Термин «задается» здесь понимается не в том смысле, что такие неравенства существуют — они существуют всегда, а в том смысле, что способ задания области G в постановке задачи — алгоритм, выясняющий по x, лежит ли точка x в G или нет — основан на явном указании списка неравенств, которым должны удовлетворять точки G. Но в таком случае ничто не мешает нам перевести описывающие G неравенства в разряд ограничений решаемой задачи. После этого можно считать задачу определенной в подходящей содержащей G области G_1 простой (с точки зрения вычисления проекции $\pi_{G_1}(x)$) формы.

Конечно, при таком расширении G мы увеличиваем радиус области определения задачи. Тем самым увеличивается и трудоемкость ее решения с заданной абсолютной точностью. Это — цена, которую приходится платить за желание иметь дело с простыми областями. Вообще говоря, трудно сказать в общем случае, приемлема ли эта цена. Вопрос о рациональном выборе G_1 должен изучаться специально в каждой конкретной ситуации. Для ЗС-методов он довольно важен, ибо их трудоемкость обычно степенным образом зависит от требуемой относительной точности, так что увеличение размера G в k раз приводит к увеличению трудоемкости в k^s раз, s зависит от p, $s \geqslant 2$. Напротив, применительно к методам с линейной сходимостью (типа ММЦТ, в котором требуется решать задачу проектирования для евклидова случая) даже заметное «раздувание» G не приводит к существенному росту трудоемкости.

Что же касается задачи (III), то она тесно связана с задачей (II). Действительно, ясно, что при $x \notin G$ $\mu_G(x)$ есть *опорный* функционал к вектору $x - \pi_G(x) \equiv \Delta_G(x)$ (т. е. такой функционал $\varphi \in E^*$, что $\|\varphi\|_* = 1$ и $\langle\varphi | \Delta_G(x)\rangle = \|\Delta_G(x)\|$). Если мы умеем решать задачу (II), т. е., в частности, находить по данному x и вектор $\Delta_G(x)$, то, как правило, легко можем построить и $\mu_G(x)$.

Пусть, например, $1 < p < \infty$. Тогда $\| u \|_p$ — дифференцируемая при $u \neq 0$ функция, так что опорный функционал к ненулевому вектору $\Delta_G(x)$ (при $x \notin G$ $\Delta_G(x) \neq 0$) есть просто производная по u функции $\| u \|_p$ в точке $u = \Delta_G(x)$. Эта производная явно выписывается. Ее вычисление по заданному u требует $O(n)$ элементарных операций.

При $p = 1$ дело обстоит несколько сложнее из-за возможной неединственности опорного функционала к $\Delta_G(x)$ (не всякий такой функционал при этом опорен к $\rho_{\|\cdot\|_1}(\cdot, G)$ в x), но в приводимых ниже примерах эта трудность преодолевается.

Укажем теперь простейшие конкретные виды областей G, для которых задачи (II), (III) допускают явное решение. Мы считаем E, $\| \cdot \|$ совпадающим с $l_p^{(n)}$, $1 \leqslant p < \infty$.

1°. G — $l_p^{(n)}$-шар $\{\| x - x_0 \|_p \leqslant r\}$. Можно считать $r = 1$, $x_0 = 0$. В этом случае при $x \notin G$ имеем $\pi_G(x) = x / \| x \|_p$ (для $x \in G$, разумеется, $\pi_G(x) = x$). Далее, для $x \in G$ можно взять $\mu_G(x) = 0$. Если же $x \notin G$, то

$$\mu_G(x) = \{(\mu_G(x))_i\}_{i=1}^n,$$

где

$$(\mu_G(x))_i = |\bar{x}_i|^{p-1} \operatorname{sgn}(\bar{x}_i), \quad \bar{x} = \frac{x}{\| x \|_p}.$$

(Здесь принято $0^0 = 0$.)

2°. G — «долька» $l_p^{(n)}$-шара $G = \{\| x - x_0 \|_p \leqslant r,\ x \geqslant x_0\}$. Можно считать $x_0 = 0$, $r = 1$. Положим для $x \in l_p^{(n)}$ $x_+ = (\max(0, x_1), \ldots, \max(0, x_n))$. Тогда

$$\pi_G(x) = \begin{cases} \dfrac{x_+}{\| x_+ \|_p} & \text{при } \| x_+ \|_p > 1, \\ x_+ & \text{при } \| x_+ \|_p \leqslant 1 \end{cases}$$

и

$$(\mu_G(x))_i = |\tilde{x}_i|^{p-1} \operatorname{sgn} \tilde{x}_i, \quad \tilde{x} = \frac{x - \bar{x}}{\| x - \bar{x} \|_p},$$

где

$$\bar{x} = \begin{cases} \dfrac{x_+}{\| x_+ \|_p}, & \text{если } \| x_+ \|_p > 1, \\ x_+ & \text{при } \| x_+ \|_p \leqslant 1. \end{cases}$$

Видно, что в ситуациях пп. 1° — 2° вычисление $\pi_G(x)$ и $\mu_G(x)$ требует $O(n)$ элементарных операций. Ясно, что в этом случае реализация ЗС-метода требует $O(n)$ элементарных операций на шаг (при этом не учитываются операции, производимые оракулом, и операции сравнения, необходимые для вычисления $j(i)$. Число последних зависит от количества ограничений решаемой задачи).

Список «простых тел» G можно было бы увеличить, но нам представляется, что в свете всего изложенного достаточно уже приведенных примеров.

Разумеется, замечания этого пункта относятся ко всем версиям ЗС-методов, рассматриваемым в работе (главы III, V, VI), а также и к другим, использующим операцию проектирования, методам.

3.4. Прежде, чем обосновывать описанные методы, рассмотрим характеристики их конкретных версий, отвечающих случаю, когда $E = L_p$, $1 \leqslant p < \infty$, и $\| \cdot \|$ на E есть $\| \cdot \|_p$. Гильбертов случай (т. е. случай $p = 2$) мы уже разобрали (3.3.2). Соответствующая оценка трудоемкости

$$M(\nu - \nu_0) = \left] \frac{1}{4(\nu - \nu_0)^2} \right[+ 2.$$

В общем случае, в соответствии с (3.1), (2.3), (2.5) получаем следующие *верхние* оценки трудоемкости ЗС-методов решения липшицевых выпуклых задач на пространствах L_p с точностью $\nu > \nu_0$.

При $1 < p < \infty$ (метод ассоциирован с $V_p(\cdot)$, см. п. 2.3) трудоемкость метода не превосходит величины

$$M_p(\nu - \nu_0) = \begin{cases} \left] \dfrac{d(p)}{(\nu - \nu_0)^2} \right[+ 2, & 1 < p \leqslant 2, \\ \left] \dfrac{d(p)}{(\nu - \nu_0)^p} \right[+ 2, & p > 2. \end{cases} \tag{3.3}$$

Здесь $d(p)$ зависит лишь от p. При этом для $1 < p \leqslant 2$ справедлива оценка $d(p) \leqslant d/(p-1)$, d — абсолютная константа (см. (2.3) и оценку $c(p)$ из § 5).

При $p = 1$ и $\dim L_1 = n$, $1 < n < \infty$ (т. е. при $E = l_1^{(n)}$), для метода, ассоциированного с $V_{1,n}(\cdot)$ (см. п. 2.3), трудоемкость не превосходит величины

$$M_{1,n}(\nu - \nu_0) \leqslant \left] \frac{d(1) \ln n}{(\nu - \nu_0)^2} \right[+ 2. \tag{3.4}$$

ЗС-методы решения липшицевых выпуклых задач на L_p, $1 < p < \infty$, ассоциированные с $V(\cdot) = V_p(\cdot)$, будем обозначать ЗС$_p$. Аналогично, ЗС-метод решения задач на $l_1^{(n)}$, ассоциированный с $V(\cdot) = V_{1,n}(\cdot)$, обозначим ЗС$_{1,n}$*).

В § 3 гл. IV будет показано, что если $E = L_p$ и $n = \dim E$ достаточно велико, а G — шар в E, то трудоемкость ЗС$_p$ (ЗС$_{1,n}$ при $p = 1$) в принципе не может быть снижена более чем в константу (зависящую лишь от p) раз. Таким образом, ЗС-методы на классах липшицевых выпуклых задач теоретически субоптимальны,

*) Отметим, что ЗС$_{1,n}$-метод, обещанный в п. 1.3.

если область определения задач класса — тело большой размерности в L_p. Точные формулировки см. в §§ 2, 3 гл. IV.

Упражнение 1. Пусть E, $\|\cdot\|$ есть L_p $\|\cdot\|_p$, $1 \leqslant p \leqslant 2$, и $\dim E = n$, $1 < n < \infty$. Докажите, что задачи из $\boldsymbol{C}_{\mathrm{Lip}} \in \boldsymbol{C}_{\mathrm{Lip}}^{\nu_0}(G, E, \|\cdot\|, m)$ можно решать с оценкой трудоемкости $\bar{c} \ln n/(\nu - \nu_0)^2$, c — абсолютная константа. ⟨Примените для решения задач данного класса метод ЗС$_V$, ассоциированный с функцией вида V^τ из (2.4). Подберите нужное τ, используя явный вид величин $c(p)$ из (2.3).⟩

3.5. Теперь перейдем к обоснованию описанных методов.

Теорема. *Справедливы следующие утверждения.*

(*i*) *Пусть* E, $\|\cdot\|$ —*регулярное пространство*, V — *соответствующая регулярная функция. Обе версии методов* ЗС$_V$, *настроенные на точность* ν, $1 > \nu > \nu_0$, *решают всякую задачу из* $\boldsymbol{C}_{\mathrm{Lip}} \in \boldsymbol{C}_{\mathrm{Lip}}^{\nu_0}(G, E, \|\cdot\|, m)$ *с относительной погрешностью*, $\leqslant \nu$, *за число шагов*, $\leqslant M_V(\nu - \nu_0)$. *Аналогичное утверждение (с заменой* $M_V(\cdot)$ *на* $M(\cdot)$*) верно и для градиентного метода* Γ *в случае гильбертова* E, $\|\cdot\|$.

(*ii*) *В ситуации* п. 3.3.3 *описанные там методы, настроенные на погрешности* $\varepsilon_j > 0$, $0 \leqslant j \leqslant m$, *решают всякую задачу* f *из* $\boldsymbol{C}_{\mathrm{Lip}} \in \boldsymbol{C}_{\mathrm{Lip}}^{\nu_0}(G, E, \|\cdot\|, m)$ *с абсолютными погрешностями*, $\leqslant \varepsilon_j + 2\mathsf{L}_{\|\cdot\|}(f_j)\,\rho_{\|\cdot\|}(G)\,\nu_0$, *при трудоемкости, не превосходящей соответственно* $M_V(\nu(f, \vec{\varepsilon}))$, $M(\nu(f, \vec{\varepsilon}))$, *где*

$$\nu(f, \vec{\varepsilon}) = \min_{0 \leqslant j \leqslant m} \frac{\varepsilon_j}{2\mathsf{L}_{\|\cdot\|}(f_j)\,\rho_{\|\cdot\|}(G)}. \tag{3.5}$$

Доказательство всех утверждений теоремы проведем одновременно. Наметим схему доказательства, а заодно поясним механизм ЗС-методов. Предположим, что решается совместная задача $f \in \boldsymbol{C}_{\mathrm{Lip}} \in \boldsymbol{C}_{\mathrm{Lip}}^{\nu_0}(G, E, \|\cdot\|, m)$, а x^* — ее решение. Прежде всего выясняется, что процесс решения действительно обрывается в некоторый момент M_f, допускающий нужную оценку сверху. Этот факт следует из того, что, как можно доказать, числа γ_i все больше некоторой положительной, не зависящей от i величины $\bar{a}$. Между тем правило ЗС1.7 прекращает работу метода на шаге i, на котором $V(\varphi_i) - \|\varphi_i\|_* > -b_i \leqslant -i\bar{a}$. Нижняя грань левой части этого неравенства по всем φ_i конечна и равна, скажем, $-T$. Поэтому метод заведомо прекратит работу до момента$]\, T/\bar{a}\, [$. Явное вычисление T и $\bar{a}$ для рассматриваемых методов и приводит к анонсированным выше оценкам трудоемкости.

Теперь поясним, как обосновывается утверждение о точности. Пусть $V_*(\varphi) = V(\varphi) - \langle \varphi \mid x^* \rangle$. Выясняется, что *если метод не обеспечивает нужной точности*, то при всех $i \leqslant M_f$ справедливы неравенства $\langle \zeta_i \mid x_i - x^* \rangle \geqslant \delta_i$, т. е. $\langle V_*'(\varphi_{i-1}) \mid -\zeta_i \rangle \leqslant -\delta_i$.

Таким образом, $(-\zeta_i)$ — хорошее направление спуска для V_*. Точка φ_i получается из φ_{i-1} сдвигом вдоль этого направления. Правила выбора величины сдвига и γ_i таковы, что убывание V_* при переходе от φ_{i-1} к φ_i не меньше γ_i (таким образом, содержательно γ_i есть априорная оценка убывания V_* за один шаг). Стало быть, $V_*(\varphi_i) \leqslant -b_i$, а так как $V_*(\varphi) \geqslant V(\varphi) - \|\varphi\|_*$ (ибо $\|x^*\| \leqslant 1$), то подавно и $V(\varphi_i) - \|\varphi_i\|_* \leqslant -b_i$. Но мы знаем, что при $i = M_f$ это не так (правило окончания). Таким образом, допущение о том, что метод не обеспечивает нужной точности, приводит нас к потиворечию.

Приведем теперь строгую аргументацию утверждений теоремы.

1°. Докажем, что трудоемкость методов допускает нужные оценки. Для этого убедимся, что для всех рассматриваемых методов справедливо неравенство $\delta_i \geqslant a > 0$, $i < M_f$. Это же верно для $i = M_f$, если только $\eta_i^{(i)} \neq 0$ и $\bar{\zeta}_i \neq 0$. Здесь a — параметр, зависящий от ситуации (т. е. от того, какое утверждение из (i) — (ii) и для какой именно версии доказывается). Именно, для обеих версий метода ЗС$_V$ можно взять

$$a = \nu - \nu_0. \tag{3.6}$$

⟨Действительно, из правила выбора $j(i)$ и правила определения $\hat{\delta}_i$ следует, что $\hat{\delta}_i \geqslant 2(\nu - \nu_0)$. Кроме того, $\|\bar{\zeta}_i\|_* \leqslant 2$, так что и $\delta_i \geqslant \hat{\delta}_i/2 \geqslant \nu - \nu_0$.⟩ Для градиентного метода Γ можно взять

$$a = 2(\nu - \nu_0) \tag{3.7}$$

⟨по тем же, что и выше, причинам. Следует только учесть, что в этом случае $\hat{\delta}_i = \delta_i$, ибо $\mu_G(x_i) = 0$⟩.

В ситуации утверждения (ii) можно взять $a = \nu(f, \vec{\varepsilon})$, если применяются модификации ЗС$_V$ ⟨ибо в этом случае

$$\hat{\delta}_i \geqslant \frac{\varepsilon_{j(i)}}{\|\eta_i^{j(i)}\|_*} = \frac{2\varepsilon_{j(i)}}{2\|\eta_i^{j(i)i}\|_*\,\rho_{\|\cdot\|}(G)} \geqslant \frac{2\varepsilon_{j(i)}}{2L_{j(i)}(f)\,\rho_{\|\cdot\|}(G)} \geqslant 2\nu(f, \vec{\varepsilon}),$$

а $\delta_i \geqslant 1/2\,\hat{\delta}_i$; здесь и далее $L_j(f) = \mathsf{L}_{\|\cdot\|}(f_j)$⟩ и $a = 2\nu(f, \vec{\varepsilon})$, если применяется Γ ⟨по тем же причинам, что и выше⟩.

2°. Предположим теперь, что работа рассматриваемого метода не закончилась после $(N+1)$-го шага, $N \geqslant 1$. Это означает, что на первых N шагах все время было

$$V(\varphi_i) - \|\varphi_i\|_* \leqslant -b_i, \quad i \leqslant N. \tag{3.8}$$

Выведем отсюда, что N не может быть слишком большим.

1) Для версий методов $ЗС_V$, отвечающих ситуациям (i) и (ii) и использующих правило ЗС1.6, имеем

$$\gamma_i = \frac{\rho_i \delta_i}{2} \geqslant \frac{a\gamma_V(a/2)}{2}$$

в силу правила выбора ρ_i ($\rho_i = \gamma_V(\delta_i/2)$) и доказанного выше неравенства $\delta_i \geqslant a$, $i < M_f$. Стало быть,

$$V(\varphi_N) - \|\varphi_N\|_* \leqslant -\frac{1}{2} N a\gamma_V\left(\frac{a}{2}\right).$$

Но левая часть этого неравенства не меньше $-1/2$ (ввиду (2.1) (ii)). Стало быть,

$$N \leqslant \frac{1}{a\gamma_V(a/2)},$$

т. е.

$$M_f \leqslant \frac{1}{a\gamma_V(a/2)} + 1$$

и трудоемкость исследуемых методов не превосходит

$$M_f + 1 \leqslant \frac{1}{a\gamma_V(a/2)} + 2,$$

что в силу определения a дает требуемую оценку трудоемкости.

2) Для методов $ЗС_V$, в ситуации (i) или (ii), использующих правило ЗС 1.6′, имеем, ввиду предложения 2.4, при $i \leqslant N$

$$r_i(t) = (\delta_i - \langle x_i \mid \zeta_i\rangle)\, t + V(\varphi_{i-1}) - V(\varphi_i - t\zeta_i) \geqslant$$
$$\geqslant (\delta_i - \langle x_i \mid \zeta_i\rangle)\, t + \langle V'(\varphi_{i-1} \mid \zeta_i)\rangle\, t - \omega_V(t)\, t \geqslant \delta_i t - \omega_V(t)\, t$$

Отсюда ясно, что

$$\max_{t\geqslant 0} r_i(t) \geqslant r_i\left(\gamma_V\left(\frac{\delta_i}{2}\right)\right) \geqslant \frac{\delta_i}{2}\gamma_V\left(\frac{\delta_i}{2}\right),$$

т. е. и в этом случае

$$\gamma_i \geqslant \frac{a}{2}\gamma_V\left(\frac{a}{2}\right),$$

что приводит к нужной оценке M_f.

3) Наконец, для градиентного метода требуемая оценка проводится в точности так же, как и выше, скажем, в п. 2).

3°. Итак, оценки трудоемкости доказаны. Перейдем к доказательству утверждений о погрешности. Прежде всего рассмотрим ситуацию (ii) и предположим, что результат был получен на «нулевом шаге» (см. п. 3.3.3). Этот результат есть 0, если

$$g^f_{j,0}(0) \leqslant \varepsilon_j,\ 1 \leqslant j \leqslant m, \ \text{ и } \ \min_{y\in G} g^f_{j,0}(y) \geqslant g^f_{0,0}(0) - \varepsilon_0;$$

в этом случае

$$g^f_{j,0}(0) \leqslant \varepsilon_j,\ 1 \leqslant j \leqslant m, \ \text{ и } \ \min_G f_0(y) \geqslant g^f_{0,0}(0) - \varepsilon_0.$$

Следовательно, в этом случае

$$f_0(0) \leqslant f_* + \varepsilon_0 + r_0(f)\,\nu_0,\ \ f_j(0) \leqslant \varepsilon_j + r_j(f)\,\nu_0,$$

что и было обещано. Этот результат, далее, есть $*$, если

$$\exists j \geqslant 1:\ \min_{y\in G} g^f_{j,0}(y) > 0,$$

а в этом случае f несовместна. Итак, если результат получен на нулевом шаге, то он удовлетворяет нужным требованиям точности.

Теперь предположим, что это не так, и в ситуациях (i) — (ii) рассматриваемые методы работают «нетривиально»— в соответствии с правилами ЗС0—ЗС2.

4°. Предположим, что решаемая задача f несовместна. По правилу ЗС2 выдачи результата $\bar{x}$ последний есть либо ответ (правильный) о несовместности f, либо точка вида $\bar{x}_{i_0}$, такая, что при $1 \leqslant j \leqslant m$

$$g^j_{i_0}(\bar{x}_{i_0}) \leqslant \begin{cases} 2(\nu - \nu_0)\, L_j(i) & \text{в ситуации (i)},\\ \varepsilon_j & \text{в ситуации (ii)}.\end{cases}$$

Так как $L_j(i) \leqslant L_j(f) = 1/2\, r_j(f)$ (напомним, что $\rho_{\|\cdot\|}(G) = 1$), то отсюда и из свойств оракула следует, что при $\bar{x} \neq *$ и всех $j \geqslant 1$

$$f_j(\bar{x}) \leqslant \begin{cases} \nu r_j(f) & \text{в ситуации (i)},\\ \varepsilon_j + \nu_0 r_j(f) & \text{в ситуации (ii)}.\end{cases} \tag{3.9}$$

Для несовместной задачи f эти неравенства и означают выполнение требований к точности результата.

5°. Предположим теперь, что f совместна, и пусть x^* — ее решение. Докажем, что если при некотором $i \leqslant M_f$ было

$$\eta_i^{j(i)} = 0 \ \text{(это возможно лишь при } i = M_f) \tag{3.10}$$

или

$$\langle \eta_i^{j(i)} \mid \bar{x}_i - x^* \rangle < \hat{\delta}_i \|\eta_i^{j(i)}\|_*,\ \ \|\eta_i^{j(i)}\|_* \neq 0, \tag{3.11}$$

то метод имеет нужную точность.

Действительно, во втором случае имеем по определению $\hat{\delta}_i$

$$g_i^{j(i)}(\bar{x}_i) - g_i^{j(i)}(x^*) = \langle \eta_i^{j(i)} \mid \bar{x}_i - x^* \rangle <$$
$$< \begin{cases} g_i^0(\bar{x}_i) - a_i + 2(\nu - \nu_0)\, L_0(i), & j(i) = 0,\ \text{ситуация (i)},\\ g_i^0(\bar{x}_i) - a_i + \varepsilon_0, & j(i) = 0,\ \text{ситуация (ii)},\\ g_i^{j(i)}(\bar{x}_i), & j(i) > 0.\end{cases}$$

Допустим, что $j(i) \neq 0$. Тогда

$$g_i^{j(i)}(\bar{x}_i) - g_i^{j(i)}(x^*) \geqslant g_i^{j(i)}(\bar{x}_i)$$

и рассматриваемое неравенство дает

$$g_i^{j(i)}(\bar{x}_i) < g_i^{j(i)}(\bar{x}_i),$$

что невозможно. Итак, $j(i) = 0$. В этом случае

$$g_i^0(\bar{x}_i) - f_0(x^*) < \begin{cases} g_i^0(\bar{x}_i) - a_i + 2(\nu - \nu_0) L_0(i), & \text{ситуация (i)}, \\ g_i^0(\bar{x}_i) - a_i + \varepsilon_0, & \text{ситуация (ii)}. \end{cases}$$

Итак, в случае (3.11) имеем для данного i

$$a_i \leqslant \begin{cases} f_0(x^*) + (\nu - \nu_0) r_0(f), & \text{ситуация (i)}, \\ f_0(x^*) + \varepsilon_0, & \text{ситуация (ii)}. \end{cases} \tag{3.12}$$

Легко видеть, что эти же утверждения справедливы и при $\eta_i^{j(i)} = 0$, т. е. в случае (3.10). Выведем из них требуемую оценку погрешности результата. В условиях (3.12) результат $\bar{x}$ есть точка $\bar{x}_{i_0}$, где $i_0 \leqslant M_f$ таково, что $j(i_0) = 0$ и $g_{i_0}^0(\bar{x}_{i_0}) = a_{M_f} \leqslant a_i$. Из соотношения $j(i_0) = 0$ в п. 4° уже было выведено соотношение (3.9). Далее,

$$f_0(\bar{x}_{i_0}) \leqslant g_{i_0}^0(\bar{x}_{i_0}) + r_0(f)\nu_0 \leqslant a_{M_f} + r_0(f)\nu_0,$$

что с учетом (3.12) дает

$$f_0(\bar{x}_{i_0}) \leqslant \begin{cases} f_* + \nu r_0(f), & \text{ситуация (i)}, \\ f_* + \varepsilon_0 + \nu_0 r_0(f), & \text{ситуация (ii)}. \end{cases} \tag{3.13}$$

Соотношения (3.9) и (3.13) доказывают нужное утверждение о точности результата.

6°. Для завершения доказательства достаточно убедиться, что при некотором $i \leqslant M_f$ действительно выполняется (3.10) или (3.11). Предположим противное — при всех $i \leqslant M_f$

$$\eta_i^{j(i)} \neq 0 \quad \text{и} \quad \langle \eta_i^{j(i)} | \bar{x}_i - x^* \rangle \geqslant \hat{\delta}_i \| \eta_i^{j(i)} \|_*, \tag{3.14}$$

и приведем (3.14) к противоречию. Действительно, из (3.14) следует

$$\langle \xi_i \mid \bar{x}_i - x^* \rangle \geqslant \hat{\delta}_i.$$

Покажем, что тогда

$$\xi_i + \mu_G(x_i) = \bar{\zeta}_i \neq 0 \quad \text{и} \quad \langle \bar{\zeta}_i | x_i - x^* \rangle \geqslant \hat{\delta}_i. \tag{3.15}$$

В гильбертовом случае это очевидно, поскольку там $x_i = \bar{x}_i$ и $\mu_G(x_i) = 0$, т. е. $\xi_i = \bar{\zeta}_i$. В общем случае имеем

$$\begin{aligned} \langle \bar{\zeta}_i | x_i - x^* \rangle &= \langle \xi_i | x_i - x^* \rangle + \langle \mu_G(x_i) | x_i - x^* \rangle = \\ &= \langle \xi_i | \bar{x}_i - x^* \rangle + \langle \xi_i | x_i - \bar{x}_i \rangle + \langle \mu_G(x_i) | x_i - x^* \rangle \geqslant \\ &\geqslant \langle \xi_i | \bar{x}_i - x^* \rangle - \| x_i - \bar{x}_i \| + \rho_{\|\cdot\|}(x_i, G) = \langle \xi_i | \bar{x}_i - x^* \rangle \geqslant \hat{\delta}_i \end{aligned}$$

(мы учли, что по определению μ_G имеем $\langle \mu_G(x_i) \mid x_i - u \rangle \geqslant \rho_{\|\cdot\|}(x_i, G)$ при всех $u \in G$). Итак, (3.15) выполнено. В первом пункте доказательства мы видели, что $\hat{\delta}_i \neq 0$, если оно вообще определено (т. е. если $\eta_i^{j(i)} \neq 0$). Итак, $\bar{\zeta}_i \neq 0$ и $\langle \zeta_i | x_i - x^* \rangle \geqslant \delta_i$, или, что то же,

$$\langle V_*'(\varphi_{i-1}) \mid \zeta_i \rangle \geqslant \delta_i. \tag{3.16}$$

Выведем из (3.16), что при всех $i \leqslant M_f$

$$-V_*(\varphi_i) + V_*(\varphi_{i-1}) \geqslant \gamma_i, \tag{3.17}$$

т. е. в силу $V_*(0) = 0$

$$V_*(\varphi_i) \leqslant -b_i, \ i \leqslant M_f. \tag{3.18}$$

Так как $V_*(\varphi_i) \geqslant V(\varphi_i) - \| \varphi_i \|_*$, то (3.18) означает, что

$$V(\varphi_i) - \| \varphi_i \|_* \leqslant -b_i. \tag{3.19}$$

Последнее неравенство при $i = M_f$ невозможно (мы уже знаем, что работа метода не может кончиться из-за $\eta_{M_f}^{j(M_f)} = 0$ или $\bar{\zeta}_{M_f} = 0$, так что в момент M_f должно быть выполнено неравенство, противоположное (3.19)).

Итак, чтобы получить желаемое противоречие, достаточно вывести (3.17) из (3.16). Это делается следующим образом.

1) Пусть речь идет о методах ЗС$_V$ в ситуациях (i) или (ii), причем используется ЗС1.6. Функция $\bar{r}_i(t) = -V_*(\varphi_{i-1} - t\zeta_i) + V_*(\varphi_{i-1})$ равна 0 при $t = 0$ и имеет производную по t в этой точке, равную $\langle \zeta_i \mid x_i - x^*, \rangle$ т. е. $\geqslant \delta_i$. Она отличается линейным слагаемым от $V(\varphi_{i-1}) - V(\varphi_{i-1} - t\zeta_i) \equiv \tilde{r}(t)$. Стало быть, по определению $\omega_V(t)$ при $0 \leqslant t$ имеем

$$\frac{d}{dt} \bar{r}_i(t) \geqslant \delta_i - \omega_V(t).$$

По определению ρ_i получаем, что при $0 \leqslant t \leqslant \rho_i$ имеем

$$\frac{d}{dt} \bar{r}_i(t) \geqslant \frac{\delta_i}{2}.$$

Отсюда немедленно следует (3.17), что и требуется.

2) Анализ ситуации с методами ЗС$_V$ в случаях (i), (ii) при правиле ЗС1.6′ ничем не отличается от предыдущего. Действительно,

функция $r_i(t)$ по своему определению отличается от $\bar r_i(t)$ линейным слагаемым, причем $\bar r_i(t) \geqslant r_i(t)$, $t \geqslant 0$. Поэтому в этом случае

$$V_*(\varphi_{i-1}) - V_*(\varphi_i) = \bar r_i(\rho_i) \geqslant r_i(\rho_i) = \gamma_i,$$

что и требуется.

3) Евклидов случай (т. е. градиентный метод) анализируется точно так же, как и п. 2).

Итак, мы вывели (3.17) из (3.16) для всех интересующих нас ситуаций. Доказательство теоремы закончено.

§ 4. Методы зеркального спуска для классов общих выпуклых задач

Покажем, что ЗС-методы можно применять и к общим выпуклым задачам, если только предположить, что G выпукло, замкнуто, ограничено и имеет внутренность (т. е. является телом).

Пусть $E, \|\cdot\|$ — регулярное пространство, $G \subset E$ — непустое выпуклое замкнутое ограниченное тело в E асферичности $\alpha_{\|\cdot\|}(G) = \alpha$. Опишем методы решения задач классов типа $C^{\nu_0}(G, E, m)$. Приведем несколько версий методов, подобно тому, как это сделано в предыдущем параграфе. Будем предполагать, что G содержится в шаре радиуса 1 с центром в 0 и содержит шар радиуса $1/2\alpha$ с тем же центром. Этого можно, очевидно, добиться подобным преобразованием G, так что принятое предположение не ограничивает общности построений.

4.1. Пусть $\nu > \nu_0$ — требуемая относительная точность решения задач из $C \in C^{\nu_0}(G, E, m)$ (далее мы рассмотрим также и случай, когда заданы не относительные, а абсолютные требования к погрешности решения). Мы увидим, что решать задачи класса C с точностью $\nu > \nu_0$ — по трудоемкости то же самое, что и решать липшицевы выпуклые задачи (задачи класса $\in C^{\nu_0}_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$) с точностью $\bar\nu(\nu) = \nu_0 + (\nu - \nu_0)/4\alpha$.

Как всегда, всякая точка $x \in G$ есть решение любой задачи $f \in C$ погрешности, $\leqslant 1$. Тем не менее по причинам, которые станут ясны позднее, в п. 4.2.3 нам будет удобно рассмотреть диапазон изменения ν от ν_0 до 2α и дать описание метода для всех значений ν из этого диапазона.

4.2. Опишем ЗС-методы решения общих выпуклых задач. Начнем с самого общего случая, когда относительно $E, \|\cdot\|$ предполагается только регулярность.

Пусть $\pi_G(x)\colon E \to G$ и $\mu_G(x)\colon E \to E^*$ определены как отображения, относящие точке $x \in E$ ближайшую (в норме E) к ней точку G, соответственно опорный в x функционал к $\rho_{\|\cdot\|}(\cdot, G)$. Пусть, далее, $V(\varphi)$ — соответствующая $(E, \|\cdot\|)$-регулярная функция. Опишем ассоциированный с ней ЗС-метод решения задач из C (обозначим его $\overline{\text{ЗС}}_V$).

Работа $\overline{\text{ЗС}}_V$, настроенного на точность ν, $2\alpha > \nu > \nu_0$, на задаче $f \in C$ состоит в построении последовательности $\varphi_i \in E^*$, $0 \leqslant i \leqslant M_f$ (M_f — число вопросов метода на f) и ее «теней» $x_i \in E$, $\bar x_i \in \operatorname{int} G$. По ходу дела строятся еще вспомогательные числовые последовательности $a_i \leqslant +\infty$, $b_i > 0$. После M_f шагов (M_f формируется самим методом) по последовательности $\bar x_i$ строится результат. Трудоемкость метода на f есть, таким образом, $M_f + 1$. Будет доказано, что эта величина не превосходит $M_{V,G}(\nu - \nu_0)$, где

$$M_{V,G}(\nu) = M_V\left(\frac{\nu}{4\alpha}\right) = \left]\frac{8\alpha}{\nu\gamma_V\left(\frac{\nu}{16\alpha}\right)}\right[+ 2. \tag{4.1}$$

Работа $\overline{\text{ЗС}}_V$ на задаче f описывается следующими правилами.

$\overline{\text{ЗС}}0$. *Начальная настройка.* Полагаем $\varphi_0 = 0$, $i = 1$, $a_0 = +\infty$, $b_0 = 0$, и пусть θ определено соотношением $(1/\theta) - 1 = \nu - \nu_0$. Тогда $0 < \theta < 1$. Переходим к $\overline{\text{ЗС}}1$.

$\overline{\text{ЗС}}1$. *i-й шаг.* К i-му шагу имеются точка φ_{i-1}, числа $a_{i-1} \leqslant +\infty$ и $b_{i-1} \geqslant 0$. На i-м шаге последовательно проводятся следующие операции.

$\overline{\text{ЗС}}1.1$. Строятся точки $x_i = V'(\varphi_{i-1})$ и $\bar x_i = \theta\,\pi_G(x_i)$. В точке $\bar x_i$ задается вопрос оракулу о решаемой задаче *). Пусть $g_i^j(y)$ — сообщенные им аффинные функционалы, η_i^j — их производные по y.

$\overline{\text{ЗС}}1.2$. Вычисляются числа

$$r_j(i) = \begin{cases} \sup\limits_{y\in G} g_i^0(y) - g_i^0(\bar x_i), & j = 0, \\ [\sup\limits_{y\in G} g_i^j(y)]_+ - [g_i^j(\bar x_i)]_+, & j \geqslant 1. \end{cases}$$

Определяется число

$$j(i) = \begin{cases} 0, & \text{если } g_i^j(\bar x_i) \leqslant (\nu - \nu_0) r_j(i), \quad 1 \leqslant j \leqslant m, \\ j \geqslant 1, & \text{такое, что } g_i^j(\bar x_i) > (\nu - \nu_0) r_j(i), \text{ если такое } j \text{ есть.} \end{cases}$$

$\overline{\text{ЗС}}1.3$. Определяется

$$a_i = \begin{cases} a_{i-1}, & j(i) > 0, \\ \min\{a_{i-1}, g_i^0(\bar x_i)\}, & j(i) = 0. \end{cases}$$

*) Это возможно, ибо $\theta < 1$ и $\theta\pi_G(x) \in \operatorname{int} G$.

$\overline{ЗС}1.4$. Если $\eta_i^{j(i)} = 0$, то перейдем к $\overline{ЗС}2$. В противном случае положим

$$\xi_i = \frac{\eta_i^{j(i)}}{\|\eta_i^{j(i)}\|_*},$$

$$\check{\delta}_i \equiv -(\nu-\nu_0)\langle\xi_i \mid \bar{x}_i\rangle + \hat{\delta}_i = \\ = \frac{1}{\|\eta_i^{(i)}\|_*}\begin{cases} g_i^0(\bar{x}_i) - a_i + (\nu-\nu_0) r_0(i), & j(i) = 0, \\ g_i^{(i)}(\bar{x}_i), & j(i) > 0. \end{cases}$$

$\overline{ЗС}1.5$. Полагаем $\bar{\zeta}_i = \xi_i + \mu_G(x_i)$. Если $\bar{\zeta}_i = 0$, то переходим к $\overline{ЗС}2$. В противном случае полагаем

$$\zeta_i = \frac{\bar{\zeta}_i}{\|\bar{\zeta}_i\|_*}, \quad \delta_i = \frac{\hat{\delta}_i}{\|\bar{\zeta}_i\|_*}.$$

$\overline{ЗС}1.6$. Полагаем

$$\rho_i = \gamma_V\left(\frac{\delta_i}{2}\right), \quad \gamma_i = \frac{\delta_i\rho_i}{2}.$$

⟨Вместо правила $\overline{ЗС}1.6$ можно использовать правило

$\overline{ЗС}1.6'$. Полагаем

$$r_i(t) = (\delta_i - \langle x_i \mid \zeta_i\rangle)\, t + V(\varphi_{i-1}) - V(\varphi_{i-1} - t\zeta_i).$$

Максимизируем $r_i(t)$ по всем $t \geqslant 0$. Если этот максимум достигается, то пусть он равен γ_i, а оптимальное значение t есть ρ_i. Если же максимум не достигается, то положим

$$\rho_i = \gamma_V\left(\frac{\delta_i}{2}\right), \quad \gamma_i = \frac{\delta_i\rho_i}{2}.⟩$$

$\overline{ЗС}1.7$. Полагаем $\varphi_i = \varphi_{i-1} - \rho_i\zeta_i$ и $b_i = b_{i-1} + \gamma_i$. Если $V(\varphi_i) - \|\varphi_i\|_* > -b_i$, то перейдем к $\overline{ЗС}2$. В противном случае закончим шаг (т. е. увеличим i на 1 и перейдем к $\overline{ЗС}1$).

$\overline{ЗС}2$. *Правило выдачи результатов.* Пусть обращение к $\overline{ЗС}2$ произошло при $i = M_f$. Работа метода на этом прекращается. Результат формируется как

$*$, если $j(i) > 0$, $1 \leqslant i \leqslant M_f$,

$\bar{x}_{i_0}$, если $j(i_0) = 0$ и $g_{i_0}^0(\bar{x}_{i_0}) = a_{M_f}$ — в противном случае.

⟨По определению a_i, см. $\overline{ЗС}1.3$, такое i_0 найдется, если $j(i) = 0$ хотя бы для одного $i \geqslant M_f$.⟩

В последнем случае положим также

$$\bar{r}_0(f) = \sup_{j(i)=0} r_0(i),$$

в первом — $\bar{r}_0(f) = 0$. Полагаем

$$\bar{r}_j(f) = \max_{1\leqslant i\leqslant M_f} r_j(i), \quad j \geqslant 1\ ^{*)}.$$

З а м е ч а н и е. При выборе $j(i)$ по правилу $\overline{ЗС}1.2$ имеется, вообще говоря, некоторая свобода. Ее целесообразно использовать так, чтобы максимизировать γ_i.

У п р а ж н е н и е. Сравните описания $ЗС_V$ и $\overline{ЗС}_V$. Чем они различаются?

4.3. Как и в предыдущем параграфе, отложим на время обоснование метода, чтобы привести серию его версий, отвечающих дополнительным предположениям о ситуации. Анонсируемые ниже свойства этих версий будут доказаны в п. 4.6.

4.3.1. В гильбертовом случае (где мы не различаем E и E^*) возможны некоторые упрощения. Эти упрощения состоят в следующем:

а) в качестве $V(\varphi)$ выбирается $\frac{1}{2}\varphi^2$;

б) в $\overline{ЗС}1.7$ правило формирования φ_i заменяется на

$$\varphi_i = \pi_G(\varphi_{i-1} - \rho_i\zeta_i)$$

(при этом автоматически будет $x_i = \varphi_{i-1}$ и можно считать $\mu_G(x_i) = 0$). В $\overline{ЗС}1.6$ правила выбора ρ_i и γ_i заменяются на $\rho_i = \delta_i$, $\gamma_i = \delta_i^2/2$ (при этом $\overline{ЗС}1.6'$ становится эквивалентным $\overline{ЗС}1.6$). Полученный метод является версией обычного градиентного метода — обозначим эту версию символом $\bar{\Gamma}$. Метод $\bar{\Gamma}$, настроенный на точность ν, $2\alpha > \nu > \nu_0$, решает все задачи из $C \in C^{\nu_0}(G, E, m)$ с точностью ν при трудоемкости, не превосходящей $M^G(\nu - \nu_0)$, где

$$M^G(\nu) = \left]\frac{4\alpha^2}{\nu^2}\right[+ 2. \tag{4.2}$$

4.3.2. До сих пор речь шла об обеспечении заданной относительной погрешности. Рассмотрим теперь случай, когда заранее задается не относительная, а абсолютная погрешность решения — вектор $\vec{\varepsilon}$ абсолютных погрешностей $\varepsilon_j > 0$, допускаемых для результата работы метода. Как и в случае липшицевых задач, каждый из трех**) описанных выше методов после надлежащей модификации можно применять и для решения задач из $C \in C^{\nu_0}(G, E, m)$ с данными абсолютными погрешностями (фраза «каждый из методов», разумеется, предполагает, что E, $\|\cdot\|$ удовлетворяет условиям применимости рассматриваемого метода).

*) Величины $\bar{r}_j(f)$ потребуются ниже, в п. 4.3.2.

**) Две версии $\overline{ЗС}_V$ и метод $\bar{\Gamma}$.

Покажем, как это делается для данного метода — обозначим его $\mathscr{B}$. Фиксируем какую-нибудь *непрерывную* функцию $N(\nu)$, оценивающую сверху трудоемкость настроенного на точность $\nu_0 + \nu$ метода $\mathscr{B}$ на классе C. Положим $\nu^0 = 2\alpha$, и пусть i_0 — первое натуральное i, для которого $2^{i_0} \geqslant N(\nu^0)$. Обозначим через ν^i, $i = 1, 2, \ldots$, минимальное ν, для которого $N(\nu) = 2^{i_0+i}$. Тогда $\nu^0 > \nu^1 > \nu^2 > \ldots$

Пусть $\mathscr{B}(\nu^0)$ — двухшаговый метод решения задач из C, описываемый следующим образом. Задается вопрос о решаемой задаче f в точке 0. Если при каком-нибудь $j \geqslant 1$ при этом получаем $g^f_{j,0}(y) > 0$ при всех $y \in G$, то задача объявляется несовместной. В противном случае точка 0 выдается в качестве ее приближенного решения. При этом пусть

$$\bar r_j^{(0)}(f) = \begin{cases} [\sup\limits_{y\in G} g^f_{j,0}(y)]_+, & j \geqslant 1, \\ \sup\limits_{y\in G} g^f_{0,0}(y) - g^f_{0,0}(0), & j = 0. \end{cases}$$

Пусть, далее, $\mathscr{B}(\nu^i)$ означает метод $\mathscr{B}$, настроенный на точность ν^i, $1 \leqslant i$.

Рассмотрим теперь метод $\overline{\mathscr{B}}$, описываемый следующим образом. Пусть $\vec\varepsilon = (\varepsilon_0, \ldots, \varepsilon_m)$ — допустимые абсолютные погрешности решения задач из $C \in C^{\nu_0}(G, E, m)$. Работа метода $\overline{\mathscr{B}}$ состоит в последовательном применении к решаемой задаче $f \in C$ процедур $\mathscr{B}(\nu^0)$, $\mathscr{B}(\nu^1)$, ... При применении s-й процедуры фиксируются числа

$$\delta_j(s, f) = \nu^s \bar r_j^{(s)}(f), \quad 0 < j \leqslant M. \tag{4.3}$$

(Числа $\bar r_j^{(s)}(f)$, $s > 0$, получаются в результате применения $\mathscr{B}(\nu^s)$ к f по правилу $\overline{\text{ЗС}}2$). Работа метода $\overline{\mathscr{B}}$ на задаче f *заканчивается* применением процедуры $\mathscr{B}(\nu^s)$, такой, что для нее впервые получается $\delta_j(s, f) \leqslant \varepsilon_j$, $0 \leqslant j \leqslant m$ (такое j имеется, поскольку из свойств оракула и определения $\bar r_j^{(s)}(f)$ ясно, что

$$\bar r_j^{(s)}(f) \leqslant r_j(f), \quad j \geqslant 1, \quad \bar r_0^{(s)}(f) \leqslant (1 + \nu_0)\, r_0(f),$$

а $\nu^s \xrightarrow[s\to\infty]{} 0$). Оказывается, что описанный метод обеспечивает решение любой задачи $f \in C$ с абсолютными погрешностями, $\leqslant \bar\varepsilon_j = \varepsilon_j + r_j(f)\,\nu_0$. Иными словами, абсолютная погрешность метода, как и всегда в аналогичной ситуации, совпадает с требуемой с точностью до «неустранимой погрешности оракула» (ср. п. 3.3.3 и п. 4.1 гл. II). Трудоемкость же указанного метода не превосходит $8\bar N(\nu(f, \vec\varepsilon)/(1 + \nu_0))$, где

$$\nu(f, \vec\varepsilon) = \min_{0\leqslant j\leqslant m} \varepsilon_j / r_j(f)$$

— максимально возможная относительная погрешность, гаранти-

рующая данные абсолютные погрешности (обратите внимание, что вполне возможно $\nu(f, \vec\varepsilon) \geqslant 1$).

Теперь мы в состоянии пояснить, почему требовалось описывать $\overline{\text{ЗС}}$-методы и для точностей $\nu > 1$. Дело в том, что такого рода *методы входят в метод* $\overline{\mathscr{B}}$ *(для некоторого конечного числа номеров* i, вполне возможно, $\nu^i \geqslant 1$). Если и $\nu(f, \vec\varepsilon) \geqslant 1$, то работа $\overline{\mathscr{B}}$, возможно, остановится на одной из процедур с $\nu^i \geqslant 1$. В этом случае мы могли бы сразу выдать в качестве решения любую точку G, но не в состоянии сделать это, так как не умеем находить $\nu(f, \vec\varepsilon)$ по известному $\vec\varepsilon$. Поэтому нам *приходится предпринимать некие действия во всех случаях.*

Конечно, мы могли бы начинать метод $\overline{\mathscr{B}}$ процедурой типа $\mathscr{B}(\nu)$ с $\nu < 1$, но при этом мы затратили бы, возможно, больше шагов на построение искомого решения. Разумеется, такой эффект мог бы иметь место лишь при $\nu(f, \vec\varepsilon) \geqslant 1$.

4.4. Выпишем верхние оценки трудоемкости $\overline{\text{ЗС}}$-методов решения общих выпуклых задач на стандартных пространствах L_p. Эти оценки непосредственно выводятся из (4.1) и (3.3)—(3.4).

$1 < p < \infty$, метод ассоциирован с $V(\cdot) = V_p(\cdot)$ (назовем его $\overline{\text{ЗС}}_p$); трудоемкость решения задач из $C \in C^{\nu_0}(G, E, m)$ на G с относительной погрешностью $\nu > \nu_0$ не превосходит

$$M_{p,G}(\nu - \nu_0) \leqslant \begin{cases} \left]\dfrac{\bar d(p)\,\alpha^2_{\|\cdot\|}(G)}{(\nu-\nu_0)^2}\right[+ 2, & 1 < p \leqslant 2, \\ \left]\dfrac{\bar d(p)\,\alpha^p_{\|\cdot\|}(G)}{(\nu-\nu_0)^p}\right[+ 2, & p > 2 \end{cases} \tag{4.4}$$

(в случае $p = 2$ при применении метода $\bar\Gamma$ оценка имеется в (4.2)).

$p = 1$ и $\dim E = n$, $1 < n < \infty$, метод ассоциирован с $V_{1,n}(\cdot)$ (назовем его $\overline{\text{ЗС}}_{1,n}$). Оценка трудоемкости $\overline{\text{ЗС}}_{1,n}$ определяется неравенством

$$M_{1,n;G}(\nu - \nu_0) \leqslant \left]\frac{\bar d(1)\,\alpha^2_{\|\cdot\|}(G)\ln n}{(\nu-\nu_0)^2}\right[+ 2. \tag{4.5}$$

В (4.4)—(4.5) $\bar d(p) < \infty$ зависит лишь от p. Отметим, что $\overline{\text{ЗС}}_{1,n}$ — метод, обещанный в п. 1.3.

4.5. Обсудим схему применения ЗС-методов. Выше мы предъявили конструкцию, сопоставляющую некоторой норме на E (вернее, некоторой функции, соответствующей этой норме) метод решения общих выпуклых задач классов $C^{\nu_0}(G, E, m)$ на выпуклых телах $G \subset E$. Заметим теперь, что эта ситуация не очень естественна: классы типа $C^{\nu_0}(G, E, m)$ определены лишь в терминах топологии E, безотносительно к выбору конкретной нормы, задающей эту топологию. Иными словами, класс $C \in C^{\nu_0}(G, E, m)$ связан не с конкретной нормой $\|\cdot\|$ на данном линейном пространстве E, а с целым классом эквивалентных друг другу норм. Стало быть, и возможности методов решения задач из C не должны зависеть от (в общем, произвольного) выбора нормы в данном классе эквивалентности. Они должны определяться линейно-топологическими (при конечномерных E — аффинными) свойствами G, а не метрическими свойствами этого тела. В то же время ЗС-методы связаны с вполне конкретными метриками. Налицо, таким образом, некоторая идейная несогласованность рекомендуемых нами методов и «духа» задач, которые мы рассматриваем.

На самом деле никакой согласованности здесь нет: при *правильном* применении ЗС-методов к решению общих выпуклых задач возможности методов определяются линейно-топологическими свойствами G. Чтобы пояснить, что означает «правильное» применение ЗС-методов, рассмотрим практически единственно важный (а формально самый простой) случай, когда E конечномерно, т. е. E есть пространство R^n, $1 < n < \infty$. В этом случае все нормы на E эквивалентны, так что любой метод зеркального спуска, ассоциированный с любой нормой на E, можно применить для решения задач из C. Фиксируем какую-нибудь норму $\|\cdot\|$ на R^n и отвечающую ей регулярную функцию V на сопряженном пространстве (которое как линейное пространство отождествляется с R^n). В выборе V, разумеется, есть неоднозначность. Мы, впрочем, знаем, что для снижения трудоемкости метода следует выбирать регулярную функцию V по возможности более гладкой (имеющей по возможности бóльшее $\gamma_V(\cdot)$).

Вместе с нормой $\|\cdot\|$ можно рассмотреть всевозможные нормы $\|x\|^A = \|Ax\|$, где A — обратимая $n \times n$-матрица. Каждой из них отвечает *регулярная функция* $V^A(\varphi) = V((A^*)^{-1}\varphi)$. Для решения задач из C можно применять любой из $\overline{\text{ЗС}}_{V^A}$-методов. При этом их оценки трудоемкости будут $N_V\left(\dfrac{\nu-\nu_0}{4\alpha_{\|\cdot\|^A}(G)}\right)$, где $N_V(\nu)$ зависит лишь от V и ν, а $\alpha_{\|\cdot\|^A}(G)$ — асферичность G относительно $\|\cdot\|^A$ (или, что то же, асферичность AG относительно $\|\cdot\|$). «Правильное» применение $\overline{\text{ЗС}}_V$ состоит в примении того из методов $\overline{\text{ЗС}}_{V^A}$, при котором оценка трудоемкости минимальна, т. е. отвечающего тому A, при котором $\alpha_{\|\cdot\|}(AG)$ минимально (иначе говоря, AG «минимально асферично» относительно $\|\cdot\|$). Соответствующее оптимальное значение $\alpha_{\|\cdot\|}(AG)$ есть чисто аффинная характеристика G. Можно сказать, таким образом, что возможности ассоциированного с V метода зеркального спуска на самом деле определяются аффинными свойствами G. Аналогичные соображения применимы и к другим версиям $\overline{\text{ЗС}}$-методов.

Посмотрим, например, что на самом деле дают нам методы $\overline{\text{ЗС}}_p$ — методы зеркального спуска, ассоциированные с $\|\cdot\|_p$ на R^n (отвечающие отождествлению R^n с $l_p^{(n)}$). Определим для всех p, $1 \leqslant p < \infty$, аффинную характеристику $\alpha_p(G)$ выпуклого ограниченного замкнутого тела $G \subset R^n$ как $\inf\{\alpha_{\|\cdot\|_p}(AG)\} \mid A$ — обратимая $n \times n$-матрица$\}$. Пусть A_G^p — соответствующее оптимальное значение A (легко видеть, что оно существует). «Правильное» применение $\overline{\text{ЗС}}_p$ к классу $C \in C^{\nu_0}(G, R^n, m)$ требует отыскания A_G^p и применения метода, ассоциированного с l_p-нормой $\|A_G^p x\|_p$. При таком подходе оценки трудоемкости $\overline{\text{ЗС}}_p$-методов на классе C будут

$$\frac{\bar d(1)\ln n\,\alpha_1^2(G)}{(\nu-\nu_0)^2}, \quad p = 1,$$

$$\frac{\bar d(p)\,\alpha_p^2(G)}{(\nu-\nu_0)^2}, \quad 1 < p \leqslant 2,$$

$$\frac{\bar d(p)\,\alpha_p^p(G)}{(\nu-\nu_0)^p}, \quad 2 < p < \infty.$$

Эти оценки, как и полагается, зависят от аффинных свойств G.

Следующий шаг оптимизации метода должен состоять в выборе оптимального (минимизирующего оценку) p. Собственно говоря, не обязательно ограничивать себя шкалой l_p-норм и отвечающих им методов, можно привлечь и другие нормы на R^n. Конечно, решение указанной задачи «правильного» выбора ЗС-метода решения данного класса общих выпуклых задач — в общем случае крайне «дорогостоящая» процедура, даже если речь идет лишь о спектре l_p-методов. Совершенно неясно, как вычислять $\alpha_p(G)$ и находить A_G^p. Впрочем, для часто встречающихся тел «простой формы» эти проблемы могут оказаться несложными, да и в общем случае никто не заставляет нас решать их точно. Следует, однако, помнить, что пренебрежение к такой «оптимизации нормы, с которой связан метод», ведет к проигрышу в трудоемкости применяемого метода по сравнению с «оптимальным» ЗС-методом.

Аналогичные замечания можно сделать и по поводу ЗС-методов решения классов липшицевых выпуклых задач. Конечно, эти классы определены в терминах конкретной нормы $\|\cdot\|$ на E. Однако *запас задач* класса типа $C_{\text{Lip}}^{\nu_0}(G, E, \|\cdot\|, m)$ по определению зависит лишь от G и от топологии E. Таким образом, если $\|\cdot\|'$ эквивалентна $\|\cdot\|$, именно,

$$0 < \underline{a} = \inf_{x \neq 0}\frac{\|x\|}{\|x\|'} \leqslant \sup_{x \neq 0}\frac{\|x\|}{\|x\|'} = \bar a < \infty,$$

то запас задач в любых двух классах типа

$$C_{\text{Lip}}^{\nu_0}(G, E, \|\cdot\|, m) \text{ и } C_{\text{Lip}}^{\nu_0'}(G, E, \|\cdot\|', m)$$

одинаков. Классы могут различаться лишь значениями ν_0 (т. е. точностью оракулов) и определением нормирующего множителя. Благодаря этому методы решения задач одного из этих классов можно применять (при соблюдении должных предосторожностей) и к задачам другого класса. Покажем, как это делается, считая для простоты, что требуется решать задачи класса типа $C^0_{\mathrm{Lip}}(G, E, \|\cdot\|', m)$ (т. е. в нашем распоряжении — точный оракул), и предполагая G телом.

Рассмотрим конкретный класс C типа $C^0_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$. Пусть ему отвечает оракул $\mathcal{O}$. Класс C, рассматриваемый как множество задач, снабженное оракулом, является в то же время некоторым классом C' типа $C^0(G, E, \|\cdot\|', m)$ (также рассматриваемым как снабженное оракулом множество задач) *). Классы C и C' отличаются друг от друга лишь нормирующим отображением.

Пусть $L_{\|\cdot\|}(f_j)$, $L_{\|\cdot\|'}(f_j)$ — константы Липшица f_j относительно $\|\cdot\|$, соответственно $\|\cdot\|'$. Ясно, что

$$L_{\|\cdot\|'}(f_j) \leqslant \bar{a} L_{\|\cdot\|}(f_j) \quad \text{и} \quad \rho_{\|\cdot\|'}(G) \leqslant \underline{a}^{-1}\rho_{\|\cdot\|}(G).$$

Пусть

$$\varkappa = \frac{\rho_{\|\cdot\|}(G)}{a\rho_{\|\cdot\|'}(G)}.$$

Тогда при всех f

$$\varkappa \leqslant \frac{L_{\|\cdot\|}(f_j)\,\rho_{\|\cdot\|}(G)}{L_{\|\cdot\|'}(f_j)\,\rho_{\|\cdot\|'}(G)} \quad \text{и} \quad \varkappa \geqslant \frac{\underline{a}}{\bar{a}} \equiv a^{-1}(\|\cdot\|, \|\cdot\|').$$

Ясно, что приближенное решение задачи $f \in C$, рассматриваемой как задача из C', относительной (в смысле второго класса) погрешности $\bar{\nu} = \varkappa\nu$, $\nu > 0$, есть, в то же время, решение f относительной погрешности ν в смысле первого класса. Таким образом, метод решения задач класса C' относительной погрешности $\varkappa\nu$ индуцирует метод решения задач C той же трудоемкости и относительной погрешности ν. В частности, если указанный метод есть $\overline{\text{ЗС}}$-метод, ассоциированный с $\|\cdot\|'$, имеющий оценку трудоемкости $M_{\|\cdot\|'}(\bar{\nu})$, $\bar{\nu}$ — точность решения задач из C', то он индуцирует метод решения задач C с оценкой трудоемкости

$$M_{\|\cdot\|'}(\varkappa\nu) \leqslant M_{\|\cdot\|'}(a^{-1}(\|\cdot\|, \|\cdot\|')\,\nu)\ **).$$

*) Строго говоря, для этого нужно было бы еще потребовать от оракула $\mathcal{O}$, чтобы множество $\{\psi(x, f) \mid x \in G\}$ содержалось в замыкании множества $\{\psi(x, f) \mid x \in \operatorname{int} G\}$ (иначе ψ может не удовлетворять условию 2) определения $(G, 0)$-оракула относительно $\|\cdot\|'$, см. п. 2.5 гл. II). Впрочем, указанное ограничение на $\mathcal{O}$ носит чисто академический характер и, разумеется, выполнено во всех разумных ситуациях.

**) Применительно к ЗС-методам конструкция такого рода возможна для всех ν_0, а не только для $\nu_0 = 0$. Оценка трудоемкости $C^{\nu_0}_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$-метода, индуцированного $C^{\varkappa\nu_0}_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$ — ЗС-методом с трудоемкостью $M_{\|\cdot\|'}(\nu - \varkappa\nu_0)$, будет $M_{\|\cdot\|'}(\varkappa(\nu - \nu_0))$.

Резюмируя, можно сказать, что хотя классы липшицевых выпуклых задач типа $C^{\nu_0}_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$, в отличие от классов общих выпуклых, определяются в терминах конкретной нормы на E, тем не менее в выборе ЗС-метода их решения есть свобода. Она сводится к возможности применять ЗС-методы, ассоциированные с любыми нормами, эквивалентными исходной. Отчасти эта свобода была очевидна с самого начала (регулярную функцию V', соответствующую $\|\cdot\|'$, можно очевидными преобразованиями подобия превратить в регулярную функцию V, соответствующую $\|\cdot\|$. При этом оценка трудоемкости ЗС_V равна примерно $M_{V'}((\nu - \nu_0)/a(\|\cdot\|, \|\cdot\|'))$). Однако свобода, доставляемая намеченным выше подходом, несколько шире — описанный там способ превращает оценку трудоемкости $M_{V'}(\nu - \nu_0)$ в $M_{V'}(\varkappa(\nu - \nu_0))$, и $\varkappa \geqslant 1/a(\|\cdot\|, \|\cdot\|')$ (может ведь случиться и так, что $\rho_{\|\cdot\|}(G)/\rho_{\|\cdot\|'}(G) \gg \underline{a}$). Впрочем, если G — $\|\cdot\|$-шар, то оба подхода совпадают, ибо тогда $\rho_{\|\cdot\|}(G)/\rho_{\|\cdot\|'}(G) = \underline{a}$ и $\varkappa = 1/a(\|\cdot\|, \|\cdot\|')$.

4.6. Теперь займемся обоснованием анонсированных выше результатов об описанных методах.

Теорема. *Справедливы следующие утверждения.*

(*i*) *Пусть E, $\|\cdot\|$ — регулярное пространство, V — соответствующая регулярная функция. Обе версии метода $\overline{\text{ЗС}}_V$, настроенные на относительную точность ν, $2\alpha > \nu > \nu_0$, обеспечивают эту точность в решении любой задачи из $C \in C^{\nu_0}(G, E, m)$ с трудоемкостью, $\leqslant M_{V,G}(\nu - \nu_0)$.*

Аналогичные результаты (с заменой $M_{V,G}$ на M^G) справедливы и для $\bar{\Gamma}$, если E, $\|\cdot\|$ гильбертово.

(*ii*) *Каждый из методов *) п. 4.3.2, настроенный на вектор абсолютных погрешностей $\vec{\varepsilon} = (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_m) > 0$, решает всякую задачу $f \in C \in C^{\nu_0}(G, E, m)$ с абсолютными погрешностями $\leqslant \bar{\varepsilon}_j = \varepsilon_j + \nu_0 r_j(f)$ при трудоемкости $\leqslant 8\bar{V}(\nu(f, \vec{\varepsilon})/(1 + \nu_0))$. Здесь $\nu(f, \vec{\varepsilon}) = \min\limits_{0 \leqslant j \leqslant m} \varepsilon_j / r_j(f)$ и $\bar{N}(\nu)$ — оценка трудоемкости соответствующего метода (из п. 4.3.2), настроенного на относительную погрешность ν.*

Доказательство. I. Начнем с доказательства (i). Оно проводится по той же самой схеме, что и в случае липшицевых выпуклых задач (см. доказательство теоремы 3.5). Именно,

1°. Оценки трудоемкости методов выводятся из очевидного **)

*) Допустимый данными E, $\|\cdot\|$.

**) Действительно, в силу $\overline{\text{ЗС}}$ 1.4 при $j(i) = 0$ по определению $r_0(i)$ имеем

$$\hat{\delta}_i = g_i^0(\tilde{x}_i - a_i)(\|\eta_i^{j(i)}\|_*)^{-1} + (\nu - \nu_0)\{\sup_{x \in G}\langle \xi_i \mid x - \tilde{x}_i\rangle + \langle \xi_i \mid \tilde{x}_i\rangle\} \geqslant (\nu - \nu_0)\sup_{x \in G}\langle \xi_i \mid x\rangle.$$

Если же $j(i) > 0$, то по определению $j(i)$ и $r_{j(i)}(i)$ имеем

$$r_{j(i)}(i) = \sup_{x \in G}\langle \eta_i^{j(i)} \mid x - \tilde{x}_i\rangle,$$

(в силу $\overline{\text{ЗС}}1.4$) неравенства

$$\delta_i \geqslant a, \text{ если } \eta_i^{j(i)} \neq 0, \tag{4.6}$$

где $a = (\nu - \nu_0)/4\alpha$ — для метода $\overline{\text{ЗС}}_V$ и $a = (\nu - \nu_0)/2\alpha$ для метода $\bar{\Gamma}$. Напомним, что именно из такого неравенства были выведены оценки трудоемкости ЗС$_V$ и Г (отличие от липшицевого случая состоит в том, что a теперь в (4α) раз меньше, чем раньше). Вывод нужных оценок из (4.6) дословно повторяет соответствующие рассуждения § 3.

2°. Что же касается до оценок погрешности, то все они в § 3 выводились из одного-единственного условного утверждения (А): если погрешность рассматриваемого метода превышает на данной совместной задаче f величину ν, то *при всех* $i \leqslant M_f$

$$\eta_i^{j(i)} \neq 0 \quad \text{и} \quad \langle \xi_i \mid \bar{x}_i - x^* \rangle \geqslant \hat{\delta}_i, \quad \bar{x}_i = \pi_G(x_i), \tag{4.7}$$

где x^* — решение f.∎

Соотношения (4.7) приводились далее к противоречию, что и доказывало, что посылка A не может иметь места. В рассматриваемой ситуации утверждение (А) также справедливо. Это-то мы и будем доказывать. Приведением же (4.7) к противоречию мы заниматься не будем — это делается дословно так же, как и в § 3.

Итак, докажем утверждение (А). Фактически мы докажем даже утверждение

(*A'*) *Пусть* $f \in C \in C^{\nu_0}(G, E, m)$ — *совместная задача, не удовлетворяющая условию*

(*B*) *Результат* $\bar{x}$ *применения рассматриваемого метода к* f *отличен от* $*$ *и для него* $g^f_{j,\bar{x}}(\bar{x}) \leqslant (\nu - \nu_0)\ \bar{r}_j(f) + f_{j*}$, где

$$f_{j*} = \begin{cases} 0, & j \geqslant 1, \\ f_*, & j = 0. \end{cases}$$

Тогда для траектории метода на f справедливо (4.7).

так что в силу $\overline{\text{ЗС}}$ 1.4 получаем

$$\hat{\delta}_i = (\nu - \nu_0)\langle \xi_i \mid \tilde{x}_i \rangle + \frac{1}{\|\eta_i^{j(i)}\|_*} g_i^{j(i)}(\tilde{x}_i) \geqslant$$

$$\geqslant (\nu - \nu_0)\langle \xi_i \mid \tilde{x}_i \rangle + \frac{1}{\|\eta_i^{j(i)}\|_*}(\nu - \nu_0) \sup_{x \in G} \langle \eta_i^{j(i)} \mid x - \tilde{x}_i \rangle =$$

$$= (\nu - \nu_0) \sup \{\langle \xi_i \mid x \rangle \mid x \in G\}.$$

Таким образом, всегда

$$\hat{\delta}_i \geqslant (\nu - \nu_0) \sup_{x \in G} \langle \xi_i \mid x \rangle.$$

Отсюда $\delta_i \geqslant (\nu - \nu_0)/2\alpha$, ибо $\|\xi_i\|_* = 1$, а G содержит шар радиуса $1/2\alpha$ с центром в 0.

Утверждение (А$'$) доказывает сформулированные в (i) оценки относительной погрешности методов. Действительно, уже отмечалось, что соотношения (4.7) не могут выполняться при всех $i \leqslant M_f$. Пусть f совместна. Ввиду (А$'$) f не может не удовлетворять условию (В), т. е. при всех j, $0 \leqslant j \leqslant m$,

$$g^f_{j,\bar{x}}(\bar{x}) \leqslant (\nu - \nu_0)\ \bar{r}_j(f) + f_{j*}.$$

Как было отмечено в п. 4.3.2, при $j \geqslant 1$ имеем $\bar{r}_j(f) \leqslant r_j(f)$, т. е. $f_j(\bar{x}) \leqslant \nu \bar{r}_j(f)$, $j \geqslant 1$. Далее, вспоминая определение $\bar{r}_0(f)$, получим из (В) при $j = 0$: для некоторого i_1 с $j(i_1) = 0$

$$g^f_{0,\bar{x}}(\bar{x}) \leqslant (\nu - \nu_0)(\sup_{y \in G} g^f_{0,\tilde{x}_{i_1}}(y) - g^f_{0,\tilde{x}_{i_1}}(\tilde{x}_{i_1})) + f_* \leqslant$$

$$\leqslant (\nu - \nu_0)(\sup_{y \in G} f_0(y) - g^f_{0,\bar{x}}(\bar{x})) + f_*$$

(мы учли, что по $\overline{\text{ЗС}}\,2$ $g^f_{0,\bar{x}}(\bar{x}) \leqslant g^f_{0,\tilde{x}_{i_1}}(\tilde{x}_{i_1})$). Отсюда

$$(1 + (\nu - \nu_0))(g^f_{0,\bar{x}}(\bar{x}) - f_*) \leqslant (\nu - \nu_0)(\sup_G f_0 - f_*) \leqslant (\nu - \nu_0) r_0(f).$$

Таким образом,

$$g^f_{0,\bar{x}}(\bar{x}) \leqslant f_* + (\nu - \nu_0)\ r_0(f) \quad \text{и} \quad f_0(\bar{x}) \leqslant f_* + \nu r_0(f).$$

Суммируя полученные оценки $f_j(\bar{x})$, $j \geqslant 1$, и $f_0(\bar{x})$, находим, что $\nu(\bar{x}, f) \leqslant \nu$, что и требуется.

Если же f несовместна, из $\overline{\text{ЗС}}$ 2 и соотношений $r_j(i) \leqslant r_j(f)$, $j \geqslant 1$, сразу следует $\nu(\bar{x}, f) \leqslant \nu$, что и доказывает (i).

3°. Итак, докажем (А$'$). Пусть f — объект, удовлетворяющий посылке А$'$, x^* — решение f. Убедимся, что при всех $i \leqslant M_f$ имеем

$$\langle \eta_i^{j(i)} \mid \tilde{x}_i - x^* \rangle \geqslant \varkappa_i, \tag{4.8}$$

где

$$\varkappa_i = \begin{cases} g_i^0(\tilde{x}_i) - a_i + (\nu - \nu_0) r_0(i), & j(i) = 0, \\ g_i^{j(i)}(\tilde{x}_i), & j(i) \neq 0. \end{cases}$$

Действительно, если бы было

$$g_{\bar{i}}^{j(\bar{i})}(\tilde{x}_{\bar{i}}) < g_{\bar{i}}^{j(\bar{i})}(x^*) + \varkappa_{\bar{i}}, \tag{4.9}$$

то было бы $j(\bar{i}) = 0$. В самом деле, при $j(\bar{i}) > 0$ второе слагаемое в правой части (4.9) равнялось бы $g_{\bar{i}}^{j(\bar{i})}(\tilde{x}_{\bar{i}})$, тогда как первое слагаемое там — не положительное, но тогда (4.9) было бы невозможно.

Итак, $j(\bar{i}) = 0$. В этом случае (4.9) дает

$$g_{\bar{i}}^0(\tilde{x}_{\bar{i}}) < f_0(x^*) + g_{\bar{i}}^0(\tilde{x}_{\bar{i}}) - a_{\bar{i}} + (\nu - \nu_0) r_0(\bar{i}),$$

т. е.

$$a_{M_f} \leqslant f_* + (\nu - \nu_0)\ \bar{r}_0(f).$$

Отсюда и из $\overline{\text{ЗС}}$ 2 сразу следует, что f удовлетворяет (B), что противоречит определению f. Легко видеть также, что к аналогичному противоречию ведет гипотеза $\eta_{\bar{i}}^{j(\bar{i})} = 0$ при каком-нибудь $\bar{i} \leqslant M_f$. Итак, можно считать, что выполнено (4.8) и

$$\eta_i^{j(i)} \neq 0, \quad i \leqslant M_f. \tag{4.10}$$

4°. Выведем из (4.8), (4.10) требуемое утверждение (4.7). В этих условиях имеем для каждого $i \leqslant M_f$

$$\langle \xi_i \,|\, \tilde{x}_i - x^* \rangle \geqslant \frac{\varkappa_i}{\|\eta_i^{j(i)}\|_*} = \check{\delta}_i.$$

Отсюда

$$\langle \xi_i \,|\, \bar{x}_i - x^* \rangle = \langle \xi_i \,|\, \tilde{x}_i - x^* \rangle + \left(\frac{1}{\theta} - 1\right) \langle \xi_i \,|\, \tilde{x}_i \rangle \geqslant$$
$$\geqslant \check{\delta}_i + (\nu - \nu_0) \langle \xi_i \,|\, \tilde{x}_i \rangle \equiv \hat{\delta}_i,$$

что и требуется (мы учли, что $(1/\theta) - 1 = \nu - \nu_0$ по определению θ). Итак, (A′), а тем самым и пункт (i) теоремы доказаны.

II. Остается доказать (ii). В обозначениях п. 4.3.2 имеем для результата $\hat{x}_i$ процедуры $\mathscr{B}(\nu'^i)$, примененной к задаче $f \in \in \boldsymbol{C} \in \boldsymbol{C}^{\nu_0}(G, E, m)$: если f совместна, то $\hat{x}_i \neq *$, и при $\hat{x}_i \neq *$ имеем

$$g^f_{0,\, \hat{x}_i}(\hat{x}_i) \leqslant f_* + \nu^i \bar{r}_0^{(i)}(f), \quad g^f_{j,\, \hat{x}_i}(\hat{x}_i) \leqslant \nu^i \bar{r}_j^{(i)}(f), \quad j \geqslant 1$$

(при $i > 0$ это следует из (i) при совместных f и из $\overline{\text{ЗС}}2$ — при несовместных; доказательство для $i = 0$ предоставляется читателю).

По правилу окончания из п. 4.3.2 остановка работы метода происходит в момент окончания первой из процедур (пусть ее номер i_f), для которого $\nu^i \bar{r}_j^{(i)}(f) \leqslant \varepsilon_j$, $0 \leqslant j \leqslant m$. Поэтому результат применения $\overline{\mathscr{B}}$ к f удовлетворяет требованиям к точности. Оценим трудоемкость $\overline{\mathscr{B}}$ на f. Возможно, что $i_f = 0$. Тогда трудоемкость $\overline{\mathscr{B}}$ на f равна 2, т. е. она заведомо меньше $8\overline{N}(\nu(f, \vec{\varepsilon})/(1 + \nu_0))$. Пусть теперь $i_f > 0$. Тогда

$$2^{i_f + i_0 - 1} = \overline{N}(\nu^{i_f - 1}) < \overline{N}\left(\frac{\nu(f, \vec{\varepsilon})}{1 + \nu_0}\right),$$

поскольку $\nu^{i_f - 1} > \nu(f, \vec{\varepsilon}) / (1 + \nu_0)$. В то же время трудоемкость $\overline{\mathscr{B}}$ на f равна

$$2 + \sum_{s=1}^{i_f} 2^{i_0 + s} \leqslant 2^{i_0 + i_f + 2} = 8 \cdot 2^{i_f - 1 + i_0} \leqslant 8\overline{N}\left(\frac{\nu(f, \vec{\varepsilon})}{1 + \nu_0}\right),$$

что и требуется. Теорема доказана.

§ 5. Некоторые дополнительные доказательства

В помещенных ниже доказательствах без специальных пояснений используются понятия и утверждения из соответствующих разделов основного текста.

Докажем оценки (2.6) § 2 и вычислим соответствующие константы $c(p)$.

1°. Пусть $E = L_p(T, \mu)$ и $1 < p < \infty$. Тогда $E^* = L_q(T, \mu)$, $q = p/(p-1)$. Обозначим через $r(\varphi)$ функцию $\|\varphi\|_q$ на E^*. Тогда $r(\varphi)$ — выпуклая однородная (степени 1) функция на E^*. Легко видеть, что опорный функционал $r'(\varphi) \in E$ к $r(\varphi)$ определен однозначно (при $\varphi \neq 0$) и $r'(\lambda\varphi) = r'(\varphi)$ при $\lambda > 0$. Если же $r(\varphi) = 1$, то

$$[r'(\varphi)](t) = |\varphi(t)|^{q-1} \operatorname{sign} \varphi(t). \tag{5.1}$$

Убедимся в справедливости следующего соотношения: если $\varphi, \psi \in L_q$ и $\|\varphi\|_q = \|\psi\|_q = 1$, то

$$\|r'(\varphi) - r'(\psi)\|_p \leqslant \begin{cases} \dfrac{2^{(2-p)/p}}{p-1} \|\varphi - \psi\|_q, & \text{если } q > 2, \\ 2\|\varphi - \psi\|_q^{1/(p-1)}, & \text{если } q \leqslant 2. \end{cases} \tag{5.2}$$

В самом деле, пусть $\|\varphi\|_q = \|\psi\|_q = 1$. Определим

$$T_+ = \{t | \operatorname{sgn} \varphi(t) = \operatorname{sgn} \psi(t)\}, \quad T_- = T \setminus T_+.$$

Обозначая

$$\Delta(t) = |[r'(\varphi)](t) - [r'(\psi)](t)|,$$

получим из (5.1), что при $t \in T_+$

$$\Delta(t) = ||\varphi(t)|^{q-1} - |\psi(t)|^{q-1}|.$$

Пусть вначале $q \leqslant 2$. Тогда функция s^{q-1} вогнута по $s \geqslant 0$, и потому при $q \leqslant 2$

$$||s_1|^{q-1} - |s_2|^{q-1}| \leqslant |s_1 - s_2|^{q-1}.$$

Таким образом, при $q \leqslant 2$ имеем для $t \in T_+$

$$\Delta(t) \leqslant |\varphi(t) - \psi(t)|^{q-1}.$$

Ясно, далее, что при $t \in T_-$ выполнено неравенство

$$\Delta(t) \leqslant 2|\varphi(t) - \psi(t)|^{q-1}.$$

Итак, всегда (при $q \leqslant 2$)

$$\Delta(t) \leqslant 2|\varphi(t) - \psi(t)|^{q-1}.$$

Отсюда

$$\int \Delta^p(t)\, d\mu(t) \leqslant 2^p \int |\varphi(t) - \psi(t)|^{(q-1)p}\, d\mu(t) = 2^p \|\varphi - \psi\|_q^q.$$

Стало быть,

$$\|\Delta(t)\|_p \leqslant 2\|\varphi-\psi\|_q^{q/p},$$

что и требуется в (5.2) для $q \leqslant 2$.

Пусть теперь $q > 2$. Тогда функция s^{q-1} выпукла по $s \geqslant 0$, и поэтому при $s_1 \geqslant s_2 \geqslant 0$

$$s_1^{q-1} - s_2^{q-1} \leqslant \left[\frac{d}{ds}\Big|_{s=s_1}(s^{q-1})\right](s_1 - s_2) = (q-1)s_1^{q-2}(s_1 - s_2).$$

Следовательно, при $t \in T_+$ имеем, полагая $h(t) = \max\{|\varphi(t)|, |\psi(t)|\}$:

$$\Delta(t) \leqslant (q-1)\, h^{q-2}(t)\, |\varphi(t) - \psi(t)|. \tag{5.3}$$

Ясно, что (5.3) выполнено и для $t \in T_-$. Поэтому

$$\int \Delta^p(t)\, d\mu(t) = (q-1)^p \int h^{(q-2)p}(t)\, |\varphi(t) - \psi(t)|^p\, d\mu(t).$$

Полагая

$$\sigma \equiv \frac{q}{p} > 1 \quad \text{и} \quad \tau \equiv \frac{\sigma}{\sigma - 1} = \frac{q}{(q-2)p}$$

и применяя неравенство Гельдера, получим

$$\int \Delta^p(t)\, d\mu(t) \leqslant$$

$$\leqslant (q-1)^p \left[\int h^{(q-2)p\tau}(t)\, d\mu(t)\right]^{\frac{1}{\tau}} \left[\int |\varphi(t) - \psi(t)|^{p\sigma} d\mu(t)\right]^{\frac{1}{\sigma}} =$$

$$= (q-1)^p \left[\int h^q(t)\, d\mu(t)\right]^{\frac{1}{\tau}} \left[\int |\varphi(t) - \psi(t)|^q d\mu(t)\right]^{\frac{1}{\sigma}} \leqslant$$

$$\leqslant (q-1)^p\, 2^{\frac{1}{\tau}} \|\varphi - \psi\|_q^{\frac{q}{\sigma}} = (q-1)^p \cdot 2^{\frac{(q-2)p}{q}} \|\varphi - \psi\|_q^p,$$

т. е.

$$\|\Delta(t)\|_p \leqslant (q-1)\cdot 2^{\frac{q-2}{q}} \|\varphi - \psi\|_q = \frac{2^{\frac{2-p}{p}}}{p-1} \|\varphi - \psi\|_q,$$

что завершает доказательство (5.2).

2°. Оценим теперь модуль гладкости ω_p функции $V_p(\varphi)$. Пусть $\varphi, \psi \in L_q$ и $\|\varphi - \psi\|_q = t$. Оценим $\|V_p'(\varphi) - V_p'(\psi)\|_p$. Можно считать, $\|\psi\|_q \geqslant \|\varphi\|_q$. Пусть $\bar{\varphi} = \varphi/\|\varphi\|_q$, $\bar{\psi} = \psi/\|\psi\|_q$ ($\bar{\varphi} = 0$ при $\varphi = 0$, $\bar{\psi} = 0$ при $\psi = 0$). Пусть вначале $\|\varphi\|_q > 0$. Тогда

$$\|\bar{\varphi} - \bar{\psi}\|_q = \left\|\frac{\|\psi\|_q \varphi - \|\varphi\|_q \psi}{\|\varphi\|_q \|\psi\|_q}\right\|_q \leqslant$$

$$\leqslant \frac{\|\varphi - \psi\|_q \|\varphi\|_q + \|\varphi\|_q \|\psi - \varphi\|_q}{\|\varphi\|_q \|\psi\|_q} = \frac{2\|\varphi - \psi\|_q}{\|\psi\|_q}.$$

Ясно, что это же неравенство верно и для $\varphi = 0$ (здесь $0/0 = 0$). Итак,

$$\|\bar{\psi} - \bar{\varphi}\|_q \leqslant \frac{2\|\psi - \varphi\|_q}{\|\psi\|_q}. \tag{5.4}$$

Пусть вначале $\|\varphi\|_q \geqslant 1$. Тогда из (5.4)

$$\|\bar{\varphi} - \bar{\psi}\|_q \leqslant 2\|\varphi - \psi\|_q.$$

В то же время $V_p'(\varphi) = r'(\bar{\varphi})$, $V_p'(\bar{\psi}) = r'(\bar{\psi})$, и, ввиду (5.2), заведомо выполнено неравенство

$$\|V_p'(\varphi) - V_p'(\psi)\|_p \leqslant 2 \begin{cases} \dfrac{2\cdot 2^{\frac{(2-p)}{p}}}{p-1} \|\varphi - \psi\|_q, & 1 < p \leqslant 2, \\ 2\cdot 2^{\frac{1}{p-1}} \|\varphi - \psi\|_q^{\frac{1}{p-1}}, & p > 2. \end{cases} \tag{5.5}$$

Пусть теперь $\|\psi\|_q \geqslant 1$, $\|\varphi\|_q \leqslant 1$. Тогда по-прежнему $\|\bar{\varphi} - \bar{\psi}\|_q \leqslant 2\|\varphi - \psi\|_q$. В то же время

$$V_p'(\psi) = r'(\bar{\psi}),\quad V_p'(\varphi) = \|\varphi\|_q r'(\bar{\varphi})$$

и

$$\|V_p'(\psi) - V_p'(\varphi)\|_p \leqslant$$

$$\leqslant (1 - \|\varphi\|_q) + \|r'(\bar{\varphi}) - r'(\bar{\psi})\|_p \leqslant \|\varphi - \psi\|_p + \|r'(\bar{\varphi}) - r'(\bar{\psi})\|_p.$$

Кроме того, ясно, что $\|V_p'(\varphi) - V_p'(\psi)\|_q \leqslant 2$ при всех φ и ψ. Отсюда и из (5.2) сразу следует (5.5).

Остается разобрать случай $\|\psi\|_q \leqslant 1$. В этом случае

$$V_p'(\varphi) - V_p'(\psi) = (\|\varphi\|_p - \|\psi\|_q)\, r'(\bar{\varphi}) + \|\psi\|_q\, (r'(\bar{\varphi}) - r'(\bar{\psi}))$$

(здесь $r'(0) = 0$). Таким образом, в силу (5.2) и (5.4)

$$\|V_p'(\varphi) - V_p'(\psi)\|_p \leqslant$$

$$\leqslant \|\varphi - \psi\|_q + \|\psi\|_q \begin{cases} \dfrac{2\cdot 2^{\frac{(2-p)}{p}}}{p-1} \dfrac{\|\varphi - \psi\|_q}{\|\psi\|_q}, & p \leqslant 2, \\ 2\cdot 2^{\frac{1}{p-1}} \dfrac{\|\varphi - \psi\|_q^{\frac{1}{p-1}}}{\|\psi\|_q^{\frac{1}{p-1}}}, & p > 2, \end{cases}$$

откуда, как и выше, следует (5.5). Итак,

$$\omega_p(t) \leqslant \begin{cases} \dfrac{2^{\frac{2+p}{p}}}{p-1}\, t, & 1 < p \leqslant 2, \\ 2^{\frac{2p-1}{p}}\, t^{\frac{1}{p-1}}, & p > 2. \end{cases} \tag{5.6}$$

Отсюда следует, что

$$\gamma_p(s) \geqslant \begin{cases} (p-1)\,2^{-\frac{2+p}{p}}s, & 1 < p \leqslant 2, \\ 2^{\frac{(1-2p)(p-1)}{p}} s^{p-1}, & p > 2, \end{cases}$$

и (2.6) доказано. При этом можно взять

$$c(p) = \begin{cases} (p-1)\,2^{-\frac{2+p}{p}} \geqslant \frac{1}{8}(p-1), & 1 < p \leqslant 2, \\ 2^{\frac{(1-2p)(p-1)}{p}} \geqslant 4^{-(p-1)}, & p > 2. \end{cases} \tag{5.7}$$

3°. Теперь займемся случаем $p = 1$. Пусть $E = l_1^{(n)}$, $1 < n < \infty$, и $\tau \geqslant 2$, а $\bar{p} = \tau/(\tau - 1)$. При $\varphi, \psi \in l_\infty^{(n)}$ в силу (5.6) мы имеем

$$\| (V^\tau)'(\varphi) - (V^\tau)'(\psi) \|_{\bar{p}} \leqslant \frac{8}{\bar{p}-1} \| \varphi - \psi \|_\tau.$$

Далее, ясно, что

$$\| \varphi - \psi \|_\tau \leqslant n^{1/\tau} \| \varphi - \psi \|_\infty$$

и

$$\| (V^\tau)'(\varphi) - (V^\tau)'(\psi) \|_1 \leqslant n^{1/\tau} \| (V^\tau)'(\varphi) - (V^\tau)'(\psi) \|_{\bar{p}}.$$

Итак,

$$\| (V^\tau)'(\varphi) - (V^\tau)'(\psi) \|_1 \leqslant n^{2/\tau} \frac{8}{\bar{p}-1} \| \varphi - \psi \|_\infty. \tag{5.8}$$

Выберем теперь τ так, чтобы минимизировать $n^{2/\tau}/(\bar{p}-1)$ при условии $\tau \geqslant 2$. Безусловная (без учета ограничения $\tau \geqslant 2$) минимизация дает

$$\tau = \ln n + \sqrt{\ln^2 n - 2\ln n}. \tag{5.9}$$

Чтобы упростить выбор τ, возьмем приближенное значение корня $\tau = 2\ln n$. Для этого выбора τ величина $n^{2/\tau} 8/(\bar{p}-1)$ будет равна

$$8e\,(\tau - 1) = 8e\,(2\ln n - 1) \leqslant 16\, e \ln n.$$

Итак, при выборе $\tau = 2\ln n$ (считаем $n > 2$) получим в (2.11) $c(1) = 1/16e$. При $n = 2$ можно взять $\tau = 2$, при этом (2.11) тоже будет иметь место с указанным $c(1)$. Итак, для пространства $l_1^{(n)}$ регулярную функцию следует выбирать из условия

$$\tau = \max\{2, 2\ln n\}; \text{ при этом } c(1) = \frac{1}{16e}. \tag{5.10}$$

Глава IV

СЛОЖНОСТЬ КЛАССОВ ОБЩИХ ВЫПУКЛЫХ ЗАДАЧ (ТОЧНЫЙ ОРАКУЛ ПЕРВОГО ПОРЯДКА)

В этой главе строятся нижние оценки сложности классов общих (липшицевых) выпуклых задач (§§ 1, 2). Эти оценки сопоставляются с верхними оценками сложности, доставляемыми результатами двух предыдущих глав о трудоемкости описанных там оптимизационных методов. Такое сопоставление позволяет охарактеризовать поведение сложностных функций рассматриваемых классов и очертить области теоретической оптимальности (точнее, субоптимальности) методов решения общих и липшицевых выпуклых задач из глав II — III (§ 3). В §§ 4—6 содержатся доказательства сформулированных в §§ 1, 2 оценок. Результаты главы в основном базируются на работе авторов [30].

§ 1. Сложность классов общих выпуклых задач

В этом параграфе формулируются и обсуждаются двусторонние оценки стохастической и детерминированной сложности классов общих выпуклых задач.

1.1. Введем вначале некоторые геометрические характеристики выпуклых тел в банаховых пространствах. Эти характеристики позволяют сравнивать указанные тела с L_p-шарами и в конечном счете дают возможность сформулировать нужные оценки сложности классов выпуклых задач. Пусть E — вещественное банахово пространство, n, $1 \leqslant n \leqslant \infty$, — его размерность, G — ограниченное выпуклое замкнутое тело в E, $\|\cdot\|$ — норма на E.

Начнем с определения *k-мерной p-асферичности* $\alpha_{p,k}(\|\cdot\|)$ *нормы* $\|\cdot\|$ *на* E. Здесь k — натуральное число, не превышающее $\dim E$, а p — параметр, $1 \leqslant p \leqslant \infty$. Величина $\alpha_{p,k}(\|\cdot\|)$ определяется следующим образом. Рассмотрим k-мерное подпространство E' в E. На E' можно многими способами задать $\|\cdot\|_p$-норму (каждый такой способ отвечает фиксации базиса в E'). Пусть такая норма фиксирована. По отношению к ней единичный шар (по норме $\|\cdot\|$) пространства E' имеет некоторую асферичность α. Нижняя грань этих асферичностей по всевозможным выборам $l_p^{(k)}$-норм на E' назовем *p-асферичностью* E' и обозначим $\alpha_p(E')$. Нижняя грань величин $\alpha_p(E')$ по всем k-мерным подпространствам E и есть *k-мерная p-асферичность* E $\alpha_{p,k}(\|\cdot\|)$.

Величина $\alpha_{p,k}(\|\cdot\|)$ есть характеристика пространства E, $\|\cdot\|$. Теперь введем аналогичную характеристику выпуклых подмножеств E. Положим

$$\alpha_{p,\|\cdot\|,k}(G) = \alpha_{p,k}(\|\cdot\|)\,\alpha_{\|\cdot\|}(G), \tag{1.1}$$

$$\alpha_{p,k}(G) = \inf\{\alpha_{p,\|\cdot\|',k}(G) \mid \|\cdot\|' \text{ — допустимая норма на } E\}. \tag{1.2}$$

Положим также

$$\alpha(G) = \inf\{\alpha_{\|\cdot\|'}(G) \mid \|\cdot\|' \text{ — допустимая норма на } E\}.$$

Заметим, что $\alpha_{p,\|\cdot\|,k}(G)$ — метрическая характеристика G (более того, она не меняется при преобразованиях подобия G), тогда как $\alpha_{p,k}(G)$ — «аффинная» характеристика G (она не меняется при непрерывных аффинных преобразованиях G). Некоторые полезные свойства этих характеристик сформулированы в следующей серии упражнений.

Упражнения. Пусть E, $\|\cdot\|$, G — те же, что и выше, $p, p' \in [1, \infty]$.

1. Докажите, что

$$\left.\begin{aligned} &\alpha_{p,\|\cdot\|,k}(G) \leqslant k^{\left|\frac{1}{p}-\frac{1}{p'}\right|}\alpha_{p',\|\cdot\|,k}(G),\\ &\alpha_{p,k}(G) \leqslant k^{\left|\frac{1}{p}-\frac{1}{p'}\right|}\alpha_{p',k}(G). \end{aligned}\right\} \tag{1.3}$$

2. Пусть $\dim E = n < \infty$. Докажите, что среди содержащихся в G симплексов существует симплекс максимального объема. Вершины этого симплекса лежат на ∂G. Кроме того, гиперплоскости, проходящие через вершины симплекса параллельно противолежащим этим вершинам граням, опорны для G.

3. Выведите из 2, что при $\dim E = n < \infty$

$$\alpha_{\infty,n}(G) \leqslant 2n. \tag{1.4}$$

4. Выведите из пп. 1—3, что всегда

$$\left.\begin{aligned} &1)\quad \alpha_{\infty,\|\cdot\|,k}(G) \leqslant 2k\alpha_{\|\cdot\|}(G),\\ &2)\quad \alpha_{p,\|\cdot\|,k}(G) \leqslant 2k^{1+\frac{1}{p}}\alpha_{\|\cdot\|}(G), \quad 1 < p < \infty; \end{aligned}\right\} \tag{1.5}$$

при $\dim E = n < \infty$

$$3)\quad \alpha_{p,n}(G) \leqslant n^{1+\frac{1}{p}}, \quad 1 \leqslant p \leqslant \infty.$$

На самом деле для $p = 2$ справедливы более точные оценки [7]: при всех $k \leqslant \dim E$

$$\alpha_{2,\|\cdot\|,k}(G) \leqslant \sqrt{k}\,\alpha_{\|\cdot\|}(G); \tag{1.6}$$

при $\dim E = n < \infty$

$$\alpha_{2,n}(G) \leqslant n. \tag{1.7}$$

Более того, как показал Дворецкий [6], для всякого целого k и числа $\delta > 0$ существует такое $n = n(k, \delta)$, что при $\dim E > n(k, \delta)$

$$\alpha_{2,\|\cdot\|,k}(G) \leqslant (1+\delta)\alpha_{\|\cdot\|}(G). \tag{1.8}$$

В частности, при $\dim E = \infty$

$$\alpha_{2,\|\cdot\|,k}(G) = \alpha_{\|\cdot\|}(G). \tag{1.9}$$

1.2. Сформулируем первый из основных результатов о сложности классов выпуклых задач. Предварительно введем некоторые обозначения. Обозначим через $\Phi_{n,p}(\nu)$, $1 \leqslant p < \infty$, верхнюю оценку трудоемкости настроенного на точность $\nu < 1$ метода ЗС_p решения липшицевых выпуклых задач класса типа

$$C^0_{\text{Lip}}(G, L_p(T, \mu), \|\cdot\|_p, m),$$

отвечающую случаю $\dim L_p(T, \mu) = n$, $n = 1, 2, \ldots, \infty$. На самом деле $\Phi_{n,p}(\nu)$ не зависит от n, если только $p \neq 1$. Пусть еще $\Phi_{n,\infty}(\nu)$ — верхняя оценка трудоемкости настроенного на точность $\nu < 1$ метода МЦТ, рассматриваемого на классе задач типа $C^0(\cdot, R^n, m)$, $n = 1, 2, \ldots, \infty$. Для $n = \infty$ $\Phi_{n,\infty}(\nu) \equiv \infty$ по определению.

В соответствии с результатами § 3 гл. II и § 3 гл. III имеем (здесь и далее $0 < \nu < 1$; только этот случай нетривиален)

■ $p = 1$,

$$\Phi_{n,1}(\nu) = \begin{cases} \dfrac{D(1)\ln(n+1)}{\nu^2}, & 1 \leqslant n < \infty,\\ +\infty, & n = \infty; \end{cases}$$

$1 < p \leqslant 2$,

$$\Phi_{n,p}(\nu) = \frac{D(p)}{\nu^2};$$

$2 < p < \infty$,

$$\Phi_{n,p}(\nu) = \frac{D(p)}{\nu^p};$$

$p = \infty$,

$$\Phi_{n,\infty}(\nu) = D(\infty)\,n\ln\frac{2}{\nu}. \quad \square \tag{1.10}$$

Здесь величина $D(p) < \infty$ зависит лишь от p, $1 \leqslant p \leqslant \infty$.

Введем еще набор функций $\varphi_{n,p}(\nu)$, определенных для $1 > \nu > 0$ при всех $n = 1, 2, \ldots, \infty$ и всех p, $1 \leqslant p \leqslant \infty$:

■ $p = 1$, $1 \leqslant n < \infty$,

$$\varphi_{n,1}(\nu) = \begin{cases} 1, & \nu \geqslant 1/4,\\ d(1)\dfrac{\ln(n+1)}{\nu^2}, & 1/4 > \nu \geqslant n^{-1/8},\\ \dfrac{d(1)}{\nu^2}, & \min\{1/4, n^{-1/8}\} > \nu \geqslant n^{-1/2},\\ d(1)\,n, & \nu < \min\{n^{-1/2}, 1/4\}; \end{cases}$$

$p=1,\ n=\infty$,

$$\varphi_{n,1}(\nu)=\begin{cases}1, & \nu\geqslant {}^1/_4,\\ +\infty, & \nu<{}^1/_4;\end{cases}$$

$1<p<\infty$,

$$\varphi_{n,p}(\nu)=d(\infty)\min(n,\varphi_p(\nu)),$$

где

$$\varphi_p(\nu)=\begin{cases}1, & \nu\geqslant {}^1/_4,\\ d(p)\,\nu^{-\max(2,\,p)}, & \nu<{}^1/_4;\end{cases}$$

$p=\infty,\ 1\leqslant n<\infty$,

$$\varphi_{n,\infty}(\nu)=\begin{cases}1, & \nu\geqslant {}^1/_4,\\ d(\infty)\,n\ln\left(\dfrac{1}{\nu}\right), & \nu<{}^1/_4;\end{cases}$$

$p=\infty,\ n=\infty$,

$$\varphi_{n,\infty}(\nu)=\begin{cases}1, & \nu\geqslant {}^1/_4,\\ \infty, & \nu<{}^1/_4.\end{cases}\qquad \square \quad (1.11)$$

В (1.11) величины $d(p)$ положительны и зависят только от p, $1\leqslant p\leqslant\infty$. Набор этих величин для всех p полностью определяет семейство функций $\varphi_{n,p}(\nu)$. Для $2<p<\infty$ $d(p)\geqslant 4^{-p}$.

Пусть C — какой-нибудь класс задач типа $C^0(G, E, m)$. Обозначим через $N(\nu)$ $(N_*(\nu))$ детерминированную сложностную (соответственно $*$-сложностную, см. п. 4.4 гл. I) функцию этого класса. Пусть, далее, $\mathcal{O}$ — произвольный локальный детерминированный оракул для поля задач $C(G, E, m)$ и $N^{\mathcal{O}}(\nu)$, $\tilde{N}^{\mathcal{O}}(\nu)$ — соответственно детерминированная и стохастическая сложность класса задач $C^{\mathcal{O}}$, получающегося из C заменой исходного оракула на оракул $\mathcal{O}$.

Имеет место следующая теорема.

Теорема об оценке сложности классов общих выпуклых задач. *Существует функция $d(p)>0$, $1\leqslant p\leqslant\infty$, такая, что для соответствующих ей (в силу (1.11)) функций $\varphi_{n,p}(\nu)$ справедливы следующие утверждения: функции $N(\nu)$, $N_*(\nu)$, $N^{\mathcal{O}}(\nu)$, $\tilde{N}^{\mathcal{O}}(\nu)$ допускают следующие оценки* *).

А. *Оценки снизу.*

А.1. При всяком натуральном $k\leqslant n$ и всяком p, $1\leqslant p\leqslant\infty$, имеем

$$N^{\mathcal{O}}(\nu)\geqslant\varphi_{k,p}(\alpha_{p,k}(G)\ \nu)\qquad (1.12)$$

и

$$\tilde{N}^{\mathcal{O}}(\nu)\geqslant\frac{\varphi_{k,p}(8\alpha_{p,k}(G)\,\nu)}{1+\left[\ \ln]\,\varphi_{k,p}(8\alpha_{p,k}(G)\,\nu)\,[\ \ \right]_+}.\qquad (1.13)$$

*) Если $\mathcal{O}$ локален, но не обязательно детерминирован, то нижние оценки $\tilde{N}^{\mathcal{O}}(\cdot)$ из п. А теоремы справедливы и в этом случае (доказывать этого мы не будем).

А.2. Пусть $E=L_p(T,\mu)$ и $1<p<\infty$. Существует такое $k_p(t)<\infty$, $t>0$, что при $n\geqslant k_p(\alpha_{\|\cdot\|_p}(G)\,\nu)$ имеем

$$\tilde{N}^{\mathcal{O}}(\nu)\geqslant\varphi_{n,p}(28\alpha_{\|\cdot\|_p}(G)\,\nu).\qquad (1.14)$$

А.3. Существует такое $\bar{k}_2(t)<\infty$ при $t>0$, что при $n>\bar{k}_2(\alpha_{\|\cdot\|}(G)\,\nu)$ имеем

$$\tilde{N}^{\mathcal{O}}(\nu)\geqslant\varphi_{\infty,2}(\alpha_{\|\cdot\|}(G)\,\nu).\qquad (1.15)$$

А.4. Если E нерефлексивно, то

$$N^{\mathcal{O}}(\nu)=+\infty \text{ при } \nu\leqslant\nu(G), \text{ где } \nu(G)>0.\qquad (1.16)$$

А.5. Если $m>0$, то.

$$N_*(\nu)=+\infty \text{ при } \nu<1.\qquad (1.17)$$

В. Оценки сверху.

В.1. При $n<\infty$

$$N(\nu)\leqslant\min\left\{\Phi_{n,\infty}(\nu);\ \inf_{1\leqslant p<\infty}\Phi_{n,p}\left(\frac{\nu}{4\alpha_{p,n}(G)}\right)\right\}.\qquad (1.18)$$

В.2. При $E=L_p(T,\mu)$, $1<p<\infty$:

$$N(\nu)\leqslant\Phi_{\infty,p}\left(\frac{\nu}{4\alpha_{\|\cdot\|}(G)}\right).\qquad (1.19)$$

Доказательству группы утверждений А сформулированной теоремы посвящены §§ 4—6. Утверждения В нам уже знакомы. Действительно, оценки (1.18) связаны с возможностью применять к задачам из C метод МЦТ из § 3 гл. II (это приводит к неравенству $N(\nu)\leqslant\Phi_{n,\infty}(\nu)$) и методы из семейства $\overline{\text{ЗС}}_p$, $1\leqslant p<\infty$ (см. п. 4.5 гл. III; отметим, что в этом пункте величины $\alpha_{p,n}(G)$ из (1.18) обозначались просто $\alpha_p(G)$). Утверждения о трудоемкости $\overline{\text{ЗС}}_p$ приводят к оценкам

$$N(\nu)\leqslant\Phi_{n,p}\left(\frac{\nu}{4\alpha_{p,n}(G)}\right),\quad 1\leqslant p<\infty.$$

Наконец, (1.19) следует из возможности применять $\overline{\text{ЗС}}_p$, см. п. 4.4 гл. III.

1.3. Теорема 1.2 содержит довольно большую информацию о сложности классов выпуклых задач, но информация эта по необходимости представлена в неудобном, весьма сжатом виде. Выпишем оценки (1.12) — (1.19) в более наглядной табличной форме. Таблица 1 оценок относится к единственно важному с практической точки зрения случаю конечномерного $E=R^n$. Носящие академический характер бесконечномерные рассмотрения проводятся в п. 1.5.

Таблица 1

G	$N(\nu) \leqslant$	$N(\nu) \geqslant$	$N^0(\nu) \geqslant$	Асимптотика $N(\nu)$ по $\nu \to 0$	Асимптотика $N(\nu)$ по $n \to \infty$
—	$\Phi_\infty(\nu) \equiv c_\infty n \, \left]\ln\left(\frac{1}{\nu}\right)\right[$	$\lambda_\infty \Phi_\infty(\nu)$ при $\nu \leqslant n^{-2}$	$\lambda_\infty \frac{\Phi_\infty(\nu)}{\ln \Phi_\infty(\nu)}$ при $\nu \leqslant n^{-2}$	$\frac{n \ln(1/\nu)}{\mu_\infty} \leqslant N(\nu) \leqslant$ $\leqslant \mu_\infty n \ln(1/\nu)$ при $\nu \leqslant n^{-2}$	—
$\alpha_{p,n}(G) \leqslant \alpha$, $1 < p < \infty$, $s = \max(2, p)$, $\bar{\alpha} = \min\{2n, \alpha n^{1/p}\}$	$\Phi_\infty(\nu) \equiv c_\infty n] \ln(1/n) [$	$\lambda_\infty \Phi_\infty(\nu)$ при $\bar{\alpha}\nu < \frac{1}{4}$ и $\nu \leqslant \bar{\alpha}^{-2}$	$\lambda_\infty \frac{\Phi_\infty(\nu)}{\ln \Phi_\infty(\nu)}$ при $\bar{\alpha}\nu \leqslant \frac{1}{32}$ и $\nu \leqslant \bar{\alpha}^{-2}$	$\frac{n \ln(1/\nu)}{\mu_\infty} \leqslant N(\nu) \leqslant$ $\leqslant \mu_\infty n \ln(1/\nu)$ при $\bar{\alpha}\nu < \frac{1}{4}$ и $\nu \leqslant \bar{\alpha}^{-2}$	$\lambda_p \Phi_p(\alpha\nu) \leqslant N(\nu) \leqslant$ $\leqslant \Phi_p(\alpha\nu)\, \alpha^{2s}$ при $\alpha\nu \geqslant n^{-1/s}$
	$\alpha^{2s} \Phi_p(\alpha\nu) \equiv c_p \frac{\alpha^s}{\nu^s}$	$\lambda_p \Phi_p(\alpha\nu)$ при $\alpha\nu \geqslant n^{-1/s}$	$\lambda_p \frac{\Phi_p(\alpha\nu)}{1 + [\ln \Phi_p(\alpha\nu)]_+}$ при $\alpha\nu \geqslant n^{-1/s}$		
			$\lambda_p \Phi_p(\alpha\nu)$ при $n \geqslant k_p(\alpha\nu)$		
$\alpha_{1,n}(G) \leqslant \alpha$	$\Phi_\infty(\nu) \equiv c_\infty n] \ln(1/\nu) [$	$\lambda_\infty \Phi_\infty(\nu)$ при $\nu \leqslant n^{-2}$	$\lambda_\infty \frac{\Phi_\infty(\nu)}{\ln \Phi_\infty(\nu)}$ при $\nu \leqslant n^{-2}$	$\frac{n \ln(1/\nu)}{\mu_\infty} \leqslant N(\nu) \leqslant$ $\leqslant \mu_\infty n \ln(1/\nu)$ при $\nu \leqslant n^{-2}$	$\lambda_1 \Phi_1^{(n)}(\alpha\nu) \leqslant N(\nu) \leqslant$ $\leqslant \alpha^4 \Phi_1^{(n)}(\alpha\nu)$ при $\alpha\nu \geqslant n^{-1/8}$ и $\frac{1}{4} > \alpha\nu$
	$\alpha^4 \Phi_1^{(n)}(\alpha\nu) \equiv c_1 \ln n \cdot \frac{\alpha^2}{\nu^2}$	$\lambda_1 \Phi_1^{(n)}(\alpha\nu)$ при $\frac{1}{4} > \alpha\nu \geqslant n^{-1/8}$	$\lambda_1 \frac{\Phi_1^{(n)}(\alpha\nu)}{1 + [\ln \Phi_1^{(n)}(\alpha\nu)]_+}$ при $\frac{1}{32} > \alpha\nu > n^{-1/8}$		
		$\frac{\lambda_1}{(\alpha\nu)^2}$ при $n^{-1/8} > \alpha\nu \geqslant n^{-1/2}$	$\frac{\lambda_1}{(\alpha\nu)^2 \ln(1/\alpha\nu)}$ при $n^{-1/8} > \alpha\nu \geqslant n^{-1/2}$		
			$\frac{\lambda_1}{(\alpha\nu)^2}$ при $n > \bar{k}_2(\alpha\nu)$		
$\alpha_{\infty,n}(G) \leqslant \alpha$	$\Phi_\infty(\nu) \equiv c_\infty n] \ln(1/\nu) [$	$\lambda_\infty \Phi_\infty(\nu)$ при $\alpha\nu < \frac{1}{4}$ и $\nu \leqslant \alpha^{-2}$	$\lambda_\infty \frac{\Phi_\infty(\nu)}{\ln \Phi_\infty(\nu)}$ при $\alpha\nu < \frac{1}{32}$ и $\nu \leqslant \alpha^{-2}$	$\frac{n \ln(1/\nu)}{\mu_\infty} \leqslant N(\nu) \leqslant$ $\leqslant \mu_\infty n \ln(1/\nu)$ при $\alpha\nu < \frac{1}{4}$ и $\nu \leqslant \alpha^{-2}$	$\frac{n \ln(1/\nu)}{\mu_\infty} \leqslant N(\nu) \leqslant$ $\leqslant \mu_\infty n \ln(1/\nu)$ при $\alpha\nu > \frac{1}{4}$ и $\nu \leqslant \alpha^{-2}$

В таблице 1 первая графа содержит характеристику геометрии G, а остальные графы — оценки детерминированной сложности $N(\nu)$ и стохастической сложности $\tilde{N}(\nu)$ класса C типа $C^0(G, R^n, m)$. Для сложности $\tilde{N}(\nu)$ приводятся только нижние оценки. Верхние оценки $\tilde{N}(\nu)$ — те же, что для $N(\nu)$, ибо, очевидно, $\tilde{N}(\nu) \leqslant N(\nu)$. Приводимые нижние оценки $N(\nu)$ и $\tilde{N}(\nu)$ верны также и для всех сложностей $N^{\mathfrak{G}}(\nu)$, $\tilde{N}^{\mathfrak{G}}(\nu)$. Через $\sigma(p)$, $\lambda(p)$, $\delta(p)$, $\mu(p)$ в таблице обозначены некоторые положительные конечные при всех p функции p. Приведенная таблица не нуждается в специальном обосновании. Она является непосредственным следствием оценок (1.12) — (1.19). Наконец, везде в таблице предполагается, что $0 < \nu < 1$.

1.4. Прокомментируем приведенные результаты на очевидных следствиях данных таблицы 1.

Сложностные функции класса C типа $C^0(G, R^n, m)$ зависят от точности ν и от параметров — от списка объектов, характеризующих класс C, т. е., в первую очередь, от G, а также и от оракула $\mathcal{O}$, отвечающего классу C, и от m.

Мы не умеем точно вычислять сложность (это, по-видимому, безнадежная задача для сколько-нибудь общих ситуаций), а можем лишь оценивать ее. Наши оценки совпадают со сложностью в лучшем случае с точностью до абсолютной мультипликативной константы. Надо заметить, что стремиться здесь к большему, по-видимому, и не стоит — следует учесть, что сама информационная сложность есть достаточно грубая характеристика интуитивно понимаемой сложности вычислительного метода.

В соответствии со сказанным естественные вопросы о сложности, на которые мы должны (и в широком диапазоне ситуаций действительно можем) отвечать,— это вопросы о характере зависимости сложности от определяющих ее параметров, т. е. об асимптотике сложности по этим параметрам. Посмотрим, что можно сказать по этому поводу.

1.4.1. Рассмотрим вначале вопрос об асимптотике сложности по $\nu \to 0$ (именно, такого рода анализом традиционно ограничивается изучение эффективности численных методов). Хорошо известно, что при малых ν $N(\nu) \geqslant O(\ln(1/\nu))$ (оценка может быть получена из анализа уже одномерной ситуации). Результаты теоремы 1.2 уточняют этот факт и дают исчерпывающую информацию об асимптотике сложности по ν: при всех достаточно малых ν (именно, при $\nu < \nu(G) > 0$) с точностью до *абсолютной* мультипликативной константы $N(\nu) \sim n \ln(1/\nu)$ *). Момент установления асимптотики (т. е. величина $\nu(G)$) зависит от аффинных свойств тела G. Для *параллелепипедов* G асимптотика устанавливается «сразу»: $\nu(G) = {}^1/_4$. В общем случае $\nu(G)$ зависит от «акубичности» G — величины $\alpha_{\infty,n}(G) \equiv \alpha$. Именно, $\nu(G) = \min\{1/4\alpha;\ 1/\alpha^2\}$. При этом имеется общий для всех тел G данной размерности момент установления асимптотики $\bar{\nu}(n) = n^{-2}$ (это происходит в силу неравенства $\alpha_{\infty,n}(G) \leqslant 2n$, см. (1.4)).

*) Запись $a \sim b$ означает, что для некоторых абсолютных положительных констант c и d имеем $d \leqslant b/a \leqslant c$.

Результаты, касающиеся стохастической сложности $\tilde{N}(\nu)$, не столь точны. Асимптотика по $\nu \to +0$ верхней оценки отличается от асимптотики нижней, правда, незначительно — при $\nu \to +0$ справедливы оценки

$$\frac{n \ln \frac{1}{\nu}}{\ln\left(n \ln \frac{1}{\nu}\right)} \lesssim \tilde{N}(\nu) \lesssim n \ln \frac{1}{\nu}$$

(здесь $a \lesssim b$ означает, что b/a не меньше абсолютной положительной константы). В частности, в асимптотике по $\nu \to +0$

$$\tilde{N}(\nu) \gtrsim \frac{N(\nu)}{\ln N(\nu)},$$

т. е. рандомизация поиска при высоких точностях если и позволяет увеличить эффективность методов, то разве что «логарифмически». Заметим еще, что независимо от вида выпуклого тела $G \subset R^n$ моменты установления всех указанных асимптотик заведомо оцениваются снизу величиной порядка n^{-2}.

1.4.2. Рассмотрим теперь зависимость сложности от других параметров. Что касается конкретного выбора (int G, 0)-оракула и числа ограничений m, то оценки теоремы вовсе на них не реагируют (что, если подумать, весьма естественно). Таким образом, следует выяснить роль G. Собственно говоря, G — весьма неудобный (нечисловой) параметр, так что отследить его роль в полном виде крайне затруднительно. Разумно выделить некоторые числовые характеристики G (они должны быть его аффинными инвариантами в силу аффинной природы рассматриваемых классов) и понять, как зависит от этих характеристик сложность.

Важнейшая (и самая простая) из таких характеристик — это размерность n тела G. И теоретически, и практически весьма полезно знать, какова асимптотика сложности по $n \to \infty$. К сожалению, оказывается, что на этот вопрос нельзя дать однозначного ответа (потому, что сложность зависит от «всего» G, а не только от размерности G).

Действительно, если G — параллелепипед размерности n (т. е. $\alpha_{\infty,n}(G) = 1$), а $\nu < {}^1/_4$, то при $n \to \infty$

$$N(\nu) \sim n \ln \frac{1}{\nu}, \quad \tilde{N}(\nu) \gtrsim \frac{n}{\ln n} \ln \frac{1}{\nu},$$

т. е. сложность растет с ростом размерности линейно (или почти линейно). С другой стороны, если G — $V_p^{(n)}$-шар, т. е. $\alpha_{p,n}(G) = 1$

и $1 < p < \infty$, то при $n \to \infty$ и $\nu < {}^1/_4$

$$N(\nu) \underset{p}{\sim} \tilde{N}(\nu) \underset{p}{\sim} \nu^{-\max(2,\,p)}$$

(здесь $a \underset{x}{\sim} b$ означает, что b/a сверху и снизу ограничено величиной, зависящей лишь от x; аналогично понимается $\underset{x}{\gtrsim}$). Таким образом, в последнем случае сложность ограничена равномерно по n (величиной, зависящей лишь от ν и p). Если же G — $l_1^{(n)}$-шар, то при $n \to \infty$ и $\nu < {}^1/_4$

$$N(\nu) \sim \tilde{N}(\nu) \sim \frac{\ln n}{\nu^2};$$

наблюдается рост сложности с ростом n, но не линейный, а логарифмический. Заметим, что в двух последних случаях значение n, начиная с которого действуют указанные асимптотики сложностей, зависит от ν (и p). При этом для стохастической сложности асимптотика устанавливается, вообще говоря, позже, чем для детерминированной.

Ввиду сказанного без дополнительных предположений о G нельзя ответить на вопрос о том, как меняется сложность с увеличением размерности. Необходимо еще должным образом фиксировать «аффинный тип» G. Рассмотрим эту проблему, считая, что G имеет тип $l_p^{(n)}$-шара (более точно, мы считаем $\alpha_{p,n}(G) \leqslant \alpha$ и рассматриваем поведение сложности при изменении G, считая α параметром). Нижние и верхние границы сложности, которые при этом получаются, зависят от G лишь через размерность G. При $n \to \infty$ и $\alpha\nu < {}^1/_{32}$ имеем

$$\frac{n \ln 1/\nu}{\ln(n \ln 1/\nu)} \underset{\alpha}{\lesssim} \tilde{N}(\nu) \leqslant N(\nu) \underset{\alpha}{\sim} n \ln \frac{1}{\nu}, \quad p = \infty,$$

$$N(\nu) \underset{\alpha,p}{\sim} \tilde{N}(\nu) \underset{\alpha,p}{\sim} \begin{cases} 1/\nu^{\max(2,\,p)}, & 1 < p < \infty, \\ \ln n/\nu^2, & p = 1, \end{cases} \tag{1.20}$$

т. е. асимптотика сложностей по $n \to \infty$ — та же, что и для $l_p^{(n)}$-шаров (разумеется, расхождение верхней и нижней границ сложностей теперь зависит и от α; оно тем больше, чем больше α).

1.4.3. Из приведенных рассуждений ясно, что сложность довольно сильно зависит от такого малообозримого фактора, как «аффинный тип G», так что в общем случае трудно рассчитывать на возможность получения конструктивных и исчерпывающих результатов. Впрочем, для некоторого достаточно широкого спектра конкретных тел G — для $l_p^{(n)}$-шаров — можно, используя оценки таблицы, получить «почти полное» представление о сложности. Приведем соответствующие результаты.

С л у ч а й $p = \infty$. Здесь картина наиболее полна: при всех ν ($\nu < {}^1/_{32}$) выполняется соотношение

$$N(\nu) \sim n \ln \frac{1}{\nu} \quad \text{и} \quad \tilde{N}(\nu) \gtrsim \frac{N(\nu)}{\ln N(\nu)}.$$

С л у ч а й $1 < p < \infty$. Здесь (как и в случае $p = 1$) результаты несколько менее точны. Состоят они в следующем (считаем $\nu < 1$):

— при высоких точностях — именно, при $\nu \leqslant n^{-2/p}$ — справедливо соотношение

$$N(\nu) \underset{p}{\sim} n \ln \frac{1}{\nu};$$

— при низких точностях — именно, при $\nu \geqslant n^{-1/\max(p,2)}$ — справедливо соотношение

$$N(\nu) \underset{p}{\sim} \frac{1}{\nu^{\max(2,\,p)}};$$

— при промежуточных точностях — именно, при $n^{-1/\max(2,p)} > \nu > n^{-2/p}$ — справедливы двусторонние оценки

$$n \underset{p}{\lesssim} N(\nu) \lesssim n \ln \frac{1}{\nu}.$$

Таким образом, при увеличении точности сложность вначале растет как $1/\nu^{\max(2,p)}$ и «не чувствует» значения n. Этот рост продолжается до тех пор, пока сложность не дойдет до уровня n. Здесь впервые начинает «ощущаться» размерность. После некоторой полосы неопределенности значений ν (неопределенности наших знаний, разумеется), в которой сложность заключена между уровнями $\underset{p}{\sim} n$ и $\underset{p}{\sim} n \ln n$, сложность начинает меняться как $n \ln 1/\nu$ и «забывает» про геометрию G (т. е., собственно, про соответствующее p).

Заметим, что относительно стохастической сложности при всех $\nu < 1$ справедливы оценки

$$\frac{N(\nu)}{\ln N(\nu) + 1} \underset{p}{\lesssim} \tilde{N}(\nu) \leqslant N(\nu), \tag{1.21}$$

т. е. все сказанное о $N(\nu)$ с некоторой поправкой верно и для $\tilde{N}(\nu)$. Заметим еще, что мы имеем оценку $N(\nu)$, точную с точностью до множителя $\underset{p}{\sim} \ln N(\nu)$.

С л у ч а й $p = 1$. Здесь обстановка качественно та же, что и в случае $p > 1$. Именно ($\nu < {}^1/_{32}$):

— при высоких точностях ($\nu \leqslant n^{-2}$) имеем

$$N(\nu) \sim n \ln \frac{1}{\nu};$$

— при очень низких точностях ($\nu \geqslant n^{-1/8}$)

$$N(\nu) \sim \frac{\ln(n+1)}{\nu^2};$$

— при промежуточных точностях ($n^{-1/8} > \nu > n^{-2}$) дело обстоит так: вначале (при $\nu > n^{-1/2}$) верны оценки

$$\frac{1}{\nu^2} \lesssim N(\nu) \lesssim \frac{\ln(n+1)}{\nu^2},$$

а затем (при $n^{-2} \leqslant \nu < n^{-1/2}$) оценки

$$n \lesssim N(\nu) \lesssim n \ln n.$$

Таким образом, отличие от случая $1 < p < \infty$ состоит в том, что размерность «чувствуется» с самого начала, с низких точностей, но слабо. «Полоса неопределенности» относительно шире, чем при $p > 1$, но и в ней (а тогда и везде) мы знаем $N(\nu)$ с точностью до множителя, $\sim \ln N(\nu)$. Наконец, при всех $\nu < 1/32$ справедлива оценка (1.21).

1.5. Обсудим бесконечномерную ситуацию. Здесь сложность классов общих выпуклых задач может быть как конечной при каждом $\nu > 0$, так и бесконечной для всех достаточно малых $\nu > 0$. Во всяком случае, если G — выпуклое ограниченное замкнутое тело в E и $\dim E = \infty$, то в силу A.3

$$N(\nu), \tilde{N}(\nu) \gtrsim \left(\frac{1}{\alpha(G)\nu}\right)^2,$$

т. е. сложность на любом бесконечномерном пространстве допускает ту же нижнюю оценку, что и в гильбертовом случае. Эта нижняя оценка заведомо указывает правильную (по характеру зависимости от ν) асимптотику сложности по $\nu \to +0$, если E регулярно и некоторая отвечающая E регулярная функция V такова, что $\omega_V(t) \leqslant O(t)$, как это имеет место в гильбертовом случае.

Далее, если E регулярно, то $N(\nu)$ конечно для всех $\nu > 0$. Наконец, в случае $E = L_p(T, \mu)$, $1 < p < \infty$, $\dim E = \infty$ можно получить более законченные результаты. Именно, в силу (1.12), (1.14) и (1.19) имеем

$$N(\nu) \underset{\alpha_{\|\cdot\|_p}(G),\, p}{\sim} \tilde{N}(\nu) \underset{\alpha_{\|\cdot\|_p}(G),\, p}{\sim} \frac{1}{\nu^{\max(2,\, p)}}.$$

§ 2. Сложность классов липшицевых выпуклых задач

В этом параграфе применительно к классам липшицевых выпуклых задач проводятся те же рассмотрения, что и проделанные выше для классов общих выпуклых задач. Результаты в основном близки к результатам § 1. Как и там, E — вещественное банахово пространство, G — ограниченное выпуклое замкнутое тело в E, $\|\cdot\|$ — норма на E. Далее $0 < \nu < 1$.

2.1. Начнем с формулировки основных результатов о сложности рассматриваемых классов задач.

Пусть C_{Lip} — какой-нибудь класс задач типа $C^{\mathrm{J}}_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$. Обозначим через $N(\nu)$ ($N_*(\nu)$) детерминированную сложностную (соответственно $*$-сложностную, см. п. 4.4 гл. I) функцию этого класса. Пусть, далее, $\mathcal{O}$ — произвольный локальный детерминированный оракул для поля задач $C_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$ и $N^{\mathcal{O}}(\nu)$, $\tilde{N}^{\mathcal{O}}(\nu)$ — соответственно детерминированная и стохастическая сложность класса задач $C^{\mathcal{O}}$, получающегося из C_{Lip} заменой исходного оракула на оракул $\mathcal{O}$.

Теорема об оценке сложности классов липшицевых выпуклых задач. *Существует функция $d(p) > 0$, $1 \leqslant p \leqslant \infty$, такая, что для соответствующих ей (в силу (1.11)) функций $\varphi_{n,p}(\nu)$ справедливы следующие утверждения. Функции $N(\nu)$, $N_*(\nu)$, $N^{\mathcal{O}}(\nu)$, $\tilde{N}^{\mathcal{O}}(\nu)$ допускают следующие оценки *).*

A. Оценки снизу.

A.1. При всяком натуральном $k \leqslant n$ и всяком p, $1 \leqslant p \leqslant \infty$, имеем

$$N^{\mathcal{O}}(\nu) \geqslant \varphi_{k,p}(\alpha_{p,\|\cdot\|,k}(G)\,\nu) \tag{2.1}$$

и

$$\tilde{N}^{\mathcal{O}}(\nu) \geqslant \frac{\varphi_{k,p}(8\alpha_{p,\|\cdot\|,k}(G)\,\nu)}{1 + \left\lfloor \ln\,]\varphi_{k,p}(8\alpha_{p,\|\cdot\|,k}(G)\,\nu)[\right\rfloor_+}. \tag{2.2}$$

A.2. Пусть $E = L_p(T, \mu)$, $\|\cdot\| = \|\cdot\|_p$, $1 < p < \infty$. Существует такое $k_p(t) < \infty$, $t > 0$, что при $n \geqslant k_p(\alpha_{\|\cdot\|_p}(G)\,\nu)$ имеем

$$\tilde{N}^{\mathcal{O}}(\nu) \geqslant \varphi_{n,p}(28\alpha_{\|\cdot\|_p}(G)\,\nu). \tag{2.3}$$

A.3. Существует такое $\bar{k}_2(t) < \infty$ при $t > 0$, что при $n > \bar{k}_2(\alpha_{\|\cdot\|}(G)\,\nu)$ имеем

$$\tilde{N}^{\mathcal{O}}(\nu) \geqslant \varphi_{\infty,2}(\alpha_{\|\cdot\|}(G)\,\nu). \tag{2.4}$$

A.4. Если E нерефлексивно, то

$$N^{\mathcal{O}}(\nu) = +\infty \ \text{при}\ \nu \leqslant \nu(G),\ \text{где}\ \nu(G) > 0. \tag{2.5}$$

A.5. Если $m > 0$, то

$$N_*(\nu) = +\infty \quad \text{при}\ \nu < 1. \tag{2.6}$$

B. Оценки сверху.

B.1. При $n < \infty$

$$N(\nu) \leqslant \min\left\{\Phi_{n,\infty}(\nu),\ \inf_{1 \leqslant p < \infty} \Phi_{n,p}\left(\frac{\nu}{\alpha_{p,n}(\|\cdot\|)}\right)\right\}. \tag{2.7}$$

*) См. сноску на стр. 136.

B.2. При $E = L_p(T, \mu)$, $1 < p < \infty$,

$$N(\nu) \leqslant \Phi_{\infty, p}(\nu). \tag{2.8}$$

Оценки сверху верны и безотносительно к тому, что G — тело.

Отметим отличие результатов этой теоремы от утверждений теоремы 1.2. В нижних оценках единственное расхождение состоит в том, что все аффинные инварианты G — величины $\alpha_{p,k}(G)$ из 1.2 — заменены в 2.1 на метрические инварианты $\alpha_{p,\|\cdot\|,k}(G)$. В верхних оценках зависящие от G характеристики $4\alpha_{p,n}(G)$, $4\alpha_{\|\cdot\|}(G)$ заменены соответственно на характеристику нормы $\|\cdot\|\,\alpha_{p,n}(\|\cdot\|)$ и на 1. И нижние, и верхние оценки теперь стали меньшими. Разумеется, так и должно быть, ибо класс задач C_{Lip} уже, чем C. Отметим, что обоснование верхних оценок B проводится так же, как и в п. 1.2, только ссылаться следует на результаты о трудоемкости МЦТ и ЗС_p не на классе общих, а на классе липшицевых выпуклых задач — именно, на п. 4.1.2 гл. II и § 3 гл. III. Доказательство оценок A проводится в §§ 4—6.

2.2. Как и в § 1, приведем развернутую сводку определяемых теоремой оценок сложности. В таблице 2 предполагается $n < \infty$. По сравнению с таблицей 1 в новой таблице имеется еще одна графа, фиксирующая гипотезу о строении рассматриваемой нормы $\|\cdot\|$ на E — нормы, в терминах которой определяется интересующий нас класс. В остальном строение и способ обоснования таблицы те же, что и в § 1.

2.3. Комментарии к таблице в основном совпадают с данными в § 1. Мы коротко изложим их, останавливаясь, главным образом, на особенностях липшицева случая по сравнению с общим.

2.3.1. Асимптотика по $\nu \to +0$. Как и в § 1, при $\nu < \nu_{\|\cdot\|}(G) > 0$ имеем $N(\nu) \sim n \ln 1/\nu$ и $\tilde{N}(\nu) \gtrsim N(\nu)/\ln N(\nu)$. Особенность липшицева случая состоит в том, что $\nu_{\|\cdot\|}(G)$ зависит как от аффинных свойств $\|\cdot\|$, так и от метрических свойств G относительно $\|\cdot\|$. В частности, $\nu_{\|\cdot\|}(G)$ не допускают положительной миноранты, зависящей лишь от n, как это было для величин $\nu(G)$ из 1.4.1. Именно, $\nu_{\|\cdot\|}(G)$, вообще говоря, тем меньше, чем больше $\alpha_{\|\cdot\|}(G)$ (можно взять $\nu_{\|\cdot\|}(G) = \min\{{}^1/_{32};\ (\alpha_{\|\cdot\|}(G)\,n)^{-2}\}$). Это и естественно. Пусть, скажем, G представляет собой δ-окрестность (в метрике, задаваемой $\|\cdot\|$) отрезка единичной длины. При малых δ решение задач класса типа $C^0_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$ с погрешностью $\nu \gg \delta$ можно свести к решению одномерных задач. Сложность при этих ν будет вести себя как $\ln 1/\nu$. Чем меньше δ, тем в большем диапазоне значений ν действует эта «патологическая» для задач большой размерности оценка сложности.

Ясно, почему такого эффекта не возникает при рассмотрении общих выпуклых задач. Функционалы таких задач могут меняться как угодно быстро (ограничено только их изменение на G). По этой причине в предыдущем примере нет никаких оснований редуцировать задачи к одномерным, сколько бы ни было мало δ. То же самое можно сказать и по-другому: класс общих выпуклых задач определен в аффинных терминах, и поэтому бессмысленно изучать, как на его сложность влияет «узость» G в каких-то направлениях — это не есть аффинный инвариант G. В аффинном смысле тело G не может быть «слишком узким» (скажем, $\alpha_{\infty,n}(G) \leqslant 2n$), что и определяет наличие «абсолютно высоких точностей» для классов общих задач в данной размерности — точностей, начиная с которых сложность этих классов ведет себя как $n \ln 1/\nu$.

2.3.2. Что касается до зависимости сложности от размерности G, то отличие от результатов п. 1.4.2 состоит в том, что теперь надо фиксировать не $\alpha_{p,n}(G)$, а $\alpha_{p,\|\cdot\|,n}(G)$, т. е. рассматривать поведение $N(\nu)$, $\tilde{N}(\nu)$ в функции от n при ограничении $\alpha_{p,\|\cdot\|,n}(G) \leqslant \alpha$. При таком условии соотношения (1.20) сохраняются и в липшицевом случае.

2.3.3. Что же касается оценки сложности на $l_p^{(n)}$-шарах, то, предполагая, что G есть $l_p^{(n)}$-шар, а $\|\cdot\|$ есть в точности $\|\cdot\|_p$, можно дословно повторить все сказанное в п. 1.4.3.

2.4. Применительно к бесконечномерной ситуации и в липшицевом случае сохраняет силу все сказанное в п. 1.5, с той, однако, оговоркой, что вместо $\alpha(G)$ там должна фигурировать $\alpha_{\|\cdot\|}(G)$.

§ 3. Рекомендации по применению методов решения общих (липшицевых) выпуклых задач

3.1. В этом параграфе данные о сложности, содержащиеся в двух предыдущих параграфах, трансформируются в рекомендации по применению МЦТ, ЗС_p. Основание для этих рекомендаций — совпадение (с точностью до теоретически малосущественных деталей) оценок трудоемкости указанных методов с оценками сложности рассматриваемого класса задач.

Данные по применению $\overline{\text{ЗС}}_p$ приводятся в предположении, что норма $\|\cdot\|_p$, с которой ассоциируется метод, оптимальна для рассматриваемого G в смысле п. 5 § 4 гл. III, т. е. для нее $\alpha_{\|\cdot\|_p}(G) = \alpha_{p,n}(G)$. Рекомендации по применению методов оформлены в виде таблиц. Первая графа фиксирует гипотезу о свойствах G. Во второй указан рекомендуемый метод, в третьей — условия, в которых он рекомендуется. Далее указывается (верхняя) оценка трудоемкости рекомендованного метода $M(\nu)$ и потенциальные границы ее снижения в классе детерминированных и стохастических методов (т. е. верхние оценки отношений $M(\nu)/N(\nu)$, $M(\nu)/\tilde{N}(\nu)$; $N(\nu)$, $\tilde{N}(\nu)$ — соответствующие сложностные функции рассматриваемого класса). Отметим, что рекомендации по применению МЦТ носят чисто теоретический

Таблица 2

$\|\cdot\|$	G	$N(\nu)\leqslant$	$N(\nu)\geqslant$	$\tilde{N}^G(\nu)\geqslant$	Асимптотика $N(\nu)$	
					по $\nu\to 0$	по $n\to\infty$
—	—	$\Phi_\infty(\nu)\equiv c_\infty n\,]\ln(1/\nu)[$	—	—	—	—
$\alpha_{p,n}(\|\cdot\|)\leqslant\alpha$, $1<p<\infty$, $s=\max(2,\ p)$, $\bar\alpha=\min\{2n, \alpha n^{1/p}\}$	$\alpha_{\|\cdot\|}(G)\leqslant\beta$	$\Phi_\infty(\nu)\equiv c_\infty n\,]\ln(1/\nu)[$	$\lambda_\infty\Phi_\infty(\nu)$ при $\beta\bar\alpha\nu<{}^1/_4$ и $\nu\leqslant(\beta\bar\alpha)^{-2}$	$\lambda_\infty\dfrac{\Phi_\infty(\nu)}{\ln\Phi_\infty(\nu)}$ при $\beta\bar\alpha\nu<{}^1/_{32}$ и $\nu\leqslant(\beta\bar\alpha)^{-2}$	$(n/\mu_\infty)\ln(1/\nu)\leqslant$ $\leqslant N(\nu)\leqslant$ $\leqslant\mu_\infty n\ln(1/\nu)$ при $\beta\bar\alpha\nu<{}^1/_4$ и $\nu\leqslant(\beta\bar\alpha)^{-2}$	$\lambda_p(\beta\alpha)^{-s}\,\Phi_p(\nu)\leqslant$ $\leqslant N(\nu)\leqslant$ $\leqslant\alpha^s\Phi_p(\nu)$ при $\alpha\beta\nu\geqslant n^{-1/s}$
		$\alpha^s\Phi_p(\nu)\equiv c_p\alpha^s/\nu^s$	$\lambda_p\Phi_p(\nu)\,(\beta\alpha)^{-s}$ при $\alpha\beta\nu\geqslant n^{-1/s}$	$\lambda_p(\beta\alpha)^{-s}\dfrac{\Phi_p(\nu)}{[\ln\Phi_p(\beta\alpha\nu)]_++1}$ при $\beta\alpha\nu\geqslant n^{-1/s}$		
				$\lambda^p(\beta\alpha)^{-s}\,\Phi_p(\nu)$ при $n\geqslant k_p(\beta\alpha\nu)$		
$\alpha_{1,n}(\|\cdot\|)\leqslant\alpha$	$\alpha_{\|\cdot\|}(G)\leqslant\beta$	$\Phi_\infty(\nu)\equiv c_\infty n\,]\ln(1/\nu)[$	$\lambda_\infty\Phi_\infty(\nu)$ при $\nu\leqslant(n\beta)^{-2}$	$\lambda_\infty\Phi_\infty(\nu)/\ln\Phi_\infty(\nu)$ при $\nu\leqslant(n\beta)^{-2}$	$(n/\mu_\infty)\ln(1/\nu)\leqslant$ $\leqslant N(\nu)\leqslant$ $\leqslant\mu_\infty n\ln 1/\nu$ при $\nu\leqslant(n\beta)^{-2}$	$\lambda_1\ln n/(\alpha\beta\nu)^2\leqslant$ $\leqslant N(\nu)\leqslant$ $\leqslant c_1\ln n\cdot\alpha^2/\nu^2$ при $\alpha\beta\nu<{}^1/_4$ и $\alpha\beta\nu\geqslant n^{-1/8}$
		$\alpha^2\Phi_1^{(n)}(\nu)\equiv c_1\ln n\alpha^2/\nu^2$	$\lambda_1\Phi_1^{(n)}(\alpha\beta\nu)$ при $\alpha\beta\nu<{}^1/_4$ и $\alpha\beta\nu\geqslant n^{-1/8}$	$\dfrac{\lambda_1\Phi_1^{(n)}(\alpha\beta\nu)}{[\ln\Phi_1^{(n)}(\alpha\beta\nu)]_++1}$ при $\alpha\beta\nu<{}^1/_{32}$ и $\alpha\beta\nu\geqslant n^{-1/8}$		
			$\lambda_1/(\alpha\beta\nu)^2$ при $\alpha\beta\nu<{}^1/_4$ и $\alpha\beta\nu\geqslant n^{-1/2}$	$\lambda_1/(\alpha\beta\nu)^2\ln 1/\alpha\beta\nu$ при $\alpha\beta\nu<{}^1/_4$ и $\alpha\beta\nu\geqslant n^{-1/2}$		
				$\lambda_1/(\alpha\beta\nu)^2$ при $n>\bar k_2(\alpha\beta\nu)$		
$\alpha_{\infty,n}(\|\cdot\|)\leqslant\alpha$	$\alpha_{\|\cdot\|}(G)\leqslant\beta$	$\Phi_\infty(\nu)\equiv c_\infty n\,]\ln(1/\nu)[$	$\lambda_\infty\Phi_\infty(\nu)$ при $\alpha\beta\nu<{}^1/_4$ и $\nu\leqslant(\beta\alpha)^{-2}$	$\lambda_\infty\Phi_\infty(\nu)/\ln\Phi_\infty(\nu)$ при $\alpha\beta\nu<{}^1/_{32}$ и $\nu\leqslant(\beta\alpha)^{-2}$	$(n/\mu_\infty)\ln(1/\nu)\leqslant$ $\leqslant N(\nu)\leqslant$ $\leqslant\mu_\infty n\ln(1/\nu)$ при $\alpha\beta\nu<{}^1/_4$ и $\nu\leqslant(\beta\alpha)^{-2}$	$(n/\mu_\infty)\ln(1/\nu)\leqslant$ $\leqslant N(\nu)\leqslant$ $\leqslant\mu_\infty n\ln(1/\nu)$ при $\alpha\beta\nu<{}^1/_4$ и $\nu\leqslant(\beta\alpha)^{-2}$

характер, так как неясно, как реализовать этот метод для сколько-нибудь заметных размерностей. Впрочем, применение ММЦТ вместо МЦТ приводит к очевидным изменениям в таблице.

3.2. Начнем с классов общих выпуклых задач. Мы рассматриваем класс задач C типа $C^0(G, R^n, m)$. Конечно, в теоретическом плане следовало бы рассмотреть и бесконечномерный случай, но здесь достаточно отослать читателя к п. 1.5, где по существу уже сформулированы рекомендации о применении ЗС-методов и (для $E = L_p(T, \mu)$) утверждения об их субоптимальности.

Рекомендации по применению методов решения общих выпуклых задач приведены в таблице 3.

3.3. Прокомментируем данные таблицы 3. Прежде всего, если заданная точность *достаточно высока*, $(\nu \leqslant \nu(G),\ \nu(G) \geqslant n^{-2})$, то теоретически субоптимален МЦТ. Его трудоемкость ($\sim n \ln (1/\nu)$) в классе детерминированных методов в принципе не снижается более чем в ~ 1 раз, а в классе стохастических методов — не более чем в $\sim \ln(n \ln(1/\nu))$ раз.

Если же заданная точность ν не столь высока, то на вопрос о теоретически субоптимальном методе решения общих выпуклых задач мы не умеем давать однозначный ответ. Мы знаем лишь, что если G имеет тип $\mathfrak{l}_p^{(n)}$-шара ($\alpha_{p,n}(G) \equiv \alpha$ не слишком велико), то в известном диапазоне изменения ν (этот диапазон есть $\alpha\nu \geqslant n^{-1/s}$, $s = \max(2,p)$, если $1 < p < \infty$ *)) целесообразно применение $\overline{ЗС}_p$. При этом трудоемкость указанного метода ($\underset{p}{\sim}\alpha^s/\nu^s$) не может быть снижена более чем в $\underset{p}{\sim}\alpha^{2s}$ раз, в классе детерминированных, и в $\underset{p}{\sim}\alpha^{2s}\ln\alpha^s/\nu^s$ раз — в классе стохастических. В диапазоне промежуточных точностей (для $1 < p < \infty$ этот диапазон есть $\{\nu | \alpha\nu < n^{-1/s},\ \nu > \nu(G)\}$) — мы рекомендовали *применять* МЦТ, хотя теоретические основания для этого слабее, чем при высоких точностях: трудоемкость МЦТ здесь не снижается более, чем в $\underset{p}{\sim} \ln(n \ln(1/\nu))$ раз в классе детерминированных и в $\underset{p}{\sim} \ln^2(n \ln(1/\nu))$ раз — в классе стохастических методов.

Заметим, что при больших $\alpha_{p,n}(G)$ основания для высказываний в пользу $\overline{ЗС}_p$ не очень надежны, так как границы потенциально возможного улучшения метода в этом случае зависят от величины $\alpha_{p,n}(G)$. Это и естественно — области G, плохо аппроксимируемые $\mathfrak{l}_p^{(n)}$-шарами при всех p, «не улавливаются» *шкалой* L_p-методов и оценок. В свое оправдание заметим, что из приведенных результатов видно, как сильно сложность рассматривае-

*) В обсуждении мы ограничимся случаем $1 < p < \infty$. Случай $p = \infty$ прост — здесь все точности высоки: $\nu(G) = \min\{1/32, \alpha^{-2}\}$. Случай $p = 1$ качественно устроен так же, как и случай $1 < p < \infty$.

Таблица 3

Условия на G	Рекомендуемый метод	Условия, при которых метод рекомендуется	Оценка трудоемкости метода, $M(\nu)$	Потенциальная граница снижения трудоемкости в классе дет. методов, $\leqslant$	Потенциальная граница снижения трудоемкости в классе ранд. методов, $\leqslant$
—	МЦТ	$\nu \leqslant n^{-2}$	$3n]\ln(1/\nu)[$	λ_∞ при $\nu \leqslant n^{-2}$	$\lambda_\infty \ln M(\nu)$ при $\nu \leqslant n^{-2}$
$\alpha_{p,n}(G) \leqslant \alpha$, $1 < p < \infty$, $s = \max(2, p)$, $\bar\alpha = \min\{2n; \alpha n^{1/p}\}$	$\overline{ЗС}_p$	$\alpha\nu \geqslant n^{-1/s}$	$c_p \dfrac{\alpha^s}{\nu^s}$	$\lambda_p \alpha^{2s}$ при $\alpha\nu \geqslant n^{-1/s}$	$\lambda_p \alpha^{2s} \ln M(\nu)$ при $\alpha\nu \geqslant n^{-1/s}$
					$\lambda_p \alpha^{2s}$ при $n \geqslant k_p(\alpha\nu)$
	МЦТ	$\alpha\nu < n^{-1/s}$	$3n]\ln(1/\nu)[$	λ_∞ при $\bar\alpha\nu < 1/32$ и $\nu \leqslant \bar\alpha^{-2}$	$\lambda_\infty \ln M(\nu)$ при $\bar\alpha\nu < 1/32$ и $\nu < \bar\alpha^{-2}$
				$\lambda_p \ln M(\nu)$ при $\alpha\nu < n^{-1/s}$	$\lambda_p \ln^2 M(\nu)$ при $\alpha\nu < n^{-1/s}$
$\alpha_{1,n}(G) \leqslant \alpha$	$\overline{ЗС}_{1,n}$	$\alpha\nu \geqslant n^{-1/2}$	$\dfrac{c_1\alpha^2 \ln(n+1)}{\nu^2}$	$\lambda_1\alpha^4$ при $1/4 > \alpha\nu \geqslant n^{-1/8}$	$\lambda_1\alpha^4 \ln M(\nu)$ при $1/32 > \alpha\nu \geqslant n^{-1/8}$
				$\lambda_1\alpha^4 \ln(n+1)$ при $\alpha\nu \geqslant n^{-1/2}$	$\lambda_1\alpha^4 \ln^2 M(\nu)$ при $\alpha\nu \geqslant n^{-1/2}$
					$\lambda_1\alpha^4 \ln M(\nu)$ при $n > k_2(\alpha\nu)$
	МЦТ	$\alpha\nu < n^{-1/2}$	$3n]\ln(1/\nu)[$	λ_∞ при $\nu \leqslant n^{-2}$	$\lambda_\infty \ln M(\nu)$ при $\nu \leqslant n^{-2}$
				$\lambda_1 \ln M(\nu)$ при $\alpha\nu < n^{-1/2}$	$\lambda_1 \ln^2 M(\nu)$ при $\alpha\nu < n^{-1/2}$
$\alpha_{\infty,n}(G) \leqslant \alpha$	МЦТ		$3n]\ln(1/\nu)[$	λ_∞ при $\nu \leqslant \alpha^{-2}$ и $\alpha\nu < 1/4$	$\lambda_\infty \ln M(\nu)$ при $\nu \leqslant \alpha^{-2}$ и $\alpha\nu < 1/32$

мого класса зависит от аффинных свойств G. Вместе с тем столь расплывчатое понятие, как «аффинные свойства G», трудно учесть и конструктивно, и достаточно полно.

Заметим еще, что приведенные рекомендации по применению $\overline{\text{ЗС}}_p$ даже на $l_p^{(n)}$-шарах «обоснованы с точностью до множителя $\lambda(p)$», который, вообще говоря, может быть велик. Для теоретического анализа этого, возможно, достаточно, но на практике требуется более подробный анализ (это замечание относится и к последующим обсуждениям). Приведем пример. Пусть требуется решать общие выпуклые задачи на $l_p^{(n)}$-шаре. Предполагается, что точность решения ν не слишком велика, n достаточно велико и p близко к 1. Оценка трудоемкости $\overline{\text{ЗС}}_p$ в этом случае $\sim 1/(p-1)\nu^2$. Между тем в силу результатов упражнения 1 § 3 гл. III при $1 \leqslant p \leqslant 2$ указанные задачи можно решать с трудоемкостью $\sim \ln n/\nu^2$. Ясно, что если $1/(p-1) \gg \ln n$, то второй метод фактически предпочтительнее первого.

3.4. Сделаем по поводу методов $\overline{\text{ЗС}}_p$, $1 < p < \infty$, еще одно замечание. Их субоптимальность на телах $G \subset l_p^{(n)}$ асферичности α была установлена (при больших n) с точностью до множителя вида $f(p, \alpha)$. Иными словами, они неулучшаемы лишь по порядку зависимости от ν, если n достаточно велико. Между тем трудоемкость $\overline{\text{ЗС}}_p$ сильно зависит от α. Естественно спросить, можно ли снизить «чувствительность методов к α». Оказывается, что, вообще говоря, этого сделать нельзя. Именно, пусть G — выпуклое замкнутое ограниченное тело в R^n p-асферичности, $\leqslant \alpha$, а $\boldsymbol{C}$ — класс типа $\boldsymbol{C}^0(G, R^m, m)$, отвечающий локальному детерминированному оракулу. Обозначим через $N(\nu)$ детерминированную сложностную функцию этого класса, и пусть $N_p(\nu, \alpha)$ есть верхняя грань функций $N(\nu)$ по всевозможным классам $\boldsymbol{C}$ описанного выше типа (т. е. классам $\boldsymbol{C}$, отвечающим всевозможным n и $G \subset R^n$, $\alpha_{p,n}(G) \leqslant \alpha$). Из свойств $\overline{\text{ЗС}}_p$ следует, что

$$N_p(\nu, \alpha) \underset{p}{\lesssim} \left(\frac{\alpha}{\nu}\right)^{\max(2,\, p)}.$$

Оказывается, что верна и нижняя оценка $N_p(\nu, \alpha)$ такого же типа:

$$N_p(\nu, \alpha) \underset{p}{\gtrsim} \left(\frac{\alpha}{\nu}\right)^{\max(2,\, p)}, \quad \nu < \frac{1}{4}. \tag{3.1}$$

Таким образом, снижение трудоемкости $\overline{\text{ЗС}}_p$ *сразу на всех телах* p-асферичности, $\leqslant \alpha$, возможно не более чем в $\underset{p}{\sim} 1$ раз. Конечно, на *некоторых* телах такого рода возможно и более сильное снижение трудоемкости $\overline{\text{ЗС}}_p$.

Таким образом, $\overline{\text{ЗС}}_p$-методы, в некотором смысле, реализуют максимум возможного в L_p-шкале характеризации выпуклых тел.

Упражнение 1. Докажите (3.1).

⟨Пусть вначале $p \geqslant 2$. Для $\alpha = 1$ (3.1) сразу следует из (1.12), так что можно ограничиться случаем $\alpha > 8$. Пусть $\bar\alpha = \alpha/8$. Найдем p' из уравнения $(1/4\nu)^{p'} = (\bar\alpha/\nu)^p$. Ясно, что это уравнение имеет корень $p' > p$. Пусть $n =]\,(\bar\alpha/\nu)^p\,[$ и G — $l_{p'}^{(n)}$-шар. Проверим, что $\alpha_{p,n}(G) \leqslant \alpha$. Действительно, в силу (1.3) имеем

$$\alpha_{p,n}(G) \leqslant n^{\frac{1}{p} - \frac{1}{p'}} \leqslant 2^{\frac{1}{p} - \frac{1}{p'}} \left(\frac{1}{4\nu}\right)^{\frac{p'-p}{p}} = 2\,(4^{p'} \cdot \bar\alpha^p)^{\frac{1}{p}} \cdot 4^{-\frac{p'-p}{p}} =$$
$$= 2 \cdot 4 \cdot 4^{-\frac{p'}{p}} \cdot 4^{\frac{p'}{p}}\, \bar\alpha = 8\bar\alpha = \alpha.$$

С другой стороны, ввиду утверждения А.1 теоремы 1.2 и (1.11), сложность решения задач класса $\boldsymbol{C}$ типа $\boldsymbol{C}^0(G, R^n, m)$ не меньше

$$d(\infty) \min\left(n, \frac{4^{-p'}}{\nu^{p'}}\right) = d(\infty) \min\left\{\Big]\left(\frac{1}{4\nu}\right)^{p'}\Big[, \left(\frac{1}{4\nu}\right)^{p'}\right\} = d(\infty)\,\frac{\bar\alpha^p}{\nu^p},$$

что и требуется в (3.1).

Пусть теперь $1 < p < 2$. Опять-таки достаточно рассмотреть случай $\alpha \geqslant 16$. Пусть $G_0^{(n)}$ есть единичный шар пространства $l_p^{(n)}$. Как было указано в (1.8), для данного k существует натуральное $n(k)$, такое, что $G_0^{(n)}$ допускает k-мерное сечение $G_0^{n,k}$ (плоскостью $E^{n,k}$, проходящей через центр $G_0^{(n)}$), асферичность которого относительно подходящей евклидовой нормы $|\cdot|$ на $E^{n,k}$ не выше 2. Можно считать, что на $E^{n,k}$ справедливо неравенство $2|x| \geqslant \|x\|_p \geqslant |x|$.

Пусть теперь $\bar\alpha = \alpha/16$ и p' выбрано так, что $(1/4\nu)^{p'} = (\bar\alpha/\nu)^2$. Тогда $p' > 2$. Положим $k =]\,(\bar\alpha/\nu)^2\,[$ и введем на $E^{n,k}$ $(n > n(k))$ норму $\|\cdot\|_{p'}$ так, чтобы единичный шар $E^{n,k}$ по этой норме — обозначим его $G_1^{n,k}$ — содержал $|\cdot|$-шар радиуса 1 с центром в 0 и содержался в $|\cdot|$-шаре радиуса $k^{\frac{1}{2} - \frac{1}{p'}}$ с тем же центром. Очевидно, $G_1^{n,k} \supset G_0^{n,k}$. Обозначим через $G^{(n)}$ выпуклую оболочку $G_0^{(n)}$ и $G_1^{n,k}$. Очевидно, $G^{(n)}$ есть центрально симметричное выпуклое тело $\|\cdot\|_p$-асферичности,

$$\leqslant 2k^{\frac{1}{2} - \frac{1}{p'}} \leqslant 2\left(2\,\frac{\bar\alpha^2}{\nu^2}\right)^{\frac{1}{2} - \frac{1}{p'}} \leqslant 4\left(\frac{1}{4\nu}\right)^{p'\left(\frac{1}{2} - \frac{1}{p'}\right)} = 4 \cdot 4^{\frac{p'}{2}}\,\bar\alpha \cdot 4^{-\frac{p'}{2}+1} =$$
$$= 16\bar\alpha = \alpha.$$

Итак, $\alpha_{p',n}(G^{(n)}) \leqslant \alpha$. С другой стороны, $G^{(n)}$ есть единичный шар R^n по норме $\|\cdot\|$, для которой $\alpha_{p',k}(\|\cdot\|) = 1$ (ибо сечение $G^{(n)}$ плоскостью $E^{n,k}$ есть в точности $G_1^{n,k}$). Стало быть, $\alpha_{p',k}(G^{(n)}) = 1$ и, ввиду утверждения А.1 теоремы 1.2, сложность $N(\nu)$ класса типа $\boldsymbol{C}^0(G^{(n)}, R^n, m)$ допускает оценку

$$N_p(\alpha, \nu) \geqslant N(\nu) \geqslant$$
$$\geqslant d(\infty) \min\left\{k, \frac{1}{(4\nu)^p}\right\} = d(\infty) \min\left\{\Big]\left(\frac{\bar\alpha}{\nu}\right)^2\Big[, \left(\frac{1}{4\nu}\right)^{p'}\right\} \geqslant d(\infty)\left(\frac{\bar\alpha}{\nu}\right)^2,$$

что и требуется в (3.1).⟩

Упражнение 2. Пусть G — $\mathrm{I}_p^{(n)}$-шар, $1 \leqslant p < \infty$. Существует ли функция $\varphi(n)$, $\varphi(n)/n \underset{n\to\infty}{\longrightarrow} 0$, такая, что для сложности $N(\nu)$ классов типа $C^0(G, R^n, m)$ справедлива оценка $N(\nu) \leqslant f(\nu)\varphi(n)$ (равномерно по p)? Тот же вопрос в диапазоне $1 \leqslant p \leqslant 2$.

3.5. Теперь приведем таблицу 4 рекомендаций по применению методов решения липшицевых выпуклых задач (классов типа $C^0_{\mathrm{Lip}}(G, R^n, \|\cdot\|, m)$; по поводу бесконечномерной ситуации см. п. 2.5). В таблице предположено, что применение ЗС$_p$-методов к решению задач рассматриваемых классов проводится так, как указано в п. 5 § 4 гл. III.

Комментарии к этой таблице с качественной точки зрения близки к приведенным выше для классов общих выпуклых задач. Предоставляем читателю найти и объяснить различия в результатах, связанные с сужением класса задач по сравнению с общими.

3.6. Ознакомившись с данными выше рекомендациями по применению методов решения общих выпуклых и липшицевых выпуклых задач, читатель вправе поинтересоваться тем, почему мы ничего не говорим об условиях применимости известных традиционных методов выпуклого программирования, делая исключение лишь для градиентного метода. Дело здесь в том, что стандартных методов, пригодных для решения произвольных негладких выпуклых задач, совсем *не так много*; как правило, сходимость традиционных методов — во всяком случае «конструктивная», с оценками скорости — устанавливается в предположении гладкости компонент задачи. Насколько нам известно, в негладком случае, помимо подробно рассмотренного выше градиентного метода, «стандартным» можно считать еще только один метод — метод Келли. Как мы сейчас убедимся, он недопустимо плох с точки зрения характеризации «трудоемкость—погрешность».

Опишем простейший вариант этого метода, применимый к решению задач без ограничений — скажем, к задачам класса типа $C^0_{\mathrm{Lip}}(G, R^n, 0)$. Метод работает следующим образом. Пусть решается задача минимизации выпуклой липшицевой функции f в выпуклой области $G \subset R^n$, и пусть мы уже задавали вопросы о f в точках $x_1, \ldots, x_k$ и получили значения $f(x_i)$ и опорные функционалы $f'(x_i)$, $i = 1, 2, \ldots, k$, функции f в этих точках. Рассмотрим функцию

$$f^k(x) = \max_{1\leqslant i\leqslant k} \{f(x_i) + \langle f'(x_i) \mid x - x_i\rangle\};$$

очевидно, она удовлетворяет условию $f^k(x) \leqslant f(x)$ и совпадает с f в «опрошенных» точках. В методе Келли в качестве очередной точки x_{k+1} выбирается точка минимума f^k на G. Собственно говоря, это описание еще не задает метод — надо указать, как выбирать x_{k+1} в случае, когда точка минимума f^k неединственна.

Таблица 4

Условия на $\|\cdot\|$, G	Рекомендуемый метод	Условия, при которых метод рекомендуется	Оценка трудоемкости метода, $M(\nu)$	Потенциальная граница снижения трудоемкости в классе детерм. методов, $\leqslant$	Потенциальная граница снижения трудоемкости в классе рандом. методов, $\leqslant$
$\alpha_{p,n}(\|\cdot\|) \leqslant \alpha$, $\alpha_{\|\cdot\|}(G) \leqslant \beta$, $1 < p < \infty$, $s = \max(2, p)$, $\bar{\alpha} = \min\{2n, \alpha n^{1/p}\}$	ЗС$_p$	$\nu\alpha\beta \geqslant n^{-1/s}$	$c_p\left(\frac{\alpha}{\nu}\right)^s$	$\lambda_n\alpha^{2s}\beta^s$ при $\alpha\beta\nu \geqslant n^{-1/s}$	$\lambda_p\alpha^{2s}\beta^s \ln M(\nu)$ при $\alpha\beta\nu \geqslant n^{-1/s}$
					$\lambda_p\alpha^{2s}\beta^s$ при $n > k_p(\alpha\beta\nu)$
	МЦТ	$\nu\alpha\beta < n^{-1/s}$	$c_\infty n]\ln(1/\nu)[$	$\lambda_p \ln(\beta n)$ при $\alpha\beta\nu < n^{-1/s}$	$\lambda_p \ln(\beta n)\ln M(\nu)$ при $\alpha\beta\nu < n^{-1/s}$
				λ_∞ при $\bar{\alpha}\beta\nu < {}^1/_4$, $\nu \leqslant (\bar{\alpha}\beta)^{-2}$	$\lambda_\infty \ln M(\nu)$ при $\bar{\alpha}\beta\nu < {}^1/_{32}$, $\nu < (\bar{\alpha}\beta)^{-2}$
$\alpha_{1,n}(\|\cdot\|) \leqslant \alpha$, $\alpha_{\|\cdot\|}(G) \leqslant \beta$	ЗС$_{1,n}$	$\nu\alpha\beta \geqslant n^{-1/2}$	$c_1\alpha^2 \frac{\ln(n+1)}{\nu^2}$	$\lambda_1\alpha^4\beta^2$ при ${}^1/_4 > \alpha\beta\nu > n^{-1/8}$	$\lambda_1\alpha^4\beta^2 \ln^2 M(\nu)$ при $\alpha\beta\nu \geqslant n^{-1/2}$
				$\lambda_1\alpha^4\beta^2 \ln(n+1)$ при $\alpha\beta\nu \geqslant n^{-1/2}$	$\lambda_1\alpha^4\beta^2 \ln M(\nu)$ при ${}^1/_{32} > \alpha\beta\nu \geqslant n^{-1/8}$
	МЦТ	$\nu\alpha\beta < n^{-1/2}$	$c_\infty n]\ln(1/\nu)[$	$\lambda_1 \ln(\beta n)$ при $\alpha\beta\nu \leqslant n^{-1/2}$	$\lambda_1 \ln(\beta n)\ln M(\nu)$ при $\alpha\beta\nu < n^{-1/2}$
				λ_∞ при $\nu \leqslant (\beta n)^{-2}$	$\lambda_\infty \ln M(\nu)$ при $\nu \leqslant (\beta n)^{-2}$
$\alpha_{\infty,n}(\|\cdot\|) \leqslant \alpha$, $\alpha_{\|\cdot\|}(G) \leqslant \beta$	МЦТ		$c_\infty n]\ln(1/\nu)[$	λ_∞ при $\alpha\beta\nu < {}^1/_4$ и $\nu \leqslant (\alpha\beta)^{-2}$	$\lambda_\infty \ln M(\nu)$ при $\alpha\beta\nu < {}^1/_{32}$ и $\nu \leqslant (\alpha\beta)^{-2}$

Для дальнейшего удобно считать, что если множество минимумов f^k пересекается с границей G, то x_{k+1} выбирается на границе. Оказывается, что при этом метод Келли имеет экспоненциально зависящую от размерности задачи трудоемкость.

Упражнение 3. Пусть $n \geqslant 3$ и $\overline{C}$ — класс задач вида $f(x) \to \min \mid x \in V_n$ на единичном шаре V_n пространства E^n, порожденных выпуклыми липшицевыми с константой 1 функциями на V_n. Докажите, что если $M(\varepsilon)$ — число шагов «усеченного» метода Келли, решающего все задачи из $\overline{C}$ с абсолютной погрешностью $\varepsilon < 10^{-2}$, то

$$M(\varepsilon) \geqslant c(n)\left(\frac{1}{\varepsilon}\right)^{\frac{n-2}{2}},$$

где $c(n) > 0$.

⟨Пусть $x^1, \ldots, x^n$ — декартовы координаты в E^n. Рассмотрим задачу

$$f_\varepsilon(x) = \max\{|x^1|,\ g_\varepsilon(x)\},$$

где $g_\varepsilon(x) = [-1 + 2\varepsilon + \sqrt{(x^2)^2 + \ldots + (x^n)^2}]_+$. Пусть $e_1, \ldots, e_n$ — соответствующие выбранным координатам базисные орты, и пусть метод начинает свою работу в точке e_1. Обозначим через S множество $\{x \in V_n \mid x^1 = 0, \sum_{i=2}^{n} (x^i)^2 = 1\}$.

Легко видеть, что работа метода Келли на f_ε происходит следующим образом. Первая опрашиваемая точка есть e_1, вторая — $(-e_1)$. После этого метод начинает «зондировать» точки S. «Узнав» значение f_ε в точке $x \in S$, метод исключает из дальнейшего рассмотрения часть S, описываемую соотношением $\{y \in S \mid \langle y \mid x\rangle \geqslant 1 - 2\varepsilon\}$; очередная точка, в которой метод «опросит» f_ε, есть некая из неисключенных еще точек S. Так будет продолжаться, по крайней мере, до тех пор, пока на S будут еще неисключенные точки. В течение всего этого «зондирования» погрешность очередных приближений в качестве решений задачи f_ε, очевидно, не меньше 2ε.

Таким образом, «усеченный по точности ε» метод Келли обязан полностью выполнить весь этап «зондирования». С другой стороны, элементарные соображения, связанные с оценкой общей площади исключаемых частей S, показывают, что число вопросов на этапе «зондирования» допускает требуемую нижнюю оценку.⟩

Интересно отметить, что метод центров тяжести устроен «почти» так же, как и метод Келли; единственное отличие состоит в том, что в качестве очередной точки выбирается не точка минимума f^k, а центр тяжести области $\{x \in G \mid f^k(x) \leqslant a_k\}$, где a_k — рекордное (минимальное среди найденных к данному шагу) значение f. Казалось бы, правило выбора x_{k+1} в методе Келли более естественно, чем в МЦТ. Удивительно, что это более естественное правило приводит к несравненно худшему результату!

3.7. В заключение приведем серию упражнений, содержащих ответы на некоторые естественные вопросы о сложности классов общих (липшицевых) выпуклых задач.

Упражнение 4. *Сложность класса строго совместных выпуклых задач.* Пусть G — выпуклое замкнутое ограниченное тело в R^n. Докажите, что $*$-сложность $N_\theta^*(\nu)$ класса $C_\theta(G, R^n, m)$ по отношению к любому локальному детерминированному оракулу удовлетворяет оценке

$$\Phi_{n,\infty}(\theta\nu) \geqslant N_\theta^*(\nu) \geqslant \max\{\varphi_{n,\infty}(\alpha_{\infty,n}(G)\,\nu),\ \varphi_{n,\infty}(\alpha_{\infty,n}(G)\,\theta)\}.$$

Выведите отсюда, что МЦТ$_*$ — субоптимальный метод решения задач всех классов C_θ в диапазоне $\theta \leqslant \alpha_{\infty,n}^{-2}(G)$, $\nu \leqslant \alpha_{\infty,n}^{-2}(G)$.

Упражнение 5. *Сложность решения выпуклых задач «по аргументу» бесконечна при* $n > 1$. Рассмотрите класс A функций двух переменных на квадрате

$$K^2 = \{x \mid |x^1| \leqslant 1,\ |x^2| \leqslant 1\},$$

образованный функциями вида

$$f_\varphi(x^1, x^2) = \max\{\varphi(x^1),\ ax^2 + b\},$$

где $|a|, |b| \leqslant 1$, φ — липшицева функция на оси с константой Липшица 1. Докажите, что, каковы бы ни были детерминированные локальный оракул $\mathcal{O}$ для класса A и M-шаговый метод $\mathcal{B}^M$ минимизации функций из A, использующий $\mathcal{O}$, с любым натуральным M, найдется функция $f \in A$, удовлетворяющая следующему свойству: f достигает минимума на K^2 в единственной точке x_f, и если x^* — результат применения $\mathcal{B}^M$ к f, то $|x^* - x_f| \geqslant 1$. Докажите аналогичный результат для тел G размерности не меньше 2.

⟨Рассмотрим подкласс $\bar{A} \subset A$, образованный функциями, не зависящими от x^2. Найдется *) $\varphi \in \bar{A}$, такая, что вся траектория $\mathcal{B}^M$ на φ лежит в множестве $\varphi > 0$, тогда как $\min\limits_{|x^1| \leqslant 1} \varphi(x^1)$ достигается в единственной точке $x^1 = c_\varphi$ и отрицателен, скажем, равен $(-\varepsilon)$, $\varepsilon > 0$. Пусть x^* — результат применения $\mathcal{B}^M$ к φ и $(x^*)^2 \geqslant 0$ (случай $(x^*)^2 < 0$ рассматривается аналогично). Положим $a = \varepsilon/2$, $b = -\varepsilon/2$ и рассмотрим функцию

$$f(x^1, x^2) = \max\{\varphi(x^1),\ ax^2 + b\}.$$

Ясно, что f достигает минимума на K^2 в единственной точке $(c_\varphi, -1) = x_f$ и f совпадает с φ в области $\varphi > 0$. В силу последнего x^* есть результат применения $\mathcal{B}^M$ к f, так что $|x_f - x^*| \geqslant 1$.⟩

Упражнение 6. *Сложность решения липшицевых задач в неограниченных областях бесконечна.*

Пусть A — класс функций f на полуоси $\{t \geqslant 0\}$, липшицевых с константой 1 и выпуклых, таких, что $f(0) = 0$ и $\lim\limits_{t \to +\infty} f(t) = -1$. Докажите, что, каковы бы ни были детерминированный локальный оракул $\mathcal{O}$ для класса A и использующий $\mathcal{O}$ детерминированный метод $\mathcal{B}^M$ минимизации функций из A, найдется $f \in A$, такая, что $f(x^*) > -1/2$, где x^* — результат применения $\mathcal{B}^M$ к f. Распространите результат на случай произвольных неограниченных выпуклых тел G (отличных от всего пространства).

⟨Докажем индукцией по p для всякого $\mathcal{B}^p$ существование в A такой функции f_p, что на всей траектории $\{x_i\}$ работы $\mathcal{B}^p$ на f_p будет

$$f(x_i) > -\frac{1}{2} + \frac{1}{10p}$$

(результат применения $\mathcal{B}^p$ к f_p считается p-й точкой траектории).

* См. ниже, п. 5.1.

База $p = 1$ очевидна. Шаг индукции $p \Rightarrow p+1$.

Пусть дан метод $\mathcal{B}^{p+1}$ и f_p — та функция из A, для которой траектория первых p точек x_i работы $\mathcal{B}^{p+1}$ лежит в множестве

$$\left\{f_p(x) > -\frac{1}{2} + \frac{1}{10p}\right\}.$$

Пусть x_{p+1} есть $(p+1)$-я точка работы $\mathcal{B}^{p+1}$ на f_p. Возможно, что $x_{p+1} \leqslant \max\limits_{1 \leqslant i \leqslant p} x_i$. В этом случае положим $f_{p+1} = f_p$. Если же $x_{p+1} > \max\limits_{1 \leqslant i \leqslant p} x_i$, то найдется липшицева с константой 1 линейная функция φ, убывающая с ростом x, такая, что

$$\varphi(x_{p+1}) > -\frac{1}{2} + \frac{1}{10(p+1)},$$

тогда как

$$\varphi(x_i) < -\frac{1}{2} + \frac{1}{10p} \quad 1 \leqslant i \leqslant p.$$

Положим $f_{p+1} = \max\{f_p, \varphi\}$. Тогда f_{p+1} будет требуемой функцией.⟩

§ 4. Доказательство нижних оценок сложности. I

4.1. В этом параграфе мы начинаем обоснование нижних оценок сложности классов общих (липшицевых) выпуклых задач, сформулированных в теоремах 1.2 и 2.1. Сделаем по этому поводу несколько общих замечаний. Прежде всего имеется некая связь между оценками А.1.—А.5 для липшицева случая (теорема 2.1) и соответствующими результатами теоремы 1.2. Действительно, пусть $\mathcal{O}$ — произвольный детерминированный локальный оракул для поля задач $C(G, E, m)$. Какова бы ни была норма $\|\cdot\|$ на E (из числа задающих топологию), класс задач, получающийся оснащением $C(G, E, m)$ стандартными нормирующими множителями и оракулом $\mathcal{O}$, содержит класс задач, получающийся аналогичной операцией из поля задач $C_{\text{Lip}}(G, E, \|\cdot\|, m)$. При этом нормирующие множители второго класса на задачах из этого класса мажорируют нормирующие множители для первого класса. Ясно поэтому, что справедливая для всех указанных оракулов $\mathcal{O}$ нижняя оценка сложностей второго класса сохраняется и для первого. Поэтому, как бы ни выбиралась согласованная с топологией E норма $\|\cdot\|$ на E, отвечающие ей оценки А.1—А.5 теоремы 2.1 верны и для функций $N^{\mathcal{G}}(\cdot)$, $\tilde{N}^{\mathcal{G}}(\cdot)$, $N_*(\cdot)$ из теоремы 1.2. С другой стороны, верхние грани этих оценок по всевозможным нормам $\|\cdot\|$ на E в точности совпадают с правыми частями оценок А.1—А.5 теоремы 1.2 (следует иметь в виду, что $\varphi_{n,p}(t)$ непрерывны справа по $t > 0$). Таким образом, достаточно рассмотреть липшицеву ситуацию. Ясно, далее, что решать задачи с ограничениями ($m > 0$) никак не проще, чем без них ($m = 0$), т. е. при доказательстве А.1—А.4 (начиная с этого места, мы все время имеем в виду оценки теоремы 2.1) можно считать $m = 0$, что мы и будем делать.

И последнее замечание. Обоснование полученных оценок довольно громоздко. Чтобы не загромождать изложение второстепенными деталями, будем конспективно излагать чисто технические этапы доказательств, полностью описывая при этом идеи конструкций.

4.2. Перейдем теперь от общих замечаний к обоснованию нижних оценок сложности. Введем одну полезную для последующего изложения характеристику $k_{E,\|\cdot\|}(\varepsilon)$ банахова пространства E с нормой $\|\cdot\|$. Фиксируем E, $\|\cdot\|$. Пусть k — натуральное число. Рассмотрим набор $\bar{\xi} = (\xi_1, \ldots, \xi_k)$ k функционалов из E^* нормы $\leqslant 1$, и образуем 2^k функций на E вида

$$f_{\omega,\bar{\xi}}(x) = \max_{1 \leqslant i \leqslant m} \{\omega_i \langle \xi_i | x \rangle\}, \quad \omega_i = \pm 1, \ \omega = (\omega_1, \ldots, \omega_k). \tag{4.1}$$

Пусть

$$g(\bar{\xi}, \omega) = \inf_{x, \|x\| \leqslant 1} f_{\omega,\bar{\xi}}(x), \tag{4.2}$$

$$\bar{g}(\bar{\xi}) = \sup_{\omega} g(\bar{\xi}, \omega), \tag{4.3}$$

$$h_{E,\|\cdot\|}(k) = -\inf_{\bar{\xi} = (\xi_1, \ldots, \xi_k),\ \|\xi_i\|_* \leqslant 1} \bar{g}(\bar{\xi}).$$

Заметим, далее, что в силу леммы фон Неймана (см. пп. 1.1 гл. VI)

$$g(\bar{\xi}, \omega) = \inf_{\|x\| \leqslant 1} \max_{1 \leqslant i \leqslant k} \{\langle \omega_i \xi_i | x \rangle\} = \inf_{\|x\| = 1} \sup_{\substack{\lambda = (\lambda_1, \ldots, \lambda_k) \\ \lambda \geqslant 0,\ \sum \lambda_i = 1}} \{\langle \textstyle\sum \lambda_i \omega_i \xi_i | x \rangle\} =$$

$$= \sup_{\lambda \geqslant 0,\ \sum \lambda_i = 1} \inf_{\|x\| = 1} \{\langle \textstyle\sum \lambda_i \omega_i \xi_i | x \rangle\} = \sup_{\lambda \geqslant 0,\ \sum \lambda_i = 1} (-\|\textstyle\sum \lambda_i \omega_i \xi_i\|_*),$$

так что

$$\bar{g}(\bar{\xi}) = -\inf_{\sum |\lambda_i| = 1} \|\textstyle\sum \lambda_i \xi_i\|_*. \tag{4.4}$$

Итак, функция $\bar{g}(\bar{\xi})$ допускает простую геометрическую интерпретацию: ее значение на наборе $\bar{\xi}$ есть расстояние в E^* от 0 до границы «гипероктаэдра» с вершинами $\pm \xi_i$.

Из сказанного ясно, что $h_{E,\|\cdot\|}(k)$ есть невозрастающая функция k. Кроме того, $h_{E,\|\cdot\|}(k) = 0$, если $k > \dim E$ (почему?), тогда как $h_{E,\|\cdot\|}(1) = 1$. В силу последнего обстоятельства для всякого $\varepsilon < {}^1/_4$ существуют k, такие, что $h_{E,\|\cdot\|}(k) > 4\varepsilon$.

Положим

$$k_{E,\|\cdot\|}(\varepsilon) = \sup\{k \mid h_{E,\|\cdot\|}(k) > 4\varepsilon\}. \tag{4.5}$$

Если же $\varepsilon \geqslant {}^1/_4$, то положим $k_{E,\|\cdot\|}(\varepsilon) = 1$. Функция $k_{E,\|\cdot\|}(\varepsilon)$ есть невозрастающая функция ε, определенная при $\varepsilon > 0$ и принимающая целые неотрицательные значения и, возможно, значение $+\infty$. Ее роль в оценке сложности выпуклого программирования связана с основным в параграфе предложением, приведенным в следующем пункте.

4.3. Обозначим через W — единичный шар пространства E, $\|\cdot\|$, а $A(E)$ — множество функций, представимых в виде

$$f(x)=\max_{1\leqslant i\leqslant N}\theta_i(\langle\xi_i\,|\,x\rangle), \tag{4.6}$$

где $\xi_i\in E^*$, $\|\xi_i\|_*\leqslant 1$, а $\theta_i(t)$ — выпуклые липшицевы с константой 1 функции на оси. Имеют место следующие предложения.

Предложение 1. Пусть $\mathcal{O}$ — произвольный детерминированный локальный оракул для поля задач $A(E)$, $\varepsilon>0$, M натуральное и $\mathcal{B}^M$ — использующий $\mathcal{O}$ M-шаговый детерминированный метод минимизации функций из $A(E)$. Тогда, если

$$M<k_{E,\|\cdot\|}(\varepsilon), \tag{4.7}$$

то в классе $A(E)$ найдется функция f, обладающая следующими свойствами:

$$\inf_W f(x)<-4\varepsilon \quad\text{и}\quad f(\bar x)>0, \tag{4.8}$$

где $\bar x$ — результат применения $\mathcal{B}^M$ к f.

Предложение 2. При тех же предположениях пусть $\widetilde{\mathcal{B}}^M$ — любой стохастический метод минимизации функций из $A(E)$, использующий оракул $\mathcal{O}$ и такой, что $l(\widetilde{\mathcal{B}}^M,f)\leqslant M$ для всех $f\in A(E)$. Тогда, если

$$64M<\frac{k_{E,\|\cdot\|}(8\varepsilon)}{1+\log_2 k_{E,\|\cdot\|}(8\varepsilon)}\equiv\bar k(\varepsilon), \tag{4.9}$$

то в классе $A(E)$ найдется функция f, обладающая свойствами

$$\inf_{x\in W}f(x)<0 \quad\text{и}\quad \mathsf{M}\max(f(\bar x_f),0)\equiv\tilde\varepsilon(\widetilde{\mathcal{B}}^M,f)>4\varepsilon. \tag{4.10}$$

Здесь $\bar x_f$ — (случайный) результат применения $\widetilde{\mathcal{B}}^M$ к f, а M — усреднение по распределению этого результата, индуцированному распределением траекторий $\widetilde{\mathcal{B}}^M$ на f. При $\bar x_f=*$ или $\bar x_f=\phi$ считается $f(\bar x_f)\equiv 1$.

Доказательство.

1. Можно считать $\varepsilon<{}^1/_4$ — в противном случае (4.7) невозможно. Пусть k — натуральное число, не превышающее $k_{E,\|\cdot\|}(\varepsilon)$. Это значит, что $h_E(k)>4\varepsilon$. Выберем $\delta>0$ так, чтобы было $h_E(k)>4\varepsilon+\delta$. Последнее неравенство по определению h_E означает существование такого набора $\xi_1,\ldots,\xi_k\in E^*$, $\|\xi_i\|_*\leqslant 1$, что при всех $\omega=(\omega_1,\ldots,\omega_k)$, $\omega_i=\pm 1$, имеем

$$\min_{\|x\|\leqslant 1}\max_{1\leqslant i\leqslant k}\{\langle\omega_i\xi_i\,|\,x\rangle\}<-4\varepsilon-\delta. \tag{4.11}$$

Выберем k чисел $\delta_1,\ldots,\delta_k$ так, чтобы было $\delta>\delta_1>\ldots>\delta_k>0$. Пусть $\mathcal{B}^M$ — метод минимизации функций из $A(E)$, детерминированный и использующий оракул $\mathcal{O}$, с $M<k$. Достаточно показать, что в классе $A(E)$ найдется функция, обладающая по отношению к $\mathcal{B}^M$ свойством (4.8).

Построим эту функцию с помощью k-шаговой конструкции. Очевидно, $\mathcal{B}^M$ можно рассмотреть как k-шаговый метод, результат применения которого к любой функции из $A(E)$ есть последняя (k-я) точка, в которой метод задает вопрос о задаче. Соответственно будем теперь обозначать метод $\mathcal{B}^k$. Пусть $I_0=\{1,\ldots,k\}$.

Первый шаг. Пусть x_1 — первая точка, в которой метод задает вопрос. Выберем среди чисел $\langle\xi_i|x_1\rangle$, $i\in I_0$, наибольшее по модулю. Пусть соответствующее i есть $i(1)$. Положим

$$\omega_i=\begin{cases}+1, & \langle\xi_{i(1)}\,|\,x_1\rangle\geqslant 0,\\ -1, & \langle\xi_{i(1)}\,|\,x_1\rangle<0,\end{cases}$$

$I_1=I_0\setminus\{i(1)\}$, и пусть $f_1(x)=\omega_1\langle\xi_{i(1)}|x\rangle+\delta_1$. Шаг закончен.

$s+1$*-й шаг.* Пусть после s шагов построения имеются

a) попарно различные числа $i(j)$, $1\leqslant j\leqslant s$, из множества I_0 и множество $I_s=I_0\setminus\{i(1),\ldots,i(s)\}$;

b) числа $\omega_j=\pm 1$, $1\leqslant j\leqslant s$;

c) функции $f_j(x)=\max\limits_{1\leqslant t\leqslant j}\{\omega_t\langle\xi_{i(t)}|x\rangle+\delta_t\}$;

d) точки $x_1,x_2,\ldots,x_{s+1}$, где x_{j+1} — $(j+1)$-я точка работы $\mathcal{B}^k$ на f_j.

На $(s+1)$-м шаге найдем максимальное по модулю среди чисел $\langle\xi_i|x_{s+1}\rangle$, $i\in I_s$,— пусть ему отвечает номер $i(s+1)$. Обозначим

$$\omega_{s+1}=\begin{cases}+1, & \langle\xi_{i(s+1)}\,|\,x_{s+1}\rangle\geqslant 0,\\ -1, & \langle\xi_{i(s+1)}\,|\,x_{s+1}\rangle<0,\end{cases}\quad I_{s+1}=I_s\setminus\{i(s+1)\},$$

и

$$f_{s+1}(x)=\max_{1\leqslant t\leqslant s+1}\{\omega_t\langle\xi_i(t)\,|\,x\rangle+\delta_t\}.$$

Пусть x_{s+2} — $(s+2)$-я точка работы $\mathcal{B}^k$ на f_{s+1} (x_{s+2} определяется лишь при $s+1<k$). Шаг построения закончен.

После k шагов такого построения будут получены функции $f_1,\ldots,f_k$, точки $x_1,\ldots,x_k$ и числа $\omega_s=\pm 1$, $1\leqslant s\leqslant k$. При этом $f_i\in A(E)$ и

1) $f_k(x)=\max\limits_{1\leqslant s\leqslant k}\{\omega_s\langle\xi_{i(s)}\,|\,x\rangle+\delta_s\}$, $f_1\leqslant f_2\leqslant\ldots\leqslant f_k$;

2) x_{s+1} есть $(s+1)$-я точка работы $\mathcal{B}^k$ на f_s, $1\leqslant s\leqslant k-1$; x_1 — первая точка работы $\mathcal{B}^k$;

3) $\langle\omega_l\xi_{i(l)}|x_l\rangle\geqslant\max\limits_{l\leqslant s\leqslant k}\langle\omega_s\xi_{i(s)}|x_l\rangle$, $1\leqslant l\leqslant k$, и

$$\langle\omega_l\xi_{i(l)}\,|\,x_l\rangle\geqslant 0$$

(это следует из правил формирования $i(l)$ и ω_l).

Убедимся, что x_k есть результат применения $\mathcal{B}^k$ к f_k и f_k — требуемая предложением функция. Проверим первое. По лемме о неразличении (п. 5.3 гл. I) достаточно показать, что $f_s = f_l$ в окрестности x_s, $l \geqslant s$. Докажем, что $f_s \equiv f_k$ в окрестности x_s, $1 \leqslant s \leqslant k$. Так как $f_s \leqslant f_{s+1} \leqslant \ldots \leqslant f_k$, то отсюда будет следовать, что при $s, s' \geqslant t$ $f_s \equiv f_{s'}$ в окрестности x_t, что нам и требуется. Имеем

$$f_s(x) \geqslant \langle \omega_s \xi_{i(s)} | x \rangle + \delta_s,$$
$$f_k(x) = \max \{f_s(x), \max_{s<t\leqslant k} (\langle \omega_t \xi_{i(t)} | x \rangle + \delta_t)\} \equiv \max \{f_s(x), f^s(x)\}.$$

Из первого соотношения следует, что

$$f_s(x_s) \geqslant |\langle \xi_{i(s)} | x_s \rangle| + \delta_s,$$

тогда как при всех $t > s$

$$\omega_t \langle \xi_{i(t)} | x \rangle \leqslant |\langle \xi_{i(s)} | x_s \rangle|$$

в силу п. 3). Так как еще и $\delta_s \leqslant \delta_t$ при $s < t$, то $f_s(x_s) > f^s(x_s)$. Стало быть, в окрестности x_s $f_k \equiv \max \{f^s, f_s\}$ совпадает с f_s, что и требуется.

Теперь убедимся, что f_k — искомая функция. Ясно, что $f_k \in A(E)$. Далее, x_k есть результат применения $\mathcal{B}^k$ к f_k и

$$f_k(x_k) \geqslant \omega_k \langle \xi_{i(k)} | x_k \rangle + \delta_k > 0.$$

С другой стороны,

$$\inf_{\|x\|\leqslant 1} f_k(x) = \inf_{\|x\|\leqslant 1} \max_{1\leqslant t\leqslant k} \{\langle \omega_t \xi_{i(t)} | x \rangle + \delta_t\} \leqslant$$
$$\leqslant \inf_{\|x\|\leqslant 1} \max_{1\leqslant t\leqslant k} \{\langle \omega_t \xi_{i(t)} | x \rangle\} + \delta < -4\varepsilon - \delta + \delta = -4\varepsilon$$

(мы учли, что $i(1), \ldots, i(k)$ — перестановка чисел $1, \ldots, k$, и соотношение (4.11)). Предложение доказано.

2. *Докажем теперь Предложение* 2.

1°. Пусть $\widetilde{\mathcal{B}}^M$, ε — объекты из формулировки этой части. Выполнение (4.9) невозможно для $\varepsilon \geqslant 1/_{32}$, так что будем считать $\varepsilon < 1/_{32}$. Пусть натуральное k и функционалы $\xi_1, \ldots, \xi_k \in E^*$, $\|\xi_i\|_* \leqslant 1$, а также число $\delta > 0$ выбраны так, что

$$\frac{k}{64 \log_2(2k)} > M$$

и

$$\min_{x\in W} \max_{1\leqslant i\leqslant k} \langle \omega_i \xi_i | x \rangle < -32\varepsilon - \delta.$$

В силу последнего при

$$\Delta = \{\bar\delta \mid \bar\delta = (\delta_1, \ldots, \delta_k),\ 0 < \delta_i < \delta\}$$

и

$$\Omega = \{\bar\omega = (\omega_1, \ldots, \omega_k) \mid \omega_i = \pm 1\}$$

имеем

$$\min_{x\in W} f_{\bar\omega,\bar\delta}(x) \equiv \min_{x\in W} \{\max_{1\leqslant i\leqslant k} (\omega_i \langle \xi_i | x \rangle + 32\varepsilon + \delta_i)\} < 0. \tag{4.12}$$

Обозначим через $p(\bar\omega)$ равновероятное распределение на Ω.

2°. Пусть X^R — произвольный набор из R попарно различных точек интервала $(0, \delta)$, а $\Delta_R = \{\bar\delta \in \Delta \mid \delta_i \in X^R,\ 1 \leqslant i \leqslant k\}$. Рассмотрим множество задач $A^R = \{f_{\bar\omega,\bar\delta} \mid \bar\omega \in \Omega,\ \bar\delta \in \Delta_R\}$ и снабдим это множество равновероятным распределением $\sigma(\bar\omega, \bar\delta) = p(\bar\omega) q_R(\bar\delta)$. Пусть $\mathcal{O}$ — оракул из формулировки предложения, а $\bar{\mathcal{O}}$ — оракул, ответ которого о функции $f_{\bar\omega,\bar\delta}$ в точке x состоит в указании числа $f_{\bar\omega,\bar\delta}(x)$ и опорного функционала $f'_{\bar\omega,\bar\delta}(x)$ вида $\pm\xi_1, \ldots, \pm\xi_k$ к функции $f_{\bar\omega,\bar\delta}(x)$ (такой опорный функционал всегда найдется) в точке x.

Пусть

$$\bar{\bar\varepsilon} = \max_{\bar\omega,\bar\delta} \{\bar\varepsilon(\widetilde{\mathcal{B}}^M, f_{\bar\omega,\bar\delta}) \mid \bar\omega \in \Omega,\ \bar\delta \in \Delta\}.$$

Тогда

$$\left.\begin{aligned} &\int_{A^R} l(\widetilde{\mathcal{B}}^M, f_{\bar\omega,\bar\delta})\, dp(\bar\omega)\, dq_R(\bar\delta) \leqslant \frac{k}{64 \log_2(2k)}, \\ &\int_{A^R} \bar\varepsilon(\widetilde{\mathcal{B}}^M, f_{\bar\omega,\bar\delta})\, dp(\bar\omega)\, dq_R(\bar\delta) \leqslant \bar{\bar\varepsilon}. \end{aligned}\right\} \tag{4.13}$$

Как и в предложении 5.2 гл. I, найдется детерминированный правильный метод $\mathcal{B}^{tR}$, такой, что

$$\left.\begin{aligned} &\int_{A^R} l(\mathcal{B}^{tR}, f_{\bar\omega,\bar\delta})\, dp(\bar\omega)\, dq_R(\bar\delta) \leqslant \frac{k}{32 \log_2(2k)}, \\ &\int_{A^R} \bar\varepsilon(\mathcal{B}^{tR}, f_{\bar\omega,\bar\delta})\, dp(\bar\omega)\, dq_R(\bar\delta) \leqslant 2\bar{\bar\varepsilon}. \end{aligned}\right\} \tag{4.14}$$

3°. Построим по методу $\mathcal{B}^{tR}$, работающему с оракулом $\mathcal{O}$, детерминированный метод $\mathcal{B}_R$, использующий оракул $\bar{\mathcal{O}}$, следующим образом. Первая поисковая точка у $\mathcal{B}_R$ — та же, что и у $\mathcal{B}^{tR}$. Опишем построение $(s+1)$-й поисковой точки. Пусть $x_1, \ldots, x_s$ — начальный фрагмент траектории $\mathcal{B}_R$ на $f \in A^R$, $h_1(y), \ldots, h_s(y)$ — аффинные функционалы, сообщенные оракулом $\bar{\mathcal{O}}$ в точках $x_1, \ldots, x_s$ соответственно (т. е. функционалы $f(x_j) + \langle f'(x_j) | y - x_j \rangle$, где $f'(x)$ — сообщаемый оракулом $\bar{\mathcal{O}}$ опорный функционал). Образуем функцию $f^s(y) = \max_{1\leqslant i\leqslant s} h_i(y)$ и возьмем в качестве $(s+1)$-й точки траектории $\mathcal{B}_R$ на f $(s+1)$-ю точку траектории $\mathscr{B}^{tR}$ на f^s (последняя однозначно определяется по f^s, т. е. по информации, полученной $\mathscr{B}_R$ от $\overline{\mathcal{O}}$ на первых s шагах работы $\mathscr{B}_R$). Если при этом получится $x_{s+1} = *$ или $x_{s+1} = \phi$, то, начиная с номера $j = s+2$, положим $x_j = \phi$.

4°. Докажем, что если R достаточно велико, то начальные фрагменты траекторий $\mathscr{B}_R$ и $\mathscr{B}^{tR}$ (длины k) на «большинстве» задач из A^R совпадают друг с другом. Именно, если $a(R)$ есть вероятность, вычисленная по распределению $\sigma(\cdot)$, события

$$Q(R) \equiv \{f_{\overline{\omega},\overline{\delta}} \mid \text{начальные фрагменты длины } k \text{ траекторий } \mathscr{B}_R \text{ и } \mathscr{B}^{tR} \text{ на } f_{\overline{\omega},\overline{\delta}} \text{ отличаются друг от друга}\},$$

то $a(R) \underset{R\to\infty}{\longrightarrow} 0$.

Откладывая пока доказательство этого факта, извлечем из него желаемое утверждение. Действительно, пусть

$$A_1 = \left\{(\overline{\omega}, \overline{\delta}) \in \Omega \times \Delta_R \,\middle|\, l(\mathscr{B}^{tR}, f_{\overline{\omega},\overline{\delta}}) \leqslant \frac{k}{4\log_2(2k)}\right\}.$$

Ввиду (4.14), тогда $\sigma(A_1) \geqslant {}^7/_8$. Пусть теперь $A_2(R) = A_1 \setminus Q(R)$. Так как $\sigma(Q(R)) = a(R) \underset{R\to\infty}{\longrightarrow} 0$, то существует $\overline{R}$, для которого $\sigma(A_2(\overline{R})) > {}^6/_7$. Из построения ясно, что результат применения $\mathscr{B}_{\overline{R}}$ к $f_{\overline{\omega},\overline{\delta}} \in A_2(\overline{R})$ получается не позже, чем на шаге с номером $\overline{s} \equiv [k/4\log_2(2k)]$, и при этом

$$\int\limits_{A_2(\overline{R})} \tilde{\varepsilon}(\mathscr{B}_{\overline{R}}, f_{\overline{\omega},\overline{\delta}})\, d\sigma(\overline{\omega}, \overline{\delta}) = \int\limits_{A_2(\overline{R})} \tilde{\varepsilon}(\mathscr{B}^{t\overline{R}}, f_{\overline{\omega},\overline{\delta}})\, d\sigma(\overline{\omega}, \overline{\delta}) \leqslant 2\overline{\overline{\varepsilon}}.$$

В силу последнего неравенства и соотношения $\sigma(A_2(\overline{R})) \geqslant {}^6/_7$ при некотором $\overline{\delta}_0 \in \Delta_{\overline{R}}$ имеем, полагая $\Omega_{\overline{\delta}_0} = \{\overline{\omega} \mid f_{\overline{\omega},\overline{\delta}_0} \in A_2(\overline{R})\}$:

$$\left.\begin{aligned} &\int\limits_{\Omega_{\overline{\delta}_0}} \tilde{\varepsilon}(\mathscr{B}_{\overline{R}}, f_{\overline{\omega},\overline{\delta}_0})\, dp(\overline{\omega}) \leqslant 3\overline{\overline{\varepsilon}}, \\ &\int\limits_{\Omega_{\overline{\delta}_0}} dp(\overline{\omega}) \geqslant {}^1/_2. \end{aligned}\right\} \tag{4.15}$$

5°. Рассмотрим теперь работу $\mathscr{B}_{\overline{R}}$ на множестве задач $A_{\overline{\delta}_0} = \{f_{\overline{\omega},\overline{\delta}_0} \mid \overline{\omega} \in \Omega\}$. На этом множестве, очевидно,

$$f_{\overline{\omega},\overline{\delta}_0}(x) = \langle f'_{\overline{\omega},\overline{\delta}_0}(x) \mid x\rangle + \overline{\delta}_{i(f'_{\overline{\omega},\overline{\delta}_0}(x)),\,0} + 32\varepsilon,$$

где $i(f'_{\overline{\omega},\overline{\delta}_0}(x)) = j$, если $f'_{\overline{\omega},\overline{\delta}_0}(x) = \pm\xi_j$. Итак, на $A_{\overline{\delta}_0}$ весь ответ оракула $\overline{\mathcal{O}}$ на вопрос в точке x однозначно восстанавливается по одной своей компоненте — опорному функционалу $f'(x)$ (и, разумеется, по точке x). Таким образом, при изучении работы $\mathscr{B}_{\overline{R}}$ на

$A_{\overline{\delta}_0}$ можно считать, что этот метод использует оракул $\overline{\overline{\mathcal{O}}}$, сообщающий не пары $f(x)$, $f'(x)$, а только $f'(x)$. Соответственно ответ $\overline{\overline{\mathcal{O}}}$ на любой вопрос есть точка фиксированного множества из $2k$ элементов. Так как $\mathscr{B}_{\overline{R}}$ детерминирован, то отсюда следует, что различных начальных фрагментов длины s его траекторий на задачах $f \in A_{\overline{\delta}_0}$ не более $(2k)^s$. Вспоминая определение $\Omega_{\overline{\delta}_0}$, выводим отсюда, что различных результатов, выдаваемых $\mathscr{B}_{\overline{R}}$ на всевозможных задачах $f_{\overline{\omega},\overline{\delta}_0}$, $\overline{\omega} \in \Omega_{\overline{\delta}_0}$, может быть не более

$$(2k)^{k/4\log_2(2k)} = 2^{k/4}.$$

С другой стороны, ясно, что множества $\{x \in W \mid f_{\overline{\omega},\overline{\delta}_0}(x) < 32\varepsilon\}$ при различных $\overline{\omega}$ попарно не пересекаются, т. е. не более чем $2^{k/4}$ задач f из $\Omega_{\overline{\delta}_0}$ может быть выполнено неравенство $\tilde{\varepsilon}(\bar{x}_f, f) < 32\varepsilon$, где $\bar{x}_f$ —результат применения $\mathscr{B}_{\overline{R}}$ к f. Отсюда и из второго соотношения в (4.15) следует

$$\int\limits_{\Omega_{\overline{\delta}_0}} \tilde{\varepsilon}(\mathscr{B}_{\overline{R}}, f_{\overline{\omega},\overline{\delta}_0})\, dp(\overline{\omega}) \geqslant 32\varepsilon\left(\frac{1}{2} - \frac{2^{k/4}}{2^k}\right),$$

т. е. в силу первого соотношения в (4.15)

$$\overline{\overline{\varepsilon}} \geqslant 10\,\varepsilon\left(\frac{1}{2} - 2^{-3k/4}\right).$$

Отсюда при $k > 5$ сразу следует $\overline{\overline{\varepsilon}} > 4\varepsilon$. Последнее в силу (4.12) и определения $\overline{\overline{\varepsilon}}$ доказывает (4.10). Если же $k \leqslant 5$, то (4.9), очевидно, невозможно, так что (4.10) доказано.

6°. Для завершения доказательства остается проверить, что $a(R) \underset{R\to\infty}{\longrightarrow} 0$. С этой целью введем понятие t-правильной ($t = 0, 1, \ldots$) функции $f \in A^R$. По определению всякая f 0-правильна и f $(t+1)$-правильна тогда и только тогда, когда она t-правильна и $f^{(t+1)} \equiv f$ в окрестности x_{t+1}, где x_{t+1} — $(t+1)$-я точка работы $\mathscr{B}_R$ на f (по определению все $f \in A^R$ совпадают друг с другом в окрестности $*$ и ϕ). Иными словами, f *t-правильна*, если при всех s, $1 \leqslant s \leqslant t$, $f = f^s$ в окрестности x_s. Из леммы о неразличении (п. 5.3 гл. I) и определения $\mathscr{B}_R$ сразу следует, что если f t-правильна, то начальные фрагменты длины $t+1$ траекторий $\mathscr{B}^{tR}$ и $\mathscr{B}_R$ на f совпадают друг с другом.

Пусть $\eta^s = (\eta_1, \ldots, \eta_s)$ означает набор ответов *) оракула $\overline{\mathcal{O}}$ на вопросы $\mathscr{B}_{\overline{R}}$ на первых s шагах работы $\mathscr{B}_{\overline{R}}$, H^s — множество

*) Каждый такой ответ есть аффинный функционал вида

$$\eta(y) = \omega\langle\xi_j \mid y\rangle + \delta,\ \omega = \pm 1,\ j \in \overline{1, k},\ \delta \in \mathbb{R}.$$

Мы отождествляем η с тройкой (ω, j, δ).

всех таких наборов,

$\mathcal{F}_R(\eta^s) = \{f \in A^R \mid$ **информация о** f **на первых** s **шагах работы** $\mathcal{B}_R$ на f есть $\eta^s\}$.

Пусть еще для ответа $\eta = \{\omega, j, \delta\}$ оракула $\overline{\mathcal{O}}$ на вопрос о $f \in A^R$ $i(\eta) = j$, и пусть $i(\eta^s)$ есть подмножество $\{j \in (1, 2, \ldots, k) \mid j = i(\eta_l)$ для какого-нибудь $l \leqslant s\}$.

Подсчитаем число $L_R(t)$ функций $f \in A^R$, не являющихся t-правильными. Напомним, что требуется доказать, что

$$a(R) \equiv L_R(k)/2^k R^k \xrightarrow[R\to\infty]{} 0.$$

Пусть $N_R(t)$ есть число t-неправильных $(t-1)$-правильных функций, а $N_R(\eta^{t-1}, t)$ — число таких функций в $\mathcal{F}_R(\eta^{t-1})$. Тогда, очевидно, $L_R(0) = 0$ и

$$L_R(t+1) = L_R(t) + \sum_{\eta^t \in H^t} N_R(\eta^t, t+1). \tag{4.16}$$

Далее, если f — $(t+1)$-неправильная t-правильная функция в $\mathcal{F}_R(\eta^t)$, то это значит, что при некоторых $i_1 \neq i_2$ имеем

$$\omega_{i_1}\langle \xi_{i_1} | x_{t+1}\rangle + \delta_{i_1} = \omega_{i_2}\langle \xi_{i_2} | x_{t+1}\rangle + \delta_{i_2}, \tag{4.17}$$

причем, по крайней мере, одно из чисел i_1, i_2 не лежит в $i(\eta^t)$. Фиксация η^t однозначно определяет δ_j, ω_j, $j \in i(\eta^t)$. Поэтому легко сообразить, что (4.17) может (при фиксированном η^t) иметь место не более чем для $\varphi(k)R^{k-|i(\eta^t)|-1}$ функций $f \in \mathcal{F}_R(\eta^t)$. Здесь $|i(\eta^t)|$ — число элементов в $i(\eta^t)$, $\varphi(k) < \infty$ не зависит от R. Замечая, что число различных η^t с заданным $i(\eta^t)$ и непустым $\mathcal{F}_R(\eta^t)$ не превосходит $\varphi(k)R^{|i(\eta^t)|}$, получаем, что

$$L_R(t+1) \leqslant L_R(t) + \varphi^2(k)R^{k-1}N(k),$$

где $N(k)$ есть число различных подмножеств множества $1, 2, \ldots, k$. Итак,

$$L_R(t) \leqslant t\varphi^2(k)N(k)R^{k-1},$$

что и дает требуемое соотношение $a(R) \xrightarrow[R\to\infty]{} 0$. Предложение доказано.

4.4. В качестве следствия получаем оценки $N^{\mathcal{O}}(\nu)$, $\widetilde{N}^{\mathcal{O}}(\nu)$ — детерминированной и стохастической сложностей класса задач, получающегося из класса $\boldsymbol{C}_{\mathrm{Lip}}$ типа $\boldsymbol{C}^0_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$ заменой оракула класса произвольным детерминированным локальным оракулом $\mathcal{O}$.

Следствие. *В обозначениях п.* 4.2 *справедливы оценки*

$$N^{\mathcal{O}}(\nu) \geqslant k_{E,\|\cdot\|}(\alpha_{\|\cdot\|}(G)\,\nu) \tag{4.18}$$

и

$$\widetilde{N}^{\mathcal{O}}(\nu) \geqslant \frac{1}{64}\,\frac{k_{E,\|\cdot\|}(8\alpha_{\|\cdot\|}(G)\,\nu)}{\log_2(2k_{E,\|\cdot\|}(8\alpha_{\|\cdot\|}(G)\,\nu))}. \tag{4.19}$$

Доказательство. Не ограничивая общности, можно считать, что G содержит единичный шар W пространства E, $\|\cdot\|$ и содержится в шаре радиуса $\alpha_{\|\cdot\|}(G)$. (Строго говоря, в этой фразе следовало бы заменить $\alpha_{\|\cdot\|}(G)$ произвольным $\alpha > \alpha_{\|\cdot\|}(G)$. Но такая замена, ввиду возможности выбирать α сколь угодно близким к $\alpha_{\|\cdot\|}(G)$ и непрерывности справа по ν правых частей (4.18) — (4.19), не изменяет результата последующих рассуждений.) Как отмечалось в п. 4.1, можно считать $m = 0$.

Рассмотрим теперь задачи из поля $A(E)$. Пусть $\nu > 0$ и $\varepsilon = \alpha_{\|\cdot\|}(G)\nu$. Обозначим $N = N^{\mathcal{O}}(\nu)$. Пусть (4.18) неверно. Это значит, что существует детерминированный метод $\mathcal{B}^M$ с трудоемкостью на классе $\boldsymbol{C}_{\mathrm{Lip}}$, меньшей $k_{E,\|\cdot\|}(\varepsilon)$, решающий все задачи из $\boldsymbol{C}_{\mathrm{Lip}}$ с погрешностью, $\leqslant \nu$.

Применим теперь к объектам $\mathcal{O}$, $\mathcal{B}^M$, ε утверждение предложения 1 п. 4.3. По этому утверждению для некоторого $f \in A(E)$

$$\inf_W f(x) < 0 \quad \text{и} \quad f(\bar{x}) > 4\varepsilon = 4\alpha_{\|\cdot\|}(G)\nu,$$

где $\bar{x}$ — результат работы $\mathcal{B}^M$ на f. Так как $W \subset G$ по условию, то

$$\nu(\mathcal{B}^M, f) > 4\varepsilon/2\mathrm{L}_{\|\cdot\|}(f)\rho_{\|\cdot\|}(G).$$

Но $\rho_{\|\cdot\|}(G) \leqslant 2\alpha_{\|\cdot\|}(G)$ и $\mathrm{L}_{\|\cdot\|}(f) \leqslant 1$ (последнее в силу свойств $A(E)$), т. е. $\nu(\mathcal{B}^M, f) > \nu$, что противоречит определению $\mathcal{B}^M$. Полученное противоречие доказывает (4.18); (4.19) доказывается аналогично (в доказательстве используется предложение 2 п.4.3 вместо предложения 1, обосновывающего (4.18)). Следствие доказано.

4.5. Теперь мы в состоянии доказать оценки А.1 (для $\infty > p \geqslant 1$). Для этого достаточно обосновать следующие два факта.

4.5.1. Характеристика $k_{E,\|\cdot\|}(t)$ обладает следующим свойством. Пусть E' — подпространство E, $\|\cdot\|'$ — норма на E', относительно которой $\|\cdot\|$ — норма на E' имеет асферичность β. Тогда

$$k_{E,\|\cdot\|}(t) \geqslant k_{E',\|\cdot\|'}(\beta t). \tag{4.20}$$

4.5.2. Если $1 \leqslant p < \infty$, то при надлежащем выборе величин $d(p) > 0$ для соответствующих функций $\varphi_{n,p}$ (определяемых из (1.11)) справедливы оценки $(1 \leqslant n < \infty)$:

$$k_{l_p^{(n)},\|\cdot\|_p}(t) \geqslant \varphi_{n,p}(t). \tag{4.21}$$

Поясним, как из пп. 4.5.1 — 4.5.2 и из (4.18) — (4.19) следует справедливость оценок (2.1) — (2.2) при $1 \leqslant p < \infty$. Во-первых, из определения $\alpha_{p, \|\cdot\|, k}(G)$ и непрерывности справа по ν правых частей (4.18) — (4.19) из п. 4.5 и (4.18) — (4.19) получаем

$$N^{\mathfrak{G}}(\nu) \geqslant k_{\mathfrak{l}_p^{(k)}, \|\cdot\|_p}(\alpha_{p, \|\cdot\|, k}(G)\nu)$$

и

$$\widetilde{N}^{\mathfrak{G}}(\nu) \geqslant \frac{k_{\mathfrak{l}_p^{(k)}, \|\cdot\|_p}(8\alpha_{p, \|\cdot\|, k}(G)\nu)}{64 \log_2 (2k_{\mathfrak{l}_p^{(k)}, \|\cdot\|_p}(8\alpha_{p, \|\cdot\|, k}(G), \nu))}$$

при всех натуральных $k \leqslant \dim E$. Во-вторых, подстановка в последние оценки оценок п. 4.5.2 приводит к (2.1) — (2.2). Остается обосновать пп. 4.5.1 и 4.5.2.

Доказательство п. 4.5.1. Ясно, что единичный шар V'_* пространства $(E')^*$ относительно нормы $\|\cdot\|'_*$ имеет асферичность β относительно $\|\cdot\|_*$-единичного шара V_* того же пространства. Без ограничения общности можно считать, что $V_* \subset V'_* \subset \beta V_*$. Пусть теперь $k_{E', \|\cdot\|'}(\beta\varepsilon) = m$. Если $m = 1$, то (4.20) выполнено, если же $m > 1$, т. е. $\beta\varepsilon < {}^1/_4$, то из определения $k_{E', \|\cdot\|'}$ *следует, что найдется* m *функционалов* $\bar{\xi}_1, \ldots, \bar{\xi}_m \in V'_*$, *таких,* что при всех λ, $\sum\limits_0^m |\lambda_i| = 1$, $\|\Sigma\lambda_i\bar{\xi}_i\|'_* > 4\beta\varepsilon$. Так как $V'_* \subset \beta V_*$, то $\|\bar{\xi}_i\|_* \leqslant \beta$. Положим $\xi_i = \bar{\xi}_i/\beta$. Тогда ξ_i — функционал нормы $\leqslant 1$ на $(E', \|\cdot\|)$. Стало быть, ξ_i можно продолжить с E' до функционала (обозначим его $\hat{\xi}_i$) на $(E, \|\cdot\|)$ нормы $\leqslant 1$. Очевидно, в силу $V_* \subset V'_*$ имеем

$$\|\sum \lambda_i\hat{\xi}_i\|_* \geqslant \|\sum \lambda_i\xi_i\|'_* > \frac{4\beta\varepsilon}{\beta} = 4\varepsilon,$$

если $\sum|\lambda_i| = 1$, что дает $k_{E, \|\cdot\|}(\varepsilon) \geqslant m$. Утверждение (4.20) доказано.

Доказательство п. 4.5.2.

1°. Пусть $p = \infty$. Тогда $k_{\mathfrak{l}_p^{(n)}}(t) = n$ при $t < {}^1/_4$, что непосредственно следует из рассмотрения функционалов $\langle \xi_i \mid x\rangle = x^i$, $1 \leqslant i \leqslant n$.

2°. Пусть $2 < p < \infty$. Тогда

$$k_{\mathfrak{l}_p^{(n)}}(t) \geqslant \frac{1}{2}\min\left\{n, \frac{4^{-p}}{t^p}\right\}.$$

В самом деле, достаточно проверить, что

$$k_{\mathfrak{l}_p^{(n)}}(t) \geqslant \frac{4^{-p}}{t^p}\cdot\frac{1}{2}$$

при $2 < t^{-p}\cdot 4^{-p} \leqslant n$. Действительно, пусть l есть наибольшее целое, меньшее $t^{-p}4^{-p}$. Тогда $l > {}^1/_2 t^{-p}4^{-p}$ и $l < n$. Рассмотрим

l координатных функционалов $\langle \xi_i \mid x\rangle = x^i$, $1 \leqslant i \leqslant l$ на $\mathfrak{l}_p^{(n)}$. Пусть $\sum\limits_1^l |\lambda_i| = 1$. Положим $\xi = \sum\limits_1^l \lambda_i\xi_i$. Тогда при $q = p/(p - 1) < 2$ имеем

$$\|\xi\|_q = \left(\sum_1^l |\lambda_i|^q\right)^{1/q} \geqslant \left(l\left(\frac{1}{l}\right)^q\right)^{1/q} = l^{-(q-1)/q} = l^{-1/p} > 4t,$$

т. е.

$$k_{\mathfrak{l}_p^{(n)}, \|\cdot\|_p}(t) \geqslant l > \frac{1}{2}(t^{-p}4^{-p}),$$

что и требуется.

3°. Пусть $1 \leqslant p \leqslant 2$. Тогда

$$k_{\mathfrak{l}_p^{(n)}}(t) \geqslant \frac{1}{4}\min\left\{n, \frac{1}{(4t)^2}\right\}.$$

Действительно, достаточно проверить, что

$$k_{\mathfrak{l}_p^{(n)}}(t) \geqslant \frac{1}{4}\left(\frac{1}{4t}\right)^2$$

при $4 < (1/4t)^2 \leqslant n$. Пусть s — максимальная степень двойки, такая, что $2^s < (1/4t)^2$. Тогда $2^s \geqslant {}^1/_4 (1/4t)^2$. Рассмотрим подпространство $R^{(s)} = \mathfrak{l}_q^{(n)}$, $q = p/(p - 1)$, состоящее из векторов, у которых все координаты, кроме первых 2^s, равны 0. Заметим, что в этом подпространстве найдется ортогональный базис из векторов, у которых первые 2^s координат равны ± 1. Действительно, можно считать $n = 2^s$. Для $s = 0$ такой базис существует. Если $\xi_1, \ldots, \xi_{2^s}$ — такой базис для данного s, то $\{\xi_1, \xi_1\}$, $\{\xi_1, -\xi_1\}, \ldots, \{\xi_{2^s}, \xi_{2^s}\}$, $\{\xi_{2^s}, -\xi_{2^s}\}$ — нужный базис для $s + 1$, так что доказываемое утверждение верно по индукции.

Пусть $\xi_1, \ldots, \xi_m$, $m = 2^s$, — соответствующий базис для $R^{(s)}$ и $\bar{\xi}_i = m^{-1/q}\xi_i$, $1 \leqslant i \leqslant m$, так что $\|\bar{\xi}_i\|_q = 1$. Пусть далее, $\sum\limits_1^m |\lambda_i| = 1$ и $\xi = \sum\limits_1^m \lambda_i\bar{\xi}_i$. Тогда

$$\|\xi\|_2^2 = \sum_1^m \lambda_i^2 \|\bar{\xi}_i\|_2^2 = m^{1-2/q}\sum_i^m \lambda_i^2 \geqslant m^{-2/q}.$$

Обозначая через θ_j модули координат ξ, получаем

$$\sum_1^m \theta_j^q = \sum_1^m (\theta_j^2)^{q/2} \geqslant m\left(\frac{\sum\theta_j^2}{m}\right)^{q/2} \geqslant m^{1+q(-1-2/q)/2} = m^{-q/2}$$

5.2. Оставляя на время в стороне доказательство предложения, выведем из него оценки (2.1) — (2.2) и А.5. Пусть $G \subset E$ и k натурально, причем $\alpha_{\|\cdot\|}(G) = \alpha$ и $\alpha_{\infty,k}(\|\cdot\|) = \beta$. Как и в п. 4.4, можно считать, что G содержится в шаре радиуса α и содержит единичный шар с центром в 0. Можно, далее, считать, что на некотором k-мерном подпространстве $E' \subset E$ введена l_∞-норма $\|\cdot\|_\infty$ так, что $\frac{1}{\beta}\|x\|_\infty \leqslant \|x\| \leqslant \|x\|_\infty$. Рассмотрим k функционалов $\bar\xi_1, \ldots, \bar\xi_k$ из $(E')^*$, являющихся координатными функционалами во введенной $l_\infty^{(k)}$-структуре на E'. Тогда $\|\bar\xi_i\|_{\infty,*} = 1$, так что $\bar\xi_i$, рассматриваемые как функционалы на E', $\|\cdot\|$, имеют норму, $\leqslant \beta$. Стало быть, $\bar\xi_i$ можно продолжать с E' на E до функционалов ξ_i нормы, $\leqslant \beta$. Поскольку сужения ξ_i на E' линейно-независимы, то и сами ξ_i, $i = 1, \ldots, k$, линейно-независимы. Пусть теперь $N^{\mathcal{O}}(\nu)$, $\widetilde N^{\mathcal{O}}(\nu)$ — детерминированная и стохастическая сложности класса задач, получающегося переоснащением класса C_{Lip} типа $C^0_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$ с помощью детерминированного локального оракула $\mathcal{O}$. Докажем, что

$$N^{\mathcal{O}}(\nu) \geqslant \varphi_{k,\infty}(\alpha\beta\nu) \tag{5.5}$$

и

$$\widetilde N^{\mathcal{O}}(\nu) \geqslant \frac{\varphi_{k,\infty}(8\alpha\beta\nu)}{1 + \left]\ln\right] \varphi_{k,\infty}(8\alpha\beta\nu)\,[}. \tag{5.6}$$

Разумеется, можно считать $m = 0$. Пусть $\mathcal{B}^M$ — детерминированный метод решения задач рассматриваемого класса с точностью ν и трудоемкостью M. Приведем к противоречию гипотезу $[M] < \varphi_{k,\infty}(\alpha\beta\nu)$ (это и докажет (5.5)). Пусть гипотеза верна. Применим к $\mathcal{O}$, $\bar\xi^n = (\xi_1, \ldots, \xi_n)$ и $\mathcal{B}^M$ первое из утверждений предложения 5.1. В соответствии с ним найдется $\bar f \in A(\bar\xi^n) \subset C_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$, для которой выполнено (5.2). Заметим, что для $f \in A(\bar\xi^n)$ имеем

$$\inf_{K(\xi^n)} f(x) = \inf_{K(\bar\xi^n)\cap E'} f(x) = \inf_{x\in E',\, \|x\|_\infty \leqslant 1} f(x) \geqslant \inf_{x \in E',\, \|x\| \leqslant 1} f(x) \geqslant \inf_G f(x),$$

т. е. и $\inf_G \bar f(x) < 0$. Но тогда из (5.2) следует $\varepsilon(\mathcal{B}^M, \bar f) > 4\alpha\beta\nu$. С другой стороны, из определения $A(\bar\xi^n)$ вытекает, что $\bar f$ липшицева с константой $\leqslant \max_i \|\xi_i\|_* \leqslant \beta$. Так как G лежит в $\|\cdot\|$-шаре радиуса α, то $\rho_{\|\cdot\|}(G) \leqslant 2\alpha$, т. е.

$$\nu(\mathcal{B}^M, \bar f) \geqslant \frac{\varepsilon(\mathcal{B}^M, \bar f)}{2\cdot 2\alpha\cdot\beta} > \nu,$$

что противоречит определению $\mathcal{B}^M$. Полученное противоречие доказывает (5.5).

Соотношение (5.6) доказывается дословно так же. При этом используется второе утверждение предложения 5.1 вместо первого, обосновывающего (5.5). Попутно мы видим, что, каков бы ни был детерминированный метод $\mathcal{B}^M$ конечной трудоемкости на рассматриваемом классе, он не в состоянии получить для каждой f из класса C_{Lip} с $\inf_G f < 0$ результат $\bar x$, для которого $f(\bar x) \leqslant 0$. Ясно отсюда, что при $m > 0$ никакой детерминированный метод конечной трудоемкости не в состоянии найти допустимый план у всякой совместной задачи класса. Это доказывает А.5.

Итак, мы выполнили намеченную для данного параграфа программу с точностью до доказательства предложения 5.1. К этому последнему мы и переходим.

5.3. Начнем с доказательства утверждения I предложения. Достаточно доказать его в предположении, что M натурально. При этом достаточно доказать, что утверждение I верно для

$$\varphi_{n,\infty}(\varepsilon) = \begin{cases} 1, & \varepsilon \geqslant 1/4, \\ \max\left\{n, \dfrac{n}{3}\ln\dfrac{1}{80\varepsilon}\right\}, & \varepsilon < 1/4. \end{cases} \tag{5.7}$$

1°. Утверждение I тривиально верно при $\varepsilon \geqslant 1/4$. Пусть теперь $\varepsilon < 1/4$. Убедимся, что (5.2) верно при $M < n$. Действительно, при всех $\omega_i = \pm 1$

$$\inf_{K(\xi^n)} \max_{1\leqslant i\leqslant n} \{\langle \omega_i \xi_i \,|\, x\rangle\} = -1$$

в силу линейной независимости ξ_i и определения $K(\xi^n)$. С помощью конструкции из доказательства предложения 4.3.1, примененной к набору $\xi_1, \ldots, \xi_n$, убеждаемся, что при $M < n$ искомая функция найдется уже среди функций вида

$$\max_{1\leqslant i\leqslant n} \{\langle \omega_i \xi_i \,|\, x\rangle + \delta_i\}, \quad \omega_i = \pm 1.$$

2°. Теперь достаточно доказать, что (5.2) верно при

$$M < \frac{n}{3}\ln\frac{1}{80\,\varepsilon},$$

если правая часть последнего неравенства больше n. Докажем даже, что оно верно при

$$M < \frac{n}{3\ln 2}\ln\frac{1}{80\,\varepsilon}. \tag{5.8}$$

3°. Рассмотрим некоторое семейство функций на оси. Пусть $\delta > 0$, $\delta < 10^{-4}$. Положим

$$\left.\begin{aligned} \gamma &= \gamma(\delta) = \frac{1}{4} - \delta, \\ c &= c(\delta) = \frac{3}{56} - \frac{3}{4}\delta, \\ q &= q(\delta) = \left(\frac{1}{2} - 7\delta\right)\Big/(1 - 4\delta). \end{aligned}\right\} \tag{5.9}$$

Пусть, далее, τ таково, что $0 < \tau < \delta$. Положим

$$\theta_\tau(t) = \begin{cases} -c + q\left|t + \frac{1}{4}\right|, & -\infty < t \leqslant -\delta, \\ -c + q\left|t + \frac{1}{4}\right| + \tau\,[t + \delta]_+, & t \geqslant -\delta. \end{cases}$$

Здесь $[t]_+ = \max\{0, t\}$. График функции $\theta_\tau(t)$ изображен на рис. 4.1. Непосредственной выкладкой проверяется, что

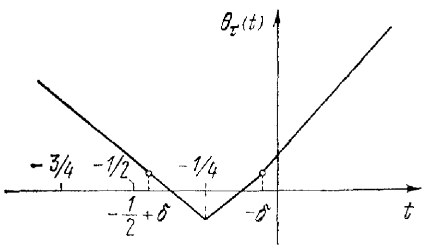

Рис. 4.1.

(i) $\theta_\tau(t)$ — выпуклая липшицева с константой 1 функция, имеющая изломы в точках $-{}^1/_4$, $-\delta$,

(ii) справедливы неравенства

$$\left.\begin{aligned} &\theta_\tau(t) < |t| - {}^3/_7, \quad |t| \geqslant 1, \\ &\theta^\tau(t) \geqslant |t| - {}^3/_7, \quad |t + {}^1/_4| \leqslant \gamma; \end{aligned}\right\} \tag{5.10}$$

(iii) $\min\limits_t \theta_\tau(t) = \theta_\tau(-{}^1/_4) = -c(\delta)$. В точках $(-{}^1/_2 + \delta)$ и $-\delta$ имеем $\theta_\tau(t) = {}^4/_3\, c(\delta)$;

(iv) если $0 \leqslant t$, $\tau' < \tau$, то $\theta_{\tau'}(t) < \theta_\tau(t)$ и $\theta_\tau(t) > \theta_{\tau'}(-t)$. Кроме того, $\theta_\tau(t)$ возрастает при $t \geqslant 0$.

4°. Теперь сформулируем и докажем лемму, из которой уже легко вытекает сформулированное предложение.

Лемма. *Пусть $\mathcal{O}$ — произвольный детерминированный локальный оракул для класса $A(\bar{\xi}^n)$, $0 < \delta < 10^{-4}$, и q, γ, c определены по* (5.9). *Пусть M — натуральное число и $\mathcal{B}^M$ — использующий $\mathcal{O}$ детерминированный метод минимизации функций из $A(\bar{\xi}^n)$. Тогда для каждого целого p, такого, что $0 \leqslant pn \leqslant M$, можно указать набор объектов K_p, f_p, $x_1, \ldots, x_{pn}$, состоящий из:*

— содержащегося в $K(\bar{\xi}^n)$ множества K_p вида $\{x \mid |\langle \xi_i \mid x\rangle - c_p^i| \leqslant \gamma^p\}$;

— функций $f_p \in A(\bar{\xi}^n)$, таких, что

$$f_p(x) - \left\{q^p \max_{1\leqslant i\leqslant n}\left|\langle \xi_i \mid x\rangle - c_p^i\right| - \frac{3}{7}\left(\frac{7c}{3}\right)^p\right\} \begin{cases} \geqslant 0 & \text{вне } K_p, \\ = 0 & \text{на } K_p; \end{cases} \tag{J_p}$$

— точек $x_1, \ldots, x_{pn}$, являющихся первыми pn точками реализации $\mathcal{B}^M$ на f_p и не лежащих в K_p.

Доказательство леммы проведем индукцией по p. Для $p = 0$ утверждение леммы тривиально верно: можно взять $K_p = K(\bar{\xi}^n)$ (т. е. $c_0^i \equiv 0$) и

$$f_0(x) = \max_{1\leqslant i\leqslant n}|\langle \xi_i \mid x_i\rangle| - \frac{3}{7}.$$

Пусть теперь утверждение леммы верно для $p = l$, $(l + 1)n \leqslant M$ и K_l; f_l; $x_1, \ldots, x_{ln}$ — соответствующий набор объектов. Построим набор, отвечающий $p = l + 1$. Для сокращения обозначений сопоставим точке $x \in E\,n$ ее «координат» $x^i = \langle \xi_i \mid x\rangle - c_p^i$ (это не настоящие координаты, соответствие $x \to (x^1, \ldots, x^n)$ не взаимно однозначно). Требуемый набор будет построен в результате n шагов, на j-м из которых будет иметься некоторые функции $f^1, \ldots, f^j$, числа $i(1), \ldots, i(j) \in I_0 = \{1, \ldots, n\}$, попарно различные, и множество $I_j = I_0 \setminus \{i(1), \ldots, i(j)\}$, а также числа $\omega_j = \pm 1$ и точки $y_1, \ldots, y_j \in E$.

Положим $f^0 = f_l$, и пусть τ_i выбраны так, что $\delta > \tau_1 > \ldots > \tau_n > 0$. Пусть уже сделаны s шагов построения и имеются объекты f^s, I_s. На $(s + 1)$-м шаге проводятся следующие операции.

1) Определяется точка y_{s+1} — как $(nl + s + 1)$-я точка работы $\mathcal{B}^M$ на f^s.

2) Среди координат y_{s+1}^i точки y_{s+1} с номерами $i \in I_s$ выбирается максимальная по модулю. Пусть $i(s + 1)$ — ее номер,

$$\omega_{s+1} = \begin{cases} -1, & y_{s+1}^{i(s+1)} < 0, \\ +1, & y_{s+1}^{i(s+1)} \geqslant 0. \end{cases}$$

Положим $I_{s+1} = I_s \setminus \{i(s + 1)\}$.

3) Определяется функция

$$f^{s+1}(x) = \max\{f^s(x), (\gamma q)^l\, \theta_{\tau_{s+1}}(\omega_{s+1} x^{i(s+1)}/\gamma^l)\}. \tag{5.11}$$

Шаг закончен. Отметим, что в силу (5.11)

$$f_l = f^0 \leqslant f^1 \leqslant \ldots \leqslant f^n. \tag{5.12}$$

После n шагов построения будет получена функция $f^n \equiv f_{l+1}$ и точки $y_1, \ldots, y_n$, а также числа $\omega_1, \ldots, \omega_n$ и подстановка $i(1), \ldots, i(n)$. Положим

$$K_{l+1} = \left\{x \,\middle|\, \left|\omega_j x^{i(j)} + \frac{\gamma^l}{4}\right| \leqslant \gamma^{l+1}, \quad 1 \leqslant j \leqslant n\right\}, \tag{5.13}$$

т. е. $c_{l+1}^i = -\omega_{j^{-1}(i)} \dfrac{\gamma^l}{4} + c_l^i$, $j^{-1}(i)$ — подстановка, обратная $i(j)$ и $x_{ln+i} = y_i$, $1 \leqslant i \leqslant n$.

Непосредственной проверкой, которую мы опустим, можно убедиться, что объекты K_{l+1}; f_{l+1}; $x_1, \ldots, x_{(l+1)n}$ — искомые. Лемма доказана.

Теперь уже легко доказать утверждение I предложения 5.1. Действительно, пусть M — натуральное число и $\mathcal{B}^M$ — метод минимизации функций из $A(\bar{\xi}^n)$, использующий оракул $\mathcal{O}$. Найдем минимальное $M' > M$, кратное n. Без ограничения общности можно считать $\mathcal{B}$ M'-шаговым методом, результат применения которого к любой задаче из $A(\bar{\xi}^n)$ есть последняя (M'-я) опрашиваемая точка. Применяя к методу $\mathcal{B}^{M'} = \mathcal{B}^M$ доказанную выше лемму при $p = M'/n$ и каком-нибудь δ, найдем функцию $f_p \in A(\bar{\xi}^n)$ и «куб» K_p, такие, что выполнено (J_p) и вся траектория $\mathcal{B}^{M'}$ на f_p лежит вне K_p. В силу (J_p) $f > 0$ вне K_p, так как на границе этого множества

$$f = (\gamma q)^p - \frac{3}{7}\left(\frac{7c}{3}\right)^p = \left(\frac{7c}{3}\right)^p - \frac{3}{7}\left(\frac{7c}{3}\right)^p > 0.$$

Вместе с тем в «центре» $\bar{x}$ множества K_p — в любой точке, в которой $\langle \xi_i | \bar{x}\rangle = c_p^i$, $1 \leqslant i \leqslant n$, имеем в силу того же (J_p) $f_p(\bar{x}) = -{}^3/_7\,(7c/3)^p$. Поскольку $M' \leqslant M + n$, то

$$f_p(\bar{x}) \leqslant -\frac{3}{7}\left(\frac{7c}{3}\right)^{M/n+1}.$$

Пусть теперь задано $\varepsilon > 0$ и M таково, что

$$\frac{3}{28\varepsilon} > 8^{\frac{M}{n}+1}. \tag{5.14}$$

Тогда и при достаточно малом $\delta > 0$ будет

$$\frac{3}{7}\left(\frac{7c(\delta)}{3}\right)^{\frac{M}{n}+1} < 4\varepsilon.$$

При таком δ предыдущее рассуждение доказывает существование $f \in A(\bar{\xi}^n)$, такой, что

$$\inf_{K(\bar{\xi}^n)} f(x) \leqslant -\frac{3}{7}\left(\frac{7c(\delta)}{3}\right)^{\frac{M}{n}+1} < -4\varepsilon$$

и $f > 0$ вдоль траектории работы $\mathcal{B}^M$ на f (с добавленным к ней результатом этой работы). Остается заметить, что (5.14) эквивалентно неравенству

$$\frac{M}{n} < \frac{1}{3\ln 2}\ln\frac{3}{224\varepsilon},$$

заведомо выполненному при

$$M < \frac{n}{3\ln 2}\ln\frac{1}{80\varepsilon},$$

что и требуется. Утверждение I предложения 5.1 доказано.

5.4. Докажем теперь утверждение II.

1°. Определим функцию $\theta(t)$ соотношением

$$\theta(t) = \begin{cases} |t| - \frac{1}{2} & |t| \geqslant \frac{1}{2}, \\ \frac{1}{8}t - \frac{1}{16}, & -\frac{1}{8} \leqslant t \leqslant \frac{1}{2}, \\ \frac{1}{16}\left|t + \frac{1}{4}\right| - \frac{11}{128}, & -\frac{3}{8} \leqslant t \leqslant -\frac{1}{8}, \\ -\frac{5}{8}\left(t + \frac{1}{2}\right), & -\frac{1}{2} \leqslant t \leqslant -\frac{3}{8}. \end{cases}$$

Тогда $\theta(t)$ — липшицева выпуклая с константой 1 функция на оси.

2°. Пусть для данного отрезка вещественной оси Δ, Δ^{+1} и Δ^{-1} суть отрезки длины $|\Delta|/4$ с центрами в серединах правой, соответственно левой половин Δ. Пусть N — фиксированное натуральное число и $T = \{\omega = (\omega_1, \ldots, \omega_N)\colon \omega_i = \pm 1\}$. Назовем *вложением* ω последовательность отрезков $\Delta_1(\omega), \ldots, \Delta_N(\omega)$, определяемую следующим рекуррентным способом:

$$\Delta_0(\omega) \equiv [-{}^1/_2,\ {}^1/_2], \quad \Delta_{i+1}(\omega) = \Delta^{\omega_{i+1}}(\Delta_i(\omega)).$$

Пусть еще $t_i(\omega)$ — середина отрезка $\Delta_i(\omega)$.

Обозначим через T^n множество $\{\bar{\omega} = (\omega^1, \ldots, \omega^n) \mid \omega^i \in T\}$. Пусть, далее, $\delta > 0$ фиксировано и R — натуральное число. Обозначим через X_R какое-нибудь R-элементное подмножество множества $(-\delta, 0)$ и через S_R — множество $n \times (N+1)$-мерных векторов, координаты которых лежат в X_R. Векторы из S_R будем обозначать $\tau = \{\tau_j^i,\ 0 \leqslant j \leqslant N,\ 1 \leqslant i \leqslant n\}$. Пусть x^i — функция, определенная как $\langle \xi_i \mid x\rangle$, $1 \leqslant i \leqslant n$.

Каждой паре $(\bar{\omega}, \tau) \in T^n \times S_R$ поставим в соответствие функции $f^j_{\bar{\omega},\tau}(x)$, $0 \leqslant j \leqslant N$, определяемые как

$$f^0_{\bar{\omega},\tau}(x) = \max_{1\leqslant i\leqslant n}\left\{|x^i| - \frac{1}{2} + \tau_0^i\right\} \equiv \max_{1\leqslant i\leqslant n}\{\theta^0_{+1}(x^i) + \tau_0^i\} \tag{5.15}$$

и

$$f^j_{\bar{\omega},\tau}(x) = \max_{1\leqslant i\leqslant n}\{\theta^j_{\omega^i_j}(x^i - t_{j-1}(\omega^i)) + \tau^i_j\}, \quad j \geqslant 1. \tag{5.16}$$

Здесь при $j \geqslant 1$

$$\theta^j_{+1}(t) \equiv \theta^j_{-1}(-t) \equiv$$

$$\equiv \frac{1}{8^{2(j-1)}\alpha_0 \ldots \alpha_{j-1}}\theta(-4^{j-1}t) + \sum_{s=0}^{j-2}\frac{8^{-2s}}{\alpha_0 \ldots \alpha_s}\theta\left(-\frac{1}{8}\right). \tag{5.17}$$

В этом выражении $\alpha_0 = 1 < \alpha_1 < \ldots$ — фиксированные числа, $\sum\limits_0^{-1} a_i \equiv 0$, $\prod\limits_1^0 b_i = 1$.

3°. Установим некоторые свойства функций $f^j_{\bar\omega,\tau}$. Пусть $\bar f^j_{\bar\omega}$ определены по тем же формулам, что и $f_{\bar\omega,\tau}$, но при $\tau = 0$. Тогда функции $\bar f^j_{\bar\omega}(x)$ таковы, что при $0 \leqslant j \leqslant N-1$

$$\bar f^{j+1}_{\bar\omega} - \bar f^j_{\bar\omega} \begin{cases} \geqslant 0 & \text{на } K_j(\bar\omega), \\ \leqslant 0 & \text{вне } K_j(\bar\omega). \end{cases} \tag{5.18}$$

Здесь $K_j(\bar\omega) = \{x \in E \mid x^i \in \Delta_j(\omega^i),\ 1 \leqslant i \leqslant n\}$. Неравенство в (5.18) вне $K_j(\bar\omega)$ является строгим при $j > 0$. Пусть, далее,

$$\bar f_{\bar\omega}(x) = \max_{0\leqslant j\leqslant N} \bar f^j_{\bar\omega}(x), \quad a = \min_{x\in K(\bar\xi^n)} \bar f_{\bar\omega}(x)$$

(как легко видеть, правая часть не зависит от $\bar\omega$),

$$\tilde f_{\bar\omega}(x) = \bar f_{\bar\omega}(x) - \frac{1}{2}a, \quad \tilde f^j_{\bar\omega}(x) = \bar f^j_{\bar\omega}(x) - \frac{1}{2}a.$$

Тогда

$$\tilde f_{\bar\omega}(x) = \tilde f^{j+1}_{\bar\omega}(x) \quad \text{в} \quad K_j(\bar\omega)\setminus K_{j+1}(\bar\omega), \quad 0 \leqslant j \leqslant N-1, \tag{5.19}$$

$$\left.\begin{aligned} &\min_{x\in K(\bar\xi^n)} \tilde f_{\bar\omega}(x) < 0, \\ &\tilde f_{\bar\omega}(x) > \frac{1}{4\cdot 64^N \alpha_0 \ldots \alpha_{N-1}} \quad \text{при } x \notin K_N(\bar\omega). \end{aligned}\right\} \tag{5.20}$$

Неравенства (5.18) — (5.20) доказываются непосредственной проверкой, исходя из определения функции $\theta(t)$. Эта работа предоставляется читателю.

4°. Обозначим $\mathcal{F}_R = \{(\bar\omega, \tau) \mid \bar\omega \in T^k,\ \tau \in S_R\}$. Снабдим $\mathcal{F}_R$ равновероятным распределением $p(\bar\omega, \tau) = r(\bar\omega)\, q_R(\tau)$. Для каждого $(\bar\omega, \tau) \in \mathcal{F}_R$ пусть

$$f_{\bar\omega,\tau}(x) = \max_{0\leqslant j\leqslant N} f^j_{\bar\omega,\tau}(x) - \frac{a}{2} \equiv \max_{0\leqslant j\leqslant N} \tilde f^j_{\bar\omega,\tau}(x).$$

Легко видеть, что $f_{\bar\omega,\tau}(x) \in A(\bar\xi^n)$.

Пусть теперь

$$\widetilde{\mathcal{B}}^M = \int_0^1 \mathcal{B}^t dt$$

— метод из формулировки предложения 5.1. Обозначим через $\bar\varepsilon$ величину $\sup\limits_{(\bar\omega,\tau)\in\bigcup_R \mathcal{F}_R} \tilde\varepsilon(\widetilde{\mathcal{B}}^M, f_{\bar\omega,\tau})$. Тогда при всяком R

$$\int_{\mathcal{F}_R} l(\widetilde{\mathcal{B}}^M, f_{\bar\omega,\tau})\, dr(\bar\omega)\, dq_R(\tau) \leqslant M$$

и

$$\int_{\mathcal{F}_R} \tilde\varepsilon(\widetilde{\mathcal{B}}^M, f_{\bar\omega,\tau})\, dr(\bar\omega)\, dq_R(\tau) \leqslant \bar\varepsilon. \tag{5.21}$$

Действуя, как и при доказательстве предложения 5.2 гл. I, мы извлекаем из (5.21) существование детерминированного правильного метода $\mathcal{B}^{t_R}$, для которого

$$\int_{\mathcal{F}_R} l(\mathcal{B}^{t_R}, f_{\bar\omega,\tau})\, dr(\bar\omega)\, dq_R(\tau) \leqslant 2M$$

и

$$\int_{\mathcal{F}_R} \tilde\varepsilon(\mathcal{B}^{t_R}, f_{\bar\omega,\tau})\, dr(\bar\omega)\, dq_R(\tau) \leqslant 2\bar\varepsilon. \tag{5.22}$$

5°. Пусть теперь $\bar{\mathcal{O}}$ — оракул для поля задач $\{f_{\bar\omega,\tau} \mid (\bar\omega, \tau) \in \bigcup_R \mathcal{F}_R\} \equiv \mathfrak{A}$, устроенный следующим образом. Ответ оракула $\bar{\mathcal{O}}$ на вопрос о задаче $f \in \mathfrak{A}$ в точке $x \in E$ есть четверка величин $\bar j = \bar j(\bar\omega, \tau, x)$, $\bar i = \bar i(\bar\omega, \tau, x)$, $\tau^{\bar i}_{\bar j}$, $\omega^{\bar i}_{\bar j}$, где $\bar j = \bar j(\bar\omega, \tau, x)$ есть наименьший среди номеров j, для которых $\tilde f_{\bar\omega,\tau}(x) = \tilde f^j_{\bar\omega,\tau}(x)$, а $\bar i(\bar\omega, \tau, x)$ есть наименьшее i, для которого

$$\tilde f^{\bar j}_{\bar\omega,\tau}(x) = -\frac{a}{2} + \theta^{\bar j}_{\omega^i_{\bar j}}(x^i - t_{\bar j-1}(\omega^i)) + \tau^i_{\bar j}.$$

Здесь принято $\omega^i_0 = +1$, $t_{-1}(\omega^i) = 0$.

Действуя по той же схеме, что и при доказательстве предложения 2 п.4.3, и используя (5.18) — (5.19), можно убедиться, что для данных M, $\widetilde{\mathcal{B}}^M$, N и отвечающего этим объектам $\bar\varepsilon$ можно взять $\delta > 0$ столь малым, а R столь большим, что для некоторого *детерминированного правильного использующего оракул* $\bar{\mathcal{O}}$ метода $\mathcal{B}_R$ будет выполняться следующее утверждение. Пусть

$$A(R) = \{(\bar\omega, \tau) \in \mathcal{F}_R \mid l(\mathcal{B}_R, f_{\bar\omega,\tau}) \leqslant l(\mathcal{B}^{t_R}, f_{\bar\omega,\tau}) < 8M$$

и

$$\tilde\varepsilon(\mathcal{B}^{t_R}, f_{\bar\omega,\tau}) = \tilde\varepsilon(\mathcal{B}_R, f_{\bar\omega,\tau}) < 4\bar\varepsilon\}.$$

Тогда

$$p(A(R)) > \frac{1}{8}. \tag{5.23}$$

Далее будем считать, что δ выбрано столь малым, а R столь большим, что сформулированное утверждение имеет место.

6°. В силу сказанного ясно, что можно найти такое $\bar\tau \in S_R$, что, обозначая для этого $\bar\tau$

$$A_{\bar\tau} = \{\bar\omega \mid l(\mathcal{B}_R, f_{\bar\omega,\tau}) < 8M \quad \text{и} \quad \tilde\varepsilon(\mathcal{B}_R, f_{\bar\omega,\tau}) < 4\bar\varepsilon\},$$

получим

$$r(A_{\bar\tau}) > \frac{1}{8}. \tag{5.24}$$

Из определения оракула $\bar{\mathscr{O}}$ ясно, что на задачах из поля $\mathscr{F}_R^{\bar\tau} = \{f_{\bar\omega, \tau} \mid \bar\omega \in T^n, \tau = \bar\tau\}$ он поставляет в точности столько же информации, что и оракул $\bar{\bar{\mathscr{O}}}$, сообщающий в точке x величины $\bar j = \bar j(\bar\omega, \bar\tau, x)$, $\bar i = \bar i(\bar\omega, \bar\tau, x)$, $\omega_{\bar j}^{\bar i}$. Поэтому при рассмотрении работы $\mathscr{B}_R$ на $A_{\bar\tau}$ можно считать, что $\mathscr{B}_R$ использует оракул $\bar{\bar{\mathscr{O}}}$. Но ответы $\bar{\bar{\mathscr{O}}}$ на вопрос о любой задаче в любой точке x суть элементы фиксированного множества с не более чем $2(N+1)n$ элементами. Отсюда следует, что различных результатов, доставляемых методом $\mathscr{B}_R$ при решении задач из $A_{\bar\tau}$ не позднее $[8M]$-го шага, не более $(2n(N+1))^{[8M]}$. С другой стороны, из (5.20) следует, что если δ достаточно мало и результат применения $\mathscr{B}_R$ к $f_{\bar\omega, \tau}$ не лежит в $K_N(\bar\omega)$, то

$$\tilde\varepsilon(\mathscr{B}_R, f_{\bar\omega, \tau}) > \frac{1}{4 \cdot 64^N \alpha_0 \ldots \alpha_{N-1}}.$$

Кроме того, множества $K_N(\bar\omega)$, отвечающие различным $\bar\omega$, попарно не пересекаются.

Из сказанного ясно, что если

$$4\bar\varepsilon \leqslant \frac{1}{4 \cdot 64^N \alpha_0 \ldots \alpha_{N-1}}, \tag{5.25}$$

то множество $A_{\bar\tau}$ не может содержать более $(2n(N+1))^{[8M]}$ элементов. Соединяя этот факт с (5.24), получаем, что в условиях (5.25) справедливо неравенство

$$(2n(N+1))^{[8M]} > \frac{1}{8} 2^{nN}, \tag{5.26}$$

т. е.

$$M > \frac{nN - 3}{8 \log_2 (2n(N+1))}. \tag{5.27}$$

7°. В частности, пусть задано $\varepsilon > 0$ и $\tilde{\mathscr{B}}^M$ — такой метод из формулировки утверждения предложения 5.3, для которого при всех $f \in A(\bar\xi^n)$ (5.4) не имеет места. Это значит, что для него $\bar\varepsilon \leqslant 4\varepsilon$, т. е. левая часть (5.25) не больше 16ε. Пусть вначале $\varepsilon \leqslant 10^{-3}$. Тогда, выбирая $\alpha_0, \ldots, \alpha_{N-1}$ достаточно близкими к 1, видим, что (5.25) будет выполнено при

$$N = \left(\left[\left(\ln \frac{1}{64\varepsilon}\right) \Big/ (\ln 64)\right] - 1\right),$$

что в силу (5.27) доставляет оценку

$$[M] \geqslant \frac{\varphi_{n,\infty}(8\varepsilon)}{1 + [\ln \varphi_{n,\infty}(8\varepsilon)]_+}$$

при надлежаще выбранном $d(\infty)$. Стало быть, в условиях (5.3) при этом $d(\infty)$ соотношение (5.4) обязано выполняться хотя бы для одного $f \in A(\bar\xi^n)$.

8°. Чтобы завершить доказательство утверждения II, надо еще проверить, что оно справедливо при $\varepsilon > 10^{-3}$. Ясно, что оно тривиально верно при $8\varepsilon \geqslant 1/4$, т. е. при $\varepsilon \geqslant 1/32$. Нам остается доказать теперь, что при $1/32 > \varepsilon > 10^{-3}$ и некоторой абсолютной константе $c > 0$ в условиях

$$M > c \frac{n}{\ln(n+1)}$$

соотношение (5.4) выполнено для некоторой $f \in A(\bar\xi^n)$. Это легко делается с помощью конструкции из доказательства предложения 2 п.4.3, примененной к набору $\xi_1, \ldots, \xi_n$. Предложение 5.3 доказано.

§ 6. Доказательство нижних оценок сложности. III

Нам осталось доказать утверждения А.2, А.3, А.4 теоремы 2.1. Последнее из них имеет чисто академический интерес, и мы не будем излагать его доказательства. Из вида функций $\varphi_{n,p}(\cdot)$ следует, что утверждение А.2 для значений p из интервала $1 < p < 2$ есть частный случай утверждения А.3. Поэтому достаточно доказать А.2 для $p \geqslant 2$ и А.3.

6.1. Начнем с доказательства А.2 для $p \geqslant 2$. Фактически требуется доказать следующее утверждение: *существует такая непрерывная справа принимающая натуральные значения неубывающая функция $k_p(t)$, что в условиях* А.2 *при $\alpha' > \alpha_{\|\cdot\|}(G)$ и $n \geqslant k_p(\alpha'\nu)$ справедливо неравенство*

$$\tilde N^{\mathscr{G}}(\nu) > c(p)(\alpha'\nu)^{-p}, \tag{6.1}$$

где $c(p) > 0$ зависит лишь от p, $p \geqslant 2$.

Для доказательства заметим, что если k натурально и $\dim E \geqslant k$ $(E, \|\cdot\| = L_p, \|\cdot\|_p)$, то найдутся k функционалов $\xi_1, \ldots, \xi_k \in E^*$ и k-мерное подпространство $E^k \subset E$, такие, что

$$\left.\begin{aligned}
&\|\xi_i\|_* \leqslant 2, && 1 \leqslant i \leqslant k,\\
&\Big\{\sum_{i=1}^k |\langle \xi_i \mid x\rangle|^p\Big\}^{1/p} \leqslant 2\|x\|, && x \in E,\\
&\min\Big\{\Big\|\sum_1^j \lambda_i \xi_i\Big\| \,\Big|\, \sum |\lambda_i| = 1\Big\} \geqslant j^{-1/p}, && 1 \leqslant j \leqslant k.
\end{aligned}\right\} \tag{6.2}$$

В интересах п. 6.8 заметим, что ниже мы получим нужную оценку сложности, отправляясь только от того обстоятельства, что для подходящих k на E существует система k функционалов, обладающих свойством (6.2). При этом в последующем изложении не используется явно условие о том, что E, $\|\cdot\|$ есть L_p, $\|\cdot\|_p$.

Пусть теперь $\xi_1, \ldots, \xi_k$ — система функционалов, обладающих свойствами (6.2), а μ — вероятностная мера, сосредоточенная на единичном шаре V пространства E. Тогда

$$2^p \geqslant \int\limits_V \sum_{i=1}^k |\langle \xi_i | x \rangle|^p d\mu(x).$$

Стало быть, при всяких $\delta, \varkappa > 0$ число функционалов ξ_i, для которых выполнено условие

$$\mu\{x \in V \mid |\langle \xi_i \mid x \rangle| \geqslant \delta\} \geqslant \varkappa, \tag{6.3}$$

не превосходит

$$N_p(\delta, \varkappa) = 2^p \delta^{-p} \varkappa^{-1}. \tag{6.4}$$

В частности, если $k > N_p(\delta, \varkappa)$, то не менее $k - N_p(\delta, \varkappa)$ из функционалов ξ_i таковы, что для множества $\Pi_\delta(\xi_i) = \{x \in V \mid |\langle \xi_i \mid x \rangle| \geqslant \delta\}$ имеем $\mu(\Pi_\delta(\xi_i)) < \varkappa$.

6.2. Пусть теперь r и N — натуральные числа, $2r \leqslant 2N \leqslant \dim E$. Тогда в рассуждениях п. 6.1 можно взять $k = 2N$. Рассмотрим всевозможные перестановки $\sigma(i)$ из $2N$ элементов и образуем функции

$$f^{(r)}_{\sigma,\gamma}(x) = \max_{1 \leqslant i \leqslant r} \{\langle \xi_{\sigma(i)} | x \rangle - i\gamma\}, \tag{6.5}$$

где $\gamma > 0$ удовлетворяет условию $r\gamma \leqslant 1$. Пусть еще

$$g^{(r)}_{\sigma,\gamma}(x) = \max\{f^{(r)}_{\sigma,\gamma}(x), 3\|x\| - 6\}. \tag{6.6}$$

Из (6.2) легко увидеть, что $g^{(r)}_{\sigma,\gamma}(x)$ обладает следующими свойствами:

(i) $g^{(r)}_{\sigma,\gamma}$ — выпуклая липшицева с константой 3 функция на E;

(ii) $g^{(r)}_{\sigma,\gamma}(x) = f^{(r)}_{\sigma,\gamma}(x)$ при $\|x\| \leqslant 1$;

(iii) $g^{(r)}_{\sigma,\gamma}(x) = 3\|x\| - 6 > g^{(r)}_{\sigma,\gamma}(0)$ при $\|x\| > 6$;

(iv) $\min\{g^{(r)}_{\sigma,\gamma}(x) \mid \|x\| \leqslant 1\} < -r^{-1/p}$.

6.3. Пусть теперь $\widetilde{\mathscr{B}}$ — какой-нибудь метод минимизации выпуклых функций на E, использующий локальный детерминированный оракул $\mathcal{O}$. Будем рассматривать работу $\widetilde{\mathscr{B}}$ на семействе функций $\mathfrak{A}$, образованном всевозможными функциями описанного выше вида. Можно считать, что $\widetilde{\mathscr{B}} = \int\limits_0^1 \mathscr{B}^t dt$ есть смесь правильных детерминированных методов. Обозначим

$$\tilde{\varepsilon}(x, g^{(r)}_{\sigma,\gamma}) = [g^{(r)}_{\sigma,\gamma}(x) - \inf\{g^{(r)}_{\sigma,\gamma}(x) \mid \|x\| \leqslant 1\}]_+,$$

и пусть $\tilde{\varepsilon}(\widetilde{\mathscr{B}}, g^{(r)}_{\sigma,\gamma})$ получается усреднением $\tilde{\varepsilon}(\tilde{x}, g^{(r)}_{\sigma,\gamma})$ по распределению результата $\tilde{x}$ применения $\widetilde{\mathscr{B}}$ к g. Пусть

$$\varepsilon = \sup\{\tilde{\varepsilon}(\widetilde{\mathscr{B}}, g^{(r)}_{\sigma,\gamma}) \mid g^{(r)}_{\sigma,\gamma} \in \mathfrak{A}\}$$

и

$$l = \sup\{l(\widetilde{\mathscr{B}}, g^{(r)}_{\sigma,\gamma}) \mid g^{(r)}_{\sigma,\gamma} \in \mathfrak{A}\}.$$

Покажем, что если ε мало, то l не может быть малым. Не ухудшая характеристик ε и l метода $\widetilde{\mathscr{B}}$ на множестве задач $\mathfrak{A}$, можно считать, что все методы $\mathscr{B}^t$ не задают вопросов (и не выдают результатов) вне шара $W = \{x \in E \mid \|x\| \leqslant 6\}$. Действительно, в силу (iii) все функции из $\mathfrak{A}$ вне W совпадают друг с другом и больше своих значений в точке $x = 0$.

Далее удобно считать, что траектории всех $\mathscr{B}^t$ на задачах из $\mathfrak{A}$ после момента выдачи результата состоят из повторений этого результата, а не из символов ϕ. Кроме того, можно, разумеется, считать, что результат всегда есть точка E, а не $*$. Таким образом, траектории $\mathscr{B}^t$ на задачах из $\mathfrak{A}$ суть последовательности точек W.

6.4. Пусть теперь $1/8 > \bar{\varepsilon} > 0$, а $r = r(\bar{\varepsilon})$ — максимальное целое, для которого $r^{-1/p} > 8\bar{\varepsilon}$. Положим $\gamma = \gamma(\bar{\varepsilon}) = \bar{\varepsilon}/2r$, $\delta = \delta(\bar{\varepsilon}) = \gamma(\bar{\varepsilon})/12$ и найдем натуральное $N = N_p(\bar{\varepsilon})$ как наименьшее целое, $\geqslant r$, и такое, что $N \geqslant N_p(\delta(\bar{\varepsilon}), 1/2r^3)$. Предположим, что $\dim E \geqslant 2N$ и рассмотрим следующую связанную с методом $\widetilde{\mathscr{B}}$ из предыдущего пункта конструкцию из r шагов. Пусть $\varkappa = 1/2r^3$ и $\xi_1, \ldots, \xi_{2N}$ — функционалы из п. 6.1, отвечающие выбору $k = 2N$.

Первый шаг конструкции состоит в следующем. Пусть $x_1(t)$ — первая точка работы $\mathscr{B}^t$, $\mu_1 = \mu^1$ — ее распределение (т. е. распределение значений $x_1(t)$, индуцированное мерой Лебега на $[0,1]$). Найдем l_1, $1 \leqslant l_1 \leqslant 2N$, так, чтобы было $\mu^1(\Pi_{6\delta}(\xi_{l_1})) < \varkappa$ (по выбору N это возможно). Определим σ^1 как перестановку, для которой $\sigma^1(1) = l_1$, и положим $f_1 = g^{(1)}_{\sigma^1,\gamma}$. Шаг закончен.

Пусть уже сделано $i < r$ шагов этой конструкции и в результате построены функции $x_1(t), \ldots, x_i(t)$ со значениями в W, функции $f_1, \ldots, f_i \in \mathfrak{A}$, перестановки $\sigma^1, \ldots, \sigma^i$ из $2N$ элементов и меры $\mu_1, \ldots, \mu_i$ на W, такие, что

(a) $x_j(t)$ есть j-я точка работы $\mathscr{B}^t$ на f_{j-1}, $2 \leqslant j \leqslant i$;

(b) μ_j есть распределение значений $x_j(t)$, $1 \leqslant j \leqslant i$;

(c) $\sigma^s(j) = \sigma^l(j)$ при всех j, s, l, таких, что $j \leqslant s \leqslant l \leqslant i$;

(d) $f_j = g^{(j)}_{\sigma^j,\gamma}$, $\quad j \leqslant i$,

(e) $\mu_j(\Pi_{6\delta}(\xi_{\sigma^l(l)})) \leqslant r\varkappa$ при всех j и l, таких, что $j \leqslant l \leqslant i$.

На $(i+1)$-м шаге делаются следующие построения. Определяется $x_{i+1}(t)$ в соответствии с (a), а затем μ_{i+1} в соответствии с (b). Строится мера

$$\mu^{i+1} = \frac{1}{i+1}\sum_{j=1}^{i+1}\mu_j.$$

Далее, среди номеров $s \in \{1, 2, \ldots, 2N\}$, отличных от i номеров $\sigma^i(1), \ldots, \sigma^i(i)$, находится номер l_{i+1}, такой, что

$$\mu^{i+1}(\Pi_{6\delta}(\xi_{l_{i+1}})) \leqslant \varkappa$$

(по выбору N такое l_{i+1} существует, см. заключительное замечание п. 6.1). Определяем σ^{i+1} как (любую) перестановку, для которой $\sigma^{i+1}(j) = \sigma^i(j)$, $1 \leqslant j \leqslant i$, и $\sigma^{i+1}(i+1) = l_{i+1}$. После этого находим f_{i+1} по (d). Ясно, что при этом (e) выполнится при всех $j \leqslant l \leqslant i+1$.

6.5. Пусть описанная конструкция проведена. Обозначим через T_i, $1 \leqslant i \leqslant r$, множество $\{t \mid x_i(t) \in \bigcup\limits_{r \geqslant j \geqslant i}\Pi_{6\delta}(\xi_{\sigma^j(j)})\}$. Оценим лебегову меру $\operatorname{mes} T_i$ множества T_i. Очевидно,

$$\operatorname{mes} T_i \leqslant \sum_{j=i}^{r}\operatorname{mes}\{t \mid x_i(t) \in \Pi_{6\delta}(\xi_{\sigma^j(j)})\} = \sum_{j=i}^{r}\mu_i(\Pi_{6\delta}(\xi_{\sigma^j(j)})) \leqslant r^2\varkappa$$

(мы использовали (e)). Стало быть, полагая $T = [0, 1] \setminus \bigcup\limits_{i=1}^{r} T_i$, получим $\operatorname{mes} T \geqslant 1 - r^3\varkappa \geqslant {}^1/_2$.

Убедимся, что при $t \in T$ $x_1(t), \ldots, x_r(t)$ суть первые r точек траектории $\mathscr{B}^t$ на $f \equiv f_r$. По лемме о неразличении (§ 5 гл. I) достаточно проверить, что $f_j = f_s$ в окрестности x_j при $r \geqslant s \geqslant j$. Используя (d), видим, что для последнего нужно лишь убедиться, что $\langle \xi_{\sigma^s(l)} \mid x_j(t)\rangle - l\gamma < \langle \xi_{\sigma^j(j)} \mid x_j(t)\rangle - j\gamma$ при $j < l \leqslant s$. Так как $\sigma^s(l) = \sigma^l(l)$ при $l \leqslant s$, то можно считать $s = l$. Поскольку

$$t \notin T_j, \text{ то } \langle \xi_{\sigma^l(l)} \mid x_j(t)\rangle < 6\delta \text{ и } \langle \xi_{\sigma^j(j)} \mid x_j(t)\rangle > -6\delta.$$

Поэтому доказываемое неравенство заведомо имеет место, если $\gamma \geqslant 12\delta$, что и в самом деле так.

6.6. Теперь мы в состоянии завершить доказательство (6.1). Пусть (в обозначениях п.п. 6.3—6.5) $r > 4l$. Тогда $x_r(t)$ есть результат применения $\mathscr{B}^t$ к f_r для всех $t \in T$, кроме, возможно,

множества $\tilde{T}$ значений t, меры $\leqslant {}^1/_4$, ибо по определению l

$$\int_0^1 l(\mathscr{B}^t, f_r)\,dt \leqslant l.$$

Но $\tilde{\varepsilon}(x_r(t), f_r) > 4\bar{\varepsilon}$ при $t \in T$. Действительно, в силу свойства (iv) п. 6.2 и правила выбора r уже

$$\inf_{\|x\| \leqslant 1} f_r(x) < -r^{-1\,p} < -8\bar{\varepsilon},$$

тогда как при $t \in T$, как мы видели выше,

$$\langle \xi_{\sigma^r(r)} \mid x_r(t)\rangle \geqslant -6\delta$$

и

$$\langle \xi_{\sigma^r(r)} \mid x_r(t)\rangle - r\gamma \leqslant f_r(x_r(t)).$$

Отсюда $f_r(x_r(t)) \geqslant -6\delta - r\gamma > -4\bar{\varepsilon}$, что и дает требуемую оценку $\tilde{\varepsilon}(x_r(t), f_r)$. Таким образом,

$$\tilde{\varepsilon} \equiv \int_0^1 \tilde{\varepsilon}(\mathscr{B}^t, f_r)\,dt > \operatorname{mes}(T \setminus \tilde{T}) \cdot 4\bar{\varepsilon} > \bar{\varepsilon}.$$

Мы получаем следующий результат. Пусть $\bar{\mathscr{B}}$ — метод из п. 6.3, а ε, l — его характеристики, введенные в этом пункте. Предположим далее, что

$$r = \left[\frac{9^{-p}}{\varepsilon^p}\right], \quad \varepsilon \leqslant {}^1/_9, \quad N = N_p(\delta(\varepsilon), \varkappa(\varepsilon))$$

и

$$\dim E \geqslant 2N \ (E, \|\cdot\| = L_p, \|\cdot\|_p).$$

Тогда справедливо неравенство

$$l \geqslant \frac{1}{4}\left[\frac{9^{-p}}{\varepsilon^p}\right]. \tag{6.7}$$

6.7. Пусть теперь G и α' — объекты из п. 6.1. Не ограничивая общности, можно считать, что рассматриваемый в A.2 класс $\boldsymbol{C}_{\text{Lip}}$ имеет $m = 0$, а G содержит единичный шар E и содержится в шаре радиуса α'. Задачи минимизации функций из $\mathfrak{A}$ можно рассматривать как задачи из $\boldsymbol{C}_{\text{Lip}}$. Если теперь $\nu > 0$ и $12\alpha'\nu \leqslant {}^1/_9$, а $\tilde{N}^{\mathscr{G}}(\nu) < l$, то для некоторого метода $\tilde{\mathscr{B}}$ решения задач из $\boldsymbol{C}_{\text{Lip}}$ имеем: $\nu(\tilde{\mathscr{B}}, f) \leqslant \nu$ и $l(\tilde{\mathscr{B}}, f) \leqslant l$ при всех $f \in \mathfrak{A}$. Но при $f \in \mathfrak{A}$ $\mathsf{L}_{\|\cdot\|_p}(f) \leqslant 3$, тогда как $\rho_{\|\cdot\|_p}(G) \leqslant 2\alpha'$. Далее, $\{x \mid \|x\|_p \leqslant 1\} \subset G$, так что

$$\varepsilon(\tilde{\mathscr{B}}, f) = \mathsf{L}_{\|\cdot\|_p}(f) \cdot 2\rho_{\|\cdot\|_p}(G)\,\nu(\tilde{\mathscr{B}}, f) \leqslant 12\alpha'\nu \quad \text{при } f \in \mathfrak{A}.$$

Отсюда и из утверждения п. 6.6 в силу соотношения (6.7) получаем

$$l \geqslant \frac{1}{4}\left[\left(\frac{1}{9\cdot 12\alpha'\nu}\right)^p\right],$$

т. е. и

$$\tilde{N}^{\mathfrak{S}}(\nu) \geqslant \frac{1}{4}\left[\left(\frac{1}{9\cdot 12\alpha'\nu}\right)^p\right].$$

Таким образом, (6.1) имеет место при $\alpha'\nu < 1/108$. Ясно, что (6.1) верно и при $\alpha'\nu \geqslant {}^1/_{108}$. Утверждение А.2 для $p \geqslant 1$ доказано.

6.8. Докажем утверждение А.3. Достаточно разобрать случай $m = 0$. Пусть $\alpha' > \alpha_{\|\cdot\|}(G)$. Можно считать, что G содержит шар радиуса 2 с центром в E и содержится в шаре радиуса $2\alpha'$. Положим $\tilde{k}(t) = k_2(t)$, где $k_2(\cdot)$ — функция из п. А.2. По теореме Дворецкого (см. упражнение 4 § 1) существует $n(k) < \infty$, такое, что если $\dim E \geqslant n(k)$, то $\alpha_{2,k}(\|\cdot\|) \leqslant 2$. Положим $\bar{k}(t) = n(\tilde{k}(t))$. Нам достаточно проверить, что если $\dim E \geqslant \bar{k}(\alpha'\nu)$, то

$$\tilde{N}^{\mathfrak{S}}(\nu) \geqslant \frac{c}{(\alpha'\nu)^2} \tag{6.8}$$

с некоторой абсолютной константой $c > 0$.

Действительно, пусть $\dim E > \bar{k}(\alpha'\nu)$. Тогда и $\dim E^* \geqslant \bar{k}(\alpha'\nu)$, так что $\alpha_{2,k}(\|\cdot\|_*) \leqslant 2$ при $k = \tilde{k}(\alpha'\nu)$. Стало быть, существуют k-мерное подпространство $L \in E^*$ и такая евклидова норма $\|\cdot\|_2$ на L, что при $\xi \in L$

$$\|\xi\|_2 \leqslant \|\xi\|_* \leqslant 2\|\xi\|_2. \tag{6.9}$$

Пусть $\xi_1, \ldots, \xi_k$ — элементы L, образующие ортонормированную систему относительно $\|\cdot\|_2$. Рассмотрим k-мерное евклидово пространство R^k, $|\cdot|$ и оператор $T: E \to R^k$, определяемый формулой $Tx = \{\langle\xi_1 \mid x\rangle, \ldots, \langle\xi_k \mid x\rangle\}$. Заметим, что при $x \in E$

$$|Tx| \leqslant 2\|x\|. \tag{6.10}$$

Это сразу следует из (6.9) и соотношения

$$|Tx| = \sup\left\{\left\langle\sum_{i=1}^k \varphi^i\xi_i \,\middle|\, x\right\rangle \,\middle|\, \sum_{i=1}^k (\varphi^i)^2 \leqslant 1\right\}.$$

Из (6.9) и (6.10) следует, что система функционалов $\xi_1, \ldots, \xi_k$ удовлетворяет (6.2) с $p = 2$. При этом $k \geqslant k_2(\alpha'\nu)$. Конструкция пп. 6.3—6.7, применяемая к этим функционалам, доставляет требуемую оценку $\tilde{N}^{\mathfrak{S}}(\nu)$ снизу (заметим, что в этой конструкции использовалось лишь существование такой системы функционалов для указанного k, а не условие «E, $\|\cdot\|$ совпадает с L_p, $\|\cdot\|_p$»). Утверждение А.3, а вместе с ним и нижние оценки сложности из теорем §§ 1—2 доказаны.

Глава V

ЗАДАЧИ ВЫПУКЛОГО СТОХАСТИЧЕСКОГО ПРОГРАММИРОВАНИЯ

В предыдущих главах были описаны методы первого порядка решения выпуклых задач, работающие с точной, или, во всяком случае, с мало (по сравнению с требуемой точностью решения) искаженной информацией о задачах.

Часто рассматривают более сложную ситуацию, когда точная информация о решаемой задаче недоступна и вместо нее приходится пользоваться информацией, искаженной случайной помехой. Такая ситуация может возникать, например, тогда, когда функционалы задачи содержательно представляют собой средние значения случайных величин, зависящих как от управлений (по которым проводится оптимизация), так и от неконтролируемых случайных факторов (по которым проводится усреднение). Если при этом информация о задаче, необходимая для оптимизации, получается в ходе работы со стохастической имитационной моделью рассматриваемого объекта, то ясно, что эта информация сама носит случайный характер. Можно представить себе и ситуации, когда в исходной задаче физически осмысленного шума нет, а помехи связаны с условиями получения нужной информации о задаче. Как бы то ни было, анализ численных методов решения задач стохастического программирования типа широко распространенной стохастической аппроксимации можно свести к изучению методов решения обычных — детерминированных — экстремальных задач; достаточно считать, что стохастичен лишь источник информации — оракул, к которому могут обращаться методы. Именно такой подход к численным методам стохастического программирования и принимается ниже.

Наша цель — научиться применять в рассматриваемой ситуации методы зеркального спуска. Оказывается, что это действительно возможно, причем в весьма широких предположениях ЗС-методы оказываются субоптимальными — существенно неулучшаемыми по трудоемкости.

В этой главе вводятся классы задач стохастического программирования (§ 1) и строятся ЗС-методы решения задач этих классов без ограничений (§ 2). Этим вопросам посвящена обширная литература. Укажем в этой связи на важную работу Ю. М. Ермольева [9]. В отличие от известных работ, мы рассматриваем более общую ситуацию, чем гильбертова, а именно — регулярную. Описываемые далее методы находятся в том же отношении к традиционным методам стохастической аппроксимации, что и ЗС-методы из главы III — к обычному градиентному спуску.

В следующей главе будет рассмотрен значительно менее изученный случай задач с ограничениями. Они будут редуцированы к играм. Развитие ЗС-методов применительно к проблеме аппроксимации седловых точек выпукло-вогнутых игр позволит получить практически законченные результаты о (выпуклых) условных стохастических задачах. Обе главы базируются (в части новых результатов) на работах авторов [21, 22].

§ 1. Классы выпуклых задач стохастического программирования

1.1. Начиная с этого параграфа, в этой и следующей главе рассматриваются задачи, определенные на выпуклых замкнутых ограниченных подмножествах вещественных банаховых пространств. Все эти пространства E, $\|\cdot\|$ предполагаются регулярными и сепарабельными. Предполагается, далее, что рассматриваемые пары E, $\|\cdot\|$ обладают следующим свойством: каково бы ни было выпуклое замкнутое ограниченное подмножество $G \subset E$, найдутся *борелевы* отображения $\pi_G(x): E \to G$ и $\mu_G(x): E \to E^*$, такие, что для всяких x $\pi_G(x)$ есть ближайшая к x точка G, а $\mu_G(x)$ — опорный функционал в точке x в к выпуклой функции $\rho_{\|\cdot\|}(x, G)$. Всегда считается, что $\mu_G(x) = 0$ при $x \in G$.

Указанные свойства E, $\|\cdot\|$ обеспечивают борелевость поисковых правил строящихся далее методов — обстоятельство существенное, ибо мы должны иметь возможность использовать вероятностные характеризации методов. С другой стороны, сформулированные свойства E, $\|\cdot\|$ фактически обеспечиваются сепарабельностью и рефлексивностью E (действительно, при этом условии нужные требования к E, $\|\cdot\|$ заведомо удовлетворяются, если $\|\cdot\|$ строго выпукла и гладка, т. е. единичная $\|\cdot\|$-сфера не содержит не сводящихся к точке отрезков, а единичный шар в каждой точке границы допускает единственную опорную гиперплоскость. В то же время указанные свойства нормы всегда можно обеспечить, должным образом (и притом как угодно мало) «пошевелив» исходную норму на E).

Упражнение 1. Докажите сформулированные выше утверждения.

1.2. Пусть E, $\|\cdot\|$ — банахово пространство, а G — выпуклое ограниченное замкнутое подмножество E. Рассмотрим поле задач $C_{\text{Lip}} \equiv C_{\text{Lip}}(G, E, \|\cdot\|, m)$ (см. § 2 гл. II). Наша цель — оснастить C_{Lip} «стохастическим оракулом первого порядка» и определить после этого классы задач выпуклого стохастического программирования. Традиционный для этого круга вопросов оракул есть стохастический оракул $\mathcal{O} = ((\Omega, \mathsf{F}_\omega); I; \psi(x, f, \omega))$ (см. § 3 гл. I), у которого информационное пространство I то же, что и у детерминированного оракула первого порядка, т. е. $I = \underbrace{(\mathbb{R}\times\ldots\times\mathbb{R})}_{m+1} \times \underbrace{(E^*\times\ldots\times E^*)}_{m+1}$. Иными словами, функция наблюдения $\psi(x, f, \omega)$ имеет следующую структуру:

$$\psi(x, f, \omega) = \{\psi_0^0(x, f, \omega), \ldots, \psi_m^0(x, f, \omega); \psi_0^1(x, f, \omega), \ldots, \psi_m^1(x, f, \omega)\},$$

где $\psi_j^0(x, f, \omega)$ — скалярные функции (интерпретируемые как наблюдения величин $f_j(x)$ при шуме оракула ω), а $\psi_j^1(x, f, \omega)$ принимают значения в E^* (и интерпретируются как наблюдения опорных (на G) функционалов $f_j'(x)$ к функциям f_j в точке $x \in G$). Оракул такого сорта будем называть *естественным*.

Ясно, что при таких общих гипотезах об оракуле никакой содержательной теории построить нельзя. Следующая традиционная гипотеза состоит в том, что ответы оракул несмещены:

$$\mathsf{M}_{\mathsf{F}_\omega}\psi_j^0(x, f, \omega) = f_j(x)$$
и
$$\mathsf{M}_{\mathsf{F}_\omega}\psi_j^1(x, f, \omega) \equiv f_j'(x) \in \partial_G f_j(x), \tag{1.1}$$

при всех $x \in G$ и $f \in C_{\text{Lip}}$. Как и всегда, M означает среднее по соответствующему распределению, указанному в нижнем индексе.

Но и одной несмещенности еще недостаточно. Необходимо должным образом ограничить интенсивность помех. Имеется много способов такого ограничения. Например, можно наложить ограничения сверху на вторые центральные моменты случайных величин $\psi_j^i(x, f, \cdot)$, $0 \leqslant j \leqslant m$. Это, однако, не слишком удобно: ситуация будет описываться большим числом параметров, к тому же абсолютные значения этих моментов ничего не говорят. Для разумной теории должны быть существенны их безразмерные отношения к «характерным» для задачи величинам.

Выберем следующий, по нашему мнению, самый простой, путь характеризации помех. Пусть $L > 0$ — некоторое число. Скажем, что «интенсивность» оракула на задаче f с $m = 0$ не превосходит L, если при всех $x \in G$

$$\left.\begin{aligned} &\mathsf{M}_{\mathsf{F}_\omega}\|\psi_j^1(x, f, \omega)\|_*^2 \leqslant \frac{L^2}{4\rho_{\|\cdot\|}^2(G)} \\ \text{и}\quad &\mathsf{M}_{\mathsf{F}_\omega}|\psi_j^0(x, f, \omega) - f(x)|^2 \leqslant L^2. \end{aligned}\right\} \tag{1.2}$$

Мы видим, что (1.2) налагает определенным образом согласованные ограничения на моменты наблюдений $f(x)$ и $f'(x)$. Именно такого типа ограничения оказываются в дальнейшем наиболее естественными.

Для общего случая ($m>0$) можно ограничить (своей константой) «интенсивность оракула» на каждой компоненте задачи. Удобно, однако, наложить не вполне «покомпонентные» ограничения:

$$\left.\begin{aligned} &\mathsf{M}_{F_\omega}\|\psi_j^1(x,f,\omega)\|_*^2 \leqslant \frac{L_j^2}{4\rho^2_{\|\cdot\|}(G)}, \quad 0\leqslant j\leqslant m,\ x\in G,\\ &\mathsf{M}_{F_\omega}\left(\max_{0\leqslant j\leqslant m}\left\{\frac{|\psi_j^0(x,f,\omega)-f_j(x)|^2}{L_j^2}\right\}\right)\leqslant 1. \end{aligned}\right\} \qquad (1.3)$$

При данном оракуле $\mathcal{O}$ и данных $L_0,\ldots,L_m>0$ соотношения (1.1) и (1.3) выделяют из C_{Lip} некоторое поле задач. Его удобно оснастить нормирующими множителями

$$r_0(f)=L_0,\quad r_j(f)=\max\{0, \min_{x\in G} f_j(x)\}+L_j,\quad 1\leqslant j\leqslant m.$$

Заметим, что для совместных f $r_j(f)=L_j$, $1\leqslant j\leqslant m$. Полученное нагруженное поле задач (§ 2 гл. I), снабженное оракулом $\mathcal{O}$, представляет собой некий класс задач. Классы такого типа мы и будем изучать. При этом параметры $L_0,\ldots,L_m$, задающие класс, будут считаться известными априори. Заметим, что в таком случае можно перейти от решения задач $f=(f_0,\ldots,f_m)$ к решению «нормированных» задач $\bar f=(f_0/L_0,\ldots,f_m/L_m)$ и добиться этой нормировкой соотношений $L_0=\ldots=L_m=1$. Таким образом, условия (1.2) выделяют класс, по существу не зависящий от параметров шума оракула.

В дальнейшем будет удобно рассматривать классы с $L_0=\ldots=L_m=L$. Как только что отмечалось, при принятом подходе общий случай сводится к этому (даже с $L=1$).

1.3. Приведенные наводящие соображения позволяют перейти к формальному определению подлежащих исследованию классов. Будем задавать их условиями типа (1.1), (1.3), обобщив последние в двух направлениях: от рассмотрения вторых моментов перейдем к рассмотрению произвольных и, кроме того, в (1.1) откажемся от требования несмещенности и разрешим определенную (небольшую) систематическую ошибку.

Пусть G, E, $\|\cdot\|$, C_{Lip} — те же объекты, что и в п. 1.2, а $\mathcal{O}=((\Omega,\mathsf{F}_\omega),\ I,\ \psi(\cdot))$ — естественный оракул для C_{Lip}. Пусть, далее, $\nu_0\geqslant 0$, $L>0$ и $r>1$. Определим поле задач $\widetilde{C}_{r,L}^{\mathcal{O},\nu_0}=\widetilde{C}_{r,L}^{\mathcal{O},\nu_0}(G,E,\|\cdot\|,m)$ как множество всех задач $f\in C_{\mathrm{Lip}}$, удовлетворяющих следующим требованиям А — В.

А. При всех $x\in G$ и j, $0\leqslant j\leqslant m$, аффинные функционалы

$$g_j(y)=\mathsf{M}_{\mathsf{F}_\omega}\psi_j^0(x,f,\omega)+\langle \mathsf{M}_{\mathsf{F}_\omega}\psi_j^1(x,f,\omega)\,|\,y-x\rangle$$

удовлетворяют условиям

$$f_j(y)\geqslant g_j(y),\ y\in G,\quad \text{и}\quad f_j(x)\leqslant g_j(x)+\nu_0 L.$$

Б. При всех j, $0\leqslant j\leqslant m$, и $x\in G$

$$\mathsf{M}_{\mathsf{F}_\omega}\|\psi_j^1(x,f,\omega)\|_*^r\leqslant\left(\frac{L}{2\rho_{\|\cdot\|}(G)}\right)^r.$$

В. При всех $x\in G$

$$\mathsf{M}_{\mathsf{F}_\omega}\|\psi^0(x,f,\omega)-f(x)\|_\infty^r\leqslant L^r.$$

Здесь

$$\psi^0(x,f,\omega)=(\psi_0^0(x,f,\omega),\ldots,\psi_m^0(x,f,\omega))$$

и $f(x)=(f_0(x),\ldots,f_m(x))$ рассматриваются как отображения в $l_\infty^{(m+1)}$. В качестве нормирующих множителей для поля $\widetilde{C}_{r,L}^{\mathcal{O},\nu_0}$ выберем

$$r_j(f)=\begin{cases}L, & j=0,\\ L+[\min_G f_j]_+, & j\geqslant 1.\end{cases}$$

Снабдим полученное нагруженное поле оракулом $\mathcal{O}$. Получим класс задач, который также обозначим $\widetilde{C}_{r,L}^{\mathcal{O},\nu_0}(G,E,\|\cdot\|,m)$.

Заметим, что, в отличие от ситуаций предыдущих глав, поля задач введенных классов явно зависят от оракула. Можно сказать, что раньше поле задач задавалось заранее и при описании методов предъявлялись соответствующие требования к оракулу. Теперь, наоборот, заранее задан оракул, а А — В — требования к задаче, при выполнении которых мы беремся ее решать с данным оракулом.

Важно заметить, что будет предложен один и тот же метод для решения задач всего семейства классов $\widetilde{C}_{r,L}^{\mathcal{O},\nu_0}(G,E,\|\cdot\|,m)$, получающихся из всевозможных $\mathcal{O}$ при данных G, m, L, ν_0, r, E, $\|\cdot\|$. Таким образом, мы не нуждаемся в априорном знании $\mathcal{O}$. Необходима лишь уверенность в том, что решаемая задача лежит в $C_{\mathrm{Lip}}(G,E,\|\cdot\|,m)$ и связана с используемым оракулом условиями А — В (конечно, G, E, $\|\cdot\|$, r, L, m, ν_0 должны быть известны априори).

1.4. З а м е ч а н и е. В важном частном случае $\nu_0=0$, $m=0$ к оракулу фактически можно не предъявлять требование В. Более точно, если $m=0$, $\nu_0=0$ и имеется функция f, удовлетворяющая по отношению к оракулу $\mathcal{O}$ свойствам А — Б, то из $\mathcal{O}$ легко получить оракул $\widetilde{\mathcal{O}}$, удовлетворяющий всем условиям А — В применительно к функции $\bar f(x)=f(x)-f(x_0)$, x_0 — центр G.

Ясно, что все равно, что именно минимизировать — $f(x)$ или $\tilde{f}(x)$.

Оракул $\overline{\mathcal{O}}$ строится следующим образом. Пусть требуется наблюдать опорный функционал и значение $\tilde{f}$ в точке $x \in G$. Обратимся для этого дважды к оракулу $\mathcal{O}$ — первый раз в точке x, а второй раз — в наудачу выбранной (по равномерной мере) точке ξ отрезка $[x_0, x]$. Пусть шумы $\mathcal{O}$ при этих обращениях будут ω и ω'. Величину $\psi_0^1(x, f, \omega)$ примем за доставляемое $\overline{\mathcal{O}}$ наблюдение $\tilde{f}'(x)$, а величину

$$\langle \psi_0^1(\varepsilon, f, \omega') \mid x - x_0 \rangle \equiv \overline{\psi}_0^0(x, f, \omega, \omega', \xi) —$$

за наблюдение $\tilde{f}(x)$. Один ответ оракула $\overline{\mathcal{O}}$ стоит, таким образом, двух обращений к $\mathcal{O}$.

Упражнение 2. Докажите, что

$$\mathsf{M}\psi_0^1(x, f, \omega) \in \partial_G \tilde{f}(x)$$

и

$$\mathsf{M}\overline{\psi}_0^0(x, f, \omega, \omega', \xi) = \tilde{f}(x).$$

При этом

$$\mathsf{M}\,|\overline{\psi}_0^0(x, f, \omega, \omega', \xi) - \tilde{f}(x)|^r \leqslant L^r.$$

Таким образом, $\tilde{f}$ по отношению к $\overline{\mathcal{O}}$ удовлетворяет всем условиям А — В.

1.5. Прежде, чем переходить к описанию методов, введем одну полезную в интересующем нас круге вопросов величину. Пусть V — $(E, \|\cdot\|)$-регулярная функция, $\omega_V(t)$ — ее модуль гладкости (§ 2 гл. III) и $r > 1$. Определим $A(r)$ как семейство всех распределений F_τ а на полуоси $\{\tau \geqslant 0\}$, таких, что

$$\int_0^\infty \tau^r d\mathsf{F}_\tau \leqslant 1.$$

Введем функцию

$$\omega_{V,r}(t) = \sup\left\{\int_0^\infty \tau\omega_V(t\tau)\,d\mathsf{F}_\tau \,\middle|\, \mathsf{F}_\tau \in A(r)\right\}.$$

Функция $\omega_{V,r}(t)$ обладает теми же качественными свойствами, что и $\omega_V(t)$: она непрерывна и ограничена при $t \geqslant 0$, $\omega_{V,r}(0) = 0$, $\omega_{V,r}(t)$ не убывает. Можно определить обратную к $\omega_{V,r}(t)$ функцию $\gamma_{V,r}(s)$:

$$\gamma_{V,r}(s) = \sup\{t \mid \omega_{V,r}(t) \leqslant s\}.$$

Легко видеть, что $\gamma_{V,r}(s)$ — неубывающая положительная (при $s > 0$) функция s, принимающая и значение $+\infty$, причем $\omega_{V,r}(\gamma_{V,r}(s)) \leqslant s$ для всех $s > 0$ (здесь $\omega_{V,r}(+\infty) = \lim\limits_{t\to\infty} \omega_{V,r}(t)$).

Упражнение 3. Докажите сформулированные выше утверждения. ◇

Роль функции $\omega_{V,r}(t)$ в последующем основана на следующей паре утверждений.

Предложение. Пусть E, $\|\cdot\|$, V, r, $\omega_{V,r}(t)$ — описанные выше объекты, $(\Omega, \mathsf{F}_\omega)$ — польское пространство с регулярной борелевой вероятностной полной по Лебегу мерой F_ω. Пусть, далее, ω_i, $i = 1, 2, \ldots, s$, — независимые распределенные по F_ω случайные величины, $\omega^i = (\omega_1, \ldots, \omega_i)$, а $\xi_i(\omega^i)\colon \Omega \to E^*$ — борелевы функции, такие, что

$$\mathsf{M}_{\mathsf{F}_{\omega^i}} \|\xi_i(\omega^i)\|_*^r \leqslant L^r, \quad L < \infty,$$

где F_{ω^i} — совместное распределение $(\omega_1, \ldots, \omega_i)$, индуцированное F_ω. Тогда

(1) при всех $\varphi \in E^*$ и $x \in E$ для функции $V_*(\xi) = V(\xi) - \langle \xi \mid x \rangle$ справедливо неравенство

$$V_*(\varphi + \xi_1(\omega_1)) \leqslant V_*(\varphi) + \langle \xi_1(\omega_1) \mid V_*'(\varphi) \rangle + r(\omega_1), \tag{1.4}$$

где $r(\omega_1) \geqslant 0$ определяется по ξ_1 и φ, причем $\mathsf{M}_{\mathsf{F}_\omega} r(\omega_1) \leqslant L\omega_{V,r}(L)$;

(2) если $\mathsf{M}_{\mathsf{F}_{\omega_i}} \xi_i(\omega^i) = 0$ при всех ω^{i-1}, $1 \leqslant i \leqslant N$, а $\rho_i \geqslant 0$, то

$$\mathsf{M}_{\mathsf{F}_{\omega^N}} \left\| \sum_1^N \rho_i \xi_i(\omega^i) \right\|_* \leqslant L \sum_{i=1}^N \rho_i \omega_{V,r}(L\rho_i) + \frac{1}{2}. \tag{1.5}$$

Упражнение 4. Докажите предложение

⟨(1). В силу утверждения предложения 2.4 гл. III первое неравенство в (1.4) верно при выборе $r(\omega_1) = \|\xi_1(\omega_1)\|_* \omega_V(\|\xi_1(\omega_1)\|_*)$.

Полагая $\overline{\xi}_1(\omega_1) = \xi_1(\omega_1)/L$, получим, обозначая через F_τ распределение величины $\|\overline{\xi}_1(\omega_1)\|_*$: $\mathsf{F}_\tau \in A(r)$ и

$$\int r(\omega_1)\,d\mathsf{F}_{\omega_1} = \int L\tau\omega_V(L\tau)\,d\mathsf{F}_\tau \leqslant L\omega_{V,r}(L),$$

что и требуется вторым неравенством в (1.4). Утверждение (1) доказано.

(2) Оценим $\mathsf{M}\mathsf{F}_{\omega^i} V(s_i(\omega^i))$, где

$$s_i(\omega^i) = \sum_1^i \rho_j \xi_j(\omega^j), \quad s_0(\omega^0) = 0.$$

В силу (1.4) при $0 \leqslant i < N$

$$V(s_{i+1}(\omega^{i+1})) \leqslant V(s_i(\omega^i)) + \rho_{i+1} \langle \xi_{i+1}(\omega^{i+1} \mid V'(s_i(\omega^i)) \rangle + r_i(\omega^{i+1}).$$

Усредняя обе части этого неравенства вначале по ω_{i+1}, а затем — по ω^i и используя $\mathsf{M}_{\mathsf{F}\omega_{i+1}} \xi_{i+1}(\omega^{i+1}) = 0$ и (1.4), получим

$$\mathsf{M}_{\mathsf{F}_{\omega^{i+1}}} V(s_{i+1}(\omega^{i+1})) \leqslant \mathsf{M}_{\mathsf{F}_{\omega^i}} V(s_i(\omega^i)) + \rho_{i+1} L\omega_{V,r}(\rho_{i+1}L).$$

Отсюда

$$M_{F_{\omega^N}} V(s_N(\omega^N)) \leqslant \sum_{i=1}^{N} \rho_i L \omega_{V,r}(\rho_i L).$$

Чтобы получить из этого неравенства (1.5), достаточно заметить, что по определению V имеем $\|\varphi\|_* \leqslant V(\varphi) + {}^1/_2$ для всех $\varphi \in E^*$.⟩

1.6. В заключение приведем оценки величин $\gamma_{V,r}(\cdot)$ для стандартных L_p-регулярных функций. Ниже $c(p, r) > 0$ — величины, зависящие лишь от r и p, $r > 1$.

Для $1 < p < \infty$, $V = V_p$ (см. § 2 гл. III):

$$\gamma_{V_p, r}(s) \equiv \gamma_{p, r}(s) \geqslant c(p, r)\, s^{\frac{1}{\bar r - 1}}, \text{ где } \bar r = \min\left\{2, r, \frac{p}{p-1}\right\}. \tag{1.6}$$

Для $p = 1$, $V = V_{1,n}$ (см. § 2 гл. III):

$$\gamma_{V_{1,n}; r}(s) \equiv \gamma_{1,n;r}(s) \geqslant \frac{c(1, r)\, s^{\frac{1}{\bar r - 1}}}{\ln(n+1)}, \tag{1.7}$$

где $\bar r = \min\{2, r\}$. Заметим, что при $r = 2$ приведенные оценки функций $\gamma_{p,r}$ и $\gamma_{1,n,r}$ по порядку зависимости от ν (ν и n) совпадают с аналогичными оценками для γ_p и $\gamma_{1,n}$ из § 2 гл. III. Для $p > 2$ это обстоятельство имеет место и при $r \geqslant p/(p-1)$.

Упражнение 5. Используя оценки γ_p, $\gamma_{1,n}$ из § 2 гл. III, докажите (1.6) — (1.7). Указание. При выводе (1.7) примите во внимание, что $\sup_{t \geqslant 0} \omega_{V,n}(t) \leqslant a < \infty$, a не зависит от n, см. § 5 гл. III.

Теперь мы подготовлены к построению ЗС-методов решения задач стохастического программирования.

§ 2. Оптимизация без ограничений

В этом параграфе описываются методы решения задач классов $\widetilde{C}^{\circledcirc, r}_{L, \nu_0}(G, E, \|\cdot\|, 0)$, т. е. задач вида

$$f(x) \to \min \mid x \in G \subset E.$$

Для решения этих задач, оказывается, можно приспособить метод зеркального спуска, в основном аналогичный ЗС-методу решения липшицевых выпуклых задач при точном оракуле первого порядка (§ 3 гл. III). Пусть $V(\cdot)$ — $(E, \|\cdot\|)$-регулярная функция. Ассоциированный с ней метод решения задач указанного класса будем обозначать $\widetilde{\text{ЗС}}_V$. Класс $\widetilde{C}^{\circledcirc, r}_{L, \nu_0}(G, E, \|\cdot\|, 0)$ для краткости обозначим $\widetilde{C}$. При описании метода будем, как обычно, считать, что G нормировано условием: 0 есть центр G, $\rho_{\|\cdot\|}(G) = 1$. Ясно, что такой нормировки можно достичь простыми преобразованиями подобия.

2.1. Начнем с описания «сходящейся», «бесконечношаговой» версии метода. Работа метода определяется выбором последовательности смещений $\{\rho_i\}_{i=0}^{\infty}$, $\rho_i > 0$, и состоит в построении последовательностей $\{\varphi_i \in E^*\}_{i=0}^{\infty}$ и $\{x_i \in E\}_{i=1}^{\infty}$, $\{\bar x_i \in G\}_{i=1}^{\infty}$. Эти последовательности строятся по следующим рекуррентным правилам, в которых $f \in \widetilde{C}$ — решаемая задача, $\psi(x, f, \omega) = (\psi_0^0(x, f, \omega), \psi_0^1(x, f, \omega))$ — функция наблюдения оракула $\mathcal{O}$, ω_i — шум оракула на i-м шаге, $\omega^i = (\omega_1, \ldots, \omega_i)$:

$$\varphi_0 = 0, \tag{2.1}$$

$$x_i \equiv x_i(\omega^{i-1}) = V'(\varphi_{i-1}(\omega^{i-1})); \quad \bar x_i \equiv \bar x_i(\omega^{i-1}) = \pi_G(x_i), \tag{2.2}$$

$$\varphi_i \equiv \varphi_i(\omega^i) = \varphi_{i-1}(\omega^{i-1}) - \rho_i\{\psi_0^1(\bar x_i, f, \omega_i) + \|\psi_0^1(x_i, f, \omega_i)\|_* \mu_G(x_i)\}. \tag{2.3}$$

Если бы оракул $\mathcal{O}$ был точным детерминированным оракулом первого порядка, а смещения ρ_i удовлетворяли бы естественным условиям $\rho_i \underset{i\to\infty}{\to} 0$, $\sum_i \rho_i = +\infty$, то траектория $\{\bar x_i\}$ сходилась бы к точке минимума f на G в следующем смысле. Пусть x^i есть «лучшая» (по значению f) среди точек $x_1, \ldots, x_i$. Тогда $f(x^i) \underset{i\to\infty}{\to} \inf_G f$ (ср. с правилами выдачи результатов методов глав II, III). В стохастическом случае, как легко видеть, верно то же самое (разумеется, после усреднения по случайным реализациям методов). Таким образом, правило выдачи результата после i шагов типа указанного выше в состоянии обеспечить сходимость метода, тогда как сама траектория $\{\bar x_i\}$ без дополнительных предположений может и не сходиться (даже в среднем по функционалу). Однако мы не в состоянии реализовать такое правило. Действительно, выбор «лучшей» среди точек $x_1, \ldots, x_i$ предполагает точное знание чисел $f(x_1), \ldots, f(x_i)$, а мы ими не располагаем.

Оказывается (и это соображение лежит в основе метода), что указанное затруднение легко преодолеть: траектория $\{\bar x_i\}_{i=1}^{\infty}$ всегда сходится к оптимуму f (в среднем по функционалу) в *чезаровском* смысле. Рассмотрим траекторию выпуклых комбинаций $\{x^i\}$ точек $\bar x_j$ вида

$$x^i \equiv x^i(\omega^{i-1}) = \Big(\sum_{j=1}^{i} \rho_j\Big)^{-1} \sum_{j=1}^{i} \rho_j \bar x_j(\omega^{j-1}) \tag{2.4}$$

(справа ω^{j-1} — начальные фрагменты ω^{i-1}). Оказывается, что при сформулированных выше требованиях к ρ_i траектория x^i сходится к точке оптимума f (в среднем по функционалу), причем скорость сходимости можно явно оценить.

Теорема. *Пусть смещения* ρ_i *удовлетворяют условиям* $\rho_i > 0$, $\sum_i \rho_i = \infty$, $\lim_{i\to\infty} \rho_i = 0$ *и* $f \in \widetilde{C}$. *Тогда для траекторий процесса*

(2.1) — (2.4) *справедливо соотношение*

$$\bar{\nu}_i(f) \equiv \mathsf{M}_{\mathsf{F}_{\omega^{i-1}}}\nu(x^i(\omega^{i-1}), f) \leqslant \nu_0 + \frac{1}{2L\sum_{j=1}^{i}\rho_j} + \frac{\sum_{j=1}^{i}\rho_j\omega_{V,r}(L\rho_j)}{\sum_{j=1}^{i}\rho_j}. \tag{2.5}$$

В частности, $\overline{\lim_{i\to\infty}}\,\bar{\nu}_i \leqslant \nu_0$.

Доказательство. Ясно, что при $i\to\infty$ правая часть (2.5) стремится к ν_0, так что второе утверждение теоремы сразу следует из первого, к доказательству которого мы и переходим. Пусть x^* есть точка минимума f на G. Положим $V_*(\varphi) = V(\varphi) - \langle\varphi \mid x^*\rangle$ и проследим за изменением V_* вдоль траектории $\{\varphi_i\}$. Положим

$$\Delta_i(\omega^i) = \psi_0^1(\bar{x}_i(\omega^{i-1}), f, \omega_i) + \|\psi_0^1(\bar{x}_i(\omega^{i-1}), f, \omega_i)\|_*\,\mu_G(x_i(\omega^{i-1}))$$

и

$$\delta_i(\omega^{i-1}) = \mathsf{M}_{\mathsf{F}_{\omega_i}}\Delta_i(\omega^i) \equiv \psi_f(\bar{x}_i(\omega^{i-1})) + b_i(\omega^{i-1})\,\mu_G(x_i(\omega^{i-1})),$$

где

$$\psi_f(x) = \mathsf{M}_{\mathsf{F}_\omega}\psi_0^1(x, f, \omega)$$

и

$$b_i(\omega^{i-1}) = \mathsf{M}_{\mathsf{F}_{\omega_i}}\|\psi_0^1(\bar{x}_i(\omega^{i-1}), f, \omega_i)\|_*.$$

Ясно, что

$$b_i(\omega^{i-1}) \geqslant \|\psi_f(\bar{x}_i(\psi^{i-1}))\|_* \tag{2.6}$$

и

$$\mathsf{M}_{\mathsf{F}_{\omega_i}}\|\Delta_i(\omega^i)\|_*^r \leqslant \frac{2^r L^r}{[2\rho_{\|\cdot\|}(G)]^r} = L^r \tag{2.7}$$

(в последней оценке следует использовать свойство Б оракула $\mathcal{O}$). Имеем в силу (1.4)

$$V_*(\varphi_j(\omega^j)) \leqslant V_*(\varphi_{j-1}(\omega^{j-1})) + \rho_j\langle\Delta_i(\omega^i) \mid -V_*'(\varphi_{j-1}(\omega^{j-1}))\rangle + r_j(\omega^j),$$

или

$$V_*(\varphi_j(\omega^j)) \leqslant V_*(\varphi_{j-1}(\omega^{j-1})) + \rho_j\langle\Delta_i(\omega^i) \mid x^* - x_j(\omega^{j-1})\rangle + r_j(\omega^j), \tag{2.8}$$

где в силу (2.7)

$$\mathsf{M}_{\mathsf{F}_{\omega_j}} r_j(\omega^j) \leqslant L\rho_j\omega_{V,r}(L\rho_j). \tag{2.9}$$

Усредняя обе части (2.8) по ω_j, получим

$$\mathsf{M}_{\mathsf{F}_{\omega_j}}V_*(\varphi_j(\omega^j)) \leqslant V_*(\varphi_{j-1}(\omega^{j-1})) + \rho_j\langle\psi_f(\bar{x}_j(\omega^{j-1})) + b_j(\omega^{j-1})\mu_G(x_j(\omega^{j-1})) \mid x^* - x_j(\omega^{j-1})\rangle + L\rho_j\omega_{V,r}(L\rho_j).$$

В силу (2.6) правая часть последнего неравенства не больше

$$V_*(\varphi_{j-1}) + \rho_j\langle\psi_f(\bar{x}_j) \mid x^* - \bar{x}_j\rangle + L\rho_j\omega_{V,r}(L\rho_j)$$

(ср. с аналогичным моментом в п. 3.5 гл. III). Таким образом,

$$\mathsf{M}_{\mathsf{F}_{\omega_j}}V_*(\varphi_j(\omega^j)) \leqslant V_*(\varphi_{j-1}(\omega^{j-1})) + \rho_j\langle\psi_f(\bar{x}_j(\omega^{j-1})) \mid x^* - \bar{x}_j(\omega^{j-1})\rangle + L\rho_j\omega_{V,r}(L\rho_j). \tag{2.10}$$

Средний член суммы в правой части (2.10) в силу свойства А оракула $\mathcal{O}$ не превосходит $f(x^*) - f(\bar{x}_j(\omega^{j-1})) + \nu_0 L$, т. е. (2.10) дает

$$\mathsf{M}_{\mathsf{F}_{\omega_j}}V_*(\varphi_j(\omega^j)) \leqslant V_*(\varphi_{j-1}(\omega^{j-1})) + \rho_j\nu_0 L + \rho_j(f(x^*) - f(\bar{x}_j(\omega^{j-1}))) + \rho_j L\omega_{V,r}(\rho_j, L). \tag{2.11}$$

Усредняя обе части (2.11) по ω^{j-1}, получим

$$\mathsf{M}_{\mathsf{F}_{\omega^j}}V_*(\varphi_j(\omega^j)) \leqslant \mathsf{M}_{\mathsf{F}_{\omega^{j-1}}}V_*(\varphi_{j-1}(\omega^{j-1})) + \rho_j\nu_0 L + \rho_j(f(x^*) - \mathsf{M}_{\mathsf{F}_{\omega^{j-1}}}f(\bar{x}_j(\omega^{j-1}))) + \rho_j L\omega_{V,r}(\rho_j L). \tag{2.12}$$

Суммируя (2.12) по $j = 1, 2, \ldots, i$ и учитывая, что $V_*(\varphi_0(\omega^0)) = 0$, найдем

$$\sum_{j=1}^{i}\rho_j\{\mathsf{M}_{\mathsf{F}_{\omega^{j-1}}}f(\bar{x}_j(\omega^{j-1}))\} - \Big(\sum_{j=1}^{i}\rho_i\Big)f(x^*) \leqslant \Big(\sum_{j=1}^{i}\rho_j\Big)\nu_0 L + \sum_{j=1}^{i}\rho_j L\omega_{V,r}(\rho_j L) - \mathsf{M}_{\mathsf{F}_{\omega^i}}V_*(\varphi_i(\omega^i)). \tag{2.13}$$

Далее, ввиду выпуклости f, имеем

$$\sum_{j=1}^{i}\rho_j f(\bar{x}_j(\omega^{j-1})) \geqslant \Big(\sum_{j=1}^{i}\rho_j\Big)f(x^i(\omega^{i-1})),$$

а в силу свойств V

$$V_*(\varphi) = V(\varphi) - \langle\varphi \mid x^*\rangle \geqslant V(\varphi) - \|\varphi\|_* \geqslant -\frac{1}{2}$$

(мы учли, что $\|x^*\| \leqslant 1$ вследствие условий нормировки на G). Используя эти неравенства в (2.13), получим

$$\Big(\sum_{j=1}^{i}\rho_j\Big)(\mathsf{M}_{\mathsf{F}_{\omega^{i-1}}}(f(x^i(\omega^{i-1})) - f(x^*))) \leqslant \Big(\sum_{j=1}^{i}\rho_j\Big)\nu_0 L + \frac{1}{2} + \sum_{j=1}^{i}\rho_j L\omega_{V,r}(\rho_j, L), \tag{2.14}$$

что и дает требуемое неравенство (2.5). Теорема доказана.

2.2. Обсудим полученный результат.

2.2.1. Прежде всего заметим, что в гильбертовом случае (E, $\|\cdot\|$ — гильбертово пространство) при $r \geqslant 2$ описанный метод, отвечающий $V(\cdot) = V_2(\cdot)$, см. п. 2.3 гл. III, можно несколько упростить (подобно тому, как это делалось в § 3 гл. III). После упрощения метод приобретает следующий вид ($\varphi_{i-1} = x_i = \bar{x}_i$):

$$x_1 = 0; \quad x_{i+1} = \pi_G(x_i - \rho_i \psi_0^1(x_i, f, \omega_i)), \tag{2.15}$$

$$x^i = \Big(\sum_{j=1}^{i} \rho_j\Big)^{-1} \sum_{j=1}^{i} \rho_j x_j. \tag{2.16}$$

При этом оценка (2.5) может быть заменена оценкой

$$\bar{\nu}_i \leqslant \nu_0 + \frac{1}{2L\sum_{j=1}^{i}\rho_j} + \frac{\sum_{j=1}^{i} L\rho_j^2}{8\sum_{j=1}^{i}\rho_j}. \tag{2.17}$$

Все перечисленные факты доказываются так же, как и в общем случае. При этом надлежит рассмотреть изменение функции $\bar{V}_*(x) = {}^1/_2 (x - x^*)^2$.

Упражнение 1. Докажите (2.17).

Заметим, что уравнения (2.15) описывают хорошо известный процесс стохастической аппроксимации Ю. М. Ермольева. Таким образом, в гильбертовом случае предлагаемый метод отличается от традиционного лишь правилом формирования результата (2.16). Это отличие, однако, весьма существенно с точки зрения анализа метода. Для самой траектории (2.16) сходимость (в том или ином вероятностном смысле) и оценки типа (2.17), насколько известно авторам, удается доказать лишь при дополнительных предположениях относительно смещений ρ_i и функций f. Мы видим, что небольшое (и с практической точки зрения ничего не стоящее) изменение метода приводит к существенному упрощению ситуации.

2.2.2. Обсудим вопрос о разумном выборе смещений ρ_i. Пусть функция $\omega_{V,r}(t)$ допускает оценку $\omega_{V,r}(t) \leqslant ct^\sigma$, $\sigma \leqslant 1$. Нас будут интересовать качественные результаты, так что примем $L = 1$. Ограничимся рассмотрением правил выбора смещений вида $\rho_i \sim i^{-x}$, $0 < x \leqslant 1$. Из (2.5) легко извлечь, что следует выбирать $x = 1/(1+\sigma)$. При этом правая часть (2.5) будет убывать с ростом i как $\nu_0 + i^{-\sigma/(1+\sigma)} \ln i$. В простейшем — гильбертовом — случае для $r \geqslant 2$ получаем правило выбора смещений вида $\rho_i \sim 1/\sqrt{i}$. При этом правая часть (2.5) ведет себя как $\nu_0 + O(\ln i/\sqrt{i})$. Заметим, что можно несколько уточнить правило выбора смещений, если отказаться от смещений вида $O(i^{-x})$: например, брать

$$\rho_i \sim i^{-\frac{1}{1+\sigma}} (\ln i)^{-\frac{1}{1+\sigma}}.$$

Это дает оценку правой части (2.5) вида

$$\nu_0 + O(i^{-\frac{\sigma}{1+\sigma}} (\ln i)^{\frac{1}{1+\sigma}} \ln\ln i).$$

Мы не пойдем по этому пути дальше, имея в виду, что чуть позже будут рассмотрены «конструктивные» — конечношаговые версии описанного метода, для которых подобных осложнений с выбором шага не возникает.

2.2.3. Скажем еще несколько слов о рекомендуемом правиле выбора шага. Ограничимся простейшим случаем: E, $\|\cdot\|$ гильбертово, $r = 2$, $\nu_0 = 0$. В этом случае мы рекомендуем выбирать шаги $\rho_i = O(1/\sqrt{i})$, тогда как «традиционные» правила выбора шага $\rho_i = O(1/i)$. С нашей точки зрения второе правило крайне плохое — для него правая часть (2.5) ведет себя как $O(1/\ln i)$, т. е. крайне медленно сходится к 0 при $i \to \infty$.

Происхождение правила $\rho_i = O(1/i)$ можно объяснить следующим образом. Чем больше смещения ρ_i, тем, грубо говоря, быстрее сходился бы метод при точной информации, по крайней мере, вначале. С другой стороны, большие смещения приводят к сильному влиянию помех на траекторию, так что желательно делать смещения малыми. Компромисс между этими двумя тенденциями традиционно устанавливался в предположении, что функция f «хорошая» — гладкая и строго выпуклая. С точки зрения предельного поведения траекторий (традиционно служащего показателем работы метода) такое предположение о f в достаточной мере оправдано. Вместе с тем при таком подходе к делу действительно целесообразен выбор $\rho_i = O(1/i)$, ибо здесь как раз помехи и определяют скорость сходимости.

Рассмотрим модельный пример: минимизацию функции на оси $f_a(x) = {}^1/_2 (x-a)^2$. Предполагается, что в i-й момент времени доступно наблюдение $a_i = a + \omega_i$ величины a, где помехи ω_i скалярны, независимы от шага к шагу и, скажем, нормальны со средним 0 и дисперсией 1. С обсуждаемой точки зрения оракул сообщает реализации случайной (несмещенной) оценки

$$\psi_0^1(x, f_a, \omega) = x - a - \omega,$$

где $\omega \in \mathbb{R}$ распределено указанным выше образом. Процесс (2.15) (считаем, в виде исключения $G = \mathbb{R}$) имеет в данном случае вид

$$x_{i+1} = x_i - \rho_i (x_i - a - \omega_i),$$

или, полагая $\Delta_i = x_i - a$,

$$\Delta_{i+1} = \Delta_i - \rho_i\Delta_i + \rho_i\omega_i = (1 - \rho_i)\Delta_i + \rho_i\omega_i.$$

Отсюда при не ограничивающем общности условии $\rho_1 = 1$ следует, что Δ_i, $i \geqslant 2$, есть комбинация случайных величин $\omega_1, \ldots, \omega_i$

с суммой коэффициентов 1, так что минимальная дисперсия Δ_i (т. е. среднее значение $2\,(f_a(x_i) - \min\limits_x f_a(x))$) отвечает случаю, когда все коэффициенты этой комбинации равны между собой. В свою очередь, последнее эквивалентно условию $\rho_i = 1/i$, $i \geqslant 1$.

Таким образом, выбор $\rho_i = 1/i$ действительно оптимален для рассматриваемого примера. Сам же этот пример можно с известным основанием считать подходящей моделью для «далекой» от начала части траектории метода стохастической аппроксимации и в общем случае. В то же время это правило оказывается неприемлемо плохим на начальных участках траектории, если не предполагать функцию f гладкой и строго выпуклой.

Действительно, пусть f_ε — выпуклая функция на оси, минимум которой достигается при $x = 0$, и $f_\varepsilon(x) = 4\varepsilon x$ при $x > 0$. Предположим, что никаких помех нет, и мы начинаем движение из точки $x_1 = 1$, причем смещения имеют вид $\rho_i = 1/i$. После i шагов мы придем в точку

$$x_i = 1 - 4\varepsilon \sum_{j=1}^{i-1} \rho_j \geqslant 1 - 4\varepsilon \ln i.$$

Обеспечение (абсолютной) точности ε требует попадания в точку x с $|x| \leqslant 1/4$, т. е. число шагов N_ε, необходимое для этого, удовлетворяет оценке $4\varepsilon \ln N_\varepsilon \geqslant 3/4$, т. е. $N_\varepsilon \geqslant e^{3/16\varepsilon}$. При малых ε получаемое число шагов фантастически велико. Дело здесь в том, что в рассмотренном выше модельном примере за счет выбора $\rho_1 = 1$ мы сразу получаем точку x_1 с $\mathsf{M}x_1 = a$ (в случае точной информации сразу попадаем в точку минимума f), и все дальнейшее поведение траектории представляет собой вызванные помехами колебания относительно этой точки. Ничего похожего на эти эффекты квадратичной ситуации, как показывает второй пример, в общем случае нет.

Рекомендуемое здесь правило выбора смещений $\rho_i = 1/\sqrt{i}$ обеспечивает несравненно более быстрый, чем «классическое», выход траектории x_i в точку оптимума (на описанной выше функции f_ε мы достигаем его за время $O(1/\varepsilon^2)$ вместо $\exp(O(1/\varepsilon))$). Зато этот метод сильно «чувствует» помехи; из-за них траектория x_i сильно колеблется относительно оптимума. Тут вступает в силу вторая особенность предлагаемого метода, которой мы в этом обсуждении еще не касались — то, что мы усредняем эту колеблющуюся траекторию (переходим от x_i к x^i в соответствии с (2.16)) и считаем результирующей именно эту траекторию средних. Усреднение, проводимое не в ходе поиска (грубо говоря, именно то, что делают шаги $\rho_i = 1/i$), а независимо от него, позволяет одновременно и приемлемо быстро выходить в окрестность оптимума независимо от того, гладка ли минимизируемая функция или нет, и сгладить флуктуации траектории, связанные с наличием помех.

2.3. Перейдем теперь к «конструктивным» — конечношаговым версиям метода, считая, что заранее задана (относительная) точность ν, $\nu_0 < \nu < 1$, которая требуется от решения.

В этой ситуации можно рассмотреть метод $\widetilde{\text{ЗС}}_V(\nu)$, состоящий в построении начального фрагмента длины N траектории (2.1) — (2.3) и выдающего в качестве результата N-ю точку траектории (2.4). Необходимо лишь должным образом выбрать смещения ρ_i, $1 \leqslant i \leqslant N$, и само число N. С точки зрения оптимизации трудоемкости следует взять все ρ_i равными друг другу в соответствии с соотношением

$$\rho_i = \frac{1}{L}\,\rho_{V,r}(\nu - \nu_0),$$

где

$$\rho_{V,r}(\nu) = \gamma_{V,r}\left(\frac{\nu}{2}\right). \tag{2.18}$$

При этом число шагов следует выбирать по формуле

$$N \equiv N_{V,r}(\nu - \nu_0),$$

где

$$N_{V,r}(\nu) = \left]\frac{1}{\nu\rho_{V,r}(\nu)}\right[. \tag{2.19}$$

При таком выборе параметров ρ и N правая часть (2.5) для $i = N$ не превосходит ν. Таким образом, построенный метод обеспечивает точность ν в решении задач класса C при трудоемкости $N_{V,r}(\nu - \nu_0)$ (а не $N_{V,r}(\cdot) + 1$; дело в том, что построение x^N требует не N, а $N - 1$ вопроса о решаемой задаче).

В гильбертовом случае при $r = 2$ вместо (2.1) — (2.4) можно использовать (2.15) — (2.17) при выборе параметров в виде

$$\rho_i = \frac{4}{L(\nu - \nu_0)}, \quad N = \left]\frac{1}{4l(\nu - \nu_0)^2}\right[.$$

2.4. Посмотрим, какие оценки трудоемкости может доставить метод $\widetilde{\text{ЗС}}_V(\nu)$. Пусть, как и в п. 2.2.2, $\omega_{V,r}(t) \leqslant ct^\sigma$. Тогда

$$N_{V,r}(\nu) = O\left\{\left(\frac{1}{\nu}\right)^{\frac{\sigma}{1+\sigma}}\right\},$$

т. е. при настройке метода на заранее заданную точность последнюю можно обеспечить немного быстрее $\Big($за $O\left(\frac{1}{\nu - \nu_0}\right)^{\frac{\sigma}{1+\sigma}}$ вместо $O\left(\left(\frac{1}{\nu - \nu_0}\right)^{\frac{\sigma}{1+\sigma}} \ln\left(\frac{1}{\nu - \nu_0}\right)\right)$ шагов$\Big)$, чем при рациональном выборе ρ_i в «бесконечношаговой» версии п. 2. 1.

Перечислим оценки, отвечающие методам $\widetilde{\text{ЗС}}^{p,r}$, ассоциированным с функциями $V_p(\cdot)$ на пространствах $(E, \|\cdot\|) = (L_p(T, \mu),$

$\|\cdot\|_p$). Приводимые ниже оценки трудоемкости непосредственно следуют из (2.18) — (2.19) и оценок $\gamma_{V_{p,r}}(\cdot)$ из (1.6) — (1.7).

Пусть вначале $\infty > p > 1$. Тогда трудоемкость $\widetilde{\text{ЗС}}^{p,r}(\nu)$ не превосходит — при $1 > \nu > \nu_0$ — величины $M_{p,r}(\nu - \nu_0)$, где

$$\left.\begin{aligned} M_{p,r}(\nu) &= D(p,r)\left(\frac{1}{\nu}\right)^{\frac{\bar r}{\bar r - 1}}, \\ \bar r &= \min\left\{2, r, \frac{p}{(p-1)}\right\}. \end{aligned}\right\} \qquad (2.20)$$

Если же $p = 1$ и $(E, \|\cdot\|) = (l_1^{(n)}, \|\cdot\|_1)$, $1 < n < \infty$, то трудоемкость метода $\widetilde{\text{ЗС}}^{1,n;r}$, ассоциированного с $V = V_{1,n}(\cdot)$, при $1 > \nu > \nu_0$ не превосходит величины $M_{1,n;r}(\nu - \nu_0)$, где

$$M_{1,n;r}(\nu) = D(1,r)\left(\frac{1}{\nu}\right)^{\frac{\bar r}{\bar r - 1}} \ln n, \quad \bar r = \min\{2, r\}. \qquad (2.21)$$

В (2.20) — (2.21) $D(p, r)$ — конечная при всех $r > 1$ функция своих аргументов.

Заметим, что при $r \geqslant 2$ (фактически даже и при $r \geqslant \min\{2, p/(p-1)\}$) оценка (2.20) по характеру зависимости от ν (а (2.21) — по характеру зависимости и от ν, и от n) совпадает с аналогичной оценкой трудоемкости ЗС-метода решения липшицевых выпуклых задач при точном оракуле первого порядка (см. п. 3.4 гл. III). Таким образом, в наиболее типичном случае $r = 2$ трудоемкость предъявленных методов стохастического программирования, грубо говоря, такая же, как и у ЗС-методов обычного выпуклого программирования. Иными словами, ЗС-методы почти не реагируют на шумы в поставляемой информации.

§ 3. Сложность задач стохастического программирования

В этом параграфе строятся нижние оценки сложности введенных в § 1 классов задач стохастического программирования. Эти оценки, с одной стороны, позволяют судить о том, в какой мере эффективны уже построенные методы, и, с другой стороны, показывают, к чему следует стремиться в общем случае — при наличии ограничений, и значения, и производные которых наблюдаются с помехами.

Переходя к построению нижних оценок сложности интересующих нас классов, заметим следующее. В главе IV были построены оценки, действующие для любых (локальных детерминированных или даже любых локальных) оракулов. При рассмотрении стохастических оракулов таких универсальных результатов

получить нельзя. Действительно, рассматриваемый оракул может случайно оказаться и точным детерминированным, так что «универсальная» *нижняя* оценка сложности необходимо должна игнорировать стохастичность оракула. В соответствии со сказанным мы будем, грубо говоря, интересоваться верхней гранью сложностных функций данного поля задач, взятой по всевозможным его оснащениям оракулами данного типа.

3.1. Начнем с рассмотрения задач без ограничений. Здесь имеет место следующий простой результат.

Теорема. *Существует функция $c(r) > 0$, обладающая следующим свойством. Каковы бы ни были $r > 1$, банахово пространство E, $\|\cdot\|$ и ограниченное выпуклое замкнутое не сводящееся к точке множество $G \subset E$, существует естественный (в смысле § 1) оракул $\mathcal{O}$ для поля задач $C_{\text{Lip}}(G, E, \|\cdot\|, 0)$, такой, что стохастическая сложность $\widetilde N(\nu)$ класса задач $\widetilde C = \widetilde C^{\mathcal{O},0}_{r,L}(G, E, \|\cdot\|, 0)$ удовлетворяет оценке*

$$\widetilde N(\nu) \geqslant c(r)\left(\frac{1}{\nu}\right)^{\frac{\bar r}{\bar r - 1}},$$

где

$$\bar r = \min\{2, r\}\text{*)}. \qquad (3.1)$$

Доказательство. Не ограничивая общности, можно считать $L = 1$. Можно далее считать, что $\rho_{\|\cdot\|}(G) = 1$ и что G содержит пару точек $h, -h$ с $\|h\| > 1/4$. Тогда G содержит 0 и содержится в шаре радиуса 2 с центром в 0. Пусть $\varphi \in E^*$ таково, что $\langle \varphi \mid h\rangle = \|h\|$ и $\|\varphi\|_* = 1$. Для всех t, $|t| \leqslant 1/4$, примем, что f_t — задача, порожденная функцией $f_t(x) = t\langle \varphi \mid x\rangle$.

1°. Пусть вначале $r \leqslant 2$. Определим оракул $\mathcal{O}$ следующим образом. На всех задачах, отличных от задач семейства $\{f_t\}_{|t| \leqslant 1/4}$, $\mathcal{O}$ есть точный оракул первого порядка. Ответ $\mathcal{O}$ на вопрос о задаче f_t в точке $x \in G$ есть случайная величина со значениями в $\mathbb{R} \times E^*$, принимающая значение (0, 0) с вероятностью $1 - p(r, t)$ и значение $(p(r,t))^{-1}(f_t(x), f'_t(x))$ — с вероятностью $p(r, t)$. Здесь

$$p(r, t) = (4|t|)^{\frac{r}{r-1}}. \qquad (3.2)$$

Непосредственной выкладкой проверяется, что при $|t| \leqslant 1/4$ $f_t \in \widetilde C \equiv \widetilde C^{\mathcal{O},0}_{r,1}(G, E, \|\cdot\|, 0)$. Пусть теперь $\widetilde{\mathcal{B}}$ — произвольный метод решения задач класса $\widetilde C$ (средних) точности ν и трудоемкости M на этом классе. Рассмотрим подкласс $\widetilde C_t$ класса $\widetilde C$, образованный парой задач $f_{+|t|}$ и $f_{-|t|}$, $|t| \leqslant 1/4$, и снабдим $\widetilde C_t$ равновероят-

*) Указанная оценка «набегает» уже на подклассе простейших линейных задач.

ным распределением. По теореме 5.2 гл. I существует *детерминированный* метод $\mathcal{B}_t$ решения задач класса $\widetilde{C}$, такой, что

$$l(\mathcal{B}_t, f_{+|t|}) + l(\mathcal{B}_t, f_{-|t|}) \leqslant 4M$$

и

$$\nu(\mathcal{B}_t, f_{+|t|}) + \nu(\mathcal{B}_t, f_{-|t|}) \leqslant 4\nu. \tag{3.3}$$

Посмотрим, что может делать метод $\mathcal{B}_t$ в случае, если на первых $]4M[= M_1$ шагах его работы он все время получает нулевую информацию. Заметим, что при решении каждой из задач $f_{+|t|}$, $f_{-|t|}$ последнее происходит с вероятностью $p_M = (1 - p(r, t))^{M_1}$.

Возможно, что при такой информации $\mathcal{B}_t$ выдает результат $\bar{x}$ не позднее M_1-го шага. Очевидно, либо для $f = f_{+|t|}$, либо для $f = f_{-|t|}$ имеем $\nu(\bar{x}, f) > {}^1/_4\,|t|$ (важно заметить, что $\bar{x}$ по определению не зависит от решаемой задачи). В силу второго неравенства из (3.3) получаем ${}^1/_4\,|t|\,p_M < 4\nu$, т. е.

$$p_M < \frac{16\nu}{|t|}. \tag{3.4}$$

Возможно, далее, что $\mathcal{B}_t$ не выдает результата на первых M_1 шагах работы метода при получении все время нулевой информации. Тогда и $l(\mathcal{B}_t, f_{+|t|})$, и $l(\mathcal{B}_t, f_{-|t|})$ больше $p_M M_1$, т. е. из первого неравенства в (3.3) $p_M]4M[< 2M$ или $p_M < {}^1/_2$. Выберем теперь $t = 32\nu$ и будем считать, что $\nu \leqslant {}^1/_{256}$. Тогда из проведенных рассуждений следует, что

$$p_M \equiv (1 - p(r, t))^{]4M[} < {}^1/_2,$$

что сразу дает

$$]4M[> \frac{c}{p(r, t)},$$

$c > 0$ — абсолютная константа. Вспоминая определение $p(r, t)$ и учитывая, что всегда $M \geqslant 1$, получаем оценку

$$M \geqslant c(r)\left(\frac{1}{\nu}\right)^{\frac{r}{r-1}}$$

с некоторым (зависящим лишь от r) $c(r) > 0$. Так как $\widetilde{\mathcal{B}}$ был произвольным методом решения задач $\widetilde{C}$ с точностью ν, то

$$\widetilde{N}(\nu) \geqslant c(r)\left(\frac{1}{\nu}\right)^{\frac{r}{r-1}},$$

что и требуется.

2°. Пусть теперь $r \geqslant 2$. Изменим построенный выше оракул $\mathcal{O}$ на задачах f_t, $|t| \leqslant {}^1/_4$, следующим образом. Ответ $\mathcal{O}$ на вопрос о f_t в точке $x \in G$ есть с вероятностью ${}^1/_2 + 2t$ ответ точного оракула первого порядка на вопрос в точке x о задаче $f_{1/4}$ и с дополнительной вероятностью — ответ того же оракула на вопрос о задаче $f_{-1/4}$. Легко видеть, что при таком определении $\mathcal{O}$ при всех t, $|t| \leqslant {}^1/_4$, имеем $f_t \in \widetilde{C} \equiv \widetilde{C}^{\mathcal{O},0}_{r,1}(G, E, \|\cdot\|, 0)$.

Пусть теперь объекты $\widetilde{\mathcal{B}}$, ν, M, $\widetilde{C}_t$, $\mathcal{B}_t$ определены так же, как и в п. 1°, $M_1 =]16M[$, $\overline{\Omega}$ — пространство шумов оракула на первых $]16M[$ шагах работы метода $\mathcal{B}_t$, $\widehat{\Omega} = \overline{\Omega} \times \widetilde{C}_t = \{(\overline{\omega}, \varkappa),\ \overline{\omega} \in \overline{\Omega},\ \varkappa = \pm 1\}$; $\widehat{\Omega}$ рассматривается как произведение вероятностных пространств. Здесь $\varkappa = +1$ отвечает функции $f_{+|t|}$, $\varkappa = -1$ — функции $f_{-|t|}$. Определим пару случайных величин — функций на пространстве $\widehat{\Omega}$: z и w. Именно, $z(\overline{\omega}, \varkappa) = +1$, если метод $\mathcal{B}_t$, решая задачу $\varkappa$ при шуме оракула $\overline{\omega}$, выдает результат на первых M_1 шагах своей работы, и этот результат $\bar{x}$ таков, что $\langle \varphi \mid \bar{x} \rangle \leqslant 0$, либо если метод $\mathcal{B}_t$ при этих же условиях не выдает результата на первых M_1 шагах. В остальных случаях $z(\overline{\omega}, \varkappa) = -1$. Далее, пусть $w(\overline{\omega}, \varkappa) = \varkappa$.

Предположим, что

$$\nu < \frac{1}{256} \quad \text{и} \quad t = 64\nu. \tag{3.5}$$

Убедимся, что тогда $z = w$ с вероятностью, $\geqslant {}^3/_4$. Действительно, если $z(\overline{\omega}, \varkappa) \neq w(\overline{\omega}, \varkappa)$, то либо трудоемкость соответствующей $\overline{\omega}$, $\varkappa$ реализации $\mathcal{B}_t$ больше $16M$, либо погрешность ее результата в качестве решения соответствующей задачи больше 16ν (это сразу следует из выбора t и определения z). В силу (3.3) вероятности каждого из указанных событий не превосходят ${}^1/_8$, что и требуется.

Итак, $\mathsf{Pr}\,\{z = w\} \geqslant {}^3/_4$. Кроме того, ясно, что $\mathsf{Pr}\,\{w = 1\} = \mathsf{Pr}\,\{w = -1\} = {}^1/_2$ (вероятности вычислены по распределению на $\widehat{\Omega}$). Из этих соотношений легко извлечь, что шенноновская взаимная информация $I(z, w)$ случайных величин z и w не слишком мала:

$$I(z, w) \geqslant c_0 > 0, \tag{3.6}$$

где c_0 — абсолютная константа.

С другой стороны, ясно, что ответы оракула $\mathcal{O}$ о задачах $f_{+|t|}$ и $f_{-|t|}$ в любой точке x получаются однозначным преобразованием (зависящим, как от параметра, от x, но не зависящим от того, о какой задаче ($f_{+|t|}$ или $f_{-|t|}$) идет речь и какой шум реализован в оракуле) из его ответов о той же задаче в точке, скажем, h. Поэтому информация, поставляемая оракулом $\mathcal{O}$ методу $\mathcal{B}_t$, получается однозначным преобразованием (независимо от решаемой задачи $f \in \widetilde{C}_t$) из последовательности отвечающих данной серии шумов оракула реализаций его ответов на вопрос о f в точке h. Тем самым и случайная величина z является однозначной функцией серии $\overline{\zeta}$ ответов оракула на вопрос о решаемой задаче в точке h. Так как z есть однозначная функция $\overline{\zeta}$, то $I(z, w) \leqslant I(\overline{\zeta}, w)$.

Последняя величина вовсе не зависит от $\mathscr{B}_t$, а полностью определяется устройством оракула $\mathcal{O}$; $I(\bar{\zeta}, w)$ может быть явно вычислено. Выкладка, которую мы опускаем, приводит в условиях (3.5) к оценке

$$I(\bar{\zeta}, w) \leqslant ((1 + c\nu^2))^{M_1} - 1),$$

$c > 0$ — абсолютная константа. Сопоставляя эту оценку с неравенством $I(\bar{\zeta}, w) \geqslant I(z, w)$, получим в силу (3.6) и определения M_1, что при выполнении (3.5)

$$M \geqslant \frac{c_2}{\nu^2}, \tag{3.7}$$

$c_2 > 0$ — абсолютная константа.

Отсюда, как и в п. 1°, следует справедливость доказываемой оценки (3.1). Теорема доказана.

3.2. Утверждение теоремы 3.1 позволяет очертить область субоптимальности ЗС-метода решения задач стохастического программирования без ограничений из § 2. Прежде всего мы видим, что если речь идет о решении задач класса $\widetilde{C} = \widetilde{C}^{\mathscr{O},0}_{r,L}(G, E, \|\cdot\|, 0)$ на регулярном пространстве $E, \|\cdot\|$, для которого надлежащим выбором $V(\cdot)$ можно обеспечить $\omega_V(t) \leqslant c_V t$, то трудоемкость $\widetilde{\text{ЗС}}_V$ в принципе — для некоторых $\mathcal{O}$ — не снижаема более чем в зависящую лишь от V и r константу $d_{V,r}$ раз. Действительно, трудоемкость $\widetilde{\text{ЗС}}_V$, настроенного на точность ν, в этом случае не превосходит в силу (2.18) — (2.19) величины $b_V\left(\frac{1}{\nu}\right)^{\frac{\bar{r}}{\bar{r}-1}}$, где $\bar{r} = \min\{r, 2\}$, а нижняя оценка (3.1) имеет тот же порядок зависимости от ν. Указанная ситуация, в частности, имеет место при $(E, \|\cdot\|) = (L_p, \|\cdot\|_p)$ с $1 < p \leqslant 2$. Фактически в последнем случае трудоемкость $\widetilde{\text{ЗС}}^{p,r}$, ввиду (2.20), не снижается более чем в $b(p, r)$ раз и тогда, когда $p > 2$, но $p/(p-1) \geqslant r$.

Продолжим анализ вопроса об оптимальности методов $\widetilde{\text{ЗС}}^{p,r}$. Считаем, разумеется $(E, \|\cdot\|) = (L_p, \|\cdot\|_p)$. Следует разобрать еще случай $p/(p-1) < r$, а также случай $p = 1$. Если $p/(p-1) < r$, то, как отмечалось в п. 2.4 (и как сразу следует из (2.20)), трудоемкость метода $\widetilde{\text{ЗС}}^{p,r}$ с точностью до множителя, зависящего лишь от p и r, та же, что и у использующего точный оракул первого порядка метода решения липшицевых выпуклых задач из § 3 гл. III. Стало быть, при указанных p и r методы $\widetilde{\text{ЗС}}^{p,r}$ заведомо субоптимальны в тех ситуациях, в которых (при точном оракуле) субоптимальны методы ЗС_p — именно, для задач большой размерности на областях G типа L_p-шаров.

Наконец, если $p = 1$ и $\dim E = n < \infty$, то, как следует из сопоставления (2.21) и (3.1), метод $\widetilde{\text{ЗС}}^{1,n;r}$, если и может быть улучшен по трудоемкости, то разве что в $b(r)\ln(n+1)$ раз.

Подчеркнем, что в проведенном обсуждении фраза типа «трудоемкость такого-то метода решения задач класса $\widetilde{C}^{\mathscr{O},0}_{r,L}(G, E, \|\cdot\|, 0)$ в принципе снижается не более чем во столько-то раз» имеет следующий точный смысл. Для данных $G, E, \|\cdot\|, r, L$ найдется такой естественный оракул $\mathcal{O}$, что высказанное утверждение справедливо для отвечающего $\mathcal{O}$ класса задач.

3.3. Теперь построим нижнюю оценку сложности для классов задач стохастического программирования с ограничениями.

Теорема. *Существует функция $c(r) > 0$ $(r > 1)$ и число $\bar{\nu}_0 > 0$, такие, что справедливо следующее утверждение. Каковы бы ни были $r > 1$, банахово пространство $E, \|\cdot\|$ и ограниченное выпуклое замкнутое не сводящееся к точке множество $G \subset E$, существует естественный (в смысле § 1) оракул $\mathcal{O}$ для поля задач $\boldsymbol{C}_{\text{Lip}} = \boldsymbol{C}_{\text{Lip}}(G, E, \|\cdot\|, m)$, такой, что стохастическая сложность $\widetilde{N}(\nu)$ класса задач $\widetilde{C} = \widetilde{C}^{\mathscr{O},0}_{r,L}(G, E, \|\cdot\|, m)$ удовлетворяет при $\nu < \nu_0$ оценке*

$$\widetilde{N}(\nu) \geqslant c(r)\ln(m+2)\left(\frac{1}{\nu}\right)^{\frac{\bar{r}}{\bar{r}-1}}, \quad \bar{r} = \min\{r, 2\}. \tag{3.8}$$

При $m > 0$ оценка (3.8) справедлива для подкласса задач $\widetilde{C}'$, образованного всеми совместными задачами f из $\widetilde{C}$ с $f_0 \equiv 0$.

Доказательство. Имея в виду теорему 3.1, достаточно доказать второе утверждение сформулированной теоремы. При $m = 1$ оно доказывается с помощью простой модификации конструкции из доказательства теоремы 3.1. Эту модификацию мы описывать не будем, предоставляя ее реализацию читателю. Пусть теперь $m > 1$. Без ограничения общности можно считать, что $m = 2k$ четно. Пусть, как и в п. 3.1, $\rho_{\|\cdot\|}(G) = 1$, G содержит точки h и $-h$ с $\|h\| > 1/4$, $\varphi \in E^*$ таково, что $\|\varphi\|_* = 1$ и $\langle\varphi \mid h\rangle = \|h\|$. Сконструируем оракул $\mathcal{O}$ следующим образом. Для каждого t, $0 < t < 1/4$, рассмотрим $m+1$ задачу $f^{t,s}$, $0 \leqslant s \leqslant m$, вида

$$f_0^{t,s} \equiv 0, \quad f_j^{t,s}(x) = \begin{cases} -t - \langle\varphi \mid x\rangle + 2\delta_{js}t, & 1 \leqslant j \leqslant k, \\ -t + \langle\varphi \mid x\rangle + 2\delta_{js}t, & k+1 \leqslant j \leqslant m, \end{cases}$$

где

$$\delta_{js} = \begin{cases} 0, & j \neq s, \\ 1, & j = s \end{cases}$$

— символы Кронекера. Пусть $\bar{h} = h/\|h\|$.

Ясно, что $f^{t,s} \in \boldsymbol{C}_{\text{Lip}}(G, E, \|\cdot\|, m)$ и при $s > 0$ задача $f^{t,s}$ имеет допустимые планы — именно, для $s \leqslant k$ допустимыми планами

$f^{t,s}$ являются $\{x \in G \mid \langle\varphi \mid x\rangle = t\}$, а для $s > k$ допустимые планы $f^{t,s}$ — точки $\{x \in G \mid \langle\varphi \mid x\rangle = -t\}$.

Пусть теперь оракул $\mathcal{O}$ на всех задачах, кроме задач семейства $\{f^{t,s}\}_{1\leqslant s\leqslant m,\, 0<t<1/4}$, есть точный оракул первого порядка, а ответы $\mathcal{O}$ на вопросы о задаче $f^{t,s}$ в точке x устроены следующим образом. $\mathcal{O}$ точно собщает опорные функционалы к компонентам $f_j(x)$ задачи в x. Далее, ответ $\mathcal{O}$ о значении вектора $f^{t,s}(x)$, $s \geqslant 1$, имеет вид $f^{t,0}(x) + \xi^{s,t}$, где случайный вектор $\xi^{s,t}$ распределен следующим образом:

— с вероятностью $1 - p(t, r)$ $\xi^{s,t} = 0$;

— условное распределение вектора $\xi^{s,t}$ при условии $\xi^{s,t} \neq 0$ есть распределение с независимыми координатами; нулевая из них принимает значение 0, а j-я, $1 \leqslant j \leqslant m$, — значение $a(t, r)$ с вероятностью $1/2 + \delta_{js}\Delta(t, r)$ и значение $(-a, (t, r))$ — с вероятностью $1/2 - \delta_{js}\Delta(t, r)$. Здесь

$$\left.\begin{aligned} p(t,r) &= \begin{cases}(2t)^{\frac{r}{r-1}}, & 1<r<2,\\ 1, & r\geqslant 2;\end{cases}\\ a(t,r) &= \begin{cases}(1/2t)^{\frac{1}{r-1}}, & 1<r<2,\\ 1, & r\geqslant 2;\end{cases}\\ \Delta(t,r) &= \begin{cases}1/2, & 1<r<2,\\ t, & r\geqslant 2.\end{cases}\end{aligned}\right\} \tag{3.9}$$

Непосредственно проверяется, что при таком определении $\mathcal{O}$ задачи $f^{t,s}$, $0 < t < 1/4$, $s \in \overline{1, m}$, попадают в класс $\widetilde{C}^{\mathcal{O},0}_{r,L}(G, E, \|\cdot\|, m)$ с $L = 2$. Не ограничивая общности, можно считать, что именно об этом классе задач и идет речь в утверждении теоремы. Пусть $\widetilde{C}'$ — соответствующий подкласс. Тогда $f^{t,s} \in \widetilde{C}'$, $0 < t < 1/4$, $1 \leqslant s \leqslant m$. Пусть теперь $\nu_0 = 10^{-2}$, $\nu < \nu_0$, и $\mathcal{B}$ — метод решения задач класса $\widetilde{C}'$ точности ν и трудоемкости M на этом классе. Рассмотрим подкласс $\widetilde{C}_t$ класса $\widetilde{C}'$, образованный m задачами $f^{t,s}$, $1 \leqslant s \leqslant m$, и снабдим его равновероятным распределением вероятностей $p(s)$ (задача $f^{t,s} \in \widetilde{C}_t$ отождествляется с ее «номером» s). Как и в доказательстве теоремы 3.1, найдется детерминированный метод $\mathcal{B}_t$, для которого

$$\int_{\widetilde{C}_t} l(\mathcal{B}_t, f^{t,s})\,dp(s) \leqslant 2M \quad\text{и}\quad \int_{\widetilde{C}_t} \nu(\mathcal{B}_t, f^{t,s})\,dp(s) \leqslant 2\nu. \tag{3.10}$$

Пусть $\overline{\Omega} = \{\overline{\omega}\}$ — пространство наборов шумов оракула на первых $16]M[$ шагах метода, а $\hat{\Omega} = \overline{\Omega} \times \widetilde{C}_t$ (справа стоит произведение вероятностных пространств). Рассмотрим две случайные величины z, w — функции на $\hat{\Omega}$, определенные следующим образом: $z(\overline{\omega}, s) = +1$, если метод B_t, решая задачу $f^{t,s}$ при наборе шумов оракула на первых $16]M[$ шагах $\overline{\omega}$, не выдает результата на первых $16]M[$ шагах или выдает результат $\bar{x}$ с $\langle\varphi \mid \bar{x}\rangle > 0$; в противном случае $z(\overline{\omega}, s) = -1$. Далее,

$$w(\widetilde{\omega}, s) = \begin{cases}+1, & 1 \leqslant s \leqslant k,\\ -1, & k+1 \leqslant s \leqslant m.\end{cases}$$

Пусть $t = 32\nu$. Тогда, как и п. 2° доказательства теоремы 3.1, из (3.10) следует $\Pr\{z = w\} \geqslant 3/4$, вследствие чего взаимная информация случайных величин z, w $I(z, w)$ удовлетворяет оценке

$$I(z, w) \geqslant c_0 > 0, \tag{3.11}$$

где c_0 — абсолютная константа. С другой стороны, ясно, что z является однозначной функцией от $16]M[$ реализаций случайной величины $\xi^{s,t}$ (т. е. для некоторой функции $\varphi_t(\cdot)$ имеем $z(\overline{\omega}, s) = \varphi_t(\bar{\xi}^t(\overline{\omega}, s))$, где $\bar{\xi}^t(\overline{\omega}, s)$ — серия $16]M[$ последовательных реализаций величины $\xi^{s,t}$, отвечающих шуму $\overline{\omega}$). Стало быть, $I(\bar{\xi}^t, w) \geqslant I(z, w) \geqslant c_0$. Но величина $I(\bar{\xi}^t, w)$ не зависит от метода $\mathcal{B}_t$, а определяется лишь устройством оракула $\mathcal{O}$. Ее оценка, которую мы предоставляем читателю, дает

$$I(\bar{\xi}^t, w) \leqslant \frac{c}{2m}\{(1 + 4\Delta^2(t, r)\,p(t, r))^{16]M[} - 1\},$$

$c > 0$ — абсолютная константа. Неравенство $I(\bar{\xi}^t, w) \geqslant c_0$ вместе с этой оценкой и соотношениями (3.9) сразу приводит к (3.8). Теорема доказана.

3.4. Мы видим, что сложность классов задач стохастического программирования не может не зависеть от числа ограничений, хотя эта зависимость довольно слаба. В следующей главе мы покажем, что полученная *нижняя* оценка сложности в достаточно общих предположениях на самом деле точна. Заметим еще, что нижние оценки сложности, доставляемые теоремами 3.1 и 3.3, «набегают» уже на простейших (линейных) задачах, так что никакое сужение рассматриваемых классов за счет увеличения требований к гладкости образующих классы задач в стохастическом случае не приводит к существенному снижению трудоемкости (в случае детерминированного оракула это не так, см. главу VII). Если же налагать на задачи класса не только ограничения гладкости, но и ограничения типа сильной выпуклости (считая E, $\|\cdot\|$ гильбертовым), то для задач без ограничений можно иногда добиться снижения трудоемкости, а для задач с ограничениями $(m \geqslant 2)$ — *нельзя. Второе сразу следует из простой модификации* доказательства теоремы 3.3 (добавление квадратичного члена вида $1/20\, x^2$ к формулам для $f_j^{t,s}$ ничего по существу не меняет в выводе (3.8)).

Возможность снизить сложность стохастического программирования *без ограничений* за счет предположений о гладкости и строгой выпуклости оптимизируемой функции хорошо известна (см., например, [9]). Ее иллюстрирует результат, сформулированный в следующем упражнении.

Упражнение. Пусть $E, \|\cdot\|$ — гильбертово пространство. $G \subset E$ — выпуклое замкнутое ограниченное тело, $\rho_{\|\cdot\|}(G) = 1$, $\widetilde{C}_{c_1,c_2}$ — подкласс класса $\widetilde{C}^{\mathscr{O},0}_{2,1}(G, E, \|\cdot\|, 0)$, образованный функциями f, такими, что если x^* — точка минимума f на G, то при $y \in G$

$$f(x^*) + c_2\|y - x^*\|^2 \geqslant f(y) \geqslant f(x^*) + c_1\|y - x^*\|^2.$$

Докажите, что метод (2.15) — (2.16) при надлежащем выборе числа шагов и смещений ρ_i в состоянии решать задачи из $\widetilde{C}_{c_1,c_2}$ с (средней) точностью ν, при трудоемкости, $\leqslant a(c_1, c_2)/\nu$ (вместо $O(1/\nu^2)$ в общем случае). Результатом работы метода после i шагов считайте точку x_i. Рассматривая простейшие квадратичные задачи $(x - a)^2 \to \min$, убедитесь, что трудоемкость $O(1/\nu)$ — лучшее, на что можно рассчитывать в описанной ситуации (независимо от того, какой метод используется для решения).

Глава VI

РЕШЕНИЕ ВЫПУКЛО-ВОГНУТЫХ ИГР И ЗАДАЧ СТОХАСТИЧЕСКОГО ПРОГРАММИРОВАНИЯ С ОГРАНИЧЕНИЯМИ

В этой главе описано применение ЗС-методов к решению выпукло-вогнутых игр, наблюдаемых как точно, так и в помехах. Показано, как сводить задачи стохастического программирования с ограничениями к задачам решения игр. Построены методы решения стохастических задач с ограничениями. Результаты предыдущей главы позволяют обосновать субоптимальность указанных методов. Последующее изложение основано на работах авторов [21, 22].

§ 1. Классы выпукло-вогнутых игровых задач

В этом параграфе описываются классы задач, которыми мы будем заниматься на протяжении всей главы. Речь будет идти о задачах аппроксимации решений выпукло-вогнутых игр или, что то же, седловых точек выпукло-вогнутых функций.

1.1. Пусть $E, \|\cdot\|$ и $E_I, \|\cdot\|_I$ — регулярные пространства, удовлетворяющие условиям, сформулированным в начале § 1 предыдущей главы, а $G \subset E$ и $G_I \subset E_I$ — ограниченные выпуклые непустые замкнутые множества (более общие ситуации у нас не возникнут). Пусть $F(x, l)$: $G \times G_I \to \mathbb{R}$ — непрерывная функция. Функция F называется *выпукло-вогнутой*, если при каждом $l \in G_I$ F, как функция $x \in G$, выпукла на G и при каждом $x \in G$ F, как функция $l \in G_I$, вогнута по $l \in G_I$. Точка $(x^*, l^*) \in G \times G_I$ называется *седловой точкой* F, если в этой точке F, как функция x, достигает минимума, а как функция l — максимума:

$$F(x, l^*) \geqslant F(x^*, l^*) \geqslant F(x^*, l) \quad \text{для всех } (x, l) \in G \times G_I. \qquad (1.1)$$

Понятие седловой точки удобно интерпретировать в терминах игр. Сопоставим функции F *игру* F двух лиц с нулевой суммой, в которой стратегии первого игрока отождествляются с точками $x \in G$, стратегии второго — с точками $l \in G_I$, а исход партии (x, l), в которой первый игрок избрал стратегию x, а второй — стратегию l, состоит в выплате первым игроком второму суммы

$F(x, l)$. При такой интерпретации седловые точки функции F называют *решениями игры* F; это в точности пары стратегий игроков, обладающие тем свойством, что отход каждого из игроков от соответствующей стратегии не приносит ему пользы, если другой игрок продолжает придерживаться отвечающей ему стратегии.

То же самое можно сформулировать следующим образом. Пусть $\bar{F}(x) = \sup_{l \in G_I} F(x, l)$ — гарантированная плата первого игрока, а $\underline{F}(l) = \inf_{x \in G} F(x, l)$ — гарантированный доход второго. Если функция F имеет седловую точку, то

$$\inf_{x \in G} \bar{F}(x) = \min_{x \in G} \bar{F}(x) = \sup_{l \in G_I} \underline{F}(l) = \max_{l \in G_I} \underline{F}(l). \tag{1.2}$$

При этом множество седловых точек F есть $X^*(F) \times L^*(F)$, где

$$X^*(F) = \{x \in G \mid \bar{F}(x) = \inf_{y \in G} \bar{F}(y)\}$$

и

$$L^*(F) = \{l \in G_I \mid \underline{F}(l) = \sup_{s \in G_I} \underline{F}(s)\}. \tag{1.3}$$

Наоборот, если выполнено (1.2), то функция F имеет седловую точку. Возникает вопрос о том, всегда ли выпукло-вогнутая непрерывная функция F имеет седловую точку. В нашей ситуации (E, E_I рефлексивны, G, G_I выпуклы, замкнуты, ограничены) это всегда так в силу фундаментальной теоремы фон Неймана [19].

1.2. Наша цель состоит в построении методов первого порядка, позволяющих отыскивать приближенные решения выпукло-вогнутых игр. Чтобы говорить о таких методах, надо фиксировать следующие объекты:

— поле задач, т. е. множество игр, подлежащих решению;

— меру погрешности точек $(x, l) \in G \times G_I$ в качестве приближенных решений игр;

— доступный методу источник информации о решаемой игре — оракул.

Опишем эти объекты следующим образом.

1.2.1. Поле задач, которое мы будем рассматривать, это множество всех игр с выпукло-вогнутыми *липшицевыми* платежными функциями $F: G \times G_I \to \mathbb{R}$. (Далее мы используем понятия «игра с платежной функцией F», «функция F», «задача отыскания решения игры с платежной функцией F» как синонимы и говорим просто о «задаче F».) Множество таких игр с данными G и G_I обозначим $\boldsymbol{D} \equiv \boldsymbol{D}(G \times G_I, E, E_I, \|\cdot\|, \|\cdot\|_I)$.

1.2.2. Погрешность точки $(x, l) \in G \times G_I$ в качестве приближенного решения игры $F \in \boldsymbol{D}$ определяется следующим образом. Точное решение игры есть пара решений экстремальных задач

$$\bar{F}(x) \to \min \mid x \in G \quad \text{и} \quad \underline{F}(l) \to \max \mid l \in G_I. \tag{1.4}$$

Естественно измерять погрешность точки $(x, l) \in G \times G_I$ в качестве приближенного решения игры F полусуммой погрешностей ее компонент в качестве приближенных решений «своих» экстремальных задач, т. е. величиной

$$\varepsilon(x, l; F) = \frac{1}{2}\{\varepsilon(x, F) + \varepsilon_I(l, F)\},$$

где

$$\varepsilon(x, F) = \bar{F}(x) - \inf_G \bar{F}(x),$$

$$\varepsilon_I(l, F) = -\underline{F}(l) + \sup_{G_I} \underline{F}(l).$$

Наряду с введенными абсолютными мерами погрешностей полезно рассматривать и относительные

$$\nu(x, F) \equiv \frac{\varepsilon(x, F)}{r(F)}, \quad \nu_I(l, F) \equiv \frac{\varepsilon_I(l, F)}{r(F)},$$

$$\nu(x, l; F) = \frac{\varepsilon(x, l; F)}{r(F)} = \frac{1}{2}\{\nu(x, F) + \nu_I(l, F)\}.$$

В качестве нормирующего множителя в случае детерминированного оракула (см. ниже) удобно взять величину

$$r(F) = 2\max\{\text{Ч}_{\|\cdot\|}(F)\rho_{\|\cdot\|}(G), \quad \text{Ч}_{\|\cdot\|_I}(F)\rho_{\|\cdot\|_I}(G_I)\},$$

где $\text{Ч}_{\|\cdot\|}(F)$ есть верхняя грань (по $l \in G_I$) констант Липшица функций $F(\cdot, l)$, получающихся из F фиксацией l, относительно $\|\cdot\|$, а $\text{Ч}_{\|\cdot\|_I}(F)$ определено «симметричным» образом. Ясно, что $\nu(x, l; F) \geqslant 0$ и $\nu(x, l; F) = 0 \Leftrightarrow \{(x, l)$ есть решение игры $F\}$. При указанном выше выборе $r(F)$ $2r(F)$ — естественная верхняя оценка изменения F на $G \times G_I$, так что в этом случае $\nu(x, l; F) \leqslant 1$ для всех $(x, l) \in G \times G_I$ и $F \in \boldsymbol{D}$. Тем самым $\nu(x, l; F)$ действительно имеет смысл относительной погрешности.

1.2.3. Мы будем рассматривать два типа оракулов — детерминированные и стохастические оракулы первого порядка. Как и в математическом программировании, детерминированный оракул есть частный случай стохастического. Последний в интересующем нас случае есть набор $\mathcal{O} = ((\Omega, \mathsf{F}_\omega); \psi_x(x, l; F), \psi_l(x, l; F))$, где $(\Omega, \mathsf{F}_\omega)$ — польское пространство с регулярной борелевой полной по Лебегу вероятностной мерой, ψ_x: $(G \times G_I) \times \boldsymbol{D} \to E^*$ и $\psi_l: (G \times G_I) \times \boldsymbol{D} \to E_I^*$ — функции наблюдения, сопоставляющие паре $(x, l) \in G \times G_I$ и задаче $F \in \boldsymbol{D}$ наблюдения опорных функционалов к F как выпуклой функции x (вогнутой функции l) соответственно. Эти наблюдения при данных x, l, F представляют собой случайные величины — функции на пространстве элементарных событий Ω. В определение входит требование борелевости ψ_x, ψ_l по x, l, ω.

Как и для методов оптимизации, считается, что на i-м шаге метод может задать оракулу вопрос в любой (по выбору метода) точке $(x_i, l_i) \in G \times G_I$ и получить в ответ значения функций $\psi_x(x_i, l_i, \omega_i)$, $\psi_l(x_i, l_i, \omega_i)$, где ω_i — случайный шум, реализовавшийся в оракуле на i-м шаге. Шумы разных шагов предполагаются независимыми в совокупности и одинаково распределенными.

По образцу §§ 2—4 гл. I можно определить понятия «метод решения задач из поля $\boldsymbol{D}$, использующий оракул $\mathcal{O}$», «траектория метода», «результаты работы метода», «средняя трудоемкость и погрешность (относительная и абсолютная)». Мы не будем выписывать эти определения, предоставляя эту работу читателю.

Наметив путь описания понятий, которыми мы будем оперировать, вернемся к определению двух типов оракулов, которые мы будем рассматривать в связи с играми. Первый — детерминированный оракул заданной (относительной) точности ν_0. Для такого оракула Ω есть одноточечное множество. При этом по определению выполнены условия

$$\left.\begin{aligned} &\|\psi_x(x,l;F)\|_* \leqslant \mathsf{L}_{\|\cdot\|}(F), \quad \|\psi_l(x,l;F)\|_{I^*} \leqslant \mathsf{L}_{\|\cdot\|_I}(F),\\ &F(y,l)-F(x,l) \geqslant \langle \psi_x(x,l;F)\,|\,y-x\rangle - \nu_0 r(F),\\ &F(x,s)-F(x,l) \leqslant \langle \psi_l(x,l;F)\,|\,s-l\rangle + \nu_0 r(F) \end{aligned}\right\} \tag{1.5}$$

при всех $F \in \boldsymbol{D}$, $x, y \in G$, $s, l \in G_I$. Классы задач, получающиеся оснащением $\boldsymbol{D}$ (рассматриваемого как поле задач вместе с фиксированной мерой погрешности $\nu(\cdot, \cdot)$, отвечающей описанному выше выбору $r(F)$) такого типа оракулами, будем называть классами типа $\boldsymbol{D}^{\nu_0}(G \times G_I, E, \|\cdot\|, E_I, \|\cdot\|_I)$.

Второй тип оракулов — это стохастический оракул, несмещенный, точности ν_0. Его определение аналогично определению стохастического оракула первого порядка из § 1 гл. V. Именно, пусть $\mathcal{O}((\Omega, \mathsf{F}_\omega); \psi_x(\cdot), \psi_l(\cdot))$ — оракул указанного выше типа. Пусть, далее, $L > 0$ и $r > 1$. Свяжем с оракулом $\mathcal{O}$ поле задач $\widetilde{\boldsymbol{D}}^{\mathcal{O},\nu_0}_{r,L}(G \times G_I, E, \|\cdot\|, E_I, \|\cdot\|_I)$, образованное всеми задачами $F \in \boldsymbol{D}$, для которых при всех $x, y \in G$, $l, s \in G_I$ справедливы соотношения

$$\text{(A)}\ \begin{cases} F(y,l)-F(x,l) \geqslant \langle \mathsf{M}_{\mathsf{F}_\omega}\psi_x(x,l;F,\omega)\,|\,y-x\rangle - \nu_0 L,\\ F(x,s)-F(x,l) \leqslant \langle \mathsf{M}_{\mathsf{F}_\omega}\psi_l(x,l;F,\omega)\,|\,s-l\rangle + \nu_0 L, \end{cases}$$

$$\text{(B)}\ \begin{cases} \mathsf{M}_{\mathsf{F}_\omega}\|\psi_x(x,l;F,\omega)\|_*^r \leqslant \left(\dfrac{L}{2\rho_{\|\cdot\|}(G)}\right)^r,\\ \mathsf{M}_{\mathsf{F}_\omega}\|\psi_l(x,l;F,\omega)\|_{I^*}^r \leqslant \left(\dfrac{L}{2\rho_{\|\cdot\|_I}(G_I)}\right)^r. \end{cases}$$

Поле задач $\widetilde{\boldsymbol{D}}^{\mathcal{O},\nu_0}_{r,L}$, снабженное нормирующим множителем $r(F) \equiv L$ (т. е. мерой относительной погрешности $\nu(x, l; F) = \frac{1}{L}\varepsilon(x, l; F)$) и оракулом $\mathcal{O}$, образует класс задач, обозначаемый, как и исходное поле, через $\widetilde{\boldsymbol{D}}^{\mathcal{O},\nu_0}_{r,L}(G \times G_I, E, \|\cdot\|, E_I, \|\cdot\|_I)$.

Заметим, что если одно из множеств G, G_I, скажем, G_I есть точка, то класс задач решения игр вида $\widetilde{\boldsymbol{D}}^{\mathcal{O},\nu_0}_{r,L}(G \times G_I, E, \|\cdot\|, E_I, \|\cdot\|_I)$ по существу превращается в класс $\widetilde{\boldsymbol{C}}^{\mathcal{O},\nu_0}_{r,L}(G, E, \|\cdot\|)$ задач стохастического программирования без ограничений.

§ 2. ЗС-методы решения игр: случай детерминированного оракула

В этом параграфе описывается применение ЗС-методов к решению игр *при точной* (вернее, слабо искаженной помехами) информации — к решению задач классов $\boldsymbol{D}^{\nu_0}(G \times G_I, E, \|\cdot\|, E_I, \|\cdot\|_I)$.

2.1. Начнем с выделения особенностей ситуации с играми по сравнению с более простой проблемой решения выпуклых экстремальных задач без ограничений. Посмотрим, какие трудности нас подстерегают в простейшей — гильбертовой ситуации (E, $\|\cdot\|$ и E_I, $\|\cdot\|_I$ гильбертовы). В этом случае ЗС-методы должны превращаться в некоторый аналог градиентного метода.

Градиентные методы решения выпукло-вогнутых игр хорошо известны. Простейший из них — метод Эрроу — Гурвица [29] — состоит в построении последовательных приближений (x_i, l_i) к искомому решению по формулам

$$\left.\begin{aligned} x_{i+1} &= \pi_G(x_i - \rho_i \nabla_x F(x_i, l_i)),\\ l_{i+1} &= \pi_{G_I}(l_i + \rho_i \nabla_l F(x_i, l_i)). \end{aligned}\right\} \tag{2.1}$$

Здесь ∇_x, ∇_l означают субградиенты по x, l соответственно, а $\rho_i > 0$ — надлежащим образом выбранные смещения. К сожалению, *указанная версия, вообще говоря, не сходится*, как бы ни выбирались смещения ρ_i (простейший пример: $F(x, l) = xl$, $|x| \leqslant 1$, $|l| \leqslant 1$). Для обеспечения сходимости метода при надлежащих ρ_i приходится налагать дополнительные ограничения на F типа сильной выпукло-вогнутости. Имеется много способов модифицировать метод так, чтобы он стал применим к любой $F \in \boldsymbol{D}$ (регуляризация F, выбор различных смещений по x- и по l-компонентам и т. п.). Насколько известно авторам, все такие модификации приводят к методам, трудоемкость которых на $\boldsymbol{D}$ выше по порядку, чем у методов типа градиентной выпуклой оптимизации.

Оказывается, что перечисленные трудности легко преодолеть: при весьма общих предположениях относительно ρ_i $\big(\rho_i \xrightarrow[i\to\infty]{} 0,\ \sum_i \rho_i = \infty\big)$

процедура (2.1) всегда сходится к седлу F, только не в обычном, а в чезаровском смысле. Именно, к седлу сходится последовательность

$$(x^i, l^i) = \frac{1}{\sum_{j=1}^{i} \rho_j} \sum_{j=1}^{i} \rho_j (x_j, l_j) \tag{2.2}$$

(ср. с ситуацией § 2 гл. V). Скорость сходимости траектории (2.2), доставляемая рациональным выбором смещений, оказывается той же, что и у градиентных методов решения липшицевых выпуклых экстремальных задач.

Отметим, что усреднение типа (2.2) в гильбертовой ситуации впервые, по-видимому, предложено Р. Бруком (применительно к близкой играм проблеме решения вариационных неравенств с монотонными операторами) в [3] и несколько позже, но независимо — авторами в [22, 21] (в последней работе речь идет об общей — регулярной ситуации).

Методы, описываемые ниже, получаются из метода Эрроу — Гурвица заменой градиентного спуска зеркальным и добавлением усреднения (2.2). Перейдем к их описанию.

2.2. Фиксируем $(E, \|\cdot\|)$-регулярную функцию $V(\varphi)$ и $(E_I, \|\cdot\|_I)$-регулярную функцию $V_I(\lambda)$. Опишем отвечающий V, V_I «сходящийся» — бесконечношаговый метод решения игр класса $\boldsymbol{D}^{\nu_0}(G \times G_I, E, \|\cdot\|, E_I, \|\cdot\|_I)$. При описании метода будем считать G и G_I нормированными условиями $0 \in G$, $0 \in G_I$, $\rho_{\|\cdot\|}(G) = \rho_{\|\cdot\|_I}(G_I) = 1$ (эти условия на G, G_I предполагаются выполненными на протяжении всего параграфа).

Метод (он будет обозначаться $\text{ЗС}_{V, V_I}$) состоит в построении последовательностей $\{\varphi_i \in E^*\}_{i=0}^{\infty}$, $\{\lambda_i \in E_I^*\}_{i=0}^{\infty}$ и их «теней» $\{x_i \in E,\ \bar{x}_i \in G\}_{i=1}^{\infty}$, $\{l_i \in E_I,\ \bar{l}_i \in G_I\}_{i=1}^{\infty}$, а также последовательности результатов $\{(x^i, l^i) \in G \times G_I\}_{i=1}^{\infty}$ в соответствии со следующими рекуррентными формулами (в которых $\rho_i > 0$ — некоторые смещения, ψ_x, ψ_l — функции наблюдения, $F \in \boldsymbol{D}^{\nu_0}$ — решаемая задача):

$$\left.\begin{aligned} \varphi_0 = 0, \quad \lambda_0 = 0; \quad x_i &= V'(\varphi_{i-1}), \quad \bar{x}_i = \pi_G(x_i), \\ l_i &= V_I'(\lambda_{i-1}), \quad \bar{l}_i = \pi_{G_I}(l_i), \end{aligned}\right\} \tag{2.3}$$

$$\left.\begin{aligned} \varphi_i &= \varphi_{i-1} - \rho_i (\psi_x(\bar{x}_i, \bar{l}_i; F) + \|\psi_x(\bar{x}_i, \bar{l}_i; F)\|_* \mu_G(x_i)), \\ \lambda_i &= \lambda_{i-1} + \rho_i (\psi_l(\bar{x}_i, \bar{l}_i; F) - \|\psi_l(\bar{x}_i, \bar{l}_i; F)\|_{I*} \mu_{G_I}(l_i)); \end{aligned}\right\} \tag{2.4}$$

$$(x^i, l^i) = \frac{1}{\sum_{j=1}^{i} \rho_j} \sum_{j=1}^{i} \rho_j (\bar{x}_j, \bar{l}_j). \tag{2.5}$$

Теорема. *Пусть смещения ρ_i удовлетворяют условиям $\rho_i > 0$, $\rho_i \underset{i\to\infty}{\to} 0$, $\sum_i \rho_i = \infty$. Тогда процесс (2.3) — (2.5) сходится на любой задаче $F \in \boldsymbol{D}^{\nu_0}$ в том смысле, что $\overline{\lim_{i\to\infty}}\, \nu(x^i, l^i; F) \leqslant \nu_0$. Более того, при всех i справедлива оценка*

$$\nu(x^i, l^i; F) \leqslant \nu_0 + \frac{1 + \sum_{j=1}^{i} \rho_j r(F)(\omega_V(\rho_j r(F)) + \omega_{V_I}(\rho_j r(F)))}{2r(F) \sum_{j=1}^{i} \rho_j}. \tag{2.6}$$

Доказательство. Первое утверждение теоремы, как легко видеть, следует из второго, которое мы и будем доказывать. Пусть $x \in G$ и $l \in G_I$. Образуем функции $V_x(\varphi) = V(\varphi) - \langle x \mid \varphi \rangle$ и $V_{Il}(\lambda) = V_I(\lambda) - \langle \lambda \mid l \rangle$ и будем следить за изменением функции $\overline{V}_{x,l}(\varphi, \lambda) = V_x(\varphi) + V_{Il}(\lambda)$ вдоль траектории $\{(\varphi_i, \lambda_i)\}$. Заметим, что в силу условий нормировки на G, G_I и свойства оракула (1.5) имеем

$$\|\psi_x(x, l; F)\|_* \leqslant \frac{1}{2} r(F)$$

и

$$\|\psi_l(x, l; F)\|_{I*} \leqslant \frac{1}{2} r(F).$$

В силу стандартных для обоснования ЗС-методов соображений (см. § 3 гл. III) имеем

$$\begin{aligned} V_x(\varphi_{i+1}) &\leqslant V_x(\varphi_i) + \rho_i \langle \psi_x(\bar{x}_i, \bar{l}_i; F) + \\ &\quad + \|\psi_x(\bar{x}_i, \bar{l}_i; F)\|_* \mu_G(x_i) \mid x - x_i \rangle + \rho_i r(F)\, \omega_V(\rho_i r(F)) \leqslant \\ &\leqslant V_x(\varphi_i) + \rho_i \langle \psi_x(\bar{x}_i, \bar{l}_i; F) \mid x - \bar{x}_i \rangle + \rho_i r(F)\, \omega_V(\rho_i r(F)) \leqslant \\ &\leqslant V_x(\varphi_i) + \rho_i F(x, \bar{l}_i) - \rho_i F(\bar{x}_i, \bar{l}_i) + \nu_0 \rho_i r(F) + \rho_i r(F)\, \omega_V(\rho_i r(F)) \end{aligned}$$

(мы учли соотношение (1.5)). Равным образом

$$\begin{aligned} V_{Il}(\lambda_{i+1}) &\leqslant V_{Il}(\lambda_i) + \rho_i F(\bar{x}_i, \bar{l}_i) - \rho_i F(\bar{x}_i, l) + \\ &\quad + \nu_0 \rho_i r(F) + \rho_i r(F)\, \omega_{V_I}(\rho_i r(F)). \end{aligned}$$

Отсюда

$$\begin{aligned} V_{x,l}(\varphi_{j+1}, \lambda_{j+1}) &\leqslant V_{x,l}(\varphi_j, \lambda_j) + \rho_j F(x, \bar{l}_j) - \\ &- \rho_j F(\bar{x}_j, l) + \rho_j r(F)(\omega_V(\rho_j r(F)) + \omega_{V_I}(\rho_j r(F))) + 2\nu_0 r(F) \rho_j. \end{aligned} \tag{2.7}$$

Суммируя отношение (2.7) по $j = 1, 2, \ldots, i$, получим, учитывая, что $V_{x,l}(\varphi, \lambda) \geqslant -1$, $V_{x,l}(\varphi_0, \lambda_0) = 0$:

$$\begin{aligned} \sum_{j=1}^{i} (F(\bar{x}_j, l) - F(x, \bar{l}_j)) \rho_j &\leqslant 1 + \sum_{j=1}^{i} \rho_j r(F)(\omega_V(\rho_j r(F)) + \omega_{V_I}(\rho_j r(F))) + \\ &+ 2\nu_0 \sum_{j=1}^{i} \rho_j r(F). \end{aligned} \tag{2.8}$$

Ввиду выпуклости F по x и вогнутости F по l, левая часть (2.8) не меньше $(\sum_{j=1}^{i}\rho_j)\,(F(x^i, l) - F(x, l^i))$, так что (2.8) дает

$$F(x^i, l) - F(x, l^i) \leqslant \frac{1 + \sum_{j=1}^{i} \rho_j r(F)(\omega_V(\rho_j r(F)) + \omega_{V_I}(\rho_j r(F)))}{\sum_{j=1}^{i} \rho_j} + 2\nu_0 r(F). \tag{2.9}$$

Максимум левой части (2.9) по $x \in G$, $l \in G_I$ равен

$$\bar F(x^i) - \underline{F}(l^i) = \bar F(x^i) - \inf_G \bar F + \sup_{G_I} \underline{F} - \underline{F}(l^i) = \varepsilon(x^i, F) + \varepsilon_I(l^i, F),$$

так что (2.6) сразу следует из (2.9). Теорема доказана.

2.3. Обсудим полученный результат.

2.3.1. Начнем со случая, когда E, $\|\cdot\|$ и E_I, $\|\cdot\|_I$ гильбертовы. Здесь метод может быть упрощен. Результирующий метод определяется соотношениями

$$\left.\begin{array}{ll} x_1 = 0; & x_i = \pi_G(x_{i-1} - \rho_{i-1}\psi_x(x_{i-1}, l_{i-1}; F)), \\ l_1 = 0; & l_i = \pi_{G_I}(l_{i-1} + \rho_{i-1}\psi_l(x_{i-1}, l_{i-1}; F)); \end{array}\right\} \tag{2.10}$$

$$(x^i, l^i) = \frac{1}{\sum_{j=1}^{i} \rho_j} \sum_{j=1}^{i} \rho_j (x_j, l_j). \tag{2.11}$$

Это в точности метод Эрроу — Гурвица, дополненный усреднением (2.11). Для него также справедлива теорема 2.2 с лучшей, чем в теореме, оценкой

$$\nu(x^i, l^i, F) \leqslant \nu_0 + \frac{1 + \frac{1}{4} r^2(F) \sum_{j=1}^{i} \rho_j^2}{2r(F) \sum_{j=1}^{i} \rho_j}. \tag{2.12}$$

2.3.2. Вопрос о рациональном выборе смещений ρ_i решается примерно так же, как и в п. 2.2.2 гл. V. Именно, если $\omega_V(t) + \omega_{V_I}(t) \leqslant ct^\sigma$, $\sigma \leqslant 1$, то в классе смещений вида $\rho_i = i^{-x}$ наилучшим является выбор $x = 1/(1+\sigma)$. При этом правая часть (2.6) с ростом i ведет себя как

$$\nu_0 + O(i^{-\frac{\sigma}{1+\sigma}}) \ln i.$$

В частности, в гильбертовом случае следует выбирать $\rho_i = O(1/\sqrt{i})$. При этом правая часть (2.6) убывает как $\nu_0 + O(\ln i/\sqrt{i})$.

2.4. Теперь опишем «конструктивную» версию $\text{ЗС}_{V, V_I}(\nu)$ метода $\text{ЗС}_{V, V_I}$, призванную обеспечить заданную точность ν, $1 > \nu > \nu_0$, в решении задач из класса D^{ν_0}. Эта версия получается из «бесконечношаговой» добавлением правила «автоматического» выбора смещений ρ_i (типа правила ЗС1.6′ из § 3 гл. III) и правила определения момента остановки.

Решая задачу $F \in \boldsymbol{D}^{\nu_0}$, метод $\text{ЗС}_{V,V_I}(\nu)$ строит конечные последовательности

$$\{\varphi_i \in E^*,\quad \lambda_i \in E_I^*\}_{i=0}^{M_F},\quad \{x_i \in E,\ \bar x_i \in G,\ l_i \in E_I,\ \bar l_i \in G_I\}_{i=1}^{M_F}$$

и результат $(\bar x, \bar l)$. Здесь M_F — формируемое методом число вопросов, за которое он решает задачу F.

Работа метода $\text{ЗС}_{V,V_I}(\nu)$ описывается следующими правилами.

$\text{ЗС}_{V,V_I}$ *0.* Полагаем $i = 1$, $\varphi_0 = 0$, $\lambda_0 = 0$. Переходим к $\text{ЗС}_{V,V_I}1$.

$\text{ЗС}_{V,V_I}$ 1. i-й шаг. К i-му шагу имеются $\varphi_{i-1} \in E^*$, $\lambda_{i-1} \in E_I^*$. На i-м шаге производятся следующие действия.

$\text{ЗС}_{V,V_I}$ 1.1. Полагаем

$$x_i = V'(\varphi_{i-1}),\quad \bar x_i = \pi_G(x_i),\quad l_i = V_I'(\lambda_{i-1}),\quad \bar l_i = \pi_{G_I}(l_i).$$

$\text{ЗС}_{V,V_I}$ 1.2. В точке $(\bar x_i, \bar l_i)$ задаем вопрос оракулу о решаемой задаче F. Пусть $\eta_i = \psi_x(\bar x_i, \bar l_i; F)$ и $\eta_{Ii} = \psi_l(\bar x_i, \bar l_i; F)$ — сообщенные им ответы. Полагаем $r(i) = 2\max\{\|\eta_i\|_*, \|\eta_{Ii}\|_{I*}\}$. Если $r(i) = 0$, переходим к $\text{ЗС}_{V,V_I}$ 2, в противном случае — к $\text{ЗС}_{V,V_I}$ 1.3.

$\text{ЗС}_{V,V_I}$ 1.3. Полагаем

$$\bar\eta_i = \eta_i + \|\eta_i\|_* \mu_G(x_i),\quad \bar\eta_{Ii} = \eta_{Ii} - \|\eta_{Ii}\|_{I*} \mu_{G_I}(l_i).$$

$\text{ЗС}_{V,V_I}$ 1.4. Пусть ρ_i — верхняя грань тех ρ, для которых

$$\{V(\varphi_{i-1} - \rho\bar\eta_i) - V(\varphi_{i-1}) + \langle\bar\eta_i | x_i\rangle \rho\} + \{V_I(\lambda_{i-1} + \rho\bar\eta_{Ii}) - V_I(\lambda_{i-1}) - \langle\bar\eta_{Ii} | l_i\rangle \rho\} \leqslant (\nu - \nu_0)\rho r(i). \tag{2.13}$$

Если $\rho_i = \infty$, то переходим к $\text{ЗС}_{V,V_I}$ 2, в противном случае — к $\text{ЗС}_{V,V_I}$ 1.5.

$\text{ЗС}_{V,V_I}$ 1. 5. Полагаем $\varphi_i = \varphi_{i-1} - \rho_i\bar\eta_i$, $\lambda_i = \lambda_{i-1} + \rho_i\bar\eta_{Ii}$.

Если $\sum_{j=1}^{i} \rho_j > 0$ и

$$\|\varphi_i\|_* + \|\lambda_i\|_{I*} - V(\varphi_i) - V_I(\lambda_i) \leqslant (\max_{1 \leqslant j \leqslant i} r(j))(\nu - \nu_0) \sum_{j=1}^{i} \rho_j, \tag{2.14}$$

то переходим к $ЗС_{V,V_I}$ 2; в противном случае — увеличим i на 1 и перейдем к следующему шагу (т. е. к $ЗС_{V,V_I}$ 1.1).

$ЗС_{V,V_I}$ 2. Правило выдачи результата. Пусть M_F — момент обращения к $ЗС_{V,V_I}$ 2. В этот момент работа метода прекращается выдачей результата

$$(\bar{x}, \bar{l}) = \begin{cases} (\bar{x}_{M_F}, \bar{l}_{M_F}), & r(M_F) = 0 \text{ или } \rho_{M_F} = +\infty, \\ \dfrac{1}{\sum_{j=1}^{i} \rho_j} \sum_{j=1}^{i} \rho_j (\bar{x}_j, \bar{l}_j) & \text{в остальных случаях.} \end{cases}$$

Теорема. *Описанный метод решает всякую задачу $F \in \mathbf{D}^{\nu_0}$ с относительной погрешностью $\leqslant \nu$, $1 > \nu > \nu_0$, при трудоемкости $M_F + 1$, не превосходящей $M_{V,V_I}(\nu - \nu_0)$, где*

$$M_{V,V_I}(\nu) \equiv \max \left\{ \left] \frac{1}{\nu \gamma_V \left(\frac{\nu}{2}\right)} \right[+ 2, \quad \left] \frac{1}{\nu \gamma_{V_I} \left(\frac{\nu}{2}\right)} \right[+ 2 \right\} = \max \{M_V(2\nu), M_{V_I}(2\nu)\} \text{ *)}.$$

Доказательство.

1°. Оценка трудоемкости. Пусть метод, решая задачу F, не остановился после N-го шага. Тогда на первых N шагах было $r(i) > 0$, $\rho_i < \infty$. В соответствии с неравенством (2.6) гл. III левая часть (2.13) не превосходит $\omega_V(r(i)\rho) + \omega_{V_I}(r(i)\rho)$. Стало быть, (2.13) выполнено при

$$\rho = \frac{1}{r(i)} \min \left\{ \gamma_V \left(\frac{\nu - \nu_0}{2} \right), \ \gamma_{V_I} \left(\frac{\nu - \nu_0}{2} \right) \right\} \equiv \frac{\bar{\rho}}{r(i)}.$$

Итак, $\rho_i \geqslant \bar{\rho}/r(i)$. Но тогда правая часть (2.14) при $i = N$ не меньше

$$(\max_{j \leqslant N} r(j)) \sum_{j=1}^{N} \rho_j (\nu - \nu_0) \geqslant \bar{\rho} N (\nu - \nu_0).$$

Левая же часть (2.14) не превосходит 1. При $i = N$ (2.14) не выполняется по определению N, что приводит к $\bar{\rho} N(\nu - \nu_0) < 1$. Это и дает требуемую оценку N и тем самым M_F.

*) Функции $M_V(\cdot)$ были введены в § 3 гл. III в связи с оценками трудоемкости ЗС-методов решения липшицевых выпуклых задач.

2°. Погрешность метода. Пусть вначале $r(M_F) = 0$. В этом случае в силу свойств оракула $F(x, \bar{l}_{M_F}) \geqslant F(\bar{x}_{M_F}, \bar{l}_{M_F}) - \nu_0 r(F)$ и $F(\bar{x}_{M_F}, l) - F(\bar{x}_{M_F}, \bar{l}_{M_F}) \leqslant r(F)\nu_0$ при всех $(x, l) \in G \times G_I$, т. е.

$$\underline{F}(\bar{l}_{M_F}) - F(\bar{x}_{M_F}, \bar{l}_{M_F}) \geqslant -\nu_0 r(F)$$

и

$$\bar{F}(\bar{x}_{M_F}) - F(\bar{x}_{M_F}, \bar{l}_{M_F}) \leqslant \nu_0 r(F).$$

Отсюда, как и при доказательстве теоремы 2.2, $\varepsilon(\bar{x}_{M_F}, \bar{l}_{M_F}; F) \leqslant \nu_0 r(F)$, что и доказывает неравенство $\nu(\bar{x}_{M_F}, \bar{l}_{M_F}; F) \leqslant \nu_0 < \nu$.

Пусть теперь $r(M_F) > 0$. Пусть $x \in G$, $V_x(\varphi) = V(\varphi) - \langle \varphi \mid x \rangle$, $l \in G_I$, $V_{Il}(\lambda) = V_I(\lambda) - \langle \lambda \mid l \rangle$. При $j \leqslant M_F$ и при всяком $\rho > 0$, для которого выполнено (2.13), имеем

$$V_x(\varphi_{j-1} - \rho\bar{\eta}_j) = \{V(\varphi_{j-1} - \rho\bar{\eta}_j) - V(\varphi_{j-1}) + \rho\langle\bar{\eta}_j \mid x_j\rangle\} + V_x(\varphi_{j-1}) + \rho\langle\bar{\eta}_j \mid x - x_j\rangle$$

и

$$V_{Il}(\lambda_{j-1} + \rho\bar{\eta}_{Ij}) = \{V_I(\lambda_{j-1} + \rho\bar{\eta}_{Ij}) - V_I(\lambda_{j-1}) - \rho\langle\bar{\eta}_{Ij} \mid l_j\rangle\} + V_{Il}(\lambda_{j-1}) + \rho\langle\bar{\eta}_{Ij} \mid l_j - l\rangle.$$

Складывая эти равенства и учитывая, что ρ удовлетворяет (2.13), получаем, обозначая $V_{x,l}(\varphi, \lambda) \equiv V_x(\varphi) + V_{Il}(\lambda)$:

$$V_{x,l}(\varphi_{j-1} - \rho\bar{\eta}_j, \lambda_{j-1} + \rho\bar{\eta}_{Ij}) \leqslant (\nu - \nu_0) r(j)\rho + V_{x,l}(\varphi_{j-1}, \lambda_{j-1}) + \rho\langle\bar{\eta}_j \mid x - x_j\rangle + \rho\langle\bar{\eta}_{Ij} \mid l_j - l\rangle. \tag{2.15}$$

Как всегда, правая часть (2.15) не больше

$$(\nu - \nu_0) r(j)\rho + V_{x,l}(\varphi_{j-1}, \lambda_{j-1}) + \rho\{F(x, \bar{l}_j) - F(\bar{x}_j, l) + 2\nu_0 r(F)\},$$

т. е. (2.15) дает

$$V_{x,l}(\varphi_{j-1} - \rho\bar{\eta}_j, \lambda_{j-1} + \rho\bar{\eta}_{Ij}) \leqslant (\nu - \nu_0) r(j)\rho + V_{x,l}(\varphi_{j-1}, \lambda_{j-1}) + \rho\{F(x, \bar{l}_j) - F(\bar{x}_j, l) + 2\nu_0 r(F)\}. \tag{2.16}$$

При $j < M_F$ (2.16) выполнено при $\rho = \rho_j$. То же верно и при $j = M_F$, если $\rho_{M_F} < \infty$. Складывая неравенства (2.16), отвечающие выбору $\rho = \rho_j$, по $j = 1, 2, \ldots, M_F - 1$ и добавляя к полученному неравенству неравенство (2.16) с $j = M_F$ и каким-нибудь ρ, удовлетворяющим (2.13) для этого j, получим, как и при доказательстве теоремы 2.2,

$$\sum_{j=1}^{M_F-1} \rho_j (F(\bar{x}_j, l) - F(x, \bar{l}_j)) + \rho (F(\bar{x}_{M_F}, l) - F(x, \bar{l}_{M_F})) \leqslant$$
$$\leqslant (\nu - \nu_0)\Big(\sum_{j=1}^{M_F-1} r(j)\rho_j + r(M_F)\rho\Big) -$$
$$- V_{x,l}(\varphi_{M_F-1} - \rho\bar{\eta}_{M_F}, \lambda_{M_F-1} + \rho\bar{\eta}_{IM_F}) +$$
$$+ 2\nu_0 r(F)\Big(\sum_{j=1}^{M_F-1} \rho_j + \rho\Big). \qquad (2.17)$$

Если $\rho_{M_F} = +\infty$, то (2.17) выполнено для некоторой стремящейся к бесконечности последовательности значений ρ. Деля обе части (2.17) на ρ и переходя к пределу по указанной последовательности значений ρ, получим, что

$$F(\bar{x}, l) - F(x, \bar{l}) \leqslant 2\nu_0 r(F) + (\nu - \nu_0) r(M_F)$$

для всех $(x, l) \in G \times G_I$. Из свойств оракула ясно, что $r(j) \leqslant r(F)$ для всех j. Таким образом, в рассматриваемом случае

$$F(\bar{x}, l) - F(x, \bar{l}) \leqslant (2\nu_0 + (\nu - \nu_0)) r(F) \leqslant 2\nu r(F),$$

откуда, как и при $r(M_F) = 0$, $\nu(\bar{x}, \bar{l}; F) \leqslant \nu$.

Пусть теперь $\rho_{M_F} \neq +\infty$. Тогда в (2.17) можно взять $\rho = \rho_{M_F}$. При этом, как и в доказательстве теоремы 2.2, левая часть (2.17) не меньше $\left(\sum_{j=1}^{M_F}\rho_j\right)(F(\bar{x}, l) - F(x, \bar{l}))$, а правая, ввиду (2.14), не больше

$$(\nu - \nu_0) \sum_{j=1}^{M_F} r(j)\rho_j + \|\varphi_{M_F}\|_* + \|\lambda_{M_F}\|_{I*} - V(\varphi_{M_F}) -$$
$$- V_I(\lambda_{M_F}) + 2\nu_0 r(F) \sum_{j=1}^{M_F} \rho_j \leqslant ((\nu - \nu_0) + (\nu - \nu_0) + 2\nu_0) r(F) \sum_{j=1}^{M_F} \rho_j.$$

Отсюда $F(\bar{x}, l) - F(x, \bar{l}) \leqslant 2\nu r(F)$, $(x, l) \in G \times G_I$, что, как и выше, дает $\nu(\bar{x}, \bar{l}; F) \leqslant \nu$.

Теорема доказана.

2.5. Обсудим полученные результаты.

2.5.1. Начнем с гильбертовой ситуации. Как и всегда, в этом случае метод может быть упрощен. Результирующая версия получается из исходной следующими изменениями:

— правило $\text{ЗС}_{V, V_I}$ 1.1 замечается на $\varphi_{i-1} = x_i = \bar{x}_i$, $\lambda_{i-1} = l_i = \bar{l}_i$;

— правило $\text{ЗС}_{V, V_I}$ 1.3 заменяется на $\bar{\eta}_i = \eta_i$, $\bar{\eta}_{I_i} = \eta_I$:

— правило $\text{ЗС}_{V, V_I}$ 1.4 выбора ρ_i заменяется на

$$\rho_i = \frac{2(\nu - \nu_0) r(i)}{\eta_i^2 + \eta_{Ii}^2}; \qquad (2.18)$$

— правило построения φ_i, λ_i заменяется на

$$\varphi_i = \pi_G(\varphi_{i-1} - \rho_i\bar{\eta}_i), \quad \lambda_i = \pi_{G_I}(\lambda_{i-1} + \rho_i\bar{\eta}_{Ii}),$$

а неравенство (2.14) — на неравенство

$$\|\varphi_i\| + \|\lambda_i\| - \frac{1}{2}\|\varphi_i\|^2 - \frac{1}{2}\|\lambda_i\|^2 \leqslant (\max_{1\leqslant j\leqslant i}(r(j)))(\nu - \nu_0) \sum_{j=1}^{i} \rho_j. \qquad (2.19)$$

Полученный метод решает всякую задачу $F \in D^{\nu_0}$ с точностью ν, $1 > \nu > \nu_0$, при трудоемкости, не превосходящей

$$\left]\frac{1}{4(\nu - \nu_0)^2}\right[+ 2 \equiv \bar{M}(\nu - \nu_0).$$

Доказательство этого утверждения проводится так же, как и доказательство теоремы 2.2. Роль $V_{x,l}$ при этом играет

$$\frac{1}{2}(\varphi - x)^2 + \frac{1}{2}(\lambda - l)^2.$$

Рекомендуем читателю проделать все необходимые выкладки. Заметим, что описанный метод совпадает с методом (2.10)—(2.11) со специальным выбором смещений и момента остановки.

2.5.2. До сих пор мы излагали версии методов, призванные решать задачи с заданной относительной погрешностью. Можно модифицировать эти версии с тем, чтобы достичь заданной абсолютной погрешности ε. Для этого достаточно заменить правую часть (2.13) на $\varepsilon\rho$, а правую часть (2.14) — на $\varepsilon \sum_{j=1}^{i} \rho_j$. (Для гильбертова случая надо заменить (2.18) на $\rho_i = 2\varepsilon/(\eta_i^2 + \eta_{Ii}^2)$, а правую часть (2.19) — на $\varepsilon \sum_{j=1}^{i} \rho_j$.) Полученные методы решают задачи $F \in D^{\nu_0}$ с абсолютной погрешностью $\varepsilon(\bar{x}, \bar{l}; F) \leqslant \varepsilon + \nu_0 r(F)$ (т. е. не превосходящей заданной с точностью до неустранимой погрешности оракула). При этом их трудоемкость не превосходит соответственно $M_{V, V_I}(\nu(F, \varepsilon))$, $\bar{M}(\nu(F, \varepsilon))$, где $\nu(F, \varepsilon) = \varepsilon/r(F)$ — (максимальная) относительная погрешность, обеспечивающая заданную абсолютную погрешность. Доказательство этих фактов аналогично предыдущему. Оно предоставляется читателю.

2.5.3. Мы видим, что решать выпукло-вогнутые игры на $E \times E_I$ «не сложнее», чем решать ЗС-методами выпуклые экстремальные задачи на «худшем» из пространств E, E_I.

§ 3. ЗС-методы решения игр: случай стохастического оракула

3.1. В этом параграфе мы покажем, что методы § 2 пригодны и для решения игр при стохастическом оракуле — для решения задач классов $\widetilde{D}^{\mathcal{O},\nu_0}_{r,L}(G \times G_I,\ E,\ \|\cdot\|,\ E_I,\ \|\cdot\|_I)$. При описании методов будем считать G и G_I нормированными условиями из начала п. 2.2. Пусть $V(\cdot)$, $V_I(\cdot)$ — те же, что и в п. 2.2, ω_i — шум оракула

$$\mathcal{O} = ((\Omega,\ \mathsf{F}_\omega),\ \psi_x(x,\ l;\ F,\ \omega)\ \psi_l(x,\ l;\ F,\ \omega))$$

на i-м шаге, $\omega^i = (\omega_1, \ldots, \omega_i)$.

Начнем с описания «бесконечношаговой» версии метода (обозначим ее $\widetilde{3C}_{V,V_I}$). Метод, решая задачу $F \in \widetilde{D}^{\mathcal{O},\nu_0}_{r,L}$, строит случайные последовательности

$$\begin{aligned}
&\{\varphi_i = \varphi_i(\omega^i) \in E^*, \quad \lambda_i = \lambda_i(\omega^i) \in E_I^*\},\\
&\{x_i = x_i(\omega^{i-1}) \in E, \quad \bar{x}_i = \bar{x}_i(\omega^{i-1}) \in G,\\
&x^i = x^i(\omega^{i-1}) \in G; \quad l_i = l_i(\omega^{i-1}) \in E_I,\\
&\bar{l}_i = \bar{l}_i(\omega^{i-1}) \in G_I, \quad l^i = l^i(\omega^{i-1}) \in G_I\}
\end{aligned}$$

в соответствии с правилами, аналогичными (2.4) — (2.5):

$$\begin{aligned}
&\varphi_0 = 0, \quad \lambda_0 = 0;\\
&x_i = V'(\varphi_{i-1}); \quad \bar{x}_i = \pi_G(x_i); \quad l_i = V_I(\lambda_{i-1}); \quad \bar{l}_i = \pi_{G_I}(l_i),\\
&\varphi_i = \varphi_{i-1} - \rho_i \{\psi_x(\bar{x}_i, \bar{l}_i; F, \omega_i) + \|\psi_x(\bar{x}_i, \bar{l}_i; F, \omega_i)\|_* \mu_G(x_i)\},\\
&\lambda_i = \lambda_{i-1} + \rho_i \{\psi_l(\bar{x}_i, \bar{l}_i; F, \omega_i) - \|\psi_l(\bar{x}_i, \bar{l}_i; F, \omega_i)\|_{I*} \mu_{G_I}(l_i)\},
\end{aligned} \quad \square \tag{3.1}$$

$$(x^i, l^i) = \frac{1}{\sum_{j=1}^{i} \rho_j} \sum_{j=1}^{i} \rho_j (\bar{x}_j, \bar{l}_j). \tag{3.2}$$

В (3.1) — (3.2) $\rho_i > 0$ — последовательность (детерминированных) смещений.

Теорема. *При выполнении условий* $\rho_i > 0$, $\lim\limits_{i\to\infty} \rho_i = 0$, $\sum\limits_i \rho_i = \infty$ *процесс* (3.1) — (3.2) *сходится (в среднем) на любой задаче* $F \in \widetilde{D}^{\mathcal{O},\nu_0}_{r,L}$ *в следующем смысле:*

$$\varlimsup_{i\to\infty} \mathsf{M}_{\mathsf{F}_{\omega^{i-1}}} \nu(x^i, l^i; F) \leqslant \nu_0.$$

При этом справедлива оценка

$$\mathsf{M}_{\mathsf{F}_{\omega^{i-1}}} \nu(x^i, l^i; F) \leqslant \nu_0 + \frac{2 + 3L\left\{\sum_{j=1}^{i} \rho_j (\omega_{V,r}(2L\rho_j) + \omega_{V_I,r}(2L\rho_j))\right\}}{2L \sum_{j=1}^{i} \rho_j}. \tag{3.3}$$

Доказательство. Пусть F — решаемая задача. Как всегда, достаточно доказать (3.3). Обозначим

$$d_i(\omega^i) = \psi_x(\bar{x}_i(\omega^{i-1}), \bar{l}_i(\omega^{i-1}); F, \omega_i) + \|\psi_x(\bar{x}_i, \bar{l}_i; F, \omega_i)\|_* \mu_G(x_i),$$
$$\bar{d}_i(\omega^{i-1}) = \mathsf{M}_{\mathsf{F}_{\omega_i}} d_i(\omega^i) = \bar{\psi}_x(\bar{x}_i, \bar{l}_i; F) + a_i(\omega^{i-1}) \mu_G(x_i),$$

где

$$\bar{\psi}_x(x,\ l;\ F) = \mathsf{M}_{\mathsf{F}_\omega} \psi_x(x,\ l;\ F,\ \omega).$$

Пусть, далее, $\tilde{d}_i(\omega^i) = d_i(\omega^i) - \bar{d}_i(\omega^{i-1})$. Тогда в силу свойства В оракула $\mathcal{O}$ (см. п. 1.2.3) и определения $a_i(\omega^{i-1})$ имеем

$$\left.\begin{aligned}
&\mathsf{M}_{\mathsf{F}_{\omega_i}} \|d_i(\omega^i)\|_*^r \leqslant L^r, \quad \mathsf{M}_{\mathsf{F}_{\omega_i}} \|\tilde{d}_i(\omega^i)\|_*^r \leqslant (2L)^r,\\
&\mathsf{M}_{\mathsf{F}_{\omega_i}} \tilde{d}_i(\omega^i) = 0, \qquad a_i(\omega^{i-1}) \geqslant \|\bar{\psi}_x(\bar{x}_i, \bar{l}_i; F)\|_*.
\end{aligned}\right\} \tag{3.4}$$

Аналогично определим

$$\delta_i(\omega^i) = \psi_l(\bar{x}_i, \bar{l}_i; F, \omega_i) - \|\psi_l(\bar{x}_i, \bar{l}_i; F, \omega_i)\|_{I*} \mu_{G_I}(l_i),$$
$$\bar{\delta}_i(\omega^{i-1}) = \mathsf{M}_{\mathsf{F}_{\omega_i}} \delta_i(\omega^i), \quad - \tilde{\delta}_i(\omega^i) = \delta_i(\omega^i) - \bar{\delta}_i(\omega^{i-1}).$$

Тогда

$$\left.\begin{aligned}
&\mathsf{M}_{\mathsf{F}_{\omega_i}} \|\delta_i(\omega^i)\|_{I*}^r \leqslant L^r, \quad \mathsf{M}_{\mathsf{F}_{\omega_i}} \|\tilde{\delta}_i(\omega^i)\|_{I*}^r \leqslant (2L)^r,\\
&\mathsf{M}_{\mathsf{F}_{\omega_i}} \tilde{\delta}_i(\omega^i) = 0.
\end{aligned}\right\} \tag{3.5}$$

Пусть теперь

$$x \in G, \quad l \in G_I, \quad V_x(\varphi) = V(\varphi) - \langle \varphi \mid x \rangle,$$
$$V_{Il}(\lambda) = V_I(\lambda) - \langle \lambda \mid l \rangle, \quad V_{x,l}(\varphi, \lambda) = V_x(\varphi) + V_{Il}(\lambda).$$

Поступая в точности так же, как и при доказательстве теоремы 2.1. гл. V, получим

$$V_x(\varphi_j(\omega^j)) \leqslant V_x(\varphi_{j-1}(\omega^{j-1})) + \rho_j \langle d_j(\omega^j) \mid x - x_j \rangle + b_j(\omega^j), \tag{3.6}$$

где в силу (3.4)

$$\mathsf{M}_{\mathsf{F}_{\omega_i}} b_i(\omega^i) \leqslant \rho_i L \omega_{V,r}(\rho_i L). \tag{3.7}$$

Далее,

$$\begin{aligned}
\langle d_i(\omega^i) \mid x - x_i \rangle &= \langle \bar{d}_i(\omega^{i-1}) \mid x - x_i \rangle + \langle \tilde{d}_i(\omega^i) \mid x \rangle - \langle \tilde{d}_i(\omega^i) \mid x_i \rangle \leqslant\\
&\leqslant \langle \bar{\psi}_x(\bar{x}_i, \bar{l}_i; F) \mid x - \bar{x}_i \rangle + \langle \tilde{d}_i(\omega^i) \mid x \rangle - \langle \tilde{d}_i(\omega^i) \mid x_i \rangle.
\end{aligned}$$

Кроме того,

$$\langle \psi_x(\bar{x}_i, \bar{l}_i; F) \,|\, x - \bar{x}_i \rangle \leqslant \nu_0 L + F(x, \bar{l}_i) - F(\bar{x}_i, \bar{l}_i).$$

Используя эти неравенства в (3.6), получаем

$$V_x(\varphi_j(\omega^j)) \leqslant V_x(\varphi_{j-1}(\omega^{j-1})) + \rho_j(\nu_0 L + F(x, \bar{l}_j) - F(\bar{x}_j, \bar{l}_j)) + \\ + \rho_j \langle \tilde{d}_j(\omega^j) \,|\, x \rangle + c_j(\omega^j), \tag{3.8}$$

где обозначено

$$c_j(\omega^j) = b_j(\omega^j) - \rho_j \langle \tilde{d}_j(\omega^j) \,|\, x_j(\omega^{j-1}) \rangle$$

и в силу (3.4) и (3.7)

$$\mathsf{M}_{\mathsf{F}_{\omega_j}} c_j(\omega^j) \leqslant \rho_j L \omega_{V,\,r}(\rho_j L). \tag{3.9}$$

Аналогично,

$$V_l(\lambda_j(\omega^j)) \leqslant V_l(\lambda_{j-1}(\omega^{j-1})) + \rho_j(\nu_0 L + F(\bar{x}_j, \bar{l}_j) - F(\bar{x}_j, l)) + \\ + \rho_j \langle \tilde{\delta}_j(\omega^j) \,|\, l \rangle + c_{I_j}(\omega^j), \tag{3.10}$$

где

$$\mathsf{M}_{\mathsf{F}_{\omega_j}} c_{I_j}(\omega^j) \leqslant \rho_j L \omega_{V_I,\,r}(\rho_j L). \tag{3.11}$$

Существенно, что c_j, c_{I_j} не зависят от параметров x и l. Складывая (3.8) и (3.10) и суммируя полученные неравенства по $j = 1, 2, \ldots, i$, получим, как и при доказательстве теоремы 2.2,

$$\Big(\sum_{j=1}^{i} \rho_j\Big)\{F(x^i(\omega^{i-1}), l) - F(x, l^i(\omega^{i-1}))\} \leqslant \\ \leqslant 1 + 2\Big(\sum_{j=1}^{i} \rho_j\Big)\nu_0 L + \Big\|\sum_{j=1}^{i} \rho_j \tilde{\delta}_j(\omega^j)\Big\|_{I*} + \Big\|\sum_{j=1}^{i} \rho_j \tilde{d}_j(\omega^j)\Big\|_* + \\ + \sum_{j=1}^{i} (c_j(\omega^j) + c_{I_j}(\omega^j)) \equiv A(\omega^i). \tag{3.12}$$

При этом $A(\omega^i)$ не зависит от x, l. Взяв максимум обеих частей (3.12) по $(x, l) \in G \times G_I$, получим

$$\Big(\sum_{j=1}^{i} \rho_j\Big) \varepsilon(x^i(\omega^{i-1}), l^i(\omega^{i-1}); F) \leqslant \frac{1}{2} + \Big(\sum_{j=1}^{i} \rho_j\Big)\nu_0 L + \\ + \frac{1}{2}\Big\|\sum_{j=1}^{i} \rho_j \tilde{\delta}_j(\omega^j)\Big\|_{I*} + \frac{1}{2}\Big\|\sum_{j=1}^{i} \rho_j \tilde{d}_j(\omega^j)\Big\|_* + \frac{1}{2}\sum_{j=1}^{i}(c_j(\omega^j) + c_{I_j}(\omega^j)). \tag{3.13}$$

Усредняя (3.13) по ω^i, получим, учитывая (3.9) и (3.11),

$$L\mathsf{M}_{\mathsf{F}_{\omega^{i-1}}} \nu(x^i, l^i; F) \leqslant \\ \leqslant \frac{1}{2\sum_{j=1}^{i} \rho_j} + \nu_0 L + \frac{1}{2\sum_{j=1}^{i} \rho_j} \mathsf{M}_{\mathsf{F}_{\omega^i}} \Big\|\sum_{1}^{i} \rho_j \tilde{\delta}_j(\omega^j)\Big\|_{I*} + \\ + \frac{1}{2\sum_{j=1}^{i} \rho_j} \mathsf{M}_{\mathsf{F}_{\omega^i}} \Big\|\sum_{1}^{i} \rho_j \tilde{d}_j(\omega^j)\Big\|_* + \\ + \frac{1}{2\sum_{j=1}^{i} \rho_j}\Big(\sum_{j=1}^{i} \rho_j L(\omega_{V,\,r}(\rho_j L) + \omega_{V_I,\,r}(\rho_j L))\Big). \tag{3.14}$$

Ввиду (3.4), для оценки $\mathsf{M}_{\mathsf{F}_{\omega^i}} \Big\|\sum_{j=1}^{i} \rho_j \tilde{d}_j(\omega^j)\Big\|_*$ можно применить предложение 1.4 гл. V, что дает оценку

$$\mathsf{M}_{\mathsf{F}_{\omega^i}} \Big\|\sum_{j=1}^{i} \rho_j \tilde{d}_j(\omega^j)\Big\|_* \leqslant \sum_{j=1}^{i} 2L\rho_j \omega_{V,\,r}(2L\rho_j) + \frac{1}{2}.$$

Аналогично оценивается и $\mathsf{M}_{\mathsf{F}_{\omega^i}} \Big\|\sum_{j=1}^{i} \rho_j \tilde{\delta}_j(\omega^j)\Big\|_{I*}$. Подставляя эти оценки в (3.14), получим требуемую оценку (3.3). Теорема доказана.

3.2. Обсудим полученный результат.

3.2.1. В гильбертовом случае при $r \geqslant 2$, как всегда, возможны упрощения. Результирующий метод имеет вид

$$\left.\begin{aligned} &x_1 = 0, \quad l_1 = 0, \\ &x_{i+1} = \pi_G(x_i - \rho_i \psi_x(x_i, l_i; F, \omega_i)); \\ &l_{i+1} = \pi_{G_I}(l_i + \rho_i \psi_l(x_i, l_i; F, \omega_i)), \end{aligned}\right\} \tag{3.15}$$

$$(x^i, l^i) = \frac{1}{\sum_{j=1}^{i} \rho_j} \sum_{j=1}^{i} \rho_j(x_j, l_j). \tag{3.16}$$

Для полученного метода верна теорема типа 3.1. При этом оценка (3.3) уточняется и имеет вид

$$\mathsf{M}_{\mathsf{F}_{\omega^{i-1}}} \nu(x^i, l^i; F) \leqslant \nu_0 + \frac{1 + \frac{1}{4} L^2 \sum_{j=1}^{i} \rho_j^2}{L \sum_{j=1}^{i} \rho_j}. \tag{3.17}$$

3.2.2. Вопрос о рациональном выборе смещений ρ_j решается так же, как и в п. 2.2.2 гл. V.

3.3. Опишем теперь «конструктивную» версию $\widetilde{\text{ЗС}}_{V,V_I}(\nu)$ метода $\widetilde{\text{ЗС}}_{V,V_I}$, призванную решать задачи из $\widetilde{\boldsymbol{D}}_{r,L}^{\mathfrak{S},\nu_0}$ с заданной (средней по реализациям метода) относительной погрешностью ν, $1>\nu>\nu_0$. Метод $\widetilde{\text{ЗС}}_{V,V_I}(\nu)$ состоит в осуществлении N шагов процесса (3.1) — (3.2), в котором все смещения равны некоторой величине $\rho>0$. Параметры N и ρ должны выбираться в функции от требуемой точности. Результатом применения указанного метода служит точка (x^N, l^N). Трудоемкость такого метода (равная числу вопросов оракулу $+$ 1 шаг на выдачу результата) равна N, поскольку построение (x^N, l^N) требует $N-1$ вопроса оракулу.

В силу оценки (3.3) для обеспечения точности $\nu>\nu_0$ можно выбирать ρ и N в виде

■ $$\rho=\frac{1}{2L}\rho_{V,V_I,r}(\nu-\nu_0),$$

где

$$\rho_{V,V_I,r}(\nu)=\min\left\{\gamma_{V,r}\left(\frac{\nu}{6}\right),\gamma_{V_I,r}\left(\frac{\nu}{6}\right)\right\},$$

$$N=N_{V,V_I,r}(\nu-\nu_0),$$

где

$$N_{V,V_I,r}(\nu)=\left]\frac{4}{\nu\rho_{V,V_I,r}(\nu)}\right[. \qquad \square \quad (3.18)$$

В гильбертовом случае при $r\geqslant 2$ следует аналогичным способом применить процесс (3.15) — (3.16), выбирая параметры ρ и N в виде

$$\rho=2\nu/L \text{ и } N=]1/\nu^2[. \qquad (3.19)$$

3.4. Сопоставляя (3.18) — (3.19) с оценкой трудоемкости ЗС-методов решения стохастических выпуклых экстремальных задач без ограничений (§ 2 гл. V), видим, что решать игры при стохастическом оракуле, грубо говоря, не сложнее, чем решать стохастические экстремальные задачи на «худшем» из пространств E, E_I. С другой стороны, задачи последнего типа суть простые частные случаи решения игр, отвечающие вырождению G (или G_I) в точку.

Приведем для справки оценки трудоемкости $\widetilde{\text{ЗС}}_{V\,V_I}(\nu)$ (т. е. функции $N_{V,V_I,r}(\nu)$) для случая, когда $(E,\|\cdot\|)=(L_p,\|\cdot\|_p)$ и $(E_I,\|\cdot\|_I)=(L_{p_I},\|\cdot\|_{p_I})$. Эти оценки получаются сопоставлением (3.18) и соотношений (1.6) — (1.7) гл. V. Пусть $p>1$, $p_I>1$. Тогда при стандартном выборе $V=V_p(\cdot)$, $V_I=V_{p_I}(\cdot)$ (см. § 2

гл. III) получим (считаем $\nu<1$)

$$N_{V,V_I,r}(\nu)\leqslant\max\left\{D(p,r)\left(\frac{1}{\nu}\right)^{\bar r(p)},\ D(p_I,r)\left(\frac{1}{\nu}\right)^{\bar r(p_I)}\right\}, \quad (3.20)$$

где $\bar r(q)=\min\{2,r,q/(q-1)\}$, а $D(q,r)<\infty$ зависит лишь от $q>1$ и $r>1$. Если $p=1$ и $\dim E\equiv n<\infty$ (и аналогично если $p_I=1$ и $\dim E_I\equiv n_I<\infty$), то в (3.20) следует заменить $D(p,r)$ на $D(1,r)\ln(n+1)$ (соответственно $D(p_I,r)$ надо заменить на $D(1,r)\ln(n_I+1)$).

Стоит заметить, что если $r\geqslant\min\{2,p/(p-1)\}$ и $r\geqslant\min\{2,p_I/(p_I-1)\}$, то трудоемкость «стохастической» версии ЗС-метода решения игр на $L_p\times L_{p_I}$ по порядку зависимости от ν и $\dim E$, $\dim E_I$ та же, что и для версии, использующей точный оракул первого порядка.

§ 4. Решение выпуклых операторных неравенств

Начиная с этого параграфа и до конца главы, мы будем заниматься применением ЗС-методов решения игр к решению разного рода выпуклых стохастических задач. Вопросы, рассматриваемые далее, можно ставить и в детерминированном случае, но там их не имеет смысла специально рассматривать (для части вопросов это уже сделано выше, в главах II, III, а остающиеся из них сразу сводятся к постановкам глав II, III).

В этом параграфе мы рассматриваем задачи решения операторных неравенств. Эти задачи интересны сами по себе, а также и ввиду того, что к ним в § 5 будут редуцированы условные задачи выпуклого стохастического программирования.

4.1. Ниже речь пойдет о приближенном решении задач вида

$$(H)\ \text{найти } x\in G, \text{ такое, что } H(x)\leqslant 0. \qquad (4.1)$$

Здесь G — замкнутое выпуклое ограниченное подмножество банахова пространства $E,\|\cdot\|$, $H\colon G\to E'$ — отображение G в банахово пространство $E',\|\cdot\|'$, причем E' снабжено структурой упорядоченного линейного пространства, так что свойство $H(x)\leqslant 0$ имеет смысл.

Упорядоченность E', с которой мы будем иметь дело, эквивалентна заданию *неотрицательного конуса* $K\subset E'$ — замкнутого непустого подмножества E', такого, что из $x\in K$, $y\in K$ следует $x+y\in K$ и $\lambda x\in K$ для всех $\lambda\geqslant 0$. Такое множество необходимо выпукло и содержит 0. Фиксация неотрицательного конуса K позволяет определить в E' отношение порядка (вернее, предпорядка) $\underset{K}{\geqslant}$. Именно, считаем $x\underset{K}{\geqslant}y$, если $x-y\in K$. Это отношение обладает естественными свойствами порядка: $x\underset{K}{\geqslant}x$ для всех x, $(x\underset{K}{\geqslant}y,\ y\underset{K}{\geqslant}z)\Rightarrow x\underset{K}{\geqslant}z$; $x\underset{K}{\geqslant}y\Rightarrow\lambda x\underset{K}{\geqslant}\lambda y$ при всех $\lambda\geqslant 0$,

$(x \underset{K}{\geqslant} y$ и $z \underset{K}{\geqslant} w) \Rightarrow x + z \underset{K}{\geqslant} y + w$. Мы будем писать $x \underset{K}{\leqslant} y$, если $y \underset{K}{\geqslant} x$. Вообще говоря, $x \underset{K}{\geqslant} y$ и $y \underset{K}{\geqslant} x$ не влечет $x = y$. Последнее верно лишь при дополнительном предположении, что $K \cap (-K) = \{0\}$.

Приведем стандартный пример такого упорядочивания. Пусть $E' = R^m$ и $K = \{x = (x', \ldots, x^m) \in R^m \mid x^i \geqslant 0\}$. Порядок $\underset{K}{\geqslant}$ есть обычный порядок $x \geqslant y \Leftrightarrow$ (для всех $i \leqslant m$ $x^i \geqslant y^i$). Если E' в (4.1) имеет указанную структуру, то (4.1) представляет собой обычную задачу решения системы скалярных неравенств.

Другой пример — это $E' = L_p(T, \mu)$ с $K = \{x = x(t) \mid x(t) \geqslant 0$ μ-почти наверное$\}$. Здесь неравенство $x \underset{K}{\geqslant} y$ есть обычное (верное μ-почти всюду) неравенство между функциями $x(t)$ и $y(t)$.

Нам потребуется ввести еще некоторые понятия. Пусть K — неотрицательный конус E и $(E')^*$ — сопряженное к E' пространство. В $(E')^*$ имеется двойственный к K конус $K^* = \{\varphi \in (E')^* \mid \langle \varphi \mid x \rangle \geqslant 0$ для всех $x \in K\}$. Легко видеть, что K^* — действительно конус. Он задает порядок $\underset{K^*}{\geqslant}$ в $(E')^*$.

Далее, пусть S: $E \to E'$ — ограниченный линейный оператор. Ему можно поставить в соответствие также ограниченный сопряженный оператор S^*: $(E')^* \to E^*$, действующий на элемент $\varphi \in (E')^*$ по правилу: $S^*\varphi$ есть функционал на E, значение которого на $x \in E$ есть $\langle \varphi \mid Sx \rangle$.

4.2. Пояснив постановку проблемы, перейдем к описанию классов задач (4.1), которые мы будем рассматривать. Это будут липшицевы выпуклые задачи типа (4.1). Уточним эти понятия.

4.2.1. Отображение H: $G \to E'$ назовем *липшицевым* с константой L, если для некоторого $L < \infty$ будет $\| H(x) - H(y) \|' \leqslant L \| x - y \|$ для всех $x, y \in G$. Назовем такое отображение *выпуклым*, если

$$H\left(\frac{x+y}{2}\right) \underset{K}{\leqslant} \frac{1}{2}(H(x) + H(y))$$

для всех $x, y \in G$. Например, если E' есть E^m со стандартным (указанным выше) порядком, то выпуклое липшицево отображение G в E' есть просто покоординатно липшицева и выпуклая m-мерная вектор-функция на G. Множество липшицевых выпуклых отображений G в E' обозначим $\boldsymbol{U} \equiv \boldsymbol{U}(G, E, \|\cdot\|, E', \|\cdot\|', K)$. Каждое такое отображение H будет отождествляться с соответствующей задачей (4.1).

4.2.2. Определим теперь меру погрешности точки $x \in G$ в качестве решения задачи $H \in \boldsymbol{U}$. Точное решение есть точка $x \in G$, такая, что $H(x) \in -K$. Естественно теперь измерять (абсолютную) погрешность точки $x \in G$ в качестве решения задачи H расстоянием от x до $(-K)$, т. е. величиной

$$\varepsilon(x, H) = \rho_{\|\cdot\|'}(H(x), \ -K). \tag{4.2}$$

Например, если E' есть R^m со стандартным упорядочением и $\|\cdot\|'$ есть $\|\cdot\|_\infty$ (так что E' есть просто $l_\infty^{(m)}$ со стандартными порядком и нормой), то при $H = (H^1(x), \ldots, H^m(x))$, $H^i(x)$ скалярны, $\varepsilon(x, H) = \max\limits_i H_+^i(x)$, где $H_+^i = \max\{0, H^i\}$. Таким образом, в этом случае погрешность есть просто максимальная по неравенствам из системы невязка в рассматриваемой точке. Мы будем писать $y \underset{K}{\leqslant} \varepsilon + z$, $y, z \in E'$, $\varepsilon \geqslant 0$, если $\rho_{\|\cdot\|'}(y - z, -K) \leqslant \varepsilon$.

4.2.3. Теперь опишем информационную базу процесса решения задачи. Пусть $\mathcal{O} = ((\Omega, \mathsf{F}_\omega); \psi(H, x, \omega))$ — оракул с пространством шумов $(\Omega, \mathsf{F}_\omega)$, функцией наблюдения $\psi(H, x, \omega) = \{\psi^0(H, x, \omega); \psi^1(H, x, \omega)\}$. Первая компонента $\psi^0(H, x, \omega)$ принимает значения в E' и служит оценкой (зависящей от реализации ω шума оракула) величины $H(x)$, а вторая компонента $\psi^1(H, x, \omega)$ принимает значения в пространстве $L(E, E')$ ограниченных линейных операторов, действующих из E в E', и служит, грубо говоря, оценкой дифференциала H в точке x. Рассматриваемые методы, как и в § 1, в состоянии на каждом шаге своей работы задать вопрос оракулу о решаемой задаче H в точке $x_i \in G$ и получить в ответ значения оценки $\psi(H, x_i, \omega_i)$, искаженной реализовавшимся на i-м шаге шумом оракула ω_i. Шумы разных шагов предполагаются независимыми в совокупности и распределенными по мере F_ω. Разумеется, $(\Omega, \mathsf{F}_\omega)$ предполагается польским пространством с регулярной борелевой полной по Лебегу вероятностной мерой, E и E' — сепарабельными, $\psi^0(H, x, \omega)$ и $\psi^1(H, x, \omega)$ — борелевыми по x и ω для всех $H \in \boldsymbol{U}$. При этом борелевость ψ^1 понимается как борелевость по x и ω всех функций вида $\langle \eta \mid \psi^1(H, x, \omega)\, e \rangle$, $e \in E$, $\eta \in (E')^*$.

4.2.4. Пусть $r > 1$, $L > 0$, $\nu_0 > 0$. Свяжем с параметрами r, L, ν_0 и данным оракулом $\mathcal{O}$ класс задач $\boldsymbol{U}_{r,L}^{\mathcal{O},\nu_0} \equiv \boldsymbol{U}_{r,L}^{\mathcal{O},\nu_0}(G, E, \|\cdot\|, E', \|\cdot\|', K)$, описываемый следующим образом.

Поле задач класса $\boldsymbol{U}_{r,L}^{\mathcal{O},\nu_0}$ образовано всеми задачами $H \in \boldsymbol{U}$, для которых верны следующие два соотношения UA и UB:

$$\left.\begin{aligned} &\| \mathsf{M}_{\mathsf{F}_\omega}\psi^0(H, x, \omega) - H(x) \|' \leqslant \frac{\nu_0 L}{2} \\ \text{и}\quad & \\ &\mathsf{M}_{\mathsf{F}_\omega}\{\psi^1(H, x, \omega)(y - x)\} \underset{K}{\leqslant} H(y) - H(x) + \nu_0 L, \end{aligned}\right\}\text{UA}$$

для всех $x, y \in G$. Смысл первого из соотношений UA ясен. Второе означает, грубо говоря, что $\mathsf{M}_{\mathsf{F}_\omega}\psi^1(H, x, \omega)$ (с точностью $\nu_0 L$)

есть опорное линейное отображение к выпуклому отображению $H(x)$ (ср. со случаем $E' = \mathbb{R}$):

$$\left.\begin{array}{l} \mathsf{M}_{\mathsf{F}_\omega}(\|\psi^0(H, x, \omega)\|')^r \leqslant \left(\dfrac{L}{4}\right)^r \quad \text{при всех } x \in G \\ \text{и} \\ \mathsf{M}_{\mathsf{F}_\omega}\|[\psi^1(H, x, \omega)]^* l\|_*^r \leqslant \left(\dfrac{L}{2\rho_{\|\cdot\|}(G)}\right)^r, \\ \text{при всех } x \in G \text{ и } l \in K_*, \text{ таких, что } \|l\|'_* \leqslant 1. \end{array}\right\} \mathrm{UB}$$

Погрешность (относительная) точки $x \in G$ в качестве приближенного решения задачи H определена как $\nu(x, H) = \frac{1}{L}\varepsilon(x, H)$. *Оракул* класса есть $\mathcal{O}$.

Далее мы будем заниматься решением задач классов $U_{r,L}^{\mathcal{O},\nu_0}$.

Упражнение 1. Пусть $E' = \mathfrak{l}_\infty^{(m)}$ с естественным порядком, так что операторы $H \in U$ представляют собой вектор-функции $H^1(x), \ldots, H^m(x)$. Докажите, что «перевод» описания класса $U_{r,L}^{\mathcal{O},\nu_0}(G, E, \|\cdot\|, \mathfrak{l}_\infty^{(m)})$ на скалярный язык дает следующее определение:

— задачи класса $U_{r,L}^{\mathcal{O},\nu_0}$ — это задачи решения систем выпуклых липшицевых неравенств $H^1(x) \leqslant 0, \ldots, H^m(x) \leqslant 0$, $x \in G$. При этом функция наблюдения оракула $\mathcal{O}$ имеет вид $(\psi_1^0(H, x, \omega), \ldots, \psi_m^0(H, x, \omega); \psi_1^1(H, x, \omega), \ldots, \psi_m^1(H, x, \omega))$, где ψ_j^0 — скалярные наблюдения $H^j(x)$, а ψ_j^1 — наблюдения опорных функционалов к H^j в точке x (т. е. функции со значениями из E^*) борелевы по x, ω. При этом борелевость ψ_j^1 по определению есть борелевость по x, ω всех функций $\langle\psi_j^1(H, x, \omega) \mid e\rangle$, $e \in E$;

— при этом $U_{r,L}^{\mathcal{O},\nu_0}$ состоит в точности из всех задач указанного вида, для которых выполнены следующие четыре условия:

$$|\mathsf{M}_{\mathsf{F}_\omega}\psi_j^0(H, x, \omega) - H_j(x)| \leqslant \frac{1}{2}\nu_0 L, \quad x \in G, \quad 1 \leqslant j \leqslant m;$$

$$H^j(y) - H^j(x) \geqslant \mathsf{M}_{\mathsf{F}_\omega}\{\langle\psi_j^1(H, x, \omega) \mid y - x\rangle - \nu_0 L\}, \quad x, y \in G, \quad 1 \leqslant j \leqslant m;$$

$$\mathsf{M}_{\mathsf{F}_\omega}\max_j\{|\psi_j^0(H, x, \omega)|\}^r \leqslant \left(\frac{L}{4}\right)^r, \quad x \in G;$$

$$\mathsf{M}_{\mathsf{F}_\omega}\|\psi_j^1(H, x, \omega)\|_*^r \leqslant \left(\frac{L}{2\rho_{\|\cdot\|}(G)}\right)^r, \quad x \in G, \quad 1 \leqslant j \leqslant m;$$

— относительная погрешность точки $x \in G$ в качестве решения задачи H определена как $\frac{1}{L}\max\{\max\limits_{1 \leqslant j \leqslant m} H^j(x), 0\}$.

4.3. Теперь опишем способ решения операторных неравенств классов $U_{r,L}^{\mathcal{O},\nu_0}$. Пусть E, $\|\cdot\|$ и $(E')^*$, $\|\cdot\|'_*$ — регулярные пространства, удовлетворяющие условиям п. 1.1 гл. V. В этих условиях мы редуцируем подлежащие решению задачи к играм.

4.3.1. Пусть $y \in E'$ и $-y \notin K$. По теореме Хана — Банаха существует функционал $\varphi \in (E')^*$ нормы 1, такой, что

$$\langle\varphi \mid y\rangle \geqslant \sup_{z \in -K}\langle\varphi \mid z\rangle + \rho_{\|\cdot\|'}(y, -K).$$

В частности, $\sup\limits_{z \in -K}\langle\varphi \mid z\rangle$ конечно, так что $\varphi \underset{K^*}{\geqslant} 0$. Но тогда $\sup\limits_{z \in -K}\langle\varphi \mid z\rangle = 0$. Итак, для всякого $y \notin -K$ существует $\varphi \in G_I = \{\varphi \in (E')^*: \|\varphi\|'_* \leqslant 1, \varphi \underset{K^*}{\geqslant} 0\}$, такое, что $\langle\varphi \mid y\rangle \geqslant \rho_{\|\cdot\|'}(y, -K)$. С другой стороны, ясно, что при $\varphi \in G_I$ $\langle\varphi \mid y\rangle \leqslant \rho_{\|\cdot\|'}(y, -K)$. При $y \in -K$, очевидно, $\sup\{\langle\varphi \mid y\rangle \mid \varphi \in G_I\} = 0$. Итак,

$$s(y) \equiv \sup\{\langle\varphi \mid y\rangle \mid \varphi \in G_I\} = \rho_{\|\cdot\|'}(y, -K) \text{ при всех } y. \tag{4.3}$$

4.3.2. Зафиксировав G_I, сопоставим задаче H задачу

$$\overline{H}: \overline{H}(x) = s(H(x)) \to \min \mid x \in G.$$

Ясно, что задача $\overline{H}$ эквивалентна задаче H в следующем точном смысле: H разрешима тогда и только тогда, когда $\overline{H}_* \equiv \inf\limits_{x \in G}\overline{H}(x) \leqslant 0$. В последнем случае всякое решение задачи $\overline{H}$ есть решение задачи H. Более того, если положить $\varepsilon(x, \overline{H}) = \overline{H}(x) - \inf\limits_G \overline{H}$, то при всяком H имеем

$$\varepsilon(x, H) \leqslant \varepsilon(x, \overline{H}) + \overline{H}_*. \tag{4.4}$$

4.3.3. Остается научиться решать задачу $\overline{H}$. Для этого заметим, что по определению $s(y)$ имеем

$$\overline{H}(x) = \sup_{l \in G_I}\langle l \mid H(x)\rangle. \tag{4.5}$$

Рассмотрим функцию $F_H(x, l) = \langle l \mid H(x)\rangle$. Легко понять, что при $H \in U$ эта функция липшицева по x, l, выпукла по x (ибо H выпукло, а $\varphi \underset{K^*}{\geqslant} 0$ при $\varphi \in G_I$) и линейна (а тогда и вогнута) по $l \in G_I$. Ввиду соотношения (4.5), решения задачи $\overline{H}$ — это в точности первые компоненты седловых точек F_H на $G \times G_I$, причем

$$\varepsilon(x, H) = \varepsilon(x, F_H). \tag{4.6}$$

Убедимся, что с помощью оракула $\mathcal{O}$ можно сконструировать оракул $\overline{\mathcal{O}}$ типа, описанного в п. 1.2.3, требуемого методами решения игр из § 3. Определим $\overline{\mathcal{O}}$ как оракул, ответы которого следующим образом строятся по ответам оракула $\mathcal{O}$:

$$\left.\begin{array}{l} \psi_x(x, l; F_H, \omega) = [\psi^1(H, x, \omega)]^* l, \\ \psi_l(x, l; F_H, \omega) = \psi^0(H, x, \omega). \end{array}\right\} \tag{4.7}$$

В последнем из соотношений (4.7) $\psi^0(H, x, \omega)$ интерпретируется как отображения со значениями в $((E')^*)^*$ (последнее пространство канонически отождествляется с E').

Покажем, что при таком определении $\overline{\mathcal{O}}$ игры F_H, порожденные задачами $H \in U_{r,L}^{\mathcal{O},\nu_0}(\cdot)$, попадают в класс

$$\widetilde{D}_{r,L}^{\overline{\mathcal{O}},\nu_0} \equiv \widetilde{D}_{r,L}^{\overline{\mathcal{O}},\nu_0}(G \times G_I, E, \|\cdot\|, (E')^*, \|\cdot\|_*').$$

Тем самым решение зедачи $\overline{H}$ (средней абсолютной) погрешности $\varepsilon > 2\nu_0 L$ можно получить следующим способом. Имитируя ответы оракула $\overline{\mathcal{O}}$ об игре F_H с помощью ответов оракула $\mathcal{O}$ об H, решаем задачу F_H с относительной точностью $\varepsilon/2L$ методом § 3. После этого первую компоненту полученной аппроксимации седловой точки F_H считаем приближенным решением задачи $\overline{H}$. Описанный метод решает задачу $\overline{H}$ с (средней абсолютной) погрешностью ε.

4.3.4. Убедимся теперь, что игра F_H действительно лежит в $\widetilde{D}_{r,L}^{\overline{\mathcal{O}},\nu_0}$. Мы должны проверить выполнение условий А и В п. 1.2.3.

Пусть $x, y \in G$, $l', l \in G_I$. Имеем

$$F_H(y, l) - F_H(x, l) = \langle l \mid H(y) - H(x)\rangle =$$
$$= \langle l \mid \mathsf{M}_{\mathsf{F}_\omega}\{\psi^1(H, x, \omega)(y - x)\}\rangle + \langle l \mid \{H(y) - H(x) - \mathsf{M}_{\mathsf{F}_\omega}\psi^1(H, x, \omega)(y - x)\}\rangle.$$

Второе слагаемое, ввиду $l \in K^*$ и $\|l\|_*' \leqslant 1$, а также в силу второго соотношения в UA, не меньше $(-\nu_0 L)$. Итак,

$$F_H(y, l) - F_H(x, l) \geqslant \langle l \mid \mathsf{M}_{\mathsf{F}_\omega}\psi^1(H, x, \omega)(y - x)\rangle - \nu_0 L =$$
$$= \langle \mathsf{M}_{\mathsf{F}_\omega}[\psi^1(H, x, \omega)]^* l \mid y - x\rangle - \nu_0 L.$$

Это есть первое из неравенств в А. Далее,

$$F_H(x, l') - F_H(x, l) = \langle l' - l \mid H(x)\rangle \leqslant$$
$$\leqslant \|l' - l\|_*' \|H(x) - \mathsf{M}_{\mathsf{F}_\omega}\psi^0(H, x, \omega)\|_*' +$$
$$+ \langle l' - l \mid \mathsf{M}_{\mathsf{F}_\omega}\psi^0(H, x, \omega)\rangle \leqslant \nu_0 L + \langle l' - l \mid \mathsf{M}_{\mathsf{F}_\omega}\psi_l(x, l;\ F_H,\ \omega)\rangle$$

(мы учли, что $\|l' - l\|_*' \leqslant 2$ ввиду $l, l' \in G_I$, а также первое из соотношений UA). Итак, А доказано.

Далее,

$$\mathsf{M}_{\mathsf{F}_\omega}\|\psi_x(x, l; F_H, \omega)\|_*^r = \mathsf{M}_{\mathsf{F}_\omega}\|[\psi'(H, x, \omega)]^* l\|_*^r \leqslant \left(\frac{L}{2\rho_{\|\cdot\|}(G)}\right)^r$$

в силу UB и ввиду того, что $l \in G_I$. Наконец, ввиду UB,

$$\mathsf{M}_{\mathsf{F}_\omega}\|\psi_l(x, l, F_H, \omega)\|_{I*}^r =$$
$$= \mathsf{M}_{\mathsf{F}_\omega}\{\|\psi^0(H, x, \omega)\|'\}^r \leqslant \left(\frac{L}{2}\right)^r \leqslant \left(\frac{L}{2\rho_{\|\cdot\|_*'}(G_I)}\right).$$

Остается заметить, что решение задачи $\overline{H}$ с абсолютной точностью $\varepsilon > 2L\nu_0$ можно свести к решению задачи F_H с относительной точностью $\varepsilon/2L$. С другой стороны, если H совместна, то решение $\overline{H}$ с абсолютной точностью ε есть в то же время и решение H с относительной точностью ε/L. Стало быть, мы умеем решать *совместные* задачи $H \in U_{r,L}^{\mathcal{O},\nu_0}$ с относительной точностью $\nu > 2\nu_0$, и решение таких задач сводится к решению игр F_H с относительной точностью $\nu/2$.

Везде выше речь, разумеется, идет о *средних* значениях точностей. Что же делать, если решаемая задача несовместна? Грубо говоря, описанная процедура делает в этом случае максимум возможного — результат $\bar{x}$ ее применения (в среднем с относительной точностью ν) таков, что $H(\bar{x})$ реализует минимально (по $x \in G$) возможное расстояние от $H(x)$ до $-K$. Таким образом, можно сказать, что описанный метод пригоден и для анализа неразрешимых задач.

Интересы дальнейшего изложения предъявляют к описанной конструкции дополнительное требование: решая задачу H, отследить величину $\overline{H}_*$. Этому требованию можно удовлетворить следующим образом. Пусть N — число шагов применяемого к F_H метода решения игр из п. 3.3, $\omega^i = (\omega_1, \ldots, \omega_i)$ — набор шумов оракула на первых $i \leqslant N$ шагах работы метода, $(\bar{x}_i, \bar{l}_i)$ — точки, в которых метод задавал вопросы о решаемой задаче, а $(\psi_l(\bar{x}_i(\omega^{i-1}), \bar{l}_i(\omega^{i-1}), F_H, \omega_i)) \equiv \bar{H}_i(\omega^i)$ — сообщенные оракулом $\overline{\mathcal{O}}$ наблюдения $H(\bar{x}_i(\omega^{i-1}))$. Образуем случайные величины

$$\bar{H}(\omega^N) = \frac{1}{N}\sum_{i=1}^{N} \bar{H}_i(\omega^i) \quad \text{и} \quad \bar{a}(\omega^N) = s(\bar{H}(\omega^N)).$$

Заметим, что после реализации работы метода, отвечающей шуму ω^N оракула, нам становится известной величина $\bar{a}(\omega^N)$. Оказывается, $\bar{a}(\omega^N)$ — достаточно хорошая оценка $\overline{H}_*$: если $\tilde{\nu} > \nu_0$ — точность, на которую настроен рассматриваемый метод решения игр, то

$$\mathsf{F}_{\omega^N}\{|\bar{a}(\omega^N) - \overline{H}_*| > 12\tilde{\nu}L\} \leqslant \frac{1}{4}. \tag{4.8}$$

Чтобы убедиться в этом, достаточно проследить за доказательством теоремы 3.1 применительно к линейной по l игре F_H. Точно так же надо действовать и применительно к гильбертовой версии метода п. 3.3.

Упражнение 2. Докажите (4.8).

⟨Складывая неравенства (3.8) и (3.10), а затем суммируя полученные неравенства по $j = 1, 2, \ldots, N$, получим, считая $\rho_j = \rho$, $1 \leqslant j \leqslant N$:

$$\sum_{j=1}^{N} \rho(F_H(\bar{x}_j(\omega^{j-1}), l) - F_H(x, \bar{l}_j(\omega^{j-1}))) \leqslant A(\omega^N), \tag{4.9}$$

где $A(\omega^N)$ не зависит от x, l и

$$M_{F_\omega N} A(\omega^N) \leqslant 2\tilde{\nu} L N \rho. \tag{4.10}$$

В качестве $A(\omega^N)$ можно взять правую часть (3.13), и (4.10) при этом будет следовать из правил выбора ρ и N для рассматриваемого метода.

Подставляя в (4.9) в качестве x какую-нибудь компоненту x^* седловой точки F_H и учитывая, что тогда при всех $l' \in G_I$ имеем

$$F_H(x^*, l') \leqslant \overline{F}_H(x^*) = \inf_{x \in G} \overline{F}_H(x) = \overline{H}_*,$$

получим из (4.9)

$$\sum_{j=1}^N \rho\,(F_H(\bar{x}_j(\omega^{j-1}), l) - \overline{H}_*) \leqslant A(\omega^N).$$

Отсюда

$$\frac{1}{N} \sum_{j=1}^N F_H(\bar{x}_j(\omega^{j-1}), l) \leqslant \overline{H}_* + \frac{1}{N\rho} A(\omega^N). \tag{4.11}$$

Вспоминая определение F_H, получаем отсюда при всех $l \in G_I$

$$\left\langle \frac{1}{N} \sum_{j=1}^N H(\bar{x}_j(\omega^{j-1})) \mid l \right\rangle \leqslant \overline{H}_* + \frac{1}{N\rho} A(\omega^N),$$

или

$$s\left(\frac{1}{N} \sum_{j=1}^N H(\bar{x}_j(\omega^{j-1}))\right) \leqslant \overline{H}_* + \frac{1}{N\rho} A(\omega^N).$$

Вместе с тем в силу выпуклости $H(x)$ имеем

$$s\left(\frac{1}{N} \sum_{j=1}^N H(\bar{x}_j(\omega^{j-1}))\right) \geqslant s\left(H\left(\frac{1}{N} \sum_{j=1}^N \bar{x}_j(\omega^{j-1})\right)\right) \geqslant \overline{H}_*.$$

Итак, при всех ω^N

$$\overline{H}_* \leqslant s\left(\frac{1}{N} \sum_{j=1}^N H(\bar{x}_j(\omega^{j-1}))\right) \leqslant \overline{H}_* + \frac{1}{N\rho} A(\omega^N). \tag{4.12}$$

В то же время в силу свойств оракула $\mathscr{O}$ и п. 2 предложения 1.5 гл. V имеем

$$M_{F_\omega N} \left\| \sum_{j=1}^N \rho\,(H(\bar{x}_j(\omega^{j-1})) - \widetilde{H}_j(\omega^j)) \right\|' \leqslant \frac{\rho}{2} \nu_0 L N + \frac{N}{2} \rho L \omega_{V_I, r}\left(\frac{\rho L}{2}\right) + \frac{1}{2},$$

откуда в силу липшицевости s относительно $\|\cdot\|'$ с константой 1

$$M_{F_\omega N} \left| s\left(\frac{1}{N} \sum_{j=1}^N H(\bar{x}_j(\omega^{j-1}))\right) - s\left(\frac{1}{N} \sum_{j=1}^N \widetilde{H}_j(\omega^j)\right) \right| \leqslant$$
$$\leqslant \frac{\nu_0 L}{2} + \frac{L}{2} \omega_{V_I, r}\left(\frac{\rho L}{2}\right) + \frac{1}{2N\rho} \leqslant \tilde{\nu} L \tag{4.13}$$

(мы учли правило выбора ρ и N из п. 3.3). Соединяя (4.12) с (4.13), получаем, используя (4.10),

$$M_{F_\omega N} \{|\overline{H}_* - \tilde{a}(\omega^N)|\} \leqslant 3\tilde{\nu} L. \tag{4.14}$$

Ясно, что (4.8) есть прямое следствие (4.14).

З а м е ч а н и е. Константа в левой части (4.8) несколько завышена. Мы могли бы понизить ее, действуя более аккуратно, но не делаем этого, чтобы не загромождать изложения. Аналогичным образом мы будем действовать и дальше. Дело в том, что в этом и следующих параграфах нас интересует принципиальная сторона дела, а не непосредственное построение практических рекомендаций. При необходимости читатель сам сможет провести надлежащие уточнения констант и снизить за счет этого трудоемкость описываемых методов.

§ 5. Решение экстремальных задач с операторными ограничениями

5.1. Рассмотрим задачи, более общие, чем (4.1),— выпуклые экстремальные задачи с операторными ограничениями вида

$$f \equiv (f_0, H_f): \quad f_0(x) \to \min \mid x \in G, \quad H_f(x) \leqslant 0. \tag{5.1}$$

Здесь $H_f\colon G \to E'$ предполагается отображением из множества $\boldsymbol{U}(G, E, \|\cdot\|, E', \|\cdot\|', K) \equiv \overline{\boldsymbol{U}}$, а f_0 — выпуклой липшицевой вещественнозначной функцией, определенной на G. Относительно $E, \|\cdot\|$ и $E', \|\cdot\|'$ будем предполагать выполнение условий начала п. 4.3; G, как всегда, считается непустым выпуклым замкнутым ограниченным подмножеством E. Выпуклые условно-экстремальные задачи являются частным случаем задач (5.1), получающимся при $E' = l_\infty^{(m)}$ и естественном упорядочивании этого пространства. Поле всех задач указанного выше типа будем обозначать

$$\boldsymbol{W} \equiv \boldsymbol{W}(G, E, \|\cdot\|, E', \|\cdot\|', K).$$

5.2. Наша цель — научиться решать задачи из $\boldsymbol{W}$, используя стохастический оракул первого порядка. Этот оракул будет, грубо говоря, сообщать (искаженные помехой) наблюдения значений и дифференциалов целевого функционала f_0 и ограничения H. Будем рассматривать оракул $\mathscr{O} = ((\Omega, F_\omega);\ \varphi^0(x, f, \omega),\ \varphi^1(x, f, \omega);\ \psi^0(x, f, \omega),\ \psi^1(x, f, \omega))$, у которого φ^0 скалярно, φ^1 принимает значения в E^*, ψ^0 принимает значения в E' и ψ^1 — в $L(E, E')$. При этом φ^0 есть оценка f_0, φ^1 — оценка опорного к f_0 функционала, ψ^0 — оценка H_f и ψ^1 — оценка дифференциала H_f. Как всегда, на i-м шаге работы метода можно обратиться к оракулу $\mathscr{O}$ с вопросом о решаемой задаче в точке $x_i \in G$ и получить в ответ значения

$$\varphi^0(x_i, f, \omega_i), \quad \varphi^1(x_i, f, \omega_i), \quad \psi^0(x_i, f, \omega_i), \psi^1(x_i, f, \psi_i).$$

Здесь ω_i — независимые и распределенные по мере F_ω шумы оракула. Естественно, Ω, F_ω предполагается польским пространством с регулярной борелевой полной по Лебегу вероятностной мерой, а φ^0, φ^1, ψ^0, ψ^1 — борелевыми по x, ω функциями.

Опишем теперь классы задач, которыми мы будем заниматься. Фиксируем $r > 1$ и $L > 0$, а также $\nu_0 > 0$ и обозначим через

$$\widetilde{W}^{\mathcal{O},\nu_0}_{r,L} \equiv \widetilde{W}^{\mathcal{O},\nu_0}_{r,L}(G, E, \|\cdot\|, E' \|\cdot\|', K)$$

класс задач вида (5.1), описываемый следующим образом.

Поле задач класса $\widetilde{W}^{\mathcal{O},\nu_0}_{r,L}$ состоит из всех задач

$$f = (f_0, H_f) \in W\ (G, E, \|\cdot\|, E', \|\cdot\|', K),$$

для которых

(1) отображение H_f попадает в класс $U^{\overline{\mathcal{O}},\nu_0}_{r,L}$ $(G, E, \|\cdot\|, E', \|\cdot\|', K)$ (см. п. 4.2.3), где $\overline{\mathcal{O}}$ — индуцированный $\mathcal{O}$ оракул $((\Omega, F_\omega); \psi^0, \psi^1)$.

(2)
$$\left.\begin{aligned} |M_{F_\omega}\varphi^0(x, f, \omega) - f_0(x)| &\leqslant \frac{\nu_0 L}{2};\\ M_{F_\omega}|\varphi^0(x, f, \omega)|^r &\leqslant \left(\frac{L}{4}\right)^r, \end{aligned}\right\} \tag{5.2}$$

$$\left.\begin{aligned} f_0(y) - f_0(x) &\geqslant \langle M_{F_\omega}\varphi^1(x, f, \omega) \mid y - x\rangle - \nu_0 L;\\ M_{F_\omega}\|\varphi^1(x, f, \omega)\|_*^r &\leqslant \left(\frac{L}{2\rho_{\|\cdot\|}(G)}\right)^r \end{aligned}\right\} \tag{5.3}$$

для всех $x, y \in G$.

Условия (1) — (2) — это стандартного для наших рассмотрений типа требования к возможному смещению и моментам стохастической информации, поставляемой оракулом.

Оракул класса есть $\mathcal{O}$.

Мера абсолютной погрешности точки $x \in G$ в качестве решения задачи f определена как

$$\varepsilon(x, f) = \max\ \{f_0(x) - f_*;\ s\ (H_f(x))\}.$$

Здесь

$$f_* = \begin{cases} +\infty, & \text{для всех } x \in G \quad H(x) \not\leqslant_K 0,\\ \inf\{f(x) \mid x \in G, H(x) \leqslant 0\} & \text{в противном случае} \end{cases}$$

— оптимальное значение целевого функционала задачи f, $s(x)$ — функция, введенная в п. 4.3.

Мера относительной погрешности точки $x \in G \cup \{*\}$ в качестве приближенного решения задачи f ($*$, как обычно, интерпрети-

руется как ответ о несовместности f) определена формулой

$$\nu(x, f) = \begin{cases} 0, & x = *,\ f \text{ несовместна},\\ 1, & x = *,\ f \text{ совместна},\\ \dfrac{1}{L}\varepsilon(x, f) & \text{в случае } x \in G. \end{cases}$$

5.3. Фиксируем класс задач

$$\widetilde{W}^{\mathcal{O},\nu_0}_{r,L} \equiv \widetilde{W}^{\mathcal{O},\nu_0}_{r,L}(G, E, \|\cdot\|, E', \|\cdot\|', K)$$

и построим метод решения задач этого класса с относительной точностью ν. Будем считать, что погрешность оракула ν_0 достаточно мала по сравнению с ν, скажем,

$$800\nu_0 < \nu. \tag{5.4}$$

Идея реализуемой ниже конструкции состоит в редукции оптимизационной задачи (5.1) к решению серии операторных неравенств. Объясним, как это делается. При описании удобно считать, что $\rho_{\|\cdot\|}(G) = 1$. Это мы отныне и предполагаем.

5.3.1. Начнем с самого простого случая: требуется решить задачу $f \in \widetilde{W}^{\mathcal{O},\nu_0}_{r,L}$, причем априори известно, что $f_* = 0$. В этом случае решение задачи f сразу сводится к решению операторного неравенства — именно, системы неравенств

$$f_0(x) \leqslant 0,\ \ H_f(x) \leqslant 0,\ \ x \in G. \tag{5.5}$$

В свою очередь, (5.5) можно записать в виде

$$H^f(x) \leqslant 0, x \in G, \tag{5.6}$$

где $H^f = (f_0, H_f)$ есть отображение G в $\widetilde{E} = \mathbb{R} \times E'$, а порядок в $\widetilde{E}$ задается конусом $\widetilde{K} = \{(t, y) \mid t \geqslant 0,\ y \in K\}$. При этом, очевидно,

$$H^f \in U\ (G,\ E, \|\cdot\|, \widetilde{E}, \|\cdot\|^\sim, \widetilde{K}) \equiv U;$$

в качестве $\|t, y\|^\sim$ условимся выбирать $\max\{|t|, \|y\|'\}$.

Задача H^f эквивалентна задаче f в следующем смысле: H^f совместна одновременно с f и для всякого $x \in G$ $\varepsilon(x, H^f) = \varepsilon(x, f)$. Таким образом, решить задачу f с заданной средней (абсолютной) точностью ε — это все равно, что решить в том же смысле задачу H^f.

Посмотрим теперь, располагаем ли мы источником информации о задаче H^f, требуемым методом из § 4. Очевидно, ответ на этот вопрос положителен. Действительно, компоненты φ^0, ψ^0 функции наблюдения оракула $\mathcal{O}$ вместе образуют нужную оценку $H^f(x)$, а компоненты φ^1, ψ^1 — нужную оценку дифференциала H^f. Таким образом, оракул $\mathcal{O}$ можно интерпетировать как оракул для поля задач U из п. 42.3. При этом сопоставление определения класса задач $\widetilde{U}^{\mathfrak{S},\nu_0}_{r,L}$ и класса задач $\widetilde{W}^{\mathfrak{S},\nu_0}_{r,L}$ приводит к включению $H^f \in \widetilde{U}^{\mathfrak{S},\nu_0}_{r,2L}$. При этом $\bar{E}$ регулярно вместе с $(E')^*$, так что к задаче H^f применим метод § 4. Решая H^f этим методом, настроенным на среднюю относительную точность $\nu/2$, получим решение задачи f средней относительной точности ν.

Ясно, что если $f_* < \infty$ известно априори (но не обязательно равно 0), то можно действовать так же, как и в случае $f_* = 0$, но применительно к задаче $\bar{f} = \{f_0 - f_*, H_f\}$. Единственное отличие от предыдущего случая состоит в том, что соответствующее операторное неравенство, как легко видеть, попадет *), вообще говоря, в $\widetilde{U}^{\mathfrak{S},\nu_0}_{r,4L}$, а не в $\widetilde{U}^{\mathfrak{S},\nu_0}_{r,2L}$, так что решение задачи f с относительной точностью ν потребует применения к H^f метода, настроенного на точность $\nu/4$.

5.3.2. Рассмотрим теперь общий случай, когда f_* априори неизвестно. Можно попытаться вначале найти f_*, а затем поступить в соответствии с рекомендациями п. 5.3.1. Опишем процедуру отыскания f_*. Пусть $f \in \widetilde{W}^{\mathfrak{S},\nu_0}_{r,L}$. Для всякого $t \in \mathbb{R}$ пусть f^t есть задача $(f_0 - t, H_f)$. Очевидно, задача f совместна и только тогда, когда при некотором t совместна задача H^{f^t}. При этом

$$f_* = \max \{t \mid H^{f^t} \text{ совместна}\}.$$

Заметим, что при $2|t| \leqslant L$ задача H^{f^t} попадает в класс $U^{\mathfrak{S},\nu_0}_{r,4L}$. Если теперь применить к H^{f^t} описанный в § 4 метод решения задач этого класса, настроенный на среднюю относительную точность $\tilde{\nu} > 2\nu_0$ (обозначим этот метод $\mathcal{B}_{\tilde{\nu}}$), то в соответствии со сказанным в п. 4.4, получим (случайный) результат $\bar{x}_t \in G$, для которого

$$\mathsf{M}\varepsilon(\bar{x}_t, H^{f^t}) \leqslant 4\tilde{\nu}L + \overline{H^{f^t}_*}, \tag{5.7}$$

а также (случайную) оценку $\tilde{a}_t$ числа $\overline{H^{f^t}_*}$, для которой

$$\mathsf{Pr}\{|\tilde{a}_t - \overline{H^{f^t}_*}| > 48\tilde{\nu}L\} \leqslant {}^1/_4. \tag{5.8}$$

Здесь M означает среднее, а Pr — вероятность, вычисленные по распределению реализаций $\mathcal{B}_{\tilde{\nu}}$ на задаче H^{f^t} (см. (4.4), (4.8)).

Пусть, далее, $g(t) = \overline{H^{f^t}_*}$. По определению $\varepsilon(x, f)$ имеем

$$\varepsilon(x, f) = \max\ \{f_0(x) - f_*,\ s(H_f(x))\} \leqslant [t - f_*]_+ + \\ + \max\{f_0(x) - t,\ s(H_f(x))\} \leqslant [t - f_*]_+ + \varepsilon(x, H^{f^t}).$$

*) При несущественной для дальнейшего оговорке $\nu_0 < {}^1/_2$.

Отсюда и из (5.7) получаем

$$\mathsf{M}\ \varepsilon(\bar{x}_t, f) \leqslant [t - f_*]_+ + g(t) + 4\tilde{\nu} L, \tag{5.9}$$

или

$$\mathsf{M}\nu(\bar{x}_t, f) \leqslant \frac{[t - f_*]_+}{L} + \frac{g(t)}{L} + 4\tilde{\nu}. \tag{5.10}$$

Отсюда ясно, что для решения задачи f с (средней относительной) точностью ν достаточно отыскать аппроксимацию t^* величины f_*, такую, что

$$\frac{1}{L}(t^* - f_*) = O(\nu) \quad \text{и} \quad \frac{1}{L} g(t^*) = O(\nu),$$

после чего применить к $H^{f^{t^*}}$ метод $\mathcal{B}_{\tilde{\nu}}$ с $\tilde{\nu} = O(\nu)$.

Остается понять, как найти требуемое t^*. Заметим, что $g(t)$ в силу своего определения липшицева с константой 1 невозрастающая функция. При этом f совместна $\Leftrightarrow$ уравнение $g(t) = 0$ разрешимо, и в последнем случае f^* есть наименьший из корней этого уравнения. Из сказанного ясно, что в качестве t^* можно выбирать достаточно точную аппроксимацию минимального корня уравнения $g(t) = 0$. Последнюю можно получить «делением отрезка пополам». При этом для определения значений $g(t)$ можно использовать оценку $\tilde{a}_t$, доставляемую применением $\mathcal{B}_{\tilde{\nu}}$ (с подходящим $\tilde{\nu}$) к H^{f^t}.

5.3.3. Реализуем намеченный план. Опишем процедуру $\mathcal{B}^\nu$ решения задач из $\widetilde{W}^{\mathfrak{S},\nu_0}_{r,L}$ с заданной (средней относительной) точностью ν, удовлетворяющей (5.4). Заметим, что из определения класса следует, что при $f \in \widetilde{W}^{\mathfrak{S},\nu_0}_{r,L}$ изменение функций f_0 и максимум функции $s(H_f)$ на G не превосходят $(1 + 2\nu_0)\,L/2$, так что интерес представляет лишь случай $\nu < 1 + \nu_0$.

Итак, считаем, что заданная точность ν удовлетворяет условию

$$800\nu_0 < \nu < 1 + \nu_0. \tag{5.11}$$

В этом предположении процедура $\mathcal{B}^\nu$ в применении к задаче $f \in \widetilde{W}^{\mathfrak{S},\nu_0}_{r,L}$ состоит из $N_1 \equiv N_1(\nu)$ поисковых этапов и одного заключительного (вид функции $N_1(\nu)$ будет указан позднее). К i-му поисковому этапу имеются точки $\underline{t}_{i-1}$, $\bar{t}_{i-1}$ — очередные границы отрезка, локализующего t^*. При этом $\underline{t}_0 = -L/2$, $\bar{t}_0 = L/2$.

На i-м поисковом этапе строится точка $t_i = {}^1/_2\,(\underline{t}_{i-1} + \bar{t}_{i-1})$. Далее к задаче $H^{f^{t_i}}$ $N_2 \equiv N_2(\nu)$ раз применяется процедура решения операторных неравенств $\mathcal{B}_{\tilde{\nu}}$ с неким $\tilde{\nu} > \nu_0$. Вид функции $N_2(\nu)$ и значение $\tilde{\nu}$ будут указаны ниже.

При j-м применении $\mathcal{B}_{\tilde{\nu}}$ к $H^{f^{t_i}}$ фиксируется наблюденная реализация $\tilde{a}^j_{t_i}$ оценки $\tilde{a}$ величины $\overline{H^{f^{t_i}}_*} = g(t_i)$. После получения

$\tilde{a}^1_{t_i}, \ldots, \tilde{a}^{N_2}_{t_i}$ строится число

$$\gamma_i = \begin{cases} 1, & \text{если более половины } \tilde{a}^j_{t_i} \text{ больше } 50\tilde{\nu}L, \\ 0 & \text{в противном случае.} \end{cases}$$

Затем строятся числа

$$\underline{t}_i = \begin{cases} \underline{t}_{i-1}, & \gamma_i = 0, \\ t_i, & \gamma_i = 1, \end{cases} \qquad \bar{t}_i = \begin{cases} t_i, & \gamma_i = 0, \\ \bar{t}_{i-1}, & \gamma_i = 1. \end{cases}$$

Шаг закончен.

При $i < N_1$ перейдем к следующему поисковому этапу, а при $i = N_1$ — к заключительному этапу. В последнем случае положим $t^* = \underline{t}_{N_2}$.

На заключительном этапе проверяется, было ли все время (т. е. при $1 \leqslant i \leqslant N_1$) $\gamma_i = 1$. Если да, то полагаем результат работы $\mathcal{B}^\nu$ на f равным $*$. В противном случае применяем $\mathcal{B}_{\tilde{\nu}}$ к H^{tt^*} и полученный результат объявляем результатом применения $\mathcal{B}^\nu$ к f.

5.3.4. Проанализируем работу метода $\mathcal{B}^\nu$ и покажем, что при надлежащем выборе $N_1(\nu)$, $N_2(\nu)$ будет действительно обеспечена заданная средняя точность в решении любой задачи $f \in \widetilde{W}^{\mathfrak{S}, \nu_0}_{r, L}$. Фиксируем f.

1°. Скажем, что на i-м поисковом этапе имел место *успех*, если не произошло ни одно из следующих двух событий:

— было $g(t_i) \leqslant 2\tilde{\nu}L$ и $\gamma_i = 1$;

— было $g(t_i) > 100\,\tilde{\nu}L$ и $\gamma_i = 0$.

В противном случае i-й этап назовем *неуспешным*.

2°. Фиксируем реализацию шумов оракула на всех N_2 поисковых этапах. Допустим, что она такова, что выполнено условие:

(А) каждый из N_2 поисковых этапов успешен. Убедимся, что если N_1 и $\tilde{\nu}$ удовлетворяют соотношениям

$$2\tilde{\nu} - 2^{-N_1} \geqslant \nu_0, \quad 100\tilde{\nu} + 2^{-N_1} + 4\tilde{\nu} \leqslant \frac{1}{2}\nu, \tag{5.12}$$

то при условии (А) средняя (по реализациям шума оракула на заключительном этапе работы метода) погрешность результата не превосходит $\nu/2$.

Действительно, пусть выполнено условие (А). Возможно, что при этом было $\gamma_i = 1$, $1 \leqslant i \leqslant N_1$. Последнее при условии (А) означает, что при всех i было $g(t_i) > 2\tilde{\nu}\,L$ и при этом

$$0 < \frac{1}{2}L - t_{N_1} \leqslant L \cdot 2^{-N_1}.$$

Так как $g(t)$ липшицева с константой 1, то тогда

$$g\left(\frac{L}{2}\right) > 2\tilde{\nu}L - L \cdot 2^{-N_1} \geqslant \nu_0 L$$

(мы использовали первое из соотношений (5.12)). Легко видеть, что последнее неравенство *при совместной* f невозможно (надо учесть, что из определения класса сразу следует $|f_0| \leqslant (1 + 2\nu_0)\,L/4$ на G). Итак, в рассматриваемом случае выполнения (А) и $\gamma_i \equiv 1$ средняя погрешность результата работы метода равна 0.

Пусть теперь выполнено (А), но не все $\gamma_i = 1$. Тогда t^* определено, и из определения t^* и выполнения (А) следует, что

— существует некоторое i_0, такое, что $t_{i_0} \geqslant t^*$ и $\gamma_{i_0} = 0$. При этом $t_{i_0} - t^* \leqslant L \cdot 2^{-N_1}$;

— либо существует такое i_1, что $t_{i_1} = t^*$ и $\gamma_{i_1} = 1$ (случай 1), либо $t^* = -L/2$ (случай 2).

В силу первого из этих замечаний и условия (А) имеем $g(t_{i_0}) \leqslant 100\,\tilde{\nu}L$. Но тогда

$$g(t^*) \leqslant g(t_{i_0}) + t_{i_0} - t^* \leqslant 100\tilde{\nu}L + L \cdot 2^{-N_1}. \tag{5.13}$$

Далее, в случае 1 $g(t^*) > 0$, т. е. $t^* \leqslant f_*$. В случае 2, очевидно, $t^* \leqslant f_*$ (мы учли, что

$$f_* \geqslant -\frac{L}{4}(1 + 2\nu_0), \quad \nu_0 < {}^1/_2).$$

Итак, во всех случаях

$$t^* - f_* \leqslant 0. \tag{5.14}$$

Соединяя (5.13), (5.14) и (5.10), видим, что в условиях (5.12) при выполнении (А) средняя погрешность доставляемого методом результата действительно не превышает $\nu/2$.

Заметим, что неравенства (5.12) выполняются, если выбрать

$$\tilde{\nu} = c_0\nu, \quad N_1(\nu) = c_1 \ln\frac{2}{\nu} \tag{5.15}$$

с подходящими *абсолютными* константами c_0 и c_1 (при этом $c_0 \geqslant 1/200$, так что $\tilde{\nu} > 4\nu_0$). Мы могли бы выписать эти константы, но не станем этого делать, чтобы не загромождать изложения.

3°. Пусть теперь $\Pr(A) = p$. Тогда средняя относительная погрешность $\nu(\mathcal{B}^\nu, f)$ метода $\mathcal{B}^\nu$ на задаче f в условиях (5.15) не превосходит $\frac{\nu}{2} + (1 + \nu_0)(1 - p)$. Действительно, в этом случае средняя условная (при фиксированном шуме на поисковых этапах) погрешность метода не превосходит $\nu/2$ при условии (А) и $(1 + \nu_0)$ — во всех остальных случаях. Чтобы построить метод заданной точности, достаточно добавить к условиям (5.15) такое правило выбора $N_2(\nu)$, что при его использовании

$$1 - p \leqslant \frac{\nu}{2(1 + \nu_0)} \left(\geqslant \frac{\nu}{4}\right) \tag{5.16}$$

(мы учли, что из (5.4) и условия $\nu < \nu_0 + 1$ следует, конечно, $\nu_0 < 1$).

Докажем, что (5.16) обеспечивается выбором

$$N_2(\nu) = c_2 \ln \frac{2}{\nu} \tag{5.17}$$

с подходящей абсолютной константой c_2. Таким образом, выбор $\tilde{\nu}$, N_1, N_2 в соответствии с (5.15) — (5.17) позволяет получить метод решения задач из $\widetilde{\boldsymbol{W}}_{r,L}^{\mathfrak{G},\nu_0}$ точности ν. Оценим $1-p$ следующим образом. Исследуем i-й поисковый этап. Рассмотрим условное распределение серии оценок $\{\tilde{a}_{t_i}^j\}_{j=1}^{N_2}$, получающееся при фиксации «предыстории» процесса решения вплоть до i-го этапа. Ясно, что при указанном условии $\tilde{a}_{t_i}^j$ независимы в совокупности и события $A_j = \{|\tilde{a}_{t_i}^j - g(t_i)| > 48\,\tilde{\nu}\,L\}$ имеют каждое вероятность $\leqslant 1/4$. Тогда одновременное наступление $\geqslant 1/2\, N_2$ из этих событий возможно с вероятностью $\leqslant e^{-\varkappa N_2}$, $\varkappa > 0$ — абсолютная константа. Ясно, что если реализовалось менее половины событий A_j, то i-й этап успешен. Из сказанного следует, что вероятность одновременного успеха всех N_1 поисковых этапов (т. е. вероятность события A) $\geqslant 1 - N_1 e^{-\varkappa N_2}$, что доставляет оценку $1 - p \leqslant N_1 e^{-\varkappa N_2}$. Вспоминая определение $N_1(\nu)$, находим, что при надлежащем выборе c_2 (5.17) влечет (5.16).

5.4. Далее, говоря о методе решения задач из $\widetilde{\boldsymbol{W}}_{r,L}^{\mathfrak{G},\nu_0}$ с точностью ν, будем иметь в виду метод $\mathscr{B}^\nu$, доставляемый описанной выше конструкцией с параметрами (5.15) — (5.17). Трудоемкость этого метода в условиях (5.11) оценивается сверху функцией

$$M_w(\nu) = c\left]\ln^2 \frac{1}{\nu}\right[\ \widetilde{\mathscr{M}}\left(\frac{\nu}{d}\right).$$

Здесь c и d — абсолютные константы, а $\widetilde{\mathscr{M}}(\nu)$ — трудоемкость настроенного на точность ν ЗС-метода решения игр класса типа $\widetilde{\boldsymbol{D}}_{r,L}^{\cdot,0}(G \times G_I;\ E, \|\cdot\|,\ \bar{E}, \|\cdot\|^{\sim})$.

§ 6. Решение условных стохастических задач

6.1. Результаты предыдущего параграфа почти непосредственно приложимы к проблеме решения условно-экстремальных выпуклых стохастических задач (задач классов $\widetilde{\boldsymbol{C}}_{r,L}^{\mathfrak{G},\nu_0}(G, E, \|\cdot\|, m)$), введенных в § 1 гл. V. Действительно, каждую такую задачу f: $\{f_0(x) \to \min \mid x \in G \subset E,\ f_j(x) \leqslant 0,\ 1 \leqslant j \leqslant m\}$ можно рассматривать как задачу с операторным ограничением вида $f: f_0(x) \to \min \mid x \in G,\ \ H_f(x) \leqslant 0,$ где $H_f(x) = (f_1(x), \ldots, f_m(x))$ — отображение E в $E' = l_\infty^{(m)}$ ($l_\infty^{(m)}$ снабжено естественным упорядочением). Ответы оракула $\mathcal{O}$ о задаче f можно интерпретировать как ответы рассмотренного в § 5 оракула о задаче (f_0, H_f). Иными словами, поле задач $\boldsymbol{C}_{\mathrm{Lip}}(G, E, \|\cdot\|, m)$ непосредственно отождествляется с полем задач $\boldsymbol{W} \equiv \boldsymbol{W}(G, E, \|\cdot\|, l_\infty^{(m)},$

$\|\cdot\|_\infty, K_\infty)$, где K_∞ — стандартный конус в $l_\infty^{(m)}$, а оракул $\mathcal{O}$ можно считать оракулом, рассмотренным в п. 5.2 для поля задач $\boldsymbol{W}$.

Абсолютная погрешность точки $x \in G$ в качестве решения задачи $f \in \boldsymbol{C}_{\mathrm{Lip}}$ — величина $\max\ \{f_0(x) - f_*, f_1(x), \ldots, f_m(x)\} \equiv \varepsilon(x, f)$ — есть в то же время и абсолютная погрешность точки x в качестве решения той же самой задачи f, рассматриваемой как элемент $\boldsymbol{W}$. Таким образом, для решения задач классов $\widetilde{\boldsymbol{C}}_{r,L}^{\mathfrak{G},\nu_0}$, казалось бы, можно применять методы решения задач классов $\widetilde{\boldsymbol{W}}_{r,L}^{\mathfrak{G},\nu_0}$.

Однако без специальных дополнительных преобразований исходной задачи осуществить эту редукцию нельзя. Дело в том, что задача $f \in \widetilde{\boldsymbol{C}}_{r,L}^{\mathfrak{G},\nu_0}$ не попадает, вообще говоря, не только в класс $\widetilde{\boldsymbol{W}}_{r,L}^{\mathfrak{G},\nu_0}$, но даже и в класс $\widetilde{\boldsymbol{W}}_{r,\bar{L}}^{\mathfrak{G},\bar{\nu}_0}$ с $\bar{\nu}_0$ и $\bar{L}$, в абсолютную константу раз бо́льшими ν_0 и L соответственно. Связанные с определением класса $\widetilde{\boldsymbol{C}}_{r,L}^{\mathfrak{G},\nu_0}$ ограничения на наблюдения значений $f_j(x)$ касались только отклонений этих наблюдений от истинных значений, тогда как ограничения на те же наблюдения, связанные с определением классов $\widetilde{\boldsymbol{W}}_{r,\bar{L}}^{\mathfrak{G},\bar{\nu}_0}$, связаны как с величинами отклонений, так и с величинами самих наблюдений.

Упражнение 1. Укажите, где в §§ 4—5 использовались связанные с определением соответствующих классов априорные ограничения на величины наблюдений значений компонент задачи.

⟨*Ответ.* Необходимо было иметь возможность свести дело к игре, у которой эффективно оцениваются наблюдения опорных функционалов по обеим компонентам.⟩

6.2. Указанная выше трудность легко преодолима. Для этого мы применим описываемую далее процедуру препарирования. Как везде в этой главе, предполагаем G, E, $\|\cdot\|$ удовлетворяющими условиям из начала § 1 гл. V.

Пусть требуется решить задачу $f \in \widetilde{\boldsymbol{C}}_{r,L}^{\mathfrak{G},\nu_0}(G, E, \|\cdot\|, m)$. Обозначим через $\bar{x}$ «центр» множества G. Допустим, что мы располагаем семейством статистических оценок a_δ, $\delta > 0$ вектора $f(\bar{x})$, таких, что

$$\Pr\{\|a_\delta - f(\bar{x})\|_\infty > L\} \leqslant \delta. \tag{6.1}$$

Предположим, что оценка a_δ строится по $R(\delta)$ наблюдениям вектора $f(\bar{x})$, доставляемым оракулом $\mathcal{O}$ (которому нужно задать $R(\delta)$ вопросов о задаче f в точке $\bar{x}$). Фиксируем $\delta > 0$ и рассмотрим строящуюся по реализованному значению $a = (a^0, \ldots, a^m)$ оценки a_δ задачу $\bar{f} = (\bar{f}_0, \ldots, \bar{f}_m): \bar{f}_0(x) = f_0(x) - a^0$ и для $1 \leqslant j \leqslant m$

$$\bar{f}_j(x) = \begin{cases} L, & \text{если } a^j > 3L, \\ 0, & \text{если } a^j < -3L, \\ f_j(x), & \text{если } -3L \leqslant a^j \leqslant 3L. \end{cases}$$

Заметим, что задача $\tilde{f}$ *является* случайным преобразованием задачи f, определяемым реализованным значением a оценки a_δ. Будем называть это преобразование *препарированием*. Скажем, что *препарирование* прошло *успешно*, если выполнено условие $\|a - f(\bar{x})\|_\infty \leqslant L$.

Упражнение 2. Пусть препарирование задачи $f \in \widetilde{C}^{\mathscr{O}, \nu_0}_{r, L}(G, E, \|\cdot\|, m)$ прошло успешно. Докажите, что при $\nu_0 < {}^1/_2$

(i) если при некотором $j \geqslant 1$ было $a^j > 3L$, то f и $\tilde{f}$ обе несовместны; независимо от значений a^j совместность одной из этих задач влечет совместность другой;

(ii) если при всех $j \geqslant 1$ имеем $a^j \leqslant 3L$, то для всех $x \in G$ $\varepsilon(x, f) \leqslant \varepsilon(x, \tilde{f})$. В частности, метод, решающий задачу $\tilde{f}$ со средней абсолютной погрешностью $\leqslant \varepsilon$, одновременно решает с той же точностью и f.

(iii). Поставляемая оракулом $\mathscr{O}$ информация о f очевидным образом пересчитывается в информацию о $\tilde{f}$. Докажите, что получающийся таким образом оракул $\widetilde{\mathscr{O}}$ для $\tilde{f}$ обеспечивает включение

$$\tilde{f} \in \widetilde{W}^{\widetilde{\mathscr{O}}, \nu_0}_{r, 40L}(G, E, \|\cdot\|, \mathrm{l}^{(m)}_\infty, \|\cdot\|_\infty, K_\infty) \equiv \widetilde{W}^{\widetilde{\mathscr{O}}, \nu_0}_{r, 40L}.$$

⟨Из определения класса $\widetilde{C}^{\mathscr{O}, \nu_0}_{r, L}$ ясно, что изменение каждой из функций f_j на G не превосходит $(1 + \nu_0) L \leqslant 2L$. Отсюда немедленно следуют свойства (i) и (ii). Ясно, далее, что если препарирование успешно, то $|\tilde{f}_j(x)| \leqslant 5L$ при $x \in G$. Отсюда и из соотношений А — В п. 1.2.3 гл. VI следует (iii) (см. упражнение 1 § 4).⟩

6.3. Рассмотрим теперь следующий метод решения задач класса $\widetilde{C}^{\mathscr{O}, \nu_0}_{r, L}(G, E, \|\cdot\|, m)$ с заданной средней относительной точностью ν, такой, что

$$c\nu_0 < \nu < 1, \tag{6.2}$$

где $c > 12$ — достаточно большая абсолютная константа (выбираемая из тех соображений, что для величины $\tilde{\nu} = \nu/80$ выполнено (5.11). По поводу значений абсолютных констант в этих и последующих оценках см. заключительное замечание § 4).

Выберем $\delta = \nu/4$ и применим к решаемой задаче f процедуру препарирования *с параметром* δ. *Если при этом получится* $a^j > 3L$ для некоторого $j \geqslant 1$, то прекратим решение и выдадим ответ о несовместности f. В противном случае применим к задаче $\tilde{f}$ (рассматриваемой как элемент класса $\widetilde{W}^{\widetilde{\mathscr{O}}, \nu_0}_{r, 4L}$) метод $\mathscr{B}^{\nu/80}$ из предыдущего параграфа. Результат $\bar{x}$ его применения к $\tilde{f}$ будем считать результатом работы описываемого метода на f. Этот метод будет решать f со (средней относительной) погрешностью $\leqslant \nu$. Действительно, в силу свойств $\mathscr{B}^{\nu/80}$ средняя условная (при условии успеха препарирования) погрешность $\bar{x}$ в качестве решения задачи f в силу утверждений (i) — (iii) упражнения 2 не превосходит $\nu/2$. Относительная погрешность любой точки $G \cup \{*\}$ в качестве решения f не превосходит, очевидно, $1 + \nu_0 \leqslant 2$. Итак, погрешность описанного метода решения задач класса $\widetilde{C}^{\mathscr{O}, \nu_0}_{r, L}$ в условиях (6.2) не превосходит $(\nu/2) + 2\delta \leqslant \nu$. Трудоемкость же его не превосходит $R(\nu/4) + M_w(\nu/80)$.

6.4. Остается научиться строить оценки a_δ с по возможности малыми $R(\delta)$. Отметим характерную особенность этой задачи. Точность, которая требуется от оценки, низка — порядка r-го момента помехи (в степени $1/r$). Зато надежность, требуемая от оценки, может быть очень высокой (оценке разрешается не иметь нужной точности лишь с вероятностью $\leqslant \nu/4$).

Изложим общий способ построения таких оценок*). Пусть (X, Σ) — пространство с σ-алгеброй подмножеств, $\mathscr{F} = \{\mathsf{F}\}$ — некоторое семейство распределений вероятностей на (X, Σ), Q — метрическое пространство с метрикой $\rho(\cdot, \cdot)$ и $\sigma(\mathsf{F}): \mathscr{F} \to Q$ — отображение, ставящее в соответствие распределению F его параметр $\sigma(\mathsf{F})$. [В интересующем нас случае (X, Σ) есть $\mathrm{l}^{(m+1)}_\infty$ с σ-алгеброй борелевских подмножеств $\mathrm{l}^{(m+1)}_\infty$, $\mathscr{F}$ есть семейство распределений значений случайных величин $\psi^0(\bar{x}, f, \omega)$, индуцированных мерой F_ω на Ω. Здесь $\mathscr{O} = \{(\Omega, \mathsf{F}_\omega); \psi^0(x, f, \omega), \psi^1(x, f, \omega)\}$ — рассматриваемый оракул, $Q = \mathrm{l}^{(m+1)}_\infty$ с метрикой, задаваемой нормой $\|\cdot\|_\infty$, и $\sigma(\mathsf{F}) = \int\limits_{\mathrm{l}^{(m+1)}_\infty} y\, d\mathsf{F}(y)$.]

Оценкой с наблюдательным временем n назовем функцию $a(x_1, \ldots, x_n) \equiv a(x^n)$ на выборочном пространстве длины n, т. е. на пространстве $X^n = \underbrace{X \times \ldots \times X}_{n \text{ раз}}$, со значениями в Q, измеримую в том смысле, что прообраз любого шара в Q лежит в наименьшей σ-алгебре X^n, содержащей все множества вида $\{x^n | x_i \in A\}$, где $A \in \Sigma$. Скажем, что *оценка a имеет точность r и надежность α на распределении* $\mathsf{F} \in \mathscr{F}$, если вероятность $\mathsf{Pr}_\mathsf{F}\{\rho(a(x^n), \sigma(F)) > r\}$ указанного в скобках события, вычисленная по распределению на X^n, получающемуся как n-я степень F^n распределения F (т. е. по распределению на X^n с независимыми распределенными по F координатами x_i), не больше α: $\mathsf{Pr}_\mathsf{F}\{\rho(a(x^n), \sigma(\mathsf{F})) > r\} \leqslant \alpha$. Скажем, что *$a$ имеет точность r и надежность α на классе* $\mathscr{F}$, если она обладает этим свойством на любом распределении $\mathsf{F} \in \mathscr{F}$. [В нашем случае оценка с наблюдательным временем n строится по n ответам оракула на вопрос о значении f в точке $\bar{x}$. При этом в качестве оценки a_δ можно взять любую оценку точности ${}^1/_2\, L$ и надежности δ.]

Оказывается, что с помощью оценки точности порядка r (именно, $r/3$) и сравнительно низкой надежности (любой, меньшей ${}^1/_2$) можно получить оценку точности r и наперед заданной надежности.

*) Текст в прямых скобках конкретизирует общую схему применительно к нуждам этого параграфа.

Предложение. Пусть $a\,(x^n)$ — оценка с наблюдательным временем n, имеющая на $\mathcal{F}$ точность $r/3$ и надежность $p < {}^1/_2$. Тогда для некоторого зависящего лишь от p (но не от $\mathcal{F}$, $a\,(\cdot)$, r и n) целого $c\,(p)$ и всякого $\alpha < p$ можно указать оценку a^α точности r и надежности α с наблюдательным временем

$$n^* = nc\,(p)\,\Big]\ln\frac{1}{\alpha}\Big[\,. \tag{6.3}$$

Оценка a^α строится следующим образом. Разбиваем выборку x^{n*} в последовательные блоки $x^{(1)},\ldots,\ x^{(s)}$ длины n каждый; $s = c\,(p)]\ln 1/\alpha\,[$. К каждому из блоков $x^{(i)}$ применяем оценку a. Получаем s точек $y_1,\ldots,\ y_s$ пространства Q. Возможно, что для некоторого i неравенство $\rho\,(y_i,\ y_j) \leqslant 2r/3$ выполнено более чем для $s/2$ номеров j. В этом случае пусть $a^{(s)}(x^{ns}) \equiv a^\alpha\,(x^{n*})$ есть первое среди y_i, обладающих указанным свойством. Если таких i нет, то положим $a^\alpha\,(x^{n*}) = y_1$.

Доказательство предложения. Фиксируем $\mathsf{F} \in \mathcal{F}$. Рассмотрим натуральное k и индуцированное F распределение на X^{nk}. Пусть $x^{(1)},\ldots,\ x^{(k)}$ — последовательные блоки x^{nk} длины n каждый, и $y_i = a\,(x^{(i)})$. Величины y_i независимы в совокупности, и вероятность каждой из них не попасть в шар $W_0 = \{y \in Q\,|\,\rho\,(\sigma\,(\mathsf{F}),\ y) \leqslant r/3\}$ не больше $p < {}^1/_2$ по определению $a\,(\cdot)$. Пусть A — событие, состоящее в одновременном наступлении более $k/2$ из событий $\{y_i \in W_0\}$; $i = 1,\ldots,\ k$. Легко видеть, что вероятность наступления A не меньше $1 - e^{-\varkappa(p)k}$, где $\varkappa\,(p) > 0$ зависит лишь от p. Пусть событие A имело место. Тогда более $k/2$ точек y_i попали в W_0. В частности, все они попали в шар радиуса $2r/3$, с центром в одной из точек y_i, $i = 1,\ldots$ $\ldots,\ k$ (в качестве центра этого шара можно взять любую из точек $y_i \in W_0$). Построим теперь $a^{(k)}\,(x^{nk})$ так, как это описано в предложении. Значение $a^{(k)}$ в силу этого описания определяет центр y шара W_1 радиуса $2r/3$, содержащего более половины точек $y_1,\ldots,\ y_k$. Ясно, что $W_1 \cap W_0 \neq \phi$, что дает $\rho\,(\sigma\,(\mathsf{F}),\ y) \leqslant r$. Итак, $\rho\,(\sigma\,(\mathsf{F}),\ a^{(k)}\,(x^{nk})) \leqslant r$ при наступлении A. Стало быть, $a^{(k)}$ имеет надежность, не худшую $e^{-\varkappa(p)k}$ при точности r. Отсюда немедленно следует утверждение предложения.

Замечание. При $p = {}^1/_4$ можно взять

$$c\,(p) = \Big]2\ln\frac{3}{2\alpha}\Big/\ln\frac{4}{3}\Big[\,. \ \Diamond$$

Применим доказанный общий результат к интересующей нас проблеме надежной оценки $f\,(\bar{x})$. Для этого достаточно научиться строить оценку a величины $\sigma\,(\mathsf{F}) = \int y\,d\,\mathsf{F}\,(y)$ точности $L/6$ и надежности $< {}^1/_2$, скажем, надежности ${}^1/_4$. Про распределение F в силу свойства оракула известно, что

$$\mathsf{M}_\mathsf{F}\,\|y - \sigma\,(\mathsf{F})\|_\infty^r \leqslant [L\,(1 + \nu_0)]^r.$$

Построим оценку a в виде

$$a\,(y^n) = \frac{1}{n}\sum_{i=1}^{n} y_i$$

и подберем нужное n. В силу неравенства (1.5) гл. V, примененного к сумме $\sum\limits_1^n \frac{1}{n}\,\xi_i$, $\xi_i = 36y_i\,/L\,(1 + \nu_0)$, имеем

$$\mathsf{M}_{\mathsf{F}n}\,\|a\,(y^n) - \sigma\,(\mathsf{F})\|_\infty \leqslant \frac{L\,(1+\nu_0)}{36}\left\{\frac{1}{2} + 36\omega_{V_{1,m+1};\,r}\left(\frac{36}{n}\right)\right\} \leqslant$$
$$\leqslant L\left\{\frac{1}{36} + 2\omega_{V_{1,m+1};\,r}\left(\frac{36}{n}\right)\right\}.$$

Ясно, что используемая оценка будет иметь точность $L/6$ и надежность ${}^1/_4$, если правая часть последнего неравенства не больше $L/24$. Таким образом, n должно быть выбрано из условия

$$\frac{1}{36} + 2\omega_{V_{1,m+1};\,r}\left(\frac{36}{n}\right) \leqslant \frac{1}{24}\,, \tag{6.4}$$

что в силу (1.7) гл. V дает

$$n = \,]\,\ln\,(m + 2)\ c\,(r)\,[, \tag{6.5}$$

где $c\,(r)$ зависит лишь от r.

Таким образом, мы умеем строить требуемые в 6.2 оценки a_δ с функцией $R\,(\delta)$ вида

$$R\,(\delta) = d\,(r)\ln(m + 2)\ln\frac{2}{\delta}\,, \tag{6.6}$$

где $d\,(r)$ зависит лишь от r. Применяя эти оценки в методе п. 6.3, получим метод решения задач класса $\widetilde{C}_{r,\,L}^{\mathfrak{S},\,\nu_0}\,(G,\ E,\ \|\cdot\|,\ m)$ с точностью ν.

6.5. Итак, мы редуцировали условные стохастические экстремальные задачи к экстремальным задачам с операторным ограничением (при этом пришлось применить процедуру препарирования, а также несколько модифицировать словоупотребление).

Решение экстремальных задач с операторными ограничениями, в свою очередь, редуцировалось к решению серии из $O\left(\ln^2\frac{1}{\nu}\right)$ операторных неравенств. Каждое из таких неравенств редуцировалось к выпукло-вогнутым играм, а уж последние решались без дополнительных редукций методом, описанным в § 3.

В конечном счете задачи из $\widetilde{C}_{r,\,L}^{\mathfrak{S},\,\nu_c}$ сводятся к играм на $E \times$ $\times\ l_1^{(m+1)}$ — вернее, на произведении ограниченных подмножеств каждой из компонент. Для решения этих игр мы рекомендуем применять ЗС-метод п. 3.3, ассоциированный с парой функции

V, V_I, где V — $(E, \|\cdot\|)$-регулярная функция, а V_I — $(l_1^{(m+1)}, \|\cdot\|_1)$-регулярная функция, в качестве которой разумно брать $V_{1,\, m+1}$ (см. п. 2.5 гл. III).

Упражнение 3. Хорошо известно, что решение экстремальных выпуклых задач f при весьма общих предположениях можно свести к решению игры с платежной функцией

$$F(x, \lambda) = f_0(x) + \sum_{j=1}^{m} \lambda_j f_j(x)$$

на $G \times \{\lambda \mid \lambda_j \geqslant 0\}$. Можно ли было сразу редуцировать задачу f к такой игре?

⟨Нет, по крайней мере, в смысле редукции в описанном выше стиле. Дело в том, что игра $F(x, \lambda)$ определена на неограниченном множестве, что не дает возможности «конструктивно» решать ее ЗС-методами. Отсечение «очень больших» λ также не приносит успеха, так как размеры «урезанной» области растут с увеличением требуемой точности решения.⟩

6.6. Обсудим теперь вопрос о трудоемкости метода решения задач класса $\widetilde{C}_{r,L}^{\mathfrak{S},\nu_0}$, реализующего изложенную в §§ 4—6 схему. Его оценка трудоемкости $\widehat{M}(\nu)$ в условиях (6.2) при выборе $V_I = V_{1,\, m+1}$ имеет вид

$$\widehat{M}(\nu) \leqslant c(r)\ln(m+2)\ln 2/\nu + d(r) \;\rbrack \ln^2 \frac{1}{\nu} \lbrack \; \max\left\{\frac{1}{\nu\gamma_{V,r}\left(\frac{\nu}{c}\right)},\ \frac{\ln(m+2)}{\nu^{\frac{\bar{r}}{\bar{r}-1}}}\right\} \leqslant D(r) \;\rbrack \ln^2 \frac{1}{\nu} \lbrack \; \max\left\{\frac{1}{\nu\gamma_{V,r}\left(\frac{\nu}{c}\right)},\ \frac{\ln(m+2)}{\nu^{\frac{\bar{r}}{\bar{r}-1}}}\right\}. \tag{6.7}$$

Здесь $\bar{r} = \min(r, 2)$, c — абсолютная константа, $D(r)$ зависит лишь от r. В частности, если $(E, \|\cdot\|) = (L_p, \|\cdot\|_p)$, $1 < p < \infty$, то трудоемкость построенного метода допускает оценку

$$D(r, p) \;\rbrack \ln^2 \frac{1}{\nu} \lbrack \; \max\left\{\left(\frac{1}{\nu}\right)^{\frac{\tilde{r}}{\tilde{r}-1}},\ \ln(m+2)\left(\frac{1}{\nu}\right)^{\frac{\bar{r}}{\bar{r}-1}}\right\},$$

где $\tilde{r} = \min\{\bar{r},\ p/(p-1)\}$.

Мы видим, в частности, что в случае $\bar{r} \leqslant p/(p-1)$ (заведомо имеющем место при $1 < p \leqslant 2$) трудоемкость построенного метода в принципе не снижаема более чем в $\underset{p,r}{\frown}$ $\ln^2 1/\nu$ раз, если только $\nu < \bar{\nu}$, $\bar{\nu} > 0$ — абсолютная константа (ср. с нижней оценкой стохастической сложности классов $\widetilde{C}_{r,L}^{\mathfrak{S},0}$ из § 3 гл. V). Если же $\bar{r} > p/(p-1)$, то при всех достаточно малых ν $(\nu \leqslant \bar{\nu}(p, \bar{r}, m))$ трудоемкость построенного метода с точностью до множителя порядка $a(p, r)\ln^2 1/\nu$ — та же, что и у ЗС-метода решения задач класса C_{Lip} на L_p при точной информации (см. § 3 гл. III). Стало быть, в этом случае рекомендуемый метод существенно не улучшаем по трудоемкости (скажем, более чем в $a_1(p, r)\rbrack \ln^2 1/\nu \lbrack$ раз) там, где субоптимальны ЗС$_p$-методы, использующие точную информацию о задачах (т. е. в случае, когда G — тело типа L_p-шара большой размерности, см. главу IV).

При $(E, \|\cdot\|) = (l_1^{(n)}, \|\cdot\|_1)$, $1 < n < \infty$, оценка трудоемкости построенного метода есть

$$D(1, r)\left(\frac{1}{\nu}\right)^{\frac{\bar{r}}{\bar{r}-1}} \max\{\ln n, \ln(m+2)\},$$

и при $\nu < \bar{\nu}$ его трудоемкость в принципе не снижается более чем в

$$a(1, r) \;\rbrack \ln^2 \frac{1}{\nu} \lbrack \; \max\left\{\frac{\ln n}{\ln m}, 1\right\}$$

раз. Заметим еще, что если речь идет о решении задач $f \in \widetilde{C}_{r,L}^{\mathfrak{S},\nu_0}$ с априори известным оптимальным значением целевого функционала f_*, то такие задачи можно сводить к однократному решению операторного неравенства (см. п. 5.3.1), так что в этом случае как верхняя оценка трудоемкости результирующего метода, так и выписанные потенциальные верхние границы понижения его трудоемкости уменьшаются в $O(\ln^2 1/\nu)$ раз.

В заключение заметим, что если интересоваться только задачами класса $\widetilde{C}_{r,L}^{\mathfrak{S},\nu_0}$, то приведенное в двух последних параграфах описание метода их решения можно заметно упростить и сделать более наглядным (подчеркнем, что речь идет именно об описании, но не о фактическом строении метода). Мы избрали менее наглядный, но зато более общий способ изложения.

§ 7. Задачи «сложного» стохастического программирования

7.1. В этом параграфе мы рассмотрим стохастические задачи вида

$$\overline{\Phi}(x) \equiv \Phi_m(\Phi_{m-1}(\ldots(\Phi_0(x))\ldots)) \to \min \mid x \in G_0, \tag{7.1}$$

состоящие в минимизации суперпозиции $m + 1$ функции $\Phi_0, \ldots, \Phi_m$. При этом будем предполагать, что информация о $\overline{\Phi}$ непосредственно недоступна, хотя имеется возможность наблюдать все Φ_i (с помощью оракулов первого порядка). Если при этом информация о Φ_i точна, то, при определенных достаточно общих предположениях о Φ_i, можно извлечь из ответов оракулов обычную информацию первого порядка (значения и опорные функционалы) для $\overline{\Phi}$. Таким образом, при точном наблюдении всех Φ_i нет необходимости как-то специально выделять задачи вида (7.1) из всего семейства экстремальных задач.

Положение меняется, когда Φ_i наблюдаются с помощью стохастических оракулов. В этом случае из информации о Φ_i нельзя (по крайней мере, непосредственно) извлечь информацию о $\overline{\Phi}$, необходимую для решения этой задачи методами первого порядка типа описанных выше. В этом случае задачи вида (7.1) заслуживают специального изучения. Покажем, что при определенных гипотезах о строении Φ_i задачу $\overline{\Phi}$ можно редуцировать к решению выпукло-вогнутой игры.

7.2. Начнем с описания ситуации. Пусть имеется $m+1$ банаховых пространств $(E_0, \|\cdot\|_0), \ldots, (E_m, \|\cdot\|_m)$. Пусть, далее, G_i есть ограниченное выпуклое замкнутое подмножество в E_i. Для единообразия обозначим через E_{m+1}, $\|\cdot\|_{m+1}$ вещественную ось с обычной нормой. Рассмотрим отображения

$$\Phi_0\colon\ G_0 \to G_1,\ \ \Phi_1 : G_1 \to G_2, \ldots, \Phi_m : G_m \to E_{m+1} \equiv \mathbb{R}$$

и определим суперпозицию $\overline{\Phi}(x) = \Phi_m(\Phi_{m-1}(\ldots(\Phi_0(x)\ldots)\colon G_0 \to \mathbb{R}$. Суперпозиция $\overline{\Phi}(x)$ определяет задачу (7.1).

Разумеется, мы будем заниматься лишь выпуклым случаем, когда все отображения Φ_i и их суперпозиция $\overline{\Phi}$ выпуклы. Более точно, предположим, что все пространства E_i, $i > 0$, снабжены выпуклыми замкнутыми конусами K_i, задающими упорядочивание $\underset{i}{\geqslant}$ в E_i (упорядочивание в $E_{m+1} = \mathbb{R}$ — обычное). Предположим, что все отображения $\Phi_i : G_i \to E_{i+1}$, $i = 0, 1, \ldots, m$, липшицевы и выпуклы (относительно порядка $\underset{i+1}{\geqslant}$ в E_{i+1}), а при $i \geqslant 1$ эти отображения вдобавок монотонны (т. е. из $x, y \in G_i$, $x \underset{i}{\leqslant} y$ следует $\Phi_i(x) \underset{i+1}{\leqslant} \Phi_i(y)$). В этих предположениях суперпозиция $\overline{\Phi}$ оказывается липшицевой выпуклой скалярной функцией на G_0 (проверить!).

Предположим еще, что пространства E_i, $\|\cdot\|_i$, $0 \leqslant i \leqslant m$, сепарабельны и регулярны, и таковы же пространства E_i^*, $\|\cdot\|_i^*$ при $1 \leqslant i \leqslant m$. Кроме того, будем считать, что множества G_i имеют непустые внутренности при $1 \leqslant i \leqslant m$.

7.3. Теперь сформулируем гипотезу о возможности наблюдения компонент Φ_i задачи $\overline{\Phi}$. Грубо говоря, мы считаем, что каждое из отображений Φ_i может наблюдаться с помощью стохастического оракула типа описанного в § 4. Для простоты будем считать этот оракул несмещенным, а ситуацию — нормированной. Именно, предположим, что $\rho_{\|\cdot\|_i}(G_i) = 1$ и центр G_i есть 0. Далее, пусть $x^m = (x_0, \ldots, x_m)$ — точка $G^m = G_0 \times \ldots \times G_m$. Предположим, что в нашем распоряжении имеется оракул $\mathcal{O}$ с пространством шумов $(\Omega, \mathsf{F}_\omega)$ и функцией наблюдения, имеющей $2(m+1)$ компонент. Первые $m+1$ из них $\psi_0^i(x^m, \omega)\colon G^m \times \Omega \to E_{i+1}$ (мы опускаем указатель $\overline{\Phi}$ в их обозначении) суть доступные методу наблюдения величин $\Phi_i(x_i)$, а следующие $m+1$ компонент $\psi_1^i(x^m, \omega)\colon G^m \times \Omega \to L(E_i, E_{i+1})$ суть наблюдения дифференциалов Φ_i. Здесь $L(E_i, E_{i+1})$ есть пространство непрерывных линейных операторов из E_i в E_{i+1}. Как всегда, $(\Omega, \mathsf{F}_\omega)$ предполагается польским пространством с регулярной борелевой вероятностной мерой, а ψ_0^i, ψ_1^i считаются борелевыми по x^m, ω (ψ_1^i — в смысле п. 4.2.3). При этом предполагается, что при всех $x^m \in G^m$ и i, $0 \leqslant i \leqslant m$,

$$\mathsf{M}_{\mathsf{F}_\omega}\psi_0^i(x^m, \omega) = \Phi_i(x_i);\ \ \mathsf{M}_{\mathsf{F}_\omega}\{\psi_1^i(x^m, \omega)(y - x_i)\} \underset{i+1}{\leqslant}$$
$$\underset{i+1}{\leqslant} \Phi_i(y) - \Phi_i(x_i),\ y \in G_i;\ \ \mathsf{M}_{\mathsf{F}_\omega}\|\psi_0^i(x^m, \omega)\|_{i+1}^r \leqslant 1;$$
$$\mathsf{M}_{\mathsf{F}_\omega}\{\|[\psi_1^i(x^m, \omega)]^* l\|_{i,*}\}^r \leqslant 1 \text{ для всех } l \underset{i+1,*}{\geqslant} 0 \text{ с } \|l\|_{i+1,*} \leqslant 1.$$

Здесь $\underset{i+1,*}{\geqslant}$ — порядок в E_{i+1}^*, индуцированный двойственным к K_{i+1} конусом K_{i+1}^*, $[\psi_1^i]^*$ — сопряженный к ψ_1^i оператор (см. п. 4.1), $r > 1$ — параметр.

7.4. Покажем, что задачу $\overline{\Phi}$ можно редуцировать к игре; при этом необходимую для решения этой игры информацию можно получить от оракула $\mathcal{O}$. Именно, пусть

$$G_i^* = \{l \in E_i^* \,|\, l \underset{i,*}{\geqslant} 0,\ \|l\|_{i,*} \leqslant 1\},\ \ 1 \leqslant i \leqslant m,$$

а

$$G^{*m} = G_1^* \times \ldots \times G_m^* = \{l^m \,|\, l^m = (l_1, \ldots, l_m),\ \ l_i \in G_i^*\}.$$

Рассмотрим игру $F_{\overline{\Phi}}$, определенную на $G^m \times G^{*m}$:

$$F_{\overline{\Phi}}(x^m, l^m) = \Phi_m(x_m) + \langle l_m \,|\Phi_{m-1}(x_{m-1}) - x_m\rangle +$$
$$+ \langle l_{m-1} \,|\Phi_{m-2}(x_{m-2}) - x_{m-1}\rangle + \ldots + \langle l_1 \,|\Phi_0(x_0) - x_1\rangle.$$

Мы покажем со временем, что игра $F_{\overline{\Phi}}$ тесно связана с задачей $\overline{\Phi}$. Именно, пусть

$$\overline{F}_{\overline{\Phi}}(x^m) = \sup_{l^m \in G^{*m}} F_{\overline{\Phi}}(x^m, l^m)$$

— платежная функция первого игрока. Тогда при всех $x_0 \in G_0$

$$\overline{\Phi}(x_0) = \inf_{x_1 \in G_1, \ldots, x_m \in G_m} \overline{F}_{\overline{\Phi}}(x_0, x_1, \ldots, x_m), \tag{7.2}$$

т. е. $\overline{\Phi}(x_0)$ есть плата первого игрока за наилучший из его ходов с первой компонентой x_0. Отсюда ясно, что решив игру $F_{\overline{\Phi}}$ с (средней) абсолютной погрешностью $\varepsilon/2$ и взяв нулевую координату x^m — компоненты полученного решения $F_{\overline{\Phi}}$ в качестве приближенного решения задачи минимизации $\overline{\Phi}$, мы получим (случайную) точку $\bar{x}_0 \in G_0$, для которой

$$\mathsf{M}\overline{\Phi}(\bar{x}_0) \leqslant \min_{G_0} \overline{\Phi}(x_0) + \varepsilon \tag{7.3}$$

(M *означает усреднение по распределению* $\bar{x}_0$).

7.5. Остается научиться решать игру $F_{\overline{\Phi}}$. Для этого применим ЗС-метод из п. 3.3. Его применение возможно. Действительно, ввиду наложенных выше условий на E_i и E_i^*, пространства $E \equiv E_0 \times \ldots \times E_m$ с нормой $\|x^m\| = \max_i \|x_i\|_i$ и $E_I \equiv E_1^* \times \ldots \times E_m^*$ с нормой $\|l^m\|_I = \max_i \|l_i\|_{i,*}$ регулярны и сепарабельны. Далее, оракул $\mathcal{O}$ фактически поставляет необходимую для этого метода информацию. В самом деле, легко видеть, что $F_{\overline{\Phi}}$ — липшицева выпукло-вогнутая игра. При этом функция

$$\psi_x(x^m, l^m, \omega) = \{[\psi_1^m(x^m, \omega)]^* \mathbb{1},\ [\psi_1^{m-1}(x^m, \omega)]^* l_m, \ldots \\ \ldots, [\psi_1^0(x^m, \omega)]^* l_1\}$$

со значениями в $E_0^* \times \ldots \times E_m^*$ ($\mathbb{1}$ — стандартный линейный функционал $\langle \mathbb{1} \mid t\rangle = t$ на $E_{m+1} = \mathbb{R}$) обладает тем свойством, что $\mathsf{M}_{F_\omega}\psi_x(x^m, l^m, \omega)$ опорно к $F_{\overline{\Phi}}$ как функции x^m. Аналогично функция $\psi_l(x^m, l^m, \omega) = \{\Phi_{m-1}(x_{m-1}) - x_m, \ldots, \Phi_0(x_0) - x_1\}$ со значениями в $E_1 \times \ldots \times E_m = [E_1^* \times \ldots \times E_m^*]^*$ такова, что $\mathsf{M}_{F_\omega}\psi_l(x^m, l^m, \omega)$ является опорным функционалом к $F_{\overline{\Phi}}$ как функции l^m.

Итак, с помощью оракула $\mathcal{O}$ можно построить несмещенные оценки ψ_x, ψ_l опорных функционалов к $F_{\overline{\Phi}}$ по x^m- и по l^m-компонентам. Из свойств оракула $\mathcal{O}$ легко извлечь, что при описанном способе наблюдения $F_{\overline{\Phi}}$ эта игра попадет в класс вида $\widetilde{\mathbf{D}}_{r,L}^{\mathcal{O},0}(G^m \times G^{*m}, E, \|\cdot\|, E_I, \|\cdot\|_I)$ с $L = 4(m+1)$. Таким образом, параметр L явно выписывается, и применение к $F_{\overline{\Phi}}$ метода п. 3.3 не вызывает принципиальных затруднений.

7.6. Остается доказать (7.2). Нам потребуется для этого простая лемма.

Лемма. *Пусть* $\widetilde{\Phi}_i$: $G_i \to \mathbb{R}$ — *монотонное липшицево с константой* 1 *выпуклое отображение,* $i \geqslant 1$. *Тогда при всех* $x \in G_i$

$$\widetilde{\Phi}_i(x) = \sup_{l \in G_i^*} \inf_{y \in G_i} \{\widetilde{\Phi}_i(y) - \langle l \mid y - x\rangle\}. \tag{7.4}$$

Доказательство. По лемме фон Неймана

$$\sup_{l \in G_i^*} \inf_{y \in G_i} \{\widetilde{\Phi}_i(y) - \langle l \mid y - x\rangle\} = \inf_{y \in G_i} \sup_{l \in G_i^*} \{\widetilde{\Phi}_i(y) - \langle l \mid y - x\rangle\}. \tag{7.5}$$

Ясно, что правая часть (7.5) при данном $x \in G$ не больше $\widetilde{\Phi}_i(x)$ (можно взять $y = x$). Обозначая правую часть (7.4) через $H(x)$, получим

$$H(x) \leqslant \widetilde{\Phi}_i(x), \quad x \in G. \tag{7.6}$$

С другой стороны, по условию $\operatorname{int} G_i \neq \phi$. Пусть $x \in \operatorname{int} G_i$. Тогда из монотонности и липшицевости (с константой 1) функции $\widetilde{\Phi}_i$ следует, что опорный к ней в x функционал l_x лежит в G_i^* (почему?). Функция $\widetilde{\Phi}_i(y) - \langle l_x \mid y - x\rangle$ достигает минимума по y при $y = x$, и этот минимум есть $\widetilde{\Phi}_i(x)$. Стало быть, левая часть (7.5) при $x \in \operatorname{int} G$ не меньше $\widetilde{\Phi}_i(x)$. Отсюда и из (7.6) $H(x) = \widetilde{\Phi}_i(x)$ при $x \in \operatorname{int} G$. Обе части этого равенства непрерывны на G_i (почему?), что и доказывает (7.4).

Упражнение 1. Существенно ли условие $\operatorname{int} G_i \neq \phi$?

Теперь мы в состоянии доказать (7.2). Действительно, по определению $F_{\overline{\Phi}}$

$$\overline{F}_{\overline{\Phi}}(x_0, \ldots, x_m) = \\ = \max_{l^m \in G^{*m}} \{\Phi_m(x_m) + \langle l_m \mid \Phi_{m-1}(x_{m-1}) - x_m\rangle + \ldots + \langle l_1 \mid \Phi_0(x_0) - x_1\rangle\}. \tag{7.7}$$

Отсюда, ввиду леммы фон Неймана,

$$\inf_{x_m \in G_m} \overline{F}_{\overline{\Phi}}(x_0, \ldots, x_m) = \\ = \max_{l_1, \ldots, l_{m-1},\, l_j \in G_j^*} \{\max_{l_m \in G_m^*} \min_{x_m \in G_m} \{\Phi_m(x_m) - \langle l_m \mid x_m - \Phi_{m-1}(x_{m-1})\rangle\} + \\ + \langle l_{m-1} \mid \Phi_{m-2}(x_{m-2}) - x_{m-1}\rangle + \ldots + \langle l_1 \mid \Phi_0(x_0) - x_1\rangle\}.$$

Выражение во внутренних фигурных скобках, ввиду леммы, есть $\Phi_m(\Phi_{m-1}(x_{m-1}))$. (Указанная лемма действительно применима к Φ_m; из ее условий в проверке нуждается лишь липшицевость с константой 1, сразу следующая из свойств оракула (ввиду которых $\mathsf{M}_{F_\omega}\psi_0^m(x^m, \omega)$ есть опорный функционал к $\Phi_m(x_m)$ нормы, $\leqslant 1$).) Итак,

$$\min_{x_m \in G_m} \overline{F}_{\overline{\Phi}}(x_0, \ldots, x_{m-1}; x_m) = \\ = \max_{l_1, \ldots, l_{m-1},\, l_j \in G_j^*} \{\Phi_m(\Phi_{m-1}(x_{m-1})) + \langle l_{m-1} \mid \Phi_{m-2}(x_{m-2}) - x_{m-1}\rangle + \ldots \\ \ldots + \langle l_1 \mid \Phi_0(x_0) - x_1\rangle\}. \tag{7.8}$$

Выражение (7.8) устроено точно так же, как и выражение (7.7), только роль m играет $m - 1$, а роль Φ_m — функция $\Phi_m(\Phi_{m-1}(\cdot))$. Если бы мы знали, что $\Phi_m(\Phi_{m-1}(\cdot))$, как и $\Phi_m(\cdot)$, липшицево с константой 1 выпуклое монотонное отображение, то могли бы перейти от (7.8) к аналогичному соотношению для $\min\limits_{x_m\in G_m, x_{m-1}\in G_{m-1}} F_{\overline{\Phi}}(x_0, \ldots, x_m)$ и, поступая так же и дальше, пришли бы к требуемому соотношению (7.2).

Из сказанного ясно, что для доказательства (7.2) достаточно проверить, что если $j > 1$ и H_j: $G_j \to \mathbb{R}$ — липшицево с константой 1 монотонное выпуклое отображение, то таково же и отображение $H_j(\Phi_{j-1}(x_{j-1})) = \widetilde{H}_{j-1}(x_{j-1})$: $G_{j-1} \to \mathbb{R}$. Монотонность и выпуклость $\widetilde{H}_{j-1}$ очевидны.

Далее, $\operatorname{int} G_j \neq \phi$ и H_j имеет в каждой точке $x_j \in \operatorname{int} G_j$ опорный функционал из G_j^*. По стандартным соображениям компактности то же верно и при $x_j \in G_j$. Пусть теперь $\bar{x}_{j-1} \in G_{j-1}$ и $\bar{x}_j = \Phi_{j-1}(\bar{x}_{j-1})$. Пусть $l \in G_j^*$ опорно к H_j в $\bar{x}_j$. Ясно, что $\tilde{l} = \mathsf{M}_{\mathsf{F}_\omega}\{[\psi_1^{j-1}(\bar{x}^m, \omega)]^* l\}$ опорно к $\widetilde{H}_{j-1}$ в $\bar{x}_{j-1}$ (здесь x^m — любая точка, $(j-1)$-я координата которой есть $\bar{x}_{j-1}$). По свойствам оракула $\|\tilde{l}\|_{j-1,*} \leqslant 1$. Таким образом, в каждой точке G_{j-1} $\widetilde{H}_{j-1}$ имеет опорный функционал нормы, $\leqslant 1$. Тем самым $\widetilde{H}_{j-1}$ липшицево с константой 1 на G_{j-1}, что и требуется. Соотношение (7.2) полностью доказано.

7.7. Замечание. Фактически, мы уже использовали конструкцию, описанную выше, именно — в § 4. Действительно, решение операторного неравенства $H(x) \leqslant 0$ сводилось там к решению «сложной» стохастической задачи $s(H(x)) \to \min | x \in G$ (см. п. 4.3.1), а эта последняя решалась сведением к игре типа $F_{\overline{\Phi}}$. Специфика ситуации § 4 состоит в том, что функция $s(y)$ известна априори и не нуждается в наблюдении; к тому же она однородна и представима в виде $\max\limits_{l\in G_1^*}\langle l | y\rangle$ (G_1^* здесь — то же, что G_l в п. 4.3.1). Это специальное представление $s(y)$ (получающееся, конечно, и из (7.4) взятием нижней грани по y) использовано в § 4 вместо представления (7.4), на котором основана общая схема рассуждений этого параграфа.

Глава VII

СИЛЬНО ВЫПУКЛЫЕ ЗАДАЧИ

До сих пор мы занимались методами первого порядка решения «произвольных» выпуклых задач. Теперь перейдем к рассмотрению специальных классов таких задач, выделяемых определенными условиями гладкости и строгой выпуклости. Следует заметить, что большинство стандартных численных методов оптимизации ориентированы как раз на такие «хорошие» задачи.

Начнем с описания подлежащих рассмотрению классов задач (§ 1) и оценим снизу их сложность (§ 2). В §§ 3, 4, 5 будут построены методы, «почти» реализующие эту нижнюю оценку сложности. Результаты главы в основном базируются на работе авторов [32].

§ 1. Классы гладких и сильно выпуклых экстремальных задач

1.1. Пусть E — гильбертово вещественное сепарабельное пространство (это соглашение действует на протяжении всей главы). Норму в E будем обозначать просто $|\cdot|$. Функцию f: $E \to R$ назовем *L-гладкой*, если

(1) f непрерывно-дифференцируема;

(2) при всех $x, y \in E$ имеем

$$|\langle f'(x) - f'(y) | x - y\rangle| \leqslant L|x - y|^2.$$

Фактически (2) эквивалентно условию

(2′) $|f'(x) - f'(y)| \leqslant L|x - y|$ при всех $x, y \in E$,

т. е. L-гладкие функции — это в точности дифференцируемые функции, производная которых липшицева с константой L.

Функция $f: E \to \mathbb{R}$ называется *(l,Q)-сильно выпуклой*, $l > 0$, $Q \geqslant 1$, если

(1) f непрерывно-дифференцируема;

(2) при всех $x, y \in E$

$$l|x - y|^2 \leqslant \langle f'(x) - f'(y) | x - y\rangle \leqslant Ql|x - y|^2.$$

Поясним определение. Пусть f дважды дифференцируема. Ее второй дифференциал (гессиан) есть квадратичная форма на E. Функция f (l,Q)-сильно выпукла тогда и только тогда, когда минимальное собственное число гессиана везде не меньше l, а максимальное собственное число везде не больше Ql *).

Функцию f, являющуюся L-гладкой при каком-нибудь $L < \infty$, будем называть *гладкой*. Функцию f, являющуюся (l, Q)-сильно выпуклой при каких-нибудь $l > 0$, $Q > 1$, будем называть *сильно выпуклой*.

Пусть f гладкая. Тогда существует минимальное $L = L_f$, для которого f — L-гладкая. Равным образом, если f сильно выпуклая, то существуют $l_f > 0$ и $Q_f \geqslant 1$, такие, что из (l, Q)-выпуклости f следует $l \leqslant l_f$ и $Q \geqslant Q_f$. При этом $Q_f = L_f/l_f$.

Пусть f (l,Q)-сильно выпуклая. Тогда имеют место следующие неравенства, которыми мы постепенно будем пользоваться:

$$\left.\begin{aligned} f(x+h) &\geqslant f(x) + \langle f'(x)\,|\,h\rangle + \frac{l}{2}h^2,\\ f(x+h) &\leqslant f(x) + \langle f'(x)\,|\,h\rangle + \frac{lQ}{2}h^2. \end{aligned}\right\} \tag{1.1}$$

Обратно, если f непрерывно-дифференцируема и обладает свойством (1.1), то f (l,Q)-сильно выпукла.

Упражнение 1. Докажите сформулированные утверждения.

1.2. Пусть $G \subset E$ — выпуклое замкнутое непустое (не обязательно ограниченное) множество, $x_1 \in G$. Мы будем рассматривать классы задач вида

$$f_0(x) \to \min \,|\, x \in G, \quad f_i(x) \leqslant 0, \quad 1 \leqslant i \leqslant m, \tag{1.2}$$

задаваемые сильно выпуклыми (и гладкими) функциями f_i. Более точно, пусть $l_i > 0$, $Q_i \geqslant 1$. Обозначим через $\boldsymbol{H} = \boldsymbol{H}_{x_1}(G, m; l_0, \ldots, l_m, Q_0, \ldots, Q_m)$ следующий класс задач (назовем его *классом сильно выпуклых задач*):

— множество задач этого класса есть в точности множество всех задач (1.2) (т. е. функций $f(x) = (f_0(x), \ldots, f_m(x))$), у которых f_i (l_i, Q_i)-сильно выпуклы ($i = 0, 1, \ldots, m$);

— оракул есть точный оракул первого порядка с областью вопросов E. Иными словами, оракулу можно задать вопрос в любой точке $x \in E$ и получить в ответ $f(x)$, $f'(x)$;

— нормирующие множители $r(f)$ определены следующим образом. Пусть для каждого $x \in G$

$$\mathrm{V}_f(x) = \max\{f_0(x) - f_*, f_1(x), \ldots, f_m(x)\}$$

*) В бесконечномерном случае — нижняя, соответственно верхняя грани спектра.

(f_* — оптимальное значение целевого функционала задачи f; если f несовместна, то $f_* = +\infty$). Тогда

$$r_i(f) \equiv r(f) \equiv \mathrm{V}_f(x_1).$$

Смысл нормировки ясен. $\mathrm{V}_f(x)$ измеряет невязку x в качестве приближенного решения f, т. е. максимальное отклонение функционалов задачи f в точке x от «номинальных» (требуемых от точного решения) своих значений. В качестве $r_i(f)$ берется невязка фиксированной точки $x_1 \in G$. Фраза «метод имеет точность ν» означает теперь «невязка результата его работы, по крайней мере, в $1/\nu$ раз меньше невязки x_1». Как мы увидим далее, сложность класса $\boldsymbol{H}$ в основном определяется величиной $Q = \max\limits_i Q_i$. Q называется *модулем сильной выпуклости класса* $\boldsymbol{H}$.

Наряду с классами типа $\boldsymbol{H}_{x_1}$ мы будем рассматривать классы $\boldsymbol{H}^V_{x_1}(G, m; l_0, \ldots, l_m; Q_0, \ldots, Q_m)$. Эти классы составлены всеми задачами $f \in \boldsymbol{H}_{x_1}(G, m; l_0, \ldots, l_m; Q_0, \ldots, Q_m)$, которые либо несовместны, либо

$$\overline{\mathrm{V}}_f(x_1) \equiv \max\{|f_0(x_1) - f_*|, f_1(x_1), \ldots, f_m(x_1)\} \leqslant V.$$

Оракул для этого класса — тот же, что и для $\boldsymbol{H}_{x_1}$, а нормирующие множители для всех задач класса равны V.

Обсудим различия в определениях классов $\boldsymbol{H}_{x_1}$ и $\boldsymbol{H}^V_{x_1}$. Очевидно, применение к данной задаче $\boldsymbol{H}^V_{x_1}$-метода требует несколько большей априорной информации, чем применение к ней $\boldsymbol{H}_{x_1}$-метода. В обоих случаях требуется оценить параметры сильной выпуклости функционалов задачи f. Но в первом требуется еще априори оценить величину $\overline{\mathrm{V}}_f(x_1)$ сверху, чтобы понять, в какой из классов $\boldsymbol{H}^V_{x_1}$ попала f. Далее, $\boldsymbol{H}_{x_1}$-метод погрешности ν решит f с (абсолютной) погрешностью $\leqslant \mathrm{V}_f(x_1)\,\nu$, а $\boldsymbol{H}^V_{x_1}$-метод той же точности — с погрешностью $\leqslant V\nu$. Так как для совместных f $V \geqslant \overline{\mathrm{V}}_f(x_1) \geqslant \mathrm{V}_f(x_1)$, то вторая оценка хуже первой. Таким образом, $\boldsymbol{H}^V_{x_1}$-методы с точки зрения приложений менее удобны, чем $\boldsymbol{H}_{x_1}$-методы.

Впрочем, проигрыш здесь скорее теоретический. Дело в том, что оценить $\overline{\mathrm{V}}_f(x_1)$ сверху несложно: достаточно знать о задаче f вопрос в точке x_1 и использовать далее (1.1). Остается дефект, связанный с «более грубым» измерением погрешности $\boldsymbol{H}^V_{x_1}$-методами в сравнении с $\boldsymbol{H}_{x_1}$-методами. Но зависимость трудоемкости $\boldsymbol{H}^V_{x_1}$-методов от точности, как будет показано, логарифмическая. Поэтому во всех «разумных» случаях она мало изменится, если от $\boldsymbol{H}^V_{x_1}$-шкалы погрешностей перейти к $\boldsymbol{H}^V_{x_1}$-шкале.

Причина введения классов $\boldsymbol{H}^V_{x_1}$ состоит в том, что в некоторых случаях мы не умеем строить для $\boldsymbol{H}_{x_1}$ столь же эффективные (по оценкам) методы, как для $\boldsymbol{H}^V_{x_1}$.

1.3. Пусть G — по-прежнему выпуклое, замкнутое и непустое, но теперь уже обязательно ограниченное подмножество E, и пусть $V > 0$, а $x_1 \in G$. Обозначим через $\boldsymbol{CS}^V_{x_1}(G, m)$ класс всех задач вида (1.2), у которых f_i — выпуклые и $L_i(f)$ — гладкие функции на E, такие, что

$$\max_i \left(\frac{1}{2} L_i(f)\rho^2_{G, x_1} + \rho_{G, x_1}|f_i'(x_1)|\right) \leqslant V;$$

здесь

$$\rho_{G, x_1} = \max_{y \in G}|y - x_1|$$

(слева написаны естественные верхние оценки величин $\sup\limits_G f_i(x) - f_i(x_1)$). Оракул для $\boldsymbol{CS}^V_{x_1}$ — тот же, что и для $\boldsymbol{H}^V_{x_1}$, нормирующий множитель — V. Классы $\boldsymbol{CS}^V_{x_1}$ называются классами *гладких выпуклых задач.*

Заметим сразу же, что решение гладких выпуклых задач можно редуцировать к решению сильно выпуклых. Действительно, пусть $\nu > 0$ — требуемая точность решения задач класса $\boldsymbol{CS}^V_{x_1}(G, m)$. Сопоставим каждой задаче $f \in \boldsymbol{CS}^V_{x_1}(G, m)$ ее *регуляризацию* f^ν:

$$f_i^\nu(x) = f_i(x) + \frac{\nu V (x - x_1)^2}{2\rho^2_{G, x_1}} - \frac{\nu V}{2}.$$

Ясно, что

$$f^\nu \in H_{x_1}^{\left(1+\frac{\nu}{2}\right)V}\left(G, m; \frac{\nu V}{\rho^2_{G, x_1}}, \ldots, \frac{\nu V}{\rho^2_{G, x_1}}; \frac{2+\nu}{\nu}, \ldots, \frac{2+\nu}{\nu}\right).$$

Пусть теперь имеется детермированный метод $\mathscr{B}_\nu$ решения задач последнего класса точности $\nu/3$. Построим по $\mathscr{B}_\nu$ метод $\mathscr{B}_\nu'$ решения задач $\boldsymbol{CS}^V_{x_1}$ следующим очевидным способом: применим к задаче f^ν метод $\mathscr{B}_\nu$ и результат x_f этого применения выдадим в качестве результата работы $\mathscr{B}_\nu'$ на f.

Очевидно, мы действительно описали $\boldsymbol{CS}^V_{x_1}$-метод (информация о f^ν очевидным образом вычисляется по информации о f). Докажем, что он имеет нужную точность ν, если только $\nu < 1$ (только этот случай не тривиален).

Возможно, что $x_f \neq *$. Тогда

$$f_i^\nu(x_f) \leqslant \left(1 + \frac{\nu}{2}\right)\frac{\nu}{3}V, \quad 1 \leqslant i \leqslant m,$$

и

$$f_i(x_f) \leqslant f_i^\nu(x_f) + \frac{\nu V}{2} \leqslant \nu V, \quad 1 \leqslant i \leqslant m.$$

Следовательно, если f несовместна, то $\nu(x_f, f) \leqslant \nu$.

Пусть теперь f совместна. Тогда совместна и f^ν, так что по свойствам $\mathscr{B}_\nu$ имеем $x_f \neq *$. Кроме того, $(f^\nu)_* \leqslant f_*$, и по свойствам $\mathscr{B}_\nu$

$$f_0^\nu(x_f) \leqslant (f^\nu)_* + \frac{\nu}{3}V\left(1 + \frac{\nu}{2}\right) \leqslant f_* + \frac{\nu}{3}V\left(1 + \frac{\nu}{2}\right),$$

$$f_0(x_f) \leqslant f_0^\nu(x_f) + \frac{\nu V}{2} \leqslant f_* + \nu V.$$

Это неравенство вместе с предыдущими доказывает, что x_f есть решение задачи f с погрешностью, $\leqslant \nu$. Итак, $\mathscr{B}_\nu'$ решает задачи из $\boldsymbol{CS}^V_{x_1}(G, m)$ с точностью ν. Трудоемкость его при этом равна трудоемкости $\mathscr{B}_\nu$.

В силу сказанного мы обратим основное внимание на классы сильно выпуклых задач. Методы их решения индуцируют и методы «гладкой выпуклой» оптимизации.

1.4. Начнем с простого (по крайней мере, в идейном плане) метода решения сильно выпуклых задач — естественного обобщения классического градиентного метода. Пусть имеется класс задач $\boldsymbol{H} = \boldsymbol{H}_{x_1}(G, m; l_0, \ldots, l_m; Q_0, \ldots, Q_m)$. Описываемый метод строит на задаче $f \in \boldsymbol{H}$ последовательность точек $x_1, x_2, \ldots$, такую, что при всех i, для которых $x_i \in G$, имеем

$$\mathsf{V}_f(x_i) \leqslant \left(1 - \frac{1}{Q}\right)\mathsf{V}_f(x_{i-1}) \tag{1.3}$$

(Q — модуль сильной выпуклости класса $\boldsymbol{H}$). Если же $x_i \notin G$, то $x_{i+s} = *$, $s \geqslant 1$, и при этом $\nu(x_i, f) = 0$ (т. е. f несовместна). В частности, для обеспечения точности ν достаточно сделать

$$\left]\frac{\ln\frac{1}{\nu}}{\ln\frac{Q}{Q-1}}\right[\leqslant \left]Q\ln\frac{1}{\nu}\right[\text{ шагов.}$$

Опишем $(i-1)$-й шаг метода. Пусть к данному шагу уже имеется точка $x_{i-1} \in G \cup \{*\}$. Если $x_{i-1} = *$, то положим и $x_i = *$. Иначе пусть

$$f_{j,x}(y) = f_j(x) + \langle f_j'(x)|y - x\rangle + \frac{1}{2}Ql_j(y - x)^2.$$

Решим задачу

$$(P_{i-1}): f_{0, x_{i-1}}(y) \to \min | y \in G,$$

$$f_{j, x_{i-1}}(y) \leqslant \left(1 - \frac{1}{Q}\right) f_{j, x_{i-1}}(x_{i-1}), \quad 1 \leqslant j \leqslant m.$$

Для формирования задачи P_{i-1} достаточно задать оракулу вопрос о f в точке x_{i-1}. В качестве x_i возьмем решение P_{i-1} ($x_i = *$, если P_{i-1} несовместна).

Убедимся, что описанный метод действительно обладает сформулированным в (1.3) свойством уменьшения невязки за шаг. Пусть вначале $x_i \in G$. Ввиду (1.1), при $j \geqslant 1$

$$f_j(x_i) \leqslant f_{j, x_{i-1}}(x_i) \leqslant \left(1 - \frac{1}{Q}\right) f_{j, x_{i-1}}(x_{i-1}) = f_j(x_{i-1})\left(1 - \frac{1}{Q}\right). \quad (1.4)$$

Отсюда ясно, что если f несовместна, то

$$V_f(x_i) \equiv \max_{j \geqslant 1} f_j(x_i) \leqslant \left(1 - \frac{1}{Q}\right) \max_{j \leqslant 1} f_j(x_{i-1}) = \left(1 - \frac{1}{Q}\right) V_f(x_{i-1}),$$

что и требуется в (1.3).

Пусть теперь f совместна и x^* — ее решение. Тогда в силу (1.1) для всех $j \geqslant 0$

$$f_j(x^*) \geqslant f_j(x_{i-1}) + \langle f_j(x_{i-1}) \,|\, x^* - x_{i-1} \rangle + \frac{l_j}{2}(x^* - x_{i-1})^2. \quad (1.5)$$

Полагая

$$y = x_{i-1} + \frac{1}{Q}(x^* - x_{i-1}),$$

найдем, что $y \in G$ и

$$f_{j, x_{i-1}}(y) = f_{j, x_{i-1}}(x_{i-1}) + \frac{1}{Q}\left[\langle f_j(x_{i-1}) \,|\, x^* - x_{i-1} \rangle + \frac{1}{2} l_j (x^* - x_{i-1})^2\right] \leqslant$$
$$\leqslant f_{j, x_{i-1}}(x_{i-1}) + \frac{1}{Q}(f_j(x^*) - f_j(x_{i-1}))$$

(мы учли (1.5)). Таким образом,

$$f_{j, x_{i-1}}(y) \leqslant \frac{1}{Q} f_j(x^*) + \left(1 - \frac{1}{Q}\right) f_{j, x_{i-1}}(x_{i-1}). \quad (1.6)$$

Отсюда ясно, что P_{i-1} совместна (y — ее план), так что при совместной f $x_i \in G$. Кроме того, полагая в (1.6) $j = 0$, найдем

$$f_0(x_i) \leqslant f_{0, x_{i-1}}(x_i) \leqslant f_{0, x_{i-1}}(y) \leqslant \frac{1}{Q} f_* + \left(1 - \frac{1}{Q}\right) f_0(x_{i-1}),$$

т. е.

$$f_0(x_{i-1}) - f_* \leqslant \left(1 - \frac{1}{Q}\right)(f_0(x_{i-1}) - f_*). \quad (1.7)$$

Мы видим, что при совместной f наряду с (1.4) выполнено (1.7), откуда

$$V_f(x_i) = \max\{f_0(x_i) - f_*, f_1(x_i), \ldots, f_m(x_i)\} \leqslant$$
$$\leqslant \left(1 - \frac{1}{Q}\right) \max\{f_0(x_{i-1}) - f_*, f_1(x_{i-1}), \ldots, f_m(x_{i-1})\} =$$
$$= \left(1 - \frac{1}{Q}\right) V_f(x_{i-1}),$$

что и требуется. Доказательство закончено.

Пример 1. Пусть $G = E$, $m = 0$. Метод превращается в классический градиентный спуск с постоянным шагом:

$$x_i = x_{i-1} - \frac{1}{Q_0 l_0} f_0'(x_{i-1}).$$

Пример 2. Пусть $m = 0$, G — любое. Функция $f_{i,x}(y)$ достигает глобального (по всем $y \in E$) минимума в точке

$$y_x = x - \frac{1}{Q_0 l_0} f_0'(x).$$

При отходе от y_x она растет пропорционально $|y - y_x|^2$. Стало быть, ее минимум на G есть ближайшая к y_x точка G, т. е. $\pi_G(y_x)$. Метод, таким образом, имеет вид

$$x_i = \pi_G\left(x_{i-1} - \frac{1}{Q_0 l_0} f_0'(x_{i-1})\right).$$

Мы получаем стандартный метод проекции градиента с постоянным шагом.

Помимо этих двух случаев, описанный метод практически приемлем, по-видимому, лишь при $G = E$ и небольших m. Действительно, в этом случае шаг состоит в решении условной квадратичной задачи во всем пространстве. При этом все квадратичные члены стандартны (имеют вид λx^2). Указанную задачу можно решить, перейдя к двойственной задаче, целевой функционал которой явно выписывается. Таким образом, шаг сводится к решению выпуклой задачи небольшой (по предположению m — небольшое) размерности.

В остальных случаях применение метода ограничено большой вычислительной сложностью шага.

1.5. Мы видим, что сложность класса H_{x_i} допускает оценку через модуль его сильной выпуклости вне зависимости от прочих параметров ситуации:

$$N(\nu) \leqslant \left] Q \ln_+ \frac{1}{\nu} \right[. \quad (1.8)$$

Неясно, однако, насколько точна эта оценка. Забегая вперед, скажем, что она «правильна по ν», но «чересчур завышена» по Q. Правильная оценка есть $O(\sqrt{Q} \ln 1/\nu)$ *). Получение оценки такого рода (нижней и верхней) вместе с построением реализующего верхнюю оценку метода — цель следующих параграфов.

*) Если только размерность E достаточно велика.

§ 2. Квадратичное программирование и оценка снизу сложности сильно выпуклых классов

2.1. В этом параграфе мы изучим сложность простейшего модельного объекта сильно выпуклого программирования — класса квадратичных задач. Будем заниматься задачами вида

$$f_{A,b,c}(x) = \frac{1}{2}\langle Ax\,|\,x\rangle - \langle b\,|\,x\rangle + c \to \min\,|\,x \in E, \qquad (2.1)$$

где A — ограниченный неотрицательный эрмитов оператор (читатель, незнакомый с этими понятиями, может иметь в виду конечномерный случай; тогда A — неотрицательно определенная симметричная матрица), а $b \in E$. Задача $f_{A,b,c}$ (l_A, Q_A) -строго выпукла, где l_A — минимальная, $l_A Q_A$ — максимальная точки спектра A (в конечномерном случае — соответствующие собственные числа A). Обозначим множества всех таких задач $\boldsymbol{H}(E)$. Если $l_A = 0$, то $f_{A,b,c}$ — L_A -гладкая, L_A — максимальная точка спектра A.

Для решения задач такого рода имеется хорошо известный метод сопряженных градиентов (МСГ). Работа его, если не интересоваться алгоритмической стороной дела, выглядит следующим образом. Пусть $E_0 = \{0\}$ и для $i \geqslant 1$ $E_i(A, b) = \mathfrak{L}(b, Ab, \ldots, A^{i-1}b)$ (здесь и далее $\mathfrak{L}$ — замкнутая линейная оболочка перечисленных в скобках векторов). Тогда i-я точка траектории МСГ на $f_{A,b,c}$ есть точка минимума этой функции на $E_{i-1}(A, b)$, $i \geqslant 1$.

Оказывается, что МСГ в некотором точном смысле есть оптимальный метод решения квадратичных задач. Мы вначале докажем это, а затем оценим трудоемкость МСГ.

2.2. Пусть $\boldsymbol{H}'(E)$ — некоторый подкласс задач вида (2.1), таких, что $\min\limits_x f_{A,b,c}(x) = 0$. Скажем, что $\boldsymbol{H}'(E)$ *выдерживает вращения*, если из $f_{A,b,c} \in \boldsymbol{H}'(E)$ следует $f_{U^*AU,U^*b,c} \in \boldsymbol{H}'(E)$, каков бы ни был ортогональный оператор U. Иными словами, вместе со всякой функцией f $\boldsymbol{H}'(E)$ обязано содержать все ее «повороты». Для $f \in \boldsymbol{H}'(E)$ и $x \in E$ определим погрешности x в качестве решения f:

Относительная погрешность

$$\nu(x, f) = \frac{f(x)}{f(0)}$$

и *абсолютная погрешность*

$$\bar{\nu}(x, f) = f(x).$$

Пусть для решения задач из $\boldsymbol{H}'(E)$ применяются детерминированные методы первого порядка (т. е. оракул сообщает только $f(x)$, $f'(x)$). Обозначим через $I(\nu, \boldsymbol{H}'(E))$ (соответственно $J(\nu, \boldsymbol{H}'(E))$) сложностные функции класса $\boldsymbol{H}'(E)$ по отношению к указанному арсеналу методов, отвечающие соответственно относительному и абсолютному определениям погрешности. Пусть еще $I_0(\nu, \boldsymbol{H}'(E))$, $(J_0(\nu, \boldsymbol{H}'(E)))$ — минимальное N, при котором МСГ с N шагами еще обеспечивает относительную (соответственно абсолютную) погрешность, $\leqslant \nu$, в решении всех задач из $\boldsymbol{H}'(E)$.

Отмеченная выше субоптимальность МСГ состоит в следующем.

Теорема. *Пусть $\boldsymbol{H}'(E)$ выдерживает вращения. Тогда*

$$I_0(\nu, \boldsymbol{H}'(E)) \leqslant 2I(\nu, \boldsymbol{H}'(E)); \quad J_0(\nu, \boldsymbol{H}'(E)) \leqslant 2J(\nu, \boldsymbol{H}'(E))$$

(*т. е. никакой метод не может быть более чем вдвое лучшим, чем МСГ*).

Доказательство. Докажем первое неравенство. Предположим, что оно неверно. Это означает, что при некотором $\nu > 0$ и некотором натуральном M найдется правильный метод $\mathcal{B}$ решения задач из $\boldsymbol{H}'(E)$ с M шагами, решающий все задачи с точностью ν, тогда как МСГ с $2M$ шагами не обладает этим свойством. В частности, $\dim E \geqslant 2M$, ибо МСГ с $\dim E + 1$ шагами *точно* решает *любую* выпуклую квадратичную задачу (почему?).

Пусть $f_{A,b,c} \equiv f$ — та задача из $\boldsymbol{H}'(E)$, которую МСГ с $2M$ шагами не решает с точностью ν. Рассмотрим цепочку подпространств $E_i = E_i(A, b)$. Если при данном i $E_i = E_{i+1}$, то $E_{i+j} = E_i$ при всех $j \geqslant i$. Легко видеть (мы не будем доказывать этого факта линейной алгебры), что точка x^* минимума $f_{A,b,c}$ (т. е. $A^{-1}b$) лежит в $\mathfrak{L}(b, Ab, A^2b, \ldots)$. Поэтому при таком i $x^* \in E_i$. Стало быть, $i \geqslant 2M$ — иначе МСГ с $2M$ шагами решил бы $f_{A,b,c}$ точно.

Итак, $E_i \neq E_{i+1}$ для $0 \leqslant i \leqslant 2M - 1$. Пусть теперь g_i — единичный вектор, лежащий в E_{i+1} и ортогональный E_i, $i = 0, 1, \ldots, 2M - 1$. Поступим следующим образом. Пусть $\mathcal{B}$ задает первый вопрос в точке x_1. Найдем e_0, $|e_0| = 1$, так, чтобы было $x_1 \in \mathfrak{L}(e_0)$, и выберем в качестве e_1 любой единичный и ортогональный e_0 вектор. Пусть U_1 — любой ортогональный оператор, такой, что $U_1 e_0 = g_0$, $U_1 e_1 = g_1$.

Обозначим через x_2 вторую точку, в которой метод $\mathcal{B}$ задает вопрос о задаче $f_{U_1^* A U_1, U_1^* b, c}$. Продолжим построение. Пусть к n-му $(1 \leqslant n \leqslant M)$ шагу уже построены ортонормированные векторы $e_0, e_1, \ldots, e_{2n-4}, e_{2n-3}$ и ортогональный оператор U_{n-1}, такой, что $U_{n-1} e_j = g_j$, $0 \leqslant j \leqslant 2n - 3$. Обозначим через x_n n-ю точку, в которой $\mathcal{B}$ «задает вопрос» о задаче $f_{U_{n-1}^* A U_{n-1}, U_{n-1}^* b, c}$ (при $n = M$ — выдает результат). Выберем e_{2n-2}, e_{2n-1} так, чтобы было $x_n \in \mathfrak{L}(e_0, \ldots, e_{2n-2})$ и при этом $e_0, \ldots, e_{2n-1}$ образовывали ортонормированную систему. Такое построение возможно, если $2n \leqslant \dim E$, что заведомо так при $n \leqslant M$ (мы в этом уже убедились). Определим, далее, ортогональный оператор U_n условием $U_n e_j = g_j$, $0 \leqslant j \leqslant 2n-1$. Шаг окончен.

После M шагов описанной процедуры будут построены векторы $e_0, \ldots, e_{2M-1}$ и ортогональный оператор U_M, такой, что $U_M e_j = g_j$, $0 \leqslant j \leqslant 2M-1$. Убедимся, что $x_1, \ldots, x_M$ есть траектория $\mathscr{B}$ на задаче $f_{U_M^* A U_M, U_M^* b, c}$. По лемме о неразличении (см. § 5 гл. I) достаточно проверить, что информации, поставляемые оракулом о задачах $f_{U_M^* A U_M, U_M^* b, c}$ и $f_{U_j^* A U_j, U_j^* b, c}$, совпадают друг с другом в каждой из точек $x_1, x_2, \ldots, x_j$. По построению $x_1, \ldots, x_j \in \mathfrak{L}(e_0, \ldots, e_{2j-2})$. Стало быть, при $j \geqslant 1$ и $t \leqslant j$

$$f_{U_j^* A U_j, U_j^* b, c}(x_t) = \frac{1}{2}\langle A U_j x_t | U_j x_t\rangle - \langle b | U_j x_t\rangle + c =$$
$$= \frac{1}{2}\langle A U_M x_t | U_M x_t\rangle - \langle b | U_M x_t\rangle + c,$$

ибо U_j и U_M совпадают на $\mathfrak{L}(e_0, \ldots, e_{2j-2})$. Далее,

$$f'_{U_j^* A U_j, U_j^* b, c}(x_t) = U_j^* A U_j x_t - U_j^* b$$

и

$$f'_{U_M^* A U_M, U_M^* b, c}(x_t) = U_M^* A U_M x_t - U_M^* b,$$

U_M совпадает с U_j на $\mathfrak{L}(e_0, \ldots, e_{2j-2}) \ni x_t$ и переводит это подпространство в E_{2j-1}. A переводит E_{2j-1} в E_{2j}, и остается проверить, что $U_M^* = U_j^*$ на E_{2j}. Это так, поскольку $U_j^* = U_j^{-1}$ и $U_M^* = U_M^{-1}$, а $U_j^{-1} g_s = e_s = U_M^{-1} g_s$ при $s \leqslant 2j-1$.

Равным образом $U_j^{-1} b = U_M^{-1} b$, ибо $b \in E_1 \subset E_{2j}$ при всех $j \geqslant 1$.

Итак, x_M есть результат применения $\mathscr{B}$ к $f_{U_M^* A U_M, U_M^* b, c}$. Так как $\boldsymbol{H}'(E)$ выдерживает вращения, то по определению $\mathscr{B}$

$$\nu \geqslant \frac{f_{U_M^* A U_M, U_M^* b, c}(x_M)}{f_{U_M^* A U_M, U_M^* b, c}(0)} = \frac{f_{A, b, c}(U_M x_M)}{f_{A, b, c}(0)} \geqslant \frac{\inf\limits_{x \in E_{2M-1}} f_{A, b, c}(x)}{f_{A, b, c}(0)}$$

(мы учли, что $U_M x_M \in \mathfrak{L}(g_0, \ldots, g_{2M-2}) = E_{2M-1}$). Но последнее отношение $> \nu$ (оно равно погрешности МСГ с $2M$ шагами на $f_{A,b,c}$). Полученное противоречие доказывает первое неравенство.

Второе доказывается дословно так же. Теорема доказана.

2.3. Применим полученный результат к оценкам сложности классов гладких выпуклых и сильно выпуклых задач. Обозначим через $\boldsymbol{H}(E, Q)$ множество всех задач $f_{A,b,c} \in \boldsymbol{H}'(E)$, у которых спектр A лежит в отрезке $[1/Q, 1]$ и $|b| = 1$, а через $\overline{\boldsymbol{H}}(E)$ — множество всех задач $f_{A,b,c} \in \boldsymbol{H}'(E)$, решение которых имеет норму, $\leqslant 1$, и $|b| \leqslant 1$, $\|A\| \leqslant 1$. Отвечающие этим классам функции I_0, J_0, очевидно, зависят лишь от линейной размерности E (от того, равна ли она $+\infty$ или нет, и если нет — то какова она). Соответственно обозначим $I_0(\nu, n, Q)$ и $J_0(\nu, n)$ функции $I_0(\nu, \boldsymbol{H}(E, Q))$, $\dim E = n$, и $J_0(\nu, \overline{\boldsymbol{H}}(E))$, $\dim E = n$. Здесь $n = 1, 2, \ldots, \infty$. Имеют место следующие утверждения.

Теорема.

1) $I_0(\nu, n, Q) = \min\{n+1, I_0(\nu, \infty, Q)\}$,

$$J_0(\nu, n) = \min\{n+1, J_0(\nu, \infty)\}.$$

2) Пусть $x_1 \in \operatorname{int} G$ *и* $\rho_{|\cdot|}(x_1, \partial G) = r \leqslant \infty$. *Тогда сложность* $N(\nu)$ *класса* $\boldsymbol{H}_{x_1}(G, m; l_0, \ldots, l_m; Q_0, \ldots, Q_m)$ *с* $Q = \max\limits_i Q_i$ *допускает оценку снизу*

$$N(\nu) \geqslant \frac{1}{2} I_0(\nu, \dim E, Q) - 1.$$

3) Пусть $x_1 \in \operatorname{int} G$ *и* $\rho_{1\cdot 1}(x_1, \partial G) = r \leqslant \infty$. *Тогда сложность* $N(\nu)$ *класса* $\boldsymbol{H}_{x_1}^V(G, m; l_0, \ldots, l_m; Q_0, \ldots, Q_m)$ *допускает оценку снизу*

$$N(\nu) \geqslant \frac{1}{2} \max_{0 \leqslant i \leqslant m} I_0(\varkappa_i \nu, \dim E, Q_i) - 1,$$

где

$$\varkappa_i = \max\left\{1, \frac{2V}{l_i r^2}, \frac{l_i}{l_0}\right\}.$$

В частности, если при некотором i_0 *с* $Q_{i_0} = Q$ *имеем* $l_{i_0} \leqslant l_0$ *и* $V \leqslant {}^1/_2 l_{i_0} r^2$, *то*

$$N(\nu) \geqslant \frac{1}{2} I_0(\nu, \dim E, Q) - 1.$$

4) Пусть G *— шар с центром в* x_1. *Тогда сложность* $N(\nu)$ *класса* $\boldsymbol{CS}_{x_1}^V(G, m)$ *допускает оценку*

$$N(\nu) \geqslant \frac{1}{2} J_0(2\nu, \dim E) - 1.$$

Упражнение 1. Докажите теорему.

Приведенная теорема сводит проблему оценки сложностей рассматриваемых классов снизу к вычислению $I_0(\nu, \infty, Q)$ и $J_0(\nu, \infty)$, т. е. к оценке трудоемкости МСГ на некотором классе бесконечномерных квадратичных задач. Этим мы и займемся.

Удобнее оценивать не $I_0(\nu, \boldsymbol{H}'(E))$ для соответствующих $\boldsymbol{H}'(E)$, а обратную функцию

$\nu(n, \boldsymbol{H}'(E)) = \inf\{\nu \mid$ МСГ с n шагами решает все задачи из $\boldsymbol{H}'(E)$

с относительной погрешностью $\leqslant \nu\}$. Аналогично — с заменой относительной погрешности на абсолютную — определим $\bar{\nu}(n, \boldsymbol{H}'(E))$. Обозначим при $\dim E = \infty$

$$\nu(n, \boldsymbol{H}(E, Q)) = \nu(n, Q), \ \bar{\nu}(n, \overline{\boldsymbol{H}}(E)) = \bar{\nu}(n)$$

(определение корректно в том смысле, что левые части не зависят от реализации E с $\dim E = \infty$).

Исследование, проводимое в следующих двух пунктах, по характеру близко к проводимому в [17]. Результирующие нижние оценки типа (2.7), (2.12), по-видимому, впервые получены в [23—24].

2.4. Пусть $\Delta = [a, 1]$, $0 \leqslant a \leqslant 1$, а μ — неотрицательная борелева мера на Δ, $\mu(\Delta) \leqslant 1$. Положим $E = L_2(\Delta, \mu)$ и рассмотрим квадратичную задачу

$$f[x] = \frac{1}{2}\int_\Delta tx^2(t)\,d\mu(t) - \int_\Delta x(t)\,d\mu(t) + \frac{1}{2}\int \frac{1}{t}\,d\mu(t) \to \min \mid x \in E$$

(считаем μ такой, что

$$\left.\int \frac{d\mu(t)}{t^2} < \infty\right).$$

Посмотрим, как на ней работает МСГ. По определению i-я точка работы МСГ на f есть точка минимума этой функции на подпространстве E, натянутом на функции $1, t, t^2, \ldots, t^{i-2}$, т. е. на подпространстве $\mathscr{P}_{i-2}$ полиномов степени $\leqslant i-2$. Имеем при $1/t \in L_2(\Delta, \mu)$:

$$f[x] = \frac{1}{2}\int t\left(x(t) - \frac{1}{t}\right)^2 d\mu(t) \tag{2.2}$$

и $\inf\limits_x f[x] = 0$. Обозначим через $\nu(n, f)$ ($\bar{\nu}(n, f)$) относительную (абсолютную) погрешность МСГ с n шагами на f. Найдем

$$\left.\begin{aligned} \nu(n, f) &= \frac{\min\limits_{p \in \mathscr{P}_{n-2}} \int_\Delta t\left(p(t) - \frac{1}{t}\right)^2 d\mu(t)}{\int_\Delta \frac{1}{t}\,d\mu(t)}, \\ \bar{\nu}(n, f) &= \min_{p \in \mathscr{P}_{n-2}} \int_\Delta t\left(p(t) - \frac{1}{t}\right)^2 d\mu(t). \end{aligned}\right\} \tag{2.3}$$

Пусть $a = 1/Q$, а μ — произвольная вероятностная мера. Введем меру $\bar{\mu}(t)$:

$$d\bar{\mu}(t) = \frac{1}{t}\,d\mu(t).$$

Имеем

$$\nu(n, f) = \min_{p \in \mathscr{P}_{n-2}} \frac{\int_\Delta (tp(t) - 1)^2\,d\bar{\mu}(t)}{\int_\Delta d\bar{\mu}(t)}.$$

Далее, $f \in \boldsymbol{H}(E, Q)$, так что

$$\nu(n, Q) \geqslant \min_{p \in \mathscr{P}_{n-2}} \frac{\int_\Delta (tp(t) - 1)^2\,d\bar{\mu}(t)}{\int_\Delta d\bar{\mu}(t)} = \min_{p \in \Pi_{n-1}} \int p^2(t)\,d\bar{\bar{\mu}}(t),$$

где $\Pi_{n-1} = \{p \in \mathscr{P}_{n-1} \mid p(0) = 1\}$, а $\bar{\bar{\mu}}$ — вероятностная нормировка $\bar{\mu}$. При изменении μ $\bar{\bar{\mu}}$ пробегает множество $\mathscr{F}_\Delta$ всех вероятностных борелевых мер на Δ. Стало быть,

$$\nu(n, Q) \geqslant \max_{\gamma \in \mathscr{F}_\Delta} \min_{p \in \Pi_{n-1}} \int_\Delta p^2(t)\,d\gamma(t) =$$
$$= \min_{p \in \mathscr{F}_{n-1}} \max_{\gamma \in \mathscr{F}_\Delta} \int_\Delta p^2(t)\,d\gamma(t) = \min_{p \in \Pi_{n-1}} \max_{t \in \Delta} p^2(t)$$

(при перестановке порядка минимизации и максимизации мы использовали лемму фон Неймана). Итак,

$$\nu(n, Q) \geqslant \min_{p \in \Pi_{n-1}} \max_{t \in \Delta} p^2(t). \tag{2.4}$$

Фактически (2.4) — точное равенство, но нам этого не требуется.

Для оценки $\bar{\nu}(n)$ поступим аналогично. Пусть $a = 0$. Заметим, что если

$$\int \frac{1}{t^2}\,d\mu(t) \leqslant 1,$$

то $f \in \overline{\boldsymbol{H}}(E)$. Иными словами, пусть $d\mu(t) = t^2\,d\gamma(t)$, где γ — вероятностная мера на $[0, 1] = \Delta$, не сосредоточенная в 0. Тогда $\mu(\Delta) \leqslant 1$, так что функция f из (2.2) лежит в $\overline{\boldsymbol{H}}(E)$. Стало быть, ввиду (2.3),

$$\bar{\nu}(n, f) = \min_{p \in \mathscr{P}_{n-2}} \frac{1}{2}\int t(tp(t) - 1)^2\,d\gamma(t),$$

т. е.

$$\bar{\nu}(n) \geqslant \min_{p \in \Pi_{n-1}} \frac{1}{2}\int tp^2(t)\,d\gamma(t)$$

для всякой *вероятностной* меры γ. Следовательно, как и выше,

$$\bar{\nu}(n) \geqslant \max_{\gamma \in \mathscr{F}_{[0,1]}} \min_{p \in \Pi_{n-1}} \frac{1}{2}\int tp^2(t)\,d\gamma(t) = \min_{p \in \Pi_{n-1}} \frac{1}{2} \max_{0 \leqslant t \leqslant 1} tp^2(t).$$

Итак,

$$\bar{\nu}(n) \geqslant \frac{1}{2} \min_{p \in \Pi_{n-1}} \max_{0 \leqslant t \leqslant 1} tp^2(t). \tag{2.5}$$

Соотношение (2.5) на самом деле — точное равенство, но нам и это нужно.

2.5. Теперь мы в состоянии оценить $\nu(n, Q)$ и $\bar{\nu}(n)$. Начнем с первой из них. Из определения Π_{n-1} ясно, что

$$\min_{p\in\Pi_{n-1}} \max_{t\in\Delta} p^2(t) = \{ \max_{\substack{p\in\mathscr{P}_{n-1}\\ |p(t)|\leqslant 1 \text{ на } \Delta}} |p(0)|\}^{-2} = $$

$$= \left\{ \max_{\substack{p\in\mathscr{P}_{n-1}\\ |p(t)|\leqslant 1 \text{ на } [-1,1]}} \left| p\left(1 + \frac{2}{Q-1}\right)\right|\right\}^{-2}$$

(при последнем преобразовании мы растянули отрезок Δ на отрезок $[-1, 1]$). Таким образом, дело сводится к отысканию полинома степени $n-1$, значение которого в точке $\bar{t} = 1 + \frac{2}{Q-1}$ максимально, при условии, что на $[-1, 1]$ этот полином уклоняется от 0 не больше, чем на 1.

Решение этой задачи дается теоремой Маркова [1]. Соответствующий полином есть полином Чебышева $T_{n-1}(x)$ степени $n-1$ независимо от выбора $\bar{t} \notin [-1, 1]$. При этом для $t \geqslant 1$ справедливо представление

$$T_{n-1}(t) = \operatorname{ch}\{(n-1) \operatorname{arc} \operatorname{ch} t\}. \tag{2.6}$$

Итак,

$$\nu(n, Q) \geqslant \left[\operatorname{ch}\left\{(n-1)\operatorname{arc}\operatorname{ch}\left(1+\frac{2}{Q-1}\right)\right\}\right]^{-2},$$

что после элементарных оценок дает

$$I_0(\nu, \infty, Q) \geqslant \sqrt{\frac{Q-1}{16}} \ln\frac{1}{\nu} + 1. \tag{2.7}$$

Вычисления для $\bar{\nu}(n)$ чуть сложнее. Пусть

$$g(n) = \min_{\substack{p\in\mathscr{P}_{2n-1}\\ p(0)=0,\ p'(0)=1}} \max_{t\in[0,1]} |p(t)|. \tag{2.8}$$

Тогда **из** (2.5) следует

$$\bar{\nu}(n) \geqslant \frac{1}{2} g(n). \tag{2.9}$$

Функция $g(n)$ может быть легко вычислена. Именно, пусть $T_{2n-1}(x)$ — полином Чебышева степени $2n-1$, а x_n — его крайний правый нуль на $[-1, 1]$. Оказывается, что

$$g(n) = \frac{1}{|T'_{2n-1}(x_n)|(1+x_n)}. \tag{2.10}$$

Упражнение 2. Докажите (2.10).

Таким образом, учитывая, что $T_n(x) = \cos(n \arccos x)$, получаем

$$g(n) = \frac{\sin\left(\frac{\pi}{4(2n-1)}\right)}{2n-1}.$$

Итак,

$$\bar{\nu}(n) \geqslant \frac{\sin\frac{\pi}{4(2n-1)}}{2(2n-1)} > \frac{\pi\sqrt{2}}{16(2n-1)^2}, \tag{2.11}$$

что дает

$$J_0(\nu, \infty) \geqslant \frac{1}{4\sqrt{\nu}}. \tag{2.12}$$

Упражнение 3. Класс $\overline{\boldsymbol{H}}(E)$, сложностью которого фактически является $J_0(\nu, \dim E)$, содержит как угодно плохо обусловленные задачи (т. е. задачи, не лежащие в $\boldsymbol{H}(E, Q)$ при сколь угодно большом Q). Объясните, как согласовать неограниченный рост правой части (2.7) при $Q \to \infty$ с оценкой (2.12). ⟨Дело в том, что нормировки погрешностей при вычислении I_0 и J_0 были различными — в первом случае мы занимались относительными, во втором — абсолютными погрешностями.⟩

2.6. Учитывая утверждения теоремы 2.3 и оценки (2.7) и (2.12), приходим к следующей теореме.

Теорема. *Для некоторых абсолютных констант c_1, $c_2 > 0$ справедливы следующие нижние оценки детерминированной сложности $N(\nu)$ классов сильно выпуклых и гладких выпуклых задач на n-мерном гильбертовом пространстве.*

А. Для класса $\boldsymbol{H}_{x_1}(G, m; l_0, \ldots, l_m; Q_0, \ldots, Q_m)$ с $\max_i Q_i = Q$ при $x_1 \in \operatorname{int} G$

$$N(\nu) \geqslant c_1 \min\left\{n, \sqrt{Q-1}\ln\frac{1}{\nu}\right\} \text{ при } \nu < 1. \tag{2.13}$$

Б. Для класса $\boldsymbol{H}^V_{x_1}(G, m; l_0, \ldots, l_m; Q_0, \ldots, Q_m)$ с $\max_i Q_i = Q$ при $x_1 \in \operatorname{int} G$

$$N(\nu) \geqslant c_1 \min\left\{n, \sqrt{Q-1}\ln\frac{1}{\nu}\right\} \text{ при } \nu < \nu_0(\boldsymbol{H}^V_{x_1}). \tag{2.14}$$

Здесь $\nu_0(\boldsymbol{H}^V_{x_1}) > 0$. Если при некотором i_0 имеем

$$Q_{i_0} = Q,\ l_{i_0} \leqslant l_0,\ \frac{2V}{l_{i_0}\rho^2_{|\cdot|}(x_1, \partial G)} \leqslant 1,$$

то можно взять $\nu_0(\boldsymbol{H}^V_{x_1}) = 1$.

В. Для класса $\boldsymbol{CS}^V_{x_1}(G, m)$ в случае, когда G — шар с центром в x_1,

$$N(\nu) \geqslant c_2 \min\left\{n, \frac{1}{\sqrt{\nu}}\right\} \text{ при } \nu \leqslant 1. \tag{2.15}$$

З а м е ч а н и е. Доставляемые теоремой 2.6 оценки были установлены в предположении, что рассматриваемые классы оснащены точным оракулом первого порядка. Можно доказать, что те же оценки верны и для детерминированных сложностей классов, получающихся из описанных выше заменой этого оракула на любой локальный детерминированный оракул. Мы не будем приводить довольно утомительного доказательства этого утверждения.

2.7. Оценки теоремы 2.6 не могут правильно описывать поведение сложности при $\nu \to 0$ для случая $n \equiv \dim E < \infty$, поскольку их правые части ограничены величиной порядка $O(n)$. Это обстоятельство связано с особенностью модельных классов квадратичных задач, сложность которых по существу и есть правая часть оценок (2.13) — (2.15). В самом деле, любую квадратичную задачу можно точно решить за n шагов. Ясно, что эти оценки заведомо занижены для тех точностей ν, для которых вступает в силу «эффект размерности» (для (2.13) — (2.14) это те ν, для которых $\sqrt{Q-1}\,\ln 1/\nu \gg n$, а для (2.15) — те ν, для которых $1/\sqrt{\nu} \gg n$).

Априори могло бы оказаться, что (2.13) — (2.15) существенно занижены при всех ν. В следующих параграфах будет показано, что это на самом деле не так, по крайней мере, для классов $\boldsymbol{H}_{x_1}$ с $m = 0$, классов $\boldsymbol{H}^V_{x_1}$ и классов $\boldsymbol{CS}^V_{x_1}$. Мы научимся решать задачи этих классов с трудоемкостью соответственно

$$\sim \sqrt{Q}\cdot\Big]\ln\frac{1}{\nu}\Big[\cdot\Big]\ln Q\Big[,\quad \sim\sqrt{Q}\Big]\ln^2 Q\Big[\cdot\Big]\ln\frac{1}{\nu}\Big[,\sim\frac{1}{\sqrt{\nu}}\Big]\ln^3\frac{1}{\nu}\Big[.$$

Сопоставление этих верхних оценок сложности с нижними оценками (2.13) — (2.15) показывает, что последние «почти» (с точностью до логарифмических множителей) правильно описывают поведение сложностей $N(\nu)$ указанных классов в том диапазоне изменения ν и Q, в котором не действует «эффект размерности». Можно сказать, что оценки (2.13) — (2.15) в «чистом виде» выделяют влияние сильной выпуклости на сложность, показывают, что происходит в асимптотике по $\dim E \to \infty$. При конечных $\dim E$ могут возникать разного рода «посторонние» для рассматриваемых классов эффекты. Например, для решения задач $f \in \boldsymbol{H}_{x_1}$, $f \in \boldsymbol{H}^V_{x_1}$, $f \in \boldsymbol{CS}^V_{x_1}$ можно применять методы решения общих выпуклых задач, практически игнорирующие специфику рассматриваемых классов. Анализ этой возможности показывает, что при конечном $\dim E$ оценки (2.13) — (2.15) не могут быть точными для всех значений ν.

Приведем уточнения нижних оценок сложности классов сильно выпуклых и гладких выпуклых задач, действующие во всем диапазоне значений $\nu > 0$.

Теорема. *Существуют такие абсолютные константы* $\nu_0, c_1, c_2 > 0$, Q_0 *и функции* $n(Q, \nu) < \infty$, $n(\nu) < \infty$, *что детерминированная и стохастическая сложности классов* $\boldsymbol{H}_{x_1}(G, m; l_0, \ldots, l_m; Q_0, \ldots, Q_m)$, $\boldsymbol{H}^V_{x_1}(G, m; l_0, \ldots, l_m; Q_0, \ldots, Q_m)$, $\boldsymbol{CS}^V_{x_1}(G, m)$ *удовлетворяют следующим нижним оценкам, в которых* $n = \dim E$.

А. Для класса $\boldsymbol{H}_{x_1}(G, m; l_0, \ldots, l_m; Q_0, \ldots, Q_m)$ *с* $\max_i Q_i = Q$ *при* $x_1 \in \operatorname{int} G$ *и* $Q \geq 2$

$$N(\nu) > c_1 \frac{\min(n, \sqrt{Q})}{\ln\min(n, \sqrt{Q})}\ln\frac{1}{\nu} \quad \text{при} \quad \nu < 1. \tag{2.16}$$

При условии $n \geq (Q, \nu)$, $Q \geq Q_0$ *и* $\nu \leq \nu_0$ *справедлива также оценка*

$$\tilde{N}(\nu) \geq c_1\sqrt{Q}\ln\frac{1}{\nu}. \tag{2.17}$$

Б. Для класса $\boldsymbol{H}^V_{x_1}(G, m; l_0, \ldots, l_m; Q_0, \ldots, Q_m)$ *с* $\max_i Q_i = Q$ *при* $x_1 \in \operatorname{int} G$ *и* $Q \geq 2$

$$N(\nu) \geq c_1 \frac{\min(n, \sqrt{Q})}{\ln\min(n, \sqrt{Q})}\ln\frac{1}{\nu} \quad \text{при} \quad \nu \leq \nu_0(\boldsymbol{H}^V_{x_1}). \tag{2.18}$$

Здесь $\nu_0(\boldsymbol{H}^V_{x_1}) > 0$. *Если при некотором* i_0 *имеем*

$$Q_{i_0} = Q,\quad l_{i_0} \leq l_0,\quad \frac{2V}{l_{i_0}\rho^2_{|\cdot|}(x_1, \partial G)} \leq 1,$$

то можно взять $\nu_0(\boldsymbol{H}^V_{x_1}) = 1$. *При условии* $n \geq n(Q, \nu)$, $Q > Q_0$ *и* $\nu < \min\{\nu_0, \nu_0(\boldsymbol{H}^V_{x_1})\}$ *справедлива также оценка*

$$\tilde{N}(\nu) \geq c_1\sqrt{Q}\ln\frac{1}{\nu}. \tag{2.19}$$

В. Для класса $\boldsymbol{CS}^V_{x_1}(G, m)$ *в случае, когда* G — *шар с центром в* x_1,

$$N(\nu) \geq c_2\min\left\{\frac{1}{\sqrt{\nu}}, \frac{n}{\ln n}\ln\frac{1}{\nu}\right\} \quad \text{при} \quad \nu < \nu_0. \tag{2.20}$$

Если при этом $n > n(\nu)$, *то*

$$\tilde{N}(\nu) \geq \frac{c_2}{\sqrt{\nu}} \quad \text{при} \quad \nu < 1. \tag{2.21}$$

Отметим, что выписанные оценки $N(\nu)$, $\tilde{N}(\nu)$ верны и для классов, получающихся из рассмотренных заменой точного оракула первого порядка любым локальным детерминированным (для $\tilde{N}(\nu)$ — даже любым локальным) оракулом.

Доказательство теоремы 2.7 весьма громоздко, и мы его опускаем.

2.8. Полученные результаты показывают, за что следует бороться. Мы видим, что сильно выпуклые задачи нельзя решать сверхлинейно сходящимися методами (т. е. асимптотика сложности по $\nu \to +0$ не может быть по порядку лучшей, чем $O(\ln 1/\nu)$. С другой стороны, градиентный метод из п. 1.4 доставляет эту асимптотику. Однако сложность классов сильно выпуклого программирования определяется не только требуемой точностью, но и «обусловленностью» рассматриваемых классов — их модулем сильной выпуклости Q. Нижняя оценка сложности при этом (в случае $\dim E = \infty$) есть $O(\sqrt{Q} \ln 1/\nu)$ (считаем $Q \geqslant 2$), а оценка трудоемкости градиентного метода — $O(Q \ln 1/\nu)$.

Нижняя оценка, как уже отмечалось, точна, так что градиентный метод «чересчур чувствителен» к параметру обусловленности класса Q. Между тем плохо обусловленные задачи (с большим Q) встречаются достаточно часто, так что желательно модифицировать градиентный метод так, чтобы снизить его чувствительность к Q до теоретически возможного уровня. Эта задача практически решается в следующих трех параграфах.

Заметим еще, что результаты теоремы 2.7 показывают, что стохастическая и детерминированная сложности (вернее, их нижние оценки) рассматриваемых классов в асимптотике по $\dim E \to \infty$ ведут себя одинаково, так что и здесь, как и в случае общих выпуклых задач, детерминированные методы, «почти» реализующие нижние оценки $N(\nu)$ (а такие методы будут построены), «почти» реализуют и $\widetilde{N}(\nu)$, т. е. рандомизация их не может дать «радикального» эффекта, во всяком случае, при больших $\dim E$.

§ 3. Сильно выпуклое программирование: безусловные задачи

3.1. В этом параграфе будет построен метод решения сильно выпуклых задач вида $f_0(x) \to \min \mid x \in E$ (т. е. в наших обозначениях — задач класса $\boldsymbol{H}_{x_1}(E, 0; l_0; Q_0)$), имеющий оценку трудоемкости $O(\sqrt{Q} \ln Q \ln 1/\nu)$, что в силу теоремы 2.6 с точностью до логарифмического множителя $O(\ln Q)$ совпадает с нижней оценкой сложности классов (по крайней мере, для больших $\dim E$). Начнем с описания идеи метода. Построить метод с трудоемкостью $O(\sqrt{Q} \ln 1/\nu)$ — это все равно, что научиться вдвое уменьшать начальную невязку за $O(\sqrt{Q})$ шагов. МСГ умеет это делать на квадратичных задачах. Естественно попытаться распространить его на общий случай, сохранив это свойство.

Установим «истинные» свойства МСГ — по возможности наиболее простые, — благодаря которым он так быстро «половинит» невязку. После этого можно попытаться «дотянуть» эти свойства МСГ до требований, диктуемых общей ситуацией.

Оказывается, что МСГ «половинит» невязку за $O(\sqrt{Q})$ шагов по простым геометрическим обстоятельствам. Пусть $x_1 = 0$, $x_2, \ldots, x_M$ — первые M точек траектории МСГ на квадратичной задаче $f = f_{A, b, c} \in \boldsymbol{H}(E, Q)$, $p_i = f'(x_i)$. Тогда из описания МСГ ясно, что

1) $f(x_{i+1}) \leqslant f(x_i) - \frac{1}{2} p_i^2$;
2) p_i ортогонально $p_1 + \ldots + p_{i-1}$;
3) p_i ортогонально $x_i - x_1$;
4) $\langle p_i \mid x^* - x_i \rangle \leqslant f(x^*) - f(x_i)$, где x^* — точка минимума f;
5) если x^* — точка минимума f, то
$$V_f(x_1) = f(x_1) - f_* \geqslant \frac{1}{2Q}(x^* - x_1)^2.$$

Действительно, x_{i+1} есть точка минимума f на $E_i(A, b)$, которое содержит x_i и p_i. Поэтому
$$f(x_{i+1}) \leqslant f(x_i - p_i) \leqslant f(x_i) - p_i^2 + \frac{1}{2} p_i^2 = f(x_i) - \frac{1}{2} p_i^2$$
(мы учли (1.1)). Это доказывает п. 1). Утверждение п. 2) очевидно, ибо $p_1, \ldots, p_{i-1} \in E_{i-1}(A, b)$, тогда как p_i ортогонально $E_{i-1}(A, b)$ (поскольку x_i есть точка минимума f на $E_{i-1}(A, b)$). Наконец, $x_i \in E_{i-1}$, так что p_i ортогонально x_i, что доказывает п. 3). Неравенство п. 5) сразу следует из (1.1) при $x = x^*$, $y = x_1$, соотношение п. 4) очевидно в силу выпуклости f.

Оказывается, что уже из пп. 1) — 5) следует, что МСГ быстро «половинит» невязку. В самом деле, предположим, что
$$V_f(x_M) > \frac{1}{2} V_f(x_1).$$
Тогда в силу п. 1)
$$\sum_{i=1}^{M-1} p_i^2 \leqslant V_f(x_1). \tag{3.1}$$
Далее, в силу п. 4)
$$f(x^*) \geqslant f(x_i) + \langle p_i \mid x^* - x_i \rangle,$$
т. е. учитывая п. 3), имеем
$$\langle p_i \mid x^* - x_1 \rangle \leqslant -f(x_i) + f(x^*) = -V_f(x_i) \leqslant -V_f(x_M) < -\frac{V_f(x_1)}{2}.$$
Сложим полученные неравенства по $i = 1, \ldots, M-1$. Получим
$$-\left|\sum_{i=1}^{M-1} p_i\right| \cdot |x^* - x_1| \leqslant \left\langle \sum_{i=1}^{M-1} p_i \,\middle|\, x^* - x_1 \right\rangle \leqslant -\frac{(M-1)V_f(x_1)}{2},$$

т. е.

$$|x^* - x_1| \left| \sum_{i=1}^{M-1} p_i \right| \geqslant \frac{(M-1)\, \nu_f(x_1)}{2}.$$

В силу п. 5) это неравенство дает

$$\sqrt{2Q\nu_f(x_1)} \left| \sum_{i=1}^{M-1} p_i \right| \geqslant \frac{(M-1)\, \nu_f(x_1)}{2}. \quad (3.2)$$

Ввиду п. 2),

$$\left| \sum_{i=1}^{M-1} p_i \right| = \sqrt{\sum_{i=1}^{M-1} p_i^2} \leqslant \sqrt{\nu_f(x_1)}$$

(последнее — в силу (3.1)). Подставляя последнюю оценку в (3.2), получим

$$\sqrt{2Q}\, \nu_f(x_1) \geqslant \frac{(M-1)\, \nu_f(x_1)}{2}, \quad \text{или} \quad M-1 \leqslant 2\sqrt{2Q}.$$

Таким образом, к шагу с номером $2\sqrt{2Q}$ МСГ наверняка «споловинит» начальную невязку.

В этом рассуждении использовались лишь факты пп. 1) — 5). Посмотрим, можно ли обеспечить их справедливость в общей ситуации, когда f — любая $[1/Q,\ Q]$-сильно выпуклая функция и по-прежнему $p_i = f'(x_i)$. Соотношения пп. 4), 5) при этом выполняется автоматически — квадратичность f при их выводе из (1.1) не нужна. Чтобы выполнялся п. 1), достаточно сдвинуться из x_i в точку $\bar{x}_i = x_i - p_i$ (при этом f в силу (1.1) уменьшится не меньше, чем на $^1/_2\, p_i^2$) и после этого взять в качестве x_{i+1} какую-нибудь точку с $f(x_{i+1}) \leqslant f(\bar{x}_i)$.

Требуется лишь придумать, как добиться выполнения пп. 2) и 3) — ортогональности p_{i+1} паре векторов $x_{i+1} - x_1$ и $\sum_1^i p_j = q_i$. Для этого достаточно, чтобы x_{i+1} было точкой минимума f на каком-нибудь аффинном подпространстве $E^i \subset E$, содержащем точку x_1 и точку $x_1 + q_i$. Необходимо еще следить за выполнением неравенства $f(x_{i+1}) \leqslant f(\bar{x}_i)$, для чего достаточно, чтобы E^i содержало $\bar{x}_i$.

Итак, E^i должно содержать x_1, $x_1 + q_i$, $\bar{x}_i$, так что можно считать E^i двумерной плоскостью (аффинной), проходящей через три эти точки. Построение x_{i+1} сводится, таким образом, к минимизации f на двумерной плоскости. Но *двумерные* задачи мы умеем решать быстро (с помощью МЦТ или ММЦТ), вне зависимости от того, строго выпукла f или просто выпукла. Конечно, мы не можем решить задачу минимизации f на E^i точно, так что при реализации намеченного пути пп. 1) — 3) будут выполняться приближенно.

Нетрудно, впрочем, подсчитать точность, с которой надо решать двумерные задачи. Оказывается, что на их решение с этой точностью требуется порядка $O(\ln Q)$ шагов. Итак, построение x_{i+1} будет «стоить» $O(\ln Q)$ вопросов о задаче f — в точности цену двумерной оптимизации f. Всего точек x_i достаточно построить $O(\sqrt{Q})$ — при выполнении пп. 1) — 5) с нужной точностью за это время невязка «споловинится». Таким образом, можно «споловинить» невязку за время $O(\sqrt{Q}\ln Q)$, проигрывая квадратичному МСГ всего в $O(\ln Q)$ раз. «Споловинив» невязку, примем последнюю из построенных точек за новое значение x_1 *) и вновь повторим процесс, и т. д.

3.2. Теперь дадим точное описание метода. Метод применим к задачам класса $\boldsymbol{H}_{x_1}(E,\ 0;\ l_0;\ Q) \equiv \boldsymbol{H}_{x_1}$. Он состоит из *этапов*, этап — из *периодов*, период — из *шагов*.

1°. Пусть $\nu < 1$ — требуемая точность. Настроенный на эту точность метод состоит из

$$M_0(\nu) = \left] \log_{\frac{4}{3}} \frac{1}{\nu} \right[$$

этапов, устройство которых не зависит от ν. i-й ($i = 1, \ldots, M_0(\nu)$) этап имеет *вход* $\bar{x}_i$ и *выход* $\bar{x}_{i+1}$, являющийся (при $i < M_0(\nu)$) входом следующего этапа. Выход последнего этапа $\bar{x}_{M_0(\nu)+1}$ является результатом применения метода к решаемой задаче. Входом первого этапа является точка $\bar{x}_1 = x_1$.

2°. Все этапы устроены одинаково с точностью до значения входа. Опишем этап со входом $\bar{x}$. Обозначим через f решаемую задачу. Пусть $\bar{Q} = \max\{Q, 2\}$. Этап состоит из $M_1(Q) =]\, 2\sqrt{\bar{Q}}\,[$ периодов, i-й ($1 \leqslant i \leqslant M_1(Q)$) период имеет вход $(x_i,\ q_i)$, x_i, $q_i \in E$, и выход $(x_{i+1},\ q_{i+1})$, являющийся — при $i < M_1(Q)$ — входом $(i+1)$-го периода. При этом $x_1 = \bar{x}$, $q_1 = 0$, тогда как $x_{M_1(Q)+1}$ есть выход этапа.

3°. Период с номером i устроен следующим образом. Пусть $(x_i,\ q_i)$ — его вход. Обозначим через E^i аффинное подпространство E, натянутое на точки $\bar{x}$, x_i, $\bar{x} + q_i$, а через V_i — шар радиуса $r = (2/l_0)|f'(\bar{x})|$ с центром в $\bar{x}$. На i-м периоде проводятся следующие действия.

1) С помощью МЦТ или ММЦТ, настроенных на относительную точность $\delta = (1/96Q\bar{Q})^2$, решается задача минимизации f на $E^i \cap V_i$. Пусть $M_2(Q)$ — число шагов соответствующего метода. Тогда в силу результатов главы II $M_2(Q) \leqslant c \ln \bar{Q}$, c — абсолютная константа.

2) Пусть $\tilde{x}_i$ — результат работы «двумерного» метода. Если $f(\tilde{x}_i) > f(x_i)$, то положим $\hat{x}_i = x_i$, иначе $\hat{x}_i = \tilde{x}_i$.

*) В «аналитических» аналогах МСГ этот прием называется *восстановлением* (см. главу VIII).

Зададим в точке $\hat{x}_i$ вопрос о задаче f (фактически это уже сделано раньше — так устроены ММЦТ и МЦТ). Пусть p_i есть проекция $f'(\hat{x}_i)$ на ортогональное дополнение к $E^i - \bar{x}$. Положим

$$x_{i+1} = \hat{x}_i - f'(\hat{x}_i)/l_0 Q, \quad q_{i+1} = q_i + p_i.$$

Период закончен.

3.3. Теорема. *Описанный метод, настроенный на точность $\nu < 1$, решает все задачи из $\boldsymbol{H}_{x_1}(E, 0; l_0; Q_0)$ с погрешностью $\leqslant \nu$, при трудоемкости*

$$M(\nu) = M_0(\nu) M_1(Q) M_2(Q) \leqslant c_1] \ln \frac{1}{\nu} [\sqrt{\bar{Q}} \ln \bar{Q},$$

где $c_1 > 0$ — абсолютная константа и $\bar{Q} = \max(Q, 2)$.

Доказательство. Оценка трудоемкости ясна из описания. Докажем, что метод обеспечивает требуемую точность. Пусть f — решаемая задача. Достаточно доказать, что на этапе с входом $\bar{x}$ и выходом $\bar{y}$ имеем

$$\mathsf{V}_f(y) \leqslant {}^3/_4 \mathsf{V}_f(\bar{x}). \tag{3.3}$$

Пусть это не так. По построению

$$f(\bar{x}) \geqslant f(x_1) \geqslant f(\hat{x}_1) \geqslant \\ \geqslant f(x_2) \geqslant \ldots \geqslant f(\hat{x}_{M_1}) \geqslant f(x_{M_1+1}) = f(\bar{y}). \tag{3.4}$$

Здесь $M_1 = M_1(Q)$. Поэтому $\mathsf{V}_f(\bar{y}) \leqslant \mathsf{V}_f(\bar{x})$, так что из невыполнения (3.3) $f'(\bar{x}) \neq 0$ и $\mathsf{V}_f(\bar{x}) \neq 0$. Далее, в силу (1.1)

$$f(x+t) \leqslant f(x) + \langle f'(x) | t\rangle + {}^1/_2\, l_0 Q t^2,$$

откуда

$$f(x_{i+1}) \leqslant f(\hat{x}_i) - \frac{p_i^2}{2l_0 Q} \leqslant f(x_i) - \frac{p_i^2}{2l_0 Q},$$

что в силу (3.4) и невыполнения (3.3) приводит к

$$\sum_{i=1}^{M_1} p_i^2 \leqslant 2l_0 Q (\mathsf{V}_f(\bar{x}) - \mathsf{V}_f(\bar{y})) < {}^1/_2 l_0 Q \mathsf{V}_f(\bar{x}). \tag{3.5}$$

Кроме того, ввиду (1.1),

$$f(\bar{x} + h) \geqslant f(\bar{x}) - |f'(\bar{x})||h| + {}^1/_2\, l_0 |h|^2,$$

так что вне V_i имеем $f(x) \geqslant f(\bar{x})$. Поэтому $\min\limits_{V_i \cap E^i} f = \min\limits_{E^i} f$. С другой стороны, изменение f на V_i не превосходит, очевидно,

$$2r|f'(\bar{x})| + \frac{1}{2} l_0 Q r^2 = \frac{4}{l_0}|f'(\bar{x})|^2 + \frac{2}{l_0} Q |f'(\bar{x})|^2 = \\ = \frac{4+2Q}{l_0}|f'(\bar{x})|^2 \leqslant \frac{4\bar{Q}}{l_0}|f'(\bar{x})|^2.$$

Следовательно,

$$f(\hat{x}_i) - \inf_{E^i} f \leqslant \delta \frac{4\bar{Q}}{l_0} |f'(\bar{x})|^2. \tag{3.6}$$

Пусть теперь $e \in E^i - \bar{x}$ — любой единичный вектор. Тогда

$$f(\hat{x}_i + te) \leqslant f(\hat{x}_i) + t\langle f'(\hat{x}_i) | e\rangle + \frac{l_0 Q}{2} t^2,$$

откуда в силу (3.6)

$$\frac{\langle f'(\hat{x}_i) | e\rangle^2}{2l_0 Q} \leqslant \delta \frac{4\bar{Q}}{l_0} |f'(\bar{x})|^2,$$

т. е.

$$|P_i f'(\hat{x}_i)| \leqslant 3\sqrt{\delta}\, \bar{Q}\, |f'(\bar{x})|, \tag{3.7}$$

где P_i — ортопроектор на $E^i - \bar{x}$. Так как $f(\hat{x}^i) \leqslant f(\bar{x}_i)$, то $|x_i - \bar{x}| \leqslant r$. Аналогично, для точки x^* минимума f на E имеем $|x^* - \bar{x}| \leqslant r$. С другой стороны,

$$f(x^*) \geqslant f(\hat{x}_i) + \langle f'(x_i) | x^* - \hat{x}_i\rangle,$$

или

$$\langle p_i | x^* - \bar{x}\rangle = \langle p_i | x^* - \hat{x}_i\rangle < f(x^*) - f(\hat{x}_i) + \\ + |p_i - f'(\hat{x}_i)||x^* - \hat{x}_i| < \\ < -{}^3/_4 V_f(\bar{x}) + 6\sqrt{\delta}\,\bar{Q}\,|f'(\bar{x})|r = -{}^3/_4 V_f(\bar{x}) + \\ + \frac{12\bar{Q}}{l_0}\sqrt{\delta}\,|f'(\bar{x})|^2. \tag{3.8}$$

Заметим еще, что $\mathsf{V}_f(\bar{x}) \geqslant f(\bar{x}) - f\left(\bar{x} - \frac{f'(\bar{x})}{l_0 Q}\right) \geqslant |f'(\bar{x})|^2/2l_0 Q$ (мы учли (1.1)). Стало быть,

$$\frac{12\bar{Q}}{l_0}\sqrt{\delta}\,|f'(\bar{x})|^2 \leqslant 24 Q\bar{Q}\sqrt{\delta}\mathsf{V}_f(\bar{x}) \leqslant {}^1/_4 \mathsf{V}_f(\bar{x})$$

по выбору δ. Следовательно, из (3.8) получаем

$$\langle p_i | x^* - \bar{x}\rangle < -\frac{\mathsf{V}_f(\bar{x})}{2}.$$

Суммируя по $i = 1, 2, \ldots, M_1$, получим

$$\langle q_{M_1+1} | x^* - \bar{x}\rangle < -{}^1/_2 M_1 \mathsf{V}_f(\bar{x}). \tag{3.9}$$

Но p_i ортогонально q_i, так что, ввиду (3.5),

$$|q_{M_1+1}| \leqslant \sqrt{\sum_{i=1}^{M_1} p_i^2} \leqslant \sqrt{\frac{l_0 Q}{2} \mathsf{V}_f(\bar{x})}.$$

Кроме того, из (1.1) ясно, что $V_f(\bar{x}) \geqslant {}^1/_2\, l_0\, |x^* - \bar{x}|^2$ или

$$|x^* - \bar{x}| \leqslant \sqrt{\frac{2}{l_0} V_f(\bar{x})}.$$

Учитывая эти оценки и (3.9), получаем

$$-\sqrt{{}^1/_2 l_0 Q V_f(\bar{x})}\sqrt{\frac{2}{l_0} V_f(\bar{x})} < -{}^1/_2 M_1 V_f(\bar{x}),$$

т. е. $M_1 < 2\sqrt{Q}$, что противоречит определению M_1. Теорема доказана.

Замечание. Можно было бы настраивать этап на уменьшение невязки не в ${}^4/_3$ раза, как мы это делали, а на ее уменьшение в некоторое число раз $\theta > 1$. Выбором θ можно было бы оптимизировать абсолютную константу c_1 в оценке трудоемкости метода, но мы не будем проводить этого громоздкого анализа.

§ 4. Задача на минимакс

4.1. Распространим теперь метод предыдущего параграфа на случай задач с ограничениями. Сделаем это в два этапа — вначале научимся решать задачи на минимакс, т. е. задачи вида

$$\bar{f}(x) = \max_{0 \leqslant i \leqslant m} f_i(x) \to \min \,|\, x \in G, \tag{4.1}$$

где

$f(x) = (f_0(x), \ldots, f_m(x)) \in \boldsymbol{H}_{x_1}(G, m;\ l_0, \ldots, l_m;\ Q_0, \ldots, Q_m)$.

Затем редуцируем решение задач с ограничениями к решению задач на минимакс. Последнее будет проделано в следующем параграфе.

Заметим, что уметь хорошо решать задачи (4.1) полезно само по себе. Во-первых, к таким задачам сводится решение систем целевых неравенств — отыскание $x \in G$, для которого $f_0(x) \leqslant a_0, \ldots, f_m(x) \leqslant a_m$, a_i — заданные числа. Действительно, отыскание такого x можно свести к задаче минимизации функции $\max\limits_{0 \leqslant i \leqslant m} (f_i(x) - a_i)$. Далее, решение условно-экстремальной задачи при априори известном оптимальном значении целевого функционала сводится к решению системы целевых неравенств, т. е. в конечном счете — к задаче типа (4.1). Наконец, при $m = 0$ задача (4.1) есть задача минимизации f в области G. При $G \neq E$ мы пока еще не умели хорошо решать задачи такого рода.

4.2. Определим погрешность точки $x \in G$ в качестве решения задачи (4.1) как

$$\nu(x, f) = \frac{V_{\bar{f}}(x)}{V_{\bar{f}}(x_1)}, \tag{4.2}$$

где $V_{\bar{f}}(x) = \bar{f}(x) - \min\limits_G \bar{f}$. Требуется построить метод решения задач на минимакс $\bar{f}$, ассоциированных с функциями $f \in \boldsymbol{H}_{x_1}(G, m;\ l_0, \ldots, l_m;\ Q_0, \ldots, Q_m)$, использующий оракул первого порядка и решающий все такие задачи с заданной точностью ν при возможно меньшей трудоемкости — желательно порядка $O(\sqrt{Q} \ln(1/\nu))$, $Q = \max\limits_{0 \leqslant i \leqslant m} Q_i$ — модуль выпуклости f (результаты § 2 показывают, что большего, по крайней мере в бесконечномерной ситуации, требовать нельзя). Фактически этот метод будет иметь оценку трудоемкости $O(\sqrt{Q} \ln^2 Q \ln 1/\nu)$.

Наметим идею метода. При изложении наводящих соображений мы будем ссылаться на некоторое количество нуждающихся в обосновании фактов. Предлагаем читателю временно в них поверить. Разумеется, при обосновании метода будут даны полные доказательства.

Как и в случае задач без ограничений, достаточно уметь быстро «половинить» невязку $V_{\bar{f}}(x)$. Для этого мы будем исходить из тех же соображений, что и в § 3 — именно, предложим метод, генерирующий наряду с последовательностью x_i последовательность векторов p_i, удовлетворяющую пп. 1) — 5) п. 3.1. Пусть для простоты $l_0 = \ldots = l_m = 1/Q$, $Q_0 = \ldots = Q_m = Q$. В § 3 в качестве p_i брались векторы $f'(x_i)$. При этом п. 1) обеспечивался тем, что при сдвиге вдоль антиградиента f_0 убывала нужным образом. Условия пп. 2) и 3) достигались за счет того, что x_i было точкой минимума f_0 на проходящей через $\bar{x}$ плоскости, содержащей направление $\sum\limits_{j=1}^{i-1} p_j$. Теперь ситуация осложнилась: функция $\bar{f}$ вовсе не обязана убывать при сдвиге вдоль своего антиградиента — ведь она не является гладкой. Поэтому выбор $p_i = \bar{f}'(x_i)$ нас не устраивает.

Чтобы понять, что является «естественным» с точки зрения предъявляемых требований определением «градиента» $\bar{f}$, заметим следующее. При $m = 0$ градиент $f'(x)$ $[1/Q, Q]$-строго выпуклой функции f можно определить так: рассмотрим естественную мажоранту функции $f(y)$ и найдем ее точку минимума $\bar{x}$. Тогда градиент есть $x - \bar{x}$. Естественное обобщение этого определения на случай функции $\bar{f}(x) = \max\limits_{0 \leqslant i \leqslant m} f_i(x)$, f_i $[1/Q, Q]$-строго выпуклы, таково: для данного x пусть

$$f_x(y) = \max_{0 \leqslant i \leqslant m} \left\{ f_i(x) + \langle f_i'(x) \,|\, y - x\rangle + \frac{(y - x)^2}{2} \right\}$$

— естественная мажоранта $\bar{f}(y)$, и пусть Tx — точка минимума $f_x(y)$. Тогда «градиентом» $\bar{f}$ в точке x назовем вектор $p(x) = x - Tx$. Вектор $p(x)$ не является опорным функционалом к $\bar{f}$

в точке x. Однако можно показать, что он обладает следующими свойствами:

1) $\bar f(Tx) \leqslant \bar f(x) - \frac{1}{2} p^2(x)$;

2) $\langle p(x) \mid x^* - x\rangle \leqslant \bar f(x^*) - \bar f(Tx)$,

где x^* — точка минимума $\bar f$. Отсюда ясно, что при выборе $p_i = p(x_i)$ и таком построении x_{i+1}, что $\bar f(x_{i+1}) \leqslant \bar f(Tx_i)$, соотношение п. 1) из п. 3.1 будет выполняться. Соотношение п. 5) выполняется автоматически. Остается обеспечить пп. 2), 3) и 4). Это-то и представляет наибольшую трудность.

В ситуации § 3 (т. е. при $m = 0$ и $p(x) = f_0'(x)$) выполнение пп. 2) и 3) обеспечивалось тем, что в качестве x_{i+1} выбиралась точка плоскости E^i, проходящая через x_1, Tx_i и $x_1 + \sum_1^i p_j$, в которой $p(x_{i+1})$ было ортогонально $E^i - x_1$. При таком выборе автоматически было $f_0(x_{i+1}) \leqslant f_0(Tx_i)$. С другой стороны, указанный выбор сводился к простой задаче минимизации f_0 на E^i.

В общем случае $(m > 0)$ выбор x_{i+1} по правилу $p(x_{i+1}) \perp E^i - x_1$, $x_{i+1} \in E^i$ также нас устроил бы: при этом автоматически было бы $\bar f(x_{i+1}) \leqslant \bar f(Tx_i)$ и пп. 1) — 5) выполнялись бы. К сожалению, $p(x)$ не есть поле градиента, так что неясно, как осуществить указанный выбор x_{i+1}. Поэтому придется поступить более хитрым образом.

Предположим, что мы умеем строить семейство точек $x_i^j \in E^i$, $j = 1, \ldots, M$, так что

а) $\bar f(Tx_i) - \bar f(Tx_i^j) \geqslant \frac{1}{2} p^2(x_i^j)$;

б) $\langle p(x_i^j) \mid Tx_i - x_i^j\rangle = 0$;

в) некоторая выпуклая комбинация векторов $p(x_i^j)$ — именно, $\sum_j \mu_j p(x_i^j)$ — ортогональна $E^i - x_1$. Убедимся, что можно взять $p_{i+1} = \sum_j \mu_j p(x_i^j)$ и выбирать в качестве x_{i+1} лучшую (по значению $\bar f$) среди точек $\{Tx_i^j\}_{j=1}^M$.

Действительно, при таком выборе p_{i+1} будем иметь, в силу п. а)

$$\frac{1}{2} p_{i+1}^2 \leqslant \frac{1}{2} \sum_j p^2(x_i^j)\mu_j \leqslant \sum_j \mu_j (\bar f(Tx_i) - \bar f(Tx_i^j)) \leqslant$$
$$\leqslant \sum_j \mu_j(\bar f(Tx_i) - \bar f(x_{i+1})) \leqslant \bar f(x_i) - \bar f(x_{i+1}),$$

так что п. 1) выполнится.

Далее, p_{i+1} ортогонально $E^i - x_1$ по предположению, так что имеют место пп. 2) и 3). Остается убедиться, что выполнен п. 4), вернее, достаточная для рассуждений п. 3.1 версия п. 4') этого соотношения. Но в силу п. 2)

$$\langle p(x_i^j) \mid x^* - x_i^j\rangle \leqslant \bar f(x^*) - \bar f(Tx_i^j) \leqslant \bar f(x^*) - \bar f(x_{i+1}).$$

Ввиду п. б) левая часть равна $\langle p(x_i^j) \mid x^* - Tx_i\rangle$. Итак,

$$\langle p(x_i^j) \mid x^* - Tx_i\rangle \leqslant \bar f(x^*) - \bar f(x_{i+1}),$$

откуда

$$\langle p_{i+1} \mid x^* - Tx_i\rangle \leqslant \bar f(x^*) - \bar f(x_{i+1}),$$

или, поскольку $p_{i+1} \perp E^i - x_1 \ni Tx_i - x_1$,

$$\langle p_{i+1} \mid x^* - x_1\rangle \leqslant \bar f(x^*) - \bar f(x_{i+1}).$$

Это и есть п. 4'). Легко видеть, что п. 4' с успехом заменяет п. 4) в рассуждениях п. 3.1.

Остается научиться строить точки x_i^j. Можно заметить, что п. а) следует из п. б), так что надо следить лишь за пп. б) и в). Для каждого $e \in E^i - Tx_i$ на прямой $\{Tx_i + te \mid t \in \mathbb{R}\}$ имеется точка $x(e)$, для которой $\langle p(x(e)) \mid e\rangle = 0$. Это следует из того, что $\langle p(Tx_i + te) \mid e\rangle$ меняет знак при изменении t от $-\infty$ до $+\infty$. Точку $x(e)$ легко найти (с любой степенью точности), так как ее отыскание сводится к отысканию корня меняющей знак функции и может быть проделано «делением пополам». Таким образом, для каждого $e \in E^i - x_1$ можно отыскать $x(e)$ так, что п. б) будет выполняться при подстановке в него $x(e)$ вместо x_i^j.

Остается придумать такую последовательную процедуру выбора $e_j \in E^i - x_1$, чтобы при некотором M нам удалось удовлетворить п. в) *). Мы желаем добиться того, чтобы было

$$\Big|\sum P_i \sum_j \mu_j p(x(e_j))\Big| < \varepsilon$$

при достаточно малом ε. Здесь P_i — ортопроектор на $E^i - x_1$, $\Sigma\mu_j = 1$, $\mu_j \geqslant 0$. Заметим, что по лемме фон Неймана

$$\inf_{\mu_j \geqslant 0,\ \sum_j \mu_j = 1} \Big| P_i \sum_j \mu_j p(x(e_j))\Big| = \sup_{\substack{e \in E^i - x_1 \\ |e| \leqslant 1}} \min_{j \leqslant m} \langle e \mid p(x(e_j))\rangle. \tag{4.3}$$

Стало быть, достаточно организовать выбор e_j так, чтобы правая часть (4.3) стала меньше ε при надлежащем, по возможности меньшем M. При данном s множество

$$K_s = \Big\{e \in E^i - x_1 \,\Big|\, |e| = 1,\ \frac{\varepsilon}{2} \leqslant \langle e \mid p(x(e_j))\rangle \text{ при } j = 1, 2, \ldots, s\Big\}$$

представляет собой дугу. Если в качестве e_{s+1} выбирать орт, направленный в середину этой дуги, то дуга K_{s+1} будет, по крайней

*) Разумеется, не в точности, но достаточно хорошо.

мере, вдвое короче K_s (ибо $K_{s+1} \subset K_s$ и K_{s+1} не содержит окрестности середины K_s, ввиду $\langle e_{s+1} \mid p(x(e_{s+1}))\rangle = 0$). Таким образом, дуги K_s уменьшаются при такой дисциплине выбора каждый раз более чем вдвое. Отсюда легко вывести, что через время M, грубо говоря, пропорциональное $\ln(1/\varepsilon)$, K_M станет пусто, т. е. при таком M левая часть (4.3) станет меньше ε. Тогда п. в) выполнится «с точностью ε» при надлежащем выборе μ_j.

Таким образом, отыскание p_{i+1} и x_{i+1} требует некоторого процесса «двойного деления пополам». «Цена» этого процесса (в вопросах о f) оказывается порядка $O(\ln^2 Q)$ ($O(\ln Q)$ — «цена» отыскания $x(e_j)$ с надлежащей точностью. Таких точек, в свою очередь, должно быть найдено $O(\ln Q)$).

Реализация изложенных соображений сталкивается с известными трудностями — как из-за наличия области G, так и из-за неодинаковости l_i, $0 \leqslant i \leqslant m$. Ценой некоторых дополнительных ухищрений эти трудности удается преодолеть.

4.3. Перейдем к точному описанию метода. Фиксируем класс $\boldsymbol{H}_{x_1} \equiv \boldsymbol{H}_{x_1}(G, m; l_0, \ldots, l_m; Q_0, \ldots, Q_m)$. Положим $Q = \max\limits_{0 \leqslant i < m} Q_i$. Введем следующие обозначения.

1°. Для $f \in \boldsymbol{H}_{x_1}$ и $x \in E$ пусть

$$f_{i,x}(y) = f_i(x) + \langle f_i'(x) \mid y - x\rangle + \frac{1}{2} l_i Q (y - x)^2$$

и

$$f_x(y) = \max_{0 \leqslant i \leqslant m} f_{i,x}(y).$$

2°. Пусть Tx — точка минимума $f_x(y)$ на G. Так как $f_x(y)$ строго выпукла, то Tx определенно однозначно. Положим $p(x) = x - Tx$. Заметим, что вычисление Tx и $f_x(Tx)$ требует задания одного вопроса о f в точке x.

Описание метода.

1°. Метод, настроенный на точность $\nu < 1$, состоит из

$$M_0(\nu) = \left]\ln_{\frac{4}{3}} \frac{1}{\nu}\right[$$

этапов. Этап состоит из $M_1(Q)$ *периодов*, период — из $\leqslant M_2(Q)$ *процедур*, процедура — из $M_3(Q)$ *шагов*. При этом

$$\left.\begin{aligned} M_1(Q) &=]\sqrt{\bar{Q}}[, \\ M_2(Q) &= 4 +]^1/_2 M_3(Q)[, \\ M_3(Q) &=]8\log_2(100\bar{Q})[, \ \bar{Q} = \max(Q, 2). \end{aligned}\right\} \qquad (4.4)$$

2°. i-й этап имеет *вход* $\bar{x}_i$ и *выход* $\bar{x}_{i+1}$, $i = 1, 2, \ldots, M_0(\nu)$. При этом $\bar{x}_1 = x_1$, $\bar{x}_i \in G$. Выход i-го этапа есть вход $(i+1)$-го (при $i < M_0(\nu)$) или результат работы метода (при $i = M_0(\nu)$).

Все этапы устроены одинаково с точностью до значения входа. Опишем этап со входом $\bar{x} \in G$.

3°. Этап со входом $\bar{x}$ состоит из периодов, i-й период ($i = 1, 2, \ldots$) имеет вход (x_i, q_i) и выход (x_{i+1}, q_{i+1}), являющийся (при $i < M_1(Q)$) входом следующего периода. При этом $x_i \in G$, $q_i \in E$, $x_1 = \bar{x}$, $q_1 = 0$. Выходом этапа является точка $x_{M_1(Q)+1}$. На i-м периоде производятся следующие действия.

А. Строится двумерное аффинное подпространство E^i пространства E, содержащее точки $\bar{x}$, x_i и $\bar{x} + q_i$. Обозначим E_0^i двумерное линейное подпространство $E^i - \bar{x}$.

Б. В точке x_i задается вопрос о задаче f. Строится число $L_i = (\sqrt{5} + 1)\,\bar{Q} \mid p(x_i) \mid$, $\bar{Q} = \max\{Q, 2\}$. Если $L_i = 0$, то работа метода прекращается выдачей результата x_i. В противном случае делается $\leqslant M_2(Q)$ процедур в соответствии с правилом В.

В. *В.1.* j-я процедура i-го периода имеет вход $e_i^{j-1} \in E_0^i$, $| e_i^{j-1} | = 1$. При этом e_i^0 выбирается произвольным единичным вектором в E_0^i.

В.2. Пусть $\varphi_{ij}(t) = \langle p(x_i + te_i^{j-1}) \mid e_i^{j-1}\rangle$. Заметим, что вычисление φ_{ij} в наперед заданной точке t требует одного вопроса о f.

На j-й процедуре производятся следующие операции.

В.2.1. Вычисляется $\varphi_{ij}(-L_i)\varphi_{ij}(L_i)$. Если $\varphi_{ij}(-L_i)\varphi_{ij}(L_i) > 0$, то этап прекращается с выходом $\bar{y} = Tx_i$. В противном случае переходим к п. В.2.2.

В.2.2. Если $\varphi_{ij}(-L_i)\varphi_{ij}(L_i) \leqslant 0$, то делаем $M_3(Q) + 1$ итераций решения уравнения

$$\varphi_{ij}(t) = 0, \ |t| \leqslant L_i,$$

методом «деления пополам».

В.2.3. Пусть $\Delta_i^j = [\underline{t}_{ij}, \bar{t}_{ij}]$ — последний отрезок локализации корня, полученный после В.2.2. Положим $t_{ij} = 0$, если $\Delta_i^j \ni 0$. В противном случае выберем t_{ij} равным $\underline{t}_{ij}$ или $\bar{t}_{ij}$ так, чтобы было $t_{ij}\varphi_{ij}(t_{ij}) \leqslant 0$ (это можно сделать, ибо $\varphi_{ij}(\underline{t}_{ij})\varphi_{ij}(\bar{t}_{ij}) \leqslant 0$). После этого положим

$$p_{ij} = p(x_i + t_{ij}e_i^{j-1}), \ x_{ij} = x_i + t_{ij}e_i^{j-1},$$

j-я процедура закончена.

В.3. После окончания j-й процедуры положим

$$K_j = \left\{e \in E_0^i \,\middle|\, \langle e \mid p_{is}\rangle > \frac{16 L_i |e|}{2^{1/2} M_3(Q)}, \quad 1 \leqslant s \leqslant j\right\}.$$

Если $K_j = \{\phi\}$, или $j = M_2(Q)$, то перейдем к $Г$. В противном случае K_j есть острый угол на плоскости E_0^i. Пусть e_i^j — орт его биссектрисы. Перейдем к процедуре $j + 1$.

Г. Пусть $M_i \geqslant 1$ — число векторов p_{ij}, построенных по правилам В. Обозначим через P_i ортопроектор на плоскость E_0^i. Решим задачу

$$\left|\sum_{j=1}^{M_i} \mu_j P_i p_{ij}\right| \to \min \mid \mu_j \geqslant 0, \ \sum_j \mu_j = 1.$$

Пусть $\{\mu_j\}$ — ее решение. Далее поступим следующим образом.

Г.1. Если при некотором j было $p_{ij} = 0$, то прервем процесс решения и выдадим в качестве результата x_{ij}. Если это не так, перейдем к п. Г.2.

Г.2. Положим

$$y_{ij} = T(Tx_{ij}), \quad s_{ij} = \frac{2(\bar f(x_i) - \bar f(y_{ij}))}{p_{ij}^2},$$

и пусть x_{i+1} — лучшая (по значению $\bar f(\cdot)$) среди точек y_{ij}, $T(Tx_i)$. Заметим, что построение y_{ij}, s_{ij} требует $2(M_2(Q) + 1)$ вопросов о f.

Г.3. Далее, положим

$$\theta_i^2 = \sum_{j=1}^{M_i} \mu_j \frac{1}{s_{ij}}, \quad \theta_i > 0$$

(будет показано, что $s_{ij} > 0$, так что определение имеет смысл). Положим

$$\eta_i = \sum_{j=1}^{M_i} \mu_j p_{ij}, \quad \eta_i' = \eta_i - P_i \eta_i \ \text{ и } \ \xi_i = \frac{\eta_i'}{\theta_i}.$$

После этого положим $q_{i+1} = q_i + \xi_i$. Период закончен.

Теорема. *Описанный метод решает задачу на минимакс $\bar f$, ассоциированную с любой задачей f из $\boldsymbol{H}_{x_1}$, с точностью ν при трудоемкости*

$$M_Q(\nu) = M_0(\nu) M_1(Q)[M_2(Q)(M_3(Q) + 3) + 2(M_2(Q) + 1)] \leqslant \leqslant c\sqrt{\bar Q}\ln^2 \bar Q \left]\ln\frac{1}{\nu}\right[.$$

Здесь c — абсолютная константа, $\nu < 1$, $\bar Q = \max\{Q, 2\}$.

Доказательство. Оценка трудоемкости ясна из построения метода. Обоснование утверждения о точности дано в § 6.

4.4. Сделаем несколько замечаний о практической реализации предлагаемого метода. Из его описания ясно, что сколько-нибудь нетривиальными с вычислительной точки зрения являются две операции:

— вычисление при данном x точки Tx, т. е. решение минимаксной задачи, порожденной квадратичными функциями с простейшей квадратичной частью (пропорциональной скалярному квадрату);

— реализация правила Г, т. е. выбор выпуклой комбинации данных векторов $\xi_1, \ldots, \xi_s \in R^2$ минимальной нормы.

Первая операция по существу та же, что и шаг градиентного метода из § 1, и по поводу ее реализации можно повторить соответствующее замечание из п. 1.4. Операция сравнительно несложна лишь при $m = 0$, а также при $G = E$ и небольших m (порядка нескольких единиц). Вторая операция гораздо более проста. Она состоит в решении квадратичной задачи оптимизации на симплексе. На самом деле ситуация еще проще. Анализ доказательства теоремы о сходимости показывает, что нет необходимости точно решать соответствующую задачу, а достаточно найти выпуклую комбинацию векторов $P_i p_{ij} = \xi_j$, $1 \leqslant j \leqslant M_i$ (речь идет о периоде i этапа со входом $\bar x$) нормы, меньшей $16L_i/2^{1/2M_3(Q)}$. При этом отсутствие такой комбинации служит сигналом о том, что уже x_{i+1} удовлетворяет условию $V(x_{i+1}) \leqslant {}^3/_4 V(\bar x)$, так что этап можно в таком случае прервать выдачей в качестве выхода точки x_{i+1}.

С другой стороны, искать выпуклую комбинацию ξ_j малой нормы можно и так: пусть K — квадрат с центром в 0 и стороной $32L_i/2^{1/2(M_3(Q))}$. Отыскание выпуклой комбинации ξ_j, лежащей в K, есть задача линейного программирования с четырьмя ограничениями на $(M_i - 1)$-мерном симплексе, так что для ее решения можно применить методы линейного программирования (лучше всего — двойственные, так как число ограничений мало — их 5 (4 неравенства и 1 равенство. Ограничения неотрицательности, разумеется, не считаются). Ясно, что решение этой задачи есть выпуклая комбинация векторов ξ_j с нужной оценкой нормы. Легко сообразить, что такая модификация метода обеспечит ему требуемую точность, если увеличить на 1 правую часть в правиле определения $M_3(Q)$ в (4.3).

Заметим еще, что при описании метода мы интересовались лишь его «порядковой оптимальностью» и не ставили себе целью максимально рационализировать описываемую конструкцию — это привело бы к сильному загромождению и так не слишком простого изложения. При желании читатель сможет придумать не один способ модификации описанной конструкции, повышающей ее практическую эффективность.

§ 5. Решение условных сильно выпуклых и гладких задач

5.1. Применим метод § 4 решения задач на минимакс к решению задач классов $\boldsymbol{H}_{x_1}^V$. Получим метод с оценкой трудоемкости $O\left(\sqrt{\bar Q}\ln^2 \bar Q \ln\frac{1}{\nu}\right)$, т. е. улучшаемый по трудоемкости (по крайней мере, при больших n и малых ν, см. теорему 2.6) разве что в $O(\ln^2\bar{Q})$ раз. К сожалению, предлагаемая далее редукция задач классов $\boldsymbol{H}_{x_1}^V$ к задачам на минимакс не проходит для классов $\boldsymbol{H}_{x_1}$ (именно, эта причина вынудила нас вводить классы $\boldsymbol{H}_{x_1}^V$). Авторам неизвестен способ распространения результатов данного параграфа на классы $\boldsymbol{H}_{x_1}$ при $m > 0$, хотя такой способ, по всей вероятности, существует.

5.2. Рассмотрим класс задач $\boldsymbol{H}_{x_1}^V \equiv \boldsymbol{H}_{x_1}^V(G, m;\ l_0, \ldots, l_m;\ Q_0, \ldots, Q_m)$.

Пусть $Q = \max\limits_{0 \leqslant i \leqslant m} Q_i$ — его модуль сильной выпуклости, $\bar{Q} = \max\{Q, 2\}$. Опишем метод решения задач класса $\boldsymbol{H}_{x_1}^V$ точности $\nu < 1$. Метод устроен следующим образом.

1°. Задается вопрос о задаче f в точке x_1. Пусть

$$g = \left(g_0, \frac{3}{\nu} f_1, \ldots, \frac{3}{\nu} f_m\right),$$

где $g_0(x) = f_0(x) - f_0(x_1) + V$. Очевидно,

$$g \in \boldsymbol{H}_{x_1}\left(G, m; l_0, \frac{3l_1}{\nu}, \ldots, \frac{3l_m}{\nu};\ Q_0, \ldots, Q_m\right) \equiv \boldsymbol{H}_{x_1}.$$

2°. К задаче на минимакс $\bar{g}$, ассоциированной с задачей $g \in \boldsymbol{H}_{x_1}$, применяется метод решения задач на минимакс из § 4, отвечающий классу $\boldsymbol{H}_{x_1}$ и настроенный на точность $^1/_3\nu^2$. Пусть

$$M_Q(\nu) \leqslant c \ln^2\bar{Q}\, \sqrt{\bar{Q}}\ \left]\ln\frac{1}{\nu}\right[$$

— число шагов этого метода (c — абсолютная константа), а $\bar{x}$ — результат его применения к g. Заметим, что локальная информация о g очевидным образом находится по локальной информации о f, так что этап 2° «информационно обеспечен» возможностью задавать вопросы о f.

3°. Задается вопрос о f в точке $\bar{x}$. Если

$$\max_{1 \leqslant j \leqslant m} f_j(\bar{x}) \leqslant \nu V,$$

то точка $\bar{x}$ выдается в качестве решения задачи. В противном случае в качестве результата выдается $*$.

Предложение. Описанный метод решает все задачи из $\boldsymbol{H}_{x_1}^V$ с точностью ν при трудоемкости,

$$\leqslant \bar{c}\sqrt{\bar{Q}}\ln^2\bar{Q}\ \left]\ln\frac{1}{\nu}\right[,$$

где $\bar{c}$ — абсолютная константа.

Доказательство. Оценка трудоемкости очевидна. Оценим погрешность метода на задаче $f \in \boldsymbol{H}_{x_1}^V$. Пусть $\bar{x}$ — результат применения метода к f. Возможно, что f несовместна. Тогда уже правило 3° гарантирует неравенство

$$\nu(\bar{x}, f) \leqslant \nu. \qquad (5.1)$$

Пусть теперь f совместна и x^* — ее решение. Тогда

$$|f_0(x_1) - f_0(x^*)| \leqslant V,\ f_j(x_1) \leqslant V, \quad 1 \leqslant j \leqslant m.$$

Стало быть,

$$0 \leqslant g_0(x^*) \leqslant 2V, \quad g_j(x_1) \leqslant \frac{3}{\nu} V, \quad 1 \leqslant j \leqslant m,$$

тогда как $g_j(x^*) \leqslant 0$, $1 \leqslant j \leqslant m$. Следовательно, $\bar{g}(x^*) \leqslant 2V$ и $\bar{g}(x_1) \leqslant \frac{3}{\nu} V$. По свойствам этапа 2° имеем

$$\bar{g}(\bar{x}) \leqslant \left(1 - \frac{\nu^2}{3}\right)\min_{y \in G} \bar{g}(y) + \frac{\nu^2}{3}\bar{g}(x_1) \leqslant$$
$$\leqslant \left(1 - \frac{\nu^2}{3}\right)\bar{g}(x^*) + \frac{\nu^2}{3}\bar{g}(x_1) \leqslant \left(1 - \frac{\nu^2}{3}\right) g_0(x^*) + \nu V. \qquad (5.2)$$

Отсюда

$$g_j(\bar{x}) \leqslant \left(1 - \frac{\nu^2}{3}\right) g_0(x^*) + \nu V \leqslant 2\left(1 - \frac{\nu^2}{3}\right) V + \nu V \leqslant 3V,$$
$$1 \leqslant j \leqslant m,$$

т. е.

$$f_j(\bar{x}) \leqslant \frac{\nu}{3}\cdot 3V = \nu V, \quad 1 \leqslant j \leqslant m.$$

Следовательно, $\bar{x} = \bar{x}$ и

$$f_j(\bar{x}) \leqslant \nu V, \quad 1 \leqslant j \leqslant m. \qquad (5.3)$$

С другой стороны, (5.2) при $j = 0$, с учетом $g_0(x^*) \geqslant 0$, дает

$$g_0(\bar{x}) \leqslant \bar{g}(\bar{x}) \leqslant g_0(x^*) + \nu V,$$

т. е. $f_0(\bar{x}) \leqslant f_* + \nu V$, что вместе с (5.3) доказывает (5.1). Предложение доказано.

5.3. Построенный метод решения задач классов $\boldsymbol{H}_{x_1}^V$ сразу же применяется и к решению гладких выпуклых задач — задач классов $\boldsymbol{CS}_{x_1}^V(G, m)$. Действительно, в п. 1.3 мы научились сопоставлять методу $\mathscr{B}_\nu$ решения задач класса

$$\boldsymbol{H}_{x_1}^{\left(1+\frac{\nu}{2}\right)V}\left(G, m; \frac{\nu V}{\rho_{G,x_1}^2}, \ldots, \frac{\nu V}{\rho_{G,x_1}^2};\ \frac{2+\nu}{\nu}, \ldots, \frac{2+\nu}{\nu}\right)$$

точности $\nu/3$ и трудоемкости M метод $\mathscr{B}_{\nu'}$ решения задач $\boldsymbol{CS}_{x_1}^V(G, m)$ точности ν при трудоемкости M. Класс $\boldsymbol{H}_{x_1}^{\left(1+\frac{\nu}{2}\right)V}$ имеет модуль сильной выпуклости $(2+\nu)/\nu$. Стало быть, выбирая в качестве $\mathscr{B}_\nu$ метод из предыдущего пункта, мы сконструируем метод решения задач $CS^V_{x_1}(G, m)$ точности $\nu < 1$ и трудоемкости

$$M(\nu) \leqslant \frac{c}{\sqrt{\nu}} \Big] \ln^3 \frac{1}{\nu} \Big[,$$

c — абсолютная константа.

5.4. Оценим возможности потенциального улучшения методов §§ 3—5. В качестве «конкурирующих» будут рассматриваться любые детерминированные методы решения задач соответствующих классов, использующие детерминированные локальные оракулы. Мы увидим, что в большинстве естественных ситуаций они существенно не улучшаемы. Далее рассматриваемая точность ν считается $< 1/2$, c_i означают абсолютные константы, n означает $\dim E$. Начнем со строго выпуклых классов.

1°. $m = 0$, $G = E$, речь идет о решении задач классов $H^V_{x_1}(E, 0; l_0; Q_0)$, $H_{x_1}(H, 0; l_0; Q_0)$. Предполагается $Q = Q_0 \geqslant 2$. Для решения задач обоих классов можно применить метод § 3. Оценка трудоемкости будет

$$M_1(\nu) \leqslant c_1 \sqrt{Q} \ln Q \ln \frac{1}{\nu} .$$

В силу результатов § 2 трудоемкость метода при условии $n \geqslant \sqrt{Q} \ln 1/\nu$ в принципе снижается не более чем в $c_2 \ln Q$ раз, а при условии $n \geqslant \sqrt{Q}$ — не более чем в $c_2 \ln^2 Q$ раз.

2°. $m = 0$, G — любое допустимое принятыми соглашениями. Для решения задач классов $H_{x_1}(G, 0; l_0; Q_0)$ и $H^V_{x_1}(G, 0; l_0; Q_0)$ можно применить метод решения задач на минимакс из § 4. Его оценка трудоемкости будет (при $Q \geqslant 2$, что далее предполагается)

$$M_2(\nu) \leqslant c_3 \sqrt{Q} \ln^2 Q \ln \frac{1}{\nu} .$$

В силу результатов § 2 при условии $n \geqslant \sqrt{Q}$ и $x_1 \in \operatorname{int} G$ его трудоемкость на классе H_{x_1} в принципе снижается не более чем в $c_4 \ln^2 Q$ раз. Аналогичное утверждение верно и для класса $H^V_{x_1}$ в асимптотике по $\nu \to 0$ (и даже для всех $\nu \leqslant 1/2$, если

$$\frac{2V}{l_0 \rho^2_{|\cdot|}(x_1, \partial G)} \leqslant 1).$$

3°. $m > 0$, G — любое допустимое принятыми соглашениями. Мы умеем «хорошо» решать лишь задачи из $H^V_{x_1}(G, m; l_0, \ldots, l_m; Q_0, \ldots, Q_m)$. Соответствующий метод, описанный в п. 5.2, имеет оценку трудоемкости (при $Q = \max\limits_{0 \leqslant i \leqslant m} Q_i \geqslant 2$, что далее предполагается)

$$M_3(\nu) \leqslant c_5 \sqrt{Q} \ln^2 Q \ln \frac{1}{\nu} .$$

В силу результатов § 2 при условии $x_1 \in \operatorname{int} G$ и $n \geqslant \sqrt{Q}$ предложенный метод улучшаем по трудоемкости в асимптотике по $\nu \to 0$ в принципе не более чем в $c_6 \ln^3 Q$ раз. При достаточно большом $\rho_{|\cdot|}(x_1, \partial G)$ (например, при $G = E$) и $Q_0 = Q$ результат справедлив для всех $\nu \leqslant 1/2$.

Замечание. Методы сильно выпуклого программирования из §§ 3—5 эффективны для классов с модулем сильной выпуклости Q, не слишком близким к 1 (в пп. 1°—3° мы предполагали $Q \geqslant 2$). *Только этот случай, разумеется, и представляет интерес.* Чисто формальное замечание состоит в том, что при Q, близких к 1, в соответствующих предположениях относительно $\dim E$ «пальма первенства» переходит к градиентному методу из § 1 (ср. точные оценки его трудоемкости при малых Q с точными оценками $I_0(\nu, \infty, Q)$, $J_0(\nu, \infty)$). Вообще, при $Q \leqslant Q_0$ градиентный метод «проигрывает» методам §§ 3—5 не более чем в $\bar{c}\sqrt{Q_0}$ раз, так что его недостатки сказываются лишь в асимптотике по $Q \to \infty$.

4°. Остается выяснить, насколько эффективен предлагаемый метод решения задач классов $CS^V_{x_1}(G, m)$. Его оценка трудоемкости

$$M_4(\nu) \leqslant c_7 \frac{1}{\sqrt{\nu}} \ln^3 \frac{1}{\nu} .$$

Сопоставляя эту оценку с оценками § 3, видим, что в условиях, когда G — шар с центром в x_1 и $1/2 \geqslant \nu \geqslant 1/n^2$, предложенный метод в принципе не улучшаем более чем в $c_8 \ln^3 (1/\nu)$ раз.

Замечание. Во всех утверждениях о субоптимальности методов §§ 3—5 присутствовало требование достаточно большого значения $\dim E$. Это и понятно: при малых $\dim E$ с рассмотренными методами начинают конкурировать методы решения общих выпуклых задач типа МЦТ, оценка трудоемкости которых зависит лишь от размерности, но не от характеристик гладкости решаемых выпуклых задач. Грубо говоря, польза от гладкости (сильной выпуклости) проявляется лишь на задачах достаточно большой размерности.

5.5. В заключение рассмотрим вопрос о возможности снижения требований к априорной информации, необходимой для «запуска» построенных методов сильно выпуклого программирования. Ограничимся случаем $m = 0$. Описанные методы можно применять, если априори известно не только то, что рассматриваемая функция f сильно выпукла, но известны и ее параметры сильной выпуклости (l_f, Q_f). Конечно, не обязательно знать «самые точные» значения параметров — именно, величины l_f и $Q_f = L_f/l_f$. Достаточно иметь их оценки l ($\leqslant l_f$), L ($\geqslant L_f$). Они-то и составляют априорную информацию для «запуска» метода (наряду с ними должна быть, разумеется, указана и требуемая точность решения ν). Надо, однако, иметь в виду, что оценка параметров l_f, L_f — практически весьма сложная задача; при этом желательно решить ее по возможности точно. Действительно, «неправильная» оценка (при которой $l > l_f$ или $L < L_f$) вообще недопустима — исходя из такой неверной априорной информации, мы применим метод, который, вообще говоря, не обеспечит нужной точности. Оценка качественно правильная (с $l \leqslant l_f$ и $L \geqslant L_f$), но грубая ($Q \equiv L/l \gg L_f/l_f \equiv Q_f$), хотя и гарантирует требуемую точность, но приводит к неэкономному методу (трудоемкость за счет грубости априорной информации возрастает в $O(\sqrt{Q/Q_f})$ раз).

Таким образом, предложенные методы действительно эффективны лишь тогда, когда априорные оценки параметров сильной выпуклости минимизируемой функции достаточно близки к их истинным значениям. Между тем даже и грубая (но правильная) их оценка часто весьма затруднительна, что существенно сужает область применимости методов. В противоположность этому, стандартные методы сильно выпуклого программирования (градиентные, сопряженных градиентов и т. п.) не страдают указанным недостатком. Никакой априорной информации о параметрах сильной выпуклости минимизируемой функции для их запуска не требуется. При этом точность, достигнутая за данное время, автоматически будет тем выше, чем меньше модуль сильной выпуклости решаемой задачи (эту фразу надо понимать не буквально — на отдельных задачах может «везти», даже если их модули сильной выпуклости велики,— а в смысле оценок трудоемкости на соответствующих классах).

Таким образом, применение стандартных методов не требует никакой априорной информации. Конечно, если стоит задача обеспечения заданной точности решения, то без оценки параметров сильной выпуклости (или, по крайней мере, ее модуля) не обойтись — модуль сильной выпуклости, как мы знаем, существенно влияет на потенциально достижимое время решения. Но задача обеспечения требуемой точности обычно на практике не ставится (по-видимому, как раз из-за трудностей с априорной информацией). Наиболее типично применение «бесконечношаговых» методов. При этом задача решается в течение определенного времени, выбираемого из «потусторонних» соображений — и что получится в смысле точности — то и получится. Специально этот вопрос на практике обычно не изучают.

Вместе с тем стандартные методы далеко не всегда обеспечивают (даже по порядку) потенциально возможный при задаче с данным модулем сильной выпуклости рост точности в функции времени решения (см. примеры в главе VIII). Возникает желание так модифицировать описанный метод, чтобы перестать нуждаться

в априорной информации, сохранив в основном его эффективность. Оказывается, это легко сделать.

Сопоставим предложенному методу «бесконечношаговый» метод $\mathcal{B}_G$ сильно выпуклого программирования, обладающий следующими свойствами. Метод применим к задачам вида

$$f(x) \to \min \mid x \in G \subset E, \tag{5.4}$$

у которых f — любая сильно выпуклая функция на E. Метод $\mathcal{B}_G$, как и исходный, использует точный оракул первого порядка, причем вопросы можно задавать во всем E. Ни в какой априорной информации метод $\mathcal{B}_G$ не нуждается. При решении задачи f с *любым* модулем сильной выпуклости $Q_f = L_f/l_f$ для результата $x_n(f)$ работы метода в момент n справедливо соотношение

$$x_n(f) \in G, \ \nu(x_n(f), f), \leqslant \nu \tag{5.5}$$

при

$$n \geqslant c\sqrt{\bar{Q}_f}\ln^3 \bar{Q}_f \ln\frac{1}{\nu}\ln\left\{\sqrt{\bar{Q}_f}\ln^3 \bar{Q}_f \ln\frac{1}{\nu}\right\},$$

если только $\nu \leqslant {}^1/_2$. Здесь $c > 0$ — абсолютная константа, $\bar{Q}_f = \max\{Q_f, 4\}$. Как всегда,

$$\nu(x, f) = \frac{f(x) - \inf\limits_G f}{f(x_1) - \inf\limits_G f},$$

x_1 — раз навсегда фиксированная точка G. Заметим, что даже если бы мы *точно знали* (l_f, L_f) заранее, то мы не могли бы гарантировать точность ν в решении f за время, меньшее $\underline{c}\sqrt{Q_f}\ln(1/\nu)$, $\underline{c}$ — абсолютная константа (во всяком случае, если $Q_f \geqslant 2$, $x_1 \in \operatorname{int} G$ и $\dim E$ достаточно велика). Таким образом, при достаточно больших $\dim E$ и $x_1 \in \operatorname{int} G$ время, начиная с которого $\mathcal{B}_G$ обеспечит точность $\nu \leqslant {}^1/_2$ в решении любой сильно выпуклой задачи с нетривиальным ($\geqslant 2$) модулем сильной выпуклости, несущественно (с точностью до логарифмического множителя) больше потенциально достижимого времени решения задач с данным модулем сильной выпуклости. Заметим еще, что если $G = E$, то все $\ln^3 \bar{Q}_f$ в (5.5) можно заменить на $\ln \bar{Q}_f$.

Перейдем к конструкции $\mathcal{B}_G$. Она напоминает способ «склейки» реализующих сложность класса при данной точности конечношаговых методов в оптимальный «бесконечношаговый» метод из § 5 гл. I, однако, не сводится к этому способу. Действительно, сейчас мы решаем более сложную задачу: проводим «склейку» как по точности, так и по априорной информации.

Построим метод $\mathcal{B}_G$ в два этапа. Вначале избавимся от необходимости знать и l_f, и L_f, и ν — нам будет достаточно знать лишь Q_f и ν. Затем элиминируем и эту априорную информацию.

1°. Обозначим через $\mathcal{B}_G(l, L, \nu)$ метод решения $(l, L/l)$-сильно выпуклых задач вида (5.4) с точностью ν из § 3 (для $E = G$) или § 4 (для $E \neq G$), а через $M(L/l, \nu)$ — его трудоемкость. Фиксируем пару точек $x, y \in G$, $x \neq y$. Пусть заданы $Q \geqslant 2$ и $\nu \leqslant \leqslant {}^1/_2$. Построим сейчас метод $\mathcal{B}_G(Q, \nu)$ решения всех задач f вида (5.4) и модуля сильной выпуклости $Q_f \leqslant Q/2$ точности ν с оценкой трудоемкости $M_1(Q, \nu) \leqslant c_1 \ln QM(Q, \nu)$, c_1 — абсолютная константа. При $G = E$ можно взять $M_1(Q, \nu) = c_1 M(Q, \nu)$.

Метод $\mathcal{B}_G(Q, \nu)$ при $G \neq E$ строится следующим образом. Зададим вопросы о решаемой задаче f в точках x и y и составим число

$$\rho_f = \frac{\langle f'(x) - f'(y) \mid x - y\rangle}{|x - y|^2}.$$

Если получится $\rho_f \leqslant 0$, то откажемся от решения, и правильно сделаем — в этом случае f заведомо не является сильно выпуклой. В противном случае положим $\bar{l}_f = \rho_f/Q$, $\bar{L}_f = \rho_f Q$ и последовательно применим к f методы

$$\mathcal{B}(\bar{l}_f, Q\bar{l}_f, \nu), \mathcal{B}(2\bar{l}_f, 2Q\bar{l}_f, \nu), \ldots, \mathcal{B}(2^i\bar{l}_f, 2^iQ\bar{l}_f, \nu), \quad 0 \leqslant i \leqslant]\log_2 Q[.$$

Из результатов их работы выберем лучший по значению f. Он и будет результатом применения $\mathcal{B}_G(Q, \nu)$ к f. Очевидно, что трудоемкость указанного метода допускает нужную оценку. Покажем, что он решает любую задачу f с $Q_f \leqslant Q/2$ с точностью ν. Действительно, по определению l_f и L_f имеем $\rho_f \geqslant l_f \geqslant \bar{l}_f$. Стало быть, при некотором $i \leqslant]\log_2 Q[$ будем иметь $2^i\bar{l}_f \leqslant l_f$, но $2^{i+1}\bar{l}_f \geqslant l_f$. Отсюда

$$2^iQ\bar{l}_f \geqslant \frac{1}{2} l_f Q \geqslant l_f Q_f = L_f,$$

так что f будет $(2^i\bar{l}_f, Q)$-сильно выпукла. Поэтому уже результат i-го из составляющих $\mathcal{B}_G(Q, \nu)$ методов имеет погрешность $\leqslant \nu$, так что это же верно и для самого $\mathcal{B}_G(Q, \nu)$.

В случае $E = G$ дело обстоит совсем просто. Действительно, в методе § 3 знание оценок самих параметров сильной выпуклости l_f, L_f минимизируемой функции f, а не только верхней оценки Q их отношения Q_f требовалось всего в двух местах — в выборе радиуса круга, в котором проводилась «двумерная» минимизация f, и в выборе шага при переходе от $\hat{x}_i$ к x_i по направлению антиградиента f в точке $\hat{x}_i$ (см. п. 3.2). В первом случае фактически достаточно знать ρ_f и Q. После этого радиус круга можно взять равным $r = 2Q\rho_f^{-1}|f'(\bar{x}_i)|$. Во втором случае можно заменить один шаг по направлению антиградиента минимизацией f на отрезке $[\hat{x}_i, \hat{x}_i - 2Q\rho_f^{-1}|f'(\hat{x}_i)|]$, проводимой делением пополам и настроенной на относительную точность порядка $1/Q^4$. Читателю предоставляется дать точные формулы для относительных точностей одномерной и двумерной минимизации в этом методе и оценить число его шагов. При этом он сможет убедиться, что они по порядку зависимости от Q и ν те же, что и у исходной версии метода.

2°. Теперь «склеим» методы $\mathcal{B}_G(Q, \nu)$ по Q, ν и построим метод $\mathcal{B}_G$. Метод $\mathcal{B}_G$ состоит из последовательных этапов, причем n-й из них $(n = 1, 2, \ldots)$ состоит из n процедур. k-я процедура n-го этапа $\Gamma_{n,k}$, $1 \leqslant k \leqslant n$, состоит в применении к решаемой задаче метода $\mathcal{B}_G(Q_k, \nu_{n-k})$ с $\ln 1/\nu_{n-k} = 2^{n-k}$ и Q_k, определяемым из уравнения $\sqrt{Q_k}\ln^3 Q_k = 2^k$ при $G \neq E$ или $\sqrt{Q_k}\ln Q_k = 2^k$ при $G = E$. Результат, формируемый методом $\mathcal{B}_G$ в момент t, есть лучший (по значению f) из результатов всех завершенных к этому моменту процедур. На начальном отрезке траектории (покуда не завершена $\Gamma_{1,1}$) результат есть x_1.

Убедимся, что описанный метод обеспечивает любую наперед заданную точность $\nu \leqslant {}^1/_2$ на любой сильно выпуклой функции f за обещанное выше время. Действительно, выберем наименьшее целое k, для которого $Q_k \geqslant 2Q_f$, и наименьшее целое m, для которого $\ln 1/\nu \leqslant 2^m$. Положим $n = m + k$ и заметим, что процедура $\Gamma_{n,k}$ дает в применении к f результат погрешности, $\leqslant \nu$ (так определялся метод $\mathcal{B}_G(Q, \nu)$). Стало быть, после момента окончания n-го этапа погрешность $\mathcal{B}_G$ на f не превосходит ν. Остается подсчитать этот момент t_n. Для определенности рассмотрим случай $G = E$ (общий случай рассматривается точно так же). В силу оценок трудоемкости $\mathcal{B}_G(Q, \nu)$ процедура $\Gamma_{q,p}$ занимает время

$$\leqslant c_2\sqrt{Q_p}\left(\ln\frac{1}{\nu_{q-p}}\right)\ln Q_p,$$

ν_{q-p} — точность, на которую настроена $\Gamma_{q,p}$, c_2 — абсолютная константа. По определению Q_p и ν_{q-p} это время не больше $c_2 2^q$. Таким образом, длительность q-го этапа не выше $c_2 q 2^q$, а момент окончания q-го этапа не превосходит $c_2\sum_{j=1}^{q} j2^j \leqslant c_3 q\cdot 2^q$, c_3 — абсолютная константа. Итак, $t_n \leqslant c_3 n 2^n$. По выбору n имеем $n = m + k$. Здесь k есть минимальное целое, для которого корень Q_k уравнения $\sqrt{Q}\ln Q = 2^k$ не меньше $2Q_f$. Легко видеть, что тогда

$$2^k \leqslant c_4(\sqrt{\overline{Q}_f}\ln\overline{Q}_f), \quad k \leqslant c_5\ln(\sqrt{\overline{Q}_f}\ln\overline{Q}_f),$$

c_4, c_5 — абсолютные константы.

Аналогично из определения m

$$2^m \leqslant c_6\ln\frac{1}{\nu}, \quad m \leqslant c_5\ln\left(2\ln\frac{1}{\nu}\right).$$

Стало быть,

$$t_n \leqslant c_3c_4c_5c_6\left(\sqrt{\overline{Q}_f}\ln\overline{Q}_f\ln\frac{1}{\nu}\right)\ln\left(\sqrt{\overline{Q}_f}\ln\overline{Q}_f\ln\frac{1}{\nu}\right),$$

что и требуется.

В заключение заметим, что нас интересовали лишь «порядковые» эффекты, так что, стремясь к ясности, мы не занимались рационализацией описанной конструкции. Имеется много путей такого рода рационализации — например, оптимизации «шкалы» изменения Q_k и ν_k. Скажем, можно было бы заменить 2^k, 2^{n-k} в уравнениях, определяющих Q_k и ν_{n-k}, на a^k, a^{n-k}, $a > 1$, и подбирать a с тем, чтобы улучшить оценки. Можно выбирать лучший предыдущий результат в качестве начальной точки очередной процедуры и применить простые правила опровержения гипотезы о том, что рассматриваемая функция (l, Q)-сильно выпукла (это позволило бы сократить время «лишних» процедур в $\mathcal{B}_G(Q, \nu)$) и т. п. Читатель, желающий практически применить описанную конструкцию, наверняка придумает множество ухищрений подобного рода.

§ 6. Доказательство теоремы 4.3

В помещенном ниже доказательстве без специальных пояснений используются и утверждения из соответствующих разделов § 4.

6.1. Пусть $\Delta = \{\lambda \in (\lambda_0, \ldots, \lambda_m) \mid \lambda_i \geqslant 0, \sum_i \lambda_i = 1\}$ и

$$f_{\lambda,x}(y) = \sum \lambda_i f_{i,x}(y).$$

Очевидно, в силу леммы фон Неймана

$$\min_{y\in G} f_x(y) = \min_{y\in G}\max_{\lambda\in\Delta} f_{\lambda,x}(y) = \max_{\lambda\in\Delta}\min_{y\in G} f_{\lambda,x}(y).$$

Следовательно, для каждого $x \in E$ определено непустое множество $\Lambda(x) \in \Delta$ такое, что при $\lambda \in \Lambda(x)$ имеем

$$f_x(Tx) = f_{\lambda,x}(Tx) = \min_{y\in G} f_{\lambda,x}(y)$$

(мы учли, что $f_{\lambda,x}(y) \leqslant f_x(y)$ при всех y). Положим

$$s(x) = \min_{y\in G} f_x(y).$$

Обозначим еще

$$\rho_{x,\lambda} = \sum_{i=0}^{m} \lambda_i l_i Q.$$

Фиксируем $f \in \boldsymbol{H}_{x_1}$, и пусть x^* — точка минимума $\bar f$ на G, $\bar f_* = \bar f(x^*)$. Далее $V_{\bar f}(\cdot)$ обозначается просто $V(x) \equiv \bar f(x) - \bar f_*$.

6.2. Лемма. *Пусть $x \in G$, $y \in G$. Тогда*

1) $V(Tx) \leqslant V(x) - (f_{\lambda,x}(x) - f_{\lambda,x}(Tx)) \leqslant V(x)$

при всех $\lambda \in \Lambda(x)$;

2) если $V(Tx) \geqslant {}^1/_2 V(x)$, *то при* $\lambda \in \Lambda(x)$

$$V(x) \leqslant 2Q(f_{\lambda,x}(x) - f_{\lambda,x}(Tx));$$

3) если $V(Tx) \geqslant {}^1/_2 V(x)$ *и* $V(y) \leqslant 2V(x)$, *то*

$$|x - y| \leqslant (\sqrt{5} + 1)Q\,|p(x)|.$$

Доказательство. Соотношение 1) очевидно вследствие цепочки неравенств

$$\bar f(x) = f_x(x) \geqslant f_{\lambda,x}(x) \geqslant f_{\lambda,x}(Tx) = f_x(Tx) \geqslant \bar f(Tx).$$

2) Пусть $u \in G$ и $h = \frac{1}{Q}(u - x) + x$. Имеем в силу (1.1)

$$f_i(x) + \langle f_i'(x) \mid u - x\rangle + \frac{l_i}{2}(u - x)^2 \leqslant f_i(u),$$

или

$$Q\left[\langle f_i'(x) \mid h - x\rangle + \frac{l_i}{2}Q(h - x)^2\right] \leqslant f_i(u) - f_i(x).$$

Следовательно,

$$f_{i,x}(h) \leqslant \frac{1}{Q}f_i(u) + \left(1 - \frac{1}{Q}\right)f_{i,x}(x).$$

Тогда при $\lambda \in \Lambda(x)$

$$f_{\lambda,x}(h) \leqslant \frac{1}{Q}\bar f(u) + \left(1 - \frac{1}{Q}\right)f_{\lambda,x}(x). \tag{6.1}$$

Обозначим $z = h - Tx$. Тогда

$$f_{\lambda,x}(h) = f_{\lambda,x}(Tx + z) = f_{\lambda,x}(Tx) + \langle q_{x,\lambda} \mid z\rangle + \frac{1}{2}\rho_{x,\lambda}z^2,$$

где $q_{x,\lambda} = (f_{\lambda,x})'(Tx)$ (производная $f_{\lambda,x}(y)$ по y при $y = Tx$). Точно так же

$$f_{\lambda,x}(x) = f_{\lambda,x}(Tx) + \langle q_{x,\lambda} \mid x - Tx\rangle + \frac{1}{2}\rho_{x,\lambda}(x - Tx)^2 =$$

$$= f_{\lambda,x}(Tx) + \langle q_{x,\lambda} \mid p(x)\rangle + \frac{1}{2}\rho_{x,\lambda}p^2(x).$$

Подставляя эти соотношения в (6.1), получим

$$\frac{1}{2}\rho_{x,\lambda}z^2 \leqslant \frac{1}{Q}\bar f(u) + \left(1 - \frac{1}{Q}\right)f_{\lambda,x}(Tx) +$$

$$+ \left(1 - \frac{1}{Q}\right)\langle q_{x,\lambda} \mid x - Tx\rangle + \frac{1}{2}\rho_{x,\lambda}p^2(x)\left(1 - \frac{1}{Q}\right) -$$

$$- \langle q_{x,\lambda} \mid z\rangle - f_{\lambda,x}(Tx) = \frac{1}{Q}(\bar f(u) - f_{\lambda,x}(Tx)) +$$

$$+\langle q_{x,\lambda}\,|\,x-Tx-h+Tx\rangle-\frac{1}{Q}\langle q_{x,\lambda}\,|\,x-Tx\rangle+$$
$$+\frac{1}{2}\rho_{x,\lambda}p^2(x)\left(1-\frac{1}{Q}\right)=\frac{1}{Q}(\bar f(u)-f_{\lambda,x}(Tx))-$$
$$-\frac{1}{Q}\langle q_{x,\lambda}\,|\,u-Tx\rangle+\frac{1}{2}\rho_{x,\lambda}p^2(x)\left(1-\frac{1}{Q}\right). \qquad (6.2)$$

Далее $\langle q_{x,\lambda}\mid u-Tx\rangle\geqslant 0$ при $u\in G$, ибо Tx — точка минимума $f_{\lambda,x}(y)$ на G. Следовательно,

$$\frac{1}{2}\rho_{x,\lambda}z^2\leqslant\frac{1}{Q}(\bar f(u)-f_{\lambda,x}(Tx))+\frac{1}{2}\rho_{x,\lambda}p^2(x)\left(1-\frac{1}{Q}\right). \quad (6.3)$$

Положим здесь $u=x^*$. Тогда из (6.3) получаем

$$(f_{\lambda,x}(Tx)-\bar f(x^*))\leqslant\frac{1}{2}Q\rho_{x,\lambda}\left(p^2(x)\left(1-\frac{1}{Q}\right)\right).$$

Стало быть,

$$\frac{1}{2}V(x)\leqslant V(Tx)\leqslant f_x(Tx)-\bar f(x^*)=f_{\lambda,x}(Tx)-\bar f(x^*)\leqslant$$
$$\leqslant\frac{1}{2}Q\rho_{x,\lambda}p^2(x)\left(1-\frac{1}{Q}\right)\leqslant\frac{1}{2}Q\rho_{x,\lambda}p^2(x). \qquad (6.4)$$

С другой стороны, в силу 1) имеем

$$V(x)\geqslant f_{\lambda,x}(x)-f_{\lambda,x}(Tx)=\frac{\rho_{x,\lambda}}{2}p^2(x)+\langle q_{x,\lambda}\,|\,p(x)\rangle\geqslant\frac{\rho_{x,\lambda}}{2}p^2(x),$$

так что (6.4) дает

$$\frac{1}{2}V(x)\leqslant Q\,(f_{\lambda,x}(x)-f_{\lambda,x}(Tx)),$$

что и требуется в п. 2).

Докажем п. 3). Для этого рассмотрим (6.3) при $u=y$, т. е.

$$z=p(x)+\frac{1}{Q}(y-x).$$

Получим

$$\frac{1}{2}\rho_{x,\lambda}z^2\leqslant\frac{1}{Q}(\bar f(y)-f_{\lambda,x}(Tx))+\frac{1}{2}\rho_{x,\lambda}p^2(x)\leqslant$$
$$\leqslant\frac{1}{Q}(\bar f(y)-\bar f_*)+\frac{1}{2}\rho_{x,\lambda}p^2(x)\leqslant\frac{2}{Q}V(x)+\frac{1}{2}\rho_{x,\lambda}p^2(x)\leqslant$$
$$\leqslant 2\rho_{x,\lambda}p^2(x)+\frac{1}{2}\rho_{x,\lambda}p^2(x)$$

(мы учли (6.4)). Итак, $z^2\leqslant 5p^2(x)$, или

$$\left|p(x)+\frac{1}{Q}(y-x)\right|\leqslant\sqrt{5}\,|p(x)|.$$

Отсюда

$$\left|\frac{1}{Q}(y-x)\right|\leqslant(\sqrt{5}+1)\,|p(x)|\quad\text{и}\quad|y-x|\leqslant Q(\sqrt{5}+1)|p(x)|,$$

что и требуется.

Лемма доказана.

6.3. Лемма. *Пусть $x\in G$ и $y\in E$ таковы, что $\langle p(y)\mid x-y\rangle\geqslant 0$. Тогда при всех $\lambda\in\Lambda(y)$ $\bar f(Ty)+\frac{1}{2}\rho_{y,\lambda}p^2(y)\leqslant\bar f(x)$.*

Доказательство. Имеем, полагая $q_{y,\lambda}$ равным производной $f_{\lambda,y}(z)$ по z в точке $z=Ty$:

$$\langle q_{y,\lambda}\mid u-Ty\rangle\geqslant 0 \text{ при } u\in G$$

и

$$q_{y,\lambda}=\Sigma\lambda_i f_i'(y)-\rho_{y,\lambda}\,p(y).$$

Итак,

$$\rho_{y,\lambda}\,p(y)=\Sigma\lambda_i f_i'(y)-q_{y,\lambda},$$

так что

$$\langle\Sigma\lambda_i f_i'(y)-q_{y,\lambda}\mid x-y\rangle\geqslant 0,$$

т. е.

$$\langle\Sigma\lambda_i f_i'(y)-q_{y,\lambda}\mid x-Ty\rangle+\langle\Sigma\lambda_i f_i'(y)-q_{y,\lambda}\mid Ty-y\rangle\geqslant 0.$$

Но

$$\Sigma\lambda_i f_i'(y)-q_{y,\lambda}=\rho_{y,\lambda}\,p(y),$$

т. е.

$$\langle\Sigma\lambda_i f_i'(y)-q_{y,\lambda}\mid x-Ty\rangle\geqslant\rho_{y,\lambda}\,p^2(y),$$

или

$$\langle\Sigma\lambda_i f_i'(y)\mid x-Ty\rangle\geqslant\langle\Sigma\lambda_i f_i'(y)-q_{y,\lambda}\mid x-Ty\rangle\geqslant\rho_{y,\lambda}p^2(y).$$

Следовательно,

$$\langle\Sigma\lambda_i f_i'(y)\mid x-y\rangle+\langle\Sigma\lambda_i f_i'(y)\mid p(y)\rangle\geqslant\rho_{y,\lambda}\,p^2(y).$$

В силу выпуклости f_i отсюда

$$\Sigma\lambda_i f_i(x)-\Sigma\lambda_i f_i(y)+\langle\Sigma\lambda_i f_i'(y)\,|\,p(y)\rangle\geqslant\rho_{y,\lambda}\,p^2(y),$$

или

$$\sum\lambda_i f_i(x)\geqslant\sum\lambda_i f_i(y)+\left\langle\sum\lambda_i f_i'(y)\,\middle|\,Ty-y\right\rangle+$$
$$+\frac{\rho_{y,\lambda}}{2}(Ty-y)^2+\frac{\rho_{y,\lambda}}{2}(Ty-y)^2,$$

т. е.

$$\bar f(x)\geqslant f_{\lambda,y}(Ty)+\frac{1}{2}\rho_{y,\lambda}(Ty-y)^2\geqslant\bar f(Ty)+\frac{1}{2}\rho_{y,\lambda}p^2(y),$$

что и требуется. Лемма доказана.

6.4. Лемма. *При всех $x,\ y\in E$ справедливо неравенство*

$$|p(y)-p(x)|\leqslant 5|x-y|+\sqrt{24|x-y|^2+4|x-y|\,|p(x)|}. \quad (6.5)$$

В частности, $p(y)$ непрерывно по y.

Доказательство. Пусть вначале $s(y) \leqslant s(x)$. Имеем

$$|f_{i,x}(z) - f_{i,y}(z)| \leqslant |f_i(x) - f_i(y) - \langle f_i'(y)|x-y\rangle| + \frac{l_iQ}{2}|(x-z)^2 - (y-z^2)| + |\langle f_i'(x) - f_i'(y)|z-x\rangle| \leqslant$$
$$\leqslant \frac{l_iQ}{2}(x-y)^2 + \frac{l_iQ}{2}(2|x-y||z-x| + |x-y|^2) + l_iQ|x-y||z-x| = l_iQ|x-y|^2 + 2l_iQ|x-y||z-x|.$$

Отсюда при $\lambda \in \Lambda(x)$

$$\Sigma\lambda_i\{f_{i,y}(z) - f_{i,x}(z) + l_iQ|x-y|^2 + 2l_iQ|x-y||z-x|\} \geqslant 0. \quad (6.6)$$

Положим $z = Ty \equiv Tx + r$. Тогда из (6.6)

$$s(y) \geqslant \Sigma\lambda_i f_{i,y}(Ty) \geqslant \Sigma\lambda_i f_{i,x}(Tx+r) - \Sigma\lambda_i l_iQ|x-y|^2 - 2\Sigma\lambda_i l_iQ|x-y||Tx+r-x|.$$

Но при $Tx + r \in G$

$$\sum \lambda_i f_{i,x}(Tx+r) \geqslant s(x) + \sum \lambda_i \frac{l_iQ}{2} r^2,$$

так что

$$\sum \lambda_i f_{i,y}(Ty) \geqslant \sum \lambda_i \left\{\frac{l_iQ}{2}r^2 + s(x) - l_iQ|x-y|^2 - 2l_iQ|x-y||Tx+r-x|\right\}.$$

Отсюда сразу следует, что при некотором i

$$s(y) \geqslant f_{i,y}(Ty) \geqslant \frac{l_iQ}{2}r^2 + s(x) - l_iQ|x-y|^2 - 2l_iQ|x-y||Tx+r-x|.$$

Так как по условию $s(y) \leqslant s(x)$, то

$$\frac{1}{2}r^2 - |x-y|^2 - 2|x-y|(|p(x)| + |r|) \leqslant 0,$$

т. е. для $|r| = t$

$$\frac{1}{2}t^2 - 2|x-y|t - 2|x-y|(|p(x)| + |x-y|) \leqslant 0,$$

откуда

$$t \leqslant 2|x-y| + \sqrt{4|x-y|^2 + 4|x-y|^2 + 4|x-y||p(x)|}.$$

Но

$$|p(y) - p(x)| \leqslant |r| + |x-y|,$$

так что

$$|p(y) - p(x)| \leqslant 3|x-y| + \sqrt{8|x-y|^2 + 4|x-y||p(x)|}. \quad (6.7)$$

Итак, при выполнении условия $s(y) \leqslant s(x)$ верно (6.7). Пусть теперь $s(y) > s(x)$. Тогда (6.7) дает

$$|p(y) - p(x)| \leqslant 3|x-y| + \sqrt{8|x-y|^2 + 4|x-y||p(y)|}.$$

Возможно, что

$$|p(y) - p(x)| \equiv \Delta \leqslant 3|x-y|,$$

тогда (6.5) верно автоматически. Пусть это не так. Положим $|x-y| = s$. Тогда

$$0 \leqslant \Delta - 3s \leqslant \sqrt{8s^2 + 4s(|p(x)| + \Delta)},$$

откуда

$$\Delta^2 - 6s\Delta + 9s^2 \leqslant 8s^2 + 4s|p(x)| + \Delta\cdot 4s$$

или

$$\Delta^2 - 10s\Delta + s^2 - 4s|p(x)| \leqslant 0.$$

Отсюда

$$\Delta \leqslant 5s + \sqrt{25s^2 - s^2 + 4s|p(x)|},$$

т. е.

$$|p(y) - p(x)| \leqslant 5|x-y| + \sqrt{24|x-y|^2 + 4|x-y||p(x)|},$$

что и требуется.

6.5. Теперь мы в состоянии провести должный анализ работы метода. Докажем (и этого, разумеется, достаточно), что на этапе невязка уменьшается не меньше чем на $^1/_4$. Более точно, пусть результат применения этапа со входом $\bar{x}$ к задаче f есть $\bar{y} \in G$. Тогда

$$\vee(\bar{y}) \leqslant {}^3/_4 \vee(\bar{x}). \quad (6.8)$$

Для доказательства предположим, что (6.8) не выполняется, и приведем это допущение к противоречию. Итак, далее примем, что

$$\vee(\bar{y}) > {}^3/_4 \vee(\bar{x}). \quad (6.9)$$

6.6. Возможно, что данный этап был прерван по правилу A (т. е. при некотором i было $L_i = 0$, или $|p(x_i)| = 0$). Это значит, что $\bar{y} = x_i$ (правило A). С другой стороны, $p(x_i) = 0$ означает в силу леммы п. 6.3, в которой надо положить $y = x_i$ и в качестве x взять произвольную точку G, что $Tx_i = x_i$ есть точка минимума $\bar{f}$ на G. Но тогда (6.9) неверно. Аналогично анализируется прерывание по правилу Г.1.

6.7. Итак, $L_i > 0$ для всех i, $p_{ij} \neq 0$ для всех i, j, для которых p_{ij} определено. Возможно, что i-й период не был прерван по правилу В.2.1. Тогда x_{i+1} есть лучшая по значению $\bar{f}(\cdot)$ среди точек TTx_i, TTx_{ij}, $1 \leqslant j \leqslant M_i$, причем

$$\langle p(x_{ij}) \mid x_i - x_{ij}\rangle \geqslant 0. \tag{6.10}$$

Действительно, (6.10) имеет место ввиду п.В.2.3, ибо

$$\langle p(x_{ij}) \mid x_i - x_{ij}\rangle = -t_{ij}\varphi_{ij}(t_{ij}).$$

Но тогда по лемме п. 6.3 имеем

$$\bar f(Tx_{ij}) \leqslant \bar f(x_i),$$

а по лемме п. 6.2

$$\bar f(T(Tx_{ij})) \leqslant \bar f(x_i).$$

Стало быть, $\bar f(x_{i+1}) \leqslant \bar f(\bar x_i)$.

6.8. Пусть теперь этап был прерван по правилу В.2.1, и это случилось на i-м периоде. Тогда $\bar y = Tx_i$ и $\bar f(x_i) \leqslant f(x_1) = f(\bar x)$ (последнее — ввиду п. 6.7). В наших условиях при некотором j $\varphi_{ij}(-L_i)\varphi_{ij}(L_i) > 0$, т. е. при надлежащем выборе $\bar t = \pm L_i$ имеем

$$\langle p(x_i + \bar t e_i^{j-1}) \mid e_i^{j-1}\rangle\, \bar t < 0.$$

Положим $y = x_i + \bar t e_i^{j-1}$. Тогда $\langle p(y) \mid x_i - y\rangle > 0$, откуда по лемме п. 6.3 $\bar f(Ty) \leqslant \bar f(x_i)$, т. е. $\mathsf{V}(Ty) \leqslant \mathsf{V}(x_i)$.

С другой стороны, $\bar y = Tx_i$ и (6.9) означает, что

$${}^3/_4\mathsf{V}(\bar x) < \mathsf{V}(\bar y) = \mathsf{V}(Tx_i) \leqslant \mathsf{V}(x_i) \leqslant \mathsf{V}(\bar x),$$

т. е.

$$\mathsf{V}(Tx_i) > {}^1/_2\, \mathsf{V}(x_i).$$

Кроме того, $\mathsf{V}(Ty) \leqslant \mathsf{V}(x_i)$. По утверждению 3) леммы п. 6.2, примененной к паре (x_i, Ty), имеем

$$|x_i - Ty| \leqslant (\sqrt{5} + 1)Q\, |p(x_i)| \leqslant L_i.$$

С другой стороны,

$$|x_i - Ty| = |(x_i - y) + p(y)| > |x_i - y| = L_i$$

(мы учли, что $\langle p(y)|x_i - y\rangle > 0$). Итак, с одной стороны, $|x_i - Ty| \leqslant L_i$, а с другой, $|x_i - Ty| > L_i$. Полученное противоречие доказывает невозможность прерывания этапа по правилу В.2.1.

6.9. Итак, на всех периодах данного этапа было $L_i > 0$, а правило В.2.1 не применялось, причем $p_{ij} \neq 0$ при всех i и j. Проанализируем i-й период. Рассмотрим его j-ю процедуру: $\varphi_{ij}(t)$ меняет знак на Δ_i^j. Следовательно, при некотором $\bar t \in \Delta_i^j$ имеем $\varphi_{ij}(\bar t) = 0$. Пусть $x_{ij}' = x_i + \bar t e_i^{j-1}$, $p_{ij}' = p(x_{ij}')$. Тогда

$$\langle p(x_{ij}')|e_i^{j-1}\rangle = 0. \tag{6.11}$$

С другой стороны,

$$|\bar t - t_{ij}| \leqslant \frac{L_i}{2^{M_2(Q)}}.$$

В силу леммы п. 6.4

$$|p(x_{ij})| \leqslant 5|x_{ij} - x_i| + \sqrt{24|x_{ij} - x_i|^2 + 4|x_{ij} - x_i||p(x_i)|} + |p(x_i)| \leqslant$$
$$\leqslant \frac{L_i}{(\sqrt{5}+1)\,\bar Q} + 5L_i + \sqrt{24L_i^2 + 4L_i\frac{1}{(\sqrt{5}+1)\,\bar Q}L_i} \leqslant 11L_i, \tag{6.12}$$

т. е.

$$|p(x_{ij})| \leqslant 11L_i.$$

Снова, применяя лемму п. 6.4, получим

$$|p(x_{ij}) - p(x_{ij}')| \leqslant$$
$$\leqslant 5|x_{ij} - x_{ij}'| + \sqrt{24|x_{ij} - x_{ij}'|^2 + 4|x_{ij} - x_{ij}'||p(x_{ij})|} \leqslant$$
$$\leqslant \frac{5L_i}{2^{M_3(Q)}} + \sqrt{24\cdot\frac{L_i^2}{2^{2M_3(Q)}} + \frac{44L_i^2}{2^{M_3(Q)}}} \leqslant \frac{8L_i}{2^{1/2 M_3(Q)}},$$

поскольку $2^{M_3(Q)} > 120$. Итак,

$$|p(x_{ij}) - p(x_{ij}')| \leqslant \frac{8L_i}{2^{1/2 M_3(Q)}},$$

что с учетом (6.11) дает

$$|\langle p(x_{ij})|e_i^{j-1}\rangle| \leqslant \frac{8L_i}{2^{\frac{1}{2}M_3(Q)}}. \tag{6.13}$$

Отсюда ясно, что $e_i^{j-1} \notin K_j$. С другой стороны, из правила В.3 ясно, что $K_j \subset K_{j-1}$. Стало быть, раствор острого угла K_j, по крайней мере, вдвое меньше раствора угла K_{j-1}, $j \geqslant 2$. Отсюда следует, что раствор угла K_j не превосходит $2\pi/2^j$. Убедимся, что $K_{M_2(Q)}$ пусто. Пусть это не так. Во всяком случае, раствор угла $K_{M_2(Q)-1}$ не превосходит $2\pi/2^{M_2(Q)-1}$. Вектор $e_i^{M_2(Q)-1} = \bar e_i$ является ортом его биссектрисы, и всякий другой единичный вектор e из $K_{M_2(Q)-1}$ удовлетворяет поэтому условию

$$|e - \bar e_i| \leqslant \frac{2\pi}{2^{M_2(Q)}}.$$

Но тогда из (6.13) при $j = M_2(Q)$ следует

$$|\langle p(x_{ij})|e\rangle| \leqslant \frac{8L_i}{2^{1/2 M_3(Q)}} + \frac{2\pi}{2^{M_2(Q)}}|p(x_{ij})| \leqslant$$
$$\leqslant \frac{8L_i}{2^{1/2 M_3(Q)}} + \frac{22\pi L_i}{2^{M_2(Q)}} \leqslant \frac{16L_i}{2^{1/4 M_3(Q)}}$$

для всякого единичного $e \in K_{M_s(Q)-1}$. Но тогда $K_{M_s(Q)} \neq \phi$, что по предположению не так.

Итак, по окончании серии процедур i-го периода непременно $K_{M_i} = \phi$ (M_i — число этих процедур). Так как $K_{M_i} = \phi$, то

$$\max_{e \in E_0^i,\ \|e\| \leqslant 1} \min_{\mu_j \geqslant 0,\ \sum_{j=1}^{M_i} \mu_j = 1} \left\langle \sum \mu_j p_{ij} \mid e \right\rangle \leqslant \frac{16L_i}{2^{1/2 M_s(Q)}}.$$

По лемме фон Неймана левая часть неравенства равна

$$\min_{\mu_j \geqslant 0,\ \sum_j \mu_j = 1} \left| P_i \sum \mu_j p_{ij} \right| = |P_i \eta_i|.$$

Итак,

$$|P_i \eta_i| \leqslant \frac{16L_i}{2^{1/2 M_6(Q)}}. \tag{6.14}$$

6.10. Докажем, что

$$\frac{1}{132(\sqrt{5}+1)Q} \leqslant \frac{|p(x_{ij})|}{|p(x_i)|} \leqslant 11(\sqrt{5}+1)\bar{Q}. \tag{6.15}$$

Действительно, правое неравенство есть (6.12). Докажем левое. Пусть $y = Tx_{ij}$. По лемме п. 6.3 и в силу (6.10) $\bar{f}(y) \leqslant \bar{f}(x_i)$. С другой стороны, по правилу выбора x_{i+1} имеем

$$^3/_4 \vee(\bar{x}) < \vee(\bar{y}) \leqslant \vee(Ty) \leqslant \vee(y) \leqslant \vee(x_i) \leqslant \vee(\bar{x}),$$

откуда $\vee(Ty) > {}^1/_2 \vee(y)$ и $2\vee(y) > \vee(x_i)$. Применяя теперь к паре (y, x_i) утверждение 3) леммы п. 6.2, найдем

$$|x_i - y| \leqslant (\sqrt{5}+1)\ Q\ |p(y)|. \tag{6.16}$$

С другой стороны, по лемме п. 6.4

$$|p(y) - p(x_{ij})| \leqslant$$
$$\leqslant 5|y - x_{ij}| + \sqrt{24|y - x_{ij}|^2 + 4|y - x_{ij}||p(x_{ij})|} =$$
$$= 5|p(x_{ij})| + \sqrt{24p^2(x_{ij}) + 4p^2(x_{ij})} \leqslant 11|p(x_{ij})|,$$

что дает $|p(y)| \leqslant 12|p(x_{ij})|$. Применяя лемму п. 6.4 к паре x_i, y и учитывая (6.16), получим

$$|p(x_i) - p(y)| \leqslant 5(\sqrt{5}+1)\ Q|p(y)| +$$
$$+ \sqrt{24(\sqrt{5}+1)^2 Q^2 p^2(y) + 4(\sqrt{5}+1)\ Qp^2(y)} \leqslant$$
$$\leqslant 10(\sqrt{5}+1)\ Q\ |p(y)|,$$

откуда

$$|p(x_i)| \leqslant 11\ (\sqrt{5}+1)\ Q|p(y)|,$$

т. е.

$$|p(x_i)| \leqslant 132(\sqrt{5}+1)\ Q|p(x_{ij})|,$$

что доказывает левое неравенство в (6.15).

6.11. Фиксируем $\lambda^{ij} \in \Lambda(x_{ij})$, и пусть $\rho_{ij} = \rho_{x_{ij}, \lambda^{ij}}$. Убедимся, что

$$s_{ij} \geqslant \rho_{x_{ij}, \lambda^{ij}} \tag{6.17}$$

и

$$s_{ij} \leqslant \frac{\vee(\bar{x})}{L_i^2} \cdot 1100(\sqrt{5}+1)^4 \bar{Q}^4. \tag{6.18}$$

Действительно, по лемме п. 6.3

$$\bar{f}(Tx_{ij}) \leqslant \bar{f}(x_i) - \frac{1}{2}\rho_{ij}p_{ij}^2,$$

тогда как

$$s_{ij} = \frac{2(-\bar{f}(T(Tx_{ij})) + \bar{f}(x_i))}{p_{ij}^2} \geqslant \frac{2(-\bar{f}(Tx_{ij}) + \bar{f}(x_i))}{p_{ij}^2}.$$

Соединение двух полученных неравенств и дает (6.17). Наконец,

$$s_{ij}p_{ij}^2 = \frac{1}{2}(\bar{f}(x_i) - \bar{f}(T(Tx_{ij}))) \leqslant \frac{1}{2}(\bar{f}(x_i) - \bar{f}(x_{i+1})) \leqslant \frac{1}{8}\vee(\bar{x})$$

и (6.18) следует из (6.15).

6.12. Имеем

$$f_q(x_{ij}) + \langle f_q'(x_{ij}) \mid \bar{x} - x_{ij} \rangle + \frac{l_q}{2}(\bar{x} - x_{ij})^2 \leqslant f_q(\bar{x}),$$

$$f_q(x_{ij}) + \langle f_q'(x_{ij}) \mid x^* - x_{ij} \rangle + \frac{l_q}{2}(x^* - x_{ij})^2 \leqslant f_q(x^*),$$

откуда

$$2f_q(x_{ij}) + \langle f_q'(x_{ij}) \mid x^* - \bar{x} \rangle + 2\langle f_q'(x_{ij}) \mid \bar{x} - x_{ij} \rangle +$$
$$+ \frac{l_q}{4}(x^* - \bar{x})^2 \leqslant f_q(\bar{x}) + f_q(x^*).$$

Из полученных неравенств

$$\frac{\rho_{ij}}{4Q}(x^* - \bar{x})^2 + \left\langle \sum_q \lambda_q^{ij} f_q'(x_{ij}) \mid x^* - \bar{x} \right\rangle +$$
$$+ 2\left\langle \sum_q \lambda_q^{ij} f_q'(x_{ij}) \mid \bar{x} - x_{ij} \right\rangle \leqslant \bar{f}(\bar{x}) + \bar{f}(x^*) - 2f_{\lambda^{ij}, x_{ij}}(x_{ij}). \tag{6.19}$$

Пусть n_{ij} есть производная $f_{\lambda^{ij}, x_{ij}}(y)$ по y при $y = Tx_{ij}$. Тогда $\langle n_{ij} \mid x - Tx_{ij} \rangle \geqslant 0$ при $x \in G$ и

$$\sum_q \lambda_q^{ij} f_q'(x_{ij}) - n_{ij} = \rho_{ij} p_{ij}.$$

Кроме того,

$$f_{\lambda^{ij},\,x_{ij}}(x_{ij}) = f_{\lambda^{ij},\,x_{ij}}(Tx_{ij}) + \frac{1}{2}\rho_{ij}p_{ij}^2 + \langle n_{ij} \mid p_{ij}\rangle.$$

Отсюда и из (6.19)

$$\frac{1}{4Q}\rho_{ij}(x^*-\bar{x})^2 + \langle \rho_{ij}p_{ij}+n_{ij} \mid x^*-\bar{x}\rangle + 2\langle \rho_{ij}p_{ij}+n_{ij} \mid \bar{x}-x_{ij}\rangle \leqslant$$
$$\leqslant \bar{f}(\bar{x}) + \bar{f}(x^*) - 2f_{\lambda^{ij},\,x_{ij}}(Tx_{ij}) - \rho_{ij}p_{ij}^2 - 2\langle n_{ij} \mid p_{ij}\rangle,$$

т. е.

$$\rho_{ij}\left(\frac{(x^*-\bar{x})^2}{4Q} + \langle p_{ij} \mid x^*-\bar{x}\rangle + 2\langle p_{ij} \mid \bar{x}-x_{ij}\rangle\right) \leqslant$$
$$\leqslant \bar{f}(\bar{x}) + \bar{f}(x^*) - 2f_{x_{ij}}(Tx_{ij}) +$$
$$+ \langle n_{ij} \mid \bar{x} - x^* + 2x_{ij} - 2\bar{x} + 2Tx_{ij} - 2x_{ij}\rangle \leqslant$$
$$\leqslant \bar{f}(\bar{x}) + \bar{f}(x^*) - 2\bar{f}(x_{i+1}) - \langle n_{ij} \mid \bar{x} - Tx_{ij}\rangle - \langle n_{ij} \mid x^* - Tx_{ij}\rangle \leqslant$$
$$\leqslant \bar{f}(\bar{x}) + \bar{f}(x^*) - 2\bar{f}(x_{i+1}) \leqslant \mathsf{V}(\bar{x}) - 2\mathsf{V}(x_{i+1}) < -{}^1/_2\mathsf{V}(\bar{x})$$

(мы учли, что

$$\langle n_{ij} \mid \bar{x} - Tx_{ij}\rangle \geqslant 0, \quad \langle n_{ij} \mid x^* - Tx_{ij}\rangle \geqslant 0$$
$$\text{и } \mathsf{V}(x_{i+1}) > {}^3/_4\mathsf{V}(\bar{x})).$$

Итак,

$$\langle p_{ij} \mid x^*-\bar{x}\rangle + 2\langle p_{ij} \mid \bar{x}-x_{ij}\rangle + \frac{(x^*-\bar{x})^2}{4Q} \leqslant -\frac{\mathsf{V}(\bar{x})}{2\rho_{ij}} \leqslant -\frac{\mathsf{V}(\bar{x})}{2s_{ij}}$$

(мы учли (6.17)). Далее, $\langle p_{ij} \mid x_i - x_{ij}\rangle \geqslant 0$, откуда

$$\langle p_{ij} \mid x^*-\bar{x}\rangle + 2\langle p_{ij} \mid \bar{x}-x_i\rangle + \frac{(x^*-\bar{x})^2}{4Q} \leqslant -\frac{1}{2}\frac{\mathsf{V}(\bar{x})}{s_{ij}}.$$

Усредним эти неравенства по j с весами μ_j. Получим

$$\langle \eta_i \mid x^*-\bar{x}\rangle + 2\langle \eta_i \mid \bar{x}-x_i\rangle + \frac{(x^*-\bar{x})^2}{4Q} \leqslant -\frac{1}{2}\mathsf{V}(\bar{x})\theta_i^2,$$

т. е.

$$\langle \eta_i' \mid x^*-\bar{x}\rangle + \frac{(x^*-\bar{x})^2}{4Q} \leqslant$$
$$\leqslant -\frac{1}{2}\mathsf{V}(\bar{x})\theta_i^2 + |\Delta_i|\,|x^*-\bar{x}| + 2|\Delta_i|\,|\bar{x}-x_i|, \tag{6.20}$$

где $\Delta_i = P_i\eta_i$. Заметим, что в силу (6.9) $\mathsf{V}(Tx_i) > {}^1/_2\,\mathsf{V}(x_i)$ и $\mathsf{V}(x^*) \leqslant \mathsf{V}(x_i)$, а также $\mathsf{V}(\bar{x}) \leqslant 2\mathsf{V}(x_i)$, откуда по утверждению 3) леммы п. 6.2

$$|x^* - x_i| \leqslant (\sqrt{5}+1)\,Q\,|p(x_i)| = L_i$$

и точно так же $|\bar{x} - x_i| \leqslant L_i$. Стало быть, $|x^* - \bar{x}| \leqslant 2L_i$ и $|\bar{x} - x_i| \leqslant L_i$, так что, ввиду (6.14), правая часть (6.20) не больше

$$-\frac{1}{2}\mathsf{V}(\bar{x})\theta_i^2 + \frac{64L_i^2}{2^{1/_2 M_3(Q)}}. \tag{6.21}$$

Заметим теперь, что в силу (6.18)

$$\frac{1}{s_{ij}} \geqslant \frac{L_i^2}{\mathsf{V}(\bar{x})\cdot 1100(\sqrt{5}+1)^4\,Q^4},$$

так что

$$\mathsf{V}(\bar{x})\theta_i^2 \geqslant \frac{L_i^2}{1100(\sqrt{5}+1)^4\,Q^4}. \tag{6.22}$$

Отсюда ясно, что при сделанном выборе $M_3(Q)$ величина (6.21) не больше $-{}^1/_4\,\mathsf{V}(\bar{x})\,\theta_i^2$.

Итак, (6.20) переписывается в виде

$$\langle \eta_i \mid x^*-\bar{x}\rangle + \frac{(x^*-\bar{x})^2}{4Q} \leqslant -\frac{1}{4}\mathsf{V}(\bar{x})\theta_i^2,$$

или

$$\langle \xi_i \mid x^*-\bar{x}\rangle + \frac{(x^*-\bar{x})^2}{4Q}\frac{1}{\theta_i} \leqslant -\frac{1}{4}\mathsf{V}(\bar{x})\theta_i. \tag{6.23}$$

Заметим теперь, что

$$\sum_{i=1}^{M_1(Q)} \xi_i^2 < \frac{1}{4}\mathsf{V}(\bar{x}). \tag{6.24}$$

В самом деле, по определению s_{ij}

$$\eta_i^2 = \Big(\sum_j \mu_j p_{ij}\Big)^2 = \Big(\sum_j \sqrt{\frac{\mu_j}{s_{ij}}}\sqrt{\mu_j s_{ij}}\,p_{ij}\Big)^2 \leqslant \sum_j \frac{\mu_j}{s_{ij}} \sum_j \mu_j s_{ij} p_{ij}^2 \leqslant$$
$$\leqslant \theta_i^2 \sum_j \mu_j(\bar{f}(x_i) - \bar{f}(x_{i+1})) = \theta_i^2(\bar{f}(x_i) - \bar{f}(x_{i+1})),$$

т. е. $\xi_i^2 \leqslant \bar{f}(x_i) - \bar{f}(x_{i+1})$, что и доказывает (6.24).

Суммируя (6.23) и учитывая, что $\xi_i \perp \sum_{j=1}^{i-1}\xi_j$, найдем

$$-\sqrt{\frac{1}{4}\mathsf{V}(\bar{x})}\,|x^*-\bar{x}| + \frac{|x^*-\bar{x}|^2}{4Q}\sum_{i=1}^{M_1(Q)}\frac{1}{\theta_i} < -{}^1/_4\,\mathsf{V}(\bar{x})\sum_{i=1}^{M_1(Q)}\theta_i,$$

так что квадратный трехчлен

$$\frac{1}{4Q}\Big(\sum_{i=1}^{M_1(Q)}\frac{1}{\theta_i}\Big)t^2 - \sqrt{\frac{1}{4}\mathsf{V}(\bar{x})}\,t + \frac{1}{4}\mathsf{V}(\bar{x})\sum_{i=1}^{M_1(Q)}\theta_i$$

имеет вещественные различные корни. Отсюда

$$^1/_4 \mathbf{V}(\bar{x}) > \frac{\mathbf{V}(\bar{x})}{4Q} \sum_{i=1}^{M_1(Q)} \theta_i \sum_{i=1}^{M_1(Q)} \frac{1}{\theta_i} \geqslant \frac{\mathbf{V}(\bar{x}) M_1^2(Q)}{4Q},$$

поскольку при положительных θ_i

$$\sum_1^M \frac{1}{\theta_i} \sum_1^M \theta_i \geqslant M^2.$$

Итак, $M_1^2(Q) < Q$, что противоречит определению $M_1(Q)$. Полученное противоречие доказывает утверждение.

Глава VIII

ЭФФЕКТИВНОСТЬ СТАНДАРТНЫХ МЕТОДОВ СИЛЬНО ВЫПУКЛОГО ПРОГРАММИРОВАНИЯ

В этой главе приведен анализ трудоемкости нескольких стандартных методов выпуклого программирования. Цель анализа — выяснить, в состоянии ли эти методы реализовать сложность классов сильно выпуклых задач. Ответ на этот вопрос во всех рассматриваемых случаях оказывается отрицательным.

§ 1. О стандартных методах сильно выпуклого программирования

Традиционные методы выпуклой оптимизации в первом приближении можно разделить на два класса — методы, для которых доказаны чистые «теоремы сходимости» без каких-либо оценок ее скорости, и методы, для которых устанавливается скорость сходимости (как правило, в асимптотике по точности). Подавляющее большинство методов второго класса описывается применительно к решению сильно выпуклых задач безусловной оптимизации — задач вида

$$f(x) \to \min \mid x \in E, \tag{1.1}$$

f — сильно выпукла. Цель главы — оценить трудоемкость нескольких методов такого рода.

1.1. Будем рассматривать только методы первого порядка решения задачи (1.1) (т. е. использующие информацию о значениях и градиентах f в просматриваемых точках). Традиционная классификация этих методов по скорости сходимости такова:

— метод сходится линейно, если строящиеся им последовательные приближения x_i сходятся к точке минимума x^* со скоростью геометрической прогрессии: $|x_i - x^*| \leqslant Aq^i$, $q < 1$;

— метод сходится сверхлинейно, если $|x_i - x^*|$ убывает быстрее любой геометрической прогрессии: $|x_i - x^*| \leqslant A_q q^i$ при всех $q < 1$;

— метод сходится квадратично, если

$$|x_i - x^*| \leqslant Aq^{(b^i)},\ 0 < q < 1,\ b > 1.$$

В первом случае оценка трудоемкости метода при точности ν имеет вид $a \ln(1/\nu)$, во втором — $\bar{O}(\ln(1/\nu))$, в третьем —

$a \ln \ln (1/\nu)$. Множитель, который мы обозначили через a, не зависит от требуемой точности, но зависит от других параметров класса (модуля сильной выпуклости решаемой задачи f, ее размерности, скорости изменения f'' и т. п.).

Традиционно считается, что различие в «рангах» двух методов по указанной классификации автоматически позволяет выделить лучший из них. Так, квадратично сходящиеся методы считаются более эффективными, чем линейно сходящиеся. Нам представляется, что такая оценка методов — по характеру асимптотической зависимости их трудоемкости от точности — далеко не всегда отвечает сути дела. Помимо точности, ситуация характеризуется и другими, по меньшей мере столь же существенными параметрами — модулем сильной выпуклости, размерностью решаемой задачи и т. п., и изучать влияние последних на трудоемкость, вообще говоря, не менее интересно, чем влияние на нее точности. Более того, сама «сверхлинейная», а тем более «квадратичная» сходимость методов первого порядка является, до некоторой степени, патологическим явлением.

Во-первых, методы сильно выпуклого программирования могут сходиться «сверхлинейно» лишь при определенных дополнительных (по сравнению с простой сильной выпуклостью) требованиях к гладкости решаемых задач. Обычно требуется некоторая непрерывность (или даже липшицевость) гессиана минимизируемой функции, и это не случайно. Действительно, в § 2 гл. VII мы видели, что сложность $N(\nu)$ класса (l, Q)-сильно выпуклых задач ($Q > 2$) по отношению к *любому* локальному детерминированному оракулу удовлетворяет оценке

$$N(\nu) \geqslant c(l, Q, \dim E) \ln \frac{1}{\nu}.$$

Стало быть, ни о какой «сверхлинейной сходимости» методов не только первого, но даже и любого, сколь угодно высокого, порядка на классе всех (l, Q)-сильно выпуклых задач не может быть и речи: сверхлинейная сходимость может возникнуть только при дополнительных требованиях к гладкости задачи, и соответствующие оценки трудоемкости будут чувствительны к значениям параметров, описывающих эту «повышенную» гладкость. С практической точки зрения именно эта последняя особенность «сверхлинейных» методов не очень приятна. Как правило, практические сильно выпуклые задачи сколь угодно гладки, однако априорные оценки «параметров гладкости», скажем, константы Липшица второй производной — часто весьма трудоемки *). Поэтому с практической точки зрения не всегда удобно использовать методы, чувствительные к «трудно наблюдаемым» свойствам задачи.

*) Конечно, и оценки параметров сильной выпуклости бывают трудоемкими, так что, возможно, стоит применять методы решения общих выпуклых задач и к сильно выпуклым, идя на потери в быстродействии, но снижая зато требования к априорной информации о решаемой задаче.

Но для «возникновения» сверхлинейных методов первого порядка одной только высокой гладкости решаемой задачи недостаточно. «Сверхлинейная» сходимость всегда чувствительна к размерности задачи. Действительно, мы видели, что при решении «самых гладких» — квадратичных (l, Q)-сильно выпуклых задач на R^n — точность ν недостижима для методов первого порядка за число шагов, меньшее $c\sqrt{Q} \ln (1/\nu)$ (считаем $Q \geqslant 2$), где $c > 0$ — абсолютная константа, если только $n > c\sqrt{Q} \ln (1/\nu)$. Таким образом, равномерная по размерности оценка трудоемкости методов первого порядка даже квадратичного программирования не может быть лучше линейной $O(\ln 1/\nu)$. Между тем зависимость трудоемкости метода от размерности, по крайней мере, столь же существенна, как и зависимость от точности. В этом смысле «квадратичная оценка скорости сходимости» может даже «незаслуженно очернить» метод. Пусть, например, имеется метод первого порядка решения (1,2)-строго выпуклых квадратичных задач, про который известно лишь, что его оценка трудоемкости имеет вид $a(\dim E) \ln \ln (1/\nu)$. Мы видели, что при $\sqrt{2} \ln (1/\nu) \leqslant 2 \dim E$ сложность указанного класса по отношению к оракулу первого порядка не меньше $c \ln 1/\nu$, $c > 0$ — абсолютная константа. Пусть ν определено соотношением $\ln 1/\nu = n = \dim E$. Тогда непременно $a(n) \ln \ln (1/\nu) \geqslant cn$, т. е. $a(n) \geqslant cn/\ln n$. Итак, оценка трудоемкости метода не может быть лучше $(cn/\ln n) \ln \ln (1/\nu)$. Эта оценка быстро растет с ростом n, так что для больших $n = \dim E$ метод (в той мере, в какой он характеризуется исходной оценкой трудоемкости) «как угодно» хуже простейшего градиентного метода (оценка трудоемкости которого на рассматриваемом классе $3 \ln (1/\nu)$).

Таким образом, «равномерной по размерности сверхлинейной сходимостью» могут обладать лишь методы, работающие с оракулами более высокого, чем первый, порядка, и то лишь на классах задач, «более гладких», чем сильно выпуклые. В естественных ситуациях такого рода действительно существуют «сверхлинейно сходящиеся» методы типа метода Ньютона.

Упражнения.

1. Пусть G — выпуклое замкнутое тело в гильбертовом пространстве, $x_1 \in G$, $F_m^{x_1}$ — класс всех задач вида

$$f_0(x) \to \min \mid x \in G,\ f_j(x) \leqslant 0,\ 1 \leqslant j \leqslant m,$$

таких, что $f_i(x)$ дважды непрерывно-дифференцируемы, (l_j, Q_j)-сильно выпуклы, $|f_j'(x_1)| \leqslant b_j$ и f_j'' удовлетворяют условию Липшица с константами L_j (здесь b_j, Q_j, l_j, L_j — заданные числа). Докажите, что сложность решения задач из $F_m^{x_1}$ с абсолютной погрешностью ε при использовании оракула второго порядка (сообщающего $f(x)$, $f'(x)$, $f''(x)$) допускает оценку

$$N(\varepsilon) \leqslant c(\rho_{|\cdot|}(G), \bar{l}, \bar{Q}, \bar{b}, \bar{L}) \ln\ln\frac{1}{\varepsilon}, \quad \varepsilon < 10^{-1};$$

здесь $(\bar{l}, \bar{Q}, \bar{b}, \bar{L})$ — список задающих класс параметров.

2. Рассмотрите класс C_k задач вида $f(x) \to \min \mid x \in \mathbb{R}$, у которых

$$|f'(0)| \leqslant 1, \quad 1 \leqslant f''(t) \leqslant 2, \; |f^s(t)| \leqslant 1, \quad 3 \leqslant s \leqslant k,$$

и докажите, что сложность решения задач этого класса с абсолютной погрешностью ε по отношению к любому детерминированному локальному оракулу допускает оценку снизу

$$N(\varepsilon) \geqslant c(k) \ln\ln\frac{1}{\varepsilon}, \quad c(k) > 0.$$

Таким образом, «квадратичная сходимость» — это лучшее, чего в естественных ситуациях можно ожидать от численных методов.

1.2. В соответствии со сказанным будем оценивать традиционные методы сильно выпуклого программирования на классах (l, Q)-сильно выпуклых задач типа (1.1) и интересоваться при этом характером зависимости трудоемкости от модуля строгой выпуклости Q. Мы не будем специально изучать влияние размерности на трудоемкость методов, ограничиваясь, когда это существенно, случаем достаточно высокой размерности. В этих условиях нас будет интересовать, реализует ли данный метод «оптимальную» оценку трудоемкости $O(\sqrt{Q} \ln 1/\nu)$. Напомним, что эта оценка потенциально неулучшаема и «почти» реализуется (с точностью до множителя $\ln Q$) методом § 3 гл. VII. Забегая вперед, укажем, что для тех традиционных методов, для которых ответ на поставленный вопрос нам известен, он отрицателен.

1.3. Методы, которые мы будем рассматривать, выглядят следующим образом. К i-му шагу уже построена точка x_{i-1} и очередное направление спуска p_{i-1}. На i-м шаге ($i \geqslant 2$) в качестве точки x_i выбирается точка минимума минимизируемой функции f на луче $\{x_{i-1} + tp_{i-1} \mid t \geqslant 0\}$, x_1 есть некоторая заранее фиксированная точка.

Методы отличаются друг от друга правилами построения p_{i-1} в функции от накопленной информации (т. е. от $x_1, \ldots, x_{i-1}$, $f'(x_1), \ldots, f'(x_{i-1})$ *). Собственно, указанная схема дает не метод, а идеализацию метода: одномерную задачу минимизации f на луче нельзя решить точно за конечное время. Такая идеализация основана на том, что одномерные задачи можно решать (и при достаточно высоких точностях) довольно быстро. Мы, как это обычно делается, будем считать, что точное решение одномерной задачи «занимает» один шаг.

Простейший метод такого рода — градиентный: $p_{i-1} = -f'(x_{i-1})$. Этот метод «вовсе не имеет памяти» — очередная точка

*) Значения $f(x_i)$ обычно не используются.

ищется на луче, выбираемом по информации о задаче в одной только предыдущей точке. Из рассуждений п. 1.5 гл. VII легко извлечь, что на (l, Q)-сильно выпуклой задаче f градиентный метод обеспечивает точность ν (т. е. в $1/\nu$ раз сокращает начальную невязку $f(x_1) - \inf_E f(x)$) за время

$$M(\nu) = \left] \frac{\ln\frac{1}{\nu}}{\ln\left(1 + \frac{1}{Q-1}\right)} \right[.$$

За существенно меньшее время метод не обеспечивает точности ν на классе $\boldsymbol{H}_{x_1}(E, 0, l, Q)$, $Q > 1$, если только $\dim E \geqslant 2$. Таким образом, трудоемкость градиентного метода, обеспечивающего точность $\nu < 1$ на классе $\boldsymbol{H}_{x_1}(E, 0, l, Q)$ при $Q \geqslant 2$ и $\dim E \geqslant 2$, не меньше $cQ \ln(1/\nu)$ ($c > 0$ — абсолютная константа). Градиентный метод «плохо ведет себя с ростом Q». Последний его недостаток — излишняя чувствительность к степени обусловленности задачи — хорошо известнен.

Упражнение. Докажите нижнюю оценку трудоемкости градиентного метода, проанализировав его работу на двумерных квадратичных задачах. ⟨Пусть $f(x) = \frac{1}{2}\langle Ax|x\rangle - \langle b|x\rangle$, $x_1 = 0$, а x_i — i-я точка работы метода на f. Тогда $x_i = x_{i-1} - t_i(Ax_{i-1} - b)$, t_i определено соотношением

$$\langle (Ax_{i-1} - b) - t_i A(Ax_{i-1} - b) | Ax_{i-1} - b\rangle = 0,$$

т. е.

$$t_i = \frac{(Ax_{i-1} - b)^2}{\langle A(Ax_{i-1} - b) \mid Ax_{i-1} - b\rangle},$$

и

$$x_i = x_{i-1} - \frac{(Ax_{i-1} - b)^2 (Ax_{i-1} - b)}{\langle A(Ax_{i-1} - b) \mid Ax_{i-1} - b\rangle}.$$

Положим $y_i = Ax_i - b$. Тогда

$$y_i = Ax_{i-1} - b - \frac{(Ax_{i-1} - b)^2 A(Ax_{i-1} - b)}{\langle A(Ax_{i-1} - b) \mid Ax_{i-1} - b\rangle} = y_{i-1} - \frac{y_{i-1}^2}{\langle Ay_{i-1} \mid y_{i-1}\rangle} Ay_{i-1}.$$

Пусть

$$A = \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad q = \frac{1-\alpha}{1+\alpha}.$$

Легко видеть, что векторы

$$y_i = \begin{pmatrix} -q^{i-1} \\ (-1)^i q^{i-1} \end{pmatrix}$$

удовлетворяют выписанным рекуррентным соотношениям. Таким образом, $-Ax_{i+1} + b = q^i \begin{pmatrix} 1 \\ (-1)^i \end{pmatrix}$. Но

$$f(x) - \inf_y f(y) = \frac{1}{2}\langle A^{-1}(Ax - b) \mid Ax - b\rangle,$$

так что

$$\frac{f(x_{i+1}) - \inf\limits_E f(x)}{f(x_1) - \inf\limits_E f(x)} = q^{2i}.$$

Итак, погрешность градиентного метода с M шагами на $(1/Q, Q)$-сильно выпуклой задаче (отвечающей $\alpha = 1/Q$) составляет $((Q-1)/(Q+1))^{2M-2}$. Следовательно, число шагов, обеспечивающее точность ν на $\boldsymbol{H}_{x_1}(R^2, 0; l, Q)$, не меньше

$$\left]\frac{\ln\frac{1}{\nu}}{2\ln\left(1+\frac{2}{Q-1}\right)}\right[,$$

что и требуется.⟩

1.4. Методы сопряженных градиентов. Так называется большая группа методов сильно выпуклого программирования, получающаяся так сказать, «аналитическим продолжением» МСГ с квадратичных задач на общие. Имеется в виду следующее. Метод сопряженных градиентов квадратичной минимизации может быть записан в виде $x_{i+1} = x_i + t_i p_i$, где t_i — точка минимума $f(x_i + tp_i)$ по всем $t \in \mathbb{R}$, p_i— очередное направление спуска, строящееся на основании накопленной информации по надлежащим формулам. Мы не указывали этих формул, когда занимались МСГ,— нам была важна его геометрия, а не аналитический вид. Важно заметить, что описывающие МСГ формулы можно записать многими эквивалентными (на квадратичных задачах) способами. Каждая из таких формальных схем МСГ может далее рассматриваться как описание некоего алгоритма общей выпуклой минимизации.

Ясно, что разные формальные схемы порождают теперь, вообще говоря, разные алгоритмы (напомним, что схемы были эквивалентны — формировали одни и те же последовательности направлений спуска — только на квадратичных задачах). Такой, до некоторой степени формальный, способ продолжения МСГ с квадратичных задач на общие и имелся в виду, когда речь шла об «аналитическом продолжении». Идея такого рода обобщений понятна. Если минимизируемая функция достаточно гладка, то в малой окрестности своего минимума она «почти квадратична». Стало быть, если рассматриваемый метод семейства вообще сходится, то после некоторого числа шагов (когда он начнет работать вблизи минимума) работа метода будет почти такой же, что и у обычного МСГ на кадратичной задаче. Но если $\dim E = n$, то МСГ с n шагами точно решает любую квадратичную задачу. Рассматриваемое его «аналитическое продолжение» и вблизи минимума не будет, конечно, гарантировать точного решения за n шагов.

Можно, однако, ожидать, что за эти n шагов оно сильно сократит невязку. И действительно, для большинства «аналитических продолжений» МСГ такого рода соображения могут быть реализованы и дают «сверхлинейные» или «квадратичные» оценки скорости сходимости (разумеется, в предположении достаточной гладкости решаемой задачи) *).

Как уже говорилось, нас интересует поведение анализируемых методов на задачах класса $\boldsymbol{H}_{x_1}(E, 0; l, Q)$ при достаточно высоких $\dim E$. Более точно, мы желаем понять, реализует ли какой-нибудь из этих методов «оптимальную» оценку трудоемкости $O(\sqrt{Q}\ln 1/\nu)$. С этой точки зрения нам мало помогают указанные результаты о сходимости, фиксирующие по существу лишь ее характер в окрестности оптимума и не касающиеся (по крайней мере, явно) связи скорости сходимости с параметром Q. Нам придется поэтому рассматривать каждый метод в отдельности и явно строить для него пример «медленно решаемой» задачи. Эту громоздкую работу мы проделаем для трех методов сопряженных градиентов. Отбор рассматриваемых методов в известной мере произволен. При описании методов будут приняты следующие обозначения: x_i — — i-я точка работы метода на рассматриваемой задаче f (нам удобно нумеровать их, начиная с 0, а не с 1), $g_i = -f'(x_i)$; p_i — направление спуска i-го шага, т. е.

$$x_{i+1} = x_i + t_i p_i, \tag{1.2}$$

где t_i выбрано как точка минимума $f(x_i + tp_i)$ по t;

$$\Delta x_i = x_{i+1} - x_i;\ \Delta g_i = g_{i+1} - g_i.$$

§ 2. Метод Флетчера — Ривса**)

В этом методе p_i выбирается по формуле

$$p_i = \frac{g_0}{|g_0|^2} + \ldots + \frac{g_i}{|g_i|^2}. \tag{2.1}$$

Мы построим для всякого достаточно большого Q и всех $l > 0$ такую (l, Q)-сильно выпуклую функцию на плоскости R^2, что ее минимизация методом ФР (Флетчера — Ривса) с точностью c/Q требует $\geqslant dQ^2$ шагов (c, $d > 0$—абсолютные константы). Таким образом, метод ФР, исходя из принятых нами критериев оценки, даже хуже градиентного метода.

2.1. Сразу сделаем одну оговорку. ФР, как и другие методы сопряженных градиентов, часто описывают как процесс «с восстановлением». Последнее означает, что процесс (1.2) — (2.1) ведется

*) Читатель может ознакомиться со схемой конструкции большого числа методов сопряженных градиентов и оценками их скорости сходимости по работе [4].

**) Описание метода заимствовано из [25], гл. II.

не все время, а некоторое число $N = k \dim E$ (k натурально) шагов. После этого процесс начинается из точки x_N заново, как из начальной точки x_1, и так — еще N шагов. Таким образом, через каждые N шагов («период восстановления») вся накопленная информация «забывается», и решение начинается из «текущей точки», как из начальной. «Восстановление» применяется для того, чтобы снять неограниченное последействие эффектов, связанных с неквадратичностью решаемой задачи. Иногда только за счет восстановления удается достичь сходимости метода при любом начальном приближении.

Между прочим, метод § 3 гл. VII тоже содержал восстановление (его работа на периоде между двумя восстановлениями называлась там этапом). Однако период восстановления этого метода был связан не со «случайной» размерностью задачи, а с ее внутренним свойством — модулем сильной выпуклости.

Мы не будем проводить анализ методов раздельно для версий с восстановлением и без него. Дело в том, что мы интересуемся работой методов лишь на задачах достаточно большой размерности; ее всегда можно выбрать столь большой, чтобы на рассматриваемом отрезке траектории восстановлений не было. Таким образом, мы всегда будем ограничиваться методами без восстановления. При этом сами примеры могут иметь и малую (в данном случае — 2) размерность. Дело в том, что (l, Q)-сильно выпуклой функции $f(x_1 \ldots, x_k)$ можно поставить в соответствие (l, Q)-сильно выпуклую функцию $\bar f(x_1, \ldots, x_n)$ любого числа переменных $n \geqslant k$ следующим образом:

$$\bar f(x_1, \ldots, x_n) = f(x_1, \ldots, x_k) + \frac{l}{2} \sum_{j=k+1}^{n} x_j^2.$$

Если начальное приближение лежит в плоскости первых k переменных (т. е. его координаты, начиная с $(k+1)$-й, равны 0), то работа любого метода сопряженных градиентов на $\bar f$ происходит все время в плоскости первых k переменных и совпадает с работой метода на f (это свойство всегда легко извлекается из описания метода). Итак, работу метода без восстановления на любой задаче размерности k можно интерпретировать как работу того же метода, но уже с восстановлением, на «столь же» сильно выпуклой задаче большой размерности. Поэтому достаточно заниматься методами без восстановления в любой, какая удобнее, размерности. Это замечание относится и к примерам в следующих параграфах.

2.2. Вернемся к методу Флетчера — Ривса. Пусть работа метода не заканчивается в конечное время. Чтобы построить пример, выразим все параметры траектории метода через углы между соседними смещениями. Более точно, пусть φ_{i+1} есть угол между Δx_i и Δx_{i+1} (лежащий между 0 и π). Легко проверить, что справедливы следующие соотношения:

$$|g_{i+1}| = \frac{t_0 \cos\varphi_1 \ldots \cos\varphi_{i+1}}{|\Delta x_0| \sin\varphi_{i+1}}, \tag{2.2}$$

$$|g_0| = \frac{t_0}{|\Delta x_0|}. \tag{2.3}$$

Предположим, что на плоскости R^2 задана последовательность точек $x_0, x_1, \ldots, x_i, \ldots$ и векторов $g_1, g_2, \ldots$, таких, что

(i) $\langle \Delta x_i \mid \Delta x_{i+1} \rangle > 0$;

(ii) $\langle g_i \mid \Delta x_i \rangle > 0$;

(iii) $g_0 = \Delta x_0$, $|g_0| = 1$;

(iv) $|g_i| = \dfrac{\cos\varphi_1 \ldots \cos\varphi_i}{\sin\varphi_i}$, $i \geqslant 1$;

(v) $\langle g_i \mid \Delta x_{i-1} \rangle = 0$, $i \geqslant 1$.

Здесь φ_i — угол между Δx_i и Δx_{i-1}, $i \geqslant 1$. Предположим далее, что существует выпуклая функция $f(x)$, такая, что $f'(x_i) = -g_i$, $i \geqslant 0$. Убедимся, что $x_0, \ldots, x_i$ — траектория ФР на f. Проведем индукцию по i. Пусть уже известно, что $x_0, \ldots, x_i$ — первые $i+1$ точки траектории ФР на f. Для $i = 1$ это так в силу (iii) и (v). Пусть теперь $i > 1$. Пусть y_{i+1} — $(i+2)$-я точка траектории ФР на f. Чтобы доказать, что $y_{i+1} = x_{i+1}$, заметим, что $y_{i+1} - x_i$ должно составлять острый угол с Δx_{i-1} и острый угол с g_i.

Ввиду (2.2) и предположения индукции, этот угол ψ определен соотношением

$$\operatorname{ctg}\psi = \frac{|g_i|\,|\Delta x_0|}{t_0 \cos\varphi_1 \ldots \cos\varphi_{i-1}}.$$

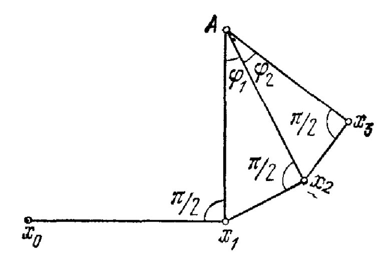

Рис. 8.1.

В силу (iii) и определения $t_0 \mid \Delta x_0 \mid = t_0 = 1$. Учитывая (iv), получаем $\operatorname{ctg}\psi = \operatorname{ctg}\varphi_i$, так что $y_{i+1} - x_i$ коллинеарно Δx_i. Но тогда $y_{i+1} = x_{i+1}$, так как в силу описания метода y_{i+1} должна быть точкой луча $\{x_i + t\Delta x_i\}$, в которой $f'(y_{i+1})$ ортогонально Δx_i. Такой точкой по (v) и является x_{i+1}.

Теперь рассмотрим следующее правило построения x_i (рис. 8.1): $\{x_i\}$ — ломаная «спираль», наматывающаяся на точку A; φ_i — углы между направлениями из A на x_i, причем Δx_i ортогонально Ax_{i+1}. Векторы g_k определены (для $k \geqslant 1$) как $\lambda_k (A - x_k)$. При этом $|x_1 - x_0| = 1$, $g_0 = x_1 - x_0$. Далее считаем $A = 0$ и $g_k = -\lambda_k x_k$, $\lambda_k \geqslant 0$, $k \geqslant 1$. Выберем φ_k так, чтобы было $|g_k| = \alpha + 2\beta\,|x_k|$ и чтобы при этом выполнялось (iv). Пусть

$|x_1| = 1/_2$. Требуется выбрать углы $\varphi_1, \varphi_2, \ldots$ так, чтобы удовлетворялись уравнения

$$\frac{\cos\varphi_1\ldots\cos\varphi_k}{\sin\varphi_k} = \alpha + \beta\cos\varphi_1\ldots\cos\varphi_{k-1}, \quad k \geqslant 1,$$

ибо $|x_k| = |x_1|\cos\varphi_1\ldots\cos\varphi_{k-1}$ (для $k=1$ $\cos\varphi_1\ldots\cos\varphi_{k-1} = 1$). Иными словами, углы φ_k находятся по рекуррентной формуле

$$\operatorname{ctg}\varphi_k = \frac{\alpha}{\cos\varphi_1\ldots\cos\varphi_{k-1}} + \beta, \quad \operatorname{ctg}\varphi_1 = \alpha + \beta. \tag{2.4}$$

Полагая $t_i = \cos\varphi_i$, получим для $k \geqslant 1$

$$\frac{\sqrt{1-t_k^2}}{t_k} = \frac{t_1\cdots t_{k-1}}{\alpha + \beta t_1\cdots t_{k-1}}$$

(при $k = 1$ правая часть равна $1/(\alpha+\beta)$), или

$$\frac{1}{t_k^2} = 1 + \frac{t_1^2 t_2^2\cdots t_{k-1}^2}{(\alpha+\beta t_1\cdots t_{k-1})^2},$$

т. е.

$$t_k^2 = \frac{(\alpha+\beta t_1\cdots t_{k-1})^2}{(\alpha+\beta t_1\cdots t_{k-1})^2 + t_1^2\cdot\ldots\cdot t_{k-1}^2}. \tag{2.5}$$

Наряду с последовательностью, удовлетворяющей (2.5), рассмотрим последовательность $\sigma_k = (t_1,\ldots t_k)^{-1}$. Тогда

$$\left(\frac{\sigma_k^2}{\sigma_{k-1}^2}\right)^{-1} = \frac{\left(\alpha + \frac{\beta}{\sigma_{k-1}}\right)^2}{\left(\alpha+\frac{\beta}{\sigma_{k-1}}\right)^2 + \frac{1}{\sigma_{k-1}^2}},$$

или

$$\sigma_k^2 = \sigma_{k-1}^2\left(1 + \frac{1}{(\alpha\sigma_{k-1}+\beta)^2}\right) = \sigma_{k-1}^2 + \frac{\sigma_{k-1}^2}{(\alpha\sigma_{k-1}+\beta)^2} \leqslant \sigma_{k-1}^2 + \frac{1}{\alpha^2}.$$

Кроме того,

$$\sigma_1^2 = \frac{1}{t_1^2} = (\alpha+\beta)^2 + 1.$$

Следовательно,

$$\sigma_k^2 \leqslant \frac{k-1}{\alpha^2} + (\alpha+\beta)^2 + 1$$

и

$$(t_1\ldots t_k)^2 \geqslant \frac{\alpha^2}{k-1+\alpha^2(\alpha+\beta)^2+\alpha^2}, \quad k \geqslant 1.$$

Таким образом, можно построить последовательности $\{x_i\}$ и $\{g_i\}$ так, чтобы было

(1) $\langle\Delta x_i | g_{i+1}\rangle = 0, \quad i = 0, 1, 2, \ldots;$

(2) $|x_{i+1}| = |x_i|\cos\varphi_i, \quad i \geqslant 1;$

(3) $g_i = -\dfrac{\alpha + 2\beta|x_i|}{|x_i|}x_i, \quad i \geqslant 1;$

(4) $|g_i| = \dfrac{\cos\varphi_1\ldots\cos\varphi_i}{\sin\varphi_i}, \quad i \geqslant 1;$

(5) $(\cos\varphi_1\ldots\cos\varphi_i)^2 \geqslant \dfrac{\alpha^2}{(i-1)+\alpha^2(\alpha+\beta)^2+\alpha^2}, \quad i \geqslant 1;$

(6) $g_0 = \Delta x_0, \quad |g_0| = 1.$

При этом φ_i — угол между Δx_i и Δx_{i-1}. Ясно, что функция $f_0(x) = \alpha|x| + \beta x^2$ обладает тем свойством, что $f_0'(x_i) = -g_i$, $i \geqslant 1$. Зададимся теперь параметром ε, $0 < \varepsilon < 1/_2$, и рассмотрим функцию

$$f_\varepsilon(x) = \begin{cases} \alpha|x| + \beta x^2, & |x| \geqslant \varepsilon, \\ \dfrac{\alpha}{2\varepsilon}x^2 + \dfrac{\alpha\varepsilon}{2} + \beta x^2, & |x| \leqslant \varepsilon. \end{cases}$$

Легко видеть, что $f_\varepsilon(x)$ $\left(\beta, \dfrac{4\alpha}{\varepsilon\beta} + 2\right)$ сильно выпукла. Кроме того, $f_\varepsilon(x)$ совпадает с $f_0(x)$ при $|x| \geqslant \varepsilon$. Поэтому при $i \geqslant 1$, для которых $|x_i| \geqslant \varepsilon$, имеем

$$f_\varepsilon'(x_i) = f_0'(x_i) = -g_i. \tag{2.6}$$

Так как $|x_i| = |x_1|\cos\varphi_1\ldots\cos\varphi_{i-1} = \frac{1}{2}\cos\varphi_1\ldots\cos\varphi_{i-1}$, то

$$|x_i| \geqslant \frac{1}{2}\sqrt{\frac{\alpha^2}{(i-2)+\alpha^2(\alpha+\beta)^2+\alpha^2}}, \quad i \geqslant 2,$$

так что (2.6) выполнено при

$$i < i(\varepsilon) = \left]2 + \left(\frac{\alpha^2}{4\varepsilon^2} - \alpha^2(\alpha+\beta)^2 - \alpha^2\right)\right[. \tag{2.7}$$

Чтобы превратить f_ε в нужную «плохую задачу» $\bar f_\varepsilon$ для ФР, надо еще обеспечить выполнение условия $\bar f_\varepsilon'(x_0) = -g_0$. Будем искать $\bar f_\varepsilon$ в виде $\bar f_\varepsilon(x) = \Delta(x) + f_\varepsilon(x)$, где $\Delta(x)$ — выпуклая функция, равная 0 в круге $|x| \leqslant 1/_2$, такая, что

$$\bar f_\varepsilon'(x_0) = -g_0,$$

т. е.

$$\Delta_\varepsilon'(x_0) = -g_0 - f_\varepsilon'(x_0) = -g_0 - \frac{\alpha x_0}{|x_0|} - 2\beta x_0 \equiv p.$$

Пусть теперь $|x| \leqslant {}^1/_2$. Тогда

$$\langle p \mid x - x_0\rangle \leqslant {}^1/_2 \mid p \mid + \langle x_0 \mid g_0\rangle + \alpha \mid x_0 \mid + 2\beta \mid x_0 \mid^2;$$

$$|x_0|^2 = 1 + {}^1/_4 = {}^5/_4;$$

$$\langle x_0 | g_0\rangle = -1; \quad |p|^2 = 1 + (\alpha + 2\beta |x_0|)^2 + \\ + \left(\frac{\alpha}{|x_0|} + 2\beta\right)\langle x_0 | g_0\rangle \leqslant 1 + (\alpha + 4\beta)^2.$$

Итак,

$$\langle p \mid x - x_0\rangle \leqslant {}^1/_2 (1 + (\alpha + 4\beta)) + 2\alpha + 4\beta - 1 \leqslant -{}^1/_2 + \\ + 3\alpha + 6\beta.$$

Положим $\alpha = {}^1/_{24}$, $\beta = {}^1/_{48}$. Тогда $\langle p \mid x - x_0\rangle \leqslant -{}^1/_4$ и $|p| \leqslant 2$.

Рассмотрим гладкую функцию θ на оси, равную 0 при $t \leqslant -{}^1/_4$ и $t + 1$ в окрестности точки $t = 0$, монотонную и выпуклую. Тогда можно взять $\Delta_\varepsilon(x) = \theta(\langle p \mid x - x_0\rangle)$. Ясно, что Δ_ε c-гладка (c — абсолютная константа) и $\Delta_\varepsilon'(x_0) = p$. Итак, для каждого ε, $0 < \varepsilon < {}^1/_2$, определена функция $\bar f_\varepsilon(x) = \Delta_\varepsilon(x) + f_\varepsilon(x)$, являющаяся $({}^1/_{48}, c/\varepsilon)$-сильно выпуклой и такой, что $\bar f_\varepsilon'(x_i) = -g_i$, $i = 0, 1, 2, \ldots, i(\varepsilon) - 1$. Сравнивая (1) — (6) с (i) — (v), видим, что $x_0, \ldots, x_{i(\varepsilon)-1}$ — *начальный отрезок траектории* ФР на $\bar f_\varepsilon$. Далее, $\bar f_\varepsilon$, очевидно, имеет модуль сильной выпуклости $Q(\varepsilon) \leqslant \bar c/\varepsilon$. Кроме того,

$$\bar f_\varepsilon(x_0) - \min \bar f_\varepsilon \leqslant c_1,$$

c_1 не зависит от ε, тогда как при $|x| \geqslant \varepsilon$ имеем

$$\bar f_\varepsilon(x) - \min \bar f_\varepsilon \geqslant c_2\varepsilon, \quad c_2 > 0]$$

не зависит от ε.

Итак, метод ФР на функции $\bar f_\varepsilon(x)$, имеющей модуль сильной выпуклости $Q(\varepsilon) \leqslant \bar c/\varepsilon$, обеспечивает точность $c_3\varepsilon$ ($c_3 > 0$ — абсолютная константа) не ранее чем через $i(\varepsilon)$ шагов, $i(\varepsilon) =]c_4/\varepsilon^2[$, $c_4 > 0$ не зависит от ε. Если теперь дано $Q \geqslant 2\bar c$, то можно взять $\varepsilon = \bar c/Q$. При этом ε функция $\bar f_\varepsilon$ доставит пример функции двух переменных с модулем сильной выпуклости Q, на которой метод ФР обеспечивает точность c/Q лишь за dQ^2 шагов ($c, d > 0$ — абсолютные константы).

Остается еще заметить, что траектория ФР не меняется при умножении функции на константу. Поэтому построенную функцию можно отнормировать так, чтобы она стала (l, Q)-сильно выпукла, каково бы ни было данное $l > 0$. Требуемый пример построен.

Мы видим, между прочим, как пагубно может сказываться на ФР «последействие»: при всех $i \geqslant 1$ $\bar f_\varepsilon'(x_0)$ точно указывает на точку минимума $\bar f_\varepsilon$, а метод из-за «памяти» строит траекторию, чем дальше, тем «более ортогональную» направлению на минимум (разумеется, до определенного момента — именно, до момента $i(\varepsilon)$).

§ 3. Метод Полака — Рибьера*)

В этом методе $p_0 = g_0$, $p_{i+1} = g_{i+1} + \gamma_{i+1}p_i$, где

$$\gamma_i = \frac{\langle g_{i+1} - g_i \mid g_{i+1}\rangle}{g_i^2}.$$

Покажем, что для любого $l > 0$ и всех достаточно больших Q существует (l, Q)-сильно выпуклая функция трех переменных, на которой метод Полака — Рибьера (ПР) обеспечивает точность $\nu_0 > 0$ не ранее чем после cQ шагов (ν_0, $c > 0$ — абсолютные константы). Таким образом, ПР не реализует «оптимальную» оценку трудоемкости, и, исходя из принятых в настоящей работе критериев оценки, он не лучше градиентного метода.

Перейдем к конструкции примера.

3.1. Докажем, что

$$\langle p_{i+1} \mid g_{i+1} - g_i\rangle = 0, \quad i \geqslant 0. \tag{3.1}$$

Заметим, что из $\langle p_i \mid g_{i+1}\rangle = 0$, $i \geqslant 0$, следует $\langle p_{i+1} \mid g_{i+1}\rangle = g_{i+1}^2$, $i \geqslant 0$; кроме того, $p_0 = g_0$, так что $\langle p_0 \mid g_0\rangle = g_0^2$. Итак, при всех $i \geqslant 0$

$$\langle p_i \mid g_i\rangle = g_i^2. \tag{3.2}$$

Следовательно,

$$\langle p_{i+1} | g_{i+1} - g_i\rangle = \langle g_{i+1} + \gamma_{i+1}p_i | g_{i+1} - g_i\rangle = \\ = g_{i+1}^2 - \langle g_{i+1} | g_i\rangle + \gamma_{i+1}\langle p_i | g_{i+1}\rangle - \gamma_{i+1}\langle p_i | g_i\rangle = \\ = g_{i+1}^2 - \langle g_{i+1} | g_i\rangle - g_i^2 \frac{\langle g_{i+1} - g_i | g_{i+1}\rangle}{g_i^2} = 0,$$

что доказывает (3.1). Это значит, что работу метода можно описать следующим образом **): очередное — $(i + 1)$-е — направление спуска лежит в плоскости, проходящей через Δx_i и g_{i+1}. В этой плоскости указанное направление выделяется двумя условиями: (1) оно образует острый угол с g_{i+1}; (2) оно ортогонально $g_{i+1} - g_i$. Эти два условия действительно однозначно определяют направление спуска. В самом деле, иначе было бы $g_{i+1} - g_i$ ортогонально $\mathfrak{L}(\Delta x_i, g_{i+1})$, и так как Δx_i ортогонально g_{i+1}, то было бы $\langle \Delta x_i \mid g_i\rangle = 0$, т. е. $g_i = 0$ (ибо при $g_i \neq 0$ $\Delta x_i = t_ip_i$, $t_i > 0$), что по предположению не так.

3.2. Теперь можно обратить наши рассуждения. Именно, пусть заданы последовательность точек $x_0, x_1 \ldots$ и последователь-

*) Описание метода взято из [25], гл. II.

**) Предполагается, что метод не останавливается через конечное число шагов выходом в точку минимума решаемой задачи ность векторов $g_0, g_1, \ldots$, причем

(i) $t_0 g_0 = \Delta x_0$, $t_0 > 0$, $g_0 \neq 0$;

(ii) $\langle g_{i+1} \mid \Delta x_i\rangle = 0$, $i \geqslant 0$;

(iii) $\Delta x_{i+1} \in \mathfrak{L}(\Delta x_i, g_{i+1})$, причем $\langle \Delta x_{i+1} \mid g_{i+1}\rangle > 0$ и $\langle \Delta x_{i+1} \mid g_{i+1} - g_i\rangle = 0$.

Пусть, кроме того, существует выпуклая функция f, такая, что $f'(x_0) = -g_i$, $i \geqslant 0$. Тогда $\{x_i\}$ — траектория ПР на f.

Это утверждение доказывается простой индукцией.

3.3. Теперь рассмотрим следующую конструкцию. Пусть (e_1, e_2, e_3) — ортобазис в трехмерном пространстве R^3. Рассмотрим цилиндрическую поверхность, направляющей которой является правильный пятиугольник в плоскости $\mathfrak{L}(e_1, e_2)$, а образующая коллинеарна e_3. Рассмотрим «обмотку» $x_0\, x_1\, x_2\, x_3\, x_4\, x_5 \ldots$ этой призмы, пересекающую образующие под постоянным углом ψ (рис. 8.2). После этого подберем векторы g_i так, чтобы последовательности $\{x_i\}$, $\{g_i\}$ удовлетворяли всем условиям (ii) — (iii). Оказывается, ψ можно выбрать так, чтобы было $g_{i+1} = t \mid g_i \mid$, $i \geqslant 1$, $t < 1$ достаточно близко к 1. После этого мы научимся строить функцию, имеющую антиградиентом в x_i вектор g_i, и исправим ее так, чтобы выполнялось условие (i). При надлежащем t эта функция дает нужный пример.

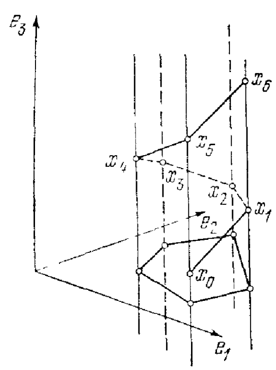

Рис. 8.2.

Перейдем к описанию конструкции. Обозначим через θ угол $2\pi/5 = 72°$, а через u_θ — вращение на угол θ в плоскости $\mathfrak{L}(e_1, e_2)$. Пусть $q_0 \in \mathfrak{L}(e_1, e_2)$, $\mid q_0 \mid = 1$, и $q_i = (u_\theta)^i q_0$, $-\infty < i < \infty$. Введем параметр $\beta > 0$. Положим

$$\left.\begin{aligned} &\Delta x_i = q_i + \beta e_3, && -\infty < i < \infty,\\ \text{и}\quad &r_i = (q_i - \cos\varphi q_{i-1}) + \beta(1 - \cos\varphi) e_3, && -\infty < i < \infty. \end{aligned}\right\} \tag{3.3}$$

Здесь φ выбрано так, что

$$0 < \varphi < \frac{\pi}{2} \quad \text{и} \quad \cos\varphi = \frac{\beta^2 + \cos\theta}{\beta^2 + 1}.$$

Отметим следующие факты.

(А) $r_i \in \mathfrak{L}(\Delta x_{i-1}, \Delta x_i)$ и $\Delta x_i \in \mathfrak{L}(\Delta x_{i-1}, r_i)$ $\langle$в самом деле, $r_i = \Delta x_i - \cos\varphi \Delta x_{i-1}\rangle$.

(Б) φ есть угол между Δx_{i-1} и Δx_i $\langle$в самом деле,

$$(|\Delta x_i|\,|\Delta x_{i-1}|)^{-1}\langle \Delta x_i \mid \Delta x_{i-1}\rangle = \frac{\cos\theta + \beta^2}{1+\beta^2} = \cos\varphi\rangle.$$

(В) $\langle r_i \mid \Delta x_{i-1}\rangle = 0 < \langle r_i \mid \Delta x_i\rangle$ $\langle$в самом деле,

$$\langle \Delta x_{i-1} \mid \Delta x_i - \Delta x_{i-1}\cos\varphi\rangle = \\ = |\Delta x_i|\,|\Delta x_{i-1}|\cos\varphi - |\Delta x_{i-1}|^2\cos\varphi = 0,$$

ибо $|\Delta x_i|$ не зависит от i; далее,

$$\langle \Delta x_i \mid \Delta x_i - \Delta x_{i-1}\cos\varphi\rangle = (\Delta x_i^2)(1 - \cos^2\varphi) = \Delta x_i^2 \sin^2\varphi\rangle.$$

Теперь докажем, что при надлежащем $t > 0$ и $\bar g_i = t^i r_i$ будет $\langle \Delta x_{i+1} \mid \bar g_{i+1} - \bar g_i\rangle = 0$. Действительно, мы должны выбрать t из условия: для всех i

$$\langle q_{i+1} + \beta e_3 \mid t^{i+1}\{(q_{i+1} - \cos\varphi q_i) + \beta(1-\cos\varphi)e_3\} - \\ - t^i\{(q_i - \cos\varphi\, q_{i-1}) + \beta(1-\cos\varphi)e_3\}\rangle = 0,$$

или

$$tq_{i+1}^2 - t\cos\varphi\langle q_{i+1} \mid q_i\rangle - \langle q_{i+1} \mid q_i\rangle + \\ + \cos\varphi\langle q_{i+1} \mid q_i\rangle + t\beta^2(1-\cos\varphi) - \beta^2(1-\cos\varphi) = 0,$$

т. е.

$$t - t\cos\varphi\cos\theta - \cos\theta + \cos\varphi\cos 2\theta + \\ + t\beta^2(1-\cos\varphi) - \beta^2(1-\cos\varphi) = 0,$$

или

$$t = \frac{\cos\theta - \cos\varphi\cos 2\theta + \beta^2(1-\cos\varphi)}{\beta^2(1-\cos\varphi) + 1 - \cos\varphi\cos\theta}.$$

Замечая, что

$$\cos 2\theta = \cos\frac{4\pi}{5} = \cos\left(\pi - \frac{\pi}{5}\right) = -\cos\frac{\theta}{2},$$

получим

$$t = \frac{\cos\theta + \cos\varphi\cos\frac{\theta}{2} + \beta^2(1-\cos\varphi)}{\beta^2(1-\cos\varphi) + 1 - \cos\varphi\cos\theta}. \tag{3.4}$$

Убедимся, что имеется такое $\Delta > 0$, что для всех t, $1 - \Delta \leqslant t \leqslant 1$, существуют β_t и φ_t, такие, что

$$\left.\begin{aligned} \cos\varphi_t &= \frac{\beta_t^2 + \cos\theta}{\beta_t^2 + 1},\\ t &= \frac{\cos\theta + \cos\frac{\theta}{2}\cos\varphi_t + \beta_t^2(1-\cos\varphi_t)}{\beta_t^2(1-\cos\varphi_t) + 1 - \cos\varphi_t\cos\theta}, \end{aligned}\right\} \tag{3.5}$$

и при этом $\beta_t > 0$, $0 < \varphi_t < \pi/2$ и β_t, φ_t непрерывно зависят от t. Действительно, при $t = 1$ система (3.5) имеет решение с нужными

свойствами. Она переписывается в виде

$$\cos\varphi_1 = \frac{1-\cos\theta}{\cos\theta+\cos\frac{\theta}{2}}, \qquad (3.5')$$

$$\cos\varphi_1 = \frac{\beta_1^2+\cos\theta}{\beta_1^2+1}, \qquad (3.5'')$$

и прямое вычисление $(1-\cos\theta)/\left(\cos\theta+\cos\frac{\theta}{2}\right)$ дает, что эта величина лежит между $^1/_2$ и 0,9. Отсюда ясно, что (3.5′) разрешимо в интервале $(0, \pi/2)$ и $\cos\varphi_1 > \cos\theta$. В силу последнего соотношения (3.5″) имеет решение $\beta_1 > 0$. Остается заметить, что для исходной (параметрически зависящей от t) системы в точке φ_1, β_1 выполнены условия теоремы о неявной функции, так что она имеет непрерывно зависящее от t решение при всех t, достаточно близких к 1.

Пусть теперь $1-\Delta \leqslant t \leqslant 1$. Положим $\varphi = \varphi_t$, $\beta = \beta_t$, и пусть Δx_i, r_i определены по (3.3) с выбранными φ и β.

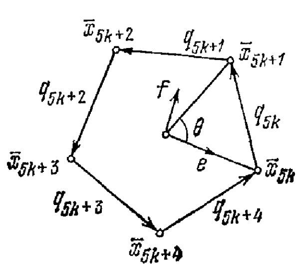

Рис. 8.3.

3.4. Рассмотрим траекторию $\{x_i\}_{i\geqslant 0}$, у которой

$$x_i = x_0 + \sum_{j=0}^{i-1}\Delta x_i,$$

x_0 выбрано в плоскости $\mathfrak{L}(e_1, e_2)$ так, что проекции $\bar{x}_i$ точек x_i на эту плоскость суть вершины правильного пятиугольника с центром в 0 и стороной 1 (очевидно, это можно сделать; на рис. 8.3 изображена проекция траектории на $\mathfrak{L}(e_1, e_2)$). Положим $v_i = q_i - q_{i-1}\cos\varphi$ (так что v_i есть проекция r_i на $\mathfrak{L}(e_1, e_2)$). Убедимся, что $\langle v_i \mid \bar{x}_i\rangle < 0$. Действительно, выбрав ортонормированный базис (e, f) на плоскости так, как показано на рис. 8.3 (где $i = 5k$), получим

$$q_i = -\cos\left(\frac{\pi-\theta}{2}\right)e + \sin\left(\frac{\pi-\theta}{2}\right)f = -\sin\frac{\theta}{2}e + \cos\frac{\theta}{2}f,$$

$$q_{i-1} = \cos\left(\frac{\pi-\theta}{2}\right)e + \sin\left(\frac{\pi-\theta}{2}\right)f = \sin\frac{\theta}{2}e + \cos\frac{\theta}{2}f$$

и

$$\langle v_i \mid e\rangle = -\sin\frac{\theta}{2}(1+\cos\varphi) < 0,$$

т. е. и $\langle v_i \mid \bar{x}_i\rangle < 0$. Проверим еще, что при t, достаточно близких к 1, прямая, ортогональная v_i и проходящая через $\bar{x}_i$, опорна к пятиугольнику (т. е. если $\bar{x}$—точка пятиугольника, то $\langle \bar{x}-\bar{x}_i \mid v_i\rangle \geqslant$ $\geqslant 0$). Достаточно доказать, что $\langle v_i \mid q_i\rangle > 0$ и $\langle v_i \mid -q_{i-1}\rangle > 0$. Имеем

$$\langle v_i \mid q_i\rangle = q_i^2 - \langle q_i \mid q_{i-1}\rangle\cos\varphi_t = 1-\cos\theta\cos\varphi_t > 0$$

и

$$\langle v_i \mid -q_{i-1}\rangle = -\cos\theta + \cos\varphi_t.$$

При $t = 1$ имеем

$$\cos\varphi_t = \frac{1-\cos\theta}{\cos\theta+\cos\frac{\theta}{2}} > \cos\theta$$

(в этом можно убедиться прямым вычислением). Стало быть, $\langle v_i \mid -q_{i-1}\rangle > 0$ при $t = 1$, а тогда это неравенство верно и для всех достаточно близких к 1 значений t. Отсюда следует, что функция $\psi_i(\bar{x}) = \langle v_i \mid -\bar{x}\rangle$ на плоскости $\mathfrak{L}(e_1, e_2)$ достигает на пятиугольнике максимума в точке $\bar{x}_i$.

Рассмотрим выпуклую функцию на $\mathfrak{L}(e_1, e_2)$:

$$\psi(\bar{x}) = \max_{0\leqslant i\leqslant 4}\psi_i(\bar{x}).$$

Она равна $\langle -v_i \mid x\rangle$ в окрестности каждой точки $\bar{x}_i$, и поэтому ее антиградиент в $\bar{x}_i$ равен v_i. Ясно, что эту функцию можно сгладить и превратить в $(\bar{l}, \bar{Q})$-сильно выпуклую функцию $\psi_t(\bar{x})$, сохранив соотношение

$$\psi_t(\bar{x}_{i+1}) = \psi_t(\bar{x}_i), \quad \psi_t'(\bar{x}_i) = -v_i \qquad (3.6)$$

и условие $\psi_t(0) = 0$, $\psi_t'(0) = 0$. При этом числа $\bar{l}$, $\bar{Q}$ можно считать не зависящими от t при условии $1 \geqslant t \geqslant 1-\Delta$, $\Delta > 0$ — абсолютная, достаточно малая, константа.

3.5. Пусть теперь $t < 1$ и $t \geqslant 1-\Delta$. Определим $\varkappa > 0$ из условия $e^{-\varkappa\beta_t} = t$ $(\varkappa = \varkappa(t))$. Образуем функцию $F(x) = F(\bar{x}, y)$ на R^3 ($\bar{x}$ — проекция x на $\mathfrak{L}(e_1, e_2)$, $y = \langle x \mid e_3\rangle$) по формуле

$$F(x) = \frac{\beta_t(1-\cos\varphi_t)}{\varkappa(t)}e^{-\varkappa(t)y} + e^{-\varkappa(t)y}(\psi_t(\bar{x}) - \psi_t),$$

где $\psi_t = \psi_t(\bar{x}_i)$ (величина справа не зависит от i). Покажем, что $F'(x_i) = -g_i$, $i \geqslant 0$. Действительно, если x таково, что $\psi_t(\bar{x}) = \psi_t$, то

$$F_{\bar{x}}'(x) = e^{-\varkappa y}\psi_t'(\bar{x}),$$

$$F_y'(x) = -a_t e^{-\varkappa y}, \quad \text{где } a_t = \beta_t(1-\cos\varphi_t).$$

Стало быть,

$$F_{\bar{x}}'(x_i) = e^{-\varkappa i\beta_t}\psi_t'(\bar{x}_i) = -t^i v_i,$$

$$-F_y'(x_i) = e^{-\varkappa i\beta_t}a_t = t^i(1-\cos\varphi_t)\beta_t,$$

так что

$$-F'(x_i) = t^i(v_i + (1 - \cos\varphi_t)\beta_t e_3) = t^i r_i = \bar{g}_i.$$

Далее,

$$F''(x) = e^{-\varkappa y}\left(\begin{array}{c:c} \psi_t''(\bar{x}) & -\varkappa\psi_t'(\bar{x}) \\ \hdashline -\varkappa\psi_t'(\bar{x}) & a_t\varkappa + \varkappa^2(\psi_t(\bar{x}) - \psi_t) \end{array}\right).$$

Легко видеть, что при всех t, достаточно близких к 1 (точнее, при $1 - \bar{\Delta} \leqslant t < 1$, $\bar{\Delta} > 0$), в области $G_t = \{|y| \leqslant 1/\varkappa(t)$ и $|\bar{x}| \leqslant 20\}$ функция F $(c_1\varkappa(t), c_2/\varkappa(t))$-сильно выпукла, c_1, c_2 — абсолютные константы. Легко видеть, далее, что ее можно продолжить из этой области до $(c_1\varkappa(t), c_2/\varkappa(t))$-сильно выпуклой функции на всем R^3. Обозначим это продолжение через $F_t(x)$. Заметим, что в области G_t лежат все точки рассматриваемой траектории $x_0, \ldots, x_{M(t)}$, где

$$M(t) = \left[\frac{1}{\beta_t\varkappa(t)}\right] = \left[\ln\frac{1}{t}\right]. \tag{3.7}$$

Функция $F_t(x)$ — «почти» требуемый пример. Единственное, что надо еще сделать — скорректировать $F_t(x)$ в окрестности x_0 так, чтобы у полученной функции $\bar{F}_t(x)$ было

$$\left.\begin{array}{r} -\bar{F}_t'(x_0) = \lambda\Delta x_0, \quad \lambda > 0 \\ \langle\bar{g}_1 - \lambda\Delta x_0 \mid \Delta x_1\rangle = 0. \end{array}\right\} \tag{3.8}$$

и

Второе условие дает

$$\lambda = \lambda_t = \frac{\langle\bar{g}_1 \mid \Delta x_1\rangle}{\langle\Delta x_0 \mid \Delta x_1\rangle} = \frac{t\langle r_1 \mid \Delta x_1\rangle}{(1 + \beta_t^2)\cos\varphi_t} =$$

$$= \frac{t}{(\cos\varphi_t(1 + \beta_t^2))}\langle q_1 - q_0\cos\varphi_t + \beta_t(1 - \cos\varphi_t)e_3 \mid q_1 + \beta_t e_3\rangle =$$

$$= \frac{t}{(\cos\varphi_t(1 + \beta_t^2))}\{1 - \cos\varphi_t\cos\theta + \beta_t^2(1 - \cos\varphi_t)\}.$$

Будем искать $\bar{F}_t(x)$ в виде $F_t(x) + \Delta(x)$, отыскивая выпуклую функцию $\Delta(x)$ так, чтобы удовлетворялось (3.8) и при этом $\Delta(x)$ была бы равна 0 в окрестности $x_1, x_2, \ldots$ Если нам удастся это сделать, то получим в силу нашего построения (см. (A) — (B)) $\bar{F}_t'(x_i) = -\bar{g}_i$, $M(t) \geqslant i \geqslant 0$, где $\bar{g}_i = g_i$, $i \geqslant 1$, и $g_0 = \lambda_t\Delta x_0$, причем g_i и x_i удовлетворяют (i) — (iii) при $i \leqslant M(t)$. Таким образом, $x_0, \ldots, x_{M(t)}$ будет начальным отрезком траектории ПР на $\bar{F}_t$.

Чтобы удовлетворить (3.8), мы должны иметь

$$-\Delta'(x_0) + \bar{g}_0 = \lambda_t\Delta x_0,$$

т. е.

$$\Delta'(x_0) = \bar{g}_0 - \lambda_t\Delta x_0 \equiv p(t). \tag{3.9}$$

Чтобы построить удовлетворяющую (3.9) и условию $\Delta'(x_i) = 0$, $0 = \Delta(x_i)$, $i \geqslant 1$, выпуклую функцию $\Delta(x)$, достаточно проверить, что при некотором $a > 0$ (a не зависит от t хотя бы для достаточно близких к 1 значений t) имеем $\langle p(t) \mid x_i - x_0\rangle \leqslant -a$. Достаточно показать, что $\langle p(t) \mid \Delta x_i\rangle \leqslant a_i$ (a_i не зависят от t), причем

$$a_0 < 0, \quad a_0 + a_1 < 0, \quad a_0 + a_1 + a_2 < 0,$$
$$a_0 + a_1 + a_2 + a_3 < 0, \quad a_0 + a_1 + a_2 + a_3 + a_4 < 0$$

(поскольку Δx_i периодичны: $\Delta x_{i+5} = \Delta x_i$). Если нам удастся это сделать, то при t, близких к 1, мы сможем указать удовлетворяющую (3.9) $\bar{c}$-гладкую выпуклую функцию, равную 0 в окрестности x_i, $i \geqslant 1$ (можно будет взять ее в виде $\varphi(\langle p(t) \mid x - x_0\rangle)$ с подходящим φ, см. § 2). Здесь $\bar{c}$ — абсолютная константа.

Прямое вычисление величин $\langle p(1) \mid \Delta x_i\rangle$ дает $\langle p(1) \mid \Delta x_i\rangle \leqslant 0$ и $\langle p(1) \mid \Delta x_0\rangle < 0$. Стало быть, при достаточно близких к 1 значениях t нужные a_i найдутся.

3.6. Суммируем полученные результаты. Мы нашли такое $\Delta > 0$ и такие абсолютные константы $c_0, c_1 > 0$, что для всех t, $1 > t > 1 - \Delta$, существует $(c_0\varkappa(t), c_1/\varkappa(t))$-сильно выпуклая функция $\bar{F}_t$ на R^3, удовлетворяющая следующему условию: вдоль первых $[\ln 1/t] \equiv M(t)$ точек x_i, $i \geqslant 1$, траектории ПР на $\bar{F}_t$ выполнено соотношение

$$\bar{F}_t(x_i) = \frac{a_t}{\varkappa(t)}e^{-\varkappa(t)\langle x_i \mid e_3\rangle}.$$

Здесь $\varkappa(t)$ определяется в п. 3.5, $a_t = \beta_t(1 - \cos\varphi_t)$, β_t, φ_t определяются из (3.5). Мы видим, что

$$\nu(x_{[M(t)/2]}, F_t) \geqslant \frac{\bar{F}_t(x_{[M(t)/2]}) - \bar{F}_t(x_{[M(t)]})}{\bar{F}_t(x_0) - \bar{F}_t(x_{[M(t)]})} \geqslant \nu_0,$$

$\nu_0 > 0$ не зависит от t. Кроме того, ясно, что

$$\underline{c} \leqslant \frac{\varkappa(t)}{(1 - t)} \leqslant \bar{c},$$

$\underline{c}, \bar{c} > 0$ — абсолютные константы. Следовательно, модуль сильной выпуклости $\bar{F}_t$, равно как и $M(t)$,— порядка $1/(1 - t)$, так что за время порядка модуля сильной выпуклости $\bar{F}_t$ метод ПР не в состоянии обеспечить точность ν_0 в решении $\bar{F}_t$.

В то же время указанный модуль сильной выпуклости может быть сделан каким угодно большим за счет малости $1 - t$. Таким образом, при всех достаточно больших Q существует (l_Q, Q)-сильно выпуклая функция, на которой ПР с $\bar{c}Q$ шагами ($\bar{c} > 0$ — абсолютная константа) не достигает точности $\nu_0 > 0$, ν_0 не зависит от Q.

Остается заметить, что ПР работает на функции λf в точности так же, как и на функции f ($\lambda > 0$), так что указанную «плохую» функцию можно взять (l, Q)-сильно выпуклой, каково бы ни было $l > 0$. Построение закончено.

§ 4. Метод проекций Зойтендейка *)

В этом методе $p_0 = g_0$ и p_i есть проекция g_i на ортогональное дополнение к $\mathfrak{L}(\{\Delta g_s\}_{s=1}^{i-1})$, $\Delta g_i = g_{i+1} - g_i$. Если получаем $p_i = 0$, но $g_i \neq 0$, то в момент i происходит восстановление. Покажем, что этот метод сходится *только за счет восстановлений* (имеются в виду обычные восстановления — через каждые $\dim F$ шагов). Более точно, мы построим $\bar{Q}$-сильно выпуклую функцию f на пространстве $\mathfrak{l}_2^{(\infty)}$, на которой метод Зойтендейка $\mathfrak{Z}$ вовсе не сходится. С нашей точки зрения это означает, что метод Зойтендейка не только не реализует оптимальную оценку трудоемкости решения сильно выпуклых задач, но и вообще малопригоден для решения задач большой размерности.

Перейдем к конструкции примера. Пусть $\{e_i\}_{i=1}^{\infty}$ — ортобазис в $\mathfrak{l}_2^{(\infty)}$. Положим

$$x_r = \sum_{j=r+1}^{\infty} 2^{-\frac{j}{2}}(j-r)\,e_j, \quad q_r = -g_r = \sum_{j=r+1}^{\infty} 2^{-\frac{j}{2}} e_j.$$

4.1. Убедимся, что существует $\bar{Q}$-сильно выпуклая функция $\bar{f}$ на $\mathfrak{l}_2^{\infty}$ ($\bar{Q}$ — некоторая константа), такая, что $\bar{f}'(x_r) = -g_r$, $0 \leqslant r < \infty$. Для этого будем искать выпуклую функцию $\xi(x)$ с липшицевой производной так, чтобы было $\xi'(x_r) = tq_r - x_r$, $t > 0$. Если удастся найти такое ξ для какого-нибудь $t > 0$, то можно будет взять

$$\bar{f}(x) = \frac{1}{t}\xi(x) + \frac{x^2}{2t}.$$

4.2. Перейдем к построению ξ. Фиксируем $t > 3$ и положим

$$\bar{q}_i = tq_i - x_i = \sum_{j=i+1}^{\infty} 2^{-j/2}(t-j+i)\,e_j.$$

Вычислим числа $t_{ij} = \langle \bar{q}_i - \bar{q}_{i+1} \mid x_j \rangle$, $0 \leqslant i, j < \infty$.

1) Пусть $i + 1 \leqslant j$. Имеем

$$\bar{q} - \bar{q}_{i+1} = t \cdot 2^{-(i+1)/2} e_{i+1} - \sum_{s=i+2}^{\infty} 2^{-s/2} e_s,$$

$$\langle \bar{q}_i - \bar{q}_{i+1} \mid x_j \rangle = -\sum_{s=j+1}^{\infty} 2^{-s}(s-j) = -2^{-j+1}.$$

*) Описание метода Зойтендейка взято из [27], гл. 3.

Итак, при $i + 1 \leqslant j$ имеем

$$t_{ij} = -2^{-j+1}.$$

2) Пусть теперь $i + 1 > j$, т. е. $i \geqslant j$. Тогда

$$\begin{aligned}\langle \bar{q}_i - \bar{q}_{i+1} \mid x_j \rangle &= \\ &= t \cdot 2^{-(i+1)}(i-j+1) - \sum_{s=i+2}^{\infty} 2^{-s}(s-j) = \\ &= t(i-j+1)\,2^{-(i+1)} - 2^{-(i+1)}(i-j+3) = \\ &= 2^{-(i+1)}((t-1)(i-j)+t-3).\end{aligned}$$

Таким образом, при $i \geqslant j$

$$t_{ij} = ((t-1)(i-j) + (t-3))\,2^{-(i+1)}.$$

Поскольку $t > 3$, то t_{ij} убывает при фиксированном i и изменении j от 0 до i:

$$t_{ii} = (t-3)\,2^{-(i+1)} = \min_{j \leqslant i} t_{ij}.$$

Положим $t_i \equiv t_{ii} = (t-3)\,2^{-(i+1)}$.

4.3. Пусть теперь $\varphi_i(\tau)$ — выпуклая функция, равная 0 при $\tau \leqslant 0$ и совпадающая с τ при $\tau \geqslant (t-3)\,2^{-(i+1)} = t_i$. Можно считать $\varphi_i(\tau)$ бесконечно дифференцируемой и такой, что

$$|\varphi_i'(\tau)| \leqslant c, \quad 0 \leqslant \varphi_i''(\tau) \leqslant L_i = c(t)\,2^i$$

(действительно, можно взять

$$\varphi_i(\tau) = \frac{t_i}{t_0}\varphi_0\left(\tau\frac{t_0}{t_i}\right)\right).$$

Составим ряд

$$\sum_{i=0}^{\infty} \varphi_i(\langle \bar{q}_i - \bar{q}_{i+1} \mid x \rangle) \equiv \sum_{i=0}^{\infty} \psi_i(x).$$

Функции ψ_i выпуклы, бесконечно дифференцируемы, $\psi_i(0) = 0$, $\psi_i'(0) = 0$. Имеем

$$\langle \psi_i''(x)\,h \mid h \rangle = \varphi_i''(\langle \bar{q}_i - \bar{q}_{i+1} \mid x \rangle) \sum_{k,\,l=1}^{\infty} (\bar{q}_i - \bar{q}_{i+1})_k (\bar{q}_i - \bar{q}_{i+1})_l h_k h_l.$$

Отсюда

$$\langle \psi_i''(x)\,h \mid h \rangle \leqslant L_i t^2 \sum_{k,\,l=i+1}^{\infty} 2^{-\frac{k+l}{2}} |h_k|\,|h_l|.$$

Следовательно,

$$\sum_{i=0}^{N}|\langle\psi_i''(x)h\,|\,h\rangle|\leqslant c_1(t)\sum_{i=0}^{N}\sum_{k,\,l=i+1}^{\infty}2^{i-\frac{k+l}{2}}|h_k||h_l|\leqslant$$
$$\leqslant c_1(t)\sum_{k,\,l=1}^{\infty}|h_k||h_l|\Big(\sum_{i=0}^{\min(k,\,l)}2^i\Big)2^{-\frac{k+l}{2}}\leqslant$$
$$\leqslant c_2(t)\sum_{k,\,l=1}^{\infty}|h_k||h_l|\cdot 2^{-\frac{|k-l|}{2}}\leqslant c_3(t)|h|^2.$$

Таким образом, ряд из вторых производных функций ψ_i вдоль любого направления абсолютно сходится и имеет ограниченную во всем пространстве сумму. Отсюда и из соотношений $\psi_i(0)=0$, $\psi_i'(0)=0$ следует, что ряды $\sum\limits_{i=0}^{\infty}\psi_i(x+th)$, $\sum\limits_{i=0}^{\infty}\langle\psi_i'(x+th)\,|\,h\rangle$ сходятся равномерно на любом отрезке вида $|t|\leqslant\Delta$. Поэтому функция

$$\psi(x)=\sum_{i=0}^{\infty}\psi_i(x)$$

определена, выпукла, непрерывна и при всех x, h

$$\langle\psi'(x)\,|\,h\rangle=\sum_{i=0}^{\infty}\langle\psi_i'(x)\,|\,h\rangle,$$

причем $\psi'(x)$ липшицева с константой $c_3(t)$.

В то же время

$$\langle\psi'(x_j)\,|\,h\rangle=\sum_{i=0}^{\infty}\langle\psi_i'(x_j)\,|\,h\rangle=$$
$$=\sum_{i=0}^{\infty}\varphi_i'(t_{ij})(\bar q_i-\bar q_{i+1})=\sum_{i=j}^{\infty}(\bar q_i-\bar q_{i+1})=\bar q_j.$$

Здесь мы учли, что $t_{ij}<0$ при $i<j$ и $t_{ij}\geqslant t_i$ при $i\geqslant j$, так что

$$\varphi_i'(t_{ij})=\begin{cases}0, & i<j,\\ 1, & i\geqslant j.\end{cases}$$

Итак, можно взять $t=4$ и соответствующее $\psi(x)$ принять за $\xi(x)$.

4.4. Таким образом, функция $\bar f(x)$, заказанная в п. 4.1, действительно существует. Положим $\mu_j=2^{-j/2}$ и $\bar h=\sum\limits_{j=1}^{\infty}\mu_j e_j$. Рассмотрим функцию $f(x)=-\langle\bar h\,|\,x\rangle+2\bar f(x)$. Убедимся, что $\{x_i\}$ — траектория $\mathfrak{Z}$ на f с начальной точкой x_0. Действительно, требуется доказать, что x_1-x_0 коллинеарно $f'(x_0)$, а $x_{i+1}-x_i$ коллинеарно проекции $f'(x_i)$ на ортогональное дополнение к $\mathfrak{L}(f'(x_1)-f'(x_0),\ \ldots,\ f'(x_i)-f'(x_{i-1}))$, тогда как при этом $\langle f'(x_{i+1})\,|\,x_{i+1}-x_i\rangle=0$.

Имеем

$$f'(x_i)=-\sum_{j=1}^{\infty}\mu_je_j+2\sum_{j=i+1}^{\infty}\mu_je_j=-\sum_{j=1}^{i}\mu_je_j+\sum_{j=i+1}^{\infty}\mu_je_j,$$
$$x_{i+1}-x_i=-\sum_{j=i+2}\mu_je_j-\mu_{i+1}e_{i+1}=-\sum_{j=i+1}^{\infty}\mu_je_j.$$

Полагая $i=0$, находим, что $f'(x_0)$ коллинеарно Δx_0. Далее,

$$\langle f'(x_{i+1})\,|\,x_{i+1}-x_i\rangle=\Big\langle-\mu_{i+1}e_{i+1}+\sum_{j=i+2}^{\infty}\mu_je_j\,\Big|-$$
$$-\sum_{j=i+2}^{\infty}\mu_je_j-\mu_{i+1}e_{i+1}\Big\rangle=\Big\{-\mu_{i+1}^2+\sum_{j=i+2}^{\infty}\mu_j^2\Big\}=$$
$$=\Big\{-2^{-(i+1)}+\sum_{j=i+2}^{\infty}2^{-j}\Big\}=0.$$

Наконец,

$$\Delta f_i\equiv f'(x_{i+1})-f'(x_i)=2\Big(\sum_{j=i+2}^{\infty}\mu_je_j-\sum_{j=i+1}^{\infty}\mu_je_j\Big)=-2\mu_{i+1}e_{i+1}.$$

Стало быть, $\mathfrak{L}(\Delta f_0,\ \ldots,\ \Delta f_{i-1})=\mathfrak{L}(e_1,\ \ldots,\ e_i)$, и проекция $f'(x_i)$ на ортогональное дополнение к $\mathfrak{L}(\Delta f_0,\ \ldots,\ \Delta f_{i-1})$ есть $\sum\limits_{j=i+1}^{\infty}\mu_je_j$, что коллинеарно $x_{i+1}-x_i$. Итак, $\{x_i\}$ — траектория $\mathfrak{Z}$ на f. Но

$$f'(x_i)=-\sum_{j=1}^{i}\mu_je_j+\sum_{j=i+1}^{\infty}\mu_je_j\xrightarrow[i\to\infty]{}-\bar h\neq 0,$$

так что метод не сходится даже и по функционалу. Требуемый пример построен.

§ 5. Замечания о методе Давидона—Флетчера—Пауэлла *)

5.1. Сделаем несколько замечаний еще об одном методе сопряженных градиентов — *методе Давидона — Флетчера — Пауэлла* (другое название — *метод переменной метрики*). В этом методе (для краткости назовем его ДФП) очередные направления спуска p_i строятся по формуле $p_i=H_ig_i$, где H_i — оператор, строящийся

*) Описание метода Давидона — Флетчера — Пауэлла взято из [25], гл. II.

по рекуррентным формулам:

$$H_0 = I \text{ (единичный оператор)},$$

$$H_{i+1} = H_i - \frac{H_i\Delta g_i \cdot H_i\Delta g_i}{\langle H_i\Delta g_i \mid \Delta g_i\rangle} + \frac{\Delta x_i \cdot \Delta x_i}{\langle \Delta x_i \mid \Delta g_i\rangle}. \tag{5.1}$$

Здесь $\Delta g_i = g_{i+1} - g_i$, а обозначение $u \cdot v$ (u, v — векторы) означает оператор $(u \cdot v)\, h = \langle v \mid h\rangle\, u$.

Метод ДФП считается наиболее эффективным методом сопряженных градиентов. Мы не сможем выяснить до конца — реализует ли он на самом деле «оптимальную» оценку трудоемкости на классе сильно выпуклых задач.

Единственное, что можно здесь утверждать, это то, что если ДФП и хорош, то заведомо лишь при определенных нормировках решаемых задач. Более точно, существуют такое $\overline{Q}$ и такие сильно выпуклые функции с модулем выпуклости $\overline{Q}$ от достаточно большого числа переменных, на которых ДФП работает «как угодно плохо» (т. е. не достигает точности $\nu_0 > 0$ за любое наперед заданное число шагов M; $\nu_0 > 0$, $\overline{Q}$ — абсолютные константы).

Дело вот в чем: в отличие от рассмотренных выше методов, ДФП «чувствует» изменение масштаба. Траектории ДФП на задаче f и на задаче λf, вообще говоря, различны и при общей исходной точке. Последнее связано с наличием третьего слагаемого в правой части (5.1). Ясно, что траектория ДФП на λf совпадает *) с траекторией на f метода ДФП$_\lambda$ с правилами

$$p_i = H_i g_i, \quad H_0 = I,$$

$$H_{i+1} = H_i - \frac{H_i\Delta g_i \cdot H_i\Delta g_i}{\langle H_i\Delta g_i \mid \Delta g_i\rangle} + \frac{1}{\lambda}\,\frac{\Delta x_i \cdot \Delta x_i}{\langle \Delta x_i \mid \Delta g_i\rangle}.$$

Если теперь фиксировать f и рассматривать траектории ДФП на λf при больших λ, то они будут «почти такими же» **), что и траектории на f метода ДФП$_\infty$ с правилами

$$p_i = H_i g_i, \quad H_0 = I,$$

$$H_{i+1} = H_i - \frac{H_i\Delta g_i \cdot H_i\Delta g_i}{\langle H_i\Delta g_i \mid \Delta g_i\rangle}.$$

Но легко видеть, что «предельный случай ДФП» — метод ДФП$_\infty$ — есть в точности метод проекций Зойтендейка (почему?). Для последнего мы умеем строить пример функции f со «сколь угодно плохой» сходимостью и модулем сильной выпуклости $\overline{Q}$. Функция λf при достаточно большом λ будет нужным примером для ДФП.

*) Начальная точка траекторий фиксирована раз и навсегда.

**) Это высказывание не вполне корректно, однако смысл его ясен. Последующие выводы легко сделать строгими, но мы не будем этим заниматься.

Это замечание, конечно, не должно интерпретироваться как критика метода ДФП. Единственное, что мы установили,— что эффективное применение ДФП требует предварительной нормировки решаемой задачи. ДФП «чувствует» не только отношение $Q = L/l$ параметров сильной выпуклости решаемой (l, Q)-сильно выпуклой задачи, но и абсолютные значения этих параметров. Было бы интересно выяснить, какая нормировка решаемой задачи «оптимальна» для ДФП и верно ли, что при такой нормировке ДФП реализует «оптимальную» оценку трудоемкости решения задач данного модуля сильной выпуклости. Было бы еще интересней выяснить, какие из стандартных методов сильно выпуклого программирования реализуют указанную оценку (и вообще есть ли такие методы). Вряд ли стоит оговаривать, что ответы на эти вопросы авторам неизвестны.

5.2. Заканчивая наш краткий и неполный анализ стандартных методов сильно выпуклого программирования, заметим, что полученные результаты характеризуют методы не всесторонне, а лишь с определенной, выбранной нами точки зрения. Правда, эта точка зрения представляется нам наиболее естественной, во всяком случае, в математическом отношении. Следует, однако, учесть, что мы вовсе не касались таких практически важных сторон методов, как простоты их вычислительной организации и вычислительной устойчивости. Наконец, стандартные методы для своего запуска вовсе не требуют априорного знания модуля выпуклости минимизируемой функции, по крайней мере явного (фактически некоторые его оценки при аккуратном применении методов все равно необходимы: надо правильно выбрать точность решения одномерных задач и т. п.). В этом, конечно, достаточно заметное практическое преимущество стандартных методов над методом § 3 гл. VII (оно, впрочем, до известной степени гасится заключительными замечаниями § 5 гл. VII).

Глава IX

МЕТОДЫ ВЫПУКЛОГО ПРОГРАММИРОВАНИЯ НУЛЕВОГО ПОРЯДКА

В предыдущих главах мы занимались методами выпуклого программирования первого порядка — т. е. методами, использующими не только информацию о значениях функционалов решаемой задачи, но и информацию об их производных. В этой главе мы рассмотрим ситуацию, в которой информация о производных недоступна. С практической точки зрения эта ситуация, по-видимому, более типична. Вместе с тем она объективно более сложна и изучена в существенно меньшей степени, чем рассматривавшаяся ранее. По этой причине наше изложение будет более беглым, а результаты — менее законченными.

План изложения таков. Первые два параграфа посвящены методам, использующим детерминированный оракул нулевого порядка. В § 1 рассматривается возможность построения на его основе оракула первого порядка с последующим применением методов первого порядка. Основной недостаток получаемых на таком пути методов — низкая их устойчивость по отношению к ошибкам оракула. В § 2 предлагается теоретически более устойчивый метод решения выпуклых задач при оракуле нулевого порядка, основанный на идее ММЦТ. В §§ 3—4 рассматривается случай, когда используется стохастический оракул нулевого порядка. В § 3 на его основе строится стохастический оракул первого порядка и рассматривается возможность редукции проблемы к задачам стохастического программирования из глав V—VI. Полученные в итоге методы имеют трудоемкость, приемлемо зависящую от размерности задачи, но слишком быстро растущую с увеличением требований к точности решения. В § 4 на основании стохастического оракула нулевого порядка строится, грубо говоря, детерминированный оракул нулевого порядка, используемый далее применительно к методу § 2. Получаемый метод имеет трудоемкость, приемлемо зависящую от требуемой точности решения, но быстро растущую с ростом размерности задачи. Заключительная часть § 4 посвящена обсуждению результатов главы.

Мы коснемся также вопроса о влиянии смещенности оракула на возможность применения описываемых методов.

§ 1. Методы нулевого порядка: детерминированный оракул. I

Пусть $\mathfrak{A}$ — поле задач $f = (f_0, \ldots, f_m)$ какого-нибудь из рассмотренных выше классов задач выпуклого программирования. Будем считать, что $\mathfrak{A}$ снабжено соответствующим правилом измерения погрешности приближенных решений (т. е. нормирующим отображением $f \to r(f)$). Рассмотрим класс задач $\overline{\mathfrak{A}}$, получающийся оснащением $(\mathfrak{A}, r(\cdot))$ оракулом $\mathcal{O}$ с наблюдательной функцией $\psi(f, x) = f(x)$, $x \in G$. Здесь $G \subset E$ — область определения задач из $\mathfrak{A}$. Предположим, что E — конечномерное пространство. Обозначим через n его размерность. Наша цель — построить эффективные методы решения задач класса $\overline{\mathfrak{A}}$.

1.1. Первое, что приходит в голову,— это построить по данному точному оракулу нулевого порядка оракул первого порядка (уже приближенный). Идея проста: мы уже умеем эффективно решать задачи класса $\widehat{\mathfrak{A}}$, получающиеся из $\overline{\mathfrak{A}}$ заменой точного оракула нулевого порядка точным же оракулом первого порядка. Иначе говоря, если бы мы умели вычислять не только значения, но и производные компонент f_j решаемой задачи $f \in \mathfrak{A}$, мы знали бы, как ее следует решать. Естественно попытаться построить приближенные оценки производных f_j по их значениям. Самое простое — заменить эти производные разностными отношениями.

Именно, если требуется построить приближение $\bar{f}'_j(x)$ к производной f_j в точке $x \in \operatorname{int} G$, то можно взять его в виде

$$\bar{f}'_j(x) = \sum_{i=1}^{n} \frac{f_j(x + t_x e_i) - f_j(x)}{t_x} \varphi_j, \tag{1.1}$$

где $e_1, \ldots, e_n$ — какой-нибудь базис в E, φ_j — линейная форма на E, такая, что

$$\langle \varphi_j \mid e_i \rangle = \begin{cases} 0, & i \neq j, \\ 1, & i = j, \end{cases}$$

а $t_x \neq 0$ — шаг разностного отношения — достаточно мал по абсолютной величине. Можно, разумеется, использовать и другие формы разностных аппроксимаций производных.

При использовании таких аппроксимаций имитация одного ответа оракула первого порядка «стоит» $n + 1$ вопроса к оракулу нулевого порядка. Можно, таким образом, надеяться на возможность использовать для решения задач класса $\overline{\mathfrak{A}}$ методы решения задач класса $\widehat{\mathfrak{A}}$ — методы первого порядка, увеличив при этом их трудоемкость в $n + 1$ раз.

1.2. Такого рода подход сталкивается с тремя проблемами. Первые две связаны с самой его возможностью, а третья — с его эффективностью. Проблемы эти состоят в следующем.

I. Каким бы малым мы ни брали шаг разностного отношения, $\tilde{f}'_j(x)$ всегда будет лучшим или худшим, но — приближением к $f'_j(x)$. Стало быть, можно эффективно применить метод первого порядка лишь тогда, когда он устойчив по отношению к ошибкам в поставляемой ему информации первого порядка — при достаточно малых ошибках в этой информации он по-прежнему хорошо решает задачу.

II. Предположим, что желательно применять устойчивый в только что описанном смысле метод первого порядка, и требуется обеспечить точность ν в решении задачи. Предположим, что мы в состоянии, анализируя метод, узнать, какие ошибки в оценке производных при этом допустимы. Тогда требуется так выбирать очередные смещения t_{x_i}, чтобы обеспечить эти, а не бóльшие ошибки в оценке производных. Более того, такие t_{x_i} надо уметь выбирать «конструктивно», т. е. на основании имеющейся к данному шагу информации. Если такого способа выбора t_{x_i} нет, то неясно, как реализовать рассматриваемую редукцию методов нулевого порядка к методам первого порядка.

III. Предположим, что с проблемами I и II мы справились, так что намеченный план реализован. Предположим, что он применяется к эффективному (реализующему сложность класса $\widehat{\mathfrak{A}}$) методу первого порядка. Будет ли полученный метод нулевого порядка реализовывать (или почти реализовывать) сложность класса $\mathfrak{A}$? Вообще говоря, неясно, почему это должно быть так.

Рассмотрим в общих чертах каждую из этих проблем. Начнем с первой. Конечно, вообще говоря, метод первого порядка не обязан быть устойчив по отношению к ошибкам в информации. Можно представить себе метод, хорошо работающий лишь при абсолютно точной информации. Однако со всеми конкретными методами, которые нам известны, в частности, со всеми методами первого порядка, рассмотренными выше, дело в этом смысле обстоит благополучно *), по крайней мере, с теоретической точки зрения. Ни один из них не нуждается в абсолютно точной информации. Более того, для каждого из них простой анализ обоснования сходимости позволяет конструктивно указать ту точность в поступающей информации, при которой применение метода, использующего приближенную информацию, приводит к результату требуемой точности за то же (по порядку) время, что и в случае точной информации. Этот допустимый для метода (и ненулевой) уровень ошибок в текущей информации может быть явно вычислен в функции от априорной информации (т. е. от класса задач, подлежащих решению) и от требуемой точности решения.

*) Исключение составляют классы $\boldsymbol{H}_{x_1}$ (но не $\boldsymbol{H}^{V}_{x_1}$).

Иногда, правда, приходится несущественно модифицировать метод, чтобы сделать его помехоустойчивым в указанном смысле. Мы не будем разбирать вопросов устойчивости построенных нами методов первого порядка. Как уже говорилось, это сравнительно легко сделать, анализируя доказательства теорем об их сходимости. Заметим, что для классов общих выпуклых задач (как при детерминированном, так и при стохастическом оракуле) эта работа фактически проделана: соответствующие классы с самого начала снабжались не точным, а приближенным оракулом.

Итак, в интересующих нас случаях первая из намеченных выше проблем решается положительно.

Что касается второй проблемы, то здесь дело обстоит сложнее. Необходимо выяснить, для каких классов задач можно конструктивно выбрать шаг разностного отношения так, чтобы обеспечить заданную точность аппроксимации опорного функционала (эта точность, как объяснено выше, находится из анализа устойчивости соответствующего метода первого порядка).

Предположим, что f дифференцируема в окрестности интересующей нас точки $x \in \operatorname{int} G$. Тогда производная $f'_j(x)$ есть

$$\lim_{t\to+0} \sum_{i=1}^{n} \frac{f_j(x+te_j)-f_j(x)}{t}\varphi_j.$$

Пусть априори известен модуль непрерывности производной f_j, т. е. такая функция $\omega_j(s) \xrightarrow[s\to+0]{} 0$, что

$$\| f'_j(x) - f'_j(y) \|_* \leqslant \omega_j(\| x - y \|)$$

при $x, y \in \operatorname{int} G$ (здесь $\|\cdot\|$ — какая-нибудь норма в соответствующем пространстве). Ясно, что

$$\left| \frac{f_j(x+te_i)-f_j(x)}{t} - \frac{d}{dt}\Big|_{t=0} f_j(x+te_i) \right| \leqslant \omega_j(t\,\|e_i\|)\|e_i\|,$$

так что можно явно указать то значение шага t_x, для которого разностное отношение с требуемой точностью аппроксимирует производную. Такова, в частности, ситуация для классов гладких (сильно выпуклых) выпуклых задач. Там $f'_j(x)$ липшицевы с априори известными (входящими в описание класса) константами. Таким образом, для этих двух классов вторая из намеченных выше проблем также решается положительно.

Иначе обстоит дело с классами общих выпуклых задач. Здесь никакой априорной информации о модуле непрерывности производных нет (да и быть не может — класс содержит задачи с разрывной производной). В этой ситуации никакой выбор шагов разностного отношения не гарантирует получения достаточно точной оценки производной. В этом легко убедиться, рассмотрев следующий пример. Пусть E есть плоскость с ортобазисом (e_1, e_2), а

$f_0(x)$ — выпуклая функция, одна из линий уровня которой — граница квадрата. При вычислении производной f_0 в одной из его вершин по формуле (1.1) при всех положительных значениях шага t_x оценка производной оказывается нулевой, хотя все опорные функционалы к f_0 в 0 могут быть отличны от 0.

Таким образом, непосредственный ответ на интересующий нас вопрос для классов общих выпуклых задач отрицателен, так что на первый взгляд решение этих задач при оракуле нулевого порядка нельзя редуцировать к их решению с помощью методов первого порядка. Оказывается, однако, что теоретически эта трудность легко преодолима — явления, подобные описанному выше, могут происходить лишь в некоторых «исключительных» точках x, и простая рандомизация метода позволяет «почти наверняка» обойти эти точки. Подробное рассмотрение этого вопроса будет дано ниже.

Теперь скажем несколько слов о последней из сформулированных проблем. Для полного исследования вопроса об эффективности получаемых намеченным путем методов нулевого порядка следовало бы сравнить оценки трудоемкости этих методов с нижними (по возможности точными) оценками сложности класса задач $\overline{\mathfrak{A}}$. Некоторые качественные результаты, однако, можно получить и без громоздкого анализа сложности $\overline{\mathfrak{A}}$.

Проведем соответствующий анализ для детерминированных методов и сложностей. Аналогичные рассмотрения можно провести и для стохастических методов и соответствующих им стохастических сложностей. Пусть, как и выше, $\widehat{\mathfrak{A}}$ — класс задач, получающийся из $\overline{\mathfrak{A}}$ заменой точного оракула нулевого порядка точным оракулом первого порядка. Обозначим через $N_{\overline{\mathfrak{A}}}(\nu)$, $N_{\widehat{\mathfrak{A}}}(\nu)$ детерминированные сложностные функции этих классов. Пусть $\{\mathcal{B}(\nu)\}_{\nu>0}$ — семейство методов решения задач класса $\widehat{\mathfrak{A}}$, причем $\mathcal{B}(\nu)$ решает эти задачи с точностью ν. Предположим, что трудоемкость $\mathcal{B}(\nu)$ есть $l(\nu)$. Рассмотрим *функцию эффективности* рассматриваемого семейства — отношение $\theta(\nu) = N_{\widehat{\mathfrak{A}}}(\nu)/l(\nu)$ (чем оно ближе к 1, тем «лучше» семейство $\{\mathcal{B}(\nu)\}$). Предположим, далее, что намеченная схема имитации оракула первого порядка оракулом нулевого в применении к методам семейства $\{\mathcal{B}(\nu)\}$ реализуема и приводит к семейству $\overline{\mathfrak{A}}$-методов $\{\overline{\mathcal{B}}(\nu)\}$, причем $\overline{\mathcal{B}}(\nu)$ решает задачи из $\overline{\mathfrak{A}}$ с точностью ν. Трудоемкость $\overline{\mathcal{B}}(\nu)$ есть в силу принятой конструкции $(n+1)\,l(\nu)$, а сложность $N_{\overline{\mathfrak{A}}}(\nu)$, очевидно, $\geqslant N_{\widehat{\mathfrak{A}}}(\nu)$. Стало быть, эффективность рассматриваемого семейства $\overline{\mathfrak{A}}$-методов $\overline{\theta}(\nu)$ удовлетворяет неравенству

$$\overline{\theta}(\nu) \geqslant \frac{\theta(\nu)}{n+1}.$$

Итак, предлагаемая конструкция при условии своей применимости снижает эффективность разве что в $n+1$ раз. Таким образом, с точностью до множителей порядка $O(n)$ рассматриваемая конструкция сохраняет эффективность методов. В частности, если семейство $\{\mathcal{B}(\nu)\}$ асимптотически по $\nu \to 0$ субоптимально (т. е. $\theta(\nu) > \theta_0 > 0$ при всех $\nu > 0$), то таково же и семейство $\{\overline{\mathcal{B}}(\nu)\}$. Вообще, «степень неоптимальности» $\overline{\mathcal{B}}(\nu)$ может разве что в $n+1$ раз быть бóльшей, чем «степень неоптимальности» $\mathcal{B}(\nu)$.

При анализе поведения методов в условиях большой размерности (скажем, при исследовании их эффективности в асимптотике по n) множитель $O(n)$ уже существен, особенно в ситуации, когда его появление приводит к качественным эффектам: трудоемкость методов, прежде не зависевшая от размерности, теперь начинает линейно расти с ростом размерности (такие ситуации возникают при рассмотрении классов гладких выпуклых (сильно выпуклых) задач и общих выпуклых задач на $\mathfrak{l}_p$-телах большой размерности). Можно, впрочем, показать, что это явление неизбежно, так как сама сложность $N_{\overline{\mathfrak{A}}}(\nu)$ во всех естественных ситуациях растет с ростом размерности, по крайней мере, линейно, в отличие от сложности $N_{\widehat{\mathfrak{A}}}(\nu)$, для которой это не всегда так.

Упражнение 1. Рассмотрите класс задач вида

$$f(x) = \langle \varphi | x \rangle \to \min | x \in V \subset E^n, \quad V = \{x \,|\, \|x\| \leqslant 1\},$$

порожденных векторами φ с $\|\varphi\| = 1$. Докажите, что детерминированная (и стохастическая) сложности решения задач этого класса с абсолютной погрешностью ε не меньше cn, если $\varepsilon < \varepsilon_0$ (здесь c, $\varepsilon_0 > 0$ — абсолютные константы).

Ввиду сформулированного в упражнении 1 результата, линейный рост сложности выпуклого программирования по отношению к методам нулевого порядка с ростом размерности — явление неизбежное, и рассматриваемая конструкция не связана с особо неприятными (с грубо-качественной точки зрения) потерями в эффективности. В силу сказанного мы не будем специально исследовать сложность выпуклого программирования при оракуле нулевого порядка, довольствуясь оценками эффективности методов с точностью до множителя $O(n)$.

1.3. Перейдем теперь от изложения общих соображений к реализации намеченного плана применительно к классам общих (липшицевых) выпуклых задач (мы уже отмечали, что для классов гладких (и сильно выпуклых) выпуклых задач его реализация не вызывает затруднений). Очевидно, достаточно уметь имитировать оракул первого порядка точности $\nu_0 > 0$ для класса общих (липшицевых) выпуклых задач с помощью оракула нулевого порядка. Достаточно уметь это делать в случае, когда $G \subset R^n$ — выпуклое тело, и требуется имитировать ответ оракула во внутренних точках G. Пусть $f(x)$ — выпуклая непрерывная функция на G, $\nu_0 > 0$, $x_0 \in \operatorname{int} G$, $V_f = \sup_G f - \inf_G f$. Рассмотрим следующую стохастическую процедуру оценки опорного функционала к f в x_0. *)

1°. Пусть Δ — симплекс с центром в x_0, целиком лежащий в G. Вычислим значения f в вершинах $y_0, \ldots, y_n$ этого симплекса и в точке x_0. Построим число $r = r_f = \max_i f(y_i) - f(x_0)$. Очевидно, $r_f \geqslant 0$ и $r \leqslant V_f$. Кроме того, если $r_f = 0$, то f линейна в пределах Δ, и по ее значениям в y_i можно вычислить *точно* ее опорный функционал в x_0 (почему?). Будем теперь считать, что $r > 0$.

2°. Пусть $e_1, \ldots, e_n$ — фиксированный базис в R^n, $x^1, \ldots, x^n$ — координаты x в этом базисе. Для простоты обозначений примем, что $x_0^i = 0$, $1 \leqslant i \leqslant n$. Пусть $T_0 > 0$ столь мало, что куб $K_0 = \{x \mid |x^i| \leqslant 3T_0\}$ содержится в Δ. Выберем $T \leqslant T_0$, и пусть $\bar{x}$ выбирается в кубе $K = \{x \mid |x^i| \leqslant T, 1 \leqslant i \leqslant n\}$ наудачу (в соответствии с равномерным распределением). Подсчитаем разностную аппроксимацию опорного функционала к f в $\bar{x}$ вида

$$\bar{\varphi} = \sum_{1}^{n} \frac{f(\bar{x} + \tau e_i) - f(\bar{x})}{\tau} \varphi_i,$$

где φ_i — линейный функционал, такой, что

$$\langle \varphi_i \mid e_j \rangle = \begin{cases} 0, & i \neq j, \\ 1, & i = j. \end{cases}$$

Примем $\bar{\varphi}$ за доставляемую описанной процедурой оценку опорного функционала к f в x_0. Более точно, будем считать, что ответ имитируемого оракула первого порядка на вопрос об f в точке x_0 есть аффинная функция

$$g(y) = f(\bar{x}) + \langle \bar{\varphi} \mid y - \bar{x} \rangle - \rho,$$

где $\rho > 0$ — параметр процедуры.

Описанная процедура имеет параметры T, τ, $\rho > 0$ и требует $1 + (n + 1) + (n + 1) = 2n + 3$ вычислений функции f. Подчеркнем, что она носит рандомизированный характер, так что функция g — результат процедуры — случайна.

Покажем, что надлежащим выбором параметров T, τ, $\rho > 0$ можно достичь того, что g с заданной, сколь угодно близкой к 1, вероятностью будет удовлетворять условию

$$A(\nu_0): \ g(y) \leqslant f(y), \quad g(x_0) \geqslant f(x_0) - \nu_0 V_f$$

*) Такая процедура впервые предложена А. М. Гупалом (Гупал А. М. Об одном методе минимизации почти-дифференцируемых функций.— Кибернетика, 1977, № 1).

(т. е. смоделированный оракул первого порядка «почти наверняка» доставляет оценки опорного функционала требуемой точности).

Удобно ввести на R^n евклидову структуру, в которой $e_1, \ldots, e_n$ — ортонормированный базис. Соответствующую евклидову норму обозначим $\|\cdot\|$. Пусть d_G — диаметр G в полученной евклидовой метрике.

Предложение. Пусть $\alpha > 0$ и параметры описанной выше процедуры (рассматриваемой при условии $r > 0$) удовлетворяют соотношениям

$$\left.\begin{aligned} T &= \frac{\nu_0 r}{2}\left\{\frac{2n^2 r}{T_0} + \frac{2n\rho}{d_G}\right\}^{-1}, \\ \tau &= \frac{\alpha T T_0 \rho}{3n^2 \sqrt{n}\, d_G r}, \quad \rho = \frac{\nu_0 r}{2}. \end{aligned}\right\} \qquad (1.2)$$

Тогда вероятность невыполнения $A(\nu_0)$ для результата применения этой процедуры к f не превосходит α.

Упражнение 2. Докажите предложение.

Заметим, что процедуру указанного типа можно одновременно применить сразу к нескольким выпуклым функциям $f_0, \ldots, f_m$ и получить «почти достоверные» оценки (нужной точности) их опорных функционалов. При этом $2n + 3$ точки, в которых требуется вычислять значения этих функций, можно выбрать общими для всех этих $m + 1$ функций, а параметры процедуры ρ, T, τ выбрать «в расчете на худший случай» — на наименьшие среди ненулевых величин $r_{f_0}, \ldots, r_{f_m}$. При этом вероятность невыполнения $A(\nu_0)$ хотя бы для одной из функций f (недостоверность имитации) не превзойдет $(m + 1)\alpha$, где α — параметр, по которому выбираются ρ, T, τ в (1.2). Таким образом, надлежащим выбором параметров процедуры можно сделать эту вероятность произвольно малой.

1.4. Ясно, как применить полученный результат для построения рандомизированных методов нулевого порядка решения общих (липшицевых) выпуклых задач на G. Пусть $\nu > 0$ — требуемая (относительная) точность решения, скажем, общих выпуклых задач с m ограничениями на G. Положим $\nu_0 = {}^1/_4\nu$ и обозначим через $\mathcal{B}_\nu$ настроенный на точность $\nu/2$ метод решения задач классов типа $C^{\nu_0}(G, E, m)$ (см. § 2 гл. III) (для простоты, детерминированный); пусть $l(\nu)$ — верхняя оценка его трудоемкости на этом классе. Выберем произвольно малое $\alpha > 0$ и будем имитировать оракул первого порядка относительной точности ν_0 с помощью оракула нулевого порядка, обеспечивая недостоверность имитации, $\leqslant \alpha / l(\nu)$. Рассмотрим метод $\overline{\mathcal{B}}$, получающийся применением $\mathcal{B}(\nu)$ в сочетании с указанным имитатором оракула для

$\mathscr{B}$ (ν). Его трудоемкость будет $\leqslant (2n+3)\, l(\nu)$. Ясно вместе с тем, что с вероятностью $\geqslant 1-\alpha$ доставляемый им результат будет решением задачи относительной погрешности $\leqslant \nu/2$. В частности, при $\alpha = \nu/4$ *средняя* относительная погрешность $\overline{\mathscr{B}}$ на рассматриваемом классе не превзойдет ν при трудоемкости $l(\nu)\,(2n+3)$. Итак, простая рандомизация методов первого порядка позволяет применять их и в случае, когда доступна лишь информация нулевого порядка. Эта рандомизация позволяет избежать осложнений, связанных с возможной негладкостью компонент задачи.

§ 2. Методы нулевого порядка: детерминированный оракул. II

Простые соображения, использованные в предыдущем параграфе, с практической точки зрения малопригодны. Дело в том, что выше оракул нулевого порядка предполагался абсолютно точным (как, впрочем, и все вычисления с поставляемой им информацией). Малые ошибки в вычислении значений компонент задачи могут приводить к колоссальным ошибкам в результирующих оценках производных в силу малости шагов разностного отношения. Последние оцениваются *сверху* величинами порядка $O(\nu_0^2\alpha)$. Для успеха конструкции должно быть $\nu_0 \leqslant \nu$ и $\alpha \leqslant \nu$, т. е. $\tau \leqslant O(\nu^3)$. При этом допустимые ошибки в вычислении разностных отношений порядка ν, т. е. допустимые погрешности во входной информации — порядка $\tau\nu = O(\nu^4)$. Таким образом, изложенная конструкция гарантирует успешное решение лишь при условии, что точность входной информации по порядку значительно выше требуемой точности решения.

2.1. Опишем другой, детерминированный метод нулевого порядка решения общих выпуклых задач. Трудоемкость этого метода по характеру зависимости от точности ν та же, что и у оптимального (в асимптотике при $\nu \to 0$) метода МЦТ. Что же касается до чувствительности трудоемкости к размерности, то она для предлагаемого метода существенно выше, чем для МЦТ (и для рандомизированного метода нулевого порядка, получающегося из МЦТ с помощью описанной в предыдущем пункте конструкции). Таким образом, новый метод заведомо не реализует сложность в асимптотике по размерности. Зато он существенно более помехоустойчив, чем метод из предыдущего пункта, по крайней мере, в формальном понимании помехоустойчивости: допустимые для него ошибки в текущей информации имеют тот же порядок, что и требуемая точность решения задачи (точные формулировки будут даны ниже).

Начнем с описания интересующего нас класса задач. Пусть G — выпуклое замкнутое ограниченное тело в R^n, а $\boldsymbol{C}(G, R^n, m)$, как и выше, в § 2 гл. II,—множество выпуклых непрерывных задач $f = (f_0, \ldots, f_m)$ на G с теми же, что и в § 2 гл. II, нормирующими множителями. Оснастим это множество оракулом нулевого порядка

$$\psi(f, x) = f(x),\ x \in G,\ f \in \boldsymbol{C}(G, R^n, m).$$

Обозначим полученный класс $\bar{\boldsymbol{C}}(G, R^n, m)$. Предъявим метод решения задач этого класса. Дальнейшее изложение основано на работе авторов [31].

2.2. Идея метода является развитием идеи ММЦТ (§ 5 гл. II) и состоит в следующем. Пусть, для большей ясности, $m = 0$. Ответ оракула первого порядка на вопрос об f в точке x позволяет локализовать решение задачи в одной из частей, на которые G делится некоторой, проходящей через x, гиперплоскостью. В другой из этих частей f не меньше, чем в x. Выбросив из G эту часть, мы уменьшаем область локализации решения. Ответ оракула нулевого порядка на вопрос об f в точке $x \in G$ не позволяет сузить область локализации решения. Однако, если мы «опросим f» в вершинах $x_0, \ldots, x_n$ некоторого симплекса Δ, то некоторое уменьшение области локализации мы себе обеспечим.

Пусть x_{i_0}— та из вершин симплекса, в которой f максимальна. На противолежащей x_{i_0} грани Δ_0 симплекса Δ f не больше, чем в x_{i_0} (ибо f выпукла, так что ее максимум на Δ совпадает с максимумом по вершинам Δ). Стало быть, f не убывает при переходе от $x \in \Delta_0$ к x_{i_0}, но тогда она не убывает на луче $[x, x_{i_0}]$ за точкой x_{i_0}. Стало быть, в многогранном конусе K с вершиной x_{i_0}, получающемся отражением от x_{i_0} конуса направлений из x_{i_0} в Δ, имеем

$$f(x) \geqslant f(x_{i_0}).$$

Поэтому указанный конус можно выбросить из G, уменьшив таким образом область локализации решения.

К сожалению, новая область локализации лишь «немногим меньше» старой и к тому же (что хуже всего) невыпукла, что создает существенные трудности для реализации намеченной идеи «в чистом виде».

Тут на помощь приходят соображения, лежащие в основе ММЦТ. Предположим, что G — шар (эллипсоид), а Δ — симплекс, расположенный вблизи центра G (так что конус K можно, грубо говоря, считать имеющим вершину в центре G). Тогда $G \setminus K$ заведомо содержится в выпуклом теле $\bar{G}_1$, получающемся отсечением от G некоторой «шапочки». $\bar{G}_1$ можно было бы взять за новую область локализации решения. Однако итерировать процесс мы не можем — граница $\bar{G}_1$ содержит плоский кусок, так что следующий «выбрасываемый конус» может опираться как раз на этот кусок, и новой «шапочки» отсечь не удастся. Целесообразно поэтому попытаться заключить $\bar{G}_1$ в эллипсоид G_1 минимального объема и считать последний новой областью локализации решения. При таком способе действий ситуация на всех шагах геометрически одинакова. Однако успех намеченные действия принесут лишь тогда, когда G_1 окажется меньшим (по объему), чем G,— только тогда объемы очередных областей локализации будут убывать с ростом номера шага со скоростью геометрической прогрессии. Легко понять, каков должен быть раствор конуса K, чтобы объем G_1 был меньше объема G. К сожалению, оказывается, что многогранный конус, отвечающий правильному симплексу, для этого слишком мал.

Мы встаем перед новой проблемой. Чтобы реализовать намеченный план, надо быть уверенным, что выбрасываемый конус K достаточно велик. Нельзя выбрать симплекс Δ так, чтобы любой из $n+1$ отвечающих ему конусов имел нужную величину. Так как мы заранее (до обращения к оракулу) не знаем, какой из этих $n+1$ конусов можно будет удалить, то вполне может случиться, что удаляемый конус окажется недостаточно большим, и мы не сможем уменьшить область локализации решения, сохранив ее форму.

Проще всего выйти из создавшегося положения так: подсчитать значения f не в вершинах малого симплекса, расположенного вблизи центра G, а в вершинах многогранника Γ с большим числом крайних точек. По-прежнему можно будет удалить из G конус, получающийся отражением от максимальной (по значению f) вершины Γ конуса направлений, ведущих из этой вершины в Γ. Если взять этот многогранник «хорошо аппроксимирующим шар», то конус такого типа будет достаточно велик, какая бы вершина Γ не оказалась максимальной.

Подобный способ (в несколько другом контексте) применяется в известном методе Кузовкина — Тихомирова [14] решения выпуклых экстремальных задач при оракуле нулевого порядка. Его недостаток состоит в том, что число вершин Γ, а вместе с ним и трудоемкость метода, оказывается, должны экспоненциально расти с ростом n. Чтобы избежать этого, применим другой подход — менее очевидный, но более экономный.

Пусть Δ — правильный симплекс с центром тяжести в центре x_0 шара G, а y^0 — максимальная из его вершин. Рассмотрим правильную пирамиду Δ_0 с вершиной y^0, нужным раствором конуса при вершине и высотой, коллинеарной $y^0 - x_0$. Основание ее выберем так, чтобы для прочих вершин $y_1, \ldots, y_n$ векторы $y_i - y^0$ были ортогональны $y_i - x_0$ (т. е. y_i есть проекция $x_0 - y^0$ на соответствующее ребро Δ_0). Возможно, что y^0 — максимальная вершина в пирамиде Δ_0. Тогда мы нашли подлежащий удалению конус нужного размера. Если это не так, то выберем максимальную вершину Δ_0 (обозначим ее y^1) и построим пирамиду Δ_1 с вершиной y^1 в точности так же, как по y^0 строилась пирамида Δ_0.

Если y^1 будет максимальна в Δ_1, все в порядке; если нет — выберем максимальную вершину Δ_1 и построим Δ_2 и т. д.

Центральное обстоятельство состоит в том, что этот процесс не может длиться долго. Действительно, $f(y^0) < f(y^1) < f(y^2) < \ldots$, но $|y^{i+1} - x_0| = (\cos\psi)^{i+1} |y^0 - x_0|$ (угол ψ есть угол между $y^0 - x^0$ и боковым ребром Δ_0). Стало быть, при некотором явно оцениваемом i будет $|y^{i+1} - x_0| \leqslant \frac{1}{n} |y^0 - x_0|$, т. е. $y^{i+1} \in \Delta$ (если только процесс не оборвался ранее). Но это невозможно, ибо по построению $f(y^{i+1}) > f(y^0)$, а y^0 была максимальной вершиной Δ. Итак, в ходе описываемого процесса не позднее, чем через некоторое, заранее оцениваемое число шагов будет построен подлежащий удалению конус нужного размера.

2.3. Перейдем теперь к описанию метода. Рассмотрим случай, когда требуется обеспечить заданные абсолютные погрешности $\vec{\varepsilon} = (\varepsilon_0, \ldots, \varepsilon_m) > 0$ в решении задач класса $\overline{C}(G, R^n, m)$. Предпошлем изложению метода лемму, обобщающую лемму 5.2 гл. II.

Лемма. *Пусть W_r — евклидов шар с центром в* 0 *радиуса r в n-мерном евклидовом пространстве R^n с ортобазисом $e_1, \ldots, e_n$, $n > 1$. Пусть $Q(\varphi)$ — конус раствора* 2φ:

$$Q(\varphi) = \{x \mid \langle x \mid e_n\rangle \geqslant \cos\varphi\, |x|\}, \quad 0 \leqslant \varphi \leqslant \frac{\pi}{2}.$$

Если $\cos\varphi < 1/n$, то множество $W_r \setminus Q(\varphi)$ можно заключить в эллипсоид объема $\beta^n(\varphi) \mid W_r \mid_n$, имеющий центр в точке $(-r\gamma(\varphi) e_n)$, где

$$\gamma(\varphi) = \frac{1 - n\cos\varphi}{1+n},$$

$$\beta(\varphi) = 2\left(\sin\frac{\varphi}{2}\right)^{\frac{n-1}{n}} \left(\cos\frac{\varphi}{2}\right)^{\frac{n+1}{n}} \frac{n\left(\frac{n-1}{n+1}\right)^{\frac{1}{2n}}}{\sqrt{n^2-1}}.$$

Отметим, что лемма 5.2 гл. II является частным случаем этого утверждения, получающимся из него при $\varphi = \pi/2$.

Упражнение 1. Докажите лемму.

Нас будет интересовать частный случай леммы, отвечающий $\varphi = \varphi_n = \arccos\left(\frac{1}{2n}\right)$.

Можно видеть, что

$$\beta(\varphi_n) = 1 - \frac{d_n}{n^2}, \quad d_n > 0,$$

$$\lim_{n\to\infty} d_n = 1/8, \quad \gamma(\varphi_n) = \frac{1}{2(n+1)}.$$

Ясно, что существует такое $\theta_n = \theta(n) > 0$, что утверждение леммы при $\varphi = \varphi_n$ остается верным и тогда, когда $\beta(\varphi_n)$ в ее формулировке заменено на $\overline{\beta}(\varphi_n) = \frac{1}{2}(1 + \beta(\varphi_n))$, а конус $Q(\varphi_n)$ разрешается сдвигать произвольно с тем условием, что его вершина лежит в $\theta(n)$ r-окрестности центра W_r. При этом соответствующее $\theta(n) \geqslant c/n$, $c > 0$ — абсолютная константа.

Пусть $\vec{\varepsilon} > 0$ — заданная абсолютная точность. Предположим, далее, что известны два подобных эллипсоида $\overline{W}_0$ и W_0. такие, что $\overline{W}_0 \subset G \subset W_0$, и пусть $|W_0| = \beta^n |\overline{W}_0|$. Известно (см. п. 1.1 гл. IV), что $\overline{W}_0$ и W_0 можно выбрать так, что $\beta \leqslant n$. Пусть O_0 — система координат, в которой W_0 — шар радиуса r_0 с центром в 0. Описываемый метод состоит из этапов. Нулевой этап является подготовительным. На i-м ($i \geqslant 1$) этапе строится очередная область локализации решения $W_i \cap G$ (W_i — эллипсоид, являющийся шаром радиуса r_i с центром в 0 в координатах O_i, имеющих тот же элемент объема, что и исходные координаты O_0). Ответы оракула о значениях компонент решаемой задачи в точках x будем обозначать $\psi_j(x)$, $0 \leqslant j \leqslant m$ (причина столь странных обозначений — резервирование возможности применять метод и в случае, когда мы располагаем приближенным, а не точным оракулом).

Нулевой этап. Рассмотрим правильный (в координатах O_0) симплекс $\hat{\Delta}$, вписанный в шар $\overline{W}_0$. Пусть $x^0, \ldots, x^n$ — его вершины. Зададим в них вопросы о решаемой задаче и сформируем числа

$$\bar{r}_j(f) = \max\{0, \max_{0 \leqslant k \leqslant n} \psi_j(x^k)\}, \quad j = 1, 2, \ldots, m.$$

Положим

$$\bar{r}_0(f) = \max_{0 \leqslant k \leqslant n} \psi_0(x^k).$$

Этап i ($1 \leqslant i \leqslant M_f$) ($M_f$ — формируемое в ходе работы метода число этапов решения задачи). К i-му этапу имеются O_{i-1}, r_{i-1} и W_{i-1}, а также центр x_i шара W_{i-1}. На i-м этапе последовательно производятся следующие действия. (Все рассмотрения будут вестись в координатах O_{i-1}, которые на i-м этапе считаются декартовыми; таким образом, евклидова структура R^n — своя на каждом этапе.) Положим $\theta_n = \theta(n)$.

1°. Возможно, что шар $\overline{W}_{i-1}$ радиуса $\theta_n r_{i-1}$ с центром в x_i не содержится в G. В этом случае находим точку $u_i \in \overline{W}_{i-1} \setminus G$ и рассматриваем опорный функционал φ_i к функции $\rho_{\|\cdot\|}(\cdot, G)$ *) в точке u_i. Тогда $\varphi_i \neq 0$ (ибо $u_i \notin G$). Рассматриваем конус $K_i = \{\langle\varphi_i \mid x - u_i\rangle \geqslant 0\}$ с вершиной в u_i. Очевидно, $K_i \cap G = \phi$. В качестве W_i выбираем эллипсоид наименьшего объема, содержащий $W_{i-1} \setminus K_i$, и в качестве O_i — ту систему координат (с тем же, что и у O_{i-1}, элементом объема), в которой W_i — шар. По лемме 2.3 радиус r_i шара W_i в координатах O_i не превышает $\overline{\beta}(n) r_{i-1}$. В качестве x_{i+1} берем центр W_i. Полагаем $j(i) = m + 1$. Переходим к п. 3°.

*) $\|\cdot\|$ — какая-нибудь раз навсегда фиксированная норма на R^n.

2°. Пусть теперь $\overline{W}_{i-1} \subset G$. Впишем в $\overline{W}_{i-1}$ правильный симплекс $\overline{\Delta}_i$. Пусть $\bar{z}_0, \ldots, \bar{z}_n$ — его вершины. Зададим вопросы о решаемой задаче f в точках x_i, $\bar{z}_0, \ldots, \bar{z}_n$ и образуем числа

$$r_i^j = \max_{0 \leqslant k \leqslant n} \psi_j(\bar{z}_k), \quad 0 \leqslant j \leqslant m$$

(ясно, что $r_i^j = \max_{x \in \overline{\Delta}} f_j(x)$ для точного оракула). Положим

$$j(i) = \begin{cases} 0, & \text{если } r_i^j \leqslant \varepsilon_j, \quad 1 \leqslant j \leqslant m, \\ j > 1, & \text{такое, что } r_i^j > \varepsilon_j, \qquad \text{если такое } j \in \overline{1, m} \text{ есть.} \end{cases}$$

Пусть $\Delta_i = \overline{\Delta}_i$ при $j(i) > 0$. В противном случае впишем в $\overline{\Delta}_i$ шар $\hat{W}_{i-1}$, касающийся граней $\overline{\Delta}_i$, и рассмотрим вписанный в в $\hat{W}_{i-1}$ правильный симплекс Δ_i. Обозначим через $\tilde{V}_i$ описанный вокруг Δ_i шар, а через $\tilde{r}_i$ — его радиус. Ясно, что $\tilde{r}_i \geqslant \frac{\theta_n}{n} r_{i-1}$ и $\tilde{V}_i \subset \overline{W}_{i-1}$.

Применим теперь следующую процедуру.

2.1. Обозначим для краткости $g(y) = \psi_{j(i)}(y)$. Для любого конечного числа точек G определим максимальную точку как ту точку набора, в которой g максимально. Далее, говоря о выборе максимальной точки среди точек данного набора, будем иметь в виду, что для этого необходимо опросить оракул в каждой из этих точек.

Пусть y^0 — максимальная среди вершин Δ_i. Построим $(n+1)$-гранную правильную пирамиду Δ^0 с вершиной y^0, высотой, коллинеарной $y^0 - x_i$, углом φ_n между высотой и боковыми гранями, и боковыми ребрами, ортогональными радиус-векторам своих концов, лежащих на основании (напомним, что изложение ведется в евклидовой структуре R^n i-го этапа — в ней декартовы координаты суть O_{i-1}). Заметим, что $\Delta^0 \subset \tilde{V}_i$. Если y^0 — максимальная среди вершин Δ^0, то скажем, что Δ^0 нормальна, положим $z_i = y^0$ и перейдем к п. 2.2. В противном случае найдем максимальную вершину y^1 у Δ^0 и положим $\delta^0 = g(y^1) - g(y^0)$. Построим для y^1 пирамиду Δ^1 по тем же правилам, что и Δ^0 для y^0. Если Δ^1 нормальна (т. е. y^1 — ее максимальная вершина), то положим $y^1 = z_i$ и перейдем к п. 2.2; в противном случае найдем максимальную вершину y^2 пирамиды Δ^1, положим $\delta^1 = g(y^2) - g(y^1)$, построим Δ^2 и т. д. Оборвем этот процесс в тот момент, когда впервые получится $y^k \in \Delta_i$ (это обязательно случится, если, конечно, он не оборвался ранее переходом к п. 2.2). Если в ходе построения очередных точек не было получено нормальной пирамиды, выберем среди построенных точек $y^0, \ldots, y^{k-1}$ ту, которой отвечает минимальное из чисел $\delta^0, \ldots, \delta^{k-1}$. Обозначим эту точку z_i и перейдем к п. 2.2.

2.2. По построению точка z_i есть некоторая точка y^s. Рассмотрим конус

$$K_i = \{y \mid \exists\lambda > 0:\ \lambda\,(y^s - y) + y^s \in \Delta^s\}.$$

По построению $\Delta^l\ K_i$ содержит круговой конус с осью $y^s - x_i$, раствором угла $2\varphi_n$ и вершиной в y^s. По построению $y^s \in \overline{V}_i \subset \subset \overline{W}_{i-1}$.

В силу леммы 1.7 область $W_{i-1} \setminus K_i$ содержится в эллипсоиде W_i объема $(\overline{\beta}\,(\varphi_n))^n \mid W_{i-1} \mid_n$. Пусть O_i — та система координат, в которой W_i есть шар с центром в 0. Пусть, кроме того, элемент объема в O_i — тот же, что и в O_{i-1}. Тогда радиус r_i шара W_i в координатах O_i есть $r_i = \overline{\beta}\,(\varphi_n)\, r_{i-1}$.

3°. Положим $a_i = \min\,\{\psi_0\,(z_s) \mid s \leqslant i,\ \ j\,(s) = 0\}$ (при $j\,(s) > 0,\ 1 \leqslant s \leqslant i$, полагаем $a_i = +\infty$). Пусть $\rho_i = \max\,\{0,\ \bar{r}_0\,(f) - a_i\}$ и

$$\nu_i = \begin{cases} 1, & j\,(i) = m+1, \\ \varepsilon_{j(i)}/\bar{r}_{j(i)}\,(f), & j\,(i) \in \overline{1, m}, \\ \dfrac{\varepsilon_0}{\rho_i + \varepsilon_0}, & j\,(i) = 0. \end{cases}$$

Положим, далее, $\nu^i = \min\limits_{0 \leqslant s \leqslant i} \nu_s$, где $\nu_0 = 1$. Если $r_i < \dfrac{\nu^i}{4n\beta}\, r_0$, перейдем к заключительному этапу, в противном случае — к этапу $i+1$.

Заключительный этап. Работа метода прекращается выдачей результата

$$\bar{x} = \begin{cases} *, \text{ если } j\,(i) > 0 \text{ при } i \leqslant M_f, \\ z_{i_0}, \text{ где } i_0 \text{ таково, что } j\,(i_0) = 0 \text{ и} \\ \quad \psi_0\,(z_{i_0}) = a_{M_f} \text{ (т. е. } \psi_0\,(z_{i_0}) \leqslant \psi_0\,(z_s) \\ \quad \text{для всех } s \text{ с } j\,(s) = 0), \text{ — в противном случае.} \end{cases}$$

Здесь M_f — номер последнего из рабочих этапов.

2.3. **Теорема.** (*i*) *Описанный метод, настроенный на точность* $\vec{\varepsilon} > 0$, *решает с этой точностью все задачи из* $\overline{C}\,(G, R^n, m)$ *при трудоемкости, не превосходящей* $\Phi^0_{n,\infty}\left(\dfrac{\nu\,(f, \vec{\varepsilon})}{1 + \nu\,(f, \vec{\varepsilon})}\right)$ (*здесь*

$$\Phi^0_{n,\infty}(\nu) = \left]\, c n^7 \ln n \ln \frac{4n\beta}{\nu}\, \right[,$$

с — абсолютная константа, а

$$\nu\,(f, \vec{\varepsilon}) = \min_{0 \leqslant j \leqslant m} \frac{\varepsilon_j}{r_j\,(f)}$$

— максимальная относительная погрешность, обеспечивающая данную абсолютную погрешность).

(*ii*) *Пусть* $\beta \leqslant n$ (*такие* $\overline{W}_0$, W_0 *всегда можно найти*). *Существует полином* $P\,(n)$, *обладающий следующим свойством. Пусть описанный метод применяется к решению задач класса* $\overline{C}^{\vec{\delta}}\,(G, R^n, m)$, *получающегося из* $\overline{C}\,(G, R^n, m)$ *заменой точного оракула нулевого порядка на приближенный оракул* $(\psi_0\,(f, x), \ldots, \psi_m\,(f, x)):\ \{f\} \times G \to R^{m+1}$ *абсолютной погрешности* $\vec{\delta}$ (*последнее означает, что* $\mid \psi_j\,(f, x) - f_j\,(x) \mid \leqslant \delta_j$ *для всех* $f \in \overline{C}\,(G, R^n, m)$, $x \in G$ *и* $j \in \overline{0, m}$). *Пусть, далее, требуемая абсолютная погрешность* $\vec{\varepsilon}$ *решения задач из* $\overline{C}^{\vec{\delta}}$ *удовлетворяет условию* $\varepsilon_j \geqslant P\,(n)\,\delta_j$, $0 \leqslant j \leqslant m$. *Тогда описанный метод, настроенный на точность* $\vec{\varepsilon}$, *решает все задачи из* $\overline{C}^{\vec{\delta}}$ *с погрешностью* $\leqslant 2\vec{\varepsilon}$, *при трудоемкости*

$$\leqslant \Phi^0_{n,\infty}\left(\frac{\nu\,(f, \vec{\varepsilon})}{1 + 2\nu\,(f, \vec{\varepsilon})}\right).$$

Обсудим утверждение теоремы. Мы построили детерминированный метод решения общих выпуклых задач, использующий оракул нулевого порядка и обладающий следующим свойством эффективности: в асимптотике (по $\nu\,(f, \vec{\varepsilon}) \to 0$) его трудоемкость совпадает со сложностью рассматриваемого класса задач (с точностью до множителя, оценивающегося сверху полиномом от размерности задачи). Метод помехоустойчив в следующем смысле: он в состоянии решить задачу с абсолютной погрешностью $\vec{\varepsilon} > 0$, если ошибки в используемой им текущей информации не превосходят величин, пропорциональных $\vec{\varepsilon}$. Правда, коэффициент пропорциональности — показатель помехоустойчивости метода — стремится к 0 с ростом размерности задачи. В § 4 мы увидим, что качественно последний эффект неустраним.

Конечно, трудоемкость описанного метода сильно зависит от n, так что реально он не может быть рекомендован при сколько-нибудь заметных значениях n.

Упражнение 2. Докажите теорему.

Упражнение 3. Постройте аналог описанного метода (и докажите для него соответствующую версию теоремы) в случае, когда требуется решать безусловные ($m = 0$) задачи с заданной относительной погрешностью $\nu > 0$. Попробуйте распространить этот результат на случай задач с ограничениями (в случае приближенного оракула последнее не удалось авторам, по крайней мере, за то время, которое они были готовы потратить на этот вопрос).

§ 3. Методы нулевого порядка: стохастический оракул. I

В предыдущих параграфах рассматривался случай, когда методу была доступна точная (или искаженная малыми ошибками) информация о значениях функционалов решаемой задачи. Теперь мы рассмотрим ситуацию, в которой указанная информация возмущена случайной (и, вообще говоря, не малой) помехой, однако помеха в среднем равна 0, а помехи разных тактов независимы.

3.1. Начнем с описания рассматриваемого класса задач. Пусть $G \subset R^n$ — выпуклое ограниченное замкнутое тело, $G_I \supset G$, содержащее G тело такого же типа, $\mathcal{F}_m(G_I)$ — множество всех $(m+1)$-мерных непрерывных покомпонентно выпуклых вектор-функций на G_I. С каждой такой функцией $f = (f_0, \ldots, f_m)$ свяжем *задачу* f:

$$f_0(x) \to \min \mid x \in G, \quad f_j(x) \leqslant 0, \quad 1 \leqslant j \leqslant m.$$

Будем считать, что задачи из $\mathcal{F}_m(G_I)$ наблюдаются с помощью стохастического оракула $\mathcal{O} = ((\Omega, \mathsf{F}_\omega); \psi(x, f, \omega))$ с функцией наблюдения $\psi(x, f, \omega)$: $\mathcal{F}_m(G_I) \times \Omega \times G_I \to R^{m+1}$. Компоненты $\psi_0, \ldots, \psi_m$ вектор-функции ψ интерпретируются как наблюдения значений соответствующих компонент f. Как всегда, $(\Omega, \mathsf{F}_\omega)$ предполагается польским пространством с регулярной борелевой полной по Лебегу вероятностной мерой F_ω, а $\psi(x, f, \omega)$ предполагается борелевой по ω, x.

Чтобы иметь возможность работать с оракулом $\mathcal{O}$, следует, как и в случае оракулов первого порядка, принять определенные гипотезы об интенсивности помех и возможных границах смещенности для оракула $\mathcal{O}$. Сделаем это следующим образом. Пусть $L_0, \ldots, L_m > 0$ и $\nu_0 \geqslant 0$. Пусть также $r > 1$. Обозначим через $\overline{C}^{\mathcal{O}}_{r,\nu_0}(G, L_0, \ldots, L_m, G_I)$ множество всех задач f из $\mathcal{F}_m(G_I)$, для которых при всех $x \in G_I$

$$(1) \qquad \mathsf{M}_{\mathsf{F}_\omega} \max_{0 \leqslant j \leqslant m} \left(\frac{|\psi_j(x, f, \omega)|}{L_j}\right)^r \leqslant 1;$$

$$(2) \qquad \max_{0 \leqslant j \leqslant m} \frac{|\mathsf{M}_{\mathsf{F}_\omega}\psi_j(x, f, \omega) - f_j(x)|}{L_j} \leqslant \nu_0.$$

С содержательной точки зрения L_j^r, грубо говоря, суть априорные оценки (абсолютных) r-х моментов ответов оракула о значениях f_j, а ν_0 — априорная оценка (относительной) смещенности оракула $\mathcal{O}$. В качестве нормирующих множителей для введенного класса выберем сами параметры L_j.

При рассмотрении методов решения задач этого класса параметры r, ν_0, L_j будем считать априори заданными. Ясно, что простой нормировкой компонент решаемых задач можно редуцировать каждый класс $\overline{C}^{\mathcal{O}}_{r,\nu_0}(G, L_0, \ldots, L_m, G_I)$ к некоторому классу $\overline{C}^{\mathcal{O}}_{r,\nu_0}(G, 1, \ldots, 1, G_I) \equiv \overline{C}^{\mathcal{O}}_{r,\nu_0}(G, G_I)$. Далее, мы рассматриваем именно этот — нормированный класс задач.

3.2. Опишем достаточно бегло два подхода к решению задач класса $\overline{C}^{\mathcal{O}}_{r,\nu_0}(G, G_I)$. Первый основан на имитации для рассматриваемых задач стохастического оракула первого порядка имеющимся стохастическим оракулом нулевого порядка с последующим применением методов типа стохастической аппроксимации (главы V — VI). При втором подходе (ему отведен следующий параграф) по стохастическому оракулу нулевого порядка строится, грубо говоря, детерминированный оракул нулевого порядка с последующим применением методов нулевого порядка, ориентированных на (приближенный) детерминированный оракул.

Начнем с первого подхода, при котором, отправляясь от стохастического оракула нулевого порядка, строят стохастический же оракул первого порядка. Идея этой конструкции неоднократно использовалась в разного рода версиях методов случайного поиска. Мы изложим ее в общих чертах, без особых подробностей. Фиксируем евклидову структуру R^n, и пусть d — диаметр и α — асферичность G в соответствующей метрике. Будем считать, что f — непрерывная выпуклая функция на G_I, а $\widehat{O}$ — *несмещенный* оракул нулевого порядка. Пусть $\rho > 0$ — параметр. Рассмотрим функцию f^ρ, получающуюся из f осреднением

$$f^\rho(x) = \frac{1}{|V_\rho|_n} \int_{\|\xi\| \leqslant \rho} f(x + \xi)\, d\xi, \tag{3.1}$$

где $V_\rho = \{\xi \mid \|\xi\| \leqslant \rho\}$, $|\cdot|_n$ — n-мерный лебегов объем. Функция $f^\rho(x)$ определена не во всей области определения G_I функции f, а в несколько меньшей области — во всех точках $x \in G_I$, расстояние от которых до границы G_I не меньше ρ (обозначим эту область G_I^ρ). В частности, если ρ/d достаточно мало, f^ρ определена в «большей части» G; если же G_I содержит ρ-окрестность G, то f^ρ определена во всем G.

Упражнение 1. Пусть G содержится в шаре радиуса R с центром в 0 и содержит шар радиуса $R/2\alpha$ с тем же центром. Докажите, что f^ρ определена в множестве θG, где $\theta = 1 - \frac{2\rho\alpha}{R}$.

В силу непрерывности f f^ρ может быть сделана равномерно (во всем G_I^ρ) как угодно близкой к f выбором достаточно малого $\rho > 0$. Наконец, G_I^ρ выпукло и f^ρ выпукла в G_I^ρ. Более того, легко видеть, что f^ρ — непрерывно-дифференцируема в G_I^ρ, причем ее градиент находится по формуле

$$\nabla f^{\rho}(x) = \frac{1}{|V_\rho|} \int\limits_{\|\xi\|=\rho} f(x+\xi)\frac{\xi}{\|\xi\|}\,ds_\rho(\xi) \tag{3.2}$$

(здесь $ds_\rho(\xi)$ — элемент поверхности сферы в R^n радиуса ρ). Переходя от интегрирования по площади сферы к интегрированию по нормированной площади (площадь всей сферы принята за 1), т. е. к интегрированию по равномерному распределению вероятностей $d\sigma(\xi)$ на s_1, получим

$$\nabla f^{\rho}(x) = \frac{n}{\rho} \int\limits_{\|\xi\|=1} \xi f(x+\rho\xi)\,d\sigma(\xi). \tag{3.3}$$

Упражнение 2. Докажите (3.2) — (3.3).

Таким образом, градиент f^ρ в G_I^ρ может быть представлен в виде интеграла от значений f с некоторым (векторным) весом.

Теперь мы в состоянии описать идею намеченного подхода. Она состоит в следующем.

1°. Выберем, на основании априорной информации, параметр ρ столь малым, чтобы область G_I^ρ определения задачи $(f_0^\rho, \ldots, f_m^\rho)$ достаточно хорошо аппроксимировала G, а сами функции f_j^ρ в этой области достаточно хорошо аппроксимировали f_j. Более точно, если ν — требуемая точность решения задачи, то подберем $\rho > 0$ и числа α_j, $1 \leqslant j \leqslant m$, так, чтобы любая точка $x \in G_I^\rho \cap G$ удовлетворяла соотношению

$$\nu(x, f) \leqslant \nu(x, \bar f^\rho) + \frac{\nu}{2},$$

где

$$\bar f^\rho: f_0^\rho \to \min \mid x \in G_I^\rho \cap G, \quad f_j^\rho(x) \leqslant \alpha_j, \quad 1 \leqslant j \leqslant m,$$

а ν — требуемая точность решения задачи f (считаем, что нормирующие множители для $\bar f^\rho$, как и для f, суть единицы).

О том, как выбрать надлежащее ρ, речь пойдет ниже.

2°. Выбрав ρ и α_j указанным способом, сведем проблему решения f с (средней) погрешностью ν к проблеме решения $\bar f^\rho$ со средней погрешностью $\nu/2$.

Укажем способ решения последней проблемы. Смоделируем для $\bar f^\rho$ оракул первого порядка следующим образом. Разыграем наудачу (по равномерной мере в V_ρ) точку $\xi \in V_\rho$ и обратимся к $\hat{\mathcal{O}}$ с вопросом от f в точке $\xi + x$. Полученный ответ примем за наблюдение $f^\rho(x)$. Обозначим этот ответ $a(\xi, \omega) = (a_0(\xi, \omega), \ldots$ $\ldots, a_m(\xi, \omega))$, где ω — шум оракула $\hat{\mathcal{O}}$ на соответствующем такте. Затем разыграем наудачу (по мере $d\sigma$) точку $\bar\xi$ на сфере $s_1 = \{\xi \mid \|\xi\| = 1\}$ и обратимся к оракулу $\hat{\mathcal{O}}$ с вопросом об f в точке

$x + \rho\bar\xi$. Пусть $b_0, \ldots, b_m$ — сообщенные им числа. Образуем векторы $(n/\rho)\, b_j\bar\xi$ и будем считать их наблюдениями градиентов f_j^ρ в точке x. Обозначим их $b_j(\bar\xi, \omega')$, где ω' — шум оракула $\hat{\mathcal{O}}$ на соответствующем такте. Покажем, что сымитированный таким образом оракул действительно является несмещенным оракулом первого порядка для задачи $\bar f^\rho$. В самом деле,

$$\mathsf{M}a = \frac{1}{|V_\rho|_n}\int \mathsf{M}_{\mathsf{F}_\omega} a(\xi, \omega)\,d\xi = \frac{1}{|V_\rho|_n}\int f(x+\xi)\,d\xi = f^\rho(x)$$

(здесь f^ρ означает вектор-функцию $(f_0^\rho, \ldots, f_m^\rho)$). Равным образом

$$\mathsf{M}b_j = \frac{n}{\rho}\int\limits_{\|\bar\xi\|=1} d\sigma(\bar\xi)\,\mathsf{M}_{\mathsf{F}_{\omega'}} a_j(\rho\bar\xi, \omega')\,\bar\xi = \frac{n}{\rho}\int\limits_{\|\bar\xi\|=1} d\sigma(\bar\xi)\, f_j(x+\rho\bar\xi)\bar\xi = \nabla f_j^\rho(x).$$

Оценим интенсивность помех для оракула, полученного таким способом. Имеем в силу свойств $\hat{\mathcal{O}}$

$$\mathsf{M}\|a\|_\infty^r = \frac{1}{|V_\rho|_n}\int \mathsf{M}_{\mathsf{F}_\omega}\|a(\xi, \omega)\|_\infty^r\,d\xi \leqslant 1 \tag{3.4}$$

и

$$\mathsf{M}\|b_j\|^r = \left(\frac{n}{\rho}\right)^r \int\limits_{\|\bar\xi\|=1} d\sigma(\bar\xi)\,\mathsf{M}_{\mathsf{F}_{\omega'}}|a_j(\rho\bar\xi, \omega')|^r \leqslant \left(\frac{n}{\rho}\right)^r. \tag{3.5}$$

Итак, при известном ρ уровень помех смоделированного оракула первого порядка (обращение к которому «стоит» двух обращений к $\hat{\mathcal{O}}$) можно явно оценить.

Теперь мы в состоянии решать задачу $\bar f^\rho$, используя построенный выше оракул первого порядка, методами первого порядка (рассчитанными на стохастический оракул).

Замечание. Ясно, что описанный выше способ имитации стохастического оракула первого порядка с помощью оракула нулевого порядка применим не только к обычным задачам математического программирования, но и к экстремальным задачам с операторными ограничениями (см. § 5 гл. VI). Предоставляем читателю сформулировать и обосновать соответствующие утверждения.

3.3. Остановимся на некоторых трудностях, связанных с намеченным подходом. Мы видим, что для его реализации надо уметь конструктивно выбирать ρ (а также и α_j) так, чтобы задача $\bar f^\rho$ с заданной точностью аппроксимировала задачу f. При этом, чем меньше ρ мы выберем, тем хуже (с точки зрения интенсивности помех) будет конструируемый оракул первого порядка (см. (3.5)) и тем больше, стало быть, будет трудоемкость использующего этот оракул метода первого порядка при заданных требованиях к точности решения задачи (*относительная точность* ν

решения исходной задачи превращается, грубо говоря, в *абсолютную* точность, требуемую от метода первого порядка).

С другой стороны, параметр ρ не может не быть мал. Действительно, даже если отвлечься от эффектов, связанных с тем, что G_I^ρ должна хорошо аппроксимировать G, надо, чтобы f_j^ρ хорошо (равномерно) приближали f_j в G_I^ρ — именно, с точностью порядка ν. Пусть, скажем, G_I содержит достаточно большую окрестность G, так что можно считать, что $G_I^\rho \supset G$. Ясно, что если ρ выбрано так, что $|f_j^\rho - f_j| \leqslant \nu/4$ в G и $\alpha_j = \nu/4$, то $\bar{f}^\rho$ доставляет требуемую аппроксимацию f.

Конечно, выбор ρ из этих соображений достаточен для наших целей, но отнюдь не необходим. Для *данной* f можно «случайно» получить интересующее нас $\bar{f}^\rho$ и при гораздо больших значениях параметров. Однако это действительно может произойти только случайно. Если требуется выбрать ρ так, чтобы для *любой* подлежащей решению задачи гарантировать приемлемое качество аппроксимации f с помощью $\bar{f}^\rho$, то нельзя, по-видимому, действовать лучше, чем выбирая ρ из тех соображений, что f_j^ρ обязаны равномерно в G_I^ρ аппроксимировать f_j с точностью порядка ν.

Посмотрим теперь, как же нужно выбирать ρ с тем, чтобы обеспечить соотношения

$$|f_j^\rho - f_j| \leqslant a\nu \text{ в } G_I^\rho, \quad 0 \leqslant j \leqslant m. \tag{3.6}$$

Соответствующий выбор ρ зависит от априорной информации о гладкости f_j. Пусть, скажем, известно, что G_I содержит $\bar{\rho}$-окрестность G и f_j липшицевы в ρ_0-окрестности G с константой L (считаем $\rho_0 \leqslant \bar{\rho}$). Никакими другими сведениями о гладкости f_j мы не располагаем. Тогда (3.6) обеспечивается выбором ρ в виде

$$\rho_\nu = \frac{a\nu}{L} \tag{3.7}$$

(считаем ν достаточно малым, так что $a\nu/L \leqslant \rho_0$). Вообще говоря, выбор ρ большим ρ_ν уже не гарантирует (3.6).

Пусть, далее, $r = 2$ (это соглашение вводится для большей ясности дальнейших рассуждений). Соединяя оценку трудоемкости соответствующего этому r (и гильбертову E) метода первого порядка, настроенного на абсолютную погрешность порядка ν, с оценкой (3.4) — (3.5) уровня интенсивности построенного выше оракула первого порядка, получим, что трудоемкость метода будет

$$\geqslant d_1 \left(\frac{nd}{\rho_\nu \nu}\right)^2 \asymp d_2 \left(\frac{nLd}{\nu^2}\right)^2$$

(здесь и далее d_i зависит от a и m, но не зависит от n, L, d, ν). Таким образом, в этом случае асимптотика трудоемкости по $\nu \to 0$ есть $O(1/\nu^4)$, а не $O(1/\nu^2)$, как для методов первого порядка

при стохастическом оракуле. Помимо этого неприятного явления, мы видим, что для успеха подхода надо еще уметь оценивать L. Если G_I, как и выше, содержит $\bar{\rho}$-окрестность G, то в $\bar{\rho}/2$-окрестности G f_j липшицевы с константой $\leqslant c/\bar{\rho}$, c — абсолютная константа, ибо $|f_j|$ не превосходит 1 в G_I в силу свойств оракула $\hat{\mathcal{O}}$. Итак, трудоемкость метода, получающегося в результате реализации рассматриваемого подхода, не меньше величины вида $d_3 \frac{n^2}{\nu^4}\left(\frac{d}{\bar{\rho}}\right)^2$. В наиболее естественном случае $d/\bar{\rho} = O(1)$ и оценка имеет вид $d_4 n^2/\nu^4$.

Конечно, если относительно гладкости f_j известно нечто большее, то оценку можно улучшить. Пусть, скажем, G_I содержит $O(d)$-окрестность G и известно, что $f_j'(x)$ существуют и *липшицевы* в G_I с константами $O(1)$. Тогда (3.6) обеспечивается выбором ρ вида $O(\sqrt{\nu})$, и трудоемкость результирующего метода будет оцениваться снизу величиной вида $d_5 n^2/\nu^3$. Эта последняя оценка уже не допускает качественного улучшения, во всяком случае, без чрезвычайно жестких и малоестественных ограничений на f_j.

В рассмотренных случаях G_I содержало окрестность G. Если это не так (скажем, $G_I = G$), то ситуация еще ухудшается (начинают влиять краевые эффекты). Принципиальных проблем здесь не возникает, но трудоемкость получающегося метода заметно растет. При $G = G_I$ для *всего* класса $C_{2,0}^{\hat{\mathcal{O}}}(G, G_I)$ на этом пути не удается сделать трудоемкость меньшей $d_6 \frac{n^2}{\nu^6 \alpha^2}$.

Таким образом, даже в самых благоприятных случаях разбираемый подход не в состоянии обеспечить трудоемкости, по порядку зависимости от точности лучшей $O(1/\nu^3)$. Правда, зависимость трудоемкости от размерности оказывается сравнительно приемлемой (с ростом n она растет как $(n \ln n)^2$).

Мы разобрали случай, когда исходный класс снабжался несмещенным оракулом. Конечно, при решении задач с точностью $\nu > 0$ можно допустить и некоторое ненулевое смещение ν_0; однако простой анализ показывает, что допустимое ν_0 должно быть по порядку меньшим ν — в наиболее благоприятных случаях должно быть $\nu_0 \leqslant O(\nu^{3/2})$ при $\nu \to 0$. Таким образом, методы, получающиеся при описанном подходе, чрезмерно требовательны к смещениям оракула (ср. ситуацию с разобранной в предыдущем параграфе, при изучении метода, получающегося при имитации оракула первого порядка оракулом нулевого).

§ 4. Методы нулевого порядка: стохастический оракул. II

4.1. Идея редукции методов нулевого порядка к методам первого порядка, на которой был основан рассматриваемый выше подход, весьма естественна и популярна. Как мы видели, она приводит к методам, трудоемкость которых приемлемым образом зависит от размерности задачи, но быстро растет с ростом требуемой точности (существенно быстрее, чем для методов первого порядка, использующих стохастический оракул). Естественность идеи могла бы навести на мысль, что этот быстрый рост трудоемкости с повышением точности закономерен и не есть дефект подхода. Оказывается, это предположение неправильно: можно предложить другой путь построения методов нулевого порядка, работающих со стохастическим оракулом и обеспечивающих точность ν в решении задач классов $\overline{C}^{\hat{\mathcal{O}}}_{r,0}(G, G_I)$ при трудоемкости, не превосходящей величины

$$\left]\ln^2\frac{1}{\nu}\left[\ln(m+2)\,R_r(n)/\nu^{\frac{\bar r}{\bar r-1}}\right.\right.,$$

где $\bar r = \min(r, 2)$, $R_r(n)$-некий полином от n. Подчеркнем, что эта оценка при ν, меньших некоторой абсолютной константы $\nu_* > 0$ с точностью до множителя порядка $O(]\ln^2 1/\nu[R_r(n))$, совпадает с *нижней* оценкой сложности классов задач выпуклой стохастической аппроксимации при оракуле первого порядка из § 3 гл. V, так что *при фиксированном n* и всех $\nu \leqslant \nu_*$, *т* получаемые методы существенно не улучшаемы по трудоемкости (именно, не более чем в $O(\ln^2 1/\nu)$ раз).

Таким образом, методы типа намеченных в § 3 действительно объективно плохи в асимптотике по $\nu \to 0$. Метод, который будет построен ниже, существенно превосходит их по характеру этой асимптотики (более того, отвечающая ему асимптотика трудоемкости по $\nu \to +0$ в принципе не улучшается более чем логарифмически). С другой стороны, полином $R_r(n)$ имеет довольно большую (заведомо большую двух) степень, так что новый метод много хуже старых по характеру зависимости трудоемкости от размерности задачи.

4.2. Идея метода, который мы собираемся построить, крайне проста. Мы имеем в своем распоряжении стохастический оракул нулевого порядка (пусть вначале точный: $\nu_0 = 0$). В то же время мы умеем решать выпуклые задачи при приближенном оракуле нулевого порядка, если только его (абсолютная) погрешность достаточно мала по сравнению с требуемой точностью решения. Именно, при рассмотрении задач класса $\overline{C}^{\hat{\mathcal{O}}}_{r,0}(G, G)$ и требуемой точности решения этих задач ν нам достаточно было бы уметь вычислять значения функционалов задачи с абсолютными ошибками, не превосходящими $\nu/4P(n)$, $P(n)$ — полином из теоремы 2.3. Если бы это нам удалось, мы могли бы решить задачу методом из § 2, настроив его на абсолютные погрешности $(\nu/4, \ldots, \nu/4)$. Трудоемкость последнего есть

$$M(n, \nu) \equiv S(n)\left]\ln\frac{1}{\nu}\right[,$$

где $S(n)$ — некоторый полином от n (мы учли, что изменения компонент f_j задачи $f \in \overline{C}^{\hat{\mathcal{O}}}_{r,0}(G, G)$ на G не превосходят 2).

Остается научиться вычислять значения функционалов задачи с ошибкой, не превосходящей $\nu/4P(n)$. Для этого поступим следующим образом: пронаблюдаем вектор $f(x)$ в интересующей нас точке $x \in G$ достаточно большое число раз Q и полученные (независимые) наблюдения $f(x)$ должным образом усредним. Конечно, какое бы Q ни взять, с некоторой, вообще говоря, положительной, вероятностью результаты усреднения не будут давать нужной аппроксимации $f(x)$. Выбрав Q достаточно большим, мы сможем сделать эту вероятность столь малой, что подобные «неприятности» не окажут заметного влияния на среднюю погрешность метода.

Более точно, имеющийся в нашем распоряжении оракул $\mathcal{O}$ позволяет, сделав Q наблюдений значений компонент f_j решаемой задачи $f \in \overline{C}^{\hat{\mathcal{O}}}_{r,0}(G, G)$ в произвольной точке $x \in G$, получить Q независимых реализаций $\xi_1, \ldots, \xi_Q$ случайной величины ξ, такой, что $\mathsf{M}\xi = f(x)$ и $\mathsf{M}\,\|\xi - f(x)\|_\infty^r \leqslant 2^r$. Применяя способ обработки наблюдений, описанный в п. 4 § 6 гл. VI, мы сможем построить оценку ξ^Q вектора $f(x)$, для которой

$$\Pr\left\{\|\xi^Q - f(x)\|_\infty > \frac{\nu}{4P(n)}\right\} \leqslant \alpha,$$

если

$$Q \geqslant Q_r(n, \nu, \alpha) = \left]\frac{W_r(n)}{\nu^{\bar r/(\bar r-1)}}\ln\frac{1}{\alpha}\ln(m+2)\right[,$$

где $W_r(n)$ — некоторый полином от n. Выберем

$$\alpha = \alpha_\nu = \nu/8M(n, \nu) \quad \text{и} \quad Q = Q(n, \nu, \alpha_\nu),$$

после чего поступим описанным выше образом (т. е. применим для решения задачи f метод $\mathcal{B}$ § 2, настроенный на абсолютные погрешности $(\nu/4, \ldots, \nu/4)$. При этом информацию для этого метода будем получать, делая по Q наблюдений вектора $f(x)$ в каждой из «опрашиваемых» точек и поставляя методу вектор ξ^Q, полученный в результате описанной выше обработки полученных наблюдений).

Ясно, что в ходе работы метода с вероятностью

$$\geqslant 1 - M(n, \nu)\,\alpha_\nu = 1 - \frac{\nu}{8}$$

будет получено приближенное решение задачи f погрешности $\leqslant 3\nu/2$ (это заведомо произойдет, если на всех $M(n, \nu)$ шагах работы метода $\mathcal{B}$ поставляемая ему информация будет отличаться от истинной покомпонентно не больше чем на $\nu/4P(n)$). С дополнительной вероятностью абсолютная погрешность результата заведомо не превзойдет 2, ибо $|f_j| \leqslant 2$ на G. Средняя погрешность метода, таким образом, не больше

$$\frac{\nu}{2}+2\cdot\frac{\nu}{4}\leqslant\nu,$$

тогда как его трудоемкость не превышает

$$Q_{l}(n,\ \nu,\ \alpha_\nu)\,M(n,\ \nu)\leqslant R_r(n)\Big]\ln^2\frac{1}{\nu}\Big[\ \Big/\ \nu^{\frac{\bar r}{\bar r-1}},$$

где $R_r(n)$ — полином по n. Это и есть результат, обещанный в начале пункта. Из приведенных рассуждений ясно также, что требование $\nu_0=0$ было не слишком существенно. Достаточно выполнения неравенств типа $\nu_0\leqslant \bar P_r^{-1}(n)\,\nu$, где $\bar P_r(n)$ — подходящий полином от n.

Таким образом, описанный метод (с теоретической точки зрения) менее чувствителен к смещенности оракула (в асимптотике по $\nu\to 0$ при фиксированном n), чем метод из § 3.

4.3. Подведем некоторые итоги. Результаты, приведенные выше, в гораздо меньшей степени отвечают на вопросы, чем вынуждают ставить новые. Действительно, ни один из описанных выше методов не может претендовать на практическую применимость, за исключением самых простых ситуаций (очень грубой точности или совсем небольшой размерности). Ведь приемлемый с практической точки зрения метод должен, во-первых, иметь оценку трудоемкости, не слишком быстро растущую с ростом размерности задачи и повышением требуемой точности решения и, во-вторых, быть не слишком чувствительным к ошибкам в информации о задаче.

Говоря об ошибках, мы не относим к ним «регулярно» устроенные ошибки (случайные ошибки с нулевым средним и независимые от шага к шагу: с такими ошибками мы умеем бороться, специально ориентируя на них наши методы). Речь идет об ошибках, относительно природы которых мы не вправе делать никаких априорных предположений, кроме естественной гипотезы об их малости по сравнению с требуемой точностью решения. В § 2 только такие ошибки и рассматривались, а в §§ 3—4 роль таких ошибок играли смещения оракула. Ясно, что определенные гипотезы о малости ошибок такого рода необходимы — если они превышают требуемую точность решения задачи, то легко понять, что решить ее с этой точностью, вообще говоря, невозможно.

Естественно измерять помехоустойчивость какого-либо метода отношением предельно допустимого для него уровня ошибок (допустимого в том смысле, что наличие ошибок не приводит к существенной потере в точности решения) к требуемой точности решения. Приведенное словесное «определение» помехоустойчивости достаточно расплывчато, но мы не будем загромождать изложение точными определениями. Ясно, что методы

с малой помехоустойчивостью, как правило, не могут быть использованы на практике.

Как же обстоит дело с трудоемкостью и помехоустойчивостью описанных выше методов нулевого порядка? К сожалению, довольно скверно. Трудоемкость методов §§ 2 и 4 весьма быстро растет с ростом размерности задачи, тогда как трудоемкость методов § 3 слишком быстро растет с ростом точности. Методы, построенные по схеме § 1, казалось бы, не имеют этого недостатка; однако их помехоустойчивость крайне мала (допустимый для них уровень ошибок должен быть порядка четвертой степени требуемой точности). Вообще, помехоустойчивость рассмотренных методов оставляет желать лучшего: даже теоретически устойчивые (в асимптотике по $\nu\to 0$) методы §§ 2 и 4 имеют помехоустойчивость, быстро падающую с ростом размерности задачи.

Можно было бы предположить, что эти трудности не являются дефектами исследуемых методов, а объективно присущи всем методам нулевого порядка. Однако у нас нет данных, чтобы утверждать это; скорее имеются основания утверждать обратное. Действительно, в случае детерминированного оракула мы имеем неустойчивые методы § 1 с приемлемой (в асимптотике по $n\to\infty$ и по $\nu\to 0$) трудоемкостью (во всяком случае, грубые оценки сложности соответствующего класса задач при оракуле нулевого порядка показывают, что особенно сильно снизить трудоемкость этих методов в принципе нельзя). С другой стороны, мы имеем качественно более устойчивый метод § 2, трудоемкость которого неприемлемо быстро растет с ростом n. Естественно предположить, что должны существовать методы, соединяющие оценки трудоемкости методов § 1 со свойствами устойчивости, по крайней мере, того же типа, что и у методов § 2. Было бы даже желательно построить методы, еще более помехоустойчивые, чем метод § 2. Последний устойчив скорее в теоретическом, чем в практическом плане: его помехоустойчивость быстро падает с ростом размерности. В этом смысле идеальной является ситуация с методами первого порядка решения общих (липшицевых) выпуклых задач из глав II, III, VI. Помехоустойчивость последних не зависит от размерности или каких-либо других параметров классов указанного типа.

К сожалению, для методов нулевого порядка такое идеальное положение невозможно: их помехоустойчивость необходимо падает с ростом размерности (для всех достаточно широких классов задач), если только трудоемкость методов остается в приемлемых границах. Обоснование этого последнего утверждения содержится в упражнении, заключающем параграф. Отметим, правда, что определяемая им оценка скорости падения помехоустойчивости с ростом размерности куда ниже, чем та, которую умеет реализовать метод § 2.

В стохастической ситуации §§ 3—4 дело обстоит примерно так же. Методы § 3 имеют трудоемкость, приемлемую по характеру зависимости от n и неприемлемую по характеру зависимости от ν, а для метода § 4 все обстоит наоборот. Естественно предположить, что существует метод, трудоемкость которого по характеру зависимости от n хотя бы та же, что и у методов § 3, а по характеру зависимости от ν — та же, что и у метода § 4.

Таким образом, полученные результаты не столько решают проблему выбора методов выпуклой оптимизации при оракуле нулевого порядка, сколько показывают, чего следует ожидать от удовлетворительного ее решения (этим, собственно, и исчерпывается их значение). Каждый из предложенных методов в каком-то отношении существенно не улучшаем, однако плох в некоторых других отношениях. Нам не удается синтезировать их достоинства, освободившись от недостатков. Можно ли это сделать, и если можно, то как,— нам неизвестно. Таким образом, и в теоретическом, и тем более в практическом отношении ситуация с оракулом нулевого порядка, по-видимому, далека от ясности.

Для сравнения заметим, что в случае оракула первого порядка дело обстоит существенно лучше (по крайней мере, в теоретическом плане). Там мы в состоянии, как правило, указать методы с верхними оценками трудоемкости, по порядку совпадающими с нижними оценками сложности соответствующих классов. С устойчивостью этих методов по отношению к ошибкам в информации о задаче дело также обстоит сравнительно благополучно. Классы общих (липшицевых) выпуклых задач в главах II, III, V, VI с самого начала рассматривались для приближенных детерминированных (смещенных стохастических) оракулов. При этом относительный уровень ошибок (измерявшийся там величиной ν_0), пропорциональный требуемой относительной точности ν (скажем, вдвое *) ее меньший), не приводил к потере устойчивости: метод, ориентированный на точный оракул и точность $\nu/2$, обеспечивал и при указанном приближенном оракуле точность ν. Помехоустойчивость (в естественном определении) этих методов, стало быть, абсолютная константа.

При рассмотрении классов строго выпуклых задач (глава VII) ситуация с помехоустойчивостью была несколько хуже. Мы предполагали там оракул точным; однако анализ построенных методов показывает, что некоторый ненулевой уровень ошибок все же допустим. Для основного из построенных в главе VII методов — метода решения безусловных сильно выпуклых задач во всем пространстве — этот уровень (в естественной нормировке) оказывается порядка $\nu/D(Q)$, где ν — требуемая относительная точность решения, Q — модуль строгой выпуклости задачи,

*) В главе VI дело обстоит несколько иначе — «вдвое» уже не годится, константа должна быть увеличена.

$D(Q)$ — подходящая функция Q. Помехоустойчивость оказывается тем меньшей, чем хуже обусловленность задачи (т. е. чем больше Q). Это и естественно: вся специфика класса связана с регулярным поведением производной минимизируемой функции. Ясно, что ошибки в наблюдении производной должны быть достаточно малы, если мы желаем, чтобы от этой регулярности еще что-то осталось. Таким образом, качественно в зависимости помехоустойчивости от обусловленности нет ничего странного.

4.4. Упражнение 1. Пусть A_n — множество «линейных» задач

$$f(x) \equiv \langle\varphi|x\rangle \to \min|\ \|x\| \leqslant 1,\ x \in E^n,$$

порожденных векторами φ с $\|\varphi\| \leqslant 1$. Снабдим A_n оракулом $\mathcal{O}_\delta$, обслуживающим лишь точки $V_n = \{x|\ \|x\| \leqslant 1\}$ и таким, что

$$\mathcal{O}_\delta(f,\ x) = \begin{cases} f(x), & |f(x)| > \delta, \\ 0, & |f(x)| \leqslant \delta. \end{cases}$$

Тогда $\mathcal{O}_\delta$ — оракул нулевого порядка абсолютной погрешности δ. Предположим, что необходимо решать задачи класса A_n^δ, получающегося оснащением A_n оракулом $\mathcal{O}_\delta$, и требуется обеспечить точность ε в решении этих задач. Докажите, что трудоемкость M любого (детерминированного) метода $\mathcal{B}$, обеспечивающего точность ε в решении задач класса A_n^δ, удовлетворяет оценке

$$M \geqslant c \exp\left\{dn\left(\frac{\delta}{\varepsilon}\right)^2\right\},$$

где $c,\ d > 0$ — абсолютные константы. Выведите отсюда, что при степенном росте трудоемкости с ростом n необходимо должно быть $\delta \leqslant O(\sqrt{\ln n/n})\,\varepsilon$. Таким образом, помехоустойчивость методов нулевого порядка с не слишком быстро растущей при увеличении n трудоемкостью необходимо должна падать с ростом n.

⟨Пусть $x_1, \ldots, x_M$ — траектория $\mathcal{B}$ на задаче $f \equiv 0$. Рассмотрим меру $\bar\mu$, сосредоточенную в точках $x_1, \ldots, x_M$. Известно, что для всякой вероятностной меры μ на шаре V_n и всякого $\varkappa > 0$ найдется единичный вектор $\varphi_{\mu,\varkappa}$, такой, что при

$$\Pi_\varkappa(\varphi) = \{x \in V_n \mid |\langle x|\varphi\rangle| > \varkappa\}$$

имеем

$$\mu(\Pi_\varkappa(\varphi_{\mu,\varkappa})) \leqslant \bar c e^{dn\varkappa^2},$$

$\bar c,\ d > 0$ — абсолютные константы. Положим $\varkappa = \delta/2\varepsilon$ и $\bar\varphi = \varphi_{\bar\mu,\varkappa}$. Тогда $\bar\mu(\Pi_\varkappa(\bar\varphi)) \leqslant \bar c \exp\{-dn\varkappa^2\}$, и при

$$\bar c e^{-dn\varkappa^2} < 1/M \qquad (4.1)$$

$\Pi_\varkappa(\bar\varphi)$ не содержит точек $x_1, \ldots, x_M$. Пусть (4.1) выполнено. Рассмотрим пару задач $f_+(x) = 2\varepsilon\langle x|\bar\varphi\rangle$ и $f_-(x) = -2\varepsilon\langle x|\bar\varphi\rangle \in A_n$. Вдоль траектории $\mathcal{B}$ на задаче $f \equiv 0$ информация, поставляемая $\mathcal{O}_\delta$, о f_+ и f_- нулевая, ибо вдоль этой траектории $x_1, \ldots, x_M$ по построению имеем $|2\varepsilon\langle\bar\varphi|x_i\rangle| \leqslant 2\varepsilon\varkappa = \delta$. Стало быть, x_M есть результат применения $\mathcal{B}$ и к f_+, и к f_-. Но тогда, очевидно, для одной из этих задач он не является приближенным решением погрешности ε, что противоречит определению $\mathcal{B}$. Итак, (4.1) неверно, т. е.

$$M \geqslant c e^{\frac{dn}{4}\left(\frac{\delta}{\varepsilon}\right)^2},$$

что и требуется.⟩

П р и л о ж е н и е

МАТЕМАТИЧЕСКИЕ ДОПОЛНЕНИЯ

Этот раздел является вспомогательным. В нем дана сводка определений и свойств некоторых относительно специальных математических понятий, использованных в книге, а также объясняются некоторые часто применяемые обозначения.

§ 1. Польские пространства, борелевы функции, меры

Предполагается знакомство с понятиями метрических пространств (в частности, компактов), открытых и замкнутых множеств, сходимости и т. п. в объеме, скажем [13], гл. I, II.

1.0. Основные обозначения, связанные с метрикой. Если X — метрическое пространство, G — подмножество X, то через $\operatorname{int} G$ обозначается *внутренность* G — множество точек $\{x \mid \exists\rho > 0\colon \{y \mid \rho(x, y) \leqslant \rho\} \subset G\}$. Через $\overline{G}$ обозначается *замыкание* G, через ∂G — граница G, т. е. множество $\overline{G} \setminus \operatorname{int} G$. Через $\rho(x, G)$ обозначается расстояние от x до G, т. е. величина

$$\rho(x, G) = \inf\{\rho(x, y) \mid y \in G\}.$$

Радиусом G (обозначение $\rho(G)$) называется нижняя грань радиусов шаров с центрами в G, содержащих G:

$$\rho(G) = \inf\{r \mid \exists x \in G\colon \forall y \in G\ \rho(x, y) \leqslant r\}.$$

Диаметром G называется величина

$$d(G) = \sup\{\rho(x, y) \mid x, y \in G\}.$$

Асферичностью G называется, грубо говоря, отношение радиусов минимального содержащего G и максимального содержащегося в G шаров:

$$\alpha(G) = \inf\{\alpha \mid \exists x, y, r\colon \{z \mid \rho(z, x) \leqslant r\} \subset \subset G \subset \{z \mid \rho(z, y) \leqslant \alpha r\}\}.$$

1.1. Польские пространства. Метрическое пространство X называется *польским*, если оно сепарабельно и полно. Класс всех польских пространств обозначается Π. Фраза «X — польское пространство» записывается как $X \in \Pi$. Метрика на X обозначается $\rho_X(\cdot, \cdot)$ (иногда просто $\rho(\cdot, \cdot)$, если из контекста ясно, о каком X идет речь). Примерами польских пространств могут служить метрические компакты, пространства R^n с обычной метрикой и сепарабельные банаховы пространства (по поводу последних см. ниже, в § 2).

Другие примеры можно получить, применяя к польским пространствам операцию перехода к замкнутому подпространству (замкнутое подпространство польского пространства — само польское пространство) и операцию *прямого умножения*. Пусть $X_i \in \in \Pi$, $i = 1, 2, \ldots$ Рассмотрим теоретико-множественное произведение $X^\infty = \prod_{i=1}^{\infty} X_i$ пространств X_i (точка X^∞ — это последовательность $x^\infty = (x_1, x_2, \ldots)$, $x_i \in X_i$). На X^∞ можно определить метрику

$$\rho_{X^\infty}(x^\infty, y^\infty) = \sum_1^\infty 2^{-i} \frac{\rho_{X_i}(x_i, y_i)}{1 + \rho_{X_i}(x_i, y_i)}.$$

Сходимость в этой метрике есть в точности «покоординатная» сходимость. Далее, говоря о прямых произведениях, всегда будем иметь в виду описанную конструкцию (разумеется, применимую и к конечному числу сомножителей).

1.2. Борелевы функции. По поводу последующего см. [5, 26].

1.2.1. Пусть X — некоторое множество, Σ — некоторое семейство его подмножеств. Σ называется *σ-алгеброй*, если

(1) $X \in \Sigma$;

(2) из $A \in \Sigma$ следует $X \setminus A \in \Sigma$;

(3) из $A_1, A_2, \ldots \in \Sigma$ следует $\bigcup_{i=1}^{\infty} A_i \in \Sigma$.

Пусть $\{\Sigma_\alpha\}_{\alpha \in J}$ — некоторое семейство σ-алгебр подмножеств X. Рассмотрим пересечение всех Σ_α по $\alpha \in J$. Это также некоторое семейство подмножеств X; очевидно, оно является σ-алгеброй. В частности, для любого семейства Σ подмножеств X существует наименьшая σ-алгебра $\overline{\Sigma}$, содержащая Σ (именно, пересечение всех содержащих Σ σ-алгебр). $\overline{\Sigma}$ называется *σ-алгеброй, натянутой на* Σ.

1.2.2. Пусть теперь $X \in \Pi$. *Борелевой σ-алгеброй* $B(X)$ *подмножеств* X называется σ-алгебра, натянутая на семейство всех замкнутых (или, что то же, всех открытых) подмножеств X. Элементы $B(X)$ называются *борелевыми подмножествами* X.

Пусть $X, Y \in \Pi$ и $f\colon X \to Y$ — отображение. *Отображение* f называется *борелевым*, если полный прообраз $f^{-1}(Y)$ всякого открытого (а тогда — и любого борелева) подмножества Y — борелево подмножество X. В частности, непрерывное отображение X в Y борелево.

Следующие свойства борелевых отображений легко следуют из определения.

Предложение. Пусть Y_i, $X_i \in \Pi$, $i = 0, 1, 2, \ldots$

(1) Если f_0: $X_0 \to X_1$ и f_1: $X_1 \to X_2$ борелевы, то таково же и $f_1 \circ f_0$: $X_0 \to X_2$ $(f_1 \circ f_0(x) = f_1(f_0(x)))$.

(2) Если f_i: $X_0 \to X_1$ борелевы и при всех $x \in X_0$ существует $\lim\limits_{i\to\infty} f_i(x) \equiv f(x)$, то f: $X_0 \to X_1$ борелево.

(3) Если f_i: $X_i \to Y_i$ борелевы, $1 \leqslant i < \infty$, то отображение $f^\infty: \prod\limits_{i=1}^{\infty} X_i \to \prod\limits_{i=1}^{\infty} Y_i$, определенное как $f^\infty(x_1, x_2, \ldots) = (f_1(x_1), f_2(x_2), \ldots)$, борелево.

(4) Отображение $f: X_0 \to \prod\limits_{i=1}^{\infty} X_i$ борелево тогда и только тогда, когда борелевы все отображения $\pi_N \circ f: X_0 \to \prod\limits_{i=1}^{N} X_i$, где $\pi_N(x^\infty) = (x_1, \ldots, x_N)$ — естественные проекции $\prod\limits_1^N X_i$ на $\prod\limits_1^N X_i$.

(5) Пусть $A_i \in B(X_0)$, $A_i \cap A_j = \phi$ при $i \neq j$, $\bigcup\limits_{i=1}^{\infty} A_i = X_0$. Пусть, далее, f_i: $X_0 \to X_1$ борелевы. Рассмотрим отображение f: $X_0 \to X_1$, определенное так: сужение f на A_i совпадает с сужением f_i на A_i, $1 \leqslant i < \infty$. Тогда f борелево.

Перечисленные в предложении факты показывают, что «естественные» операции над борелевыми функциями не выводят из класса борелевых функций. С их помощью легко доказывается борелевость всех правил, строящихся в доказательствах теорем главы I, или борелевость правил, связанных с методами из глав V—VI. Приведем модельное рассуждение такого рода. Пусть X, Y, $Z \in \Pi$ и $f(x, y)$: $X \times Y \to Z$ — борелево отображение, и g: $X \to Y$ также борелево. Требуется доказать, что функция $\bar{f}(x) = f(x, g(x))$ также борелева. Действительно, рассмотрим последовательность отображений

$$\varphi_0:\ X \to X \times X\ (\varphi_0(x) = (x, x)),$$

$$\varphi_1:\ X \times X \to X \times Y\ (\varphi_1(x, x') = (x, g(x'))),\quad f:\ X \times Y \to Z.$$

Очевидно, все они борелевы (φ_0 непрерывно, φ_1 — в силу п. 3 предложения, f — по условию). Ввиду п. 1 предложения борелево и $f \circ \varphi_1 \circ \varphi_0 \equiv \bar{f}$.

1.2.3. Специально рассмотрим скалярные борелевы функции. Пусть $X \in \Pi$ и $B(X, R)$ — множество борелевых функций на X со значениями на вещественной оси. Иногда полезна характеризация $B(X, R)$ как минимального пространства функций $\bar{B}$, содержащего все непрерывные ограниченные функции *) и замкнутого относительно секвенциальной поточечной сходимости (т. е. $B(X, R)$ — такое, что из $f_i \in A$, $i = 1, 2, \ldots$, и $\lim\limits_{i\to\infty} f_i(x) \equiv f(x)$ для всех x следует $f(x) \in A$). С другой стороны, $A \subset X$ борелево тогда и только тогда, когда характеристическая функция множества A

$$\chi_A(x) = \begin{cases} 0, & x \notin A, \\ 1, & x \in A, \end{cases}$$

лежит в $B(X, R)$. Таким образом, можно дать новое определение борелевых множеств, определив их через понятие борелевой скалярной функции (последние, как мы видели, можно определить независимо от понятия борелевых множеств).

1.2.4. Введем еще одно полезное понятие. Пусть $T(X) \subset B(X, R)$ — семейство ограниченных борелевых функций. Обозначим через $\bar{T}(X)$ минимальное содержащее $T(X)$ множество функций, содержащее линейную оболочку $T(X)$ и такое, что если $f_i \in \bar{T}(X)$, $|f_i| \leqslant c$ для некоторого $c < \infty$, и $f_i(x) \xrightarrow[i\to\infty]{} f(x)$ при всех $x \in X$, то $f \in \bar{T}(X)$. Скажем, что $T(X)$ *порождает* $C_B(X)$, если $\bar{T}(X) \supset C_B(X)$. Например, $C_B(X)$ порождает само себя. Можно доказать, что если $X_i \in \Pi$, $i = 1, 2, \ldots$, то множество функций на X^∞ вида $\prod\limits_{i=1}^{k} f_i(x_i)$, где k — произвольное натуральное число, $f_i \in C_B(X_i)$, порождает $C_B(X^\infty)$, $X^\infty = \prod\limits_1^\infty X_i$.

1.3. Меры.

1.3.1. Пусть X — множество, Σ — σ-алгебра подмножеств X. Пара (X, Σ) называется *измеримым пространством*. *Мерой* (вернее, *σ-аддитивной мерой*) на (X, Σ) называется неотрицательная функция μ, определенная на подмножествах X, лежащих в Σ (принимающая, возможно, и значение $+\infty$), обладающая следующим свойством аддитивности: если A_i, $i = 1, 2, \ldots$, лежат в Σ и $A_i \cap A_j = \phi$ при $i \neq j$, то $\mu(\bigcup A_i) = \sum\limits_{i=1}^{\infty} \mu(A_i)$ (если ряд справа расходится, то его сумма считается равной $+\infty$). Мера μ называется *σ-конечной*, если все пространство X есть объединение счетного числа множеств конечной меры. Мера μ называется *полной по Лебегу*, если из $A \in \Sigma$, $\mu(A) = 0$ и $B \subset A$ следует $B \subset \Sigma$ (тогда $\mu(B) = 0$ автоматически). Мера μ называется *вероятностной*, если $\mu(X) = 1$ (тогда для всех $A \in \Sigma$ справедливо неравенство $0 \leqslant \mu(A) \leqslant 1$).

*) Пространство которых мы обозначим $C_B(X)$.

1.3.2. Пусть $\bar{R}$ означает вещественную ось, пополненную точками $+\infty$ и $-\infty$. Пусть (X, Σ) — измеримое пространство и $f\colon X \to \bar{R}$ — функция. Она называется *σ-измеримой*, если прообраз любого множества вида $\{t > a\}$ на $\bar{R}$ лежит в Σ (тогда таков же прообраз любого борелева множества на оси). Каждую такую функцию можно попытаться следующим образом проинтегрировать по мере μ. Пусть вначале $f \geqslant 0$. Назовем *простой функцией* φ неотрицательную функцию, (X, Σ)-измеримую и принимающую конечное число значений. Пусть $t_1, \ldots, t_r$ — эти значения и $A_i = \{x \in X \mid \varphi(x) = t_i\}$. Положим

$$\int \varphi(x)\, d\mu(x) = \sum_{i=1}^{r} t_i \mu(A_i)$$

(здесь $0 \cdot (+\infty) = 0$). (X, Σ)-измеримая функция $f \geqslant 0$ называется *суммируемой*, если она есть предел (во всех точках X, кроме, возможно, точек множества меры нуль) неубывающей последовательности простых функций φ_i, с конечным $\lim\limits_{i\to\infty} \int \varphi_i(x)\, d\mu(x)$. Этот предел называется *интегралом от f по мере* μ и обозначается $\int f(x)\, d\mu(x)$ (определение корректно: доказывается, что предел не зависит от выбора $\{\varphi_i\}$ с указанными свойствами). Функция $f\colon X \to \bar{R}$, (X, Σ)-измеримая, называется *суммируемой*, если суммируемы обе функции $f_+ = \max\{f, 0\}$ и $f_- = -\min\{f, 0\}$. При этом интеграл $\int f(x)\, d\mu(x)$ определяется как

$$\int f_+(x)\, d\mu(x) - \int f_-(x)\, d\mu(x).$$

Пусть $L(X, \mu)$ — пространство μ-суммируемых функций. Оказывается, что это — линейное пространство (относительно естественных поточечных операций), содержащее вместе с каждой функцией f и ее модуль $|f|$;

$$I(f) = \int f(x)\, d\mu(x)$$

— линейный функционал на этом пространстве. Он неотрицателен ($I(f) \geqslant 0$ при $f \geqslant 0$) и непрерывен в следующем смысле: пусть $f_i \in L(X, \mu)$ и $|f_i| \leqslant f$, $f \in L(X, \mu)$. Пусть, далее, $f_i(x) \underset{i\to\infty}{\longrightarrow} f(x)$ μ-почти везде *). Тогда $f(x) \in L(X, \mu)$ и $I(f) = \lim\limits_{i\to\infty} I(f_i)$ (теорема Лебега). Кроме того, если $|f| \leqslant g \in L(X, \mu)$ и f (X, Σ)-измерима, то $f \in L(X, \mu)$.

*) Говорят, что некоторое свойство, зависящее от точки пространства (X, Σ) с мерой μ, выполнено μ-почти наверное (μ-почти везде), если оно выполнено для всех $x \in X \setminus A$, где $A \in \Sigma$ и $\mu(A) = 0$.

1.3.3. Меры на польских пространствах. Пусть (X, Σ, μ) — измеримое пространство с σ-аддитивной вероятностной мерой μ и $X \in \Pi$. Мера μ (точнее, тройка X, Σ, μ) называется *борелевой*, если $B(X) \subset \Sigma$ (т. е. мера определена для всех борелевых множеств). Мера μ называется *регулярной*, если из $A \in \Sigma$ следует, что существуют *замкнутые* множества $A_i \subset A$, такие, что $\mu(A) = \lim\limits_{i\to\infty} \mu(A_i)$ *). Известно, что всякая (вероятностная) σ-аддитивная мера на $B(X)$, $X \in \Pi$, регулярна. Далее, слово «мера» применительно к польским пространствам X означает «полная по Лебегу регулярная борелева вероятностная мера на X».

Пусть X — польское пространство, и пусть $C_B(X)$, как и выше, означает пространство непрерывных ограниченных функций на X со значениями в $\mathbb{R}$. Если μ — мера на X, то функционал

$$I(f) = \int f(x)\, d\mu(x)$$

на $C_B(X)$ обладает следующими свойствами:

(1) $I(f)$ линеен;

(2) $I(f) \geqslant 0$ при $f \geqslant 0$;

(3) $I(1) = 1$ (здесь 1 — функция, тождественно равная 1);

(4) если $f_i \in C_B(X)$ и $f_1 \geqslant f_2 \geqslant \ldots$, $\lim\limits_{i\to\infty} f_i(x) = 0$, то $\lim\limits_{i\to\infty} I(f_i) = 0$ (теорема Лебега).

Оказывается, верно и обратное: всякий функционал на $C_B(X)$, удовлетворяющий условиям (1) — (4), порождается некоторой мерой μ по формуле

$$I(f) = \int f(x)\, d\mu(x).$$

При этом различным мерам отвечают различные функционалы.

1.3.4. Композиции мер. Пусть X и Y — польские пространства и $\Phi_{y|x}$ — семейство мер на Y, зависящее от параметра $x \in X$. Это семейство называется *борелевым* по x, если для каждой $f \in C_B(Y)$ функция

$$\tilde{f}(x) = \int f(y)\, d\Phi_{y|x}$$

борелева по x. Если $\Phi_{y|x}$ борелево по x и f — ограниченная скалярная борелева функция на $X \times Y$, то

$$\tilde{f}(x) = \int f(x, y)\, d\Phi_{y|x}$$

борелево по x. Пусть теперь Φ_x — мера на X и $\Phi_{y|x}$ — борелево по x семейство мер на Y. Тогда для каждой функции

*) Тогда существуют и компакты A_i с тем же свойством.

$f \in C_B(X \times Y)$ имеет смысл функционал

$$\int \left\{ \int f(x, y)\, d\Phi_{y|x} \right\} d\Phi_x.$$

Используя теорему Лебега, легко проверить, что этот функционал обладает свойствами (1) — (4) и, следовательно, порождается некоторой мерой. Эта мера обозначается $\Phi_x \otimes \Phi_{y|x}$. Пусть, в частности, $\Phi_y \equiv \Phi_{y|x}$ не зависит от x. Тогда $\Phi_x \otimes \Phi_{y|x}$ есть мера, сопоставляемая паре мер Φ_x на X и Φ_y на Y. Итак, по паре польских пространств с мерами (X, Φ_x) и (Y, Φ_y) можно построить третье польское пространство $(X \times Y, \Phi_x \otimes \Phi_y)$, оно называется *произведением пространств с мерой* (X, Φ_x) и (Y, Φ_y).

Обратно, пусть $\Phi_{x,y}$ — мера на $X \times Y$. Построим по ней меру Φ_x на X по формуле $\Phi_x(A) = \Phi_{x,y}(A \times Y)$ (для всех A, для которых правая часть определена). Легко видеть, что получится действительно мера на X. Она называется *мерой на* X, *индуцированной* $\Phi_{x,y}$. Оказывается, что существует борелево по $x \in X$ семейство мер $\Phi_{x|y}$ на Y, такое, что $\Phi_x \otimes \Phi_{y|x} = \Phi_{x,y}$. Φ_x, $\Phi_{y|x}$ определены по $\Phi_{x,y}$ «почти однозначно»: если Φ_x, $\Phi_{y|x}$ и $\widetilde{\Phi}_x$, $\widetilde{\Phi}_{y|x}$ таковы, что $\Phi_x \otimes \Phi_{y|x} = \widetilde{\Phi}_x \otimes \widetilde{\Phi}_{y|x}$, то $\Phi_x = \widetilde{\Phi}_x$ и $\Phi_{y|x} = \Phi_{y|x}$. Φ_x-почти для всех x. Пусть теперь $X_i \subset \Pi$, $X^i = \prod_{j=1}^{i} X_j$, $X^\infty = \prod_{j=1}^{\infty} X_j$. Предположим, что на пространствах X^i заданы меры Φ_{x^i}, согласованные в следующем смысле: X^i является естественным прямым сомножителем в X^j, $j > i$ (X^i отождествляется с набором первых i сомножителей среди j сомножителей, произведение которых есть X^j), так что можно рассмотреть меры $\Phi^j_{x^i}$, индуцированные Φ_{x^j} на X^i. *Согласованность мер* Φ_{x^i} означает, что $\Phi^j_{x^i} = \Phi_{x^i}$ при всех i и $j > i$.

Примером такого семейства может служить семейство мер, индуцированных на X^i мерой Φ_{x^∞} на X^∞. Оказывается, что никаких других примеров и нет: по классической теореме Колмогорова всякое согласованное в указанном смысле семейство мер порождается описанным образом некоторой (однозначно определенной семейством) мерой на X^∞. Этот факт был использован в п. 3.3.2 гл. I.

1.3.5. О совпадении мер. Иногда полезно иметь удобные достаточные условия совпадения двух мер. Пусть X — польское пространство и $T(X)$ — какое-нибудь семейство ограниченных борелевых функций на X, порождающее $C_B(X)$. Пусть Φ_x и $\widetilde{\Phi}_x$ — две меры на X. Чтобы доказать, что $\Phi_x = \widetilde{\Phi}_x$, достаточно проверить, что

$$\int f(x)\, d\Phi_x = \int f(x)\, d\widetilde{\Phi}_x$$

для всех $f \in T(X)$ (действительно, тогда семейство функций $f \in B(X, R)$, для которых

$$\int f(x)\, d\Phi_x = \int f(x)\, d\widetilde{\Phi}_x$$

содержит $\overline{T}(X)$, т. е. $C_B(X)$, так что $\Phi_x = \widetilde{\Phi}_x$).

Приведем пример. Пусть $X, Y \in \Pi$ и $\Phi_{x,y}$; Φ_x; $\Phi_{y|x}$ — меры (семейство мер) на соответствующих пространствах. Пусть мы желаем доказать, что $\Phi_{x,y} = \Phi_x \otimes \Phi_{y|x}$. Для этого достаточно проверить, что для любых $f \in C_B(X)$ и $g \in C_B(Y)$

$$\int f(x) g(y)\, d\Phi_{x,y} = \int f(x) \left\{ \int g(y)\, d\Phi_{y|x} \right\} d\Phi_x$$

(действительно, в п. 1.2.4 отмечалось, что множество функций вида $f(x)\, g(y)$, $f \in C_B(X)$, $g \in C_B(Y)$, порождает $C_B(X \times Y)$).

§ 2. Банаховы пространства

Предполагаются известными основные понятия и факты линейной алгебры и основные определения, относящиеся к гильбертовым пространствам, в объеме, скажем, [13], гл. III, IV.

2.1. Пусть E — линейное вещественное пространство. Функция $\| x \|$ на E, принимающая вещественные значения, называется *нормой*, если

(1) $\| x \| \geqslant 0$, причем $\| x \| = 0 \Leftrightarrow x = 0$;

(2) $\| \lambda x \| = | \lambda | \, \| x \|$, $\lambda \in R$;

(3) $\| x + y \| \leqslant \| x \| + \| y \|$.

Пространство E с нормой $\| \cdot \|$ называется *нормированным*. Норма $\| \cdot \|$ порождает метрику на E: $\rho(x, y) = \| x - y \|$. При этом линейные операции в E оказываются непрерывными в соответствующей метрике.

Каждая норма $\| \cdot \|$ на E определяет некоторое множество — единичный шар E $V = \{x \mid \| x \| \leqslant 1\}$. Множество V обладает следующими свойствами:

(1) V выпукло (т. е. из $x, y \in V$ и $\theta \in [0, 1]$ следует $\theta x + (1 - \theta) y \in E$);

(2) V уравновешено ($-V \equiv \{x = -u \mid u \in V\} = V$);

(3) V поглощающее (т. е. $\forall x \in E \ \exists \lambda > 0$; $\lambda x \in V$);

(4) V ограничено по лучам (т. е. $\forall x \neq 0 \ \exists \lambda > 0$: $\lambda x \notin V$);

(5) V замкнуто по лучам (т. е. $\forall x \in E$ множество $\{\lambda \mid \lambda x \in V\}$ замкнуто на оси).

Обратно, каждое множество $V \subset E$ со свойствами (1) — (5) есть единичный шар некоторой (однозначно определенной) нормы — именно, нормы

$$\| x \| = \inf \{\lambda > 0 \mid \lambda^{-1} x \in V\}.$$

2.2. Норма на E, как уже говорилось, задает метрику (и, следовательно, выделяет класс сходящихся последовательностей элементов E). Две нормы на E называются эквивалентными, если они порождают одну и ту же топологию на E (т. е. запас сходящихся последовательностей в обеих нормах один и тот же). Можно доказать, что две нормы $\|\cdot\|$ и $\|\cdot\|'$ эквивалентны тогда и только тогда, когда единичный шар одной (каждой) из них — ограниченное множество в другой, или, что то же, для некоторого $c>0$

$$\frac{1}{c}\|x\|' \geqslant \|x\| \geqslant c\|x\|' \quad \text{для всех } x \in E.$$

2.3. Пространство E с нормой $\|\cdot\|$ называется *банаховым*, если E как метрическое пространство с метрикой $\|x-y\|$ полно. Банаховость — свойство топологии нормированного пространства, а не конкретного способа нормирования (т. е. если E, $\|\cdot\|$ — банахово пространство и $\|\cdot\|'$ эквивалентно $\|\cdot\|$, то E, $\|\cdot\|'$ — тоже банахово).

2.4. Теорема Хана — Банаха. Пусть E — линейное пространство и $p(h)$ — вещественнозначная функция на E. Функция $p(h)$ называется *сублинейной функцией*, если

(1) $p(\lambda h)=\lambda p(h)$, $\lambda \geqslant 0$;

(2) $p(h_1+h_2)\leqslant p(h_1)+p(h_2)$

(иначе говоря, $p(h)$ выпукло).

Для линейного (и выпуклого) анализа следующая теорема играет фундаментальную роль.

Теорема Хана — Банаха. *Пусть E — линейное пространство, E_1 — его подпространство, $p(h)$ — сублинейная функция на E. Пусть, далее, f — линейный функционал на E_1, такой, что $\langle f|h\rangle \leqslant p(h)$ для всех $h\in E_1$ (здесь и далее $\langle f|x\rangle$ — значение линейного функционала f на векторе x). Тогда существует линейный функционал $\bar f$ на всем E, совпадающий с f на E_1, и такой, что $\langle \bar f|h\rangle \leqslant p(h)$ для всех $h\in E$.*

2.5. Сопряженное пространство. Пусть E, $\|\cdot\|$ — нормированное пространство, а f — линейный функционал на E. Функционал f называется *непрерывным*, если $\langle f|x\rangle$ — непрерывная функция $x\in E$ (достаточно, чтобы она была непрерывна в 0). Эквивалентное определение состоит в том, что функция $\langle f|x\rangle$ ограничена на единичном шаре E. Совокупность всех линейных непрерывных функционалов на E с естественными линейными операциями сама образует линейное пространство. Оно называется *сопряженным* к E, $\|\cdot\|$ и обозначается E^*. На E^* вводится норма

$$\|f\|_* = \sup\{\langle f|x\rangle | x\in E,\ \|x\|\leqslant 1\}.$$

Относительно этой нормы E^* оказывается банаховым пространством. Важно заметить, что запас элементов (и топология) пространства E^* определяется только запасом элементов и топологией E (т. е. при замене $\|\cdot\|$ на эквивалентную $\|\cdot\|'$ запас элементов и линейная структура E^* не изменяются, а норма $\|\cdot\|_*$ заменится на эквивалентную $(\|\cdot\|')_* = \|\cdot\|'_*$).

Теперь рассмотрим вопрос о запасе элементов E^*. Оказывается, их достаточно много: для каждого $x\in E$ существует $\varphi \in E^*: \|\varphi\|_*=1$ и $\langle\varphi\mid x\rangle = \|x\|$.

⟨Действительно, рассмотрим натянутое на x (можно считать $x\neq 0$) одномерное линейное подпространство E_1. На нем определен линейный функционал $\varphi(\lambda x)=\lambda\|x\|$, $\lambda\in\mathbb{R}$. Норма этого функционала (на E_1) равна 1 и $\varphi(x)=\|x\|$. Остается заметить, что из теоремы Хана — Банаха вытекает следствие.

Следствие. *Если $E_1\subset E$ — подпространство и φ — непрерывный линейный функционал на E_1, то φ можно продолжить, не увеличивая нормы, на все E.*

В самом деле, если a есть норма φ на E_1, то $\langle\varphi\mid h\rangle\leqslant a\|h\|$ при $h\in E_1$. Функция $a\|h\|$ сублинейна на E, и по теореме Хана — Банаха φ продолжается до линейного функционала $\bar\varphi$ на всем E, такого, что $\langle\bar\varphi\mid h\rangle\leqslant a\|h\|$, $h\in E$. Ввиду последнего $\|\bar\varphi\|_*\leqslant a$, что и доказывает, что $\bar\varphi$ — требуемое следствием продолжение φ.⟩

Итак, справедливы «симметричные» соотношения

$$\sup_{x\in E\,|\,\|x\|\leqslant 1}\langle\varphi|x\rangle = \|\varphi\|_*$$

и

$$\sup_{\varphi\in E^*\,|\,\|\varphi\|_*\leqslant 1}\langle\varphi|x\rangle = \|x\|. \tag{2.1}$$

Далее, из определения линейных операций в E^* ясно, что форма $\langle\varphi|x\rangle: E^*\times E\to\mathbb{R}$ при фиксированном $x\in E$ есть линейный функционал на E^*, непрерывный в силу (2.1) (подобно тому, как эта форма при фиксированном φ есть линейный непрерывный функционал на E).

Итак, вектор $x\in E$ порождает линейный непрерывный функционал на E^*. Норма этого функционала есть $\|x\|$. Иными словами, мы построили каноническое изометрическое (сохраняющее норму) вложение E во второе сопряженное пространство $(E^*)^*\equiv E^{**}$. Пространство E называется *рефлексивным*, если образ E при этом вложении есть все E^{**} (т. е. всякий линейный непрерывный функционал на E задается вектором $x\in E$ по формуле $\varphi\to\langle\varphi|x\rangle$). Известно, что если E рефлексивно, то таково же и E^*, и обратно.

2.6. Слабые топологии. Наряду с обычной (задаваемой нормой) топологией банахова пространства E можно рассматривать слабую топологию — слабейшую топологию E, в которой непрерывны все функции вида $x\to\langle\varphi|x\rangle$, $\varphi\in E^*$. Если E^* сепарабельно, то эта топология на *ограниченных* (по норме) подмножествах E может быть задана с помощью метрики

$$\rho(x,y)=\sum 2^{-i}\,|\langle\varphi_i|x-y\rangle|,$$

где $\{\varphi_i\}$ — счетное плотное в единичном шаре E^* семейство функционалов. Линейные операции E непрерывны в слабой топологии. *Выпуклые* замкнутые (по норме) подмножества E замкнуты и в слабой топологии.

На сопряженном к E пространстве E^*, наряду с обычной и слабой топологиями, можно рассматривать и $*$-слабую топологию — слабейшую среди всех топологий, для которых все функции $\varphi \to \langle\varphi|x\rangle$, $x \in E$, непрерывны. Если E сепарабельно, $*$-слабая топология на *ограниченных* подмножествах E^* может быть задана метрикой

$$\rho(\varphi, \psi) = \sum 2^{-i} |\langle\varphi - \psi | x_i\rangle|,$$

где $\{x_i\}$ — счетное плотное в единичном шаре E множество векторов.

Весьма важно, что единичный шар E^* — всегда компакт в $*$-слабой топологии. В частности, если E рефлексивно, то единичный шар E—компакт в слабой топологии E. ⟨Действительно, ясно, что слабая топология E совпадает с $*$-слабой топологией $E^{**} = E$.⟩ Соединяя это обстоятельство с тем фактом, что выпуклые замкнутые (по норме) подмножества E слабо замкнуты, получаем следующий важный результат. Ограниченное выпуклое замкнутое подмножество рефлексивного банахова пространства компактно в слабой топологии.

Отметим еще, что в конечномерном случае все введенные топологии совпадают друг с другом. Всякое конечномерное пространство рефлексивно.

2.7. Прямые произведения банаховых пространств. Пусть E_i, $\|\cdot\|_i$, $1 \leqslant i \leqslant k < \infty$, — банаховы пространства. Их теоретико-множественное прямое произведение $E = \prod_{i=1}^{k} E_i$ наделено естественной линейной структурой (линейные операции происходят покоординатно), а естественная топология прямого произведения задается нормой — скажем, $\|x^k\| = \max_i \|x_i\|_i$, где $x^k = (x_1, \ldots, x_k)$, $x_i \in E_i$. Сопряженное к E пространство есть $\prod_{i=1}^{k} E_i^*$ с нормой $\sum_i \|\varphi_i\|_*$, $\varphi = (\varphi_1, \ldots, \varphi_k)$, $\varphi_i \in E_i^*$. Если E_i рефлексивны, $1 \leqslant i \leqslant k$, то таково же и $\prod_{i=1}^{k} E_i$.

2.8. Примеры. Приведем примеры используемых в книге конкретных банаховых пространств. Все они принадлежат шкале пространств L_p. Эта шкала определяется следующим образом. Пусть (X, Σ, μ) — измеримое пространство с σ-конечной σ-аддитивной полной по Лебегу мерой μ. Будем считать, две скалярные Σ-измеримые функции f_1, f_2 эквивалентными, если $f_1 = f_2$

μ-почти везде. Пусть $M(X, \mu)$ — пространство, элементы которого — классы эквивалентности X, Σ-измеримых скалярных функций с естественными (поточечными) линейными операциями.

Пусть, далее, $1 \leqslant p < \infty$. Пространство $L_p(X, \mu)$ состоит из всех элементов $M(X, \mu)$, для которых конечна величина

$$\|\bar f\|_p = \left\{\int |f(x)|^p d\mu(x)\right\}^{1/p}$$

(слева $\bar f$ означает элемент $M(X, \mu)$; справа f — какой-нибудь представитель класса $\bar f$; от его выбора правая часть не зависит). Для $p = \infty$ пространство $L_\infty(X, \mu)$ определяется как множество элементов $M(X, \mu)$, для которых конечна величина

$$\|\bar f\|_\infty = \operatorname{sup\,vrai} |f(x)| = \inf_{A \in \Sigma,\, \mu(A)=0} \sup_{x \in X \setminus A} |f(x)|.$$

Пространства $L_p(X, \mu)$, $1 \leqslant p \leqslant \infty$, оказываются линейными подпространствами $M(X, \mu)$, банаховыми относительно норм $\|\cdot\|_p$. При $1 \leqslant p < \infty$ сопряженное к $L_p(X, \mu)$ пространство канонически отождествляется с $L_q(X, \mu)$, где $q = p/(p-1)$, т. е.

$$\frac{1}{p} + \frac{1}{q} = 1.$$

При этом отождествлении элемент $\bar f \in L_q$ действует как функционал на вектор $\bar g \in L_p$ по формуле

$$\langle \bar f | \bar g\rangle = \int f(x) g(x) d\mu(x)$$

($f \in \bar f$, $g \in \bar g$). Указанное отождествление сохраняет норму. В частности,

$$\int |f(x) g(x)| d\mu(x) \leqslant \|f\|_q \|g\|_p, \quad \frac{1}{p} + \frac{1}{q} = 1$$

(неравенство Гельдера). Отсюда следует, что если μ — вероятностная мера и $1 \leqslant p \leqslant p' \leqslant \infty$, то $L_p(X, \mu) \supset L_{p'}(X, \mu)$ и $\|f\|_p \leqslant \|f\|_{p'}$ при $f \in L_{p'}$.

Специально выделим случай $p = 2$. При $p = 2$ пространство $L_p(X, \mu)$ гильбертово (со скалярным произведением

$$(\bar f | \bar g) = \int f(x) g(x) d\mu(x), \ f \in \bar f, \ g \in \bar g),$$

$$\|\bar f\|_2 = (\bar f | \bar f)^{1/2}.$$

Всякое гильбертово пространство H канонически отождествляется с сопряженным (вектор $x \in H$ действует на векторы $y \in H$ как функционал по формуле $\langle x|y\rangle = (x|y)$ (справа — скалярное произведение в H)). Это отождествление H и H^* при реализации H в виде $L_2(X, \mu)$ превращается в указанное выше отождествление $(L_2(X, \mu))^*$ с $L_2(X, \mu)$.

Рассмотрим еще случай конечномерных пространств $L_p(X, \mu)$. Каждое такое пространство линейной размерности n, $1 \leqslant n < \infty$, изометрически изоморфно координатному пространству R^n с нормой $\|x\|_p = (\sum |x^i|^p)^{1/p}$, $x = (x^1, \ldots, x^n) \in R^n$. При $p = \infty$ $(\sum |x^i|^p)^{1/p}$ заменяется на $\max\limits_i |x^i|$. Так нормированные R^n будут обозначаться $\mathfrak{l}_p^{(n)}$. Сопряженное к $\mathfrak{l}_p^{(n)}$ есть $\mathfrak{l}_q^{(n)}$,

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Элемент $\varphi \in \mathfrak{l}_q^{(n)}$ действует на элемент $x \in \mathfrak{l}_p^{(n)}$ как функционал по формуле

$$\langle \varphi | x \rangle = \sum_{i=1}^{n} \varphi^i x^i.$$

При $1 < p < \infty$ пространства $L_p(X, \mu)$ рефлексивны. При $p = 1$ и $p = \infty$ это (в бесконечномерном случае) не так.

2.9. Интегрирование функций со значениями в банаховом пространстве. Пусть E, $\|\cdot\|$ — сепарабельное банахово пространство и (X, μ) — польское пространство с мерой. (Напомним, что мера на польском пространстве для нас означает полную по Лебегу вероятностную борелеву регулярную меру. Векторные функции можно интегрировать и в более общей ситуации, но нам это не потребуется.)

Обозначим через $M(X, E)$ множество классов эквивалентности (по отношению совпадения μ-почти всюду) борелевых функций на X со значениями в E. Пусть $L_p(X, \mu, E)$ есть множество всех элементов $\bar{f} \in M(X, E)$, для которых конечна величина

$$\|\bar{f}\|_p = \|\{\|f(x)\|\}\|_p, \quad f \in \bar{f}.$$

Можно доказать, что $L_p(X, \mu, E)$ — линейное (относительно естественных линейных операций) и банахово в $\|\cdot\|_p$ пространство, $L_p(X, \mu, E) \supset L_{p'}(X, \mu, E)$ при $p' \geqslant p$, причем $\|\bar{f}\|_{p'} \geqslant \|\bar{f}\|_p$ при $p' \geqslant p$, $\bar{f} \in L_{p'}$.

Функции $f \in \bar{f} \in L_p(X, \mu, E)$ можно интегрировать по мере μ следующим образом. Скажем, что f — *ступенчатая* функция, если она борелева и принимает счетное число значений $t_1, \ldots, t_r, \ldots$ Для ступенчатой функции $f \in \bar{f} \in L_1(X, \mu, E)$ определим интеграл $\int f(x) d\mu(x)$ как

$$\sum_{i=1}^{\infty} \mu\{x \mid f(x) = t_i\}\, t_i$$

(ряд абсолютно сходится, ибо $\bar{f} \in L_1$). Далее, можно доказать, что всякая функция $f \in \bar{f} \in L_1$ представима в виде суммы ряда

$\sum\limits_{i=1}^{\infty} f_i$ ступенчатых функций с конечной суммой $\sum\limits_{i=1}^{\infty} \|\bar{f}_i\|_1$. Доказывается, что ряд $\sum\limits_{i=1}^{\infty} \int f_i(x) d\mu(x)$ абсолютно сходится в E, и его сумма не зависит от выбора f_i с указанными свойствами. Она по определению принимается за интеграл $\int \bar{f}(x) d\mu(x)$. Интеграл не зависит от выбора $f \in \bar{f} \in L_1$ и

$$\left\| \int \bar{f}(x) d\mu(x) \right\| \leqslant \|\bar{f}\|_1.$$

Таким образом, отображение $\bar{f} \to \int f(x) d\mu(x)$ есть линейный непрерывный оператор I из $L_1(X, \mu, E)$ в E; при этом $\|I\bar{f}\| \leqslant \|\bar{f}\|_p$, если $\bar{f} \in L_p(X, \mu, E)$.

2.10. Наряду с обозначениями, введенными выше (они стандартны для этой книги), широко используются следующие обозначения. Пусть $G \subset E$, а $\|\cdot\|$ — норма на E. Тогда $\rho_{\|\cdot\|}(x, G)$, $\rho_{\|\cdot\|}(G)$, $d_{\|\cdot\|}(G)$, $\alpha_{\|\cdot\|}(G)$ означают соответственно расстояние от x до G, радиус G, диаметр G, асферичность G (см. п. 1.0) в метрике на E, задаваемой нормой $\|\cdot\|$. Любая точка $x \in G$, такая, что G содержится в шаре радиуса $\rho_{\|\cdot\|}(G)$ с центром в x, называется *центром* G. Можно доказать, что если E рефлексивно, а G выпукло, замкнуто, ограничено и непусто, то существует хотя бы один центр G.

ОСНОВНЫЕ ОБОЗНАЧЕНИЯ

Арабские цифры — номер параграфа и пункта, римские — номер главы, П — приложение. В списке не приведены вполне стандартные обозначения, а также обозначения, область действия которых не превышает одной главы

$\mathbb{R}$	—	вещественная ось
R^n	—	n-мерное вещественное координатное пространство
E^n	—	n-мерное евклидово вещественное пространство
inf, sup	—	точные нижняя (верхняя) грани числового множества. Для пустого множества равны $+\infty$, соответственно $-\infty$
$[t]_+$	—	функция $\max\{0, t\}$
$]t[$	—	минимальное целое k, не меньшее t
$\mathfrak{L}\{e_1, \ldots, e_k\}$	—	линейная оболочка перечисленного в скобках множества векторов
int G	П.1.0	внутренность подмножества G
∂G	П.1.0	граница подмножества G
$\overline{G}$	П.1.0	замыкание подмножества G
$E, \\|\cdot\\|$	П.2	вещественное банахово пространство с нормой
$l(\mathcal{B}, \mathfrak{A})$, $\bar{l}(\mathcal{B}, \mathfrak{A})$, $\nu(\mathcal{B}, \mathfrak{A})$, $\bar{\nu}(\mathcal{B}, \mathfrak{A})$	I.4.3	обычная (сильная) трудоемкость, соответственно погрешность метода $\mathcal{B}$ на классе $\mathfrak{A}$
$N(\nu)$, $\widetilde{N}(\nu)$	I.4.3	детерминированные (стохастическая) сложностные функции рассматриваемого класса задач
$\equiv$	—	«есть по определению»
МЦТ	II.3	метод центров тяжести
ММЦТ	II.5	модифицированный метод центров тяжести
ЗС_V, ЗС_p, $\text{ЗС}_{1,n}$	III.3	разного рода версии методов зеркального спуска решения выпуклых экстремальных задач
$\overline{\text{ЗС}}_V$, $\overline{\text{ЗС}}_p$, $\overline{\text{ЗС}}_{1,n}$	III.4	
$\widetilde{\text{ЗС}}_V$, $\widetilde{\text{ЗС}}_V(\nu)$, $\widetilde{\text{ЗС}}^{p,r}$	V.2	

$\text{ЗС}_{V,V_I}$, $\text{ЗС}_{V,V_I}(\nu)$	VI.2	версии методов зеркального спуска решения выпукло-вогнутых игр
$\widetilde{\text{ЗС}}_{V,V_I}$, $\widetilde{\text{ЗС}}_{V,V_I}(\nu)$	VI.3	
$\omega_V(t)$, $\gamma_V(s)$	III.2	характеристики гладкости регулярной функции V
$\omega_{V,r}(t)$, $\gamma_{V,r}(s)$	V.1	
$\mathcal{C}_{\text{Lip}}(G, E, \\|\cdot\\|, m)$, $\mathcal{C}(G, E, m)$	П.2	нагруженное поле липшицевых (общих) выпуклых задач с m ограничениями на множестве G в пространстве E, $\\|\cdot\\|$
E^*, $\\|\cdot\\|_*$	П.2	пространство с нормой, сопряженное к E, $\\|\cdot\\|$
$\langle\varphi \mid x\rangle$	П.2	значение линейного функционала $\varphi \in E^*$ на векторе $x \in E$ (в гильбертовом случае — скалярное произведение)
$\rho_{\\|\cdot\\|}(G)$	П.2.10	радиус множества G относительно нормы $\\|\cdot\\|$
$d_{\\|\cdot\\|}(G)$	П.2.10	диаметр множества G
$\alpha_{\\|\cdot\\|}(G)$	П.2.10	$\\|\cdot\\|$ — асферичность подмножества G
$\rho_{\\|\cdot\\|}(x, G)$	П.2.10	расстояние от x до G
$\alpha_{p,k}(\\|\cdot\\|)$, $\alpha_{p,\\|\cdot\\|,k}(G)$, $\alpha_{p,k}(G)$, $\alpha(G)$	IV.1.1	разного рода характеристики асферичности
∇f	—	градиент скалярной функции на гильбертовом пространстве
$\partial_G f(x)$	II.1	множество опорных (на G) функционалов к выпуклой функции f в точке x
$\vee_G(f)$	II.1.3	изменение функции f на множестве G
$\mathrm{L}_{\\|\cdot\\|}(f)$	II.2.2	константа Липшица f относительно $\\|\cdot\\|$
$L_p(T, \mu)$	П.2	банахово пространство интегрируемых в p-й степени функций на пространстве T с σ-аддитивной σ-конечной мерой μ, $1 \leqslant p \leqslant \infty$
$\\|\cdot\\|_p$	П.2	естественная L_p-норма.
$l_p^{(n)}$	П.2	стандартное n-мерное пространство L_p
X^i	П.1	произведение пространства X на себя i раз, $0 \leqslant i \leqslant \infty$
$x^i = (x_1, \ldots, x_i)$	—	точка пространства X^i
$f = (f_0, \ldots, f_m)$	I.2.1	задача математического программирования, по-

		рожденная вектор-функцией f (и сама эта вектор-функция)
f_*	I.2.1	оптимальное значение целевого функционала задачи f
$\mathscr{O} = ((\Omega, \mathsf{F}_\omega), I, \psi(x, f, \omega))$	I.3.1	оракул с пространством шумов $(\Omega, \mathsf{F}_\omega)$, функцией наблюдения $\psi(x, f, \omega)$ и информационным пространством I
$r(f) = (r_0(f), \ldots, r_m(f))$	I.2.3	нормирующие множители рассматриваемого класса задач
$\mathfrak{A}(\mathscr{F}, G, m, E)$	I.2.2	поле задач с m ограничениями на множестве $G \subset E$, порожденных $(m+1)$-мерными функциями семейства $\mathscr{F}$
$\vec{\varepsilon}(x, f) = (\varepsilon_0(x, f), \ldots, \varepsilon_m(x, f))$	I.2.3	вектор абсолютных погрешностей точки x в качестве приближенного решения задачи f
$\nu(x, f)$	I.2.3	относительная погрешность точки x в качестве приближенного решения задачи f
$\mathscr{B}$ $(\tilde{\mathscr{B}})$	I.3.3	детерминированный (стохастический) метод решения задач рассматриваемого класса
$\int_T \mathscr{B}^t d\mathsf{F}_t$	I.3.4	смесь детерминированных методов
$l(\tilde{\mathscr{B}}, f)$, $\bar{l}(\tilde{\mathscr{B}}, f)$	I.4.1	средняя (сильная) трудоемкости метода $\tilde{\mathscr{B}}$ на задаче f
$\nu(\tilde{\mathscr{B}}, \bar{l}), \bar{\nu}(\tilde{\mathscr{B}}, f)$	I.4.1	средняя (сильная) погрешности метода $\tilde{\mathscr{B}}$ на задаче f
$\boldsymbol{C}^{\mathscr{O}}_{\mathrm{Lip}}(G, E, \Vert\cdot\Vert, m)$, $\boldsymbol{C}^{\mathscr{O}}(G, E, m)$	II.2	классы задач, получающиеся оснащением соответствующих полей оракулом $\mathscr{O}$
$\boldsymbol{C} \in \boldsymbol{C}^{V_0}_{\mathrm{Lip}}(G, E, \Vert\cdot\Vert, m)$, $\boldsymbol{C} \in \boldsymbol{C}^{V_0}(G, E, m)$	II.2	сокращенная запись высказывания «$\boldsymbol{C}$ есть класс задач типа $\boldsymbol{C}^{V_0}_{\mathrm{Lip}}(G, E, \Vert\cdot\Vert, m)$» (соответственно $\boldsymbol{C}^{V_0}(G, E, m)$)
$\boldsymbol{D}(G \times G_I, E, E_I, \Vert\cdot\Vert, \Vert\cdot\Vert_I)$	VI.1.2	поле липшицевых выпукло-вогнутых игр
$\tilde{\boldsymbol{C}}^{\mathscr{O}, V_0}_{r,L}(G, E, \Vert\cdot\Vert, m)$	V.1.3	класс стохастических выпуклых задач
$\boldsymbol{D}^{V_0}(G \times G_I, E, \Vert\cdot\Vert, E_I, \Vert\cdot\Vert_I)$, $\tilde{\boldsymbol{D}}^{\mathscr{O}, V_0}_{r,L}(G \times G_I, E, \Vert\cdot\Vert, E_I, \Vert\cdot\Vert_I)$	VI.1.2	классы задач решения игр

$\boldsymbol{H}_{x_1}(G, m, l_0, \ldots, l_m, Q_0, \ldots, Q_m)$, $\boldsymbol{H}^V_{x_1}(G, m, l_0, \ldots, l_m, Q_0, \ldots, Q_m)$, $\boldsymbol{CS}^V_{x_1}(G, m)$	VII.1	классы гладких (сильно) выпуклых экстремальных задач
$\underset{x}{\sim}, \underset{x}{\lesssim}, \sim, \lesssim$	IV.1.4	—
$g^f_{j,x}(y)$	II.2.5	—
G_*	I.2.3	—
$\pi_G(x)$	III.3.1	—
$\mu_G(x)$	III.3.1	—

ЛИТЕРАТУРА

1. Ахиезер Н. И. Лекции по теории аппроксимации.—М.: Наука, 1965.
2. Бахвалов Н. С. Об оптимальных оценках скорости сходимости квадратурных процессов и методов интегрирования типа Монте-Карло на классах функций.— В сб.: Численные методы решения дифференциальных и интегральных уравнений и квадратурные формулы.— М.: Наука, 1964, с. 5—63.
3. Брук (Bruck R). On Weak Convergence of an Ergodic Iteration for the Solution of Variational Inequalities for Monotone Operators in Hilbert Space.— J. Math. Anal. and Appl., 1977, v. 61, No. 1.
4. Данилин Ю. М. Методы сопряженных направлений для решения задач минимизации.— Кибернетика, 1971, № 5, с. 122—136.
5. Данфорд Н., Шварц Д. Линейные операторы.— М.: ИЛ, 1962.
6. Дворецкий (Dvoretzky A.) Some Results on Convex Bodies and Banach Spaces.— Proc. Internat. Symposium on Linear Spaces, Jerusalem, 1960, 123—160.
7. Джон (John F.) Extremum Problems with Inequalities as subsidiary Conditions.— Studies and essays, presented to R. Courant on his 60-th birthday, N. Y., 1948, 187—204.
8. Дэй М. Линейные нормированные пространства.— М.: ИЛ, 1961.
9. Ермольев Ю. М. Методы стохастического программирования.— М.: Наука, 1976.
10. Звонкин А. К., Левин Л. А. Сложность конечных объектов и обоснование понятий информации и случайности с помощью теории алгоритмов.— УМН, 1970, т. XXV, вып. 6, 85—127.
11. Иванов В. В. Об оптимальных алгоритмах минимизации функций некоторых классов.— Кибернетика, 1972, № 4, 81—94.
12. Карп Р. Сводимость комбинаторных задач.— Кибернетический сборник, новая серия, № 12.— М.: Мир, 1975.
13. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа.— М.: Наука, 1972.
14. Кузовкин А. И., Тихомиров В. М. О количестве вычислений для нахождения минимума выпуклой функции.— Экономика и матем. методы, 1967, т. III, вып. I, с. 95—103.
15. Левин А. Ю. Об одном алгоритме минимизации выпуклых функций.— Доклады АН СССР, 1965, т. 160, вып. 6.
16. Левин Л. А. Универсальные задачи перебора.— Проблемы передачи информации, 1973, т. IX, вып. 3, с. 115—116.
17. Мартин, Ти (Martin D. W., Tee G. I.) Iterative Methods for Linear Equations with Positive Definite Symmetric Matrix.— Computer J., 1961, v. 4, № 3.
18. Митягин Б. С. Два неравенства для объемов выпуклых тел.— Матем. заметки, 1969, т. V, вып. 1.
19. Фон Нейман Дж. К теории стратегических игр.— В сб.: Матричные игры.— М.: Физматгиз, 1961.

20. Немировский А. С., Юдин Д. Б. Методы оптимизации, адаптивные к «существенной» размерности задачи.— Автоматика и телемеханика, 1977, № 4, с. 75—87.
21. Немировский А. С., Юдин Д. Б. Эффективные методы решения задач выпуклого программирования большой размерности.— Экономика и матем. методы, 1979, т. XV, вып. 1.
22. Немировский А. С., Юдин Д. Б. Чезаровская сходимость градиентного метода аппроксимации седловых точек выпукло-вогнутых функций.— Доклады АН СССР, 1978, т. 239, вып. 5, с. 1056—1059.
23. Облонская Л. Я. Сравнение быстроты сходимости методов сопряженных градиентов и градиентного для квадратичных функционалов.— В сб.: Вопросы точности и эффективности вычислительных алгоритмов.— Труды симпозиума, т. 4. — Киев: 1969, с. 94—103.
24. Облонская Л. Я. О скорости сходимости метода сопряженных градиентов для квадратичных функционалов.— Труды 2-й зимней школы по математическому программированию и смежным вопросам (1969), вып. 3.— М.: 1969, с. 550—568.
25. Полак Э. Численные методы оптимизации. Единый подход.— М.: Мир, 1974.
26. Халмош П. Теория меры.— М.: ИЛ, 1963.
27. Химмельблау Д. Прикладное нелинейное программирование.— М.: Мир, 1975.
28. Шор Н. З. О скорости сходимости метода обобщенного градиентного спуска с растяжением пространства.— Кибернетика, 1970, № 2.
29. Эрроу К. Дж. и др. Исследования по линейному и нелинейному программированию.— М.: ИЛ, 1962.
30. Юдин Д. Б., Немировский А. С. Оценка информационной сложности задач математического программирования.— Экономика и матем. методы, 1976, т. XII, вып. I, с. 128—142.
31. Юдин Д. Б., Немировский А. С. Информационная сложность и эффективные методы решения выпуклых экстремальных задач.— Экономика и матем. методы, 1976, т. XII, вып. 2, с. 357—369.
32. Юдин Д. Б., Немировский А. С., Информационная сложность строго выпуклого программирования.— Экономика и матем. методы, 1977, т. XIII, вып. 3, с. 550—559.
33. Юдин Д. Б., Немировский А. С. Эффективность случайного поиска в задачах управления.— Известия АН СССР, Техническая кибернетика, 1977, № 3, с. 3—17.
34. Юдин Д. Б., Немировский А. С. Эффективность рандомизации управления.— В сб.: Проблемы случайного поиска.— Рига: Зинатне, 1978, № 7.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ