А.Ф. ИЗМАИЛОВ, М.В. СОЛОДОВ

ЧИСЛЕННЫЕ МЕТОДЫ ОПТИМИЗАЦИИ

Издание второе, переработанное
и дополненное

*Рекомендовано учебно-методическим советом
по прикладной математике и информатике для студентов
высших учебных заведений, обучающихся по специальности
010200 "Прикладная математика и информатика"
и по направлению 510200 "Прикладная математика и информатика"*

МОСКВА
ФИЗМАТЛИТ®
2008

УДК 519.85
ББК 22.18
И 37

Измаилов А. Ф., Солодов М. В. **Численные методы оптимизации.** — 2-е изд., перераб. и доп. — М.: ФИЗМАТЛИТ, 2008. — 320 с. — ISBN 978-5-9221-0975-8.

Современный курс численных методов оптимизации. Основное внимание уделено методам общего назначения, ориентированным на решение гладких задач математического программирования без какой-либо специальной структуры. Излагаются как «классические» методы, важные в идейном отношении, так и более изощренные «новые» алгоритмы, привлекающие в настоящее время наибольшее внимание специалистов и пользователей. Некоторые результаты в монографической литературе публикуются впервые.

Для студентов, аспирантов и научных работников, интересующихся численными методами оптимизации.

Учебное издание

ИЗМАИЛОВ Алексей Феридович
СОЛОДОВ Михаил Владимирович
ЧИСЛЕННЫЕ МЕТОДЫ ОПТИМИЗАЦИИ

Редактор *И.Л. Легостаева*
Оригинал-макет: *И.В. Шутов*
Оформление переплета: *Н.В. Гришина*

Подписано в печать 22.05.08. Формат 60×90/16. Бумага офсетная. Печать офсетная. Усл. печ. л. 20. Уч.-изд. л. 23,5. Тираж 1000 экз. Заказ №

Издательская фирма «Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117997, Москва, ул. Профсоюзная, 90
E-mail: fizmat@maik.ru, fmlsale@maik.ru;
http://www.fml.ru

Отпечатано с готовых диапозитивов
в ОАО «Ивановская областная типография»
153008, г. Иваново, ул. Типографская, 6
E-mail: 091-018@adminet.ivanovo.ru

ISBN 978-5-9221-0975-8

ISBN 978-5-9221-0975-8

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ КО ВТОРОМУ ИЗДАНИЮ

Первое издание этой книги увидело свет в 2003 г. и имело определенный успех у читателей, что в значительной степени и подвигло авторов на подготовку переработанного второго издания. За это время авторы выпустили два учебника по различным аспектам *теории* оптимизации — на португальском языке [1)] и на русском [20]. Зачем вообще нужна теория оптимизации? С одной стороны, это чрезвычайно достойная и богатая на открытия область математической деятельности, находящаяся в постоянном развитии и весьма привлекательная для каждого, кто способен оценить красоту и нетривиальность теоретических построений. С другой стороны, опосредовано, через *численные методы* оптимизации, теория оптимизации находит применения для практического (читай: численного) решения реальных задач, возникающих в приложениях, и, по мнению авторов данной книги, именно второе предназначение теории оптимизации является главным.

Например, те или иные необходимые условия оптимальности принято сравнивать между собой по их тонкости (близости к достаточным), а также по тому, насколько обременительны предположения (условия гладкости, условия регулярности ограничений), гарантирующие справедливость данных необходимых условий. Вместе с тем, следует оценивать необходимые условия оптимальности еще и с точки зрения их потенциальной практической полезности, т. е. возможности построения на их основе новых (либо обоснования и анализа свойств известных) численных методов. Самые тонкие и, казалось бы, гармоничные необходимые условия оптимальности иногда остаются совершенно невостребованными в численной оптимизации, в то время как более грубые условия находят многочисленные вычислительные приложения.

Упомянутая выше книга на португальском языке является первым томом, за которым последовал второй [2)], посвященный численной оптимизации. В книге [20] большое внимание уделено приложениям

[1)] *Izmailov A.F., Solodov M.V.* Otimização. V. 1. Condições de otimalidade, elementos de análise convexa e de dualidade.—Rio de Janeiro: IMPA, 2005.

[2)] *Izmailov A.F., Solodov M.V.* Otimização. V. 2. Métodos computacionais.—Rio de Janeiro: IMPA, 2007.

теории чувствительности именно в численной оптимизации. Что же касается настоящей книги, то одной из ее отличительных особенностей является то, что элементы теории оптимизации излагаются в ней действительно в минимально необходимом объеме (только то, что требуется для изложения и обоснования представленных здесь численных методов).

Во втором издании полностью переписаны и значительно расширены п. 3.2.3 и § 6.3, ввиду чрезвычайной практической важности рассматриваемых в них методов. На сегодняшний день квазиньютоновские методы следует считать наиболее эффективным подходом к решению гладких задач безусловной оптимизации, а методы семейства bundle — наиболее эффективным подходом к решению негладких выпуклых задач. На самом деле, если иметь ввиду лишь вычислительную практику, а не теорию, то другие методы для соответствующих классов задач можно было бы и не рассматривать, по крайней мере в деталях. В § 5.4 добавлен материал о преодолении эффекта Маратоса для методов последовательного квадратичного программирования (SQP) за счет использования так называемых поправок второго порядка, что на сегодняшний день является наиболее распространенным практическим подходом к данной проблеме.

Кроме того, имеется множество других исправлений и дополнений, большинство из которых носит характер комментариев, призванных, в частности, усилить акценты и еще яснее обозначить, какие методы действительно составляют state-of-the-art численной оптимизации, а какие представляют в основном исторический интерес, либо продолжают использоваться лишь благодаря простоте идеи и/или реализации, или в результате того, что пользователи просто не знают о наличии лучших альтернатив.

А.Ф. Измаилов, *Москва*
М.В. Солодов, *Рио-де-Жанейро*
Апрель 2008 г.

ВВЕДЕНИЕ

Наука о численных методах оптимизации весьма молода; она сформировалась во второй половине прошлого века и продолжает активно развиваться. Изначально предназначенная в основном для обслуживания исследования операций (или, более общим образом, науки о принятии решений), сейчас она представляет собой самостоятельную область деятельности на стыке теории и приложений, со своими традициями, языком и широким международным сообществом специалистов. В мире издается множество журналов по оптимизации (исключение составляет Россия, в которой в силу определенных причин никогда не существовало и не существует по сей день специализированного научного журнала по этой тематике). Все возрастающая востребованность результатов в данной области объясняется как наличием множественных естественных оптимизационных процессов в природе, так и естественным стремлением человека к оптимальной организации своей деятельности. С другой стороны, стремительное развитие численной оптимизации именно в последние десятилетия привело к появлению алгоритмов и программного обеспечения, пригодных для решения реальных прикладных задач (в том числе задач большой размерности), тем самым по-настоящему сблизив эту науку с областью ее возможных приложений.

К настоящему моменту различными авторами опубликовано множество (в том числе прекрасных) руководств и учебников по численным методам оптимизации (см., например, [1, 4, 6, 7, 9, 10, 13, 16, 17, 19, 23, 26, 28, 30–33, 36, 38, 40, 41, 44–46, 48]; здесь намеренно не упоминаются более ранние и явно устаревшие издания). Однако, как видно из приведенного списка, несмотря на чрезвычайно активное развитие этого предмета, с начала 90-х годов издавались лишь единичные курсы такого рода. В современной западной литературе по оптимизации часто цитируются книги [41, 50], но широкому кругу российских читателей они недоступны. Давно ожидаемым событием стал выход фундаментального труда [8], однако авторы надеются, что и настоящая книга займет достойное место в отечественной литературе по методам оптимизации.

В любом случае появление нового курса по методам оптимизации должно сопровождаться разъяснениями, чем данный курс отличается

от существующих и какие цели преследовались при его подготовке. Предмет этот чрезвычайно обширен, поэтому важно обозначить основные принципы, которыми руководствовались авторы при отборе материала, и, прежде всего, какие вопросы и почему не нашли отражения в курсе.

Главный объект изучения в данном курсе — гладкая задача конечномерного математического программирования без каких-либо предположений о выпуклости. Тем самым, во-первых, сознательно сводится к минимуму рассмотрение чрезвычайно важного класса негладких задач и задач с недостаточной гладкостью. Это объясняется лишь неизбежной ограниченностью объема курса. Элементы негладкого анализа и соответствующие методы обсуждаются ниже главным образом как средства решения изначально гладких задач. Такое естественное возникновение негладкости при работе с гладкими задачами наблюдается нередко. Например, множество решений системы гладких неравенств обычно негладко (в некотором интуитивном смысле).

Во-вторых, важной отличительной чертой основной части данного курса является сознательный отказ от использования аппарата выпуклого анализа. При обосновании и сравнительном анализе численных методов оптимизации в большинстве известных руководств этот аппарат используется систематически. Это удобно, поскольку наличие выпуклой структуры существенно упрощает анализ и дает возможность довести его до конца. Тем самым, получается единая картина методов оптимизации, но лишь для весьма специального класса задач с очень существенными структурными ограничениями. Можно, видимо, утверждать, что подавляющее большинство задач оптимизации, возникающих в приложениях, не являются выпуклыми (а если и являются выпуклыми, то не являются гладкими). Кроме того, представляется весьма интересным выяснить, насколько далеко можно продвинуться в обосновании методов оптимизации, не привлекая по существу предположений о выпуклости. Одной из важнейших задач авторов при подготовке этого курса была демонстрация того, что продвинуться можно достаточно далеко.

Тем не менее, негладкие выпуклые задачи не могли быть полностью исключены из рассмотрения ввиду несомненной важности и актуальности этой проблематики. Обзор некоторых современных методов решения таких задач вынесен в отдельную главу.

Выбор основного класса задач, рассматриваемого в курсе, во многом определяет круг обсуждаемых (и соответственно не обсуждаемых) численных методов: главным образом речь будет идти о методах общего назначения, ориентированных на решение гладких задач математического программирования без какой-либо специальной структуры. При этом, с одной стороны, обсуждаются методы, важные в идейном отношении, анализируются их сравнительные достоинства

и недостатки, что позволяет сделать изложение связным, показав взаимосвязь различных методов. С другой стороны, авторы надеются, что курс позволяет составить представление о state-of-the-art численной оптимизации, о тех методах, которые наиболее активно обсуждаются в современной литературе. Более того, некоторые излагаемые ниже результаты (см., например, § 3.1, 4.5, 5.4) в монографической литературе публикуются, по-видимому, впервые.

Подчеркнем, что этот курс освещает методы оптимизации с точки зрения математика, а не практика: многочисленные тонкости реализации и практического использования излагаемых методов (например, масштабирование, применимость методов к задачам большой размерности, автоматическое дифференцирование) здесь почти не обсуждаются. Нужно, однако, понимать, что для практика такие тонкости и эвристические приемы, основанные на большом опыте применения методов, зачастую значительно важнее абстрактных теорем о сходимости. В этом смысле прекрасными «руководствами пользователя» являются, например, книги [13, 42, 46, 47, 49, 50]. Много внимания вопросам реализации методов уделено также в более ранних изданиях [32, 39], однако содержащаяся в них информация по методам условной оптимизации с современной точки зрения явно недостаточна.

Совершенно не затронут в курсе еще один круг вопросов — устойчивость методов к разного рода возмущениям и, в частности, влияние неточного решения вспомогательных задач на конечный результат работы метода. Важность этих вопросов несомненна, но за ответами на них читателю следует обратиться к другим изданиям.

В книге отсутствует раздел, содержащий предварительные и вспомогательные сведения, необходимые для ее чтения. Дело в том, что такие сведения минимальны: требуется лишь знакомство читателя с линейной алгеброй и математическим анализом в объеме начальных курсов, включая дифференциальное исчисление функций многих переменных. Весьма желательно также знакомство с основами выпуклого анализа. Впрочем, в соответствии со сказанным выше, в большей части курса (по сути дела, за исключением главы 6) выпуклый анализ если и привлекается, то лишь на уровне базовых понятий и простейших результатов этой науки.

Избранные факты из анализа, которые играют в курсе особенно важную роль, приводятся по ходу изложения, иногда без доказательств (к ним относятся, например, теорема Вейерштрасса, теорема о неявной функции, теорема Радемахера). Кроме того, многие используемые понятия и результаты комментируются в сносках.

Отсутствие в книге иллюстраций объясняется тем, что, по мнению авторов, будет полезнее, если читатель сам по возможности проиллюстрирует приводимые сведения рисунками (обычно двумерными).

Подчеркнем, что авторство излагаемых ниже результатов указывается лишь в некоторых случаях. Дело в том, что вопрос об авторстве

часто бывает весьма непростым и спорным и вряд ли является важным для большинства читателей. Соответственно в списке литературы нет оригинальных журнальных статей, а приведены только монографии. Историческим вопросам много внимания уделено, например, в [6, 41].

Вообще, целью авторов было написание максимально компактного (без лишней информации) и, вместе с тем, достаточно полного введения в современную численную оптимизацию. За более детальной информацией, касающейся отдельных классов методов, читателю следует обратиться к соответствующей специальной литературе, ссылки на которую имеются в тексте.

В тексте содержится большое количество задач. Некоторые просты и даже рутинны (хотя авторы старались избегать тривиальных задач), а в некоторых приводятся вполне содержательные утверждения, являющиеся неотъемлемой частью излагаемого материала, или постановки проблем, которые могут послужить основой даже для курсовых и дипломных работ. Ряд задач и примеров заимствованы из [3, 10, 37, 41] и других источников.

Общепринятые обозначения специально не оговариваются, их пояснение вынесено в список обозначений. Для удобства ссылок в книге применяется следующая система нумерации ее разделов. Номер параграфа состоит из двух цифр, первая из которых обозначает номер главы, в которой находится этот параграф. Аналогично, номер пункта состоит из трех цифр, первые две из которых составляют номер параграфа, в котором находится этот пункт. Нумерация объектов (формул, определений, теорем и т. п.) в каждом параграфе независимая. При ссылке на объект извне параграфа используется номер, состоящий из трех цифр, первые две из которых составляют номер параграфа, а последняя — номер объекта в параграфе. Под «условиями» того или иного утверждения (теоремы, предложения, леммы) всегда понимается все то, что сказано в этом утверждении до слова «Тогда».

Эта книга написана на основе курсов лекций, которые авторы на протяжении ряда лет читали на факультете ВМиК МГУ им. М.В.Ломоносова, на факультете физико-математических и естественных наук РУДН, а также в Instituto de Matemática Pura e Aplicada (Рио-де-Жанейро, Бразилия). Авторы чрезвычайно признательны IMPA, ставшему «базой» для их сотрудничества в последние годы. Особую благодарность авторы выражают Claudia Sagastizábal за ее постоянную профессиональную и личную помощь и поддержку. Мы признательны О.А. Брежневой, а также нашим студентам и аспирантам, которые способствовали устранению неточностей и опечаток, присутствовавших в первоначальном варианте рукописи.

* * *

Авторы посвящают эту книгу памяти Владимира Георгиевича Карманова. Его замечательный учебник «Математическое программирование» сыграл выдающуюся роль в становлении и развитии отечественной численной оптимизации. Своеобразие стиля и подхода к излагаемому материалу обеспечивают этому учебнику популярность у студентов и специалистов и по сей день (первое издание увидело свет еще в 1975 г., а последнее — в 2008 г. [24]). Долгие годы Владимир Георгиевич принимал активное участие в деятельности российского оптимизационного сообщества и в подготовке кадров для него, в том числе на кафедре исследования операций факультета ВМиК МГУ, выпускниками которой являются оба автора настоящей книги. Личное удовольствие и профессиональная польза от общения с Владимиром Георгиевичем, а также его многолетняя прямая и косвенная поддержка — неоценимый вклад в эту книгу.

СПИСОК ОБОЗНАЧЕНИЙ

$\mathbf{R}$ — множество вещественных чисел.

$\mathbf{R}_+$ — множество неотрицательных вещественных чисел.

$\mathbf{R}^n$ — n-мерное арифметическое пространство, снабженное евклидовым скалярным произведением и соответствующей нормой.

$\mathbf{R}(m, n)$ — пространство вещественных $m \times n$-матриц, снабженное нормой, подчиненной нормам в $\mathbf{R}^n$ и $\mathbf{R}^m$.

$x_1, \ldots, x_n$ — компоненты вектора $x \in \mathbf{R}^n$ в стандартном базисе пространства $\mathbf{R}^n$ (если не оговорено иное).

$\langle \cdot, \cdot \rangle$ — евклидово скалярное произведение.

$|\cdot|$ — евклидова норма вектора (аналогичное обозначение используется для количества элементов конечного множества).

$|\cdot|_p$ — норма вектора, определяемая как корень степени p из суммы возведенных в степень p модулей его компонент (в частности, $|\cdot| = |\cdot|_2$).

$|\cdot|_\infty$ — норма вектора, определяемая как максимум из модулей его компонент.

$B(\overline{x}, \delta) = \{x \in \mathbf{R}^n \mid |x - \overline{x}| < \delta\}$ — открытый шар радиуса δ с центром в точке $\overline{x} \in \mathbf{R}^n$.

$\overline{B}(\overline{x}, \delta) = \{x \in \mathbf{R}^n \mid |x - \overline{x}| \leqslant \delta\}$ — замкнутый шар радиуса δ с центром в точке $\overline{x} \in \mathbf{R}^n$.

$\operatorname{dist}(x, X) = \inf\limits_{\xi \in X} |x - \xi|$ — расстояние от точки x до множества X.

$\{x^k\} = \{x^0, x^1, \ldots, x^k, \ldots\}$ — последовательность.

$\{x^k\} \to x \quad (k \to \infty)$ — последовательность $\{x^k\}$ сходится к элементу x (для числовых последовательностей $\{a_k\}$ используются также обозначения $a_k \to a \quad (k \to \infty)$ или $\lim\limits_{k \to \infty} a_k = a$).

$\liminf\limits_{k \to \infty} a_k \quad (\limsup\limits_{k \to \infty} a_k)$ — нижний (верхний) предел числовой последовательности $\{a_k\}$.

$\operatorname{int} X$ — внутренность множества X.

$\operatorname{cl} X$ — замыкание множества X.

$\operatorname{span} X$ — линейная оболочка множества X (минимальное линейное подпространство, содержащее X).

$\operatorname{conv} X$ — выпуклая оболочка множества X (минимальное выпуклое множество, содержащее X).

$\operatorname{cone} X$ — коническая оболочка множества X (минимальный выпуклый конус, содержащий X).

$\dim X$ — размерность линейного пространства X.

E^n — единичная $n \times n$-матрица.

$X^\perp = \{x \in \mathbf{R}^n \mid \langle x, \xi \rangle = 0 \quad \forall\, \xi \in X\}$ — ортогональное дополнение множества $X \subset \mathbf{R}^n$.

$K^* = \{x \in \mathbf{R}^n \mid \langle x, \xi \rangle \geqslant 0 \quad \forall\, \xi \in K\}$ — конус, (положительно) сопряженный к конусу $K \subset \mathbf{R}^n$.

Ax — умножение матрицы A на вектор x (или действие линейного оператора A на элемент x).

$\operatorname{im} A$ — образ (множество значений) матрицы (линейного оператора) A.

$\ker A$ — ядро (множество нулей) матрицы (линейного оператора) A.

A^{T} — матрица, транспонированная к матрице A.

$\det A$ — определитель матрицы A.

$\operatorname{rank} A = \dim \operatorname{im} A$ — ранг матрицы (линейного оператора) A.

$L_{f,X}(c) = \{x \in X \mid f(x) \leqslant c\}$ — множество Лебега функции f на множестве X.

$\mathcal{D}_f(x)$ — множество направлений убывания функции f в точке x.

$\partial f(x)$ — субдифференциал выпуклой (либо супердифференциал вогнутой) функции f в точке x.

$\partial_\varepsilon f(x)$ — ε-субдифференциал выпуклой функции f в точке x.

$\mathcal{F}_X(x)$ — множество возможных относительно множества X в точке x направлений.

$\pi_X(x)$ — проекция точки x на замкнутое выпуклое множество X.

$\operatorname{graph} F = \{(x, y) \in X \times Y \mid F(x) = y\}$ — график отображения F: $X \to Y$.

$\square$ — знак окончания доказательства.

Посвящается памяти
В.Г. Карманова

Глава 1
ЭЛЕМЕНТЫ ТЕОРИИ ОПТИМИЗАЦИИ

При обосновании и сравнительном анализе численных методов оптимизации обычно существенно используются теоретические результаты о задачах оптимизации: условия существования решений, необходимые и достаточные условия оптимальности, оценки расстояния до множества решений или допустимого множества задачи. Более того, нередко сама идея построения метода имеет в своей основе те или иные условия оптимальности. В настоящей главе излагаются минимально необходимые для дальнейшего теоретические сведения о задачах оптимизации. Подчеркнем, что многие важные с теоретической точки зрения результаты, не находящие пока серьезных приложений в области численных методов, намеренно опущены.

§ 1.1. Начальные сведения о задачах оптимизации

1.1.1. Постановка и классификация задач оптимизации. Основной объект изучения в данном курсе — задача об отыскании минимума функции $f\colon D \to \mathbf{R}$, определенной на заданном множестве $D \subset \mathbf{R}^n$. Эту задачу будем записывать в виде

$$f(x) \to \min, \quad x \in D, \tag{1}$$

и будем называть *задачей оптимизации* (или *задачей математического программирования*). При этом точки множества D называют *допустимыми точками*, само D — *допустимым множеством*, а функцию f — *целевой функцией* задачи (1).

Определение 1. Точка $\overline{x} \in D$ называется:
глобальным решением задачи (1), если

$$f(\overline{x}) \leqslant f(x) \quad \forall\, x \in D; \tag{2}$$

локальным решением задачи (1), если найдется окрестность U точки $\overline{x}$ такая, что

$$f(\overline{x}) \leqslant f(x) \quad \forall\, x \in D \cap U. \tag{3}$$

Разумеется, всякое глобальное решение является и локальным, но, вообще говоря, не наоборот. Если неравенство в (2) или (3) при

$x \neq \overline{x}$ выполняется строгим образом, то говорят о *строгом глобальном* или соответственно *строгом локальном решении*. Заметим, что строгое локальное решение не обязательно является изолированным (от других локальных решений).

Определение 2. *Значением* задачи (1) называется величина

$$\overline{v} = \inf_{x \in D} f(x).$$

Далее, всякую задачу максимизации

$$f(x) \to \max, \quad x \in D, \tag{4}$$

всегда можно заменить эквивалентной задачей минимизации

$$-f(x) \to \min, \quad x \in D.$$

Локальные и глобальные решения этих задач совпадают, а значения отличаются только знаками. Таким образом, с теоретической точки зрения безразлично, какой класс задач (максимизации или минимизации) рассматривать: если изучен один из этих классов, то можно считать изученным и другой. В этом курсе будут рассматриваться главным образом задачи минимизации.

Решения задач (1) и (4) в совокупности называют *экстремальными точками* или *экстремумами* функции f на множестве D, а сами задачи (1) и (4) часто называют *экстремальными задачами*.

Чаще всего допустимое множество задачи (1) задается в следующем виде:

$$D = \{x \in P \mid F(x) = 0,\ G(x) \leqslant 0\}, \tag{5}$$

где $P \subset \mathbf{R}^n$ — заданное множество, $F\colon P \to \mathbf{R}^l$ и $G\colon P \to \mathbf{R}^m$ — заданные отображения. При этом удобно предполагать, что функция f также определена на P. Условие принадлежности допустимой точки множеству P называется *прямым ограничением*, а ограничения-равенства и ограничения-неравенства в определении D — *функциональными ограничениями*.

Если $D = \mathbf{R}^n$ (соответственно $D \neq \mathbf{R}^n$), то задачу (1) называют *задачей безусловной* (соответственно *условной*) *оптимизации*. Задача безусловной оптимизации может рассматриваться как частный случай задачи (1), (5) (при $P = \mathbf{R}^n$ и отсутствии функциональных ограничений, что можно формально записать как $l = m = 0$). Теории таких задач посвящен § 1.2.

При $P = \mathbf{R}^n$ и $m = 0$ задача (1), (5) становится задачей с чистыми ограничениями-равенствами. Такие задачи рассматриваются в начальном курсе математического анализа (см., например, [23]). В основе их теории лежит классический принцип Лагранжа; этим вопросам посвящен § 1.3. Наконец, теория задач оптимизации со смешанными ограничениями (равенствами и неравенствами) излагается в § 1.4.

Напомним, что множество называется *полиэдром*, если оно может быть представлено как множество решений конечной системы линейных неравенств. Отображение называется *аффинным*, если оно может быть представлено как сумма линейного отображения и постоянного вектора. Функция называется *квадратичной*, если она может быть представлена как сумма квадратичной формы и аффинной функции.

Более специальные классы задач оптимизации связаны, например, со случаем, когда множество P является полиэдром, а отображения F и G аффинны (в этом случае само множество D является полиэдром). Если при этом f — линейная функция, то (1), (5) называют *задачей линейного программирования*, а если f — квадратичная функция, то *задачей квадратичного программирования*. Соответствующие разделы теории оптимизации называются *линейным программированием* и *квадратичным программированием*; им посвящена гл. 7.

Разумеется, приведенная классификация задач оптимизации не претендует на полноту. Например, здесь не рассматривается важный в идейном плане класс задач выпуклого программирования (по причинам, указанным во введении).

Обращаем внимание на следующее важное обстоятельство. Деление ограничений в (5) на функциональные и прямые условно. Обычно выбор того, какие ограничения отнести к прямым, а какие к функциональным, находится во власти математика, исследующего задачу. Всегда можно считать, что есть лишь прямые ограничения. Наоборот, прямые ограничения всегда можно записать как функциональные (например, с помощью индикаторной функции множества P). В зависимости от целей исследования в конкретном случае удобна та или иная форма представления ограничений. Однако, как правило, полезно возможно более детальное описание функциональных ограничений. Обычно в качестве P берется множество «простой» структуры, во всяком случае выпуклое [1], например шар или параллелепипед [2] в $\mathbf{R}^n$. Особенно часто используются множества $P = \mathbf{R}^n_+$ или просто $P = \mathbf{R}^n$ (случай, когда прямых ограничений нет).

Постановки задач, в которых нетривиальным образом присутствуют как прямые, так и функциональные ограничения, рассматриваются ниже лишь в связи с некоторыми специальными вопросами: обыч-

[1] Множество $X \subset \mathbf{R}^n$ называется *выпуклым*, если вместе с любыми двумя своими точками оно содержит соединяющий их отрезок: $tx^1 + (1-t)x^2 \in X \quad \forall\, x^1, x^2 \in X, \ \forall\, t \in [0,1]$.

[2] *Параллелепипедом* в $\mathbf{R}^n$ называется прямое произведение n числовых отрезков, т. е. множество вида $\{x \in \mathbf{R}^n \mid a_i \leqslant x_i \leqslant b_i \ \forall\, i = 1, \ldots, n\}$, где a_i и b_i — заданные числа, $a_i \leqslant b_i$, $i = 1, \ldots, n$. Иногда удобно допускать, что какие-то из чисел a_i могут быть равны $-\infty$, а b_i равны $+\infty$.

но будет предполагаться, что все ограничения отнесены либо к прямым, либо к функциональным. Такой подход представляется разумным с точки зрения подавляющего большинства излагаемых ниже численных методов.

Задача 1. Используя геометрические построения, высказать гипотезу о глобальном решении задачи

$$-x_1 + x_2 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^2 \mid x_1^2 + x_2^2 \leqslant 1,\ x_1 \geqslant x_2^2,\ x_1 + x_2 \geqslant 0\}.$$

Задача 2. То же задание для задачи

$$\max\{|x_1 - 2|,\ |x_2|\} \to \min,$$

$$x \in D = \{x \in \mathbf{R}^2 \mid 2|x_1| - |x_2| \leqslant 2\}.$$

Задача 3. То же задание для задачи

$$(x_1 - 1)^2 + x_2^2 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^2 \mid x_1^2 + x_2^2 = 2\sigma x_1 x_2\},$$

где $\sigma \in \mathbf{R}$ — параметр.

Задача 4. Используя геометрические построения, высказать гипотезу о том, при каких значениях параметра $\sigma \in \mathbf{R}$ задача

$$x_1 + \sigma x_2 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^2 \mid x_1^3 + x_2^3 = 3x_1 x_2\},$$

имеет глобальное решение, а при каких — лишь локальное.

1.1.2. Существование глобального решения. Если $\overline{v} = -\infty$, то задача (1), конечно же, не имеет глобального решения (это случай, когда целевая функция не ограничена снизу на допустимом множестве). Более того, значение задачи может быть конечно, но не достигаться ни в одной допустимой точке, и при этом глобального решения тоже нет. Большое значение имеют достаточные признаки, позволяющие судить о существовании у задачи глобального решения априори, без фактического отыскания такого решения. Простейшим результатом такого рода (что нисколько не снижает его важности) является следующая классическая теорема Вейерштрасса, доказываемая в любом курсе математического анализа (например, в [23]).

Теорема 1. *Пусть $D \subset \mathbf{R}^n$ — непустое компактное множество, $f\colon D \to \mathbf{R}$ — непрерывная на D функция.*

Тогда задачи (1) *и* (4) *имеют глобальные решения.*

Условие непрерывности функции f в этой теореме можно несколько ослабить, заменив его условием полунепрерывности снизу [1)] в случае задачи на минимум и сверху в случае задачи на максимум. В остальном же условия теоремы Вейерштрасса вполне сбалансированы: ослабить в ней условие компактности множества D (например, отказавшись от ограниченности этого множества, что бывает важно в приложениях) можно лишь за счет усиления требований на функцию f, причем получаемые на этом пути результаты обычно сами являются следствиями теоремы 1. Приведем одно такое очевидное следствие.

Следствие 1. *Пусть* $D \subset \mathbf{R}^n$ — *произвольное множество,* $f\colon D \to \mathbf{R}$ — *непрерывная на* D *функция, причем существует число* c, *для которого множество Лебега* $L_{f,D}(c) = \{x \in D \mid f(x) \leqslant c\}$ *функции* f *непусто и компактно.*

Тогда задача (1) *имеет глобальное решение.*

Пример 1. Покажем, что задача (1) с допустимым множеством

$$D = \{x \in \mathbf{R}^2 \mid x_1 + x_2 > 0\}$$

и целевой функцией

$$f\colon D \to \mathbf{R}, \quad f(x) = x_1^2 + x_2 + \frac{1}{x_1 + x_2},$$

имеет глобальное решение.

Заметим, что множество D не ограничено и не замкнуто, поэтому теорема Вейерштрасса неприменима. Тем не менее зафиксируем произвольное c, для которого $L(c) = L_{f,D}(c) \neq \varnothing$ (например, можно взять $c = f(x)$ для произвольного фиксированного $x \in D$), и докажем, что $L(c)$ компактно.

Предположим сначала, что $L(c)$ не ограничено, т. е. существует последовательность $\{x^k\} \subset L(c)$ такая, что $|x^k| \to \infty$ $(k \to \infty)$. При этом $x_1^k + x_2^k > 0$ и $f(x^k) \leqslant c$ $\forall\, k$. Если предположить, что $|x_1^k| \to \infty$ $(k \to \infty)$, то $c \geqslant f(x^k) > (x_1^k)^2 + x_2^k > (x_1^k)^2 - x_1^k \to +\infty$ $(k \to \infty)$, что невозможно. Если же последовательность $\{x_1^k\}$ ограничена, то $x_2^k \to +\infty$ $(k \to \infty)$, поэтому при достаточно больших k имеем $c \geqslant f(x^k) > x_2^k \to +\infty$ $(k \to \infty)$, что снова невозможно. Таким образом, множество $L(c)$ ограничено.

[1)] Функция $f\colon X \to \mathbf{R}$ называется *полунепрерывной снизу в точке* $x \in X \subset \mathbf{R}^n$, если для всякой последовательности $\{x^k\} \subset X$ такой, что $\{x^k\} \to x$ $(k \to \infty)$, справедливо неравенство $\liminf\limits_{k\to\infty} f(x^k) \geqslant f(x)$. Функция f *полунепрерывна снизу на множестве* X, если она обладает этим свойством в каждой точке множества X. Понятие полунепрерывности сверху вводится аналогичным образом.

Предположим теперь, что $L(c)$ не замкнуто, т. е. существует последовательность $\{x^k\} \subset L(c)$ такая, что $\{x^k\} \to x \quad (k \to \infty)$, $x \in \in \mathbf{R}^n \setminus L(c)$. Функция f непрерывна на D, поэтому $f(x) \leqslant c$. Значит, $x \in \operatorname{cl} D \setminus D$, т. е. $x_1 + x_2 = 0$. Но тогда $c \geqslant f(x^k) \to +\infty \quad (k \to \to \infty)$, что невозможно. Таким образом, множество $L(c)$ замкнуто, а значит, компактно, и требуемый результат получается применением следствия 1.

Задача 5. Последовательность $\{x^k\} \subset D \subset \mathbf{R}^n$ называется *критической* (относительно множества D), если либо $|x^k| \to \infty$, либо $\{x^k\} \to x \in \operatorname{cl} D \setminus D \quad (k \to \infty)$. Функция $f\colon D \to \mathbf{R}$ называется *бесконечно растущей* или *коэрцитивной* (на D), если для любой критической относительно D последовательности $\{x^k\}$ справедливо равенство $\limsup\limits_{k\to\infty} f(x^k) = +\infty$. Используя схему рассуждений из примера 1, доказать следующее утверждение. Пусть $D \subset \mathbf{R}^n$ — произвольное непустое множество, $f\colon D \to \mathbf{R}$ — непрерывная бесконечно растущая на D функция. Тогда задача (1) имеет глобальное решение.

Важный пример применения следствия 1 возникает в связи со следующим понятием.

Определение 3. *Проекцией* точки $y \in \mathbf{R}^n$ на множество $X \subset \subset \mathbf{R}^n$ называется точка, ближайшая к y среди всех точек множества X, т. е. глобальное решение задачи

$$|x - y| \to \min, \quad x \in X.$$

Следствие 2. *Проекция любой точки на любое непустое замкнутое множество в* $\mathbf{R}^n$ *существует.*

Задача 6. Доказать следствие 2. Доказать, что если множество выпукло, то проекция единственна.

В тех случаях, когда множество $X \subset \mathbf{R}^n$ замкнуто и выпукло, (единственная) проекция точки $y \in \mathbf{R}^n$ на X будет обозначаться через $\pi_X(y)$.

Задача 7. Пусть X — замкнутое выпуклое множество в $\mathbf{R}^n$. Доказать, что

$$\langle y - \pi_X(y),\, x - \pi_X(y)\rangle \leqslant 0 \quad \forall\, y \in \mathbf{R}^n, \quad \forall\, x \in X.$$

Доказать, что если K — замкнутый выпуклый конус [1] в $\mathbf{R}^n$, то

$$\langle y - \pi_K(y),\, x\rangle \leqslant 0 \quad \forall\, y \in \mathbf{R}^n, \quad \forall\, x \in K.$$

[1] Множество $K \subset \mathbf{R}^n$ называется *конусом*, если вместе с любой своей точкой оно содержит проходящий через эту точку луч с началом в нуле (возможно, без начала): $tx \in K \quad \forall\, x \in K, \quad \forall\, t > 0$.

Задача 8. Пусть X — замкнутое выпуклое множество в $\mathbf{R}^n$. Доказать, что оператор проектирования на X обладает неразжимающим свойством, т. е.

$$|\pi_X(y^1) - \pi_X(y^2)| \leqslant |y^1 - y^2| \qquad \forall\, y^1,\, y^2 \in \mathbf{R}^n.$$

Задача 9. Пусть X — замкнутое выпуклое множество в $\mathbf{R}^n$. Доказать, что для произвольных $x \in X$ и $d \in \mathbf{R}^n$ функция

$$\varphi_1\colon \mathbf{R}_+ \to \mathbf{R}, \quad \varphi_1(t) = |\pi_X(x + td) - x|,$$

монотонно неубывающая, а функция

$$\varphi_2\colon (\mathbf{R}_+ \setminus \{0\}) \to \mathbf{R}, \quad \varphi_2(t) = \varphi_1(t)/t,$$

монотонно невозрастающая.

Задача 10. Пусть K — замкнутый конус в $\mathbf{R}^n$, $y \in -K^*$. Доказать, что единственной проекцией y на K является 0.

Задача 11. Пусть $A \in \mathbf{R}(n, n)$ — симметрическая матрица, $b \in \in \mathbf{R}^n$, $c \in \mathbf{R}$. Рассмотрим квадратичную функцию

$$f\colon \mathbf{R}^n \to \mathbf{R}, \quad f(x) = \langle Ax,\, x\rangle + \langle b,\, x\rangle + c.$$

Доказать следующие утверждения:

а) если задача безусловной минимизации функции f имеет локальное решение, то матрица A неотрицательно определена, причем локальное решение по необходимости является глобальным;

б) если матрица A положительно определена, то функция f является бесконечно растущей на $\mathbf{R}^n$ (см. задачу 5) и, в частности, задача минимизации f на любом непустом замкнутом допустимом множестве $D \subset \mathbf{R}^n$ имеет глобальное решение.

Убедиться, что требование замкнутости D в утверждении б) существенно.

Задача 12. С помощью критерия Сильвестра и утверждений из задачи 11 найти все значения параметра $\sigma \in \mathbf{R}$, при которых задача безусловной минимизации функции

$$f\colon \mathbf{R}^2 \to \mathbf{R}, \quad f(x) = \sigma x_1^2 + 4x_1x_2 + 4x_2^2 + x_1 + x_2,$$

имеет решение.

Задача 13. Пусть задача безусловной минимизации непрерывной на $\mathbf{R}^n$ функции $f\colon \mathbf{R}^n \to \mathbf{R}$ имеет глобальное решение. Следует ли отсюда, что имеет глобальное решение задача минимизации f на любом замкнутом множестве в $\mathbf{R}^n$?

Задача 14. Пусть задачи безусловной минимизации непрерывных на $\mathbf{R}^n$ функций f_1, $f_2\colon \mathbf{R}^n \to \mathbf{R}$ имеют глобальные решения. Следует ли отсюда, что имеет глобальное решение задача безусловной минимизации функции $f_1(\cdot) + f_2(\cdot)$?

§ 1.2. Прямые условия оптимальности

1.2.1. Касательный конус. Прямые условия оптимальности. В этом пункте рассматривается задача оптимизации

$$f(x) \to \min, \quad x \in D, \tag{1}$$

где $D \subset \mathbf{R}^n$ — заданное множество, структура которого не конкретизируется, $f\colon D \to \mathbf{R}$ — заданная функция.

Определение 1. Вектор $h \in \mathbf{R}^n$ называется *касательным* к множеству D в точке $\overline{x} \in D$, если

$$\operatorname{dist}(\overline{x} + th,\, D) = o(t), \quad t \geqslant 0.$$

Очевидно, множество всех векторов, касательных к множеству D в точке $\overline{x} \in D$, является конусом. Этот конус называется *касательным конусом* к множеству D в точке $\overline{x}$ и обозначается $T_D(\overline{x})$:

$$T_D(\overline{x}) = \{h \in \mathbf{R}^n \mid \operatorname{dist}(\overline{x} + th,\, D) = o(t),\ t \geqslant 0\} =$$

$$= \{h \in \mathbf{R}^n \mid \forall\, \{t_k\} \subset \mathbf{R}_+,\ \{t_k\} \to 0+,\ \exists\, \{h^k\} \subset \mathbf{R}^n,$$

$$\{h^k\} \to h\ (k \to \infty)\colon \overline{x} + t_k h^k \in D\ \forall\, k\}.$$

Касательный конус всегда непуст, так как содержит 0.

Наряду с касательным конусом часто рассматривают так называемый *контингентный конус* (или *конус Булигана*) к множеству D в точке $\overline{x} \in D$:

$$C_D(\overline{x}) = \{h \in \mathbf{R}^n \mid \exists\, \{t_k\} \subset \mathbf{R}_+,\ \{t_k\} \to 0+\colon \operatorname{dist}(\overline{x} + t_k h,\, D) = o(t_k)\} =$$

$$= \{h \in \mathbf{R}^n \mid \exists\, \{t_k\} \subset \mathbf{R}_+,\ \{t_k\} \to 0+,\ \exists\, \{h^k\} \subset \mathbf{R}^n,$$

$$\{h^k\} \to h\ (k \to \infty)\colon \overline{x} + t_k h^k \in D\ \forall\, k\}.$$

Очевидно, $T_D(\overline{x}) \subset C_D(\overline{x})$, но, вообще говоря, не наоборот. Вместе с тем обычно конструктивное описание $T_D(\overline{x})$ или $C_D(\overline{x})$ для более конкретных D удается получить лишь в таких предположениях, в которых эти два конуса не различаются (впрочем, бывают и исключения). Если $\overline{x} \in \operatorname{int} D$, то $T_D(\overline{x}) = C_D(\overline{x}) = \mathbf{R}^n$.

Задача 1. Привести пример несовпадения касательного и контингентного конусов.

Задача 2. Доказать, что контингентный и касательный конусы к любому множеству в любой точке замкнуты.

В случае выпуклого множества D контингентный и касательный конусы к нему в любой его точке совпадают и легко вычисляются.

Предложение 1. *Пусть* $D \subset \mathbf{R}^n$ — *выпуклое множество*, $\overline{x} \in D$. *Тогда* $T_D(\overline{x}) = C_D(\overline{x}) = \operatorname{cl}\operatorname{cone}(D - \overline{x})$.

Задача 3. Доказать предложение 1.

Значение понятия контингентного конуса состоит в том, что именно его наиболее естественно использовать при формулировке необходимого и достаточного условий первого порядка оптимальности в задаче (1).

Теорема 1. *Пусть* $D \subset \mathbf{R}^n$ — *произвольное множество*, *а функция* $f\colon D \to \mathbf{R}$ *дифференцируема в точке* [1] $\overline{x} \in D$.

Тогда если $\overline{x}$ *является локальным решением задачи* (1), *то*

$$\langle f'(\overline{x}),\, h\rangle \geqslant 0 \quad \forall\, h \in C_D(\overline{x}). \tag{2}$$

Доказательство. Зафиксируем произвольный $h \in C_D(\overline{x})$ и отвечающие ему последовательности $\{t_k\} \subset \mathbf{R}_+$ и $\{h^k\} \subset D$ такие, что $\{t_k\} \to 0+$, $\{h^k\} \to h$ $(k \to \infty)$, $\overline{x} + t_k h^k \in D$ $\forall k$. Если $\overline{x}$ — локальное решение задачи (1), то для любого достаточно большого k

$$0 \leqslant f(\overline{x} + t_k h^k) - f(\overline{x}) = t_k \langle f'(\overline{x}),\, h^k\rangle + o(t_k).$$

Разделив левую и правую части этого неравенства на t_k и перейдя к пределу при $k \to \infty$, получим (2). □

Определение 2. Точка $\overline{x} \in \mathbf{R}^n$ называется *стационарной точкой* задачи (1), если $\overline{x} \in D$ и выполнено (2).

Таким образом, теорема 1 утверждает, что всякое локальное решение задачи (1), являющееся точкой гладкости ее целевой функции, является стационарной точкой этой задачи. Обратное, вообще говоря, неверно (если только не предполагать, что D — выпуклое множество, а f — выпуклая функция на D).

Необходимое условие оптимальности, приведенное в теореме 1, называют *прямым* в том смысле, что в нем фигурируют лишь переменные исходной задачи (также называемые *прямыми переменными*). При этом (2) эквивалентным образом переписывается в виде

$$f'(\overline{x}) \in (C_D(\overline{x}))^*. \tag{3}$$

Как будет показано ниже, вычисление конуса $(C_D(\overline{x}))^*$ для более конкретных D обычно связано с использованием вспомогательных *двойственных переменных*, поэтому получаемые на этом пути

[1] Здесь и далее в подобных ситуациях предполагается, что если $\overline{x} \notin \operatorname{int} D$, то функция f определена не только на D, но и на некоторой окрестности $\overline{x}$. Такое предположение позволяет говорить о дифференцируемости f в точке $\overline{x}$.

необходимые условия оптимальности называют *прямодвойственным*. Известны результаты о том, что необходимое условие (3) существенно усилено быть не может в том смысле, что нельзя существенно сузить конус в правой части (3).

В случае выпуклого допустимого множества из предложения 1 и теоремы 1 вытекает следующее необходимое условие оптимальности.

Следствие 1. *Пусть* $D \subset \mathbf{R}^n$ *— выпуклое множество, а функция* $f\colon D \to \mathbf{R}$ *дифференцируема в точке* $\overline{x} \in D$.

Тогда если $\overline{x}$ *является локальным решением задачи* (1), *то*

$$\langle f'(\overline{x}),\, x - \overline{x}\rangle \geqslant 0 \quad \forall\, x \in D. \tag{4}$$

Задача 4. Показать, что если в условиях следствия 1 множество D замкнуто, то (4) эквивалентно выполнению равенства

$$\pi_D(\overline{x} - t f'(\overline{x})) = \overline{x}$$

для некоторого $t > 0$, причем выполнение этого равенства для некоторого $t > 0$ влечет его выполнение для любого $t > 0$.

Приведенное в теореме 1 необходимое условие оптимальности близко́ к достаточному условию первого порядка в том смысле, что последнее получается заменой для $h \neq 0$ нестрогого неравенства в (2) строгим.

Теорема 2. *Пусть выполнены условия теоремы* 1.

Тогда если

$$\langle f'(\overline{x}),\, h\rangle > 0 \quad \forall\, h \in C_D(\overline{x}) \setminus \{0\}, \tag{5}$$

то $\overline{x}$ *является строгим локальным решением задачи* (1).

Доказательство. От противного: предположим, что существует последовательность $\{x^k\} \subset D \setminus \{\overline{x}\}$ такая, что $f(x^k) \leqslant f(\overline{x})$ $\forall\, k$ и $\{x^k\} \to \overline{x}$ $(k \to \infty)$. Последовательность $\{(x^k - \overline{x})/|x^k - \overline{x}|\}$ лежит на единичной сфере в $\mathbf{R}^n$, которая является компактом, а значит, эта последовательность имеет предельную точку. Без ограничения общности можем считать, что вся последовательность сходится к некоторому $h \in \mathbf{R}^n$, $|h| = 1$. Очевидно, что при этом $h \in C_D(\overline{x}) \setminus \{0\}$.

Далее, $\forall\, k$ имеем

$$0 \geqslant f(x^k) - f(\overline{x}) = \langle f'(\overline{x}),\, x^k - \overline{x}\rangle + o(|x^k - \overline{x}|).$$

Разделив левую и правую части этого неравенства на $|x^k - \overline{x}|$ и перейдя к пределу при $k \to \infty$, получим

$$\langle f'(\overline{x}),\, h\rangle \leqslant 0,$$

что противоречит (5). $\square$

Заметим, что достаточное условие (5) может выполняться только в том случае, когда конус $C_D(\overline{x})$ является *острым*, т. е. не содержит нетривиальных линейных подпространств. Подобное нетипично, например, для случая, когда допустимое множество задается скалярными функциональными ограничениями, число которых меньше числа переменных, поэтому область применимости теоремы 2 весьма ограничена.

Проблема вычисления конусов $C_D(\overline{x})$ и $(C_D(\overline{x}))^*$ (заметим, что иногда конус $(C_D(\overline{x}))^*$ может быть вычислен без вычисления самого конуса $C_D(\overline{x})$) для задач более специальных классов важна не только потому, что ее решение позволяет привести полученные условия оптимальности к пригодной для практического использования форме, но и потому, что оно лежит в основе дальнейшего развития локальной теории таких задач (например, в основе условий второго порядка оптимальности). Эти вопросы являются центральными в § 1.3 и § 1.4. В следующем пункте рассматривается простейший случай, когда $D = \mathbf{R}^n$.

1.2.2. Условия оптимальности в задаче безусловной оптимизации. Будем рассматривать задачу безусловной оптимизации

$$f(x) \to \min, \quad x \in \mathbf{R}^n, \tag{6}$$

где $f\colon \mathbf{R}^n \to \mathbf{R}$ — заданная функция. Из теоремы 1 немедленно вытекает следующий результат, впервые (в иных терминах) указанный Ферма.

Теорема 3. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *дифференцируема в точке* $\overline{x} \in \mathbf{R}^n$.

Тогда если $\overline{x}$ *является локальным решением задачи* (6), *то*

$$f'(\overline{x}) = 0. \tag{7}$$

Доказательство. Очевидно, контингентный конус к $\mathbf{R}^n$ в любой точке $\overline{x} \in \mathbf{R}^n$ совпадает со всем $\mathbf{R}^n$, поэтому согласно теореме 1

$$\langle f'(\overline{x}),\, h\rangle \geqslant 0 \quad \forall\, h \in \mathbf{R}^n.$$

Если предположить, что $\langle f'(\overline{x}),\, h\rangle > 0$ для некоторого $h \in \mathbf{R}^n$, то немедленно приходим к противоречию, поскольку $\langle f'(\overline{x}), -h\rangle < 0$. □

Определение 3. Точка $\overline{x} \in \mathbf{R}^n$ называется *стационарной точкой* задачи (6) (или *критической точкой* функции f), если выполнено (7).

Таким образом, являющееся точкой гладкости функции f локальное решение задачи (6) (впрочем, как и локальное решение соответствующей задачи на максимум) является стационарной точкой этой задачи, но, вообще говоря, не наоборот. Очевидно, теорема 2

неприменима в данном контексте, и возникает необходимость в более тонкой характеризации локального решения, в терминах второй производной функции f. Такая характеризация предлагается в двух следующих теоремах, содержащих соответственно необходимое и достаточное условия второго порядка оптимальности.

Теорема 4. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *дважды дифференцируема в точке* $\overline{x} \in \mathbf{R}^n$.

Тогда если $\overline{x}$ *является локальным решение задачи* (6), *то матрица Гессе функции* f *в точке* $\overline{x}$ *неотрицательно определена*:

$$\langle f''(\overline{x})h,\, h\rangle \geqslant 0 \quad \forall\, h \in \mathbf{R}^n. \tag{8}$$

Доказательство. Фиксируем $h \in \mathbf{R}^n$. Если $\overline{x}$ — локальное решение задачи (6), то для любого достаточно малого $t > 0$

$$0 \leqslant f(\overline{x} + th) - f(\overline{x}) = \langle f'(\overline{x}),\, th\rangle + \frac{1}{2}\,\langle f''(\overline{x})th,\, th\rangle + o(t^2) =$$

$$= \frac{t^2}{2}\,\langle f''(\overline{x})h,\, h\rangle + o(t^2),$$

где принято во внимание равенство (7). Разделив левую и правую части полученного неравенства на t^2 и перейдя к пределу при $t \to$ $\to 0+$, получим (8). $\square$

Теорема 5. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *дважды дифференцируема в точке* $\overline{x} \in \mathbf{R}^n$.

Тогда если выполнено (7) *и матрица Гессе функции* f *в точке* $\overline{x}$ *положительно определена, т. е.*

$$\langle f''(\overline{x})h,\, h\rangle > 0 \quad \forall\, h \in \mathbf{R}^n \setminus \{0\}, \tag{9}$$

то $\overline{x}$ *является строгим локальным решением задачи* (6).

Доказательство. От противного: предположим, что существует последовательность $\{x^k\} \subset \mathbf{R}^n \setminus \{\overline{x}\}$ такая, что $f(x^k) \leqslant f(\overline{x})$ $\forall\, k$ и $\{x^k\} \to \overline{x}$ $(k \to \infty)$. Последовательность $\{(x^k - \overline{x})/|x^k - \overline{x}|\}$ без ограничения общности можно считать сходящейся к некоторому $h \in \mathbf{R}^n \setminus \{0\}$.

Далее, $\forall\, k$ имеем

$$0 \geqslant f(x^k) - f(\overline{x}) = \langle f'(\overline{x}),\, x^k - \overline{x}\rangle + \frac{1}{2}\langle f''(\overline{x})(x^k - \overline{x}),\, x^k - \overline{x}\rangle +$$

$$+\ o(|x^k - \overline{x}|^2) = \frac{1}{2}\langle f''(\overline{x})(x^k - \overline{x}),\, x^k - \overline{x}\rangle + o(|x^k - \overline{x}|^2),$$

где принято во внимание (7). Разделив левую и правую части полученного неравенства на $|x^k - \overline{x}|^2$ и перейдя к пределу при $k \to \infty$, получим

$$\langle f''(\overline{x})h,\, h\rangle \leqslant 0,$$

что противоречит (9). $\square$

Подчеркнем, что условия (7), (8) не являются достаточными для оптимальности. Например, для функции $f\colon \mathbf{R} \to \mathbf{R}$, $f(x) = x^3$, эти условия выполнены в точке $\overline{x} = 0$, которая не является локальным решением задачи (6). Аналогично, условия (7), (9) не являются необходимыми для оптимальности. Для функции $f\colon \mathbf{R} \to \mathbf{R}$, $f(x) = x^4$, эти условия не выполнены в точке $\overline{x} = 0$, хотя эта точка и является решением задачи (6).

В случае задачи на максимум теоремы 4 и 5 сохранят силу, если поменять знаки неравенств в (8) и (9) на противоположные. Кроме того, теоремы 3–5 полностью сохраняют силу и для задачи условной оптимизации (1), если только дополнительно предположить, что $\overline{x} \in \operatorname{int} D$: при локальных рассмотрениях этот случай ничем не отличается от случая задачи безусловной оптимизации (в частности, как отмечалось выше, $C_D(\overline{x}) = \mathbf{R}^n$).

Пример 1. Найдем экстремальные точки функции

$$f\colon \mathbf{R}^2 \to \mathbf{R}, \quad f(x) = x_1^3 + x_2^3 - 3x_1x_2,$$

на всем $\mathbf{R}^2$.

Для определения стационарных точек имеем систему уравнений

$$\frac{\partial f}{\partial x_1}(x) = 3x_1^2 - 3x_2 = 0, \qquad \frac{\partial f}{\partial x_2}(x) = 3x_2^2 - 3x_1 = 0.$$

Ее решениями являются $x^1 = 0$ и $x^2 = (1, 1)$. Нетрудно убедиться, что глобальных экстремумов здесь нет. Для выявления локальных экстремумов воспользуемся условиями второго порядка оптимальности. Вычислим $f''(x^1)$ и $f''(x^2)$:

$$f''(x^1) = \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix}, \quad f''(x^2) = \begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix}.$$

Очевидно, матрица $f''(x^1)$ не является знакоопределенной, т. е. в силу теоремы 4 x^1 не является локальным экстремумом. В то же время матрица $f''(x^2)$ положительно определена, т. е. в силу теоремы 5 x^2 — единственная точка строгого локального минимума, а локальных максимумов нет.

Пример 2. Найдем глобальные решения задачи

$$f(x) = x_1 + \frac{x_2^2}{4x_1} + \frac{x_3^2}{x_2} + \frac{2}{x_3} \to \min,$$

$$x \in D = \{x \in \mathbf{R}^3 \mid x_1 > 0,\ x_2 > 0,\ x_3 > 0\}.$$

Множество D открыто, поэтому всякая допустимая точка принадлежит $\operatorname{int} D$. Учитывая это, легко находим единственную стационарную точку $\overline{x} = (1/2,\ 1,\ 1)$.

Задача 5. Для задачи из примера 2 доказать, что функция f является бесконечно растущей на D (см. задачу 1.5).

Из утверждений, сформулированных в задачах 1.5 и 5, следует, что единственная стационарная точка в рассматриваемом примере является глобальным решением.

Задача 6. Построить пример функции $f\colon \mathbf{R}^n \to \mathbf{R}$, дифференцируемой на $\mathbf{R}^n$ и имеющей ровно одну критическую точку, которая является ее локальным, но не глобальным экстремумом на $\mathbf{R}^n$. Существует ли такой пример при $n = 1$?

Задача 7. Доказать, что при любом значении параметра $\sigma > 1$ система уравнений

$$\sigma \cos x_1 \sin x_2 + x_1 e^{x_1^2+x_2^2} = 0, \quad \sigma \sin x_1 \cos x_2 + x_2 e^{x_1^2+x_2^2} = 0$$

относительно $x \in \mathbf{R}^2$ имеет ненулевое решение.

§ 1.3. Задача с ограничениями-равенствами

Пусть $f\colon \mathbf{R}^n \to \mathbf{R}$ — заданная функция, $F\colon \mathbf{R}^n \to \mathbf{R}^l$ — заданное отображение. В этом параграфе рассматривается задача оптимизации с чистыми ограничениями-равенствами

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in \mathbf{R}^n \mid F(x) = 0\}. \tag{2}$$

Такая задача — классическая постановка математического анализа. Принцип Лагранжа, выражающий необходимое условие первого порядка оптимальности для задачи (1), (2), несомненно входит в число наиболее выдающихся и красивых открытий в истории математики. Кроме того, этот принцип послужил идейной основой для построения теории более общих задач (со смешанными ограничениями; см. § 1.4).

1.3.1. Теорема о неявной функции. Теорема Люстерника. Прежде чем переходить собственно к задаче (1), (2), рассмотрим некоторые результаты о локальной структуре множества решений системы нелинейных уравнений, которые будут нужны ниже. Следующая классическая теорема о неявной функции — важнейший инструмент многомерного нелинейного анализа (ее доказательство см., например, в [23]).

Теорема 1. *Пусть отображение* $F\colon \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^n$ *дифференцируемо в некоторой окрестности точки* $(\overline{\sigma}, \overline{x}) \in \mathbf{R}^s \times \mathbf{R}^n$, *причем его частная производная по* x *непрерывна в точке* $(\overline{\sigma}, \overline{x})$. *Пусть* $F(\overline{\sigma}, \overline{x}) = 0$ *и* $\det \dfrac{\partial F}{\partial x}(\overline{\sigma}, \overline{x}) \neq 0$.

Тогда найдутся окрестности U точки $\overline{\sigma}$ и V точки $\overline{x}$, для которых существует единственное отображение $\chi\colon U \to \mathbf{R}^n$ такое, что $\chi(U) \subset V$ и

$$F(\sigma, \chi(\sigma)) = 0 \quad \forall\, \sigma \in U. \tag{3}$$

При этом по необходимости $\chi(\overline{\sigma}) = \overline{x}$, и если окрестность U достаточно мала, то отображение χ дифференцируемо на U, причем

$$\chi'(\sigma) = -\left(\frac{\partial F}{\partial x}(\sigma, \chi(\sigma))\right)^{-1} \frac{\partial F}{\partial \sigma}(\sigma, \chi(\sigma)), \quad \sigma \in U.$$

В частности, производная отображения χ непрерывна в точке $\overline{\sigma}$.

Известны многочисленные модификации приведенной классической теоремы. Ниже понадобится следующая модификация, называемая теоремой о существовании неявной функции.

Теорема 2. *Пусть отображение $F\colon \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l$ дифференцируемо в некоторой окрестности точки $(\overline{\sigma}, \overline{x}) \in \mathbf{R}^s \times \mathbf{R}^n$, причем его частная производная по x непрерывна в точке $(\overline{\sigma}, \overline{x})$. Пусть $F(\overline{\sigma}, \overline{x}) = 0$ и $\operatorname{rank} \dfrac{\partial F}{\partial x}(\overline{\sigma}, \overline{x}) = l$.*

Тогда найдутся окрестность U точки $\overline{\sigma}$, отображение $\chi\colon U \to \mathbf{R}^n$ и число $M > 0$ такие, что выполнено (3) и

$$|\chi(\sigma) - \overline{x}| \leqslant M|F(\sigma, \overline{x})| \quad \forall\, \sigma \in U. \tag{4}$$

Доказательство. Зафиксируем произвольную матрицу $A \in \mathbf{R}(n-l, n)$, удовлетворяющую следующим требованиям:

$$\ker \frac{\partial F}{\partial x}(\overline{\sigma}, \overline{x}) \cap \ker A = \{0\}.$$

Очевидно, этим условиям удовлетворяет любая матрица, строки которой дополняют строки матрицы $\dfrac{\partial F}{\partial x}(\overline{\sigma}, \overline{x})$ (которые линейно независимы) до базиса в $\mathbf{R}^n$. Введем отображение

$$\Phi\colon \mathbf{R}^s \times \mathbf{R}^n \to \mathbf{R}^l \times \mathbf{R}^{n-l}, \quad \Phi(\sigma, x) = \begin{pmatrix} F(\sigma, x) \\ A(x - \overline{x}) \end{pmatrix}.$$

Для этого отображения в точке $(\overline{\sigma}, \overline{x})$ выполнены все условия теоремы 1. Поэтому найдутся окрестность U точки $\overline{\sigma}$ и непрерывное в точке $\overline{\sigma}$ отображение $\chi\colon U \to \mathbf{R}^n$ такие, что $\chi(\overline{\sigma}) = \overline{x}$ и

$$\Phi(\sigma, \chi(\sigma)) = \begin{pmatrix} F(\sigma, \chi(\sigma)) \\ A(\chi(\sigma) - \overline{x}) \end{pmatrix} = 0 \quad \forall\, \sigma \in U,$$

и, в частности, выполнено (3). Заметим, что

$$|\xi| = \left| \left(\frac{\partial \Phi}{\partial x} (\overline{\sigma}, \overline{x}) \right)^{-1} \frac{\partial \Phi}{\partial x} (\overline{\sigma}, \overline{x})\, \xi \right| \leqslant$$

$$\leqslant \left\| \left(\frac{\partial \Phi}{\partial x} (\overline{\sigma}, \overline{x}) \right)^{-1} \right\| \left| \frac{\partial \Phi}{\partial x} (\overline{\sigma}, \overline{x})\, \xi \right| \quad \forall\, \xi \in \mathbf{R}^n.$$

Привлекая теорему о среднем [1)], с учетом непрерывности частной производной F по x в точке $(\overline{\sigma}, \overline{x})$ отсюда выводим, что если окрестность U выбрана достаточно малой, то

$$|F(\sigma, \overline{x})| = |\Phi(\sigma, \overline{x})| \geqslant$$

$$\geqslant \left| \frac{\partial \Phi}{\partial x} (\overline{\sigma}, \overline{x})(\chi(\sigma) - \overline{x}) \right| - \left| \Phi(\sigma, \chi(\sigma)) - \Phi(\sigma, \overline{x}) - \frac{\partial \Phi}{\partial x} (\overline{\sigma}, \overline{x})(\chi(\sigma) - \overline{x}) \right| \geqslant$$

$$\geqslant \left\| \left(\frac{\partial \Phi}{\partial x} (\overline{\sigma}, \overline{x}) \right)^{-1} \right\|^{-1} |\chi(\sigma) - \overline{x}| -$$

$$- \sup_{t \in [0,\, 1]} \left\| \frac{\partial \Phi}{\partial x} (\sigma,\, t\chi(\sigma) + (1-t)\overline{x}) - \frac{\partial \Phi}{\partial x} (\overline{\sigma}, \overline{x}) \right\| |\chi(\sigma) - \overline{x}| \geqslant$$

$$\geqslant \frac{1}{2} \left\| \left(\frac{\partial \Phi}{\partial x} (\overline{\sigma}, \overline{x}) \right)^{-1} \right\|^{-1} |\chi(\sigma) - \overline{x}| \quad \forall\, \sigma \in U.$$

Это и есть оценка (4) при $M = 2 \left\| \left(\dfrac{\partial \Phi}{\partial x} (\overline{\sigma}, \overline{x}) \right)^{-1} \right\|$. $\square$

Следствием теоремы о существовании неявной функции является теорема об оценке расстояния до множества D, заданного в (2).

Теорема 3. *Пусть отображение* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *дифференцируемо в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$, *причем его производная непрерывна в точке* $\overline{x}$. *Пусть* $F(\overline{x}) = 0$ *и* $\operatorname{rank} F'(\overline{x}) = l$.

Тогда найдутся окрестности U *точки* $\overline{x}$ *и число* $M > 0$ *такие, что для заданного в* (2) *множества* D *справедлива оценка*

$$\operatorname{dist}(x,\, D) \leqslant M |F(x)| \quad \forall\, x \in U.$$

Доказательство. Введем отображение

$$\Phi\colon \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^l, \quad \Phi(x,\, \xi) = F(x + \xi).$$

Применяя к этому отображению в точке $(\overline{x},\, 0)$ теорему 2, получаем требуемое. $\square$

Наконец, следствием теоремы об оценке расстояния является знаменитая теорема Люстерника о касательном подпространстве.

[1)] Согласно теореме о среднем, если отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ дифференцируемо в каждой точке отрезка, соединяющего точки $x^1, x^2 \in \mathbf{R}^n$, то $|F(x^1) - F(x^2)| \leqslant \sup\limits_{t \in [0,\, 1]} \|F'(tx^1 + (1-t)x^2)\| |x^1 - x^2|$.

Теорема 4. *Пусть отображение* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *дифференцируемо в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$, *причем его производная непрерывна в точке* $\overline{x}$. *Пусть* $F(\overline{x}) = 0$.

Тогда для заданного в (2) *множества* D *справедливо следующее*:
а) $C_D(\overline{x}) \subset \ker F'(\overline{x})$;
б) *если* $\operatorname{rank} F'(\overline{x}) = l$, *то* $T_D(\overline{x}) = C_D(\overline{x}) = \ker F'(\overline{x})$.

Задача 1. Используя теорему 3, доказать теорему 4 (для доказательства утверждения а) провести рассуждения, аналогичные использованным при доказательстве теоремы 1.2.1).

1.3.2. Принцип Лагранжа. Теперь обратимся собственно к задаче (1), (2). Пусть $\overline{x} \in D$ — локальное решение такой задачи, причем функция f и отображение F достаточно гладки и

$$\operatorname{rank} F'(\overline{x}) = l. \tag{5}$$

Заметим, что это условие, называемое *условием регулярности* ограничений задачи (1), (2) в точке $\overline{x}$, эквивалентным образом записывается в виде $\operatorname{im} F'(\overline{x}) = \mathbf{R}^l$, или $\ker(F'(\overline{x}))^{\mathrm{T}} = (\operatorname{im} F'(\overline{x}))^{\perp} = \{0\}$. Согласно теореме Люстерника $C_D(\overline{x}) = \ker F'(\overline{x})$.

Лемма 1. *Для всякой матрицы* $A \in \mathbf{R}(l, n)$

$$(\ker A)^* = (\ker A)^{\perp} = \operatorname{im} A^{\mathrm{T}}.$$

Задача 2. Доказать лемму 1.

Привлекая теорему 1.2.1, сопутствующие ей комментарии (см. формулу (1.2.3)), а также лемму 1, получаем прямодвойственное необходимое условие первого порядка оптимальности в задаче (1), (2): существует $\overline{\lambda} \in \mathbf{R}^l$ такой, что

$$f'(\overline{x}) = -(F'(\overline{x}))^{\mathrm{T}}\overline{\lambda}$$

(введение знака минус в правую часть совершенно не обязательно, но удобно для дальнейшего). В этом и состоит классический принцип Лагранжа, выражаемый следующей теоремой, обычно формулируемой в терминах так называемой *функции Лагранжа* задачи (1), (2)

$$L\colon \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}, \quad L(x, \lambda) = f(x) + \langle \lambda, F(x) \rangle.$$

Заметим, что

$$\frac{\partial L}{\partial x}(x, \lambda) = f'(x) + (F'(x))^{\mathrm{T}}\lambda, \quad \frac{\partial L}{\partial \lambda}(x, \lambda) = F(x),$$

$$x \in \mathbf{R}^n, \quad \lambda \in \mathbf{R}^l.$$

Теорема 5. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *дифференцируема в точке* $\overline{x} \in \mathbf{R}^n$, *отображение* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *дифференцируемо в некоторой окрестности этой точки, причем его производная непрерывна в точке* $\overline{x}$.

Тогда если $\overline{x}$ *является локальным решением задачи* (1), (2) *и выполнено* (5), *то найдется элемент* $\overline{\lambda} \in \mathbf{R}^l$ *такой, что*

$$\frac{\partial L}{\partial x}(\overline{x}, \overline{\lambda}) = 0. \tag{6}$$

В приведенной теореме, как и в теореме Люстерника, важнейшую роль играет условие регулярности (5); без него утверждения обеих теорем не имеют места. Например, для функции $F\colon \mathbf{R} \to \mathbf{R}$, $F(x) = x^2$, в точке $\overline{x} = 0$ справедливо $F'(\overline{x}) = 0$, $\ker F'(\overline{x}) = \mathbf{R}$, но $D = \{\overline{x}\}$ и $C_D(\overline{x}) = \{0\}$. Точка $\overline{x}$ одноточечного множества D является единственным решением задачи (1), (2) с произвольной целевой функцией $f\colon \mathbf{R} \to \mathbf{R}$. С другой стороны, если эта функция такова, что $f'(\overline{x}) \neq 0$, то (6) не имеет места ни при каком $\overline{\lambda} \in \mathbf{R}$. Обобщения теоремы Люстерника и принципа Лагранжа, справедливые в более слабых условиях, чем (5), можно найти в [21, 22].

Определение 1. Точка $\overline{x} \in \mathbf{R}^n$ называется *стационарной точкой* задачи (1), (2), если $\overline{x} \in D$ и выполнено (6) при некотором $\overline{\lambda} \in \mathbf{R}^l$. Элемент $\overline{\lambda}$ при этом называется *множителем Лагранжа*, отвечающим стационарной точке $\overline{x}$.

Таким образом, при выполнении определенных требований гладкости локальное решение задачи (1), (2) (как и локальное решение соответствующей задачи на максимум), в котором выполнено условие регулярности (5), является стационарной точкой этой задачи, но, вообще говоря, не наоборот. Очевидно, условие регулярности гарантирует единственность множителя Лагранжа, отвечающего стационарной точке $\overline{x}$. Очевидно также, что при выполнении (5) достаточное условие первого порядка оптимальности, даваемое теоремой 1.2.2, применимо только в том случае, когда $\ker F'(\overline{x}) = \{0\}$ (при этом по необходимости $n \leqslant l$ и из теоремы Люстерника легко следует, что $\overline{x}$ — изолированная точка допустимого множества D). Условия второго порядка оптимальности для задачи (1), (2) излагаются в следующем пункте.

Систему уравнений

$$\frac{\partial L}{\partial x}(x, \lambda) = 0, \quad F(x) = 0$$

относительно $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$, характеризующую стационарные точки задачи (1), (2) и отвечающие им множители Лагранжа, называют

системой Лагранжа. Заметим, что систему Лагранжа можно записать в эквивалентном виде:

$$L'(x,\,\lambda) = 0$$

(слева стоит полная производная). Число неизвестных в этой системе равно числу уравнений, поэтому естественно рассчитывать на то, что она имеет изолированные (друг от друга) решения. Это обстоятельство будет играть важную роль в § 4.3.

Геометрический смысл принципа Лагранжа состоит в том, что при выполнении соответствующих требований гладкости и условия регулярности (5) в локальном решении $\overline{x}$ задачи (1), (2) градиент $f'(\overline{x})$ представим в виде линейной комбинации строк матрицы Якоби $F'(\overline{x})$ ограничивающего отображения. Разумеется, теорема Ферма (теорема 1.2.3) может рассматриваться как частный случай принципа Лагранжа при $l = 0$.

В некоторых простейших случаях принцип Лагранжа позволяет решить задачу (1), (2) аналитически.

Пример 1. Пусть требуется найти экстремальные точки функции

$$f\colon \mathbf{R}^2 \to \mathbf{R}, \quad f(x) = \frac{1}{3}\,x_1^3 + x_2,$$

на единичной окружности

$$D = \{x \in \mathbf{R}^2 \,\big|\, x_1^2 + x_2^2 = 1\}.$$

Существование точек глобального максимума и минимума вытекает из теоремы Вейерштрасса. Очевидно, для функции $F\colon \mathbf{R}^2 \to \mathbf{R}$, $F(x) = x_1^2 + x_2^2 - 1$, справедливо $F'(x) \neq 0 \quad \forall\, x \in D$, т. е. условие регулярности выполнено в каждой допустимой точке. Выписываем функцию Лагранжа:

$$L(x,\,\lambda) = \frac{1}{3}\,x_1^3 + x_2 + \lambda(x_1^2 + x_2^2 - 1), \quad x \in \mathbf{R}^2, \quad \lambda \in \mathbf{R},$$

и соответствующую систему Лагранжа:

$$\frac{\partial L}{\partial x_1}\,(x,\,\lambda) = x_1^2 + 2\lambda x_1 = 0, \quad \frac{\partial L}{\partial x_2}\,(x,\,\lambda) = 1 + 2\lambda x_2 = 0,$$

$$F(x) = x_1^2 + x_2^2 - 1 = 0.$$

Возможны два случая. Если $x_1 = 0$, то из третьего уравнения следует, что $x_2 = \pm 1$, а из второго, что $\lambda = \mp 1/2$. Если же $x_1 \neq 0$, то, выражая из первых двух уравнений x_1 и x_2 через λ и подставляя результат в третье уравнение, приходим к уравнению только относительно λ: $(2\lambda)^2 + 1/(2\lambda)^2 = 1$. Элементарно проверяется, что последнее уравнение неразрешимо, т. е. этот случай стационарных точек не дает.

Таким образом, задача имеет две стационарные точки: $x^1 = (0, 1)$ и $x^2 = (0, -1)$. Сравнивая значения целевой функции в данных точках, убеждаемся, что x^1 — точка глобального максимума, а x^2 — глобального минимума, и других экстремумов нет.

1.3.3. Условия второго порядка оптимальности. Начнем с необходимого условия второго порядка.

Теорема 6. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображение* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *дважды дифференцируемы в точке* $\overline{x} \in \mathbf{R}^n$.

Тогда если $\overline{x}$ *является локальным решением задачи* (1), (2) *и выполнено* (5), *то для единственного элемента* $\overline{\lambda} \in \mathbf{R}^l$, *удовлетворяющего* (6), *имеет место следующее неравенство*:

$$\left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\,\overline{\lambda})h,\,h \right\rangle \geqslant 0 \quad \forall\, h \in \ker F'(\overline{x}). \tag{7}$$

Доказательство. Фиксируем $h \in \ker F'(\overline{x})$. Тогда, согласно теореме Люстерника, $h \in T_D(\overline{x})$, т. е. существует отображение $r\colon \mathbf{R} \to \mathbf{R}^n$ такое, что

$$F(\overline{x} + th + r(t)) = 0 \quad \forall\, t \in \mathbf{R}, \quad |r(t)| = o(t).$$

Поэтому если $\overline{x}$ — локальное решение задачи (1), (2), то для любого достаточно близкого к нулю t

$$\begin{aligned}
&0 \leqslant f(\overline{x} + th + r(t)) - f(\overline{x}) = \\
&= f(\overline{x} + th + r(t)) + \langle \overline{\lambda},\, F(\overline{x} + th + r(t))\rangle - f(\overline{x}) - \langle \overline{\lambda},\, F(\overline{x})\rangle = \\
&= L(\overline{x} + th + r(t),\, \overline{\lambda}) - L(\overline{x},\, \overline{\lambda}) = \\
&= \left\langle \frac{\partial L}{\partial x}(\overline{x}, \overline{\lambda}), th + r(t) \right\rangle + \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda})(th + r(t)), th + r(t) \right\rangle + o(t^2) = \\
&= \frac{t^2}{2}\left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\,\overline{\lambda})h,\,h \right\rangle + o(t^2)
\end{aligned}$$

(здесь учтено (6)). Разделив левую и правую части полученного неравенства на t^2 и перейдя к пределу при $t \to 0$, получим (7). □

В достаточном условии второго порядка используется не условие регулярности (5), а сам факт существования множителя Лагранжа.

Теорема 7. *Пусть выполнены условия теоремы* 6.

Тогда если $\overline{x} \in D$ *и существует удовлетворяющий* (6) *элемент* $\overline{\lambda} \in \mathbf{R}^l$, *для которого имеет место*

$$\left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\,\overline{\lambda})h,\,h \right\rangle > 0 \quad \forall\, h \in \ker F'(\overline{x}) \setminus \{0\}, \tag{8}$$

то $\overline{x}$ *является строгим локальным решением задачи* (1), (2).

Доказательство. От противного: предположим, что существует последовательность $\{x^k\} \subset D \setminus \{\overline{x}\}$ такая, что $f(x^k) \leqslant f(\overline{x})$ $\forall\, k$ и $\{x^k\} \to \overline{x}$ $(k \to \infty)$. Последовательность $\{(x^k - \overline{x})/|x^k - \overline{x}|\}$ можно считать сходящейся к некоторому $h \in \mathbf{R}^n \setminus \{0\}$. Очевидно, что $h \in C_D(\overline{x}) \setminus \{0\}$; значит, в силу утверждения а) теоремы Люстерника $h \in \ker F'(\overline{x}) \setminus \{0\}$.

Далее, $\forall\, k$ имеем

$$0 \geqslant f(x^k) - f(\overline{x}) = \\ = \langle f'(\overline{x}),\, x^k - \overline{x}\rangle + \frac{1}{2}\,\langle f''(\overline{x})(x^k - \overline{x}),\, x^k - \overline{x}\rangle + o(|x^k - \overline{x}|^2),$$

$$0 = F(x^k) - F(\overline{x}) = \\ = F'(\overline{x})(x^k - \overline{x}) + \frac{1}{2}\,F''(\overline{x})[x^k - \overline{x},\, x^k - \overline{x}] + o(|x^k - \overline{x}|^2).$$

Складывая левую и правую части первого соотношения с домноженными скалярно на $\overline{\lambda}$ соответственно левой и правой частями второго соотношения и принимая во внимание (6), получаем: $\forall\, k$

$$0 \geqslant \langle f'(\overline{x}),\, x^k - \overline{x}\rangle + \langle \overline{\lambda},\, F'(\overline{x})(x^k - \overline{x})\rangle + \\ + \frac{1}{2}\,\langle f''(\overline{x})(x^k - \overline{x}),\, x^k - \overline{x}\rangle + \\ + \frac{1}{2}\,\langle \overline{\lambda},\, F''(\overline{x})[x^k - \overline{x},\, x^k - \overline{x}]\rangle + + o(|x^k - \overline{x}|^2) = \\ = \left\langle \frac{\partial L}{\partial x}\,(\overline{x},\, \overline{\lambda}),\, x^k - \overline{x} \right\rangle + \\ + \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\, \overline{\lambda})(x^k - \overline{x}),\, x^k - \overline{x} \right\rangle + o(|x^k - \overline{x}|^2) = \\ = \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\, \overline{\lambda})(x^k - \overline{x}),\, x^k - \overline{x} \right\rangle + o(|x^k - \overline{x}|^2).$$

Разделив левую и правую части этого неравенства на $|x^k - \overline{x}|^2$ и перейдя к пределу при $k \to \infty$, получим

$$\left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\, \overline{\lambda})h,\, h \right\rangle \leqslant 0,$$

что противоречит (8). $\square$

Теоремы 1.2.4 и 1.2.5 могут рассматриваться как частные случаи соответственно теорем 6 и 7 при $l = 0$.

В случае задачи на максимум в теоремах 6 и 7 следует поменять знаки неравенств в (7) и (8) на противоположные. Кроме того, теоремы 5–7 сохраняют силу и для задачи с дополнительным прямым ограничением $x \in P$, если только $\overline{x} \in \operatorname{int} P$.

Пример 2. Пусть требуется найти экстремальные точки функции

$$f\colon \mathbf{R}^2 \to \mathbf{R}, \quad f(x) = \frac{1}{2}\,(\sigma_1 x_1^2 + \sigma_2 x_2^2),$$

на множестве

$$D = \{x \in \mathbf{R}^2 \,\big|\, x_1^3 + x_2^3 = 1\},$$

где σ_1 и σ_2 — параметры, $0 < \sigma_1 < \sigma_2$.

Аналогично примеру 1 условие регулярности выполнено в каждой допустимой точке. Система Лагранжа имеет три решения $(x^i,\ \lambda^i)$, $i = 1,\ 2,\ 3$: $x^1 = (0,\ 1)$, $\lambda^1 = -\sigma_2/3$; $x^2 = (1,\ 0)$, $\lambda^2 = -\sigma_1/3$; $x^3 = (\sigma_1/\widetilde{\sigma},\ \sigma_2/\widetilde{\sigma})$, $\lambda^3 = -\widetilde{\sigma}/3$, где $\widetilde{\sigma} = \sqrt[3]{\sigma_1^3 + \sigma_2^3}$. Исследуя полученные стационарные точки с помощью условий оптимальности второго порядка, убеждаемся, что x^1 и x^2 — точки строгого локального минимума, а x^3 — строгого локального максимума.

Существование глобального минимума в этой задаче следует из утверждения о бесконечно растущих функциях, сформулированного в задаче 1.1.5. Сравнивая значения целевой функции в точках x^1 и x^2, убеждаемся, что точкой глобального минимума является x^2.

Задача 3. Найти экстремальные точки функции

$$f\colon \mathbf{R}^2 \to \mathbf{R}, \quad f(x) = \frac{1}{3}\,x_1^3 + x_2,$$

на окружности $D = \{x \in \mathbf{R}^2 \,\big|\, x_1^2 + x_2^2 = 2\}$ (ср. с примером 1).

Задача 4. Дать геометрическую интерпретацию задачи из примера 2.

Задача 5. Пусть $A \in \mathbf{R}(n,\ n)$ — симметрическая матрица. Найти все экстремальные точки квадратичной формы

$$f\colon \mathbf{R}^n \to \mathbf{R}, \quad f(x) = \langle Ax,\ x\rangle,$$

на единичной сфере $D = \{x \in \mathbf{R}^n \,\big|\, |x| = 1\}$.

Задача 6. Среди всех треугольников заданного периметра найти тот, который имеет наибольшую площадь.

Задача 7. На заданной высоте над плоской поверхностью расположено артиллерийское орудие. Снаряд вылетает из орудия с заданной скоростью. Под каким углом к горизонту следует произвести выстрел, чтобы снаряд улетел как можно дальше?

§ 1.4. Задача со смешанными ограничениями

Этот параграф посвящен условиям оптимальности для задачи оптимизации со смешанными функциональными ограничениями

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in \mathbf{R}^n \mid F(x) = 0,\ G(x) \leqslant 0\}, \tag{2}$$

где $f\colon \mathbf{R}^n \to \mathbf{R}$ — заданная функция, $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ — заданные отображения. Более общая постановка, когда имеется еще и прямое ограничение, здесь не рассматривается по двум причинам. Во-первых, доказательство соответствующей модификации условий оптимальности весьма нетривиально, громоздко и сопряжено с привлечением глубоких результатов выпуклого анализа, чего авторы стремятся по возможности избегать. Во-вторых, для описания и обоснования обсуждаемых ниже численных методов вполне достаточно условий оптимальности для задачи в постановке (1), (2). Условия оптимальности для случая наличия как прямых, так и функциональных ограничений можно найти, например, в [10, 37].

Заметим далее, что от ограничений-неравенств всегда можно избавиться, например, с помощью следующего простого приема. Пусть отображение G имеет компоненты $g_i(\cdot)$, $i = 1, \ldots, m$; тогда, как легко видеть, локальные решения задачи (1), (2) находятся в очевидном соответствии с локальными решениями задачи

$$f(x) \to \min, \quad (x,\ \sigma) \in \widetilde{D},$$

где

$$\widetilde{D} = \{(x,\ \sigma) \in \mathbf{R}^n \times \mathbf{R}^m \mid F(x) = 0,\ g_i(x) + \sigma_i^2 = 0,\ i = 1, \ldots, m\}$$

либо

$$\widetilde{D} = \{(x,\ \sigma) \in \mathbf{R}^n \times \mathbf{R}^m_+ \mid F(x) = 0,\ g_i(x) + \sigma_i = 0,\ i = 1, \ldots, m\}$$

(в последнем случае простое ограничение $\sigma \in \mathbf{R}^m_+$ рассматривается как прямое). Этот и подобные ему приемы иногда оказываются полезными как при построении и обосновании численных методов решения задачи (1), (2) (см. § 4.7), так и при выводе условий оптимальности для такой задачи [6]. Однако в тех случаях, когда ограничения-неравенства удается учитывать непосредственно, получаемые результаты часто оказываются более точными. Можно даже сказать, что отказ от обязательного сведéния (без потери гладкости) ограничений-неравенств к равенствам как раз и позволяет говорить о нелинейной оптимизации как о самостоятельной области математики, а не просто о подразделе нелинейного анализа.

Заметим также, что возможна и обратная трансформация: всякое ограничение-равенство можно записать как два противоположных нестрогих неравенства, но этот прием используется лишь в некоторых специальных случаях, поскольку получаемая таким образом система ограничений оказывается слишком «вырожденной». В частности, допустимое множество получаемой задачи будет иметь пустую внутренность (за исключением совершенно патологических случаев),

и специфика задач с чистыми ограничениями-неравенствами все равно не будет приобретена.

1.4.1. Леммы о линейных системах. Материал этого параграфа является прямым обобщением материала § 1.3. В частности, для получения прямодвойственных условий оптимальности потребуется обобщение леммы 1.3.1 на случай наличия не только линейных уравнений, но и линейных неравенств. В свою очередь для доказательства такого обобщения потребуются некоторые факты о разрешимости систем линейных неравенств и уравнений.

Лемма 1. *Для любой матрицы* $A \in \mathbf{R}(m,n)$ *система*

$$Ax \geqslant 0, \quad A^{\mathrm{T}}y = 0, \quad y \geqslant 0, \quad Ax + y > 0$$

относительно $(x,y) \in \mathbf{R}^n \times \mathbf{R}^m$ *имеет решение.*

Доказательство. Обозначим через $a_i \in \mathbf{R}^n$ строки матрицы A, $i = 1,\dots,m$. Докажем, что для каждого $i = 1,\dots,m$ существуют элементы $x^i \in \mathbf{R}^n$ и $y^i \in \mathbf{R}^m$ такие, что

$$Ax^i \geqslant 0, \quad A^{\mathrm{T}}y^i = 0, \quad y^i \geqslant 0, \quad \langle a_i,\, x^i\rangle + y^i_i > 0. \tag{3}$$

Тогда для доказательства леммы будет достаточно взять $x = \sum\limits_{i=1}^{m} x^i$, $y = \sum\limits_{i=1}^{m} y^i$. Ясно также, что совместность системы (3) достаточно доказать для какого-то одного номера i, например $i = 1$.

Рассуждаем по индукции (по m). Пусть $m = 1$. Если $a_1 \neq 0$, то достаточно взять $x = a_1$, $y = 0$. Если же $a_1 = 0$, то можно взять $x = 0$ и $y = 1$.

Пусть доказываемое утверждение верно для всех $m \leqslant s$ и пусть $m = s+1$. Обозначим через $\widetilde{A}$ подматрицу матрицы A, состоящую из ее первых s строк. Нужно указать элементы $x \in \mathbf{R}^n$ и $y \in \mathbf{R}^{s+1}$ такие, что

$$\widetilde{A}x \geqslant 0, \quad \langle a_{s+1},\, x\rangle \geqslant 0, \quad \sum_{i=1}^{s+1} y_i a_i = 0, \quad y \geqslant 0, \quad \langle a_1,\, x\rangle + y_1 > 0. \tag{4}$$

По предположению индукции существуют $\widetilde{x} \in \mathbf{R}^n$ и $\widetilde{y} \in \mathbf{R}^s$, удовлетворяющие условию

$$\widetilde{A}\widetilde{x} \geqslant 0, \quad \sum_{i=1}^{s} \widetilde{y}_i a_i = 0, \quad \widetilde{y} \geqslant 0, \quad \langle a_1,\, \widetilde{x}\rangle + \widetilde{y}_1 > 0.$$

Если $\langle a_{s+1},\, \widetilde{x}\rangle \geqslant 0$, то условию (4) удовлетворяют $x = \widetilde{x}$ и $y = (\widetilde{y},\, 0)$.

Предположим, что $\langle a_{s+1},\, \widetilde{x}\rangle < 0$. Зададим матрицу $B \in \mathbf{R}(s,\, n)$, строками которой являются

$$b_i = a_i + \gamma_i a_{s+1}, \quad \gamma_i = -\frac{\langle a_i,\, \widetilde{x}\rangle}{\langle a_{s+1},\, \widetilde{x}\rangle} \geqslant 0, \quad i = 1,\dots,s. \tag{5}$$

Применяя к B предположение индукции, получаем существование таких элементов $u \in \mathbf{R}^n$ и $v \in \mathbf{R}^m$, что

$$Bu \geqslant 0, \quad B^{\mathrm{T}}v = 0, \quad v \geqslant 0, \quad \langle b_1,\, u\rangle + v_1 > 0. \tag{6}$$

Положим

$$x = u - \frac{\langle a_{s+1},\, u\rangle}{\langle a_{s+1},\, \widetilde{x}\rangle}\,\widetilde{x}, \quad y = \left(v,\, \sum_{i=1}^{s}\gamma_i v_i\right) \geqslant 0,$$

где неравенство следует из неравенств в (5) и (6). Заметим, что имеет место равенство $\langle a_{s+1},\, x\rangle = 0$. Кроме того, $\forall\, i = 1, \ldots, s$ имеем

$$\langle a_i,\, x\rangle = \langle b_i - \gamma_i a_{s+1},\, x\rangle = \langle b_i,\, x\rangle = \\ = \langle b_i,\, u\rangle - \frac{\langle a_{s+1},\, u\rangle}{\langle a_{s+1},\, \widetilde{x}\rangle}\,\langle b_i,\, \widetilde{x}\rangle = \langle b_i,\, u\rangle \geqslant 0,$$

где снова приняты во внимание соотношения (5) и (6). Далее,

$$\sum_{i=1}^{s+1} y_i a_i = \sum_{i=1}^{s} v_i a_i + \left(\sum_{i=1}^{s}\gamma_i v_i\right) a_{s+1} = B^{\mathrm{T}}v = 0.$$

Наконец, вновь привлекая (5) и (6), имеем

$$\langle a_1,\, x\rangle + y_1 = \langle b_1 - \gamma_1 a_{s+1},\, x\rangle + v_1 = \langle b_1,\, x\rangle + v_1 = \\ = \langle b_1,\, u\rangle + \frac{\langle a_{s+1},\, u\rangle}{\langle a_{s+1},\, \widetilde{x}\rangle}\langle b_1,\, \widetilde{x}\rangle + v_1 = \langle b_1,\, u\rangle + v_1 > 0,$$

что завершает доказательство (4). □

Следствие 1. *Для любых матриц* $A_1 \in \mathbf{R}(m_1,\, n)$ *и* $A_2 \in \mathbf{R}(m_2,\, n)$ *система*

$$A_1 x = 0, \quad A_2 x \geqslant 0,$$
$$A_1^{\mathrm{T}} y^1 + A_2^{\mathrm{T}} y^2 = 0, \quad y^2 \geqslant 0,$$
$$A_2 x + y^2 > 0$$

относительно $(x,\, y^1,\, y^2) \in \mathbf{R}^n \times \mathbf{R}^{m_1} \times \mathbf{R}^{m_2}$ *имеет решение.*

Задача 1. Доказать следствие 1.

Следующий результат называют теоремой Моцкина об альтернативе.

Лемма 2. *Для любых матриц* $A_i \in \mathbf{R}(m_i, n)$, $i = 0, 1, 2$, *имеет решение одна и только одна из следующих двух систем*:

$$A_0 x > 0, \quad A_1 x = 0, \quad A_2 x \geqslant 0$$

относительно $x \in \mathbf{R}^n$ *либо*

$$A_0^{\mathrm{T}} y^0 + A_1^{\mathrm{T}} y^1 + A_2^{\mathrm{T}} y^2 = 0, \quad y^0 \geqslant 0, \quad y^2 \geqslant 0$$

относительно $(y^0, y^1, y^2) \in (\mathbf{R}^{m_0} \setminus \{0\}) \times \mathbf{R}^{m_1} \times \mathbf{R}^{m_2}$.

Доказательство. Если предположить, что x — решение первой системы, а (y^0, y^1, y^2) — решение второй, причем $y^0 \neq 0$, то

$$0 = \langle A_0^{\mathrm{T}} y^0 + A_1^{\mathrm{T}} y^1 + A_2^{\mathrm{T}} y^2,\ x\rangle = \langle y^0,\ A_0 x\rangle + \langle y^2,\ A_2 x\rangle > 0,$$

что невозможно.

Теперь предположим, что первая система несовместна. Но согласно следствию 1 найдутся $x \in \mathbf{R}^n$ и $(y^0, y^1, y^2) \in \mathbf{R}^{m_0} \times \mathbf{R}^{m_1} \times \mathbf{R}^{m_2}$ такие, что

$$A_0 x \geqslant 0, \quad A_1 x = 0, \quad A_2 x \geqslant 0, \tag{7}$$

$$A_0^{\mathrm{T}} y^0 + A_1^{\mathrm{T}} y^1 + A_2^{\mathrm{T}} y^2 = 0, \quad y^0 \geqslant 0, \quad y^2 \geqslant 0,$$

$$A_0 x + y^0 > 0. \tag{8}$$

Остается показать, что $y^0 \neq 0$. Однако если бы это было не так, то из (7) и (8) немедленно следовало бы, что x — решение первой системы, что невозможно по предположению. □

Теперь можно доказать утверждение, которое собственно и будет использовано в этом параграфе. Речь идет о так называемой лемме о сопряженном конусе, или лемме Фаркаша.

Лемма 3. *Введем конус*

$$K = \{x \in \mathbf{R}^n \mid A_1 x = 0, A_2 x \geqslant 0\},$$

где $A_1 \in \mathbf{R}(m_1, n)$, $A_2 \in \mathbf{R}(m_2, n)$.

Тогда

$$K^* = \{x \in \mathbf{R}^n \mid x = A_1^{\mathrm{T}} y^1 + A_2^{\mathrm{T}} y^2,\ y^1 \in \mathbf{R}^{m_1},\ y^2 \in \mathbf{R}_+^{m_2}\}. \tag{9}$$

Доказательство. Обозначим правую часть (9) через Q. Тогда $\forall\, \xi \in Q$, $\forall\, x \in K$ при некоторых $y^1 \in \mathbf{R}^{m_1}$, $y^2 \in \mathbf{R}_+^{m_2}$ имеем

$$\langle \xi,\ x\rangle = \langle A_1^{\mathrm{T}} y^1 + A_2^{\mathrm{T}} y^2,\ x\rangle = \langle y^2,\ A_2 x\rangle \geqslant 0,$$

т. е. $Q \subset K^*$.

Теперь возьмем произвольный элемент $\xi \in K^*$; тогда система

$$\langle \xi,\ x\rangle < 0, \quad A_1 x = 0, \quad A_2 x \geqslant 0$$

относительно $x \in \mathbf{R}^n$ несовместна. Применяя лемму 2, можем утверждать существование $y_0 \in \mathbf{R}$, $y^1 \in \mathbf{R}^{m_1}$ и $y^2 \in \mathbf{R}^{m_2}$ таких, что

$$-y_0 \xi + A_1^{\mathrm{T}} y^1 + A_2^{\mathrm{T}} y^2 = 0, \quad y_0 > 0, \quad y^2 \geqslant 0.$$

Отсюда немедленно следует, что $\xi \in Q$, т. е. $K^* \subset Q$. □

1.4.2. Условия Каруша–Куна–Таккера. Начнем с описания касательного конуса к заданному в (2) множеству D в заданной точке. Скалярное ограничение-неравенство называется *активным* в данной

точке, если оно выполняется в этой точке как равенство. Совершенно очевидно, что при локальных рассмотрениях следует принимать во внимание лишь активные в данной точке ограничения, поскольку остальные не оказывают никакого влияния на структуру допустимого множества вблизи этой точки (здесь предполагается, что G по крайней мере непрерывно в рассматриваемой точке). *Множеством индексов активных ограничений* задачи (1), (2) в точке $\overline{x} \in \mathbf{R}^n$ называется множество

$$I(\overline{x}) = \{i = 1, \ldots, m \mid g_i(\overline{x}) = 0\}.$$

Предполагая дифференцируемость отображений F и G в точке $\overline{x} \in D$, введем конус, получаемый линеаризацией ограничения-равенства и активных ограничений-неравенств в точке $\overline{x}$:

$$H(\overline{x}) = \{h \in \ker F'(\overline{x}) \mid \langle g_i'(\overline{x}),\, h\rangle \leqslant 0 \quad \forall\, i \in I(\overline{x})\}. \tag{10}$$

Будем говорить, что в точке $\overline{x}$ выполнено *условие регулярности Мангасариана–Фромовица* [1], если

$$\operatorname{rank} F'(\overline{x}) = l$$

и существует элемент $\overline{h} \in \ker F'(\overline{x})$ такой, что

$$\langle g_i'(\overline{x}),\, \overline{h}\rangle < 0 \quad \forall\, i \in I(\overline{x}).$$

Следующая теорема обобщает теорему Люстерника.

Теорема 1. *Пусть отображение* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *дифференцируемо в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$, *причем его производная непрерывна в точке* $\overline{x}$, *а отображение* $G\colon \mathbf{R}^n \to \mathbf{R}^m$ *дифференцируемо в точке* $\overline{x}$. *Пусть* $F(\overline{x}) = 0$, $G(\overline{x}) \leqslant 0$ *и конус* $H(\overline{x})$ *определен в* (10).

Тогда для заданного в (2) *множества* D *справедливо следующее*:

а) $C_D(\overline{x}) \subset H(\overline{x})$;

б) *если в точке* $\overline{x}$ *выполнено условие регулярности Мангасариана–Фромовица, то* $T_D(\overline{x}) = C_D(\overline{x}) = H(\overline{x})$.

Доказательство. Справедливость утверждения а) проверяется непосредственно с помощью рассуждений, аналогичных использованным при доказательстве теоремы 1.2.1.

Докажем б). Пусть $h \in H(\overline{x})$, а $\overline{h}$ — элемент из условия Мангасариана–Фромовица. Зафиксируем $\theta \in (0, 1]$ и рассмотрим элемент $h(\theta) = \theta\overline{h} + (1 - \theta)h$. Очевидно, $h(\theta) \in \ker F'(\overline{x})$, поэтому, применяя к отображению F в точке $\overline{x}$ теорему Люстерника, получаем существование отображения $r\colon \mathbf{R} \to \mathbf{R}^n$ такого, что

$$F(\overline{x} + th(\theta) + r(t)) = 0 \quad \forall\, t \in \mathbf{R}, \quad |r(t)| = o(t). \tag{11}$$

[1] Общепринятая аббревиатура — MFCQ (от английского Mangasarian–Fromovitz constraint qualification).

Кроме того,

$$\langle g_i'(\overline{x}),\, h(\theta)\rangle = \theta\langle g_i'(\overline{x}),\, \overline{h}\rangle + (1-\theta)\langle g_i'(\overline{x}),\, h\rangle < 0 \quad \forall\, i \in I(\overline{x}),$$

поэтому для любого достаточно малого $t > 0$

$$g_i(\overline{x} + th(\theta) + r(t)) = t\langle g_i'(\overline{x}),\, h(\theta)\rangle + o(t) < 0 \quad \forall\, i \in I(\overline{x}). \tag{12}$$

Из (11) и (12) немедленно следует, что $\overline{x} + th(\theta) + r(t) \in D$ для любого достаточно малого $t > 0$, поэтому в силу второго соотношения в (11) $h(\theta) \in T_D(\overline{x})$. Наконец, используя замкнутость конуса $T_D(\overline{x})$ (см. задачу 1.2.2), получаем, что $h = \lim\limits_{\theta \to 0+} h(\theta) \in T_D(\overline{x})$. $\square$

Пусть теперь $\overline{x} \in D$ — локальное решение задачи (1), (2), причем функция f дифференцируема в точке $\overline{x}$, отображения F и G удовлетворяют условиям гладкости из теоремы 1 и в точке $\overline{x}$ выполнено условие Мангасариана–Фромовица. Тогда из теоремы 1.2.1 (см. также (1.2.3)) и леммы 3 получаем следующее прямодвойственное необходимое условие первого порядка оптимальности: существуют $\overline{\lambda} \in \mathbf{R}^l$ и $\overline{\mu}_i \geqslant 0$, $i \in I(\overline{x})$, такие, что

$$f'(\overline{x}) = -(F'(\overline{x}))^{\mathrm{T}}\overline{\lambda} - \sum_{i \in I(\overline{x})} \overline{\mu}_i g_i'(\overline{x}).$$

Чтобы привести полученное необходимое условие к окончательной форме, называемой теоремой Каруша–Куна–Таккера, введем *функцию Лагранжа* задачи (1), (2):

$$L\colon \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R},$$

$$L(x,\, \lambda,\, \mu) = f(x) + \langle \lambda,\, F(x)\rangle + \langle \mu,\, G(x)\rangle.$$

Очевидно,

$$\frac{\partial L}{\partial x}\,(x,\, \lambda,\, \mu) = f'(x) + (F'(x))^{\mathrm{T}}\lambda + (G'(x))^{\mathrm{T}}\mu,$$

$$x \in \mathbf{R}^n, \quad \lambda \in \mathbf{R}^l, \quad \mu \in \mathbf{R}^m.$$

Теорема 2. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображение $G\colon \mathbf{R}^n \to \mathbf{R}^m$ дифференцируемы в точке $\overline{x} \in \mathbf{R}^n$, а отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ дифференцируемо в некоторой окрестности этой точки, причем его производная непрерывна в точке $\overline{x}$.*

Тогда если $\overline{x}$ является локальным решением задачи (1), (2) и в точке $\overline{x}$ выполнено условие регулярности Мангасариана–Фромовица, то найдутся элементы $\overline{\lambda} \in \mathbf{R}^l$ и $\overline{\mu} \in \mathbf{R}^m_+$ такие, что

$$\frac{\partial L}{\partial x}\,(\overline{x},\, \overline{\lambda},\, \overline{\mu}) = 0, \tag{13}$$

$$\langle \overline{\mu},\, G(\overline{x})\rangle = 0. \tag{14}$$

Условие (14) называется *условием дополняющей нежесткости* (название устоявшееся, хотя его трудно признать удачным). Смысл этого условия состоит в следующем. При $G(\overline{x}) \leqslant 0$ и $\overline{\mu} \geqslant 0$ равенство (14) имеет место тогда и только тогда, когда

$$\overline{\mu}_i g_i(\overline{x}) = 0 \quad \forall\, i = 1, \ldots, m,$$

т. е.

$$\overline{\mu}_i = 0 \quad \forall\, i \in \{1, \ldots, m\} \setminus I(\overline{x}).$$

Таким образом, условие дополняющей нежесткости как раз «выключает» из рассмотрения те ограничения-неравенства, которые не являются активными в данной точке.

Определение 1. Точка $\overline{x} \in \mathbf{R}^n$ называется *стационарной точкой* задачи (1), (2), если $\overline{x} \in D$ и выполнены соотношения (13), (14) при некоторых $\overline{\lambda} \in \mathbf{R}^l$ и $\overline{\mu} \in \mathbf{R}^m_+$. Элементы $\overline{\lambda}$ и $\overline{\mu}$ при этом называются *множителями Лагранжа*, отвечающими стационарной точке $\overline{x}$.

Задача 2. Доказать, что выполнение условия регулярности Мангасариана–Фромовица в произвольной стационарной точке задачи (1), (2) равносильно ограниченности множества множителей Лагранжа, отвечающих этой стационарной точке.

Итак, при выполнении определенных требований гладкости локальное решение задачи (1), (2), в котором выполнено условие регулярности Мангасариана–Фромовица, является стационарной точкой этой задачи. Обратное, вообще говоря, неверно; условия второго порядка оптимальности для задачи (1), (2) излагаются в следующем пункте. Заметим, что условие Мангасариана–Фромовица не гарантирует единственности пары множителей Лагранжа, отвечающей стационарной точке $\overline{x}$. Единственность пары множителей гарантируется более сильными условиями регулярности ограничений, например, так называемым *условием линейной независимости* [1], состоящим в том, что строки матрицы $F'(\overline{x})$ и векторы $g_i'(\overline{x})$, $i \in I(\overline{x})$, образуют линейно независимую систему в $\mathbf{R}^n$.

Кроме того, само условие регулярности Мангасариана–Фромовица *вместе* с условием единственности пары множителей Лагранжа, отвечающей стационарной точке $\overline{x}$, принято называть *строгим условием регулярности Мангасариана–Фромовица* [2]. Разумеется, это условие слабее условия линейной независимости.

[1] Общепринятая аббревиатура — LICQ (от английского Linear independence constraint qualification).

[2] Общепринятая аббревиатура — SMFCQ (от английского Strict Mangasarian–Fromovitz constraint qualification).

Задача 3. Показать, что строгое условие регулярности Мангасариана–Фромовица в стационарной точке $\overline{x}$ задачи (1), (2) равносильно следующему условию: для отвечающих $\overline{x}$ множителей Лагранжа $\overline{\lambda} \in \mathbf{R}^l$ и $\overline{\mu} \in \mathbf{R}^m_+$ строки матрицы $F'(\overline{x})$ и векторы $g_i'(\overline{x})$, $i \in I_+(\overline{x}, \overline{\mu})$, образуют линейно-независимую систему в $\mathbf{R}^n$, и существует элемент $\overline{h} \in \ker F'(\overline{x})$ такой, что

$$\langle g_i'(\overline{x}), \overline{h}\rangle = 0 \quad \forall\, i \in I_+(\overline{x}, \overline{\mu}), \quad \langle g_i'(\overline{x}), \overline{h}\rangle < 0 \quad \forall\, i \in I(\overline{x}) \setminus I_+(\overline{x}, \overline{\mu}),$$

где

$$I_+(\overline{x}, \overline{\mu}) = \{i \in I(\overline{x}) \,\big|\, \overline{\mu}_i > 0\}.$$

Систему уравнений и неравенств

$$\frac{\partial L}{\partial x}\,(x, \lambda, \mu) = 0, \quad F(x) = 0, \quad \mu \geqslant 0, \quad G(x) \leqslant 0, \quad \langle \mu, G(x)\rangle = 0$$

относительно $(x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, характеризующую стационарные точки задачи (1), (2) и отвечающие им множители Лагранжа, называют *системой Каруша–Куна–Таккера* [1]).

Геометрический смысл необходимого условия Каруша–Куна–Таккера состоит в том, что при выполнении соответствующих требований гладкости и условия Мангасариана–Фромовица в локальном решении $\overline{x}$ задачи (1), (2) градиент $f'(\overline{x})$ целевой функции представи́м в виде суммы линейной комбинации строк матрицы Якоби $F'(\overline{x})$ и неположительной комбинации градиентов $g_i'(\overline{x})$, $i \in I(\overline{x})$. Принцип Лагранжа (теорема 1.3.5) является частным случаем теоремы 2 при $m = 0$.

Необходимое условие Каруша–Куна–Таккера может непосредственно использоваться для отыскания решений задачи (1), (2). Схема действий при этом такова: нужно составить и решить систему Каруша–Куна–Таккера, а потом исследовать полученные стационарные точки на оптимальность (с помощью условий второго порядка оптимальности, излагаемых в следующем пункте, либо другими средствами). Отдельно нужно исследовать точки, в которых нарушено условие регулярности Мангасариана–Фромовица. Численным методам решения системы Каруша–Куна–Таккера посвящен § 4.5 (см. также п. 4.3.1, где говорится о методах решения системы Лагранжа для случая отсутствия ограничений-неравенств). Реализовать указанную схему аналитически удается лишь в простейших случаях (см. примеры 1.3.1, 1.3.2). Подчеркнем, что перед составлением системы Каруша–Куна–Таккера рассматриваемую задачу следует привести именно к виду (1), (2). В частности, задача должна быть задачей на минимум, а знаки ограничений-неравенств должны быть направлены в нужную сторону.

[1]) Общепринятое сокращение — система ККТ.

В заключение этого пункта еще раз обратимся к фундаментальному понятию условия регулярности ограничений. Из леммы 2 вытекает, что невыполнение в точке $\overline{x}$ условия регулярности Мангасариана–Фромовица равносильно существованию множителей $\overline{\lambda} \in \mathbf{R}^l$ и $\overline{\mu} \in \mathbf{R}^m_+$, не равных нулю одновременно и таких, что

$$(F'(\overline{x}))^{\mathrm{T}}\overline{\lambda} + (G'(\overline{x}))^{\mathrm{T}}\overline{\mu} = 0 \tag{15}$$

и выполнено (14). Но тогда в теореме 2 можно отказаться от условия регулярности Мангасариана–Фромовица, если соответствующим образом модифицировать равенство (13). Общепринятой формулировкой получаемого таким образом необходимого условия оптимальности является теорема Ф. Джона. Введем *обобщенную функцию Лагранжа* задачи (1), (2):

$$L_0\colon \mathbf{R}^n \times \mathbf{R} \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R},$$

$$L_0(x,\, \lambda_0,\, \lambda,\, \mu) = \lambda_0 f(x) + \langle \lambda,\, F(x)\rangle + \langle \mu,\, G(x)\rangle.$$

Теорема 3. *Пусть выполнены условия теоремы* 2.

Тогда если $\overline{x}$ *является локальным решением задачи* (1), (2), *то найдутся число* $\overline{\lambda}_0 \geqslant 0$ *и элементы* $\overline{\lambda} \in \mathbf{R}^l$ *и* $\overline{\mu} \in \mathbf{R}^m_+$, *не равные нулю одновременно и такие, что*

$$\frac{\partial L_0}{\partial x}(\overline{x},\, \overline{\lambda}_0,\, \overline{\lambda},\, \overline{\mu}) = 0 \tag{16}$$

и выполнено (14).

Действительно, в случае выполнения условия регулярности Мангасариана–Фромовица равенство (16) имеет место при $\overline{\lambda}_0 = 1$ и $\overline{\lambda}$, $\overline{\mu}$ из теоремы 2, а в противном случае — при $\overline{\lambda}_0 = 0$ и не равных одновременно нулю $\overline{\lambda}$, $\overline{\mu}$ из (15). Подчеркнем, что в последнем случае получаемое необходимое условие оптимальности не слишком содержательно: оно выполняется при любой целевой функции f вне зависимости от того, является $\overline{x}$ локальным решением задачи (1), (2) или нет. Можно сказать, что при этом теорема Ф. Джона выражает только сам факт нарушения условия регулярности Мангасариана–Фромовица. Тем не менее в теории оптимизации теорема Ф. Джона находит важные приложения. Например, на ней основана теория необходимых условий второго порядка оптимальности, содержательных и при невыполнении условия регулярности Мангасариана–Фромовица [2]. С другой стороны, случаи прямого использования теоремы Ф. Джона для построения или обоснования численных методов оптимизации авторам не известны.

Заметим, что обобщенная функция Лагранжа L_0 линейна по совокупности множителей $(\lambda_0,\, \lambda,\, \mu) \in \mathbf{R} \times \mathbf{R}^l \times \mathbf{R}^m$, поэтому, если в утверждении теоремы 3 реализуется $\overline{\lambda}_0 \neq 0$, то соответствующей нормировкой тройки $(\overline{\lambda}_0,\, \overline{\lambda},\, \overline{\mu})$ всегда можно добиться выполнения

равенства $\overline{\lambda}_0 = 1$, что приводит к сформулированному в определении 1 понятию стационарности в смысле теоремы Каруша–Куна–Таккера. Вообще, под *условием регулярности ограничений* задачи (1), (2) часто понимают любое условие, которое гарантирует справедливость утверждения теоремы 3 при $\overline{\lambda}_0 \neq 0$ [37]. Условие Мангасариана–Фромовица и условие линейной независимости являются условиями регулярности в указанном смысле. Другой важный пример условия регулярности доставляет *условие линейности*, состоящее в том, что отображения F и G аффинны. Локальное решение задачи (1), (2), в котором выполнено то или иное условие регулярности ограничений, иногда называют *квалифицированным* [1]).

Задача 4. Доказать, что условие линейности является условием регулярности ограничений.

Задача 5. Доказать, что условие Мангасариана–Фромовица (а тем более условие линейной независимости) не только является условием регулярности ограничений, но и сообщает задаче (1), (2) следующее свойство: выполнение этого условия в точке $\overline{x} \in D$ делает невозможным одновременное выполнение (14), (16) при $\overline{\lambda}_0 = 0$ и каких-либо значениях $\overline{\lambda} \in \mathbf{R}^l$ и $\overline{\mu} \in \mathbf{R}^m_+$, не равных одновременно нулю (ср. с задачей 2).

1.4.3. Условия второго порядка оптимальности. В этом пункте приводятся необходимое и достаточное условия второго порядка оптимальности в задаче (1), (2). Для их формулировки потребуется так называемый *критический конус* (или *конус критических направлений*) задачи (1), (2) в точке $\overline{x} \in D$, являющейся точкой гладкости функции f и отображений F и G:

$$K(\overline{x}) = \{h \in H(\overline{x}) \mid \langle f'(\overline{x}),\, h\rangle \leqslant 0\}. \tag{17}$$

Лемма 4. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображения* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *и* $G\colon \mathbf{R}^n \to \mathbf{R}^m$ *дифференцируемы в точке* $\overline{x} \in \mathbf{R}^n$ *и последняя является стационарной точкой задачи* (1), (2).

Тогда для всякой пары множителей Лагранжа $\overline{\lambda} \in \mathbf{R}^l$ *и* $\overline{\mu} \in \mathbf{R}^m_+$, *отвечающих стационарной точке* $\overline{x}$, *имеют место равенства*

$$K(\overline{x}) = \{h \in H(\overline{x}) \mid \langle f'(\overline{x}), h\rangle = 0\} =$$

$$= \{h \in H(\overline{x}) \mid \overline{\mu}_i \langle g_i'(\overline{x}), h\rangle = 0 \ \forall\, i \in I(\overline{x})\}. \tag{18}$$

[1]) Общепринятый английский термин для условия регулярности ограничений — Constraint qualification (CQ).

Доказательство. Для произвольного $h \in H(\overline{x})$ в силу (13), (14) и условия $\overline{\mu} \geqslant 0$ справедливо

$$\langle f'(\overline{x}),\, h\rangle = -\langle \overline{\lambda},\, F'(\overline{x})h\rangle - \langle \overline{\mu},\, G'(\overline{x})h\rangle = -\sum_{i\in I(\overline{x})} \overline{\mu}_i \langle g_i'(\overline{x}),\, h\rangle,$$

причем в последней сумме все слагаемые неположительны. Отсюда немедленно следует требуемое. □

Если $\overline{\mu}_i > 0 \;\; \forall\, i \in I(\overline{x})$, то говорят, что в стационарной точке $\overline{x}$ выполнено *условие строгой дополнительности* (для множителей Лагранжа $\overline{\lambda}$ и $\overline{\mu}$). Заметим, что при выполнении этого условия из (18) вытекает равенство

$$K(\overline{x}) = \{h \in \ker F'(\overline{x}) \,\big|\, \langle g_i'(\overline{x}),\, h\rangle = 0 \;\; \forall\, i \in I(\overline{x})\}.$$

Теорема 4. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображения* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *и* $G\colon \mathbf{R}^n \to \mathbf{R}^m$ *дважды дифференцируемы в точке* $\overline{x} \in \mathbf{R}^n$.

Тогда если $\overline{x}$ *является локальным решением задачи* (1), (2) *и в точке* $\overline{x}$ *выполнено условие линейной независимости, то для единственной пары элементов* $\overline{\lambda} \in \mathbf{R}^l$ *и* $\overline{\mu} \in \mathbf{R}^m_+$, *удовлетворяющих* (13), (14), *имеет место следующее неравенство*:

$$\left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\, \overline{\lambda},\, \overline{\mu})h,\, h \right\rangle \geqslant 0 \quad \forall\, h \in K(\overline{x}), \tag{19}$$

где конус $K(\overline{x})$ *определен в* (17).

Доказательство. Фиксируем элемент $h \in K(\overline{x})$. Введем множество индексов

$$I(\overline{x},\, h) = \{i \in I(\overline{x}) \,\big|\, \langle g_i'(\overline{x}),\, h\rangle = 0\}.$$

Согласно теореме Люстерника существует отображение $r\colon \mathbf{R} \to \mathbf{R}^n$ такое, что $\forall\, t \in \mathbf{R}$ имеем

$$F(\overline{x} + th + r(t)) = 0, \quad g_i(\overline{x} + th + r(t)) = 0 \quad \forall\, i \in I(\overline{x},\, h),$$
$$|r(t)| = o(t).$$

Отсюда и из определения множеств $H(\overline{x})$ и $I(\overline{x},\, h)$ элементарно выводится, что $\overline{x} + th + r(t) \in D$ для любого достаточно малого $t > 0$; поэтому если $\overline{x}$ — локальное решение задачи (1), (2), то, привлекая лемму 4 и соотношения (13), (14), получаем: для любого достаточно малого $t > 0$

$$0 \leqslant f(\overline{x} + th + r(t)) - f(\overline{x}) = f(\overline{x} + th + r(t)) + \langle \overline{\lambda},\, F(\overline{x} + th + r(t))\rangle +$$
$$+ \langle \overline{\mu},\, G(\overline{x} + th + r(t))\rangle - f(\overline{x}) - \langle \overline{\lambda},\, F(\overline{x})\rangle - \langle \overline{\mu},\, G(\overline{x})\rangle =$$
$$= L(\overline{x} + th + r(t),\, \overline{\lambda},\, \overline{\mu}) - L(\overline{x},\, \overline{\lambda},\, \overline{\mu}) = \left\langle \frac{\partial L}{\partial x}(\overline{x},\, \overline{\lambda},\, \overline{\mu}),\, th + r(t) \right\rangle +$$

$$+\frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x},\,\overline{\lambda},\,\overline{\mu})(th+r(t)),\,th+r(t)\right\rangle + o(t^2) =$$

$$=\frac{t^2}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x},\,\overline{\lambda},\,\overline{\mu})h,\,h\right\rangle + o(t^2).$$

Разделив левую и правую части полученного неравенства на t^2 и перейдя к пределу при $t\to 0+$, получим (19). $\square$

Заметим, что согласно теоремам 1.2.2 и 1 равенство $K(\overline{x})=\{0\}$ является достаточным для того, чтобы точка $\overline{x}\in D$ была строгим локальным решением задачи (1), (2). Более тонкое достаточное условие, допускающее нетривиальность критического конуса, содержится в следующей теореме [1]).

Теорема 5. *Пусть выполнены условия теоремы* 4.

Тогда если $\overline{x}\in D$ *и существуют удовлетворяющие* (13), (14) *элементы* $\overline{\lambda}\in\mathbf{R}^l$ *и* $\overline{\mu}\in\mathbf{R}^m_+$, *для которых имеет место*

$$\left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\,\overline{\lambda},\,\overline{\mu})h,\,h\right\rangle > 0 \quad \forall\, h\in K(\overline{x})\setminus\{0\}, \tag{20}$$

где конус $K(\overline{x})$ *определен в* (17), *то* $\overline{x}$ *является строгим локальным решением задачи* (1), (2).

Доказательство. Схема доказательства совершенно аналогична схеме доказательства теоремы 1.3.7. От противного: предположим, что существует последовательность $\{x^k\}\subset D\setminus\{\overline{x}\}$ такая, что $f(x^k)\leqslant f(\overline{x})$ $\forall\, k$ и $\{x^k\}\to\overline{x}$ $(k\to\infty)$. Последовательность $\{(x^k-\overline{x})/|x^k-\overline{x}|\}$ будем считать сходящейся к некоторому $h\in\in\mathbf{R}^n\setminus\{0\}$. Очевидно, $h\in C_D(\overline{x})\setminus\{0\}$, значит, в силу утверждения а) теоремы 1 $h\in H(\overline{x})\setminus\{0\}$. Кроме того, h принадлежит контингентному конусу к множеству $\{x\in\mathbf{R}^n\,|\,f(x)\leqslant f(\overline{x})\}$ в точке $\overline{x}$, поэтому из утверждения а) теоремы 1 вытекает, что $\langle f'(\overline{x}),\,h\rangle\leqslant 0$. Таким образом, $h\in K(\overline{x})\setminus\{0\}$ (см. (17)).

Далее, $\forall\, k$ имеем

$$0\geqslant f(x^k)-f(\overline{x}) =$$

$$=\langle f'(\overline{x}),\,x^k-\overline{x}\rangle+\frac{1}{2}\,\langle f''(\overline{x})(x^k-\overline{x}),\,x^k-\overline{x}\rangle+o(|x^k-\overline{x}|^2),$$

$$0=F(x^k)-F(\overline{x}) =$$

$$=F'(\overline{x})(x^k-\overline{x})+\frac{1}{2}\,F''(\overline{x})[x^k-\overline{x},\,x^k-\overline{x}]+o(|x^k-\overline{x}|^2),$$

[1]) Общепринятая аббревиатура для достаточного условия, приводимого в этой теореме — SOSC (от английского Second-order sufficient condition).

$$0 \geqslant g_i(x^k) - g_i(\overline{x}) =$$

$$= \langle g_i'(\overline{x}),\, x^k - \overline{x}\rangle + \frac{1}{2}\,\langle g_i''(\overline{x})(x^k - \overline{x}),\, x^k - \overline{x}\rangle + o(|x^k - \overline{x}|^2) \quad \forall\, i \in I(\overline{x}).$$

Домножаем скалярно на $\overline{\lambda}$ левую и правую части второго соотношения, домножаем на $\overline{\mu}_i$ левую и правую части соответствующего соотношения последней группы $(i \in I(\overline{x}))$ и складываем соответствующие части всех полученных соотношений. Принимая во внимание (13), (14), в результате имеем: $\forall\, k$

$$0 \geqslant \langle f'(\overline{x}),\, x^k - \overline{x}\rangle + \langle \overline{\lambda},\, F'(\overline{x})(x^k - \overline{x})\rangle + \langle \overline{\mu},\, G'(\overline{x})(x^k - \overline{x})\rangle +$$

$$+\frac{1}{2}\,\langle f''(\overline{x})(x^k - \overline{x}),\, x^k - \overline{x}\rangle + \frac{1}{2}\,\langle \overline{\lambda},\, F''(\overline{x})[x^k - \overline{x},\, x^k - \overline{x}]\rangle +$$

$$+\frac{1}{2}\,\langle \overline{\mu},\, G''(\overline{x})[x^k - \overline{x},\, x^k - \overline{x}]\rangle + o(|x^k - \overline{x}|^2) =$$

$$= \left\langle \frac{\partial L}{\partial x}\,(\overline{x},\, \overline{\lambda},\, \overline{\mu}),\, x^k - \overline{x} \right\rangle + \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\, \overline{\lambda},\, \overline{\mu})(x^k - \overline{x}),\, x^k - \overline{x} \right\rangle +$$

$$+o(|x^k - \overline{x}|^2) = \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\, \overline{\lambda},\, \overline{\mu})(x^k - \overline{x}),\, x^k - \overline{x} \right\rangle + o(|x^k - \overline{x}|^2).$$

Разделив левую и правую части этого неравенства на $|x^k - \overline{x}|^2$ и перейдя к пределу при $k \to \infty$, получим

$$\left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\, \overline{\lambda},\, \overline{\mu})h,\, h \right\rangle \leqslant 0,$$

что противоречит (20). $\square$

Теоремы 1.3.6 и 1.3.7 являются частными случаями соответственно теорем 4 и 5 при $m = 0$. Кроме того, теоремы 2–5 останутся верны и для задачи с дополнительным прямым ограничением $x \in P$, если только $\overline{x} \in \operatorname{int} P$.

В заключение подчеркнем важнейшую роль, которую играли в этом и предыдущем параграфах условия регулярности ограничений (условие (1.3.5) в контексте задачи с ограничениями-равенствами и условия Мангасариана–Фромовица и линейной независимости в случае смешанных ограничений). Проблема получения содержательных условий оптимальности без предположений о регулярности ограничений либо в ослабленных предположениях такого рода привлекала большое внимание исследователей в последние двадцать лет. С некоторыми результатами в этом направлении можно познакомиться в работах [2, 21, 22].

Глава 2

НАЧАЛЬНЫЕ СВЕДЕНИЯ О МЕТОДАХ ОПТИМИЗАЦИИ

Назначение этой главы — дать общее представление о том, что такое численные методы (алгоритмы) оптимизации, какими они бывают и какие вопросы возникают в связи с их практическим использованием и теоретическим анализом. Здесь же рассматриваются численные методы решения одномерных задач оптимизации. Значение этих методов, прежде всего, в том, что они являются необходимыми составляющими для построения многих алгоритмов, предназначенных для решения более сложных задач.

§ 2.1. Общее понятие о методах оптимизации

Пусть $P \subset \mathbf{R}^n$ — заданное множество, $f\colon P \to \mathbf{R}$ — заданная функция, $F\colon P \to \mathbf{R}^l$ и $G\colon P \to \mathbf{R}^m$ — заданные отображения. Рассматриваемые в последующих главах численные методы ориентированы главным образом на решение задачи

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in P \mid F(x) = 0,\ G(x) \leqslant 0\}, \tag{2}$$

либо основных частных случаев этой задачи (см. п. 1.1.1). Наличие каких-то особых свойств у множества P (кроме его замкнутости и выпуклости), функции f и отображений F и G (кроме их гладкости) не предполагается. В этом смысле речь идет о методах общего назначения (кроме гл. 6, где эксплуатируются свойства выпуклости, и гл. 7, в которой рассматриваются специальные методы линейного и квадратичного программирования).

2.1.1. Классификация методов оптимизации. Понятия сходимости. Любой численный метод решения задачи (1), (2) основан на вычислении значений целевой и ограничивающих функций, а также (часто) их производных. Организация этих вычислений не должна рассматриваться как составная часть оптимизационного алгоритма:

обычно их следует интерпретировать как внешнюю процедуру (симулятор), которая с точки зрения оптимизационного алгоритма представляет собой «черный ящик». Метод называют *пассивным*, если точки, в которых производятся вычисления, выбираются независимо друг от друга и в принципе определены заранее, и *последовательным*, если такие точки выбираются последовательно, в процессе счета, на основе информации, получаемой в ходе вычислительного процесса. По понятным причинам последовательных методов подавляющее большинство.

Последовательный метод генерирует последовательность точек x^0, $x^1, \ldots, x^k, \ldots$ в $\mathbf{R}^n$, которые называются *приближениями* к решению задачи. Эта последовательность $\{x^k\}$, которую называют *траекторией* метода, может составлять лишь часть множества точек, в которых производятся вычисления. Переход от x^k к следующему приближению x^{k+1} называется *шагом* или *итерацией* метода. Способ этого перехода составляет суть метода, поэтому часто методы записывают в виде *итерационной схемы*

$$x^{k+1} = \Psi_k(x^k), \quad k = 0, 1, \ldots, \tag{3}$$

где Ψ_k — отображение со значениями в $\mathbf{R}^n$, определенное на некотором множестве $U_k \subset \mathbf{R}^n$, причем предполагается, что к шагу $k+1$ это отображение известно. Если итерационная схема фиксирована, то задание начального приближения $x^0 \in U_0$ корректно определяет траекторию метода тогда и только тогда, когда

$$\Psi_{k-1}(\Psi_{k-2}(\ldots \Psi_0(x^0)\ldots)) \in U_k \quad \forall\, k = 1, 2, \ldots.$$

Рассмотренная итерационная схема (3) называется *одношаговой*. Иногда приходится рассматривать *многошаговые* схемы, когда Ψ_k зависит не только от x^k, но и от некоторых предыдущих приближений.

Порядком метода называется максимальный порядок производных целевой и ограничивающих функций задачи (1), (2), используемых при осуществлении итерации метода. Так, если используется информация лишь о значениях функции, то говорят о методах нулевого порядка. Если же привлекаются первые (вторые и т. д.) производные, то говорят о методах первого (второго и т. д.) порядка.

Если любое начальное приближение $x^0 \in U_0$ корректно определяет траекторию метода, некоторая точка которой совпадает с искомым решением задачи, то метод называется *конечношаговым* или *конечным*. Конечные методы для задач линейного и квадратичного программирования обсуждаются в гл. 7. Для более общих задач оптимизации обычно приходится иметь дело с *бесконечношаговыми* методами, траектории которых, вообще говоря, не попадают в точное решение, а лишь аппроксимируют его в некотором смысле. В зависимости от характера такой аппроксимации говорят о разных типах

сходимости метода. Например, если

$$\{x^k\} \to \overline{x} \quad (k \to \infty),$$

где $\overline{x}$ — искомое решение задачи, то говорят о *сходимости по аргументу*, или просто о *сходимости* метода из начального приближения x^0 (разумеется, это подразумевает, что x^0 корректно определяет траекторию $\{x^k\}$). Иногда удается установить не сходимость метода к конкретному решению, а лишь более слабое свойство:

$$\operatorname{dist}(x^k, S) \to 0 \quad (k \to \infty),$$

где S — множество решений задачи. В этом случае говорят о *сходимости к множеству решений*. Если S — компакт, то такая сходимость равносильна тому, что траектория метода имеет предельные точки, любая из которых содержится в S; если же S не ограничено, то такая траектория может вообще не иметь предельных точек.

Если $x^k \in D$ для достаточно больших k и

$$f(x^k) \to \overline{v} \quad (k \to \infty),$$

где $\overline{v}$ — значение задачи (1), то говорят о *сходимости по функции*, а саму последовательность $\{x^k\}$ при этом называют *минимизирующей*. Разумеется, сама минимизирующая последовательность может не сходиться.

Часто удается показать сходимость метода не к множеству (локальных или глобальных) решений, а лишь к множеству стационарных точек задачи, либо стремление к нулю невязки соответствующих необходимых условий оптимальности. Например, в случае задачи безусловной оптимизации при выполнении

$$\{f'(x^k)\} \to 0 \quad (k \to \infty)$$

говорят о *сходимости по градиенту*. Часто удается установить, что любая предельная точка траектории метода является стационарной.

Наконец, если $U_0 = \mathbf{R}^n$ и имеет место сходимость метода из любого начального приближения $x^0 \in \mathbf{R}^n$, то говорят о *глобальной сходимости*. Если же тот или иной вид сходимости удается гарантировать только из начальных приближений, достаточно близких к искомому решению (множеству решений, стационарной точке, множеству стационарных точек), то говорят о *локальной сходимости* метода. В принципе любой локально сходящийся метод должен рассматриваться в совокупности с некоторым способом получения подходящего для него начального приближения.

Несколько слов об иерархии методов оптимизации. Во многих случаях метод решения сложной задачи состоит в ее редукции к более простой или к последовательности более простых. Разумеется, это имеет смысл только в том случае, когда для таких простых задач

известны эффективные методы решения. Например, классический метод Ньютона сводит решение системы нелинейных уравнений к последовательному решению линейных систем, для которых развиты эффективные методы вычислительной линейной алгебры. Аналогично, методы решения общих задач условной оптимизации часто используют как вспомогательные процедуры методы одномерной оптимизации, безусловной оптимизации, линейного и квадратичного программирования. Методы безусловной оптимизации сами используют методы одномерной оптимизации, а методы квадратичного программирования используют методы линейного программирования. Поэтому естественно, что развитие методов, стоящих выше в этой иерархии, невозможно без должного развития методов, стоящих ниже.

2.1.2. Оценки скорости сходимости. Правила остановки. При анализе бесконечношаговых методов чрезвычайно важен вопрос не только о сходимости, но и о *скорости* такой *сходимости*, являющейся одной из основных характеристик эффективности метода. Будем здесь говорить о скорости сходимости по аргументу; терминология для других видов сходимости совершенно аналогична.

Пусть метод локально сходится к решению $\overline{x}$. Оценки скорости сходимости выражают гарантированную скорость убывания величины $|x^k - \overline{x}|$ при $k \to \infty$ для сходящихся к $\overline{x}$ траекторий $\{x^k\}$. Здесь всюду предполагается, что начальное приближение $x^0 \in \mathbf{R}^n$ достаточно близко́ к $\overline{x}$, рассматриваемые траектории не попадают в $\overline{x}$ ни на каком конечном шаге, а все возникающие ниже константы не зависят от конкретного выбора x^0. Если

$$\limsup_{k\to\infty} \frac{|x^{k+1} - \overline{x}|}{|x^k - \overline{x}|} = q,$$

где $q \in (0, 1)$ — некоторая константа, то говорят, что метод сходится с *линейной скоростью*, а если $q = 0$, то со *сверхлинейной*. Важным частным случаем сверхлинейной является *квадратичная скорость* сходимости, когда

$$\limsup_{k\to\infty} \frac{|x^{k+1} - \overline{x}|}{|x^k - \overline{x}|^2} \leqslant C$$

(здесь и далее $C > 0$ — некоторая константа).

Задача 1. Пусть последовательность $\{x^k\} \subset \mathbf{R}^2$ сходится к точке $\overline{x} \in \mathbf{R}^2$ с квадратичной скоростью. Следует ли отсюда квадратичная или хотя бы сверхлинейная скорость сходимости последовательностей $\{x_1^k\}$ и $\{x_2^k\}$ к $\overline{x}_1$ и $\overline{x}_2$ соответственно?

Скорость сходимости более низкая, чем линейная, называется *сублинейной*. Примером оценки, гарантирующей только сублинейную

скорость сходимости, служит неравенство

$$\limsup_{k\to\infty} k|x^k - \overline{x}| \leqslant C.$$

В случае его выполнения говорят об *арифметической скорости* сходимости метода. Если же

$$\limsup_{k\to\infty} \frac{|x^k - \overline{x}|}{q^k} \leqslant C$$

при некотором $q \in (0, 1)$, то говорят о *геометрической скорости* сходимости. Очевидно, линейная скорость сходимости подразумевает геометрическую, но, вообще говоря, не наоборот.

Систематическое изучение оценок скорости сходимости последовательностей проводится, например, в [6, 31].

Оценки скорости сходимости дают важный критерий качественного сравнения различных методов друг с другом. Вместе с тем при всей своей важности скорость сходимости вовсе не единственная характеристика эффективности метода. Необходимо, например, принимать во внимание трудоемкость одной итерации, причем обычно методы с более высокой скоростью сходимости имеют более трудоемкую итерацию. Быстро сходящийся метод с дорогой итерацией может проигрывать более медленному методу, каждая итерация которого является более дешевой. Другая важная характеристика — устойчивость метода по отношению к влиянию возмущений входных данных и помех, неизбежных в реальных вычислительных процессах. Нередко на практике остро встает проблема допустимости генерируемых методом траекторий, т. е. справедливость включения $\{x^k\} \subset D$ (либо $\{x^k\} \subset \{x \in P \mid G(x) \leqslant 0\}$, либо $\{x^k\} \subset P$). Например, функция f и/или отображения F и G могут быть вообще не определены вне D, и метод, траектории которого покидают допустимое множество, оказывается просто некорректен.

При выборе метода следует принимать во внимание следующее обстоятельство. Если требуется решить одну конкретную оптимизационную задачу, то предпочтение следует отдать надежному методу, возможно, с не самой высокой скоростью сходимости. Если же речь идет о решении серии однотипных задач, например, отличающихся лишь значениями некоторых параметров, то скорость сходимости выходит на первый план. В последнем случае целесообразно потратить время и силы на выбор и настройку подходящего быстрого алгоритма для имеющейся серии задач.

Кроме того, на практике вычислительный процесс не может продолжаться бесконечно, поэтому любой бесконечношаговый метод должен сопровождаться правилом остановки, и наиболее надежные правила такого рода получаются как раз с помощью оценок скорости

сходимости. Действительно, наличие такой оценки позволяет указать количество шагов метода, гарантирующее достижение заданной точности (по аргументу либо в каком-то другом смысле). Однако, во-первых, оценки скорости сходимости, которые обычно удается получить (и, в частности, те, о которых говорилось выше), носят существенно локальный характер, а количественная характеризация требуемого качества начального приближения редко бывает конструктивной. Во-вторых, в оценках всегда фигурируют константы (такие, как q и C), определение которых является самостоятельной нетривиальной задачей. В этом смысле область практической применимости оценок скорости сходимости для выработки правил остановки методов весьма ограничена.

На практике обычно используют менее надежные правила остановки по косвенным признакам, например, по поведению полученной к текущему шагу части траектории $\{x^k\}$ или последовательности $\{f(x^k)\}$, а также по величине невязки необходимых условий оптимальности. Например, фиксируют число $\varepsilon > 0$ и останавливают метод после $(k+1)$-го шага, если

$$|x^{k+1} - x^k| \leqslant \varepsilon$$

(т. е. если шаг метода стал достаточно малым) либо

$$|f(x^{k+1}) - f(x^k)| \leqslant \varepsilon,$$

а в случае задачи безусловной оптимизации используют также неравенство

$$|f'(x^{k+1})| \leqslant \varepsilon.$$

Ясно, что такие правила остановки весьма ненадежны и не гарантируют близость получаемого приближения к искомому решению ни в каком смысле.

Например, из предельного соотношения $\lim\limits_{k\to\infty} |x^{k+1} - x^k| = 0$ сходимость последовательности $\{x^k\}$ не следует, что видно на примере последовательности $\{x^k\} \subset \mathbf{R}$, $x^k = \sqrt{k}$ или $x^k = \sin\sqrt{k}$, $k = 0, 1, \ldots$. Тем не менее эти правила используются очень часто, поскольку других конструктивных правил остановки просто нет.

Кроме того, количество шагов метода может (и должно) определяться имеющимися вычислительными ресурсами. Часто, если процесс не остановился раньше по одному из указанных выше правил, то его приходится останавливать по причине исчерпания ресурсов. Обычно задают максимальное время счета, максимальное количество итераций, максимальное количество вычислений значений целевой и ограничивающих функций и т. п. Конечно, получаемое при этом «приближение» может не иметь к искомому решению никакого отношения.

Видимо, наиболее правильным является комбинирование указанных правил остановки, что предполагает задание некоторой их иерархии.

Сделаем одно замечание, касающееся локальной и глобальной сходимости методов. Обычно глобально сходящиеся методы имеют более низкую скорость сходимости, чем локально сходящиеся. Это связано с тем, что локально сходящиеся методы обычно лучше учитывают локальную структуру задачи вблизи искомого решения. Поэтому представляется естественным применять локально сходящийся метод в сочетании с некоторым глобально сходящимся. Роль последнего при этом состоит в том, чтобы обеспечить подходящее начальное приближение для первого, а тот уже должен быстро аппроксимировать искомое решение. При этом крайне желательно, чтобы глобально сходящийся метод был оптимизационным, т. е. ориентированным на поиск решений (а не, скажем, стационарных точек) рассматриваемой задачи, и чтобы переключение с одного метода на другой осуществлялось автоматически, поскольку далеко не всегда можно заранее указать необходимое количество шагов глобально сходящегося метода до переключения. Тем самым формируется законченный алгоритм решения задачи. Типичный результат, описывающий поведение комбинированного алгоритма, выглядит так: метод глобально сходится в некотором смысле (чаще всего к стационарной точке или ко множеству стационарных точек или сходится в том смысле, что любая предельная точка его траектории является стационарной), причем справедлива некоторая локальная оценка скорости сходимости (желательно сверхлинейная). Некоторые стратегии глобализации сходимости рассматриваются в гл. 5.

Что же касается отыскания глобальных решений (без предположений о выпуклости задачи [1]), делающих любое локальное решение глобальным, а любую стационарную точку решением), то соответствующие численные методы развиты значительно хуже, чем методы поиска локальных решений, а точнее, стационарных точек. Такие

[1] Функция $f\colon X \to \mathbf{R}$ называется *сильно выпуклой* с константой $\theta \geqslant 0$ на выпуклом множестве $X \subset \mathbf{R}^n$, если $f(tx^1 + (1-t)x^2) \leqslant tf(x^1) + (1-t)f(x^2) - \theta t(1-t)|x^1 - x^2|^2 \quad \forall\, x^1, x^2 \in X, \quad \forall\, t \in [0,1]$. Сильно выпуклая с константой $\theta = 0$ функция называется просто *выпуклой*. Под *выпуклостью задачи* (1) понимается выпуклость ее допустимого множества D и целевой функции f на D. Любое локальное решение выпуклой задачи оптимизации является глобальным, а любая стационарная (в смысле определения 1.2.2) точка является решением; множество решений такой задачи всегда выпукло. Задача минимизации непрерывной сильно выпуклой с положительной константой функции на выпуклом множестве всегда имеет единственное решение. Функция f называется *вогнутой*, если функция $-f$ выпукла. Свойства задачи максимизации вогнутой функции на выпуклом множестве аналогичны свойствам выпуклой задачи минимизации.

методы глобальной оптимизации (не путать с глобально сходящимися методами!) должны основываться на глобальном взгляде на всю задачу целиком, недоступном локальным методам, работающим лишь в окрестности текущего приближения, а глобальный взгляд неминуемо сопряжен с необходимостью просмотра большого количества точек, распределенных по всему допустимому множеству.

Наиболее очевидный метод глобальной оптимизации состоит в переборе узлов достаточно мелкой сетки на D с вычислением и сравнением значений целевой функции f в этих точках. Предполагается, что D не слишком сложно устроено, и на нем можно эффективным образом задать сетку. Однако ясно, что такой подход реалистичен лишь для задач малой размерности; ниже он будет подробно рассмотрен применительно к одномерным задачам. Более экономичные методы глобальной оптимизации часто включают в себя локальные процедуры, что позволяет сократить пассивный перебор. Простейший пример доставляет *метод мультистарта*, когда из каждого узла крупной сетки на D запускается метод локальной оптимизации, после чего сравниваются значения f в найденных локальных решениях (стационарных точках). Таким образом, локальные методы играют важную роль и при построении глобальных. Ниже основное внимание уделено именно методам локальной оптимизации. О методах глобальной оптимизации см. [18, 36].

§ 2.2. Методы одномерной оптимизации

В этом параграфе рассматриваются три наиболее простых метода решения задачи

$$f(x) \to \min, \quad x \in [a,\, b], \tag{1}$$

где $a,\, b \in \mathbf{R}$, $a < b$, $f\colon [a,\, b] \to \mathbf{R}$. Все три метода нулевого порядка. Первый метод пассивный, не очень эффективный, но зато он применим к заданным на $[a,\, b]$ функциям весьма широкого класса. Два других метода последовательные, относительно эффективные, но лишь для функций из значительно более специального класса. Разумеется, известно множество других методов одномерной оптимизации [4, 10, 13, 27, 37].

Цель этого параграфа — дать представление об основных идеях, лежащих в основе методов, существенно использующих одномерную специфику задачи. Этим и определяется отбор материала. Отметим, что весьма надежные и эффективные процедуры решения задачи (1) реализованы во всех современных общематематических пакетах (таких, например, как Matlab и Maple), не говоря уже о специализированных оптимизационных пакетах [49]. Это является одной из причин, по которым более подробно методы одномерной оптимизации здесь не рассматриваются.

2.2.1. Метод перебора на равномерной сетке. Предположим, что f — непрерывная на $[a, b]$ функция и требуется найти значение $\overline{v} = \min\limits_{x\in[a,\,b]} f(x)$ задачи (1). Предположим, что для приближенного решения этой задачи разрешено вычислить значения f в N точках отрезка $[a, b]$. Эти точки можно выбрать, например, как узлы равномерной сетки, т. е. как середины отрезков, разбивающих $[a, b]$ на N равных частей.

Алгоритм 1. Полагаем $\delta_N = (b-a)/N$ и вычисляем

$$x_i^N = (a + i\delta_N) - \frac{\delta_N}{2}, \quad i = 1, \ldots, N, \qquad v_N = \min_{i=1,\ldots,N} f(x_i^N).$$

Величина v_N берется в качестве приближения к $\overline{v}$. Тем самым описан *метод перебора на равномерной сетке.*

Величину $v_N - \overline{v} \geqslant 0$ естественно назвать *погрешностью* метода (по функции). Если не накладывать на f дальнейших предположений, то при любом фиксированном N погрешность может быть сколь угодно велика. Оценить погрешность можно лишь в более узких классах функций, например, для функций, непрерывных по Липшицу [1]) на $[a, b]$ с заданной константой $L > 0$. Пусть $\overline{x} \in [a, b]$ — глобальное решение задачи (1). Заметим, что минимальное расстояние от $\overline{x}$ до узлов сетки не превосходит $\delta_N/2$. Тогда

$$v_N - \overline{v} = \min_{i=1,\ldots,N} f(x_i^N) - f(\overline{x}) \leqslant \frac{1}{2} L\delta_N = \frac{L(b-a)}{2N}.$$

Уже отмечавшаяся выше трудность практического использования подобных оценок (например, для определения N, обеспечивающего вычисление $\overline{v}$ с заданной точностью) состоит в том, что величина L известна или легко определима лишь в исключительных случаях.

Нетрудно убедиться, что предложенный метод оптимален в классе пассивных методов отыскания $\overline{v}$ для функций, непрерывных по Липшицу на $[a, b]$ с заданной константой, в следующем смысле: при любом другом выборе сетки на $[a, b]$ найдется функция указанного класса, для которой погрешность соответствующего метода будет больше, чем в случае равномерной сетки (об оптимальности методов оптимизации см. [10, 37]).

Заметим, что рассмотренный метод не позволяет находить с заданной точностью само глобальное решение $\overline{x}$, даже если известна константа L. Действительно, если в качестве приближения к $\overline{x}$ брать ту точку сетки, в которой реализуется значение v_N функции

[1]) Напомним, что функция $f\colon X \to \mathbf{R}$ называется *непрерывной по Липшицу* на множестве $X \subset \mathbf{R}^n$ с константой $L > 0$, если $|f(x^1) - f(x^2)| \leqslant L|x^1 - x^2| \quad \forall\, x^1, x^2 \in X$. Аналогичным образом это понятие вводится и для отображений $F\colon X \to \mathbf{R}^l$.

f, то для всякого N можно указать непрерывную на $[a, b]$ по Липшицу с константой L функцию f, для которой уклонение такого приближения от $\overline{x}$ будет равно $(b-a) - \delta_N/2$.

Задача 1. Пусть при некотором $c \in [a, b]$ функция $f\colon [a, b] \to \mathbf{R}$ непрерывна по Липшицу на каждом из отрезков $[a, c]$ и $[c, b]$ с константами $L_1 > 0$ и $L_2 > 0$ соответственно. Показать, что f непрерывна по Липшицу на всем $[a, b]$ с константой $\max\{L_1, L_2\}$.

Задача 2. Пусть функция $f\colon [a, b] \to \mathbf{R}$ непрерывна на $[a, b]$ и дифференцируема на $[a, b]$ всюду, за исключением, может быть, конечного числа точек $c_i \in [a, b]$, $i = 1, \ldots, s$, причем

$$L = \sup_{x \in [a, b] \setminus \{c_1, \ldots, c_s\}} |f'(x)| < \infty.$$

Показать, что f непрерывна по Липшицу на $[a, b]$ с константой L.

Задача 3. Пусть функции $f_1, f_2\colon [a, b] \to \mathbf{R}$ непрерывны по Липшицу на $[a, b]$ с константами $L_1 > 0$ и $L_2 > 0$ соответственно. Показать, что функции $f_1(\cdot) + f_2(\cdot)$ и $\max\{f_1(\cdot), f_2(\cdot)\}$ непрерывны по Липшицу на $[a, b]$ с константами $L_1 + L_2$ и $\max\{L_1, L_2\}$ соответственно.

Задача 4. Привести пример непрерывной на $[a, b]$ функции, не являющейся непрерывной по Липшицу на этом отрезке.

2.2.2. Метод дихотомии. Метод золотого сечения.

Определение 1. Функция $f\colon [a, b] \to \mathbf{R}$ называется *унимодальной* (на $[a, b]$), если она достигает на $[a, b]$ глобального минимума в единственной точке $\overline{x}$, причем слева от $\overline{x}$ строго убывает, а справа строго возрастает, т. е.

$$f(y) > f(z) \quad \forall\, y, z \in [a, \overline{x}],\ y < z,$$

$$f(y) < f(z) \quad \forall\, y, z \in [\overline{x}, b],\ y < z.$$

Непрерывная на $[a, b]$ функция унимодальна тогда и только тогда, когда она имеет на $[a, b]$ единственный локальный минимум (по необходимости являющийся глобальным). Любая непрерывная строго выпуклая [1]) на $[a, b]$ функция унимодальна, но существуют, разумеется, и невыпуклые унимодальные функции.

Задача 5. Пусть $X \subset \mathbf{R}^n$ — выпуклое множество. Функция $f\colon X \to \mathbf{R}$ называется *квазивыпуклой* (на X), если любое ее множество Лебега выпукло. Показать, что это равносильно следующему:

[1]) Функция $f\colon X \to \mathbf{R}$ называется *строго выпуклой* на выпуклом множестве $X \subset \mathbf{R}^n$, если $f(tx^1 + (1-t)x^2) < tf(x^1) + (1-t)f(x^2)$ $\forall\, x^1, x^2 \in X$, $x^1 \neq x^2$, $\forall\, t \in (0, 1)$.

$\forall x^1, x^2 \in X, \ \forall t \in [0, 1]$

$$f(tx^1 + (1-t)x^2) \leqslant \max\{f(x^1), f(x^2)\}. \qquad (2)$$

Привести пример квазивыпуклой функции, не являющейся выпуклой.

Задача 6. Пусть $X \subset \mathbf{R}^n$ — выпуклое множество. Функция $f\colon X \to \mathbf{R}$ называется *строго квазивыпуклой* (на X), если $\forall x^1, x^2 \in X, \ x^1 \neq x^2, \ \forall t \in (0,1)$ условие (2) выполнено как строгое неравенство. Доказать, что в случае, когда $X = [a, b]$, унимодальность непрерывной на X функции равносильна ее строгой квазивыпуклости.

Задача 7. Пусть функции $f_1, f_2\colon [a, b] \to \mathbf{R}$ непрерывны и унимодальны на $[a, b]$. Следует ли отсюда унимодальность функций $f_1(\cdot) + f_2(\cdot)$, $\max\{f_1(\cdot), f_2(\cdot)\}$ на $[a, b]$?

Следующее свойство унимодальных функций играет центральную роль при обосновании двух рассматриваемых ниже методов.

Лемма 1. *Пусть функция* $f\colon [a, b] \to \mathbf{R}$ *унимодальна,* $\overline{x}$ *— ее глобальный минимум на* $[a, b]$.

Тогда для произвольных точек $y, z \in [a, b]$ *таких, что* $y < z$, *справедливо следующее*:

а) *если* $f(y) \leqslant f(z)$, *то* $\overline{x} \in [a, z]$;

б) *если* $f(y) \geqslant f(z)$, *то* $\overline{x} \in [y, b]$.

Доказательство. Докажем утверждение а) (утверждение б) доказывается аналогично). Предположим, что $f(y) \leqslant f(z)$, но $\overline{x} > z$. Тогда $y < z < \overline{x}$, поэтому $f(y) > f(z)$ в силу унимодальности f, а это противоречит сделанному предположению. $\square$

Таким образом, сравнив значения унимодальной функции f в двух точках отрезка $[a, b]$, можно локализовать глобальное решение задачи (1) на отрезке меньшей длины. Повторяя эту процедуру, можно последовательно уменьшать длину отрезка локализации. Соответствующие методы различаются способами выбора точек для сравнения. Заметим, что при этом аппроксимируется не только значение задачи (1), знание которого часто бывает недостаточным, но и само ее решение.

Предполагая унимодальность f, опишем *метод дихотомии* (метод деления пополам).

Алгоритм 2. Полагаем $a_1 = a$, $b_1 = b$, $c_1 = (a + b)/2$ и вычисляем $f(c_1)$. Полагаем $k = 1$.

1. Полагаем $y_k = (a_k + c_k)/2$ и вычисляем $f(y_k)$. Если $f(y_k) \leqslant f(c_k)$, то полагаем $a_{k+1} = a_k$, $b_{k+1} = c_k$, $c_{k+1} = y_k$ и переходим к п. 3. Если же $f(y_k) > f(c_k)$, то полагаем $z_k = (c_k + b_k)/2$ и вычисляем $f(z_k)$.

2. Если $f(c_k) \leqslant f(z_k)$, то полагаем $a_{k+1} = y_k$, $b_{k+1} = z_k$, $c_{k+1}{=}c_k$. Если же $f(c_k) > f(z_k)$, то полагаем $a_{k+1} = c_k$, $b_{k+1} = b_k$, $c_{k+1} = z_k$.

3. Увеличиваем номер шага k на 1 и переходим к п. 1.

Лемма 1 гарантирует, что в любом случае после шага k алгоритма $\overline{x} \in [a_{k+1},\, b_{k+1}]$, причем $c_{k+1} = (a_{k+1} + b_{k+1})/2$ и

$$b_{k+1} - a_{k+1} = \frac{1}{2}\,(b_k - a_k) = \ldots = \frac{1}{2^k}\,(b - a).$$

В качестве очередного приближения к $\overline{x}$ можно брать любую точку $x_{k+1} \in [a_{k+1},\, b_{k+1}]$. Для реализации каждого шага метода дихотомии требуется самое большее два вычисления значений целевой функции f (нулевой шаг требует одного такого вычисления). Предположим, что всего таких вычислений разрешено сделать N; удобно считать N нечетным числом. Такое количество вычислений позволяет сделать по крайней мере $k = (N-1)/2$ шагов метода, что приводит к оценке погрешности (на этот раз по аргументу)

$$|x_{k+1} - \overline{x}| \leqslant b_{k+1} - a_{k+1} \leqslant \frac{1}{2^{(N-1)/2}}\,(b - a) \approx 0.707^{N-1}\,(b - a).$$

Таким образом, с ростом N погрешность уменьшается не медленнее, чем со скоростью геометрической прогрессии со знаменателем $q \approx 0.707$. Ниже рассматривается более эффективный метод, для которого аналогичная оценка имеет место при $q \approx 0.618$ (при больши́х N это дает весьма ощутимый выигрыш в точности).

Золотым сечением отрезка называется такое его деление на две части, при котором отношение длины всего отрезка к длине большей части равно отношению длины большей части к длине меньшей части. Эта пропорция с древности использовалась в разных областях человеческой деятельности. Таким образом, если точка y осуществляет золотое сечение отрезка $[a,\, b]$ и расположена ближе к a, чем к b, то

$$\frac{b-a}{b-y} = \frac{b-y}{y-a}\,.$$

Отсюда находим, что

$$y = y(a,\, b) = a + \frac{3\sqrt{5}}{2}\,(b - a) \approx a + 0.382(b - a).$$

Вторая точка, которая осуществляет золотое сечение $[a,\, b]$, дается формулой

$$z = z(a,\, b) = a + \frac{\sqrt{5}-1}{2}\,(b - a) \approx a + 0.618\,(b - a).$$

Будем называть y и z соответственно *меньшей* и *большей золотыми точками.*

Справедливость следующего утверждения устанавливается прямыми вычислениями.

Лемма 2. *Пусть y и z — меньшая и бóльшая золотые точки отрезка $[a, b]$ соответственно.*

Тогда:

а) $z - a = b - y = (\sqrt{5} - 1)(b - a)/2$;

б) *y — бóльшая золотая точка отрезка $[a, z]$, а z — меньшая золотая точка отрезка $[y, b]$.*

Опишем *метод золотого сечения*, предполагая унимодальность f.

Алгоритм 3. Полагаем $a_1 = a$, $b_1 = b$, $y_1 = y(a, b)$, $z_1 = z(a, b)$ и вычисляем $f(y_1)$. Полагаем $k = 1$.

1. Вычисляем то из значений $f(y_k)$ или $f(z_k)$, которое еще не вычислено. Если $f(y_k) \leqslant f(z_k)$, то полагаем $a_{k+1} = a_k$, $b_{k+1} = z_k$, $y_{k+1} = y(a_{k+1}, b_{k+1})$, $z_{k+1} = y_k$. Если же $f(y_k) > f(z_k)$, то полагаем $a_{k+1} = y_k$, $b_{k+1} = b_k$, $y_{k+1} = z_k$, $z_{k+1} = z(a_{k+1}, b_{k+1})$.

2. Увеличиваем номер шага k на 1 и переходим к п. 1.

Заметим, что в любом случае после шага k алгоритма будут выполнены равенства $y_{k+1} = y(a_{k+1}, b_{k+1})$, $z_{k+1} = z(a_{k+1}, b_{k+1})$; это следует из утверждения б) леммы 2. Кроме того, в силу леммы 1 $\overline{x} \in [a_{k+1}, b_{k+1}]$, причем согласно утверждению а) леммы 2

$$b_{k+1} - a_{k+1} = \frac{\sqrt{5} - 1}{2}(b_k - a_k) = \ldots = \left(\frac{\sqrt{5} - 1}{2}\right)^k (b - a).$$

В качестве очередного приближения к $\overline{x}$ можно взять любую точку $x_{k+1} \in [a_{k+1}, b_{k+1}]$.

Подчеркнем, что реализация каждого шага метода золотого сечения требует одного вычисления значения целевой функции f (что обеспечивается свойством золотого сечения, описанным утверждением б) леммы 2). Тогда N таких вычислений позволяют сделать $k = N - 1$ шагов метода, что приводит к оценке погрешности

$$|x_{k+1} - \overline{x}| \leqslant b_{k+1} - a_{k+1} = \left(\frac{\sqrt{5} - 1}{2}\right)^{N-1} (b - a) \approx 0.618^{N-1}(b - a).$$

С ростом N погрешность уменьшается не медленнее, чем со скоростью геометрической прогрессии со знаменателем $q \approx 0.618$.

На практике часто не задают заранее число шагов или вычислений значений целевой функции, а продолжают процесс до тех пор, пока длина очередного отрезка локализации не станет меньше заданной величины.

Нередко при реализации методов решения более общих задач приходится иметь дело с задачей одномерной минимизации не на отрезке, а, скажем, на полупрямой $[a, +\infty)$. Определение унимодальности и лемма 1 прямо распространяются и на этот случай. В предположении об унимодальности целевой функции такая задача сводится к задаче оптимизации на отрезке. Действительно, фиксируем произвольное $\delta > 0$ и выберем минимальное $k = 0, 1, \ldots,$ для которого $f(a + k\delta) \leqslant f(a + (k+1)\delta)$. Тогда по лемме 1 глобальный минимум f на $[a, +\infty)$ лежит на отрезке $[a, a+(k+1)\delta]$. Если же $f(a+ + k\delta) > f(a + (k+1)\delta) \ \ \forall\, k = 0, 1, \ldots,$ то приходим к противоречию с унимодальностью функции f.

В случае непрерывной, но не обязательно унимодальной функции f можно доказать сходимость методов дихотомии и золотого сечения к некоторому локальному решению задачи (1). В реальных приложениях унимодальность целевой функции на заданном отрезке часто не имеет места либо не известно, имеет она место или нет. Тем не менее рассмотренные методы часто приводят к успеху и в этих случаях.

Задача 8. Сделать два шага метода дихотомии для задачи

$$f(x) = 2x^2 - 7x + 1 \to \min, \quad x \in [1, 5].$$

Глава 3
МЕТОДЫ БЕЗУСЛОВНОЙ ОПТИМИЗАЦИИ

В этой главе основное внимание уделено методам, важным в идейном отношении. Некоторые практически значимые методы (например, квазиньютоновские методы, методы сопряженных направлений, методы нулевого порядка) излагаются весьма схематично. Это объясняется тем, что, во-первых, основной целью в данном курсе является построение методов для задач условной оптимизации, а во-вторых, область методов гладкой безусловной оптимизации можно считать вполне сформировавшейся, и существующая литература по ним в основном достаточна.

§ 3.1. Методы спуска

Пусть $f\colon \mathbf{R}^n \to \mathbf{R}$ — заданная функция. Один из наиболее очевидных подходов к отысканию решения задачи безусловной оптимизации

$$f(x) \to \min, \quad x \in \mathbf{R}^n, \tag{1}$$

состоит в следующем. Для текущего приближения к решению определяется направление, по которому функция f убывает, после чего делается шаг некоторой длины по этому направлению. Для полученного таким образом нового приближения процедура повторяется. Методы такого типа и называются методами спуска; о них и пойдет речь в этом параграфе.

3.1.1. Общая схема методов спуска. Начнем со строгого определения направления убывания.

Определение 1. Вектор $d \in \mathbf{R}^n$ называется *направлением убывания* функции $f\colon \mathbf{R}^n \to \mathbf{R}$ в точке $x \in \mathbf{R}^n$, если для любого достаточно малого $t > 0$ имеет место неравенство $f(x + td) < f(x)$.

Множество всех направлений убывания функции f в точке $x \in \mathbf{R}^n$ является конусом и обозначается $\mathcal{D}_f(x)$. Таким образом, $d \in \mathcal{D}_f(x)$ в том и только том случае, когда любой достаточно малый сдвиг из точки x в направлении d приводит к уменьшению значения функции. Элементарно доказывается

Лемма 1. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *дифференцируема в точке* $x \in \mathbf{R}^n$.

Тогда:

а) *для любого* $d \in \mathcal{D}_f(x)$ *выполнено* $\langle f'(x), d\rangle \leqslant 0$;

б) *если* $d \in \mathbf{R}^n$ *удовлетворяет условию* $\langle f'(x), d\rangle < 0$, *то* $d \in \mathcal{D}_f(x)$.

Задача 1. Доказать следующее уточнение леммы 1. Если функция $f\colon \mathbf{R}^n \to \mathbf{R}$ дважды дифференцируема в точке $x \in \mathbf{R}^n$, то:

а) для любого $d \in \mathcal{D}_f(x)$ такого, что $\langle f'(x),\, d\rangle = 0$, выполнено $\langle f''(x)d,\, d\rangle \leqslant 0$;

б) если $d \in \mathbf{R}^n$ удовлетворяет условиям $\langle f'(x),\, d\rangle = 0$, $\langle f''(x)d,\, d\rangle < 0$, то $d \in \mathcal{D}_f(x)$.

К *методам спуска* относятся методы, которые можно задать итерационной схемой

$$x^{k+1} = x^k + \alpha_k d^k, \quad d^k \in \mathcal{D}_f(x^k), \quad k = 0, 1, \ldots, \tag{2}$$

где *параметры длины шага* $\alpha_k > 0$ выбираются так, чтобы выполнялось по крайней мере неравенство

$$f(x^{k+1}) < f(x^k), \tag{3}$$

т. е. чтобы последовательность $\{f(x^k)\}$ монотонно убывала. Здесь предполагается, что $\mathcal{D}_f(x^k) \neq \varnothing$; если $\mathcal{D}_f(x^k) = \varnothing$, то процесс останавливают. Заметим, что поскольку $d^k \in \mathcal{D}_f(x^k)$, неравенство (3) выполняется для любого достаточно малого числа $\alpha_k > 0$. Таким образом, происходит «спуск» генерируемых приближений на поверхности все более низкого уровня целевой функции задачи, что позволяет надеяться на сходимость последовательности таких приближений к решению. Конкретный метод спуска характеризуется способом выбора направлений убывания, а также используемой процедурой выбора параметра длины шага. Такие процедуры основаны на исследовании сужения функции f на луч, исходящий из точки x^k в направлении вектора d^k, поэтому их называют процедурами *одномерного поиска* [1]. Интересно отметить следующую особенность оптимизационных алгоритмов: при выборе направления поиска d^k обычно используется некоторая приближенная модель функции f, в то время как одномерный поиск осуществляется для самой функции f.

Из леммы 1 следует, что если $f'(x^k) \neq 0$, то в качестве направления убывания можно взять $d^k = -f'(x^k)$. Соответствующие такому выбору методы спуска особенно просты в реализации, а также наиболее естественны и важны с теоретической точки зрения; именно для

[1] Общепринятый английский термин — *Linesearch*.

них в п. 3.1.2 будет проведен детальный анализ сходимости и скорости сходимости. Заметим, однако, что по причинам, обсуждаемым в конце п. 3.1.2, на практике методы с указанным способом выбора направления убывания обычно оказываются крайне неэффективными.

Значительно большее практическое значение имеют методы спуска, в которых $d^k = -Q_k f'(x^k)$, где для каждого k симметрическая матрица $Q_k \in \mathbf{R}(n, n)$ положительно определена, но выбирается специальными «умными» способами, а не просто кладется равной единичной матрице (что соответствует $d^k = -f'(x^k)$); см. п. 3.2.3. Подчеркнем, что если $f'(x^k) \neq 0$, то в силу леммы 1 направление $d^k = -Q_k f'(x^k)$ также является направлением убывания функции f в точке x^k, поскольку

$$\langle f'(x^k),\, d^k\rangle = -\langle Q_k f'(x^k),\, f'(x^k)\rangle < 0.$$

Антиградиентное же направление $d^k = -f'(x^k)$ может использоваться как «последнее средство» в тех случаях, когда более изощренные способы выбора направления убывания на данной итерации рассматриваемого алгоритма по тем или иным причинам не срабатывают.

Рассмотрим основные процедуры одномерного поиска, предполагая, что для текущего приближения x^k уже получено направление $d^k \in \mathcal{D}_f(x^k)$.

Правило одномерной минимизации. Параметр $\alpha_k > 0$ выбирается из условия

$$f(x^k + \alpha_k d^k) = \min_{\alpha \geqslant 0} f(x^k + \alpha d^k),$$

т. е. как решение одномерной задачи оптимизации

$$\varphi_k(\alpha) \to \min, \quad \alpha \in \mathbf{R}_+, \tag{4}$$

где

$$\varphi_k\colon \mathbf{R}_+ \to \mathbf{R}, \quad \varphi_k(\alpha) = f(x^k + \alpha d^k).$$

Заметим, что при этом если функция f дифференцируема в точке x^{k+1}, то

$$0 = \varphi_k'(\alpha_k) = \langle f'(x^k + \alpha_k d^k),\, d^k\rangle = \langle f'(x^{k+1}),\, d^k\rangle, \tag{5}$$

т. е. если $f'(x^{k+1}) \neq 0$, то геометрически x^{k+1} является точкой касания луча, исходящего из x^k в направлении d^k, и поверхности уровня функции f, проходящей через точку x^{k+1}.

Такой способ выбора α_k является наилучшим в том смысле, что позволяет получить следующее приближение с наименьшим значением целевой функции вдоль используемого направления убывания. Если f — квадратичная функция, т. е.

$$f(x) = \langle Ax,\, x\rangle + \langle b,\, x\rangle, \quad x \in \mathbf{R}^n,$$

где $A \in \mathbf{R}(n, n)$ — симметрическая матрица, $b \in \mathbf{R}^n$, причем матрица A положительно определена, то решение α_k задачи (4) вычисля-

ется по явной формуле:

$$\alpha_k = -\frac{\langle 2Ax^k + b,\, d^k\rangle}{2\langle Ad^k,\, d^k\rangle}.$$

Однако в общем случае этот способ определения α_k весьма трудоемок. Для решения на каждом шаге одномерной задачи оптимизации (4) могут использоваться методы, рассмотренные в § 2.2. Кроме того, можно (и нужно!) пользоваться менее трудоемкими способами выбора параметра длины шага, излагаемыми ниже.

Заметим, наконец, что иногда (а на практике фактически всегда, за исключением специальных случаев, таких, как случай квадратичной целевой функции) задачу (4) заменяют задачей

$$\varphi_k(\alpha) \to \min, \quad \alpha \in [0,\, \widehat{\alpha}],$$

где $\widehat{\alpha} > 0$ — фиксированный параметр. В дальнейшем, говоря о правиле одномерной минимизации, будем иметь в виду именно эту вспомогательную задачу, считая, что выбор $\widehat{\alpha} = +\infty$ соответствует вспомогательной задаче (4).

Правило Армихо. Этот способ предполагает дифференцируемость функции f в текущей точке x^k. Фиксируем числа $\widehat{\alpha} > 0$, ε, $\theta \in (0, 1)$. Полагаем $\alpha = \widehat{\alpha}$.

1. Проверяем выполнение неравенства

$$f(x^k + \alpha d^k) \leqslant f(x^k) + \varepsilon\alpha\langle f'(x^k),\, d^k\rangle. \tag{6}$$

2. Если (6) не выполнено, то заменяем α на $\theta\alpha$ и переходим к п. 1. В противном случае полагаем $\alpha_k = \alpha$.

Таким образом, α_k вычисляется как первое из чисел α, получаемых в результате дробления начального значения $\widehat{\alpha}$, для которого выполнится неравенство (6). Величина $\alpha\langle f'(x^k),\, d^k\rangle$ в правой части (6) имеет смысл «предсказанного» линейной моделью целевой функции убывания ее значения при шаге длины α по направлению d^k. Таким образом, неравенство (6) означает, что реальное убывание значения целевой функции должно составлять как минимум заданную (определяемую выбором параметра $\varepsilon \in (0, 1)$) долю от «предсказанного» убывания. Следующая лемма показывает, что если d^k удовлетворяет достаточному условию направления убывания, сформулированному в лемме 1, т. е.

$$\langle f'(x^k),\, d^k\rangle < 0, \tag{7}$$

то количество дроблений в правиле Армихо будет конечным.

Лемма 2. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *дифференцируема в точке* $x^k \in \mathbf{R}^n$.

Тогда если элемент $d^k \in \mathbf{R}^n$ *удовлетворяет* (7), *то неравенство* (6) *имеет место для любого достаточно малого* $\alpha > 0$.

Доказательство. Для любого достаточно малого $\alpha > 0$ имеет место

$$f(x^k + \alpha d^k) - f(x^k) = \langle f'(x^k),\, \alpha d^k\rangle + o(\alpha) =$$
$$= \varepsilon\alpha\langle f'(x^k),\, d^k\rangle + (1-\varepsilon)\alpha\langle f'(x^k),\, d^k\rangle + o(\alpha) =$$
$$= \varepsilon\alpha\langle f'(x^k),\, d^k\rangle + \alpha\left((1-\varepsilon)\langle f'(x^k),\, d^k\rangle + \frac{o(\alpha)}{\alpha}\right) \leqslant \varepsilon\alpha\langle f'(x^k),\, d^k\rangle,$$

поскольку $(1-\varepsilon)\langle f'(x^k),\, d^k\rangle + o(\alpha)/\alpha < 0$. $\square$

Очевидно, при выполнении (7) выбор α_k по правилу Армихо гарантирует выполнение условия монотонного убывания (3). Более того, неравенство (6) при $\alpha = \alpha_k$ дает оценку того, насколько $f(x^{k+1})$ меньше, чем $f(x^k)$, и эта оценка обычно оказывается достаточной для обоснования сходимости метода, в отличие от условия (3). Доказательство сходимости существенно упрощается, когда удается установить, что количество дроблений для определения α_k конечно равномерно по k.

Лемма 3. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ дифференцируема на $\mathbf{R}^n$ и ее производная непрерывна по Липшицу на $\mathbf{R}^n$ с константой $L > 0$.*

Тогда если для некоторых x^k, $d^k \in \mathbf{R}^n$ выполнено (7), *то неравенство* (6) *имеет место для любого $\alpha \in (0, \overline{\alpha}_k]$, где*

$$\overline{\alpha}_k = -\frac{2(1-\varepsilon)\langle f'(x^k), d^k\rangle}{L|d^k|^2} > 0. \tag{8}$$

Доказательство этой леммы основано на следующем свойстве функций с непрерывными по Липшицу производными.

Лемма 4. *В условиях леммы* 3

$$|f(x+\xi) - f(x) - \langle f'(x), \xi\rangle| \leqslant \frac{L}{2}\,|\xi|^2 \quad \forall\, x, \xi \in \mathbf{R}^n.$$

Доказательство. По формуле Ньютона–Лейбница имеем

$$|f(x+\xi) - f(x) - \langle f'(x),\, \xi\rangle| = \left|\int_0^1 \langle f'(x+t\xi) - f'(x),\, \xi\rangle\, dt\right| \leqslant$$
$$\leqslant \int_0^1 |f'(x+t\xi) - f'(x)||\xi|\, dt \leqslant \int_0^1 Lt|\xi|^2\, dt = \frac{L}{2}\,|\xi|^2,$$

где также использовано неравенство Коши–Буняковского–Шварца. $\square$

Доказательство леммы 3. Из леммы 4 и неравенства в (8) для любого $\alpha \in (0,\, \overline{\alpha}_k]$ получаем

$$f(x^k + \alpha d^k) - f(x^k) \leqslant \langle f'(x^k),\, \alpha d^k\rangle + \frac{L}{2}\,\alpha^2 |d^k|^2 =$$

$$= \alpha\left(\langle f'(x^k),\, d^k\rangle + \frac{L}{2}\,\alpha |d^k|^2\right) \leqslant \varepsilon\alpha\langle f'(x^k),\, d^k\rangle. \ \square$$

Таким образом, в условиях леммы 3, если

$$\frac{\langle f'(x^k),\, d^k\rangle}{|d^k|^2} \leqslant \delta < 0, \tag{9}$$

где δ — не зависящая от k константа, а числа $\widehat{\alpha}$, ε и θ берутся одними и теми же на всех итерациях, то количество дроблений в правиле Армихо будет конечным равномерно по k, т. е. величины α_k будут отделены от нуля не зависящей от k константой. На практике это обычно означает, что, начиная с некоторой итерации, α_k перестает меняться, т. е. фактически реализуется следующее правило выбора параметра длины шага.

Правило постоянного параметра. Фиксируем (не зависящее от k) число $\overline{\alpha} > 0$ и полагаем $\alpha_k = \overline{\alpha}$. Это правило самое простое; его обычно применяют в тех случаях, когда вычисление значений целевой функции задачи является трудоемкой операцией. Заметим, что если в условиях леммы 3 выполнено (9), то при достаточно малом $\overline{\alpha}$ неравенство (6) будет выполнено на каждом шаге при $\alpha = \alpha_k$. Таким образом, теоретический анализ методов с достаточно малым постоянным параметром длины шага сводится к анализу методов, использующих правило Армихо.

Заметим, наконец, что если удается определить константу Липшица L либо оценить ее сверху, то формула (8) может использоваться для явного вычисления параметров длины шага. Однако, как уже отмечалось выше, оценка константы Липшица редко бывает легко осуществима. Кроме того, чем больше L, тем меньше $\overline{\alpha}_k$ в лемме 3, т. е. тем короче будут шаги метода спуска, а значит, тем ниже будет скорость сходимости.

Приведенные три правила одномерного поиска рассматриваются ниже как основные; именно для них доказываются все утверждения. Тем не менее на практике часто пользуются и другими (более сложными) правилами. Одним из них является *правило Голдстейна*, состоящее в выборе параметра длины шага, удовлетворяющего двойному неравенству

$$\varepsilon_1 \leqslant \frac{f(x^k + \alpha d^k) - f(x^k)}{\alpha\langle f'(x^k),\, d^k\rangle} \leqslant \varepsilon_2 \tag{10}$$

при фиксированных $\varepsilon_1, \varepsilon_2 \in (0,\, 1)$, $\varepsilon_1 < \varepsilon_2$. Левое неравенство — это неравенство Армихо (6) при $\varepsilon = \varepsilon_1$; оно обеспечивает достаточное убывание значения целевой функции. Вместе с тем, согласно лемме 2, неравенство Армихо выполняется для любого достаточно малого $\alpha > 0$, в отличие от правого неравенства в (10), которое, как легко

видеть, нарушается при всех α, достаточно близких к нулю. В этом и заключается смысл введения правого неравенства: оно не позволяет выбирать слишком малые параметры длины шага, тем самым препятствуя замедлению метода.

Другой реализацией той же идеи является *правило Вулфа*, которое вводится так же, как правило Голдстейна, но вместо (10) использует неравенства

$$f(x^k + \alpha d^k) \leqslant f(x^k) + \varepsilon_1 \alpha \langle f'(x^k),\, d^k \rangle, \tag{11}$$

$$\langle f'(x^k + \alpha d^k),\, d^k \rangle \geqslant \varepsilon_2 \langle f'(x^k),\, d^k \rangle. \tag{12}$$

В тех случаях, когда вычисление производной функции f не слишком трудоемко, правило Вулфа признается наиболее эффективным известным правилом одномерного поиска. Важное свойство этого правила связано с квазиньютоновскими методами (см. п. 3.2.3).

Приведем процедуру, реализующую правило Вулфа (правило Голдстейна может быть реализовано аналогичным образом). Фиксируем числа $\varepsilon_1, \varepsilon_2 \in (0, 1)$, $\varepsilon_1 < \varepsilon_2$. Полагаем $\check{\alpha} = \widehat{\alpha} = 0$. Выбираем начальное пробное значение $\alpha > 0$.

1. Проверяем выполнение неравенств (11) и (12). Если оба они выполнены, то переходим к п. 6.

2. Если нарушено (11), то полагаем $\widehat{\alpha} = \alpha$ и переходим к п. 5.

3. Если нарушено (12), то полагаем $\check{\alpha} = \alpha$.

4. Если $\widehat{\alpha} = 0$, то выбираем новое пробное значение $\alpha > \check{\alpha}$ («экстраполяция») и переходим к п. 1.

5. Выбираем новое пробное значение $\alpha \in (\check{\alpha}, \widehat{\alpha})$ («интерполяция») и переходим к п. 1.

6. Полагаем $\alpha_k = \alpha$.

Нарушение неравенства (11) означает, что текущее пробное значение α «слишком велико», а нарушение неравенства (12) означает, что «слишком мало». Описанная процедура работает следующим образом. Сначала реализуются шаги «экстраполяции», до тех пор, пока $\widehat{\alpha}$ не станет положительным. После этого выполняются шаги «интерполяции»; при этом $\widehat{\alpha}$ может только уменьшаться, оставаясь положительным, а $\check{\alpha}$ — только увеличиваться, всегда оставаясь меньше $\widehat{\alpha}$.

«Экстраполяция» и «интерполяция» в приведенной процедуре могут быть организованы многими способами. Например, можно фиксировать числа $\theta_1 > 1$, $\theta_2 \in (0, 1)$ и при «экстраполяции» заменять α на $\theta_1 \alpha$, а при «интерполяции» полагать $\alpha = (1 - \theta_2)\check{\alpha} + {} + \theta_2 \widehat{\alpha}$.

Более изощренные способы обсуждаются, например, в [50]. Важно, чтобы в случае бесконечного числа шагов «экстраполяции» величина $\check{\alpha}$ неограниченно возрастала, а в случае бесконечного числа шагов «интерполяции» величина $\hat{\alpha} - \check{\alpha}$ стремилась к нулю.

Лемма 5. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *непрерывно дифференцируема и ограничена снизу на* $\mathbf{R}^n$.

Тогда если для некоторых x^k, $d^k \in \mathbf{R}^n$ *выполнено неравенство* (7), *то реализующая правило Вулфа процедура, в которой* $\check{\alpha} \to +\infty$ *в случае бесконечного числа шагов «экстраполяции» и* $(\hat{\alpha} - \check{\alpha}) \to 0$ *в случае бесконечного числа шагов «интерполяции», будет конечной.*

На самом деле вместо непрерывной дифференцируемости и ограниченности снизу f на $\mathbf{R}^n$ в этой лемме достаточно предполагать аналогичные свойства φ_k на $\mathbf{R}_+ \setminus \{0\}$.

Доказательство. Если предположить бесконечность числа шагов «экстраполяции», то процедурой генерируется бесконечно возрастающая последовательность значений $\check{\alpha}$, для каждого из которых

$$f(x^k + \check{\alpha} d^k) \leqslant f(x^k) + \varepsilon_1 \check{\alpha} \langle f'(x^k),\, d^k \rangle, \tag{13}$$

что в силу неравенства (7) противоречит ограниченности f снизу. Таким образом, число шагов «экстраполяции» конечно.

Предположим теперь, что бесконечно число шагов «интерполяции». Тогда монотонные последовательности значений $\check{\alpha}$ и $\hat{\alpha}$ сходятся к общему пределу $\widetilde{\alpha}$. При этом для элементов первой последовательности справедливы неравенства (13) и

$$\langle f'(x^k + \check{\alpha} d^k),\, d^k \rangle < \varepsilon_2 \langle f'(x^k),\, d^k \rangle, \tag{14}$$

а для элементов второй

$$f(x^k + \hat{\alpha} d^k) > f(x^k) + \varepsilon_1 \hat{\alpha} \langle f'(x^k),\, d^k \rangle. \tag{15}$$

Предельным переходом в (13), (15) получаем равенство

$$f(x^k + \widetilde{\alpha} d^k) = f(x^k) + \varepsilon_1 \widetilde{\alpha} \langle f'(x^k),\, d^k \rangle. \tag{16}$$

С учетом (15) и монотонного убывания значений $\hat{\alpha}$ отсюда получаем, что эти значения всегда остаются строго больше, чем $\widetilde{\alpha}$. Используя равенство (16), перепишем неравенство (15) в виде

$$f(x^k + \hat{\alpha} d^k) > f(x^k) + \varepsilon_1 (\widetilde{\alpha} + \hat{\alpha} - \widetilde{\alpha}) \langle f'(x^k),\, d^k \rangle = \\ = f(x^k + \widetilde{\alpha} d^k) + \varepsilon_1 (\hat{\alpha} - \widetilde{\alpha}) \langle f'(x^k),\, d^k \rangle,$$

т. е., с учетом неравенства $\hat{\alpha} - \widetilde{\alpha} > 0$,

$$\frac{f(x^k + \hat{\alpha} d^k) - f(x^k + \widetilde{\alpha} d^k)}{\hat{\alpha} - \widetilde{\alpha}} > \varepsilon_1 \langle f'(x^k),\, d^k \rangle.$$

Переходя к пределу и принимая во внимание неравенства $\varepsilon_1 < \varepsilon_2$ и (7), имеем

$$\langle f'(x^k + \widetilde{\alpha} d^k),\, d^k\rangle \geqslant \varepsilon_1 \langle f'(x^k),\, d^k\rangle > \varepsilon_2 \langle f'(x^k),\, d^k\rangle. \qquad (17)$$

С другой стороны, предельным переходом в (14) получаем неравенство

$$\langle f'(x^k + \widetilde{\alpha} d^k),\, d^k\rangle \leqslant \varepsilon_2 \langle f'(x^k),\, d^k\rangle,$$

противоречащее неравенству (17). $\square$

В завершение этого пункта заметим, что в последнее время существенное внимание в литературе уделяется так называемым методам с немонотонным одномерным поиском. Эти методы позволяют делать более длинные шаги, допуская даже увеличение значения целевой функции на некоторых итерациях. А именно, при выборе α_k значение $f(x^k + \alpha_k d^k)$ сравнивается не с $f(x^k)$, а с максимальным (либо средним) значением функции f за некоторое фиксированное число предшествующих итераций. Практика показывает, что такие методы могут превосходить по эффективности обычные методы спуска.

3.1.2. Градиентные методы. В этом пункте предполагается, что функция f дифференцируема на $\mathbf{R}^n$.

Градиентными методами называются методы спуска (2), в которых направление убывания d^k полагается равным антиградиенту функции f в точке x^k: $d^k = -f'(x^k)$ (если $f'(x^k) = 0$, то текущая точка x^k является стационарной точкой задачи (1), и работа метода на этом заканчивается; при этом удобно формально полагать, что $x^k = x^{k+1} = \ldots$). Таким образом, схема (2) принимает вид

$$x^{k+1} = x^k - \alpha_k f'(x^k), \quad k = 0, 1, \ldots. \qquad (18)$$

В частности, градиентные методы являются методами первого порядка.

Алгоритм 1. Выбираем $x^0 \in \mathbf{R}^n$ и полагаем $k = 0$. Выбираем одно из трех основных правил одномерного поиска из п. 3.1.1 и необходимые для реализации этого правила параметры: $\widehat{\alpha} > 0$ (либо $\widehat{\alpha} = +\infty$) в случае правила одномерной минимизации, $\widehat{\alpha} > 0$, $\varepsilon, \theta \in (0, 1)$ в случае правила Армихо и $\overline{\alpha} > 0$ в случае правила постоянного параметра.

1. Вычисляем α_k в соответствии с выбранным правилом одномерного поиска по направлению $d^k = -f'(x^k)$ (если $f'(x^k) = 0$, то α_k можно формально считать произвольным числом).

2. Вычисляем x^{k+1} по формуле (18).

3. Увеличиваем номер шага k на 1 и переходим к п. 1.

Градиентный метод, в котором параметры длины шага выбираются по правилу одномерной минимизации, называется *методом скорейшего спуска* [1)], хотя это название не следует понимать буквально. Важной отличительной особенностью этого метода (по крайней мере при $\widehat{\alpha} = +\infty$) является вытекающее из (5) равенство

$$\langle f'(x^{k+1}),\ f'(x^k)\rangle = 0,$$

согласно которому используемые на соседних итерациях направления убывания $d^k = -f'(x^k)$ и $d^{k+1} = -f'(x^{k+1})$ взаимно ортогональны. Таким образом, спуск в данном методе происходит «зигзагообразно».

Неравенство (6) из правила Армихо принимает вид

$$f(x^k - \alpha f'(x^k)) \leqslant f(x^k) - \varepsilon\alpha|f'(x^k)|^2.$$

Заметим, что если $f'(x^k) \neq 0$, то условие (9) здесь выполнено автоматически при $\delta = -1$, а равенство (8) принимает вид

$$\overline{\alpha}_k = \frac{2(1-\varepsilon)}{L}. \tag{19}$$

Правая часть не зависит от k, поэтому, если градиент функции f удовлетворяет условию Липшица на $\mathbf{R}^n$ и используется правило Армихо, то согласно лемме 3

$$\alpha_k \geqslant \breve{\alpha} > 0, \tag{20}$$

где $\breve{\alpha}$ — не зависящая от k константа.

Имеет место следующая теорема о глобальной сходимости градиентных методов.

Теорема 1. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ дифференцируема на $\mathbf{R}^n$ и ее производная непрерывна по Липшицу на $\mathbf{R}^n$ с константой $L > 0$. Пусть в случае использования в алгоритме* 1 *правила постоянного параметра $\overline{\alpha}$ удовлетворяет условию*

$$\overline{\alpha} < \frac{2}{L}. \tag{21}$$

Тогда любая предельная точка любой траектории $\{x^k\}$ алгоритма 1 *является стационарной точкой задачи* (1). *Если предельная точка существует или если функция f ограничена снизу на $\mathbf{R}^n$, то*

$$\{f'(x^k)\} \to 0 \quad (k \to \infty). \tag{22}$$

Доказательство. Сначала рассмотрим случай использования правила Армихо. Если ни для какого k равенство $f'(x^k) = 0$ не реализуется, то последовательность $\{f(x^k)\}$ монотонно убывает. Пусть

[1)] В англоязычной литературе так принято называть любые градиентные методы.

последовательность $\{x^k\}$ имеет предельную точку; тогда из монотонного убывания последовательности $\{f(x^k)\}$ следует ее ограниченность снизу, а значит, и сходимость (если функция f ограничена снизу на $\mathbf{R}^n$, то это верно и при отсутствии у последовательности $\{x^k\}$ предельной точки). В силу правила Армихо и левого неравенства в (20) для любого k имеем

$$f(x^k) - f(x^{k+1}) \geqslant \varepsilon \alpha_k |f'(x^k)|^2 \geqslant \varepsilon \breve{\alpha} |f'(x^k)|^2,$$

причем левая часть этого неравенства стремится к нулю при $k \to \infty$. Значит, имеет место (22), что и дает требуемый результат.

Пусть теперь используется правило одномерной минимизации. Для каждого k обозначим через $\widetilde{x}^{k+1}$ точку, которая получилась бы из текущего приближения x^k, если бы использовалось правило Армихо, а через $\widetilde{\alpha}_k$ — соответствующий параметр длины шага. Тогда

$$f(x^k) - f(x^{k+1}) \geqslant f(x^k) - f(\widetilde{x}^{k+1}) \geqslant \varepsilon \widetilde{\alpha}_k |f'(x^k)|^2.$$

Остается повторить предыдущую часть доказательства, заменив в ней α_k на $\widetilde{\alpha}_k$.

Пусть, наконец, используется правило постоянного параметра. Но согласно лемме 3, равенству (19) и условию (21) это равносильно тому, что используется правило Армихо при $\widehat{\alpha} = \overline{\alpha}$ и достаточно малом $\varepsilon > 0$. $\square$

В случае использования правила одномерной минимизации или правила Армихо более тонкий анализ позволяет снять в теореме 1 требование непрерывности производной f по Липшицу, заменив его требованием непрерывной дифференцируемости f на $\mathbf{R}^n$. Подчеркнем, что все трудности здесь связаны с возможным невыполнением левого неравенства в (20) при любом $\breve{\alpha} > 0$, т. е. с возможным бесконечным уменьшением параметров длины шага.

Теорема 2. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ непрерывно дифференцируема на $\mathbf{R}^n$. Пусть в алгоритме* 1 *используется правило одномерной минимизации или правило Армихо.*

Тогда любая предельная точка любой траектории $\{x^k\}$ алгоритма 1 *является стационарной точкой задачи* (1). *Если последовательность $\{x^k\}$ ограничена, то имеет место предельное соотношение* (22).

Доказательство. Случай использования правила одномерной минимизации сводится к случаю использования правила Армихо так же, как и при доказательстве теоремы 1, поэтому будем говорить только о правиле Армихо. Пусть последовательность $\{x^k\}$ имеет предельную точку $\overline{x} \in \mathbf{R}^n$, причем ни для какого k равенство $f'(x^k) = 0$ не реализуется. Пусть $\{x^{k_j}\} \to \overline{x}$ $(j \to \infty)$. Случай отделенной от нуля последовательности $\{\alpha_{k_j}\}$ рассматривается так

же, как при доказательстве теоремы 1 (только все рассуждения проводятся для $\{x^{k_j}\}$ вместо $\{x^k\}$), поэтому можем предположить, что $\{\alpha_{k_j}\} \to 0 \quad (j \to \infty)$.

Тогда для любого достаточно большого j при определении α_{k_j} начальное значение $\widehat{\alpha}$ было уменьшено по крайней мере один раз, т. е. $\alpha = \alpha_{k_j}/\theta$ не удовлетворяет неравенству (6):

$$f\left(x^{k_j} - \frac{\alpha_{k_j}}{\theta} f'(x^{k_j})\right) > f(x^{k_j}) - \varepsilon \frac{\alpha_{k_j}}{\theta} |f'(x^{k_j})|^2.$$

Обозначим $\widetilde{\alpha}_{k_j} = \alpha_{k_j}|f'(x^{k_j})|/\theta$, тогда последнее неравенство принимает вид

$$f\left(x^{k_j} - \widetilde{\alpha}_{k_j} \frac{f'(x^{k_j})}{|f'(x^{k_j})|}\right) > f(x^{k_j}) - \varepsilon \widetilde{\alpha}_{k_j} |f'(x^{k_j})|.$$

Заметим, что $\{\widetilde{\alpha}_{k_j}\} \to 0 \quad (j \to \infty)$. Отсюда и из последнего неравенства с помощью теоремы о среднем, деления на $\widetilde{\alpha}_{k_j}$ и предельного перехода легко выводится, что

$$|f'(\overline{x})| \leqslant \varepsilon |f'(\overline{x})|,$$

а это возможно только при $f'(\overline{x}) = 0$.

Последнее утверждение теоремы элементарно доказывается от противного. □

Задача 2. Доказать аналог теоремы 2 для градиентного метода с одномерным поиском по правилу Голдстейна или Вулфа.

Существование предельной точки у траектории градиентного метода в теоремах 1 и 2 не утверждается. Гарантировать существование предельной точки можно при дополнительном предположении об ограниченности множества Лебега $L(f(x^0)) = L_{f,\mathbf{R}^n}(f(x^0))$, или, более общим образом, об ограниченности самой последовательности $\{x^k\}$.

Задача 3. Пусть в дополнение к условиям теоремы 2 множество $L(f(x^0))$ ограничено. Доказать, что при этом

$$\operatorname{dist}(x^k,\, S_0(x^0)) \to 0 \quad (k \to \infty),$$

где $S_0(x^0) = \{x \in L(f(x^0)) \,\big|\, f'(x) = 0\}$.

Перейдем к локальному анализу градиентных методов, т. е. к анализу их поведения вблизи заданной стационарной точки $\overline{x} \in \mathbf{R}^n$ задачи (1). Будем предполагать двукратную дифференцируемость функции f в этой точке и выполнение в ней достаточного условия второго порядка оптимальности. Очевидно, что при этом в точке $\overline{x}$ выполнено так называемое *условие квадратичного роста*, которое будет играть центральную роль в локальном анализе: существуют окрестность U точки $\overline{x}$ и число $\gamma > 0$ такие, что

$$f(x) - f(\overline{x}) =$$
$$= \frac{1}{2}\langle f''(\overline{x})(x-\overline{x}),\, x-\overline{x}\rangle + o(|x-\overline{x}|^2) \geqslant \gamma|x-\overline{x}|^2 \quad \forall\, x \in U. \quad (23)$$

Лемма 6. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *дифференцируема в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$ *и дважды дифференцируема в этой точке. Пусть* $\overline{x}$ *является стационарной точкой задачи* (1), *причем в этой точке выполнено сформулированное в теореме* 1.2.5 *достаточное условие второго порядка оптимальности.*

Тогда для любого числа $\nu \in (0,4)$ *найдется окрестность* U *точки* $\overline{x}$ *такая, что наряду с* (23) *будет выполнено*

$$|f'(x)|^2 \geqslant \nu\gamma(f(x) - f(\overline{x})) \quad \forall\, x \in U. \quad (24)$$

Доказательство. Для $x \in \mathbf{R}^n$, достаточно близкого к $\overline{x}$, имеем

$$f'(x) = f''(\overline{x})(x-\overline{x}) + o(|x-\overline{x}|),$$

поэтому

$$f(x) - f(\overline{x}) = \frac{1}{2}\langle f''(\overline{x})(x-\overline{x}),\, x-\overline{x}\rangle + o(|x-\overline{x}|^2) =$$
$$= \frac{1}{2}\langle f'(x),\, x-\overline{x}\rangle + o(|x-\overline{x}|^2),$$

т. е.

$$\langle f'(x),\, x-\overline{x}\rangle = 2(f(x) - f(\overline{x})) + o(|x-\overline{x}|^2).$$

Тогда, привлекая соотношение (23), имеем

$$\langle f'(x),\, x-\overline{x}\rangle - \sqrt{\nu}(f(x) - f(\overline{x})) =$$
$$= (2-\sqrt{\nu})(f(x) - f(\overline{x})) + o(|x-\overline{x}|^2) \geqslant$$
$$\geqslant (2-\sqrt{\nu})\gamma|x-\overline{x}|^2 + o(|x-\overline{x}|^2) > 0,$$

поэтому

$$\langle f'(x),\, x-\overline{x}\rangle \geqslant \sqrt{\nu}(f(x) - f(\overline{x})). \quad (25)$$

Отсюда и из соотношения (23) следует, что

$$|f'(x)||x-\overline{x}| \geqslant \langle f'(x),\, x-\overline{x}\rangle \geqslant \sqrt{\nu}(f(x) - f(\overline{x})) \geqslant \sqrt{\nu}\gamma|x-\overline{x}|^2,$$

т. е.

$$|f'(x)| \geqslant \sqrt{\nu}\gamma|x-\overline{x}|.$$

Тогда с учетом (25)

$$|f'(x)|^2 \geqslant \sqrt{\nu}\gamma|x-\overline{x}|\,|f'(x)| \geqslant \sqrt{\nu}\gamma\langle f'(x),\, x-\overline{x}\rangle \geqslant \nu\gamma(f(x) - f(\overline{x})),$$

что и требовалось доказать. □

Теорема 3. *Пусть в дополнение к условиям теоремы* 1 *функция* f *дважды дифференцируема в точке* $\overline{x} \in \mathbf{R}^n$. *Пусть* $\overline{x}$ *яв-*

ляется стационарной точкой задачи (1), *причем в этой точке выполнено сформулированное в теореме* 1.2.5 *достаточное условие второго порядка оптимальности. Пусть, наконец, в случае использования в алгоритме* 1 *правила одномерной минимизации величина* $\widehat{\alpha}$ *выбрана конечной.*

Тогда:

а) *траектория алгоритма* 1, *определяемая любым начальным приближением* $x^0 \in \mathbf{R}^n$, *достаточно близким к* $\overline{x}$, *сходится к* $\overline{x}$; *скорость сходимости по функции линейная, а по аргументу геометрическая*: $\forall\, k = 0, 1, \ldots$ *имеют место оценки*

$$f(x^{k+1}) - f(\overline{x}) \leqslant q(f(x^k) - f(\overline{x})), \tag{26}$$

$$|x^k - \overline{x}| \leqslant \Gamma\sqrt{(f(x^k) - f(\overline{x}))} \leqslant (\sqrt{q})^k \Gamma\sqrt{(f(x^0) - f(\overline{x}))}, \tag{27}$$

где числа $q \in [0, 1)$ *и* $\Gamma > 0$ *не зависят ни от* k, *ни от* x^0;

б) *если используется правило одномерной минимизации, причем* $\widehat{\alpha}$ *удовлетворяет условию*

$$\widehat{\alpha} \geqslant \frac{1}{L}, \tag{28}$$

то

$$\limsup_{k\to\infty} \frac{f(x^{k+1}) - f(\overline{x})}{f(x^k) - f(\overline{x})} \leqslant 1 - \frac{\sigma}{\Sigma}, \tag{29}$$

где $\sigma > 0$ — *минимальное, а* $\Sigma > 0$ — *максимальное собственные значения матрицы* $f''(\overline{x})$;

в) *если используется правило Армихо, причем* $\widehat{\alpha}$ *удовлетворяет условию* (28), *а* θ — *условию*

$$\theta \leqslant \frac{1}{2}, \tag{30}$$

то

$$\limsup_{k\to\infty} \frac{f(x^{k+1}) - f(\overline{x})}{f(x^k) - f(\overline{x})} \leqslant 1 - 2\varepsilon(1-\varepsilon)\frac{\sigma}{\Sigma}; \tag{31}$$

если же снять условие (30) *на* θ, *но предположить, что* $\varepsilon \geqslant 1/2$, *то*

$$\limsup_{k\to\infty} \frac{f(x^{k+1}) - f(\overline{x})}{f(x^k) - f(\overline{x})} \leqslant 1 - \theta\frac{\sigma}{\Sigma}; \tag{32}$$

г) *если используется правило постоянного параметра, то*

$$\limsup_{k\to\infty} \frac{f(x^{k+1}) - f(\overline{x})}{f(x^k) - f(\overline{x})} \leqslant 1 - 2\sigma\overline{\alpha} + \sigma\Sigma\,\overline{\alpha}^2. \tag{33}$$

Доказательство. Фиксируем окрестность U точки $\overline{x}$ такую, что выполнены соотношения (23) и (24) при некоторых $\gamma > 0$ и $\nu \in (0, 4)$ (достаточное условие второго порядка оптимальности и лемма 6 гарантируют существование такой окрестности).

Пусть $x^k \in U$. В случае использования правила одномерной минимизации $\forall\, \alpha \in [0, \widehat{\alpha}]$ имеем

$$f(x^{k+1}) - f(\overline{x}) \leqslant f(x^k - \alpha f'(x^k)) - f(\overline{x}) \leqslant$$
$$\leqslant f(x^k) - f(\overline{x}) - \alpha |f'(x^k)|^2 + \frac{L}{2}\,\alpha^2 |f'(x^k)|^2,$$

где последнее неравенство следует из леммы 4. Анализируя квадратный трехчлен (относительно α) в правой части этого неравенства, убеждаемся, что он достигает минимума на $[0, \widehat{\alpha}]$ при $\alpha = \min\{\widehat{\alpha}, 1/L\}$, т. е. с учетом (24) имеем

$$f(x^{k+1}) - f(\overline{x}) \leqslant f(x^k) - f(\overline{x}) - \alpha\left(1 - \alpha\frac{L}{2}\right)|f'(x^k)|^2 \leqslant$$
$$\leqslant \left(1 - \min\left\{\widehat{\alpha},\, \frac{1}{L}\right\} \max\left\{1 - \widehat{\alpha}\,\frac{L}{2},\, \frac{1}{2}\right\} \nu\gamma\right)(f(x^k) - f(\overline{x})). \quad (34)$$

В случае использования правила Армихо в силу соотношений (20) и (24) имеем

$$f(x^{k+1}) - f(\overline{x}) \leqslant f(x^k) - f(\overline{x}) - \varepsilon\breve{\alpha}|f'(x^k)|^2 \leqslant$$
$$\leqslant (1 - \varepsilon\breve{\alpha}\nu\gamma)(f(x^k) - f(\overline{x})). \quad (35)$$

Случай использования правила постоянного параметра сводится к случаю использования правила Армихо (см. доказательство теоремы 1; в этом случае $\breve{\alpha} = \widehat{\alpha} = \overline{\alpha}$). Таким образом, в любом из рассмотренных случаев справедливо неравенство (26) при соответствующем $q < 1$.

Ближайшая задача — доказать, что если точка x^0 достаточно близка к $\overline{x}$, то траектория $\{x^k\}$ не покидает окрестность U. Для этого потребуется следующее неравенство, вытекающее из (23):

$$|f'(x)| \leqslant L|x - \overline{x}| \leqslant L\sqrt{\frac{f(x) - f(\overline{x})}{\gamma}} \qquad \forall\, x \in U. \quad (36)$$

Фиксируем число $r > 0$ такое, что $B(\overline{x}, r) \subset U$, а затем определим число $\delta > 0$, удовлетворяющее условию

$$\delta + \frac{\widehat{\alpha} L\sqrt{(f(x) - f(\overline{x}))/\gamma}}{1 - \sqrt{|q|}} \leqslant r \qquad \forall\, x \in B(\overline{x},\, \delta) \quad (37)$$

(подчеркнем, что при этом $\delta \leqslant r$). Пусть $x^0 \in B(\overline{x}, \delta)$; тогда в силу соотношений (36) и (37) имеем

$$|x^1 - \overline{x}| \leqslant |x^0 - \overline{x}| + |x^1 - x^0| \leqslant \delta + \widehat{\alpha}|f'(x^0)| \leqslant$$
$$\leqslant \delta + \widehat{\alpha} L\sqrt{\frac{f(x^0) - f(\overline{x})}{\gamma}} \leqslant r,$$

т. е. $x^1 \in B(\overline{x}, r)$. Отсюда и из соотношения (23), в частности, следует, что $q \geqslant 0$, так как иначе выполнение неравенства (26) при $k = 0$ было бы невозможно.

Предположим, что $x^k \in B(\overline{x}, r) \quad \forall\, k = 1, \ldots, s$. Тогда согласно (26)

$$f(x^k) - f(\overline{x}) \leqslant q(f(x^{k-1}) - f(\overline{x})) \leqslant \ldots$$
$$\ldots \leqslant q^k(f(x^0) - f(\overline{x})) \qquad \forall\, k = 0, \ldots, s.$$

Поэтому с учетом неравенства (36) получаем

$$|x^{k+1} - x^k| \leqslant \widehat{\alpha}|f'(x^k)| \leqslant \widehat{\alpha}L\sqrt{\frac{f(x^k) - f(\overline{x})}{\gamma}} \leqslant$$
$$\leqslant (\sqrt{q})^k \widehat{\alpha}L\sqrt{\frac{f(x^0) - f(\overline{x})}{\gamma}} \qquad \forall\, k = 0, \ldots, s.$$

Отсюда и из (37) следует, что

$$|x^{s+1} - \overline{x}| \leqslant |x^s - \overline{x}| + |x^{s+1} - x^s| \leqslant \ldots \leqslant |x^0 - \overline{x}| + \sum_{k=0}^{s} |x^{k+1} - x^k| \leqslant$$
$$\leqslant \delta + \widehat{\alpha}L\sqrt{\frac{f(x^0) - f(\overline{x})}{\gamma}} \sum_{k=0}^{s} (\sqrt{q})^k \leqslant \delta + \frac{\widehat{\alpha}L\sqrt{(f(x^0) - f(\overline{x}))/\gamma}}{1 - \sqrt{q}} \leqslant r,$$

т. е. $x^{s+1} \in B(\overline{x}, r)$.

Тем самым доказано, что $\{x^k\} \subset U$. В частности, для любого k имеет место оценка (26), откуда с учетом соотношения (23) немедленно следует оценка (27) при $\Gamma = 1/\sqrt{\gamma}$. Теперь сходимость $\{x^k\}$ к $\overline{x}$ очевидна, что завершает доказательство утверждения а).

Утверждения б)–г) легко следуют из утверждения а), формул (34), (35), асимптотических оценок числа γ и константы Липшица для производной функции f на U через σ и Σ соответственно при стягивании окрестности U в точку $\overline{x}$, а также того факта, что в лемме 6 число ν можно взять сколь угодно близким к 4. При доказательстве в) нужно указать максимальное $\breve{\alpha}$, при котором гарантировано выполнение (20). При доказательстве г) в качестве ε нужно взять число, для которого выполнено равенство (19) при $\overline{\alpha}_k = \overline{\alpha}$, т. е. $\varepsilon = 1 - \overline{\alpha}L/2$. $\square$

Задача 4. Доказать утверждения б)–г) теоремы 3.

Задача 5. Доказать, что условие дифференцируемости функции f на всем $\mathbf{R}^n$ (и соответственно условие непрерывности по Липшицу ее производной на всем $\mathbf{R}^n$) в теореме 3 можно ослабить, ограничившись аналогичным условием лишь в некоторой окрестности точки $\overline{x}$.

Более тонкий анализ позволяет уточнить оценку (29) для метода скорейшего спуска [6, 41, 50].

С точки зрения оценки (31) в правиле Армихо предпочтительно выбирать ε близким к $1/2$, однако даже при $\varepsilon = 1/2$ оценка (31) несколько хуже, чем оценка (29) для метода скорейшего спуска, как и оценка (32). С точки зрения последней предпочтительно выбирать θ как можно более близким к 1, однако при стремлении θ к 1 каждое использование правила Армихо становится, вообще говоря, все более трудоемким. Наконец, оптимальным с точки зрения оценки (33) выбором постоянного шагового параметра является $\overline{\alpha} = 1/L$; при этом оценка (33) совпадает с (29).

Дополнительные предположения о гладкости и/или выпуклости f позволяют несколько уточнить (количественно, но не качественно) полученные оценки скорости сходимости по крайней мере для метода скорейшего спуска и градиентного метода с постоянным параметром длины шага [6, 10, 34].

Одно из направлений дальнейшего развития локального анализа градиентных методов связано с отказом от достаточного условия второго порядка оптимальности в искомом локальном решении $\overline{x}$ и соответственно с использованием условия квадратичного роста в более слабой форме. Например, достаточное условие второго порядка не может выполняться, если $\overline{x}$ — глобальное решение задачи (1), не являющееся изолированным от других точек множества глобальных решений S. Вместе с тем этот случай важен, например, в связи с задачами метода наименьших квадратов для недоопределенных систем уравнений, когда

$$f(x) = \frac{1}{2}\,|F(x)|^2, \quad x \in \mathbf{R}^n, \tag{38}$$

где $F\colon \mathbf{R}^n \to \mathbf{R}^l$ — гладкое отображение, $n > l$. В такой ситуации условие квадратичного роста (23) естественно заменить следующей оценкой расстояния до множества S:

$$f(x) - \overline{v} \geqslant \gamma(\operatorname{dist}(x,\, S))^2 \quad \forall\, x \in U, \tag{39}$$

где U — некоторая окрестность точки $\overline{x}$, $\gamma > 0$ — константа, а $\overline{v} = f(\overline{x})$ — значение задачи (1). Например, если отображение F удовлетворяет в точке $\overline{x}$ условиям теоремы 1.3.3, то для введенной в (38) функции f оценка (39) следует немедленно из этой теоремы. Значение таких рассмотрений состоит еще и в том, что, в отличие от случая выполнения (23), они не подразумевают выпуклость функции f даже локально, вблизи $\overline{x}$.

В анализе такого рода обычно отказываются от чисто локальных рассмотрений, предполагая, что U в (39) — окрестность всего множества S, а не только его точки $\overline{x}$. Для соответствующего варианта (39) также используют название *условие квадратичного роста*. Часто рассматривают окрестности U специального вида, например,

$U = U(S, \delta) = \{x \in \mathbf{R}^n \mid \operatorname{dist}(x, S) \leqslant \delta\}$, либо $U = L_{f, \mathbf{R}^n}(\overline{v} + \delta)$, либо $U = L_{|f'(\cdot)|, \mathbf{R}^n}(\delta)$ при некотором $\delta > 0$. Основное содержание соответствующих теорем составляет доказательство сходимости метода к множеству S либо к точке этого множества, а также получение оценок скорости сходимости. Например, в соответствующих требованиях гладкости выполнение (39) при $U = U(S, \delta)$ и некоторых $\delta > 0$ и $\gamma > 0$ гарантирует следующее: если начальное приближение x^0 выбрано достаточно близким к S, то определяемая им траектория градиентного метода, использующего правило Армихо, сходится к точке множества S с линейной скоростью по функции и с геометрической скоростью по аргументу (см. [21]; там же можно найти анализ случаев выполнения условия квадратичного роста). Родственный, но в некоторых отношениях более общий результат доказывается ниже.

Иногда рассматривают также случай, когда вместо (39) выполнено более слабое условие

$$f(x) - \overline{v} \geqslant \gamma(\operatorname{dist}(x, S))^p \quad \forall\, x \in U$$

при $p > 2$. Для этого случая известны результаты о сходимости градиентных методов с арифметической скоростью.

При изучении сходимости и скорости сходимости градиентных методов к множеству S_0 стационарных точек задачи (1) условие квадратичного роста естественно заменить оценкой расстояния до этого множества следующего вида:

$$|f'(x)| \geqslant \gamma \operatorname{dist}(x, S_0) \quad \forall\, x \in U, \tag{40}$$

где U — некоторая окрестность множества S_0, а $\gamma > 0$ — константа.

Задача 6. Внося необходимые изменения в доказательство леммы 6, доказать следующее утверждение. Пусть существуют числа $\delta > 0$ и $\gamma > 0$ такие, что функция $f\colon \mathbf{R}^n \to \mathbf{R}$ трижды дифференцируема на $U = U(S, \delta)$, норма ее третьей производная ограничена на U и выполнено (39). Тогда для любого числа $\nu \in (0, 4)$ найдется число $r > 0$ такое, что

$$|f'(x)|^2 \geqslant \nu\gamma(f(x) - \overline{v}) \quad \forall\, x \in U(S, r).$$

Подчеркнем, что условие квадратичного роста в определенном смысле влечет оценку (40). Точнее, из результата задачи 6 следует, что если $S = S_0$, функция f обладает достаточной гладкостью и выполнено условие квадратичного роста (39) при $U = U(S, \delta)$ и некоторых $\delta > 0$ и $\gamma > 0$, то выполнена и оценка (40) при соответствующих (возможно, других) окрестности U множества S_0 и числе $\gamma > 0$.

В дальнейшем будем считать, что окрестность U в (40) имеет вид $U = L_{|f'(\cdot)|,\mathbf{R}^n}(\delta)$ при некотором $\delta > 0$ (заметим, что выполнение оценки (40) при $U = U(S_0, \delta)$ и некотором $\delta > 0$ влечет ее выполнение при $U = L_{|f'(\cdot)|,\mathbf{R}^n}(\delta)$, возможно, с другим $\delta > 0$). Кроме того, будем предполагать, что выполнены условия теоремы 1 и $\overline{v} > -\infty$, что влечет выполнение (22) для любой траектории $\{x^k\}$ алгоритма 1, поэтому все приближения, начиная с некоторого, будут лежать в U. Именно это сообщает приводимому далее результату о сходимости и скорости сходимости полностью глобальный характер.

Еще одно техническое условие, выполнение которого придется потребовать, называется *условием отделимости критических поверхностей уровня* и состоит в следующем: существует число $\chi > 0$ такое, что для любых $\overline{x}^1, \overline{x}^2 \in S_0$, для которых $f(\overline{x}^1) \neq f(\overline{x}^2)$, выполнено $|\overline{x}^1 - \overline{x}^2| \geqslant \chi$. Элементарно проверяется, что если множество S_0 гладко связно, т. е. любые две точки этого множества можно соединить гладкой кривой, то всюду на этом множестве функция f принимает одно и то же значение, и условие отделимости критических поверхностей уровня выполняется тривиальным образом.

Задача 7. Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ дифференцируема на $\mathbf{R}^n$, ее производная непрерывна по Липшицу на $\mathbf{R}^n$, и пусть f принимает на множестве своих критических точек лишь конечное число значений. Показать, что при этом выполнено условие отделимости критических поверхностей уровня.

Задача 8. Пусть $f\colon \mathbf{R}^n \to \mathbf{R}$ — квадратичная функция, имеющая критическую точку. Показать, что при этом выполнено как условие (40) (при $U = \mathbf{R}^n$ и некотором $\gamma > 0$), так и условие отделимости критических поверхностей уровня (тривиальным образом).

Теорема 4. *Пусть в дополнение к условиям теоремы* 1 *значение задачи* (1) *конечно и выполнено условие* (40) *при* $U = L_{|f'(\cdot)|,\mathbf{R}^n}(\delta)$ *и некоторых* $\delta > 0$ *и* $\gamma > 0$. *Пусть, кроме того, выполнено условие отделимости критических поверхностей уровня. Пусть в алгоритме* 1 *используется правило Армихо или правило постоянного параметра.*

Тогда любая траектория $\{x^k\}$ *алгоритма* 1 *сходится к некоторой стационарной точке задачи* (1). *Скорость сходимости по функции линейная, а по аргументу — геометрическая.*

Доказательство. Случай использования правила постоянного параметра сводится к случаю использования правила Армихо так же, как и выше, поэтому будем говорить только о последнем. Тогда $\forall\, k$

$$f(x^{k+1}) - f(x^k) \leqslant -\frac{\varepsilon}{\widehat{\alpha}}\,|x^{k+1} - x^k|^2. \tag{41}$$

Поэтому если ни для какого k не реализуется равенство $f'(x^k) = 0$, то последовательность $\{f(x^k)\}$ монотонно убывает, а значит, в силу

конечности значения задачи (1) сходится, что влечет предельное соотношение $|x^{k+1} - x^k| \to 0 \quad (k \to \infty)$. Используя соотношения (18) и (20), отсюда выводим

$$\breve{\alpha}|f'(x^k)| \leqslant \alpha_k|f'(x^k)| = |x^{k+1} - x^k| \to 0 \quad (k \to \infty),$$

поэтому без ограничения общности можно считать, что $\{x^k\} \subset U$. Тогда последнее соотношение и оценка (40) приводят к оценке

$$|x^k - \overline{x}^k| \leqslant \frac{1}{\gamma}\,|f'(x^k)| \leqslant \frac{1}{\breve{\alpha}\,\gamma}\,|x^{k+1} - x^k| \to 0 \quad (k \to \infty), \tag{42}$$

где $\overline{x}^k$ — произвольная проекция точки x^k на множество S_0 стационарных точек задачи (1) (см. следствие 1.1.2).

Из (42) имеем

$$|\overline{x}^{k+1} - \overline{x}^k| \leqslant |x^{k+1} - \overline{x}^{k+1}| + |x^{k+1} - x^k| + |x^k - \overline{x}^k| \to 0 \quad (k \to \infty).$$

Но тогда условие отделимости критических поверхностей уровня гарантирует существование числа v такого, что для любого достаточно большого k

$$f(\overline{x}^k) = v. \tag{43}$$

Тогда из того, что $\overline{x}^{k+1} \in S_0$, леммы 4, определения проекции и оценки (42) имеем

$$|f(x^{k+1}) - v| = |f(x^{k+1}) - f(\overline{x}^{k+1}) - \langle f'(\overline{x}^{k+1}),\, x^{k+1} - \overline{x}^{k+1}\rangle| \leqslant$$

$$\leqslant \frac{L}{2}\,|x^{k+1} - \overline{x}^{k+1}|^2 \leqslant \frac{L}{2}\,|x^{k+1} - \overline{x}^k|^2 \leqslant \frac{L}{2}\,(|x^{k+1} - x^k| + |x^k - \overline{x}^k|)^2 \leqslant$$

$$\leqslant \frac{L}{2}\left(1 + \frac{1}{\breve{\alpha}\,\gamma}\right)^2 |x^{k+1} - x^k|^2 \to 0 \quad (k \to \infty).$$

С учетом монотонного убывания последовательности $\{f(x^k)\}$ и (41) отсюда следует, что

$$0 \leqslant f(x^{k+1}) - v \leqslant M(f(x^k) - f(x^{k+1})),$$

где

$$M = \frac{L\widehat{\alpha}}{2\varepsilon}\left(1 + \frac{1}{\breve{\alpha}\,\gamma}\right)^2 > 0.$$

Перегруппируя слагаемые, приходим к оценке

$$0 \leqslant f(x^{k+1}) - v \leqslant q(f(x^k) - v), \tag{44}$$

где $q = M/(1 + M) \in (0, 1)$. Это и означает, что $\{f(x^k)\}$ сходится к v с линейной скоростью.

Далее, используя неравенства (41) и (44), получаем, что для любого достаточно большого k

$$|x^{k+1}-x^k| \leqslant \sqrt{\frac{\widehat{\alpha}}{\varepsilon}\,(f(x^k)-f(x^{k+1}))} \leqslant \sqrt{\frac{\widehat{\alpha}}{\varepsilon}\,(f(x^k)-v)} \leqslant$$

$$\leqslant \sqrt{\frac{\widehat{\alpha}}{\varepsilon}\,q(f(x^{k-1})-v)} \leqslant \ldots \leqslant (\sqrt{q})^k \sqrt{\frac{\widehat{\alpha}}{\varepsilon}\,(f(x^0)-v)}.$$

Обозначим $\Gamma = \sqrt{\widehat{\alpha}(f(x^0)-v)/\varepsilon}$, тогда $\forall i = 0, 1, \ldots$

$$|x^{k+i}-x^k| \leqslant \sum_{j=0}^{i-1} |x^{k+j+1}-x^{k+j}| \leqslant \sum_{j=0}^{i-1} (\sqrt{q})^{k+j}\Gamma \leqslant \frac{(\sqrt{q})^k\Gamma}{1-\sqrt{q}}\,.$$

Отсюда вытекает фундаментальность последовательности $\{x^k\}$, т. е. ее сходимость к некоторой точке $\overline{x} \in \mathbf{R}^n$. Переходя в последнем неравенстве к пределу при $i \to \infty$, приходим к оценке

$$|x^k - \overline{x}| \leqslant \frac{(\sqrt{q})^k\Gamma}{1-\sqrt{q}}\,,$$

которая означает, что скорость сходимости геометрическая. Наконец, из (42) следует, что $\operatorname{dist}(x^k, S_0) \to 0$ $(k \to \infty)$, поэтому $\overline{x} \in S_0$. $\square$

В заключение отметим главный недостаток градиентных методов. Речь идет о неприемлемо низкой скорости сходимости в случаях, когда поверхности уровня целевой функции сильно «вытянуты», т. е. имеют «овражный» характер. Количественно этот эффект можно пояснить так: если метод сходится к строгому локальному решению $\overline{x}$, в котором максимальное собственное значение матрицы $f''(\overline{x})$ значительно больше минимального, то отношение σ/Σ близко к нулю, а значит, правые части оценок (29), (31)–(33) близки к 1. Подчеркнем, что дело здесь отнюдь не в несовершенстве этих оценок: низкая скорость сходимости градиентных методов в подобных ситуациях подтверждается вычислительной практикой. Геометрически (весьма неформально) происходит следующее: генерируемые методом приближения быстро спускаются почти на «дно оврага», а затем начинают «прыгать» с одного его «склона» на другой, причем длина осуществляемых шагов все время уменьшается. Особенно сильный замедляющий эффект это явление оказывает в том случае, когда «дно оврага» изогнуто.

Известны способы ускорения сходимости градиентных методов для «овражных» функций [10]. Кроме того, в § 3.2 и § 3.3 будут рассмотрены методы, обладающие сверхлинейной скоростью сходимости и значительно менее чувствительные к «овражному» характеру целевой функции задачи.

Задача 9. Из начального приближения $x^0 = (-2, 1)$ сделать два шага метода скорейшего спуска для задачи безусловной минимизации функции

$$f\colon \mathbf{R}^2 \to \mathbf{R}, \quad f(x) = x_1^2 + x_1x_2 + 2x_2^2 + x_1.$$

Задача 10. Из начального приближения $x^0 = (1, -1)$ сделать для задачи безусловной минимизации функции

$$f\colon \mathbf{R}^2 \to \mathbf{R}, \quad f(x) = 2x_1^2 + x_1x_2 + 3x_2^2,$$

один шаг градиентного метода, использующего правило Армихо с параметрами $\widehat{\alpha} = 1$, $\varepsilon = \theta = 1/2$.

§ 3.2. Метод Ньютона. Квазиньютоновские методы

Фундаментальная идея рассматриваемого в этом параграфе метода Ньютона чрезвычайно важна: она лежит в основе всех быстро сходящихся алгоритмов для различных классов нелинейных задач, в том числе и задач оптимизации, хотя в своем базовом варианте метод Ньютона не является оптимизационным. Имеется в виду, что, например, в отличие от рассмотренных выше методов спуска, применительно к задачам безусловной оптимизации метод Ньютона одинаково хорошо локально сходится как к минимумам, так и к максимумам и вообще к любым стационарным точкам задачи. С другой стороны, метод допускает и оптимизационную трактовку, дающую почву для глобализации его сходимости.

Высокая скорость сходимости метода достигается за счет того, что он является методом второго порядка. Соответственно его итерация существенно более трудоемка, чем, например, итерации градиентных методов. К счастью, на основе метода Ньютона строятся так называемые квазиньютоновские методы, которые лишь немногим уступают методу Ньютона по скорости сходимости, и при этом их итерации лишь немногим более трудоемки, чем итерации градиентных методов.

3.2.1. Метод Ньютона для уравнений. Классический метод Ньютона вводится для уравнения

$$\Phi(x) = 0, \tag{1}$$

где $\Phi\colon \mathbf{R}^n \to \mathbf{R}^n$ — гладкое отображение. Пусть $x^k \in \mathbf{R}^n$ — текущее приближение к искомому решению $\overline{x}$ уравнения (1); тогда вблизи x^k уравнение можно аппроксимировать его линеаризацией

$$\Phi(x^k) + \Phi'(x^k)(x - x^k) = 0. \tag{2}$$

Это и есть итерационное уравнение *метода Ньютона.*

Алгоритм 1. Выбираем $x^0 \in \mathbf{R}^n$ и полагаем $k = 0$.
1. Вычисляем $x^{k+1} \in \mathbf{R}^n$ как решение уравнения (2).
2. Увеличиваем номер шага k на 1 и переходим к п. 1.

Предполагая невырожденность $\Phi'(x^k)$, итерационную схему метода Ньютона часто записывают в виде

$$x^{k+1} = x^k - (\Phi'(x^k))^{-1}\Phi(x^k), \quad k = 0, 1, \ldots, \tag{3}$$

хотя, разумеется, полное обращение матрицы $\Phi'(x^k)$ при реализации метода совершенно не требуется.

Теорема 1. *Пусть отображение* $\Phi\colon \mathbf{R}^n \to \mathbf{R}^n$ *дифференцируемо в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$, *причем его производная непрерывна в этой точке. Пусть* $\overline{x}$ *является решением уравнения* (1), *причем* $\det \Phi'(\overline{x}) \neq 0$.

Тогда любое начальное приближение $x^0 \in \mathbf{R}^n$, *достаточно близкое к* $\overline{x}$, *корректно определяет траекторию алгоритма* 1, *которая сходится к* $\overline{x}$. *Скорость сходимости сверхлинейная, а если производная* Φ *непрерывна по Липшицу в окрестности точки* $\overline{x}$, *то квадратичная.*

Доказательство. Согласно теореме о малом возмущении невырожденной матрицы [1]) найдутся окрестность U точки $\overline{x}$ и число $M > 0$ такие, что

$$\det \Phi'(x) \neq 0, \quad \|(\Phi'(x))^{-1}\| \leqslant M \quad \forall\, x \in U. \tag{4}$$

В частности, для $x^k \in U$ уравнение (2) имеет единственное решение x^{k+1}, т. е. итерация алгоритма 1 из такой точки x^k корректно определена.

Кроме того, в силу (3), (4) и теоремы о среднем окрестность U можно выбрать так, что если $x^k \in U$, то

$$\begin{aligned} |x^{k+1} - \overline{x}| &= |x^k - \overline{x} - (\Phi'(x^k))^{-1}\Phi(x^k)| \leqslant \\ &\leqslant \|(\Phi'(x^k))^{-1}\| |\Phi(x^k) - \Phi(\overline{x}) - \Phi'(x^k)(x^k - \overline{x})| \leqslant \\ &\leqslant M \sup_{t\in[0,\,1]} \|\Phi'(tx^k + (1-t)\overline{x}) - \Phi'(x^k)\| |x^k - \overline{x}|. \end{aligned} \tag{5}$$

Заметим, что

$$\sup_{t\in[0,\,1]} \|\Phi'(tx^k + (1-t)\overline{x}) - \Phi'(x^k)\| \to 0 \quad (x^k \to \overline{x}),$$

поэтому из (5) следует, что для всякого $q \in (0, 1)$ найдется число $\delta > 0$ такое, что $B(\overline{x}, \delta) \subset U$ и, если $x^k \in B(\overline{x}, \delta)$, то

$$|x^{k+1} - \overline{x}| \leqslant q|x^k - \overline{x}|;$$

[1]) Имеется в виду следующее известное утверждение [12]: если $\overline{A} \in \mathbf{R}(n, n)$, $\det \overline{A} \neq 0$, то для любой матрицы $A \in \mathbf{R}(n, n)$, удовлетворяющей неравенству $\|A - \overline{A}\| < 1/\|\overline{A}^{-1}\|$, справедливо, что $\det A \neq 0$, причем

$$\|A^{-1} - \overline{A}^{-1}\| \leqslant \frac{\|\overline{A}^{-1}\|^2 \|A - \overline{A}\|}{1 - \|\overline{A}^{-1}\| \|A - \overline{A}\|}.$$

в частности, $x^{k+1} \in B(\overline{x}, \delta)$. Таким образом, во-первых, всякое начальное приближение x^0, достаточно близкое к $\overline{x}$, корректно определяет траекторию алгоритма 1, а во-вторых, эта траектория сходится к $\overline{x}$ со сверхлинейной скоростью.

Наконец, если производная Φ непрерывна по Липшицу на U с константой $L > 0$, то из (5) имеем: если $x^k \in U$, то

$$|x^{k+1} - \overline{x}| \leqslant ML|x^k - \overline{x}|^2, \tag{6}$$

что и дает квадратичную скорость сходимости [1]. $\square$

Здесь и в других теоремах о методах ньютоновского типа для квадратичной оценки скорости сходимости вместо непрерывности производной Φ по Липшицу достаточно предполагать выполненным более слабое условие

$$\|\Phi'(x) - \Phi'(\overline{x})\| \leqslant L|x - \overline{x}| \quad \forall\, x \in V,$$

где V — окрестность точки $\overline{x}$, а $L > 0$ — некоторая константа. Тем не менее, следуя традиции, почти везде ниже в этом контексте будем предполагать липшицеву непрерывность производной в окрестности решения.

Иногда в выражении для $\Phi'(\cdot)$ можно априори выделить члены, которые обращаются в нуль в точке $\overline{x}$; особенно часто это случается, когда известно аналитическое выражение для $\Phi'(\cdot)$. Такие члены разумно сразу полагать равными нулю, заменяя $\Phi'(\cdot)$ соответствующим аппроксимирующим отображением $\Psi\colon \mathbf{R}^n \to \mathbf{R}(n, n)$, и рассматривать *неточный метод Ньютона* [2], который можно записать в виде итерационной схемы

$$x^{k+1} = x^k - (\Psi(x^k))^{-1}\Phi(x^k), \quad k = 0, 1, \ldots. \tag{7}$$

Следствие 1. *Пусть выполнены условия теоремы* 1. *Тогда для любого отображения* $\Psi\colon \mathbf{R}^n \to \mathbf{R}(n, n)$ *такого, что*

$$\|\Psi(x) - \Phi'(\overline{x})\| \to 0 \quad (x \to \overline{x}),$$

любое начальное приближение $x^0 \in \mathbf{R}^n$, *достаточно близкое к* $\overline{x}$, *корректно определяет траекторию метода* (7), *которая сходится к* $\overline{x}$. *Скорость сходимости сверхлинейная, а если производная* Φ *непрерывна по Липшицу в окрестности точки* $\overline{x}$ *и существуют окрестность* V *точки* $\overline{x}$ *и число* $N > 0$ *такие, что*

$$\|\Psi(x) - \Phi'(\overline{x})\| \leqslant N|x - \overline{x}| \quad \forall\, x \in V,$$

то квадратичная.

Задача 1. Доказать следствие 1.

---

[1] Константу в (6) можно уменьшить вдвое, если воспользоваться очевидным аналогом леммы 3.1.4 для отображений.

[2] Общепринятый английский термин — Inexact Newton method.

Наряду с неточными методами Ньютона большое практическое значение имеют так называемые *усеченные методы Ньютона* [1], особенно часто используемые для задач большой размерности. В этих методах итерационное уравнение (2) метода Ньютона решают с помощью тех или иных итерационных же методов, причем внутренние итерации продолжают не вплоть до отыскания точного решения уравнения (2), а останавливают после достаточного количества шагов, определяемого с помощью специальных критериев. Например, в качестве внутреннего итерационного процесса часто используется метод сопряженных градиентов применительно к квадрату невязки уравнения (2) (о методах сопряженных градиентов речь пойдет в п. 3.3.2).

Другая важная модификация метода Ньютона, направленная на снижение трудоемкости его итерации, носит название *метода хорд*:

$$x^{k+1} = x^k - (\Phi'(x^0))^{-1}\Phi(x^k), \quad k = 0, 1, \ldots. \tag{8}$$

Здесь нужно вычислять производную отображения Φ только один раз, а не на каждой итерации, и все решаемые линейные системы имеют одну и ту же матрицу, но, конечно, скорость сходимости такого метода ниже, чем у метода Ньютона.

Задача 2. Доказать, что в условиях теоремы 1 метод (8) локально сходится к $\overline{x}$ с линейной скоростью, причем скорость сходимости тем выше, чем ближе x^0 к $\overline{x}$.

На практике обычно используют схему, в которой $\Phi'(x^k)$ вычисляется не только при $k = 0$, но и не на каждом шаге k, а для некоторой бесконечно возрастающей последовательности значений k. Такой «компромисс» между схемами (3) и (8) имеет целью снижение суммарной трудоемкости первой и повышение скорости сходимости второй.

В течение двух последних десятилетий большое внимание исследователей привлекало изучение поведения метода Ньютона в условиях более слабых, чем невырожденность $\Phi'(\overline{x})$, а также построение модификаций метода, работоспособных и эффективных в таких ослабленных предположениях (по этим вопросам см. [21, 22]).

Задача 3. Из начального приближения $x^0 = 1$ сделать два шага метода Ньютона для уравнения

$$x^p = 0,$$

где $p \geqslant 2$ — целочисленный параметр. Проанализировать скорость сходимости; объяснить наблюдаемый эффект. Модифицировать метод Ньютона, введя параметр длины шага, равный p; проанализировать скорость сходимости.

[1] Общепринятый английский термин — Truncated Newton method.

Задача 4. Доказать следующее утверждение. Пусть функция $\Phi\colon \mathbf{R} \to \mathbf{R}$ p раз дифференцируема в точке $\overline{x} \in \mathbf{R}$, $p \geqslant 2$, причем $\overline{x}$ — корень кратности p уравнения (1), т. е.

$$\Phi(\overline{x}) = \Phi'(\overline{x}) = \ldots = \Phi^{(p-1)}(\overline{x}) = 0, \quad \Phi^{(p)}(\overline{x}) \neq 0.$$

Тогда метод Ньютона локально сходится к $\overline{x}$ с линейной скоростью, а если ввести параметр длины шага, равный p, то скорость сходимости станет сверхлинейной.

3.2.2. Метод Ньютона для задачи безусловной оптимизации. Теперь обратимся к задаче безусловной оптимизации

$$f(x) \to \min, \quad x \in \mathbf{R}^n, \tag{9}$$

где $f\colon \mathbf{R}^n \to \mathbf{R}$ — дважды дифференцируемая на $\mathbf{R}^n$ функция. Стационарные точки этой задачи описываются уравнением (1), где

$$\Phi\colon \mathbf{R}^n \to \mathbf{R}^n, \quad \Phi(x) = f'(x),$$

— *градиентное отображение*, поэтому стационарные точки можно искать, применяя метод Ньютона к такому уравнению.

Пусть $x^k \in \mathbf{R}^n$ — текущее приближение к искомой стационарной точке $\overline{x}$ задачи (9); тогда следующее приближение x^{k+1} ищется как решение линейного уравнения

$$f'(x^k) + f''(x^k)(x - x^k) = 0, \tag{10}$$

т. е.

$$x^{k+1} = x^k - (f''(x^k))^{-1} f'(x^k), \quad k = 0, 1, \ldots. \tag{11}$$

Заметим, что описанная итерация допускает следующую оптимизационную трактовку. Вблизи точки x^k задачу (9) можно аппроксимировать задачей безусловной оптимизации с квадратичной целевой функцией:

$$f(x^k) + \langle f'(x^k), x - x^k\rangle + \frac{1}{2}\langle f''(x^k)(x - x^k), x - x^k\rangle \to \min, \quad x \in \mathbf{R}^n, \tag{12}$$

причем (10) есть в точности уравнение для стационарных точек такой задачи. Таким образом, *метод Ньютона* для задачи безусловной оптимизации состоит в следующем.

Алгоритм 2. Выбираем $x^0 \in \mathbf{R}^n$ и полагаем $k = 0$.

1. Вычисляем $x^{k+1} \in \mathbf{R}^n$ как стационарную точку задачи (12).
2. Увеличиваем номер шага k на 1 и переходим к п. 1.

Из теоремы 1 немедленно вытекает

Следствие 2. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *дважды дифференцируема в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$, *причем ее вторая производная непрерывна в этой точке. Пусть* $\overline{x}$ *является стационарной точкой задачи* (9), *причем в этой точке выполнено сформулированное в теореме* 1.2.5 *достаточное условие второго порядка оптимальности.*

Тогда любое начальное приближение $x^0 \in \mathbf{R}^n$, *достаточно близкое к* $\overline{x}$, *корректно определяет траекторию алгоритма* 2, *которая сходится к* $\overline{x}$. *Скорость сходимости сверхлинейная, а если вторая производная* f *непрерывна по Липшицу в окрестности точки* $\overline{x}$, *то квадратичная.*

Напомним, что матрица Гессе дважды дифференцируемой функции является симметрической. В условиях следствия 2 наиболее рациональным способом решения итерационного линейного уравнения (10) является, видимо, алгоритм Холецкого (алгоритм квадратного корня), который позволяет получить LL^{T}-разложение положительно определенной симметрической матрицы за порядка $n^3/6$ умножений и столько же сложений [12] (L — нижняя треугольная матрица с положительными диагональными элементами). Такую процедуру можно снабдить дополнительным механизмом с тем, чтобы в случае нарушения положительной определенности текущей матрицы $f''(x^k)$ она автоматически заменялась положительно определенной матрицей (например, вида $f''(x^k) + \gamma E^n$ при достаточно большом $\gamma > 0$; см. [6, 13, 16, 41, 50]), что важно с точки зрения глобализации сходимости метода (см. п. 3.2.3 и § 5.1). Кроме того, без указанной модификации следствие 2 полностью сохраняет силу, если вместо достаточного условия оптимальности в стационарной точке $\overline{x}$ предполагать выполненным лишь более слабое условие $\det f''(\overline{x}) \neq 0$. В этом смысле метод Ньютона «не отличает» локальные минимумы от других стационарных точек задачи (9).

Главное достоинство метода Ньютона — сверхлинейная скорость сходимости. Напомним, что для рассмотренных в § 3.1 градиентных методов в лучшем случае можно гарантировать лишь линейную скорость сходимости. Основные же недостатки метода Ньютона следующие.

Прежде всего, метод обладает лишь локальной сходимостью в отличие, например, от тех же градиентных методов, которые в разумных предположениях сходятся глобально в том смысле, что любая предельная точка любой траектории такого метода является стационарной точкой задачи (9). Ясно, что если f — квадратичная функция, т. е.

$$f(x) = \langle Ax, x\rangle + \langle b, x\rangle, \quad x \in \mathbf{R}^n, \tag{13}$$

где $A \in \mathbf{R}(n, n)$ — симметрическая матрица, $b \in \mathbf{R}^n$, причем матрица A невырождена, то метод Ньютона найдет ее единственную

критическую точку из любого начального приближения за один шаг. Однако в неквадратичном случае это уже не так: траектория метода может вообще не иметь стационарные точки задачи в числе своих предельных точек, если она определяется неудачным начальным приближением. Следующий пример заимствован из [10].

Пример 1. Рассмотрим функцию

$$f\colon \mathbf{R} \to \mathbf{R}, \quad f(x) = \begin{cases} -\dfrac{x^4}{4\sigma^3} + \left(1 + \dfrac{3}{\sigma}\right)\dfrac{x^2}{2}, & \text{если} \quad |x| \leqslant \sigma, \\ \dfrac{x^2}{2} + 2|x| - \dfrac{3\sigma}{4}, & \text{если} \quad |x| > \sigma, \end{cases}$$

где $\sigma > 0$ — параметр. Нетрудно убедиться, что при любом таком σ функция f дважды непрерывно дифференцируема на $\mathbf{R}$, и задача (9) с такой целевой функцией имеет единственную стационарную точку $\overline{x} = 0$. Более того, $f''(\overline{x}) = 1 + 3/\sigma > 0$, т. е. выполнены все условия следствия 2. Возьмем $x^0 = \sigma$; тогда для траектории $\{x^k\}$, генерируемой алгоритмом 2, имеем: $x^k = 2(-1)^k$, $k = 1, 2, \ldots$, и $\overline{x}$ не является предельной точкой $\{x^k\}$, каким бы малым ни было σ.

Вопросы глобализации сходимости методов ньютоновского типа будут обсуждаться в гл. 5. Пока заметим лишь, что для глобализации может быть полезна приведенная выше оптимизационная трактовка метода, а также введение параметра длины шага, т. е. замена схемы (11) схемой

$$x^{k+1} = x^k - \alpha_k (f''(x^k))^{-1} f'(x^k), \quad k = 0, 1, \ldots,$$

где $\alpha_k > 0$ может быть отлично от 1. Вдали от решения единичный параметр длины шага, характеризующий чистый метод Ньютона, может не обеспечивать монотонного убывания последовательности $\{f(x^k)\}$, и в этом смысле чистый метод Ньютона не является методом спуска.

Кроме того, на каждой итерации метода Ньютона нужно вычислять вторую производную целевой функции и решать соответствующую линейную систему. В целом итерация метода существенно более трудоемка, чем, например, итерации градиентных методов. Модификациям метода, позволяющим в определенном смысле избавиться от этого недостатка и при этом сохранить высокую скорость сходимости, посвящен п. 3.2.3.

Задача 5. Из начального приближения $x^0 = (1, 1)$ сделать два шага метода Ньютона для задачи безусловной минимизации функции

$$f\colon \mathbf{R}^2 \to \mathbf{R}, \quad f(x) = x_1^2 + e^{x_2^2}.$$

Проанализировать скорость сходимости. Показать, что в данном случае метод Ньютона сходится глобально.

3.2.3. Квазиньютоновские методы. Целый ряд методов безусловной оптимизации может быть записан в виде итерационной схемы

$$x^{k+1} = x^k + \alpha_k d^k, \quad d^k = -Q_k f'(x^k), \quad k = 0, 1, \ldots, \tag{14}$$

где для каждого k симметрическая матрица $Q_k \in \mathbf{R}(n, n)$ положительно определена, а $\alpha_k > 0$ — параметр длины шага, выбираемый посредством одной из процедур одномерного поиска для методов спуска, описанных в п. 3.1.1. Напомним, что (14) — действительно метод спуска, поскольку если x^k не является стационарной точкой задачи (9), то

$$\langle f'(x^k), d^k \rangle = -\langle Q_k f'(x^k), f'(x^k) \rangle < 0, \tag{15}$$

т. е. согласно лемме 3.1.1 $d^k \in \mathcal{D}_f(x^k)$. При $Q_k = E^n$ получаем, что схема (14) — один из градиентных методов, а при $Q_k = (f''(x^k))^{-1}$ и $\alpha_k = 1$ — метод Ньютона.

Алгоритм 3. Выбираем $x^0 \in \mathbf{R}^n$ и полагаем $k = 0$. Выбираем одно из трех основных правил одномерного поиска из п. 3.1.1 и необходимые для реализации этого правила параметры: $\widehat{\alpha} > 0$ (либо $\widehat{\alpha} = +\infty$) в случае правила одномерной минимизации, $\widehat{\alpha} > 0$, $\varepsilon, \theta \in (0, 1)$ в случае правила Армихо и $\overline{\alpha} > 0$ в случае правила постоянного параметра.

1. Полагаем $d^k = -Q_k f'(x^k)$, где $Q_k \in \mathbf{R}(n, n)$ — симметрическая положительно определенная матрица, и вычисляем α_k в соответствии с выбранным правилом одномерного поиска по направлению d^k (если $f'(x^k) = 0$, то формально полагаем $\alpha_k = 1$).
2. Полагаем $x^{k+1} = x^k + \alpha_k d^k$.
3. Увеличиваем номер шага k на 1 и переходим к п. 1.

Квазиньютоновские методы — это методы вида (14), в которых матрицы Q_k выбираются таким образом, чтобы они в некотором смысле аппроксимировали $(f''(\overline{x}))^{-1}$ в искомом решении $\overline{x}$. А именно, матрицы Q_k должны удовлетворять условию (17) из следующей теоремы, называемой теоремой Дэнниса–Морэ.

Теорема 2. *Пусть выполнены условия следствия* 2. *Пусть, кроме того, траектория* $\{x^k\}$ *алгоритма* 3, *использующего правило Армихо при* $\widehat{\alpha} = 1$ *и* $\varepsilon \in (0, 1/2)$, *сходится к* $\overline{x}$.

Тогда следующие два утверждения эквивалентны:

а) *скорость сходимости последовательности* $\{x^k\}$ *к* $\overline{x}$ *является сверхлинейной, причем* $\alpha_k = 1$ *для любого достаточно большого* k *и*

$$|Q_k f'(x^k)| \leqslant M |f'(x^k)| \quad \forall\, k = 0, 1, \ldots, \tag{16}$$

где $M > 0$ *— не зависящая от* k *константа*;

б) *имеет место предельное соотношение*

$$(Q_k - (f''(\overline{x}))^{-1}) f'(x^k) = o(|f'(x^k)|). \tag{17}$$

Доказательство. Предположим сначала, что выполнено а); тогда в силу (14) и равенства $\alpha_k = 1$ для достаточно большого k имеем

$$\begin{aligned}((f''(\overline{x}))^{-1} - Q_k) f'(x^k) &= (f''(\overline{x}))^{-1} f'(x^k) + x^{k+1} - x^k = \\ &= (f''(\overline{x}))^{-1}(f'(x^k) - f'(\overline{x}) - f''(\overline{x})(x^k - \overline{x})) + x^{k+1} - \overline{x} = \\ &= o(|x^k - \overline{x}|).\end{aligned}$$

С учетом того, что

$$|Q_k f'(x^k)| \geqslant |x^k - \overline{x}| - |x^{k+1} - \overline{x}| = |x^k - \overline{x}| + o(|x^k - \overline{x}|),$$

последнее соотношение влечет

$$((f''(\overline{x}))^{-1} - Q_k) f'(x^k) = o(|Q_k f'(x^k)|),$$

поскольку

$$\begin{aligned}\frac{o(|x^k - \overline{x}|)}{|Q_k f'(x^k)|} \leqslant \frac{o(|x^k - \overline{x}|)}{|x^k - \overline{x}| + o(|x^k - \overline{x}|)} = \\ = \frac{o(|x^k - \overline{x}|)/|x^k - \overline{x}|}{1 + o(|x^k - \overline{x}|)/|x^k - \overline{x}|} \to 0 \quad (k \to \infty).\end{aligned}$$

Остается заметить, что в силу (16) $o(|Q_k f'(x^k)|) = o(|f'(x^k)|)$.

Пусть теперь выполнено б). Выполнение (16) очевидно. Докажем, что $\alpha_k = 1$ при достаточно большом k. В силу классической теоремы о среднем найдется число $\widetilde{t}_k \in [0, 1]$ такое, что

$$\begin{aligned}f(x^k - Q_k f'(x^k)) = \\ = f(x^k) - \langle f'(x^k), Q_k f'(x^k)\rangle + \frac{1}{2}\langle f''(\widetilde{x}^k) Q_k f'(x^k), Q_k f'(x^k)\rangle,\end{aligned}$$

где $\widetilde{x}^k = x^k - \widetilde{t}_k Q_k f'(x^k)$. Достаточно показать, что

$$\begin{aligned}\langle f'(x^k), Q_k f'(x^k)\rangle - \frac{1}{2}\langle f''(\widetilde{x}^k) Q_k f'(x^k), Q_k f'(x^k)\rangle \geqslant \\ \geqslant \varepsilon \langle f'(x^k), Q_k f'(x^k)\rangle. \end{aligned} \tag{18}$$

Заметим, что $\{\widetilde{x}^k\} \to \overline{x} \quad (k \to \infty)$, поскольку $\{x^k\} \to \overline{x}$ и $\{Q_k f'(x^k)\} \to 0$ (последнее следует из (16)). Тогда согласно (17) неравенство (18) переписывается в виде

$$\begin{aligned}(1 - \varepsilon)\langle f'(x^k), (f''(\overline{x}))^{-1} f'(x^k)\rangle \geqslant \\ \geqslant \frac{1}{2}\langle f'(x^k), (f''(\overline{x}))^{-1} f'(x^k)\rangle + o(|f'(x^k)|^2),\end{aligned}$$

или

$$\left(\frac{1}{2} - \varepsilon\right)\langle (f''(\overline{x}))^{-1} f'(x^k), f'(x^k)\rangle \geqslant o(|f'(x^k)|^2),$$

что выполнено при больших k, поскольку $\varepsilon \in (0, 1/2)$, а положительная определенность $f''(\overline{x})$ влечет положительную определенность $(f''(\overline{x}))^{-1}$.

Докажем, наконец, что скорость сходимости последовательности $\{x^k\}$ является сверхлинейной. При достаточно больших k из соотношений (14), (17) и равенства $\alpha_k = 1$ имеем

$$|x^{k+1} - \overline{x}| = |x^k - \overline{x} - Q_k f'(x^k)| = |x^k - \overline{x} - (f''(\overline{x}))^{-1} f'(x^k)| +$$
$$+ o(|f'(x^k)|) \leqslant \|(f''(\overline{x}))^{-1}\| |f'(x^k) - f'(\overline{x}) - f''(\overline{x})(x^k - \overline{x})| +$$
$$+ o(|f'(x^k) - f'(\overline{x})|) = o(|x^k - \overline{x}|),$$

что и требовалось доказать. $\square$

Задача 6. Показать, что в условиях теоремы 2 соотношение (17) равносильно условию

$$(Q_k^{-1} - f''(\overline{x}))Q_k f'(x^k) = o(|Q_k f'(x^k)|). \tag{19}$$

Задача 7. Доказать аналог теоремы 2 для случая использования правила Голдстейна с начальным пробным значением параметра длины шага $\alpha = 1$.

Важнейший вывод, который следует из теоремы Дэнниса–Морэ, таков: для построения эффективных численных методов оптимизации *необходимо* использовать информацию «второго порядка»; в частности, методы, использующие лишь линейную модель исходной задачи, не могут быть эффективными. Заметим, что для самого метода Ньютона, как и для неточных методов Ньютона, при наличии сходимости к точке $\overline{x}$ условие (17) выполняется автоматически. Однако информацию «второго порядка» можно использовать и без непосредственного вычисления вторых производных. Основная идея квазиньютоновских методов состоит в том, чтобы полностью заменить вычисление $f''(x^k)$ и решение соответствующей линейной системы прямой аппроксимацией $(f''(\overline{x}))^{-1}$, причем, вообще говоря, не в смысле выполнения предельного соотношения $Q_k \to (f''(\overline{x}))^{-1}$ ($k \to \infty$), а в значительно более слабом смысле соотношения (17). Такие аппроксимации должны рекуррентно пересчитываться без использования информации о второй производной f и, к счастью, такое построение возможно осуществить, причем многими способами.

Для каждого k положим

$$r^k = x^{k+1} - x^k, \quad s^k = f'(x^{k+1}) - f'(x^k). \tag{20}$$

Заметим, что эти векторы уже известны к тому моменту, когда нужно вычислять Q_{k+1}, и стремление к выполнению (19) можно формализовать в виде равенства

$$Q_{k+1} s^k = r^k, \tag{21}$$

которое принято называть *квазиньютоновским уравнением*. Действительно, если считать, что $\alpha_k = 1$ (см. теорему 2), то в силу (14) и (20) $r^k = -Q_k f'(x^k)$, и согласно (19) нужно выбирать Q_{k+1} таким образом, чтобы вектор $Q_{k+1}^{-1} r^k$ аппроксимировал $f''(\overline{x}) r^k$. При этом

$$s^k = \int_0^1 f''(\theta x^{k+1} + (1-\theta) x^k) r^k \, d\theta,$$

и неявное предположение о том, что матрица $f''(\theta x^{k+1} + (1-\theta)x^k)$ в правой части последнего равенства аппроксимирует $f''(\overline{x})$ (что является автоматическим в случае сходимости траектории $\{x^k\}$ к $\overline{x}$), естественным образом приводит к мысли потребовать выполнения равенства $Q_{k+1}^{-1} r^k = s^k$, которое и дает (21).

Итак, имея симметрическую положительно определенную матрицу Q_k и векторы r^k и s^k, предлагается выбирать симметрическую положительно определенную матрицу Q_{k+1} удовлетворяющей квазиньютоновскому уравнению (21). Это, однако, можно сделать множеством способов. Естественным дополнительным требованием является «минимальность» (в некотором смысле) поправки $Q_{k+1} - Q_k$, что диктуется соображениями устойчивости: от итерации к итерации матрицы Q_k должны меняться как можно меньше. В зависимости от конкретного смысла, вкладываемого в «минимальность», получаются различные квазиньютоновские методы.

Исторически первым квазиньютоновским методом является *метод Давидона–Флетчера–Пауэлла* (ДФП), в котором Q_0 — произвольная положительно определенная симметрическая матрица (например, $Q_0 = E^n$), и для каждого k

$$Q_{k+1} = Q_k + \frac{r^k (r^k)^{\mathrm{T}}}{\langle r^k,\ s^k \rangle} - \frac{(Q_k s^k)(Q_k s^k)^{\mathrm{T}}}{\langle Q_k s^k,\ s^k \rangle}. \tag{22}$$

Заметим, что генерируемые по таким формулам матрицы остаются симметрическими и удовлетворяют квазиньютоновскому уравнению (21):

$$Q_{k+1} s^k = Q_k s^k + r^k \frac{\langle r^k,\ s^k \rangle}{\langle r^k,\ s^k \rangle} - Q_k s^k \frac{\langle Q_k s^k,\ s^k \rangle}{\langle Q_k s^k,\ s^k \rangle} =$$

$$= Q_k s^k + r^k - Q_k s^k = r^k.$$

Кроме того, поправка $Q_{k+1} - Q_k$ является матрицей, ранг которой не превосходит 2, поскольку $\ker(Q_{k+1} - Q_k)$ содержит все векторы, ортогональные r^k и $Q_k s^k$, и в этом смысле поправка «мала».

Что же касается положительной определенности матрицы Q_{k+1}, то она зависит не только от самой квазиньютоновской формулы, по которой эта матрица вычисляется, но и от способа выбора параметра длины шага в (14). Точнее, имеет место следующее

Предложение 1. *Пусть* $Q_k \in \mathbf{R}(n, n)$ — *симметрическая положительно определенная матрица. Пусть функция* $f: \mathbf{R}^n \to \mathbf{R}$ *дифференцируема в точках* x^k, $x^{k+1} \in \mathbf{R}^n$, *которые связаны формулой* (14) *с некоторым параметром длины шага* $\alpha_k > 0$.

Тогда формулы (20) *и* (22) *корректно определяют положительно определенную матрицу* Q_{k+1} *в том и только том случае, если выполняется неравенство*

$$\langle f'(x^{k+1}),\, d^k\rangle > \langle f'(x^k),\, d^k\rangle. \tag{23}$$

Доказательство. Необходимость немедленно следует из (14), (20) и квазиньютоновского уравнения (21), согласно которым

$$\langle d^k,\, f'(x^{k+1}) - f'(x^k)\rangle = \frac{1}{\alpha_k}\langle r^k,\, s^k\rangle = \langle Q_{k+1}s^k,\, s^k\rangle,$$

и из того факта, что при $s^k = 0$ формула (22) была бы некорректной.

Докажем достаточность. Согласно (14), (20) и (23)

$$\langle r^k,\, s^k\rangle = \alpha_k\langle d^k,\, f'(x^{k+1}) - f'(x^k)\rangle > 0. \tag{24}$$

В частности, $s^k \neq 0$ и $\langle Q_k s^k,\, s^k\rangle > 0$ в силу положительной определенности матрицы Q_k, и поэтому матрица Q_{k+1} корректно определена.

Для произвольного $\xi \in \mathbf{R}^n$ в силу (22) имеем

$$\langle Q_{k+1}\xi,\, \xi\rangle = \langle Q_k\xi,\, \xi\rangle + \frac{\langle r^k,\, \xi\rangle^2}{\langle r^k,\, s^k\rangle} - \frac{\langle Q_k s^k,\, \xi\rangle^2}{\langle Q_k s^k,\, s^k\rangle} =$$

$$= \frac{\langle r^k,\, \xi\rangle^2}{\langle r^k,\, s^k\rangle} + \frac{|Q_k^{1/2}\xi|^2|Q_k^{1/2}s^k|^2 - \langle Q_k^{1/2}\xi,\, Q_k^{1/2}s^k\rangle^2}{|Q_k^{1/2}s^k|^2} \geqslant 0,$$

поскольку оба слагаемых в правой части неотрицательны в силу (24) и неравенства Коши–Буняковского–Шварца. Более того, равенство $\langle Q_{k+1}\xi,\, \xi\rangle = 0$ может иметь место только в том случае, когда оба эти слагаемых нулевые, т. е.

$$\langle r^k,\, \xi\rangle = 0 \tag{25}$$

и

$$|Q_k^{1/2}\xi||Q_k^{1/2}s^k| = |\langle Q_k^{1/2}\xi,\, Q_k^{1/2}s^k\rangle|.$$

Последнее равенство означает, что $Q_k^{1/2}\xi = tQ_k^{1/2}s^k$ при некотором числе t, что в силу обратимости матрицы $Q_k^{1/2}$ приводит к равенству $\xi = ts^k$. Но тогда согласно (25) $t\langle r^k,\, s^k\rangle = \langle r^k,\, \xi\rangle = 0$, что в силу (24) возможно лишь при $t = 0$, т. е. при $\xi = 0$. $\square$

В частности, если параметр длины шага в (14) выбирается по правилу одномерной минимизации, то $\langle f'(x^{k+1}),\, d^k\rangle = 0$ (см. (3.1.5)), и, как следует из (15), неравенство (23) выполняется автоматически.

Что же касается более практических способов одномерного поиска, то неравенство (23) всегда выполняется, если параметр длины шага в (14) выбирается по правилу Вулфа, а вот, скажем, правило Армихо или правило Голдстейна этим свойством не обладают. Указанное обстоятельство является причиной того, что на практике в неквадратичном случае в квазиньютоновских методах наиболее часто используется именно правило Вулфа.

Самым эффективным из известных квазиньютоновских методов общего назначения в настоящее время считается *метод Бройдена–Флетчера–Голдфарба–Шэнно* (БФГШ), в котором для каждого k

$$Q_{k+1} = Q_k + \frac{(r^k - Q_k s^k)(r^k)^{\mathrm{T}} + r^k (r^k - Q_k s^k)^{\mathrm{T}}}{\langle r^k,\ s^k \rangle} - \\ - \frac{\langle r^k - Q_k s^k,\ s^k \rangle r^k (r^k)^{\mathrm{T}}}{\langle r^k,\ s^k \rangle^2}. \tag{26}$$

Элементарно проверяется, что, как и для метода ДФП, генерируемые по таким формулам матрицы остаются симметрическими и удовлетворяют квазиньютоновскому уравнению (21), а поправки $Q_{k+1} - Q_k$ являются матрицами ранга не выше 2.

Задача 8. Показать, что методы ДФП и БФГШ являются «взаимодвойственными» в следующем смысле. Для произвольной симметрической положительно определенной матрицы $Q_k \in \mathbf{R}(n, n)$ положим $H_k = Q_k^{-1}$. Пусть матрица Q_{k+1} сгенерирована по формуле (26), а матрица H_{k+1} — по формуле

$$H_{k+1} = H_k + \frac{s^k (s^k)^{\mathrm{T}}}{\langle r^k,\ s^k \rangle} - \frac{(H_k r^k)(H_k r^k)^{\mathrm{T}}}{\langle H_k r^k,\ r^k \rangle}$$

(ср. с (22)), причем матрица H_{k+1} невырождена. Тогда матрица Q_{k+1} невырождена, причем $H_{k+1} = Q_{k+1}^{-1}$. Вывести отсюда аналог предложения 8 для метода БФГШ.

Можно показать, что если f — квадратичная функция вида (13) с положительно определенной матрицей A, а α_k выбирается по правилу одномерной минимизации, то методы ДФП и БФГШ найдут единственную критическую точку функции f (по необходимости являющуюся глобальным решением задачи (9); см. задачу 1.1.11) из любого начального приближения за $\overline{k} \leqslant n$ шагов, причем $Q_{\overline{k}} = (f''(x^{\overline{k}}))^{-1} = A^{-1}/2$ [6, 41]. Напомним, что для квадратичных функций выбор параметра длины шага по правилу одномерной минимизации сводится к вычислению по явной формуле; см. п. 3.1.1. Применительно к задачам безусловной оптимизации квадратичных функций, квазиньютоновские методы оказываются тесно связанными с методами сопряженных направлений, рассматриваемыми в § 3.3.

Результаты о сходимости и скорости сходимости методов ДФП и БФГШ в неквадратичном случае можно найти, например, в [16, 32, 34, 42, 46, 50]. Этот анализ крайне нетривиален и сопряжен с преодолением очень серьезных технических трудностей, поэтому здесь приводятся лишь некоторые комментарии.

Известные результаты о глобальной сходимости методов ДФП и БФГШ относятся к случаю выпуклой функции f. Теория квазиньютоновских методов для невыпуклых задач не завершена, хотя вычислительная практика свидетельствует об их чрезвычайно высокой эффективности и в невыпуклом случае (в первую очередь этот относится к БФГШ).

При использовании неградиентных методов спуска вида (14) иногда рекомендуют время от времени делать градиентные шаги, полагая $Q_k = E^n$ на некоторых итерациях с бесконечно возрастающими номерами k. Оказывается, что такие «встроенные» градиентные шаги обеспечивают глобальную сходимость (в некотором смысле) неградиентного в целом метода [6]. Глобальная сходимость алгоритма 3 без встроенных шагов будет установлена в § 5.1, но лишь при некотором дополнительном предположении относительно последовательности матриц $\{Q_k\}$, а именно, в случае, когда эти матрицы равномерно положительно определены (см. (5.1.3)), что для методов ДФП и БФГШ не является автоматическим и должно проверяться в тех или иных предположениях.

В неквадратичном случае ожидать конечности квазиньютоновских методов, разумеется, не приходится, однако скорость сходимости обычно оказывается весьма высокой. Доказательство сверхлинейной скорости сходимости для конкретных квазиньютоновских методов сводится к (обычно весьма нетривиальной) проверке выполнения условия (17) и применению теоремы Дэнниса–Морэ.

Квазиньютоновские методы чрезвычайно популярны среди пользователей методов оптимизации в связи с тем, что они сочетают высокую скорость сходимости с невысокой трудоемкостью итерации. Практическое значение этих методов трудно переоценить. С теоретической точки зрения любой квазиньютоновский метод есть не более чем умная реализация фундаментальной идеи метода Ньютона, однако, это тот случай, когда реализация имеет не меньшее значение, чем реализуемая идея.

Об общих схемах построения и исследования квазиньютоновских методов см. [13, 16, 32, 34, 41, 46, 50].

§ 3.3. Методы сопряженных направлений

Здесь рассматривается еще один важный класс методов, которые вводятся не как квазиньютоновские и основаны на других принципах,

однако обладают более высокой скоростью сходимости, чем градиентные методы, причем повышение скорости сходимости не связано с существенным повышением трудоемкости итерации. В частности, практически значимые реализации методов этого класса являются методами первого порядка.

3.3.1. Методы сопряженных направлений для квадратичных функций. Изначально методы сопряженных направлений вводятся для задачи безусловной минимизации

$$f(x) \to \min, \quad x \in \mathbf{R}^n, \tag{1}$$

с квадратичной целевой функцией

$$f\colon \mathbf{R}^n \to \mathbf{R}, \quad f(x) = \langle Ax,\, x\rangle + \langle b,\, x\rangle, \tag{2}$$

где $A \in \mathbf{R}(n,\, n)$ — положительно определенная симметрическая матрица, $b \in \mathbf{R}^n$. В основе методов этого класса лежит следующее понятие.

Определение 1. Система векторов $d^0,\, d^1, \ldots, d^k \in \mathbf{R}^n \setminus \{0\}$ называется *A-сопряженной*, если

$$\langle Ad^i,\, d^j\rangle = 0 \quad \forall\, i,\, j = 0,\, 1, \ldots, k, \quad i \neq j.$$

Ясно, что это понятие обобщает понятие ортогональной системы векторов (последнее отвечает случаю $A = E^n$).

Лемма 1. *Для произвольной положительно определенной симметрической матрицы $A \in \mathbf{R}(n, n)$ любая A-сопряженная система векторов линейно независима.*

Доказательство. Предположим, что один из векторов A-сопряженной системы $d^0,\, d^1, \ldots, d^k$ линейно выражается через другие, например,

$$d^0 = \sum_{i=1}^{k} \beta_i d^i$$

при некоторых числах β_i, $i = 1, \ldots, k$. Умножая скалярно обе части этого равенства на Ad^0, получим

$$0 < \langle Ad^0,\, d^0\rangle = \sum_{i=1}^{k} \beta_i \langle Ad^0,\, d^i\rangle = 0,$$

что невозможно. □

Из доказанной леммы, в частности, следует, что число векторов в A-сопряженной системе в $\mathbf{R}^n$ не может превышать n.

Общая схема *методов сопряженных направлений* для функции вида (2) выглядит так:

$$x^{k+1} = x^k + \alpha_k d^k, \quad k = 0,\, 1, \ldots, n-1, \tag{3}$$

где $d^0, d^1, \ldots, d^{n-1}$ — генерируемая тем или иным способом A-сопряженная система векторов, а числовые параметры длины шага α_k выбираются из условия

$$f(x^k + \alpha_k d^k) = \min_{\alpha \in \mathbf{R}} f(x^k + \alpha d^k), \tag{4}$$

т. е.

$$\alpha_k = -\frac{\langle 2Ax^k + b,\, d^k \rangle}{2\langle Ad^k,\, d^k \rangle}. \tag{5}$$

Минимум в (4) берется по всем $\alpha \in \mathbf{R}$, а не только по $\alpha \geqslant 0$, поскольку d^k вовсе не обязательно является направлением убывания функции f в точке x^k. Более того, векторы d^k и $-d^k$ могут оба не быть направлениями убывания, т. е. методы сопряженных направлений не являются, вообще говоря, методами спуска; в их основе лежит другой принцип.

Разумеется, суть любого метода сопряженных направлений составляет конкретный способ построения A-сопряженной системы. Тем не менее каким бы ни был этот способ, имеет место следующий фундаментальный факт: любой метод сопряженных направлений позволяет найти решение задачи безусловной минимизации квадратичной функции вида (2) с положительно определенной матрицей A не более чем за n шагов из любого начального приближения.

Теорема 1. *Пусть $A \in \mathbf{R}(n, n)$ — положительно определенная симметрическая матрица, $b \in \mathbf{R}^n$, функция $f\colon \mathbf{R}^n \to \mathbf{R}$ имеет вид* (2).

Тогда, каким бы ни было начальное приближение $x^0 \in \mathbf{R}^n$, точка x^n, получаемая по схеме (3), (4) *при любом выборе A-сопряженной системы векторов $d^0, d^1, \ldots, d^{n-1}$, является (глобальным) решением задачи* (1).

Доказательство. В силу положительной определенности матрицы A задача (1) имеет единственную стационарную точку $\overline{x} \in \mathbf{R}^n$, характеризующуюся равенством

$$f'(\overline{x}) = 2A\overline{x} + b = 0, \tag{6}$$

причем эта стационарная точка по необходимости является глобальным решением (см. задачу 1.1.11). В силу леммы 1 система векторов $d^0, d^1, \ldots, d^{n-1}$ является базисом в $\mathbf{R}^n$, поэтому вектор $\overline{x} - x^0$ можно представить в виде

$$\overline{x} - x^0 = \sum_{k=0}^{n-1} \beta_k d^k \tag{7}$$

при некоторых числах β_k, $k = 0, 1, \ldots, n-1$. С другой стороны, согласно (3) имеем

$$x^n = x^0 + \sum_{k=0}^{n-1} \alpha_k d^k,$$

где числа α_k определены в (5). Остается показать, что $\beta_k = \alpha_k$ $\forall\, k = 0, 1, \ldots, n-1$.

Из равенства (7) и A-сопряженности векторов $d^0, d^1, \ldots, d^{n-1}$ получаем $\forall\, k = 0, 1, \ldots, n-1$

$$\langle A\overline{x}, d^k\rangle = \langle Ax^0, d^k\rangle + \sum_{i=0}^{n-1} \beta_i \langle Ad^i, d^k\rangle = \langle Ax^0, d^k\rangle + \beta_k \langle Ad^k, d^k\rangle,$$

откуда следует, что

$$\beta_k = \frac{\langle A\overline{x} - Ax^0, d^k\rangle}{\langle Ad^k, d^k\rangle} = -\frac{\langle 2Ax^0 + b, d^k\rangle}{2\langle Ad^k, d^k\rangle}, \tag{8}$$

где также принято во внимание равенство (6). Вновь привлекая (3) и A-сопряженность системы $d^0, d^1, \ldots, d^{n-1}$, имеем

$$\langle Ax^k, d^k\rangle = \langle A(x^{k-1} + \alpha_{k-1} d^{k-1}), d^k\rangle =$$
$$= \langle Ax^{k-1}, d^k\rangle = \ldots = \langle Ax^0, d^k\rangle,$$

что в совокупности с (5) и (8) и дает требуемое утверждение. □

Задача 1. Доказать, что в условиях теоремы 1, каким бы ни было начальное приближение $x^0 \in \mathbf{R}^n$, для произвольного $k = 0, 1, \ldots, n-1$ точка x^{k+1}, получаемая по схеме (3), (4) при любом выборе A-сопряженной системы векторов $d^0, d^1, \ldots, d^k$, является (глобальным) решением задачи

$$f(x) \to \min, \quad x \in X_k,$$

где $X_k = x^0 + \operatorname{span}\{d^0, d^1, \ldots, d^k\}$. В частности,

$$\langle f'(x^{k+1}), d^i\rangle = 0 \quad \forall\, i = 0, 1, \ldots, k.$$

Оказывается, один из возможных способов построения A-сопряженной системы уже был упомянут выше: можно показать, что векторы $d^k = -Q_k f'(x^k)$, $k = 0, 1, \ldots, n-1$, где матрицы Q_k генерируются в соответствии с методом ДФП или с БФГШ (см. п. 3.2.3), будут обладать нужным свойством [6, 41]. Специфика методов ДФП и БФГШ состоит в том, что они позволяют одновременно с решением рассматриваемой задачи оптимизации находить A^{-1}, что может быть весьма полезно. С другой стороны, реализация методов ДФП и БФГШ требует хранения в памяти и пересчета на каждой итерации матрицы $Q_k \in \mathbf{R}(n, n)$, что при больших n может быть дорого. Более экономичная реализация метода сопряженных направлений обсуждается ниже.

3.3.2. Метод сопряженных градиентов. Важнейшим из методов сопряженных направлений является *метод сопряженных градиентов*, в котором

$$d^0 = -f'(x^0), \quad d^k = -f'(x^k) + \beta_{k-1} d^{k-1}, \quad k = 1, \ldots, n-1, \qquad (9)$$

а числа β_{k-1} выбираются из условия A-сопряженности системы «соседних» векторов d^{k-1}, d^k:

$$0 = \langle Ad^{k-1}, d^k \rangle = -\langle Ad^{k-1}, f'(x^k) \rangle + \beta_{k-1} \langle Ad^{k-1}, d^{k-1} \rangle,$$

т. е.

$$\beta_{k-1} = \frac{\langle Ad^{k-1}, f'(x^k) \rangle}{\langle Ad^{k-1}, d^{k-1} \rangle}. \qquad (10)$$

Алгоритм 1. Выбираем $x^0 \in \mathbf{R}^n$ и полагаем $k = 0$.

1. Если $f'(x^k) = 0$, то останавливаемся.
2. Если $k \geqslant 1$, то вычисляем β_{k-1} по формуле (10).
3. Вычисляем d^k по формуле (9). Вычисляем α_k по формуле (5).
4. Вычисляем x^{k+1} по формуле (3).
5. Если $k < n-1$, то увеличиваем номер шага k на 1 и переходим к п. 1.

Формула (5) для α_k подразумевает, что $d^k \neq 0$ (а формула (10) для β_{k-1} — что $d^{k-1} \neq 0$). Если $d^0 = -f'(x^0) = 0$, то алгоритм остановился бы еще до этапа определения d^0 и α_0 (и тем более β_0). Предположим, что для какого-то $k \geqslant 1$ впервые реализовалось равенство $d^k = 0$ (в частности, $d^{k-1} \neq 0$). Тогда в силу (9)

$$0 = \langle f'(x^k), d^k \rangle = -|f'(x^k)|^2 + \beta_{k-1} \langle f'(x^k), d^{k-1} \rangle = -|f'(x^k)|^2,$$

где последнее равенство следует из того, что согласно определению величины α_{k-1}

$$\langle f'(x^k), d^{k-1} \rangle = 0 \qquad (11)$$

(см. (3.1.5)). Таким образом, $f'(x^k) = 0$, и алгоритм остановился бы еще до этапа определения d^k и α_k (и тем более β_k).

Следующая теорема показывает, что предложенный алгоритм действительно является методом сопряженных направлений.

Теорема 2. *Пусть выполнены условия теоремы* 1.

Тогда, каким бы ни было начальное приближение $x^0 \in \mathbf{R}^n$, *если для произвольного* $k = 1, \ldots, n$ *алгоритм* 1 *сгенерировал векторы* $d^0, d^1, \ldots, d^{k-1}$, *то они образуют* A*-сопряженную систему, а градиенты* $f'(x^0), f'(x^1), \ldots, f'(x^{k-1})$ — *ортогональную систему в* $\mathbf{R}^n$.

Доказательство. Сразу заметим, что в силу сказанного выше, если алгоритм сгенерировал векторы $d^0, d^1, \ldots, d^{k-1}$, то все они отличны от нуля, как и градиенты $f'(x^0), f'(x^1), \ldots, f'(x^{k-1})$.

Доказательство проведем индукцией по k. При $k = 1$ утверждение теоремы очевидно, поскольку система d^0, d^1 является A-сопряженной по построению, а векторы $f'(x^0) = d^0$ и $f'(x^1)$ ортогональны в силу (11).

Пусть утверждение теоремы верно при $k = s \geqslant 2$; нужно доказать, что

$$\langle Ad^s, d^i \rangle = 0, \quad \langle f'(x^s), f'(x^i) \rangle = 0 \quad \forall\, i = 0, 1, \ldots, s-1. \tag{12}$$

В силу (9), предположения индукции и утверждения из задачи 1

$$\langle f'(x^s), f'(x^i) \rangle = \langle f'(x^s), -d^i + \beta_{i-1} d^{i-1} \rangle = 0 \quad \forall\, i = 0, 1, \ldots, s-1,$$

что доказывает второе равенство в (12).

Далее, при $i = s-1$ первое равенство в (12) верно по построению. При произвольном $i = 0, 1, \ldots, s-2$ в силу (9) и предположения индукции имеем

$$\langle Ad^s, d^i \rangle = -\langle Af'(x^s), d^i \rangle + \beta_{s-1} \langle Ad^{s-1}, d^i \rangle = -\langle Af'(x^s), d^i \rangle. \tag{13}$$

Но согласно (3)

$$f'(x^{i+1}) - f'(x^i) = 2A(x^{i+1} - x^i) = 2\alpha_i Ad^i,$$

поэтому

$$\langle Af'(x^s), d^i \rangle = \frac{1}{2\alpha_i} \langle f'(x^s), f'(x^{i+1}) - f'(x^i) \rangle = 0$$

в силу уже доказанного второго равенства в (12). Вместе с (13) это завершает доказательство первого равенства в (12). □

Согласно доказанной теореме и теореме 1 метод сопряженных градиентов находит решение задачи безусловной минимизации квадратичной функции вида (2) с положительно определенной матрицей A из любого начального приближения не более чем за n шагов.

При попытке применения метода сопряженных градиентов в неквадратичном случае первая проблема состоит в том, что в формулу (10) для β_{k-1} явно входит матрица A. К счастью, эта проблема чисто техническая. Например, формулу (10) можно переписать в виде

$$\beta_{k-1} = \frac{\langle f''(x^k) d^{k-1}, f'(x^k) \rangle}{\langle f''(x^k) d^{k-1}, d^{k-1} \rangle},$$

однако вычисления $f''(x^k)$ желательно избежать.

Задача 2. Показать, что формула (10) может быть переписана в виде

$$\beta_{k-1} = \frac{\langle f'(x^k),\, f'(x^k) - f'(x^{k-1})\rangle}{|f'(x^{k-1})|^2},$$

или

$$\beta_{k-1} = \frac{|f'(x^k)|^2}{|f'(x^{k-1})|^2}.$$

В неквадратичном случае приведенные формулы для β_{k-1} дают, вообще говоря, различные значения, а генерируемые векторы d^k не образуют A-сопряженную систему относительно какой-либо матрицы A. Ожидать конечности метода сопряженных градиентов при этом, разумеется, не приходится, однако можно ожидать, что скорость сходимости (при наличии сходимости) останется высокой, поскольку локально гладкая функция хорошо аппроксимируется квадратичной частью своего ряда Тейлора.

Вычислительная практика свидетельствует о том, что предпочтительно пользоваться первой формулой из задачи 2; объяснение этому см. в [6, 41, 50]. В литературе можно найти и другие формулы для β_{k-1}.

Напомним также, что в силу (9) и (11)

$$\langle f'(x^k),\, d^k\rangle = -|f'(x^k)|^2 < 0,$$

если $f'(x^k) \neq 0$, и в этом случае $d^k \in \mathcal{D}_f(x^k)$ согласно лемме 3.1.1. Поэтому можно брать минимум в (4) только по $\alpha \geqslant 0$ и считать, что $\alpha_k > 0$; с этими оговорками метод сопряженных градиентов является методом спуска. В неквадратичном случае точная одномерная минимизация обычно невозможна, и ее часто заменяют менее трудоемкими правилами одномерного поиска, например правилом Вулфа [50]. Для улучшения глобального поведения метода иногда рекомендуют использовать «встроенные» градиентные шаги (см. п. 3.2.3).

Результаты о сходимости и скорости сходимости метода сопряженных градиентов в неквадратичном случае имеются в [10, 24, 32, 34, 50]. Интересно отметить, что известные результаты о сходимости касаются второй, а не первой формулы для β_{k-1} в задаче 2. Таким образом, здесь имеет место следующая весьма нередкая в численной оптимизации ситуация, свидетельствующая о том, что переоценивать (как, впрочем, и недооценивать) роль теории в этой области деятельности не следует: теоретически обоснованным является один способ выбора параметров метода, в то время как практика свидетельствует о преимуществах другого способа, не обоснованного теоретически.

В силу указанных выше причин, квазиньютоновские методы и метод сопряженных градиентов являются наиболее часто используемыми на практике алгоритмами решения гладких задач безусловной оптимизации. Детальный сравнительный анализ этих методов можно

найти в [6, 34, 41, 42]. В настоящее время в неквадратичном случае преимущество отдают квазиньютоновским методам, а метод сопряженных градиентов, весьма популярный в прошлом, все чаще характеризуют как устаревший. В некотором роде, метод сопряженных градиентов является квазиньютоновским методом «для бедных». А именно, как следует из утверждения, приводимого в следующей задаче, в случае использования первой формулы из задачи 2 и точной одномерной минимизации шаг метода сопряженных градиентов можно интерпретировать как шаг метода БФГШ, в котором в правой части формулы (3.2.23) вместо матрицы Q_k, вычисленной на предыдущем шаге, используется матрица E^n. Иными словами, такой метод сопряженных градиентов — это суррогатный метод БФГШ с реинициализацией $Q_k(= E^n)$ на каждом шаге.

Задача 3. Предположим, что для точек x^k, $x^{k+1} \in \mathbf{R}^n$ и векторов d^{k-1}, $d^k \in \mathbf{R}^n$, вычисленных для некоторого $k \geqslant 1$ методом сопряженных градиентов, справедливы равенства $\langle f'(x^k), d^{k-1}\rangle = 0$ и $\langle f'(x^{k+1}), d^k\rangle = 0$. Показать, что направление $d^{k+1} \in \mathbf{R}^n$, вычисляемое согласно (9) и второй формуле в задаче 2, можно представить в виде $d^{k+1} = -Q_{k+1}f'(x^{k+1})$, где матрица Q_{k+1} вычисляется по формуле (3.2.23), в правой части которой вместо Q_k стоит E^n.

Задача 4. Решить методом сопряженных градиентов задачу безусловной минимизации функции

$$f\colon \mathbf{R}^2 \to \mathbf{R}, \quad f(x) = x_1^2 + 2x_2^2,$$

используя начальное приближение $x^0 = (1, 1)$.

§ 3.4. Методы нулевого порядка

Этот короткий параграф не содержит никаких теорем, поскольку его материал лежит несколько в стороне от основного круга вопросов, обсуждаемых в данном курсе. С другой стороны, практическая значимость методов нулевого порядка не позволяет опустить этот материал полностью. Компромиссным решением является краткий обзор, который и приводится ниже.

Пусть $f\colon \mathbf{R}^n \to \mathbf{R}$ — заданная функция. Реализация методов решения задачи безусловной минимизации

$$f(x) \to \min, \quad x \in \mathbf{R}^n, \tag{1}$$

о которых шла речь выше, требует вычисления первых или даже вторых производных целевой функции задачи. Однако на практике такие вычисления часто оказываются слишком трудоемкими либо их принципиально невозможно осуществить точно по причине отсутствия явных выражений не только для производных, но и для самой функции. Иногда самое большее, на что можно рассчитывать, это умение вычислять значение целевой функции задачи в заданной

точке (например, так обстоит дело, если значение получается в результате эксперимента, параметры которого определяются аргументом целевой функции). В таких ситуациях речь может идти либо об аппроксимации производных по информации о значениях функции, либо о построении совершенно иных методов оптимизации, не связанных с вычислением производных (такие методы иногда называют *методами прямого поиска*).

Принципиально иные трудности связаны с тем, что целевая функция может просто не иметь производных, и тогда обсуждавшиеся выше методы также неприменимы, не говоря уже об их теоретическом обосновании. Один из подходов к решению негладких задач обсуждается в конце этого параграфа. Подчеркнем, что данная проблематика связана с преодолением принципиальных трудностей, особенно когда негладкость не сопровождается наличием какой-либо дополнительной структуры, такой, как выпуклость. Методы негладкой выпуклой оптимизации значительно более развиты; им посвящена гл. 6.

Каждому из рассматривавшихся выше методов первого и второго порядков можно поставить в соответствие целый ряд методов нулевого порядка, если заменить производные их конечно-разностными аппроксимациями. Например, для приближенного вычисления градиента функции f в точке $x^k \in \mathbf{R}^n$ можно использовать *разделенные разности вперед*:

$$\frac{\partial f}{\partial x_j}(x^k) \approx g_j^k = \frac{f(x^k + t_k e^j) - f(x^k)}{t_k}, \quad j = 1, \ldots, n,$$

либо *центральные разделенные разности*:

$$\frac{\partial f}{\partial x_j}(x^k) \approx g_j^k = \frac{f(x^k + t_k e^j) - f(x^k - t_k e^j)}{2t_k}, \quad j = 1, \ldots, n,$$

где $t_k > 0$ — малое число, а $e^1, \ldots, e^n$ — стандартный базис в $\mathbf{R}^n$. Аппроксимация центральными разностями существенно точнее, однако она требует почти вдвое больше вычислений значений функции f.

Например, *конечно разностный аналог градиентного метода* задается схемой

$$x^{k+1} = x^k - \alpha_k g^k, \quad k = 0, 1, \ldots, \tag{2}$$

где векторы $g^k \in \mathbf{R}^n$ вычисляются одним из указанных выше способов, а параметры длины шага α_k определяют в соответствии с одним из правил одномерного поиска по направлению $-g^k$ (см. п. 3.1.1). Интересно отметить, что обоснование разумного поведения такого метода вряд ли возможно без предположений о гладкости функции f, хотя сам метод не использует ее производных.

Пример 1. Для функции

$$f\colon \mathbf{R}^2 \to \mathbf{R}, \quad f(x) = \max\{|x_1|, |x_2|\},$$

в точке $x^k = (-1, -1)$ вектор g^k, вычисленный с помощью разделенных разностей вперед, будет нулевым, и соответствующий конечно разностный аналог градиентного метода «застрянет» в точке x^k, тогда как единственным безусловным локальным минимумом этой функции является ее глобальный минимум 0.

Впрочем, это не удивительно: обсуждаемый метод основан на аппроксимации градиентного метода, поэтому обоснования его разумных свойств естественно ожидать лишь при корректности самого объекта аппроксимации.

Наиболее очевидным методом, не связанным с (точным или приближенным) вычислением производных, является *метод покоординатного спуска*, задаваемый схемой

$$x^{k+1} = x^k + \alpha_k d^k, \quad k = 0, 1, \ldots, \tag{3}$$

где

$$d^0 = e^1, \quad \ldots, \quad d^{n-1} = e^n, \quad d^n = e^1, \quad \ldots, \quad d^{2n-1} = e^n, \quad \ldots.$$

Числовые параметры длины шага α_k могут выбираться с помощью одномерной минимизации, т. е. из условия

$$f(x^k + \alpha_k d^k) = \min_{\alpha \in \mathbf{R}} f(x^k + \alpha d^k);$$

минимум берется по всем $\alpha \in \mathbf{R}$, поскольку d^k может не быть направлением убывания функции f в точке x^k. Тогда итерация метода состоит в решении одномерной задачи минимизации функции f по одной координате при фиксированных остальных; координаты, по которым производится минимизация, циклически меняются.

В методе покоординатного спуска могут использоваться и другие способы выбора параметров длины шага. Например, перед началом процесса фиксируют числа $\widehat{\alpha} > 0$ и $\theta \in (0, 1)$ и полагают $\alpha = \widehat{\alpha}$. На каждом шаге проверяют выполнение неравенства

$$f(x^k + \alpha d^k) < f(x^k);$$

если оно выполнено, то полагают $\alpha_k = \alpha$. В противном случае проверяют выполнение неравенства

$$f(x^k - \alpha d^k) < f(x^k);$$

если оно выполнено, то полагают $\alpha_k = -\alpha$. Если и это неравенство не выполнено, то либо полагают $\alpha_k = 0$ (это имеет смысл делать не более чем на n последовательных шагах), либо заменяют α на $\theta\alpha$ и повторяют всю процедуру для данного шага.

Теоремы о сходимости различных вариантов метода покоординатного спуска для выпуклых функций можно найти в [4, 10, 24]. Интересно, что и здесь, несмотря на то, что метод никак не связан с вычислением производных, гарантировать его «разумное» поведение

можно лишь для гладких функций; см. пример 1 выше. Кроме того, без требования выпуклости доказать глобальную сходимость метода покоординатного спуска не удается: предельные точки траектории метода могут не быть стационарными в задаче (1); см. [50].

Метод покоординатного спуска легко распространить на задачу условной минимизации, допустимое множество которой является параллелепипедом.

В *методе случайного покоординатного спуска* в качестве d^k в (3) берется e^j, где j есть реализация случайной величины, принимающей значения $1, \ldots, n$ с равными вероятностями. В *методе случайного поиска* в качестве d^k берется реализация случайного вектора, равномерно распределенного на единичной сфере в $\mathbf{R}^n$. Параметры длины шага выбираются с помощью тех или иных процедур одномерного поиска, обеспечивающих по крайней мере монотонное невозрастание последовательности $\{f(x^k)\}$. Для методов такого типа характерны утверждения о сходимости с вероятностью 1; см. [24].

Наконец, обратимся к случаю, когда функция f не предполагается гладкой. Чрезвычайно важным (и даже решающим) при работе с невыпуклыми негладкими задачами является следующее обстоятельство: если функция непрерывна по Липшицу в некоторой окрестности каждой точки в $\mathbf{R}^n$, то она дифференцируема почти всюду на $\mathbf{R}^n$ в смысле меры Лебега (см. п. 4.5.2). Поэтому можно ожидать, что функция f будет дифференцируема в точке $\widetilde{x}^k$, случайно выбранной, например, в кубе со стороной $\delta_k > 0$ с центром в текущем приближении x^k. Это соображение приводит к *стохастическому конечно разностному аналогу градиентного метода*, задаваемому схемой (2), в которой

$$g_j^k = \frac{f(\widetilde{x}^k + t_k e^j) - f(\widetilde{x}^k)}{t_k}, \quad j = 1, \ldots, n,$$

где $t_k > 0$. При определенном согласовании параметров α_k, δ_k и t_k можно установить сходимость такого метода к множеству стационарных точек задачи (1) с вероятностью 1; см. [28] (правда, в силу отсутствия требований гладкости это предполагает уточнение самого понятия стационарной точки).

Задача 1. Из начального приближения $x^0 = 0$ сделать для задачи безусловной минимизации функции

$$f\colon \mathbf{R}^2 \to \mathbf{R}, \quad f(x) = x_1^2 + 2x_1x_2 + 3x_2^2 + 4x_1,$$

два шага метода покоординатного спуска с выбором парметров длины шага с помощью одномерной минимизации.

Задача 2. То же задание для функции

$$f\colon \mathbf{R}^2 \to \mathbf{R}, \quad f(x) = 2x_1^2 - x_1x_2 + x_2^2,$$

и начального приближения $x^0 = (2, 0)$.

Глава 4
МЕТОДЫ УСЛОВНОЙ ОПТИМИЗАЦИИ

Эта глава занимает в курсе центральное место. Материал остальных глав в значительной степени является подготовительным, вспомогательным либо дополнительным по отношению к материалу этой главы.

§ 4.1. Методы решения задач с простыми ограничениями

Изложение методов условной оптимизации начнем с рассмотрения задачи с прямым ограничением

$$f(x) \to \min, \quad x \in P, \tag{1}$$

где $P \subset \mathbf{R}^n$ — множество «простой структуры» (во всяком случае замкнутое и выпуклое), а $f\colon \mathbf{R}^n \to \mathbf{R}$ — гладкая функция. Обычно такие задачи возникают как вспомогательные при реализации методов решения задач с более сложными ограничениями. Напомним, что в сделанных предположениях для всякой точки $x \in \mathbf{R}^n$ существует единственная проекция $\pi_P(x)$ на P (см. следствие 1.1.2 и задачу 1.1.6). Напомним также, что точка $\overline{x} \in P$ является стационарной в задаче (1) в смысле определения 1.2.2 в том и только том случае, когда

$$\langle f'(\overline{x}),\, x - \overline{x}\rangle \geqslant 0 \quad \forall\, x \in P, \tag{2}$$

или, что то же самое,

$$\pi_P(\overline{x} - t f'(\overline{x})) = \overline{x} \tag{3}$$

для некоторого $t > 0$, причем выполнение этого условия для какого-то $t > 0$ влечет его выполнение для любого $t > 0$ (см. предложение 1.2.1, следствие 1.2.1 и задачу 1.2.4).

4.1.1. Методы проекции градиента. *Методы проекции градиента* могут рассматриваться как результат соединения идеи градиентных методов безусловной оптимизации с проектированием генерируемых приближений на допустимое множество условной задачи. В результате приходим к схеме

$$x^{k+1} = \pi_P(x^k - \alpha_k f'(x^k)), \quad k = 0, 1, \ldots, \tag{4}$$

где $x^0 \in P$, а для выбора параметров длины шага $\alpha_k > 0$ используются процедуры одномерного поиска, являющиеся обобщением соответствующих процедур для методов спуска из п. 3.1.1. Одномерный поиск выполняется для функции

$$\varphi_k\colon \mathbf{R}_+ \to \mathbf{R}, \quad \varphi_k(\alpha) = f(x^k(\alpha)),$$

где

$$x^k(\alpha) = \pi_P(x^k - \alpha f'(x^k)).$$

Например, может использоваться *правило одномерной минимизации*, когда α_k ищется как решение одномерной задачи

$$\varphi_k(\alpha) \to \min, \quad \alpha \in \mathbf{R}_+,$$

либо

$$\varphi_k(\alpha) \to \min, \quad \alpha \in [0, \widehat{\alpha}],$$

где $\widehat{\alpha} > 0$ — параметр. Однако ввиду высокой трудоемкости этого правила его редко применяют на практике.

Значительно чаще применяется *правило Армихо*, которое реализуется по той же схеме, что и в п. 3.1.1, только вместо (3.1.6) использует неравенство

$$f(x^k(\alpha)) \leqslant f(x^k) + \varepsilon\langle f'(x^k),\, x^k(\alpha) - x^k\rangle. \tag{5}$$

Напомним, что перед началом процесса должны быть выбраны параметр $\varepsilon \in (0, 1)$, начальное значение для дробления $\widehat{\alpha} > 0$ и параметр дробления $\theta \in (0, 1)$. Применяют также различные модификации правила Армихо (например, аналоги правила Голдстейна) и *правило постоянного параметра*, когда полагают $\alpha_k = \overline{\alpha}$ при фиксированном (не зависящем от k) числе $\overline{\alpha} > 0$.

Алгоритм 1. Выбираем $x^0 \in P$ и полагаем $k = 0$. Выбираем одно из трех правил одномерного поиска и необходимые для реализации этого правила параметры: $\widehat{\alpha} > 0$ (либо $\widehat{\alpha} = +\infty$) в случае правила одномерной минимизации, $\widehat{\alpha} > 0$, ε, $\theta \in (0, 1)$ в случае правила Армихо и $\overline{\alpha} > 0$ в случае правила постоянного параметра.

1. Вычисляем α_k в соответствии с выбранным правилом одномерного поиска.
2. Вычисляем x^{k+1} по формуле (4).
3. Увеличиваем номер шага k на 1 и переходим к п. 1.

Заметим, что если для некоторого k точка x^k оказывается стационарной в задаче (1), то $x^k = x^{k+1} = \ldots$, как бы ни выбирался параметр длины шага α_k. Заметим также, что в случае, когда $P = \mathbf{R}^n$, методы проекции градиента с указанными выше способами выбора α_k превращаются в соответствующие градиентные методы.

Свойства сходимости методов проекции градиента во многом аналогичны свойствам градиентных методов. Ограничимся двумя результатами, обобщающими теоремы 3.1.1 и 3.1.4 соответственно.

Важную роль в предлагаемом анализе будет играть следующее неравенство:

$$\langle f'(x^k),\, x^k(\alpha) - x^k\rangle \leqslant -\frac{1}{\alpha}\,|x^k(\alpha) - x^k|^2 \quad \forall\,\alpha > 0. \tag{6}$$

Оно получается прямой выкладкой: $\forall\,\alpha > 0$ имеем

$$\langle f'(x^k),\, x^k(\alpha) - x^k\rangle = \frac{1}{\alpha}\,\langle x^k + \alpha f'(x^k) - x^k,\, x^k(\alpha) - x^k\rangle =$$

$$= \frac{1}{\alpha}\,\langle x^k - x^k(\alpha),\, x^k(\alpha) - x^k\rangle + \frac{1}{\alpha}\,\langle x^k(\alpha) - (x^k - \alpha f'(x^k)),\, x^k(\alpha) - x^k\rangle \leqslant$$

$$\leqslant -\frac{1}{\alpha}\,|x^k(\alpha) - x^k|^2,$$

где использован результат задачи 1.1.7.

Лемма 1. *Пусть множество $P \subset \mathbf{R}^n$ замкнуто и выпукло, функция $f\colon \mathbf{R}^n \to \mathbf{R}$ дифференцируема на $\mathbf{R}^n$, а ее производная непрерывна по Липшицу на $\mathbf{R}^n$ с константой $L > 0$.*

Тогда для произвольного $x^k \in \mathbf{R}^n$ неравенство (5) имеет место для любого $\alpha \in (0, \overline{\alpha}]$, где

$$\overline{\alpha} = \frac{2(1-\varepsilon)}{L} > 0. \tag{7}$$

Доказательство. Из леммы 3.1.4, неравенства (6) и формулы (7) для любого $\alpha \in (0, \overline{\alpha}]$ получаем

$$f(x^k(\alpha)) - f(x^k) \leqslant \langle f'(x^k),\, x^k(\alpha) - x^k\rangle + \frac{L}{2}\,|x^k(\alpha) - x^k|^2 \leqslant$$

$$\leqslant \varepsilon\langle f'(x^k),\, x^k(\alpha) - x^k\rangle + \left(\frac{L}{2} - \frac{1-\varepsilon}{\alpha}\right)|x^k(\alpha) - x^k|^2 \leqslant$$

$$\leqslant \varepsilon\langle f'(x^k),\, x^k(\alpha) - x^k\rangle. \quad \square$$

Из доказанной леммы следует, что в ее условиях, которые предполагаются выполненными в обеих приводимых ниже теоремах, количество дроблений в правиле Армихо будет конечным равномерно по k, т. е.

$$\alpha_k \geqslant \check{\alpha} > 0, \tag{8}$$

где $\check{\alpha}$ — не зависящая от k константа. Кроме того, использование правила постоянного параметра при

$$\overline{\alpha} < \frac{2}{L} \tag{9}$$

равносильно использованию правила Армихо при $\widehat{\alpha} = \overline{\alpha}$ и достаточно малом $\varepsilon > 0$ и поэтому не требует отдельного рассмотрения.

Теорема 1. *Пусть множество* $P \subset \mathbf{R}^n$ *замкнуто и выпукло, а функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *дифференцируема на* $\mathbf{R}^n$, *и ее производная непрерывна по Липшицу на* $\mathbf{R}^n$ *с константой* $L > 0$.

Тогда любая предельная точка любой траектории $\{x^k\}$ *алгоритма 1 является стационарной точкой задачи* (1). *Если предельная точка существует или если функция* f *ограничена снизу на* P, *то для любой ограниченной последовательности* $\{t_k\} \subset \mathbf{R}_+$

$$|\pi_P(x^k - t_k f'(x^k)) - x^k| \to 0 \quad (k \to \infty). \tag{10}$$

Доказательство. Пусть используется правило Армихо; тогда в силу (6) для любого k имеем

$$f(x^k) - f(x^{k+1}) \geqslant -\varepsilon \langle f'(x^k),\, x^{k+1} - x^k \rangle \geqslant$$

$$\geqslant \frac{\varepsilon}{\alpha_k} |x^{k+1} - x^k|^2 \geqslant \frac{\varepsilon}{\widehat{\alpha}} |x^{k+1} - x^k|^2. \tag{11}$$

Если ни для какого k точка x^k не является стационарной в задаче (1), то

$$|x^{k+1} - x^k| = |\pi_P(x^k - \alpha_k f'(x^k)) - x^k| > 0,$$

поскольку $\alpha_k > 0$. Поэтому из (11) вытекает монотонное убывание последовательности $\{f(x^k)\}$. Пусть последовательность $\{x^k\}$ имеет предельную точку $\overline{x} \in \mathbf{R}^n$; тогда из монотонного убывания последовательности $\{f(x^k)\}$ следует ее ограниченность снизу (величиной $f(\overline{x})$), а значит, и сходимость (если функция f предполагается ограниченной снизу на $\mathbf{R}^n$, то это верно и при отсутствии у последовательности $\{x^k\}$ предельной точки). Таким образом, левая часть (11) стремится к нулю при $k \to \infty$, поэтому

$$|x^{k+1} - x^k| \to 0 \quad (k \to \infty). \tag{12}$$

Пусть последовательность $\{t_k\} \subset \mathbf{R}_+$ ограничена сверху числом $\widehat{t}$; тогда из результата задачи 1.1.9 и соотношений (8) и (12) выводим

$$|x^k(t_k) - x^k| \leqslant \max\left\{1,\, \frac{t_k}{\alpha_k}\right\} |x^k(\alpha_k) - x^k| \leqslant$$

$$\leqslant \max\left\{1,\, \frac{\widehat{t}}{\widecheck{\alpha}}\right\} |x^{k+1} - x^k| \to 0 \quad (k \to \infty),$$

а это и есть (10).

Выберем сходящуюся к $\overline{x}$ подпоследовательность $\{x^{k_j}\}$ последовательности $\{x^k\}$; тогда из соотношения (10) при $t_k = t \ \ \forall\, k$, где $t > 0$ — произвольное число, и из непрерывности оператора проектирования (см. задачу 1.1.8) вытекает равенство (3), что и означает стационарность точки $\overline{x}$.

Случай использования правила одномерной минимизации сводится к доказанному так же, как и в теореме 3.1.1. $\square$

Обозначим множество стационарных точек задачи (1) через S_0.

Задача 1. Пусть в дополнение к условиям теоремы 1 множество $L_{f,P}(f(x^0))$ ограничено. Доказать, что при этом

$$\operatorname{dist}(x^k,\ S_0 \cap L_{f,P}(f(x^0))) \to 0 \quad (k \to \infty)$$

(ср. с задачей 3.1.3).

Теперь предположим дополнительно, что выполнено условие отделимости критических поверхностей уровня, которое формулируется точно так же, как в п. 3.1.2 (с учетом нового определения множества S_0), и оценка расстояния

$$|\pi_P(x - f'(x)) - x| \geqslant \gamma \operatorname{dist}(x,\ S_0) \quad \forall\, x \in U, \tag{13}$$

где

$$U = \{x \in \mathbf{R}^n \mid |\pi_P(x - f'(x)) - x| < \delta\}, \tag{14}$$

а $\delta > 0$ и $\gamma > 0$ — некоторые константы. В случае $P = \mathbf{R}^n$ указанные два условия обсуждались в п. 3.1.2. Не вдаваясь в детали, отметим, что оба они выполнены, если, например, f — квадратичная функция, P — полиэдр и $S_0 \neq \varnothing$ (ср. с задачей 3.1.8).

Теорема 2. *Пусть в дополнение к условиям теоремы* 1 *значение задачи* (1) *конечно и выполнено* (13) *для множества* U, *заданного в* (14), *и некоторых* $\delta > 0$ *и* $\gamma > 0$. *Пусть, кроме того, выполнено условие отделимости критических поверхностей уровня.*

Тогда любая траектория $\{x^k\}$ *алгоритма* 1 *сходится к некоторой стационарной точке задачи* (1). *Скорость сходимости по функции линейная, а по аргументу геометрическая.*

Доказательство этой теоремы получается несложной модификацией доказательства теоремы 3.1.4. Нужно только принять во внимание используемое здесь понятие стационарности, а также результаты задач 1.1.9 и 1.2.4.

Задача 2. Доказать теорему 2.

Количественно более точные (и даже в определенном смысле неулучшаемые) оценки скорости сходимости методов проекции градиента в дополнительных предположениях о выпуклости функции f на P можно найти в [10, 24, 37].

Каждая итерация любого метода проекции градиента требует вычисления проекций определенных точек на множество P (возможно, неоднократного — в зависимости от используемой процедуры одномерного поиска). Поэтому метод может представлять практический

интерес лишь в тех случаях, когда проектирование выполняется легко, что определяется простотой устройства множества P. Действительно эффективные алгоритмы на основе методов проекции градиента и их обобщений получены для задач оптимизации с линейными ограничениями, т. е. когда P — полиэдр; см. [6, 27]. Заметим что в этом случае вычисление проекции сводится к решению задачи квадратичного программирования; см. § 7.3.

Задача 3. Получить следующие формулы для проекций:
а) если $P = \overline{B}(\widehat{x}, \delta)$, где $\widehat{x} \in \mathbf{R}^n$, $\delta > 0$, то $\forall\, y \in \mathbf{R}^n$

$$\pi_P(y) = \begin{cases} y, \text{ если } y \in P, \\ \widehat{x} + \delta(\widehat{x} - y)/|\widehat{x} - y|, \text{ если } y \in \mathbf{R}^n \setminus P; \end{cases}$$

б) если $P = \mathbf{R}^n_+$, то $\forall\, y \in \mathbf{R}^n$

$$(\pi_P(y))_j = \max\{0, y_j\}, \quad j = 1, \ldots, n;$$

в) если $P = \{x \in \mathbf{R}^n \mid x_j \in [a_j, b_j], j = 1, \ldots, n\}$, где a_j и b_j — заданные числа, $a_j \leqslant b_j$, $j = 1, \ldots, n$, то $\forall\, y \in \mathbf{R}^n$

$$(\pi_P(y))_j = \begin{cases} a_j, \text{ если } y_j < a_j, \\ y_j, \text{ если } a_j \leqslant y_j \leqslant b_j, \\ b_j, \text{ если } y_j > b_j, \end{cases} \qquad j = 1, \ldots, n;$$

г) если $P = \{x \in \mathbf{R}^n \mid \langle a, x\rangle = b\}$, где $a \in \mathbf{R}^n \setminus \{0\}$, $b \in \mathbf{R}$, то $\forall\, y \in \mathbf{R}^n$

$$\pi_P(y) = y + (b - \langle a, y\rangle)\,\frac{a}{|a|^2};$$

д) если $P = \{x \in \mathbf{R}^n \mid \langle a, x\rangle \leqslant b\}$, где $a \in \mathbf{R}^n \setminus \{0\}$, $b \in \mathbf{R}$, то $\forall\, y \in \mathbf{R}^n$

$$\pi_P(y) = y + \min\{0, b - \langle a, y\rangle\}\,\frac{a}{|a|^2};$$

е) если $P = \{x \in \mathbf{R}^n \mid Ax = b\}$, где $A \in \mathbf{R}(l, n)$, $b \in \mathbf{R}^l$, причем $\operatorname{rank} A = l$, то $\forall\, y \in \mathbf{R}^n$

$$\pi_P(y) = y - A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}(Ay - b).$$

Чтобы избавиться от необходимости многократного проектирования точек на множество P на каждой итерации метода, иногда вместо схемы (4) рассматривают схему

$$x^{k+1} = (1 - \alpha_k)x^k + \alpha_k \pi_P(x^k - t_k f'(x^k)), \quad k = 0, 1, \ldots,$$

где параметры $t_k > 0$ выбираются более-менее произвольно, а α_k определяется согласно правилу одномерной минимизации либо правилу Армихо при $\widehat{\alpha} = 1$. Разумеется, в этих правилах одномерного

поиска нужно переопределить $x^k(\alpha)$, положив для каждого $\alpha \geqslant 0$

$$x^k(\alpha) = (1-\alpha)x^k + \alpha\pi_P(x^k - t_k f'(x^k)).$$

Можно показать, что если последовательность $\{t_k\}$ отделена от нуля, то так построенные методы обладают глобальной сходимостью в смысле теоремы 1. Вместе с тем линейную скорость локальной сходимости для них установить не удается, по крайней мере в обычных предположениях.

Теоретическое значение методов проекции градиента определяется главным образом тем фактом, что многие другие методы, генерирующие допустимые траектории, вдоль которых соблюдается свойство монотонного невозрастания значений целевой функции, могут интерпретироваться как возмущения или модификации соответствующих методов проекции градиента.

Задача 4. Из начального приближения $x^0 = (3, 1)$ для задачи

$$x_1 - x_2 \to \min, \quad x \in D = \{x \in \mathbf{R}^2 \mid 1 \leqslant x_1 \leqslant 3,\ 1 \leqslant x_2 \leqslant 2\},$$

сделать два шага метода проекции градиента, использующего правило постоянного параметра.

4.1.2. Возможные направления и методы спуска. Оставшаяся часть этого параграфа посвящена изложению общей схемы и некоторых конкретных примеров методов, получаемых распространением идей методов спуска, рассмотренных в § 3.1, на задачи условной оптимизации. Для таких задач наряду с понятием направления убывания целевой функции используется понятие возможного направления.

Определение 2. Вектор $d \in \mathbf{R}^n$ называется *возможным направлением* относительно множества $D \subset \mathbf{R}^n$ в точке $x \in D$, если для любого достаточно малого $t > 0$ выполняется $x + td \in D$.

Множество всех возможных относительно множества D в точке $x \in D$ направлений является конусом и обозначается $\mathcal{F}_D(x)$. Таким образом, $d \in \mathcal{F}_D(x)$ в том и только том случае, когда любой достаточно малый сдвиг из точки x в направлении d не выводит из множества D; в частности, всегда $0 \in \mathcal{F}_D(x)$.

Следующие две леммы дают очевидные признаки возможности данного направления.

Лемма 2. *Пусть множество $D \subset \mathbf{R}^n$ замкнуто и выпукло. Тогда*

$$\xi - x \in \mathcal{F}_D(x) \quad \forall\, x, \xi \in D.$$

Лемма 3. *Пусть*

$$D = \{x \in \mathbf{R}^n \mid G(x) \leqslant 0\},$$

где отображение $G\colon \mathbf{R}^n \to \mathbf{R}^m$ *дифференцируемо в точке* $x \in D$. *Тогда*:

а) *для любого* $d \in \mathcal{F}_D(x)$ *выполнено* $\langle g_i'(x), d\rangle \leqslant 0 \quad \forall\, i \in I(x)$;

б) *если* $d \in \mathbf{R}^n$ *удовлетворяет условию* $\langle g_i'(x), d\rangle < 0 \quad \forall\, i \in I(x)$, *то* $d \in \mathcal{F}_D(x)$.

Здесь, как и всюду ранее, $g_i(\cdot)$ — компоненты отображения G, $i = 1, \ldots, m$, а $I(x) = \{i = 1, \ldots, m \mid g_i(x) = 0\}$ — множество индексов активных в точке x ограничений.

Задача 5. Показать, что в случае линейных ограничений (т. е. в случае аффинного отображения G) строгие неравенства в утверждении б) леммы 3 можно заменить нестрогими.

Методы спуска вводятся для общей задачи оптимизации

$$f(x) \to \min, \quad x \in D, \tag{15}$$

допустимое множество $D \subset \mathbf{R}^n$ которой, вообще говоря, не предполагается ни замкнутым, ни выпуклым, а целевая функция гладкой. К этой группе относятся методы, укладывающиеся в следующую итерационную схему:

$$x^{k+1} = x^k + \alpha_k d^k, \quad d^k \in \mathcal{D}_f(x^k) \cap \mathcal{F}_D(x^k), \quad k = 0, 1, \ldots, \tag{16}$$

где $x^0 \in D$, а параметры длины шага $\alpha_k > 0$ выбираются так, чтобы по крайней мере выполнялось

$$f(x^{k+1}) < f(x^k), \quad x^{k+1} \in D. \tag{17}$$

Заметим, что поскольку $d^k \in \mathcal{D}_f(x^k) \cap \mathcal{F}_D(x^k)$, условие (17) будет выполнено при любом достаточно малом $\alpha_k > 0$. Подразумевается, что если $\mathcal{D}_f(x^k) \cap \mathcal{F}_D(x^k) = \varnothing$, то процесс останавливают. Таким образом, происходит «спуск» генерируемых приближений на поверхности все более низкого уровня целевой функции в пределах допустимого множества задачи. Конкретный метод спуска характеризуется способом выбора возможных направлений убывания d^k, а также используемой процедурой одномерного поиска вдоль этого направления для выбора параметра длины шага α_k. Если $D = \mathbf{R}^n$, то схема (16), (17) совпадает с общей схемой методов спуска для задач безусловной оптимизации, введенной в п. 3.1.1.

Если $D = P$ — замкнутое и выпуклое множество, то второе условие в (17) выполнено $\forall\, \alpha_k \in [0,\, \widehat{\alpha}_k]$, где

$$\widehat{\alpha}_k = \sup_{\alpha \geqslant 0\colon x^k + \alpha d^k \in P} \alpha > 0 \tag{18}$$

(если $\widehat{\alpha}_k = +\infty$, то второе условие в (17) выполнено $\forall\, \alpha_k \geqslant 0$). Для выбора α_k могут применяться те же процедуры одномерного поиска,

что рассматривались в п. 3.1.1, но с учетом дополнительного ограничения $\alpha_k \leqslant \widehat{\alpha}_k$. Процедура существенно упрощается в тех случаях, когда $\widehat{\alpha}_k$ можно явно вычислить либо оценить снизу.

Задача 6. Показать, что если

$$P = \{x \in \mathbf{R}^n \mid Ax \leqslant b\},$$

где $A \in \mathbf{R}(m, n)$ — матрица со строками a_i, $i = 1, \ldots, m$, $b \in \mathbf{R}^m$, то $\forall\, x^k \in P$, $\forall\, d^k \in \mathcal{F}_P(x^k)$ для введенной в (18) величины $\widehat{\alpha}_k$ справедливо следующее:

а) если $\langle a_i,\, d^k\rangle > 0$ хотя бы для одного номера $i = 1, \ldots, m$, то

$$\widehat{\alpha}_k = \min_{i=1,\ldots,m:\, \langle a_i,\, d^k\rangle > 0} \frac{b_i - \langle a_i,\, x^k\rangle}{\langle a_i,\, d^k\rangle};$$

б) если $\langle a_i,\, d^k\rangle \leqslant 0 \quad \forall\, i = 1, \ldots, m$, то $\widehat{\alpha}_k = +\infty$.

В следующем пункте приводятся два примера реализации методов спуска для условных задач с простыми ограничениями.

4.1.3. Методы условного градиента. Условные методы Ньютона. Будем предполагать, что допустимое множество P задачи (1) не только замкнуто и выпукло, но и ограничено. *Методы условного градиента* характеризуются следующим способом выбора возможных направлений убывания d^k в схеме (16) (в которой $D = P$). Для текущего приближения $x^k \in P$ решается задача

$$\langle f'(x^k),\, x - x^k\rangle \to \min, \quad x \in P, \tag{19}$$

получаемая линеаризацией целевой функции задачи (1) в точке x^k с последующим отбрасыванием константы $f(x^k)$. Пусть $\overline{x}^k \in P$ — любое решение задачи (19), а $v_k = \langle f'(x^k),\, \overline{x}^k - x^k\rangle$ — ее значение (решение существует в силу теоремы Вейерштрасса). Заметим, что $v_k \leqslant \langle f'(x^k),\, x^k - x^k\rangle = 0$, поскольку $x^k \in P$. Если $v_k = 0$, то

$$\langle f'(x^k),\, x - x^k\rangle \geqslant v_k = 0 \quad \forall\, x \in P,$$

т. е. точка x^k является стационарной в задаче (1) (см. (2)); в этом случае процесс останавливают. Если же $v_k < 0$, то, как следует из лемм 3.1.1 и 2, вектор $d^k = \overline{x}^k - x^k$ удовлетворяет условию $d^k \in \in \mathcal{D}_f(x^k) \cap \mathcal{F}_P(x^k)$. Такой вектор d^k называют *условным антиградиентом* функции f в точке x^k относительно множества P.

Легко видеть, что при указанном способе выбора d^k величина $\widehat{\alpha}_k$, введенная в (18), равна 1, поэтому организация процедур одномерного поиска для выбора параметров длины шага α_k не вызывает затруднений.

Для методов условного градиента имеет место аналог теоремы 1; см. [37, 41]. Более того, в дополнительных предположениях о выпуклости f на P удается установить арифметическую скорость сходимости метода (см. [10, 24, 33, 34]), причем эта оценка оказывается неулучшаемой в естественных классах задач [33, 41]. Таким образом, скорость сходимости методов условного градиента весьма невысока.

Кроме того, такие методы имеют практический смысл лишь в тех случаях, когда отыскание решения вспомогательной задачи (19) с линейной целевой функцией не вызывает затруднений. Если P — полиэдр, то (19) — задача линейного программирования; см. гл. 7.

Задача 7. Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ дифференцируема в точке $x^k \in P \subset \mathbf{R}^n$, $f'(x^k) \neq 0$. Получить следующие формулы для решения $\overline{x}^k$ задачи (19):

а) если $P = \overline{B}(\widehat{x}, \delta)$, где $\widehat{x} \in \mathbf{R}^n$, $\delta > 0$, то

$$\overline{x}^k = \widehat{x} - \delta \frac{f'(x^k)}{|f'(x^k)|};$$

б) если $P = \{x \in \mathbf{R}^n \mid x_j \in [a_j, b_j],\ j = 1, \ldots, n\}$, где a_j и b_j — заданные числа, $a_j \leqslant b_j$, $j = 1, \ldots, n$, то решением служит любая точка $\overline{x}^k$ такая, что для $j = 1, \ldots, n$

$$\overline{x}_j^k = a_j, \quad \text{если} \quad \frac{\partial f}{\partial x_j}(x^k) > 0,$$

$$\overline{x}_j^k = b_j, \quad \text{если} \quad \frac{\partial f}{\partial x_j}(x^k) < 0,$$

$$\overline{x}_j^k \in [a_j, b_j], \quad \text{если} \quad \frac{\partial f}{\partial x_j}(x^k) = 0.$$

Несмотря на весьма ограниченную область применения и низкую скорость сходимости методы условного градиента представляют определенный теоретический интерес, поскольку являются простейшей реализацией фундаментальной идеи линеаризации функций, участвующих в постановке оптимизационной задачи. Свою окончательную форму эта идея обретает в рассматриваемых в § 4.2 методах возможных направлений для задач с функциональными ограничениями-неравенствами, где линеаризуется не только целевая функция, но и ограничения.

Естественным развитием идеи линеаризации является использование не линейной, а более точной квадратичной аппроксимации целевой функции. В результате приходим к *условным методам Ньютона*, т. е. к схеме (16), в которой $d^k = \overline{x}^k - x^k$, где $\overline{x}^k$ ищется как решение задачи

$$\langle f'(x^k), x - x^k \rangle + \frac{1}{2} \langle f''(x^k)(x - x^k), x - x^k \rangle \to \min, \quad x \in P. \quad (20)$$

Если P — полиэдр, то (20) — задача квадратичного программирования; см. § 7.3. В остальных случаях отыскание решения задачи (20) может быть немногим проще, чем исходной задачи (1). Если множество P задается функциональными ограничениями, то их можно линеаризовать в текущей точке x^k. Однако при этом более эффективным оказывается использование другого квадратичного члена в целевой функции соответствующей вспомогательной задачи: для обеспечения глобальной сходимости этот член может задаваться любой положительно определенной матрицей (вместо $f''(x^k)$) в то время как для обеспечения высокой скорости локальной сходимости этот член должен вбирать в себя информацию о «кривизне» не только целевой функции, но и ограничений задачи. Получаемым на этом пути методам последовательного квадратичного программирования посвящены § 4.4 и § 5.4.

Задача 8. Из начального приближения $x^0 = 0$ для задачи

$$x_1^2 + 2x_2^2 - 3x_1 + 4x_2 \to \min, \quad x \in \overline{B}(0,\, 2),$$

сделать два шага метода условного градиента с выбором параметров длины шага по правилу одномерной минимизации.

§ 4.2. Методы возможных направлений

На каждой итерации методов условной оптимизации, рассматривавшихся в § 4.1, допустимое множество задачи используется как целое: в методах проекции градиента на него осуществляется проектирование, а в методах условного градиента и условных методах Ньютона ищется глобальное решение вспомогательной оптимизационной задачи с тем же допустимым множеством. Разумеется, возможность такого «глобального ви́дения» ограничивается лишь случаями, когда глобальное устройство допустимого множества является достаточно простым.

Для более сложных задач более реалистичным представляется строить методы, использующие вместо глобальной локальную информацию, доступную в окрестности текущего приближения к решению. Этот подход реализуется, в частности, в *методах возможных направлений*, как принято называть методы спуска для задачи с функциональными ограничениями-неравенствами:

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in \mathbf{R}^n \,|\, G(x) \leqslant 0\}, \tag{2}$$

где функцию $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображение $G\colon \mathbf{R}^n \to \mathbf{R}^m$ будем считать гладкими. Об основных идеях таких методов и пойдет речь в этом параграфе.

Напомним введенную в п. 4.1.2 общую схему методов спуска для условных задач:

$$x^{k+1} = x^k + \alpha_k d^k, \quad d^k \in \mathcal{D}_f(x^k) \cap \mathcal{F}_D(x^k), \quad k = 0, 1, \ldots, \qquad (3)$$

где $x^0 \in D$, а параметры длины шага $\alpha_k > 0$ выбираются так, чтобы по крайней мере выполнялось

$$f(x^{k+1}) < f(x^k), \quad x^{k+1} \in D. \qquad (4)$$

Подчеркнем, что если в задаче присутствуют нелинейные ограничения-равенства, то нетривиальность конуса возможных относительно допустимого множества направлений — совершенно нетипичное явление, поэтому при рассмотрении методов спуска для задач с функциональными ограничениями обычно обсуждают случай чистых ограничений-неравенств. Эти рассмотрения легко распространяются на случай наличия линейных ограничений-равенств, а также «простых» прямых ограничений.

Базовая реализация методов возможных направлений основана на следующем способе выбора d^k в схеме (3). Для текущего приближения $x^k \in D$ решается задача линейного программирования

$$\sigma \to \min, \quad (\sigma, d) \in U_k, \qquad (5)$$

$$U_k = \{u = (\sigma, d) \in \mathbf{R} \times \mathbf{R}^n \,\big|\, \langle f'(x^k), d\rangle \leqslant \sigma,$$

$$\langle g_i'(x^k), d\rangle \leqslant \sigma, \, i \in I_k, \, -1 \leqslant d_j \leqslant 1, \, j = 1, \ldots, n\}, \qquad (6)$$

где, как обычно, $g_i(\cdot)$ — компоненты отображения G, $i = 1, \ldots, m$, а $I_k = I(x^k) = \{i = 1, \ldots, m \,\big|\, g_i(x^k) = 0\}$ — множество индексов активных в точке x^k ограничений задачи (1), (2).

В определении множества U_k условие нормировки $-1 \leqslant d_j \leqslant 1$, $j = 1, \ldots, n$, служит для обеспечения гарантированной разрешимости задачи (5), (6). Действительно, $(0, 0) \in U_k$, т. е. $U_k \neq \varnothing$, и разрешимость задачи (5), (6) немедленно вытекает из следствия 1.1.1.

Пусть $u^k = (\sigma_k, d^k)$ — любое решение задачи (5), (6). Очевидно, $\sigma_k \leqslant 0$, поскольку в допустимой точке $(0, 0)$ целевая функция задачи (5), (6) принимает нулевое значение. Если $\sigma_k = 0$, то процесс останавливают.

Предложение 1. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображение $G\colon \mathbf{R}^n \to \mathbf{R}^m$ дифференцируемы в точке $x^k \in D$, где множество D введено в* (2).

Тогда если в точке x^k выполнено условие регулярности Мангасариана–Фромовица и при некотором $d^k \in \mathbf{R}^n$ точка $(0, d^k)$ является решением задачи (5), (6) *при $I_k = I(x^k)$, то x^k — стационарная точка задачи* (1), (2).

Доказательство. Напомним, что условие Мангасариана–Фромовица в данном случае состоит в существовании такого вектора $\overline{h} \in \mathbf{R}^n$, что

$$\langle g_i'(x^k), \overline{h} \rangle < 0 \quad \forall\, i \in I(x^k) = I_k. \tag{7}$$

Если точка $(0, d^k)$ является решением задачи (5), (6), то точка $(0, 0)$ также является ее решением. Применяя теорему 1.4.3, учитывая то обстоятельство, что ограничения задачи (5), (6) удовлетворяют условию линейности, и принимая во внимание утверждение из задачи 1.4.4, получаем существование чисел $\widetilde{\mu}_0 \geqslant 0$ и $\widetilde{\mu}_i \geqslant 0$, $i \in I_k$, таких, что

$$1 - \widetilde{\mu}_0 - \sum_{i \in I_k} \widetilde{\mu}_i = 0, \tag{8}$$

$$\widetilde{\mu}_0 f'(x^k) + \sum_{i \in I_k} \widetilde{\mu}_i g_i'(x^k) = 0. \tag{9}$$

Равенство (8) означает, в частности, что среди чисел $\widetilde{\mu}_0$, $\widetilde{\mu}_i$, $i \in I_k$, есть положительные. Тогда, умножая левую и правую части равенства (9) скалярно на $\overline{h}$ и используя (7), получаем, что $\widetilde{\mu}_0$ не может равняться нулю. Разделив левую и правую части (9) на $\widetilde{\mu}_0$, приходим к равенству

$$f'(x^k) + \sum_{i \in I_k} \overline{\mu}_i g_i'(x^k) = 0,$$

где $\overline{\mu}_i = \widetilde{\mu}_i / \widetilde{\mu}_0 \geqslant 0$, $i \in I_k$, а это и означает стационарность точки x^k в задаче (1), (2). $\square$

Предложение 1 может быть доказано и другими способами, например непосредственно через прямые условия оптимальности.

Если же $\sigma_k < 0$, то, как следует из лемм 3.1.1 и 4.1.3, вектор d^k удовлетворяет условию $d^k \in \mathcal{D}_f(x^k) \cap \mathcal{F}_D(x^k)$. Собственно говоря, суть вспомогательной задачи (5), (6) состоит в том, что с уменьшением σ вторая компонента d допустимой точки (σ, d) этой задачи приобретает все более ярко выраженные свойства возможного направления убывания. В частности, при $I(x^k) = \varnothing$ выбираемый вектор d^k совпадает по направлению с антиградиентом функции f в точке x^k.

Далее, выбор параметров длины шага α_k посредством процедур одномерного поиска из п. 3.1.1 должен производиться с учетом дополнительного ограничения $\alpha_k \leqslant \widehat{\alpha}_k$, где $\widehat{\alpha}_k > 0$ — максимальное число, для которого $x^k + \alpha d^k \in D \quad \forall\, \alpha \in [0, \widehat{\alpha}_k]$. Это ограничение обеспечивает выполнение второго условия в (4). Если множество D выпукло (например, если функции g_i выпуклы, $i = 1, \ldots, m$), то

указанная величина $\widehat{\alpha}_k$ дается формулой

$$\widehat{\alpha}_k = \sup_{\alpha \geqslant 0:\, x + \alpha d \in D} \alpha.$$

Но даже при этом $\widehat{\alpha}_k$ удается явно вычислить либо оценить снизу лишь в специальных случаях (см. п. 4.1.2). Если же нет оснований считать D выпуклым, то сколько-нибудь осмысленный выбор параметров длины шага вообще представляет серьезную проблему.

Кроме того, даже если считать $\widehat{\alpha}_k$ известным для каждого k и даже в очень сильных предположениях, описанные методы возможных направлений могут либо вообще не сходиться ни в каком смысле, либо «застревать» вблизи точек, не являющихся стационарными в задаче (1), (2). Соответствующий пример имеется, скажем, в [4]. Дело в том, что при $I_k = I(x^k)$ вспомогательная задача (5), (6) никак не принимает в расчет ограничения исходной задачи (1), (2), которые не активны, но «почти» активны в точке x^k. Это может приводить к тому, что по выбранному направлению d^k величина $\widehat{\alpha}_k$ оказывается слишком малой, в то время как существуют возможные направления убывания, по которым можно сделать значительно бóльший шаг.

Идея учета «почти» активных ограничений реализована в *методах Зойтендейка*, в которых перед началом процесса фиксируют числа $\delta_0 > 0$ и $\nu \in (0, 1)$ и для каждого k полагают

$$I_k = \{i = 1, \ldots, m \mid g_i(x^k) \geqslant -\delta_k\}.$$

Если для решения $u^k = (\sigma_k, d^k)$ соответствующей задачи (5), (6) выполняется неравенство $\sigma_k \leqslant -\delta_k$, то d^k принимается в качестве подходящего возможного направления убывания. В противном случае δ_k заменяют на $\nu\delta_k$ и повторяют процедуру для данного шага. Для методов Зойтендейка в определенных предположениях уже удается установить глобальную сходимость ко множеству стационарных точек (см. [10, 24]).

Заметим, что как методы проекции градиента, так и методы спуска для условных задач предполагают знание начальной точки x^0, которая допустима. В случае задачи с простым допустимым множеством, т. е. в ситуации § 4.1, отыскание такой точки обычно не составляет труда. В общем же случае применяют специальные методы поиска начальной допустимой точки [24, 41], однако нужно отметить, что этот поиск может быть сравним по сложности с поиском решения исходной задачи (1), (2).

Подробнее методы спуска здесь не обсуждаются, поскольку при всей их естественности и важности в идейном плане эти методы вряд ли могут считаться «правильным» подходом к решению задач оптимизации с нелинейными функциональными ограничениями, не говоря уже о том, что эти методы применимы лишь при отсутствии нелинейных ограничений-равенств. Значительно более эффективные

и практически востребованные методы общего назначения обсуждаются ниже.

§ 4.3. Методы решения задач с ограничениями-равенствами

В этом параграфе речь пойдет о методах решения задачи

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in \mathbf{R}^n \mid F(x) = 0\}, \tag{2}$$

где $f\colon \mathbf{R}^n \to \mathbf{R}$ — гладкая функция, $F\colon \mathbf{R}^n \to \mathbf{R}^l$ — гладкое отображение. Напомним, что задачу, в которой присутствуют также и ограничения-неравенства, всегда можно привести к виду (1), (2) за счет введения дополнительных переменных; см. начало § 1.4. Поэтому излагаемые здесь методы в принципе применимы и к задачам со смешанными ограничениями. Упомянутый прием может оказаться эффективным, особенно в тех случаях, когда, используя конструкцию метода, удается явно выразить дополнительные переменные через исходные, тем самым избежав искусственного повышения размерности задачи. В частности, этот прием будет использоваться в § 4.7 при распространении на случай смешанных ограничений материала из пп. 4.3.2, 4.3.3 настоящего параграфа. Вместе с тем, как уже отмечалось выше, непосредственный учет ограничений-неравенств часто приводит к лучшим результатам. Ограничимся простым примером: система Лагранжа для задачи с ограничениями-равенствами в равной мере описывает как локальные минимумы, так и локальные максимумы, в то время как для системы Каруша–Куна–Таккера это, вообще говоря, не так (из-за условия неотрицательности множителя, отвечающего ограничениям-неравенствам). О методах, специально разработанных для случая смешанных ограничений, речь пойдет в § 4.4–4.6.

4.3.1. Методы решения системы Лагранжа. Стационарные точки задачи (1), (2) описываются системой Лагранжа

$$L'(x, \lambda) = 0, \tag{3}$$

где

$$L\colon \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}, \quad L(x, \lambda) = f(x) + \langle \lambda, F(x) \rangle,$$

— функция Лагранжа. Но (3) есть система $n + l$ уравнений относительно $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$, поэтому стационарные точки вместе с соответствующими множителями Лагранжа можно искать, применяя к (3) метод Ньютона, рассмотренный в п. 3.2.1:

$$(x^{k+1}, \lambda^{k+1}) = (x^k, \lambda^k) - (L''(x^k, \lambda^k))^{-1} L'(x^k, \lambda^k), \quad k = 0, 1, \ldots.$$

Распишем ньютоновскую итерацию, вычисляя явно производные функции Лагранжа. Пусть x^k — текущее приближение к искомой

стационарной точке $\overline{x}$ задачи (1), (2), а $\lambda^k \in \mathbf{R}^l$ — к отвечающему $\overline{x}$ множителю Лагранжа. Тогда следующее приближение $(x^{k+1},\ \lambda^{k+1})$ ищется как решение линейной системы

$$\frac{\partial^2 L}{\partial x^2}\,(x^k,\ \lambda^k)(x-x^k)+(F'(x^k))^{\mathrm{T}}(\lambda-\lambda^k) = -f'(x^k)-(F'(x^k))^{\mathrm{T}}\lambda^k, \quad (4)$$

$$F'(x^k)(x-x^k) = -F(x^k) \quad (5)$$

относительно $(x,\ \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$.

Таким образом, *метод Ньютона* для системы Лагранжа состоит в следующем.

Алгоритм 1. Выбираем $(x^0,\ \lambda^0) \in \mathbf{R}^n \times \mathbf{R}^l$ и полагаем $k = 0$.

1. Вычисляем $(x^{k+1},\ \lambda^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^l$ как решение системы (4), (5).

2. Увеличиваем номер шага k на 1 и переходим к п. 1.

Теорема 1. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображение* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *дважды дифференцируемы в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$, *причем их вторые производные непрерывны в этой точке. Пусть в точке* $\overline{x}$ *выполнено условие регулярности ограничений, причем* $\overline{x}$ *— стационарная точка задачи* (1), (2), *а* $\overline{\lambda} \in \mathbf{R}^l$ *— однозначно отвечающий ей множитель Лагранжа. Пусть, наконец, в точке* $\overline{x}$ *выполнено сформулированное в теореме* 1.3.7 *достаточное условие второго порядка оптимальности.*

Тогда любое начальное приближение $(x^0,\ \lambda^0) \in \mathbf{R}^n \times \mathbf{R}^l$, *достаточно близкое к точке* $(\overline{x},\ \overline{\lambda})$, *корректно определяет траекторию алгоритма* 1, *которая сходится к* $(\overline{x},\ \overline{\lambda})$. *Скорость сходимости сверхлинейная, а если вторые производные* f *и* F *непрерывны по Липшицу в окрестности точки* $\overline{x}$, *то квадратичная.*

Доказательство. Требуемый результат будет немедленно следовать из теоремы 3.2.1, если показать невырожденность матрицы

$$L''(\overline{x},\ \overline{\lambda}) = \begin{pmatrix} \dfrac{\partial^2 L}{\partial x^2}(\overline{x},\ \overline{\lambda}) & (F'(\overline{x}))^{\mathrm{T}} \\ F'(\overline{x}) & 0 \end{pmatrix}.$$

Рассмотрим произвольный элемент $\widetilde{u} = (\widetilde{x},\ \widetilde{\lambda}) \in \ker L''(\overline{x},\ \overline{\lambda})$; тогда

$$\frac{\partial^2 L}{\partial x^2}\,(\overline{x},\ \overline{\lambda})\widetilde{x} + (F'(\overline{x}))^{\mathrm{T}}\widetilde{\lambda} = 0, \quad (6)$$

$$F'(\overline{x})\widetilde{x} = 0. \quad (7)$$

Умножая левую и правую части (6) скалярно на $\widetilde{x}$ и используя (7), получим

$$0 = \left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda})\widetilde{x}, \widetilde{x} \right\rangle + \langle \widetilde{\lambda}, F'(\overline{x})\widetilde{x} \rangle = \left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda})\widetilde{x}, \widetilde{x} \right\rangle.$$

Это означает, что $\widetilde{x} = 0$, поскольку иначе было бы нарушено достаточное условие второго порядка оптимальности в точке $\overline{x}$. Но тогда из (6) вытекает, что

$$(F'(\overline{x}))^{\mathrm{T}}\widetilde{\lambda} = 0. \tag{8}$$

Условие регулярности ограничений в точке $\overline{x}$ эквивалентно выполнению равенства $\ker(F'(\overline{x}))^{\mathrm{T}} = \{0\}$, поэтому (8) может иметь место лишь при $\widetilde{\lambda} = 0$. Тем самым показано, что $\widetilde{u} = (\widetilde{x}, \widetilde{\lambda}) = 0$, т. е. матрица $L''(\overline{x}, \overline{\lambda})$ невырождена. □

Заметим, что матрица линейной системы (4), (5) является симметрической, но, как нетрудно видеть, при $l > 0$ эта матрица не может быть строго знакоопределенной. Для решения такой системы могут применяться как общие методы вычислительной линейной алгебры [12], так и специальные методы, использующие особую структуру матрицы этой системы [42].

Квадратичная скорость сходимости траектории в прямодвойственных переменных, вообще говоря, не влечет сверхлинейную скорость сходимости в прямых переменных (см. задачу 2.1.1). Эта проблема, а также проблема глобализации сходимости метода Ньютона для системы Лагранжа будет рассмотрена ниже в контексте методов последовательного квадратичного программирования (см. § 4.4 и § 5.4).

Интересно отметить, что если в искомой стационарной точке выполнено условие регулярности ограничений и известно достаточно хорошее начальное приближение x^0 в прямых переменных, то хорошее начальное приближение λ^0 в двойственных переменных всегда может быть указано. Действительно, условие регулярности ограничений дает невырожденность матрицы $F'(\overline{x})(F'(\overline{x}))^{\mathrm{T}}$, поэтому из определения стационарной точки и отвечающего ей множителя Лагранжа имеем

$$\overline{\lambda} = -(F'(\overline{x})(F'(\overline{x}))^{\mathrm{T}})^{-1}F'(\overline{x})f'(\overline{x}).$$

Полагая

$$\lambda^0 = -(F'(x^0)(F'(x^0))^{\mathrm{T}})^{-1}F'(x^0)f'(x^0),$$

в силу теоремы о малом возмущении невырожденной матрицы получим требуемое.

Задача 1. Зададим множество

$$\mathcal{R} = \{x \in \mathbf{R}^n \mid \operatorname{rank} F'(x) = l\}$$

и отображение

$$\lambda_\tau\colon \mathcal{R} \to \mathbf{R}^l, \quad \lambda_\tau(x) = (F'(x)(F'(x))^{\mathrm{T}})^{-1}(\tau F(x) - F'(x)f'(x)),$$

где $\tau \in \mathbf{R}$ — параметр. Вычислить производную $\lambda_\tau(\cdot)$ в стационарной точке $\overline{x}$ задачи (1), (2), предполагая выполнение в этой точке условия регулярности ограничений. На этой основе построить и обосновать неточный метод Ньютона для уравнения

$$\frac{\partial L}{\partial x}(x,\, \lambda_\tau(x)) = 0$$

относительно $x \in \mathbf{R}^n$ при подходящем выборе τ. Сравнить трудоемкость итераций данного метода и алгоритма 1.

Разумеется, к системе Лагранжа (3) может применяться не только метод Ньютона, но и различные его модификации (см. п. 3.2.1 и [6]), а также другие методы решения систем нелинейных уравнений [31]. Большое практическое значение имеют методы [1)], в которых на шаге k вместо матрицы $\frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)$ «работает» матрица $\Xi_k^{\mathrm{T}} \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k)\Xi_k$, где столбцы матрицы $\Xi_k \in \mathbf{R}(n,\, \dim\ker F'(x^k))$ образуют базис в $\ker F'(x^k)$, гладко зависящий от x^k.

Систему (3) можно решать, применяя методы безусловной оптимизации к задаче

$$|L'(x,\, \lambda)|^2 \to \min, \quad (x,\, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l. \tag{9}$$

Глобальные решения этой задачи совпадают с решениями системы (3) (в случае разрешимости последней). В частности, глобально сходящиеся методы решения системы (3) можно получать как глобально сходящиеся методы безусловной оптимизации применительно к задаче (9). Принципиальным недостатком такого подхода, как и других методов непосредственного решения системы Лагранжа, является то, что при этом пропадает оптимизационная сущность исходной задачи (1), (2), и, в частности, локальные максимумы и минимумы этой задачи не различаются. Ниже в этом параграфе речь идет о методах, в определенном смысле свободных от указанного недостатка.

4.3.2. Метод квадратичного штрафа. Идея снятия функциональных ограничений задачи условной оптимизации за счет введения в ее целевую функцию дополнительных слагаемых, обеспечивающих «штрафование» за нарушение снимаемых ограничений, находит применение при построении многих оптимизационных алгоритмов. Непосредственно эта идея реализуется в так называемых методов штрафов, которые с более общих позиций будут рассмотрены в § 4.7. Здесь

[1)] Общепринятое английское название для методов такого типа — Reduced Hessian methods.

же речь пойдет об одном из наиболее естественных методов такого рода для задачи (1), (2) с чистыми ограничениями-равенствами.

Введем семейство *штрафных функций*, получаемых добавлением к f *квадратичного штрафа* $|F(\cdot)|^2/2$, умноженного на *параметр штрафа* $c \geqslant 0$:

$$\varphi_c\colon \mathbf{R}^n \to \mathbf{R}, \quad \varphi_c(x) = f(x) + \frac{c}{2}\,|F(x)|^2. \tag{10}$$

Заметим, что $\varphi_c(\cdot) \equiv f(\cdot)$ на D, в то время как в любой точке вне допустимого множества D значение штрафного слагаемого неограниченно возрастает с ростом c. В этом смысле задача безусловной оптимизации

$$\varphi_c(x) \to \min, \quad x \in \mathbf{R}^n, \tag{11}$$

аппроксимирует исходную задачу (1), (2) при $c \to +\infty$.

Метод квадратичного штрафа для задачи (1), (2) состоит в следующем.

Алгоритм 2. Выбираем последовательность $\{c_k\} \subset \mathbf{R}_+$ такую, что $c_k \to \infty$ $(k \to \infty)$, и полагаем $k = 0$.

1. Вычисляем $x^k \in \mathbf{R}^n$ как стационарную точку задачи (11) с целевой функцией, задаваемой формулой (10) при $c = c_k$.
2. Увеличиваем номер шага k на 1 и переходим к п. 1.

Для отыскания стационарных точек задачи (11) могут использоваться методы безусловной оптимизации, рассмотренные в гл. 3. Заметим, что достаточная гладкость функции f и отображения F обеспечивает любую нужную гладкость целевой функции задачи (11). Алгоритму 2 можно придать последовательный характер, если, например, фиксировать используемый метод безусловной оптимизации и для каждого $k = 1, 2, \ldots$ в качестве начального приближения для этого метода брать точку x^{k-1} (что вполне естественно). Чтобы такое начальное приближение было «достаточно хорошим», c_k не должно сильно отличаться от c_{k-1}; иными словами, увеличение параметра штрафа с ростом k не должно происходить слишком быстро.

Имеет место следующая теорема о сходимости метода квадратичного штрафа.

Теорема 2. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображение* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *непрерывно дифференцируемы на* $\mathbf{R}^n$.

Тогда если алгоритм 2 *генерирует последовательность* $\{x^k\}$, *то любая ее предельная точка* $\overline{x} \in \mathbf{R}^n$, *в которой выполнено условие регулярности ограничений, является стационарной точкой задачи* (1), (2). *Более того, для любой сходящейся к* $\overline{x}$ *подпоследовательности* $\{x^{k_j}\}$ *справедливо*

$$\{c_{k_j} F(x^{k_j})\} \to \overline{\lambda} \quad (i \to \infty), \tag{12}$$

где $\overline{\lambda} \in \mathbf{R}^l$ — *(единственный) отвечающий* $\overline{x}$ *множитель Лагранжа.*

Доказательство. Для каждого k, полагая $\lambda^k = c_k F(x^k)$, имеем

$$\varphi'_{c_k}(x) = f'(x^k) + (F'(x^k))^{\mathrm{T}} \lambda^k = 0, \tag{13}$$

следовательно,

$$F'(x^k) f'(x^k) + F'(x^k)(F'(x^k))^{\mathrm{T}} \lambda^k = 0.$$

Из условия регулярности ограничений в точке $\overline{x}$ и теоремы о малом возмущении невырожденной матрицы вытекает, что при достаточно больших j матрица $F'(x^{k_j})(F'(x^{k_j}))^{\mathrm{T}}$ невырождена, поэтому

$$\lambda^{k_j} = -(F'(x^{k_j})(F'(x^{k_j}))^{\mathrm{T}})^{-1} F'(x^{k_j}) f'(x^{k_j}).$$

Отсюда и из непрерывной дифференцируемости f и F следует предельное соотношение

$$\{\lambda^{k_j}\} \to \widetilde{\lambda} \quad (i \to \infty), \tag{14}$$

где

$$\widetilde{\lambda} = -(F'(\overline{x})(F'(\overline{x}))^{\mathrm{T}})^{-1} F'(\overline{x}) f'(\overline{x}).$$

Но тогда, с учетом (13), справедливо равенство

$$\frac{\partial L}{\partial x}(\overline{x}, \widetilde{\lambda}) = f'(\overline{x}) + (F'(\overline{x}))^{\mathrm{T}} \widetilde{\lambda} = 0.$$

Кроме того, в силу стремления c_k к бесконечности и (14)

$$F(\overline{x}) = \lim_{i \to \infty} F(x^{k_j}) = \lim_{i \to \infty} \frac{\lambda^{k_j}}{c_{k_j}} = 0.$$

Тем самым показано, что $\overline{x}$ — стационарная точка задачи (1), (2), а $\overline{\lambda} = \widetilde{\lambda}$ — единственный множитель Лагранжа, отвечающий этой стационарной точке. Соотношение (12) следует из (14). □

В общем случае задача (11) может не иметь стационарных точек либо иметь более одной стационарной точки. Следующая теорема содержит, в частности, достаточные условия того, что первое не имеет места.

Теорема 3. *Пусть функция* $f: \mathbf{R}^n \to \mathbf{R}$ *и отображение* $F: \mathbf{R}^n \to \mathbf{R}^l$ *непрерывны на* $\mathbf{R}^n$. *Пусть точка* $\overline{x} \in \mathbf{R}^n$ *является строгим локальным решением задачи* (1), (2).

Тогда найдется такое число $\delta > 0$, *что для* $c \geqslant 0$ *и для любого (глобального) решения* $x(c)$ *задачи*

$$\varphi_c(x) \to \min, \quad x \in \overline{B}(\overline{x}, \delta),$$

целевая функция которой задается формулой (10), *имеет место*

$$\|x(c) - \overline{x}\| \to 0 \quad (c \to +\infty). \tag{15}$$

В частности, для любого достаточно большого числа c точка $x(c)$ является локальным решением задачи (11).

Заметим, что согласно теореме Вейерштрасса (теорема 1.1.1) для любого c точка $x(c)$, о которой идет речь в этой теореме, всегда существует.

Доказательство. Выберем число $\delta > 0$ так, чтобы точка $\overline{x}$ была единственным глобальным решением задачи

$$f(x) \to \min, \quad x \in D \cap \overline{B}(\overline{x}, \delta). \tag{16}$$

Значения функции $x(\cdot)\colon \mathbf{R} \to \mathbf{R}^n$ содержатся в компакте $\overline{B}(\overline{x}, \delta)$, а значит, эта функция имеет предельную точку $\widetilde{x} \in \overline{B}(\overline{x}, \delta)$ при $c \to +\infty$. Докажем, что $\widetilde{x}$ является глобальным решением задачи (16).

Действительно, $\forall\, c \geqslant 0$ имеет место

$$\varphi_c(x(c)) \leqslant \varphi_c(x) \quad \forall\, x \in \overline{B}(\overline{x}, \delta),$$

поэтому в силу (2) и (10) имеем

$$f(x(c)) + \frac{c}{2}|F(x(c))|^2 = \varphi_c(x(c)) \leqslant \inf_{x \in D \cap \overline{B}(\overline{x}, \delta)} \varphi_c(x) = \inf_{x \in D \cap \overline{B}(\overline{x}, \delta)} f(x),$$

где в правой части стоит значение $v = v(\overline{x}, \delta)$ задачи (16). Тогда, фиксируя последовательность $\{c_k\} \subset \mathbf{R}$ такую, что $c_k \to +\infty$, $\{x(c_k)\} \to \widetilde{x}$ $(k \to \infty)$, получаем

$$f(\widetilde{x}) + \limsup_{k \to +\infty} \frac{c_k}{2}|F(x(c_k))|^2 \leqslant v. \tag{17}$$

Отсюда следует, что

$$F(\widetilde{x}) = \lim_{k \to \infty} F(x(c_k)) = 0,$$

т. е. $\widetilde{x} \in D \cap \overline{B}(\overline{x}, \delta)$. Поэтому

$$v \leqslant f(\widetilde{x}), \tag{18}$$

и из (17) вытекает неравенство

$$\limsup_{k \to \infty} c_k |F(x(c_k)|^2 \leqslant 0,$$

которое может иметь место только при выполнении равенства

$$\lim_{k \to \infty} c_k |F(x(c_k))|^2 = 0.$$

Таким образом, (17) принимает вид

$$f(\widetilde{x}) \leqslant v,$$

что в совокупности с (18) дает равенство

$$f(\widetilde{x}) = v,$$

которое, с учетом допустимости $\widetilde{x}$ в задаче (16), означает, что $\widetilde{x}$ — глобальное решение этой задачи.

Но тогда в силу выбора числа δ справедливо равенство $\widetilde{x} = \overline{x}$, т. е. единственной предельной точкой функции $x(\cdot)$ является $\overline{x}$. Поэтому выполнено (15) и, в частности, $x(c) \in B(\overline{x}, \delta)$ для достаточно больших c. Очевидно, отсюда следует, что для таких c точка $x(c)$ является локальным решением задачи (11). $\square$

Задача 2. Доказать аналог теоремы 3 для случая, когда вместо существования у задачи (1), (2) строгого локального решения предполагается существование компактного изолированного множества локальных решений. (Под изолированным множеством локальных решений понимается состоящее из локальных решений множество $S \subset \subset \mathbf{R}^n$ такое, что для некоторого числа $\delta > 0$ в окрестности $U(S, \delta) = = \{x \in \mathbf{R}^n \mid \operatorname{dist}(x, S) < \delta\}$ не содержится локальных решений, не принадлежащих S.)

Задача 3. Убедиться, что доказательство теоремы 3 полностью проходит и для штрафной функции вида

$$\varphi_c(x) = f(x) + c\psi(x), \quad x \in \mathbf{R}^n,$$

где ψ: $\mathbf{R}^n \to \mathbf{R}$ — произвольный (не обязательно квадратичный) непрерывный на $\mathbf{R}^n$ штраф, т. е. функция, удовлетворяющая условиям $\psi(x) = 0 \quad \forall\, x \in D$, $\psi(x) > 0 \quad \forall\, x \in \mathbf{R}^n \setminus D$.

Согласно теореме 3 любое строгое локальное решение $\overline{x}$ задачи (1), (2) может быть найдено как предел последовательности, генерируемой алгоритмом 2, если в этом алгоритме стационарные точки задачи (11) выбираются соответствующим образом. В некоторых дополнительных предположениях можно утверждать, что при достаточно большом k задача (11) с целевой функцией, задаваемой формулой (10) при $c = c_k$, будет иметь в окрестности $\overline{x}$ единственную стационарную точку $x(c_k)$; более того, можно оценить скорость сходимости последовательности $\{x(c_k)\}$ к $\overline{x}$. Подчеркнем, что при этом, если для каждого k в качестве начального приближения для используемого метода безусловной минимизации берется точка, достаточно близкая к $\overline{x}$, то можно ожидать, что в качестве очередного приближения алгоритм будет находить именно $x^k = x(c_k)$. Это косвенно свидетельствует в пользу того, чтобы в качестве начального приближения брать x^{k-1}: попав однажды в достаточно малую окрестность точки $\overline{x}$ при достаточно большом k (а этого можно ожидать в силу теоремы 2), при таком выборе генерируемые алгоритмом приближения x^k уже не покинут этой окрестности, а значит, будут совпадать с $x(c_k)$.

Теорема 4. *Пусть выполнены условия теоремы* 1.

Тогда существуют окрестность U *точки* $\overline{x}$ *и числа* $\overline{c} \geqslant 0$ *и* $M > 0$ *такие, что для каждого* $c > \overline{c}$ *задача* (11) *с целевой*

функцией, задаваемой формулой (10), *имеет в* U *единственную стационарную точку* $x(c)$, *причем*

$$|x(c)-\overline{x}|\leqslant M\,\frac{|\overline{\lambda}|}{c}\,,\quad |cF(x(c))-\overline{\lambda}|\leqslant M\,\frac{|\overline{\lambda}|}{c}\,. \tag{19}$$

Доказательство. Введем отображение

$$\Phi\colon \mathbf{R}\times(\mathbf{R}^n\times\mathbf{R}^l)\to\mathbf{R}^n\times\mathbf{R}^l,\quad \Phi(\sigma,\,\chi)=\begin{pmatrix}\dfrac{\partial L}{\partial x}(\xi,\,\eta)\\ F(\xi)-\sigma\eta\end{pmatrix},\ \ \chi=(\xi,\,\eta). \tag{20}$$

Пусть $\overline{\lambda}\in\mathbf{R}^l$ — однозначно отвечающий стационарной точке $\overline{x}$ множитель Лагранжа. Положим $\overline{\chi}=(\overline{x},\,\overline{\lambda})$. Тогда $\Phi(0,\,\overline{\chi})=0$ и $\dfrac{\partial\Phi}{\partial\chi}\,(0,\,\overline{\chi})=L''(\overline{x},\,\overline{\lambda})$, причем невырожденность последней матрицы в условиях настоящей теоремы установлена при доказательстве теоремы 1. Применяя к отображению Φ в точке $(0,\,\overline{\chi})$ классическую теорему о неявной функции (теорема 1.3.1), получаем существование числа $\overline{\sigma}>0$ и окрестности $\mathcal{U}$ точки $\overline{\chi}$ в $\mathbf{R}^n\times\mathbf{R}^l$ таких, что для всякого $\sigma\in(-\overline{\sigma},\,\overline{\sigma})$ уравнение

$$\Phi(\sigma,\,\chi)=0 \tag{21}$$

относительно $\chi\in\mathbf{R}^n\times\mathbf{R}^l$ имеет в $\mathcal{U}$ единственное решение $\chi(\sigma)=(\xi(\sigma),\,\eta(\sigma))$, причем функция $\chi(\cdot)$ дифференцируема на $(-\overline{\sigma},\,\overline{\sigma})$ и ее производная непрерывна в точке 0. Кроме того, из теоремы 1.3.2 следует оценка

$$|\chi(\sigma)-\overline{\chi}|\leqslant M|\Phi(\sigma,\,\overline{\chi})|=M|\overline{\lambda}|\sigma\quad\forall\,\sigma\in(-\overline{\sigma},\,\overline{\sigma}) \tag{22}$$

при некотором $M>0$, если $\overline{\sigma}$ достаточно мало.

Положим $\overline{c}=1/\overline{\sigma}$, $x(c)=\xi(1/c)$, $c>\overline{c}$. Тогда для любого такого c из (20) имеем

$$\varphi_c'(x(c))=f'(x(c))+c(F'(x(c)))^{\mathrm{T}}F(x(c))=\frac{\partial L}{\partial x}\left(\xi\left(\frac{1}{c}\right),\eta\left(\frac{1}{c}\right)\right)=0,$$

т. е. $x(c)$ является стационарной точкой задачи (11). Кроме того, из оценки (22) следуют оценки (19), поскольку по построению $F(\xi(\sigma))=\sigma\eta(\sigma)\ \ \forall\,\sigma\in(-\overline{\sigma},\,\overline{\sigma})$.

Наконец, рассмотрим произвольные последовательности $\{c_i\}\subset\subset\mathbf{R}^n$ и $\{x^i\}\subset\mathbf{R}^n$ такие, что $\{c_i\}\to\infty$, $\{x^i\}\to\overline{x}$ $(i\to\infty)$, причем для каждого i точка x^i является стационарной в задаче (11). Как легко вывести из (20), для каждого i точка $\chi^i=(x^i,\,c_iF(x^i))$ будет решением уравнения (21) при $\sigma=1/c_i$, причем из теоремы 2 следует, что

$$\{c_iF(x^i)\}\to\overline{\lambda}\quad(i\to\infty).$$

Поэтому $\chi^i\in\mathcal{U}$ при достаточно больших i, а значит, $\chi^i=\chi(1/c_i)$, и, в частности, $x^i=x(c_i)$. Это доказывает, что $x(c)$ — единственная

в некоторой окрестности U точки $\overline{x}$ стационарная точка задачи (11) при достаточно большом c. $\square$

Как указано в приведенном доказательстве, определенные в нем функции $\xi(\cdot)$ и $\eta(\cdot)$ дифференцируемы на $(-\overline{\sigma}, \overline{\sigma})$ (на самом деле, повышая требования гладкости на f и F, можно обеспечить любую нужную гладкость этих функций, а значит, и функции $x(\cdot)$ на $(\overline{c}, +\infty)$). Кроме того, как следует из (20), точка $\xi(\sigma)$ для каждого $\sigma \in (-\overline{\sigma}, \overline{\sigma})$ удовлетворяет уравнению

$$\sigma f'(\xi) + (F'(\xi))^{\mathrm{T}} F(\xi) = 0 \tag{23}$$

относительно $\xi \in \mathbf{R}^n$. Дифференцируя левую часть этого уравнения вдоль траектории $\xi(\cdot)$, получаем, что последняя удовлетворяет на $(-\overline{\sigma}, \overline{\sigma})$ следующей системе обыкновенных дифференциальных уравнений:

$$\Big((\sigma - 1) f''(\xi) + \frac{\partial^2 L}{\partial x^2}(\xi, F(\xi)) + (F'(\xi))^{\mathrm{T}} F'(\xi)\Big)\dot{\xi} + f'(\xi) = 0, \tag{24}$$

где точка обозначает производную по σ. К тому же результату приводит использование приведенной в теореме 1.3.1 формулы для производной неявной функции. При более высокой гладкости f и F можно получить уравнения и для старших производных $\xi(\cdot)$.

Сказанное мотивирует следующую «непрерывную» трактовку метода квадратичного штрафа. Фиксируем число $\sigma_0 \neq 0$ и находим точку $\xi^0 \in \mathbf{R}^n$, которая является решением уравнения (23) при $\sigma = \sigma_0$, после чего теми или иными средствами это решение непрерывно продолжаем (экстраполируем) по параметру σ до значения $\sigma = 0$. В частности, при этом могут использоваться приближенные схемы для решения начальной задачи, получаемой добавлением к (24) начального условия

$$\xi(\sigma_0) = \xi^0,$$

а также экстраполяция с помощью формулы Тейлора. Разумеется, если начальная пара (σ_0, ξ^0) далека от пары $(0, \overline{x})$, где $\overline{x}$ удовлетворяет условиям теоремы 4, то такое продолжение не всегда возможно, а если возможно, то не обязательно единственным образом. С общих позиций методы продолжения по параметру обсуждаются в § 5.3.

Далее, как следует из оценки (19), при $\overline{\lambda} = 0$ (что в условиях теоремы 4 равносильно равенству $f'(\overline{x}) = 0$) $x(c) = \xi(1/c) = \overline{x}$ для любого достаточно большого c, т. е. квадратичный штраф оказывается *точным* в следующем смысле: при достаточно большом значении параметра штрафа c стационарная точка вспомогательной задачи безусловной оптимизации (11), достаточно близкая к искомому решению $\overline{x}$ задачи (1), (2), в точности совпадает с $\overline{x}$. Штрафную функцию φ_c при этом также называют *точной*.

Если же $\overline{\lambda} \neq 0$, то из (24), условия регулярности ограничений и определения множителя Лагранжа имеем

$$|F'(\overline{x})\dot{\xi}(0)| = |\overline{\lambda}| > 0,$$

т. е. при $\sigma \to 0$ траектория $\xi(\cdot)$ «входит» в решение $\overline{x}$ трансверсально допустимому множеству D в том смысле, что касательный вектор $\dot{\xi}(0)$ к этой траектории не содержится в касательном подпространстве $\ker F'(\overline{x})$ к D в точке $\overline{x}$. В этом случае оценка (19) является неулучшаемой по порядку и, в частности, квадратичный штраф не может быть точным.

Задача 4. Доказать следующие факты, дополняющие утверждение теоремы 4:

а) для любого числа $\varepsilon > 0$ найдется число $c(\varepsilon) \geqslant \overline{c}$ такое, что

$$\frac{1}{c}\left(\|(L''(\overline{x}, \overline{\lambda}))^{-1}\|\,|\overline{\lambda}| - \varepsilon\right) \leqslant$$

$$\leqslant |x(c) - \overline{x}| \leqslant \frac{1}{c}\left(\|(L''(\overline{x}, \overline{\lambda}))^{-1}\|\,|\overline{\lambda}| + \varepsilon\right) \quad \forall\, c > c(\varepsilon);$$

б) для любого достаточно большого $c > \overline{c}$ матрица $\varphi_c''(x(c))$ положительно определена, т. е. в стационарной точке $x(c)$ задачи (11) выполнено сформулированное в теореме 1.2.5 достаточное условие второго порядка оптимальности.

Теоретически алгоритм 2 вполне надежен и сходится тем быстрее, чем быстрее возрастает последовательность $\{c_k\}$; последнее видно из оценки (19). Ясно, однако, что все трудности здесь перенесены на этап решения вспомогательных задач безусловной оптимизации. Выше уже была указана причина, по которой слишком быстрое увеличение параметра штрафа нежелательно. Основной же недостаток метода квадратичного штрафа состоит в том, что параметр штрафа должен бесконечно увеличиваться (за исключением нетипичного случая, когда в искомом решении выполнено $f'(\overline{x}) = 0$). С ростом c задача (11) становится все хуже обусловленной: ее целевая функция приобретает все более «овражный» характер (см. п. 3.1.2; подробный анализ ухудшения обусловленности вспомогательной задачи метода квадратичного штрафа с ростом параметра штрафа можно найти в [6]). Поэтому отыскание решения вспомогательной задачи с ростом k становится все более трудоемким. В определенном смысле острота этой проблемы снимается с помощью использования на заключительном этапе работы метода (т. е. при больших k) упомянутых выше экстраполяционных процедур; подробнее об этом см. [38]. Кроме того, указанная трудность может быть преодолена, если перейти от чистого метода квадратичного штрафа к рассматриваемым в п. 4.3.3 более совершенным (и соответственно несколько более сложным) схемам,

для которых бесконечное увеличение параметра штрафа оказывается ненужным.

Задача 5. Решить задачу

$$x \to \min, \quad x^p = 0,$$

где $p \geqslant 2$ — целочисленный параметр, методом квадратичного штрафа. Проанализировать скорость сходимости; объяснить наблюдаемый эффект.

Задача 6. Решить задачу

$$2x_1^2 + (x_2 - 1)^2 \to \min, \quad x \in D = \{x \in \mathbf{R}^2 \mid 2x_1 + x_2 = 0\},$$

методом квадратичного штрафа.

Задача 7. То же задание для задачи

$$x_1^2 + 2x_2^2 \to \min, \quad x \in D = \{x \in \mathbf{R}^2 \mid 3x_1 + x_2 = 3\}.$$

4.3.3. Модифицированные функции Лагранжа и точные гладкие штрафные функции. Один подход, избавляющий от необходимости бесконечного увеличения параметра штрафа, основан на добавлении гладкого (например, квадратичного) штрафного слагаемого не к целевой функции задачи (1), (2), а к ее функции Лагранжа, в результате чего получается семейство так называемых *модифицированных функций Лагранжа* задачи (1), (2):

$$L_c\colon \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}, \quad L_c(x, \lambda) = L(x, \lambda) + \frac{c}{2}|F(x)|^2, \tag{25}$$

где $c > 0$. Происхождение этого названия можно объяснить следующим образом. Стационарная точка $\overline{x}$ задачи (1), (2) является стационарной и для задачи

$$L(x, \overline{\lambda}) \to \min, \quad x \in \mathbf{R}^n, \tag{26}$$

где $\overline{\lambda}$ — отвечающий $\overline{x}$ множитель Лагранжа. Поэтому вместо решения системы Лагранжа задачи (1), (2) можно искать стационарную точку задачи (26). Разумеется, множитель $\overline{\lambda}$ заранее не известен, однако его можно так или иначе аппроксимировать, используя информацию о текущем приближении к $\overline{x}$ (см., например, задачу 1). Недостаток же такого подхода состоит в том, что, например, выполнение условий теоремы 1 не гарантирует положительной определенности матрицы $\dfrac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda})$, и обоснованное применение к задаче (26) методов безусловной оптимизации возможно лишь в неестественно сильных предположениях.

В анализе на оптимальность и при построении алгоритмов модифицированная функция Лагранжа может использоваться наряду

с обычной, поскольку, как легко видеть, $\forall\, c$ справедливо

$$\frac{\partial L_c}{\partial x}(x,\,\lambda) = \frac{\partial L}{\partial x}(x,\,\lambda) \quad \forall\, x \in D,\ \forall\, \lambda \in \mathbf{R}^l,$$

$$\frac{\partial L_c}{\partial \lambda}(x,\,\lambda) = \frac{\partial L}{\partial \lambda}(x,\,\lambda) = F(x) \quad \forall\, x \in \mathbf{R}^n,\ \forall\, \lambda \in \mathbf{R}^l,$$

и множество решений уравнения

$$L_c'(x,\,\lambda) = 0$$

совпадает с множеством решений системы Лагранжа задачи (1), (2). При этом с алгоритмической точки зрения модифицированная функция Лагранжа обладает определенными преимуществами перед обычной. Важнейшее из этих преимуществ раскрывается в утверждении, приводимом в следующей задаче.

Задача 8. Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ дважды дифференцируемы точке $\overline{x} \in \mathbf{R}^n$. Пусть $\overline{x}$ — стационарная точка задачи (1), (2), и в ней выполнено достаточное условие второго порядка оптимальности с множителем Лагранжа $\overline{\lambda} \in \mathbf{R}^l$. Доказать, что для любого достаточно большого числа c матрица

$$\frac{\partial^2 L_c}{\partial x^2}(\overline{x},\,\overline{\lambda}) = \frac{\partial^2 L}{\partial x^2}(\overline{x},\,\overline{\lambda}) + c(F'(\overline{x}))^{\mathrm{T}} F'(\overline{x})$$

положительно определена, т. е. в стационарной точке $\overline{x}$ задачи

$$L_c(x,\,\overline{\lambda}) \to \min, \quad x \in \mathbf{R}^n,$$

выполнено сформулированное в теореме 1.2.5 достаточное условие второго порядка оптимальности.

Итерация *метода модифицированных функций Лагранжа* состоит в решении для данного $c > 0$ и данного приближения λ к $\overline{\lambda}$ задачи безусловной оптимизации

$$L_c(x,\,\lambda) \to \min, \quad x \in \mathbf{R}^n, \tag{27}$$

с последующим пересчетом значения λ по специальному правилу. Происхождение этого правила несколько проясняет формула (12) (если принять во внимание, что эта формула получена для чистого метода квадратичного штрафа, а не для модифицированной функции Лагранжа), а также задача 9 ниже. Что же касается величины c, то в принципе она может оставаться одной и той же на всех итерациях; во всяком случае ее бесконечное увеличение не предполагается.

Алгоритм 3. Выбираем монотонно неубывающую последовательность $\{c_k\} \subset \mathbf{R}_+$ и точку $\lambda^0 \in \mathbf{R}^l$ и полагаем $k = 0$.

1. Вычисляем $x^k \in \mathbf{R}^n$ как стационарную точку задачи (27) с целевой функцией, задаваемой формулой (25) при $c = c_k$ и $\lambda = \lambda^k$.

2. Полагаем $\lambda^{k+1} = \lambda^k + c_k F(x^k)$.

3. Увеличиваем номер шага k на 1 и переходим к п. 1.

Существуют также варианты метода, в которых последовательность $\{c_k\}$ не фиксируется заранее, а строится по ходу вычислительного процесса в зависимости от получаемой информации. Об эвристических правилах выбора $\{c_k\}$ см. [6, 41].

Задача 9. Показать, что если для некоторого числа c и некоторого $\lambda \in \mathbf{R}^l$ точка $x \in \mathbf{R}^n$ является стационарной в задаче (27), причем $\operatorname{rank} F'(x) = l$, то

$$\lambda + cF(x) = -(F'(x)(F'(x))^{\mathrm{T}})^{-1} F'(x) f'(x)$$

(ср. с задачей 1).

Теорема 5. *Пусть выполнены условия теоремы* 1.

Тогда существуют окрестность U *точки* $\overline{x}$ *и числа* $\overline{c} \geqslant 0$, $\delta > 0$ *и* $M > 0$ *такие, что*:

а) *для любой пары* $(\lambda, c) \in \Delta(\overline{c}, \delta)$, *где*

$$\Delta(\overline{c}, \delta) = \{(\lambda, c) \in \mathbf{R}^l \times \mathbf{R} \mid |\lambda - \overline{\lambda}| < \delta c,\ c > \overline{c}\},$$

задача (27) *с целевой функцией, задаваемой формулой* (25), *имеет в* U *единственную стационарную точку* $x(\lambda, c)$, *причем*

$$\begin{aligned} |x(\lambda, c) - \overline{x}| &\leqslant M \frac{|\lambda - \overline{\lambda}|}{c}, \\ |\lambda + cF(x(\lambda, c)) - \overline{\lambda}| &\leqslant M \frac{|\lambda - \overline{\lambda}|}{c}; \end{aligned} \tag{28}$$

б) *существует число* $\overline{\delta} \in (0, \delta]$ *такое, что для любой монотонно неубывающей последовательности* $\{c_k\} \subset \mathbf{R}_+$ *и любой точки* $\lambda^0 \in \mathbf{R}^l$, *для которых выполнено*

$$(\lambda^0, c_0) \in \Delta(\overline{c}, \overline{\delta}), \tag{29}$$

формула

$$\lambda^{k+1} = \lambda^k + c_k F(x(\lambda^k, c_k)), \quad k = 0, 1, \ldots,$$

корректно определяет последовательность $\{\lambda^k\}$ (*в том смысле, что* $\{(\lambda^k, c_k)\} \subset \Delta(\overline{c}, \delta)$, *т.е. для всякого* k *точка* $x(\lambda^k, c_k)$ *корректно определена*), *причем* $\{\lambda^k\} \to \overline{\lambda}$, $\{x(\lambda^k, c_k)\} \to \overline{x}$ $(k \to \infty)$. *Скорость сходимости последовательности* $\{\lambda^k\}$ *линейная, а если* $c_k \to \infty$ $(k \to \infty)$, *то сверхлинейная.*

Полное доказательство этой теоремы выходит за рамки настоящего курса; ограничимся лишь некоторыми комментариями. Прежде всего заметим, что при $\lambda = 0$ утверждение а) совпадает с утверждением теоремы 4. Доказательство а) строится по той же схеме, что

и доказательство теоремы 4, однако требует несколько более тонких рассуждений, чем простое применение классической теоремы о неявной функции. При доказательстве утверждения б) основная трудность состоит именно в том, чтобы установить существования числа $\overline{\delta} \in (0, \delta]$ такого, что для любой пары (λ^0, c_0), удовлетворяющей (29), последовательность $\{(\lambda^k, c_k)\}$ не покидает множества $\Delta(\overline{c}, \delta)$. Если это установлено, то в силу (28) $\forall\, k$ имеем

$$|x(\lambda^k, c_k) - \overline{x}| \leqslant \frac{M}{c_k} |\lambda^k - \overline{\lambda}| \leqslant \frac{M}{\overline{c}} |\lambda^k - \overline{\lambda}|,$$

$$|\lambda^{k+1} - \overline{\lambda}| \leqslant \frac{M}{c_k} |\lambda^k - \overline{\lambda}| \leqslant \frac{M}{\overline{c}} |\lambda^k - \overline{\lambda}|,$$

и если число $\overline{c}$ выбрано достаточно большим, то последовательность $\{\lambda_k\}$ сходится к $\overline{\lambda}$ с линейной скоростью, а если $\{c_k\} \to \infty$ $(k \to \infty)$, то со сверхлинейной. Кроме того, последовательность $\{x(\lambda^k, c_k)\}$ сходится к $\overline{x}$, причем скорость ее сходимости не ниже, чем скорость сходимости $\{\lambda_k\}$, и во всяком случае не ниже геометрической. Полное доказательство можно найти в [6, 14].

Согласно теореме 5, если в алгоритме 3 начальные значения c_0 и λ^0 связаны включением (29), а x^k выбирается как стационарная точка задачи (27), ближайшая к искомому решению $\overline{x}$ задачи (1), (2), то траектория $\{(x^k, \lambda^k)\}$ алгоритма 3 корректно определяется и сходится к $(\overline{x}, \overline{\lambda})$ (о скорости сходимости см. выше). Для выбора «удачной» стационарной точки x^k на практике используют стандартную схему, уже обсуждавшуюся в связи с методом квадратичного штрафа: при каждом $k = 1, 2, \ldots$ в качестве начального приближения для используемого метода безусловной оптимизации берется точка x^{k-1}. Обычно генерируемая таким образом последовательность $\{x^k\}$ остается в окрестности одного локального решения задачи (1), (2), и этого оказывается достаточно для сходимости метода.

Важно отметить, что как бы далеко $\lambda^0 \in \mathbf{R}^l$ не отстояло от $\overline{\lambda}$, условие (29) будет выполняться, если число c_0 достаточно велико. Иными словами, не обязательно обеспечивать хорошее начальное приближения к множителю Лагранжа, но за грубость такого приближение приходится платить выбором большого начального значения параметра штрафа.

Задача 10. Доказать следующее дополнение утверждения а) теоремы 5: для любой пары $(\lambda, c) \in \Delta(\overline{c}, \delta)$, если число c достаточно велико, то матрица $\dfrac{\partial^2 L_c}{\partial x^2}(x(\lambda, c), \lambda)$ положительно определена, т. е. в стационарной точке $x(\lambda, c)$ задачи (27) выполнено сформулированное в теореме 1.2.5 достаточное условие второго порядка оптимальности (ср. с утверждением б) из задачи 4 и задачей 8).

С модифицированными функциями Лагранжа связаны так называемые *точные гладкие штрафные функции*: к функции L_c добавляется дополнительное гладкое слагаемое, штрафующее за нарушение необходимых условий первого порядка оптимальности, после чего полученная функция минимизируется по совокупности прямой и двойственной переменных. Введем, например, семейство функций

$$\varphi_{c_1,c_2}\colon \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R},$$

$$\varphi_{c_1,c_2}(x,\lambda) = L_{c_1}(x,\lambda) + \frac{c_2}{2}\left|\frac{\partial L}{\partial x}(x,\lambda)\right|^2 =$$

$$= L(x,\lambda) + \frac{c_1}{2}|F(x)|^2 + \frac{c_2}{2}\left|\frac{\partial L}{\partial x}(x,\lambda)\right|^2, \quad (30)$$

и для $c_1,\ c_2 > 0$ будем рассматривать задачу

$$\varphi_{c_1,c_2}(x,\lambda) \to \min, \quad (x,\lambda) \in \mathbf{R}^n \times \mathbf{R}^l \quad (31)$$

(ср. с задачей (9)). Легко видеть, что любое решение системы Лагранжа задачи (1), (2) является стационарной точкой задачи (31), а при определенном согласовании параметров c_1 и c_2 оказывается верным и обратное.

Лемма 1. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображение* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *дважды дифференцируемы в некоторой окрестности точки* $\widehat{x} \in \mathbf{R}^n$, *причем их вторые производные непрерывны в этой точке. Пусть* $\operatorname{rank} F'(\widehat{x}) = l$.

Тогда для произвольного $\widehat{\lambda} \in \mathbf{R}^l$ *найдется число* $\overline{c}_2 > 0$ *такое, что если* $c_2 \in (0, \overline{c}_2]$, *то окрестность* U *точки* $(\widehat{x}, \widehat{\lambda})$ *и число* $\overline{c}_1(c_2) \geqslant 0$ *могут быть выбраны таким образом, чтобы для любого* $c_1 > \overline{c}_1(c_2)$ *пара* $(\overline{x}, \overline{\lambda}) \in U$ *была стационарной точкой задачи* (31) *с целевой функцией, задаваемой формулой* (30), *в том и только том случае, когда* $\overline{x}$ *— стационарная точка задачи* (1), (2), *а* $\overline{\lambda}$ *— отвечающий ей множитель Лагранжа.*

Доказательство. Выберем число $\overline{c}_2 > 0$ настолько малым, что матрица

$$E^n + c_2 \frac{\partial^2 L}{\partial x^2}(\widehat{x}, \widehat{\lambda})$$

положительно определена $\forall\, c_2 \in (0, \overline{c}_2]$. Далее, фиксируем $c_2 \in (0, \overline{c}_2]$ и выберем число $\overline{c}_1(c_2) \geqslant 0$ настолько большим, что матрица

$$c_1 c_2 F'(\widehat{x})\left(E^n + c_2 \frac{\partial^2 L}{\partial x^2}(\widehat{x}, \widehat{\lambda})\right)^{-1}(F'(\widehat{x}))^{\mathrm{T}} - E^l$$

положительно определена $\forall\, c_1 > \overline{c}_1(c_2)$. Если c_1 и c_2 удовлетворяют указанным условиям, то, как легко убедиться, для любой пары $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$, достаточно близкой к $(\widehat{x}, \widehat{\lambda})$, матрица

$$\Lambda_{c_1,c_2}(x,\lambda)=\begin{pmatrix} E^n+c_2\dfrac{\partial^2 L}{\partial x^2}(x,\lambda) & c_1(F'(x))^{\mathrm{T}} \\ c_2F'(x) & E^l \end{pmatrix}\in \mathbf{R}(l,n)$$

невырождена (нужно показать невырожденность $\Lambda_{c_1,c_2}(\widehat{x},\widehat{\lambda})$ и применить теорему о малом возмущении невырожденной матрицы). Прямым вычислением получаем, что

$$\varphi'_{c_1,c_2}(x,\lambda)=\Lambda_{c_1,c_2}(x,\lambda)\begin{pmatrix}\dfrac{\partial L}{\partial x}(x,\lambda)\\ F(x)\end{pmatrix},$$

т. е. точка $(\overline{x},\overline{\lambda})$, достаточно близкая к $(\widehat{x},\widehat{\lambda})$, может быть стационарной в задаче (31) лишь в том случае, когда $(\overline{x},\overline{\lambda})$ — решение системы Лагранжа задачи (1), (2). □

Теорема 6. *Пусть выполнены условия теоремы* 1.

Тогда:

а) *если функция* f *и отображение* F *трижды дифференцируемы в точке* $\overline{x}$, *то для всякого* $c_2>0$ *найдется число* $\overline{c}_1(c_2)\geqslant 0$ *такое, что для любого* $c_1>\overline{c}_1(c_2)$ *матрица* $\varphi''_{c_1,c_2}(\overline{x},\overline{\lambda})$ *положительно определена* (*где функция* φ_{c_1,c_2} *задана формулой* (30)), *т.е. в стационарной точке* $(\overline{x},\overline{\lambda})$ *задачи* (31) *выполнено сформулированное в теореме* 1.2.5 *достаточное условие второго порядка оптимальности*;

б) *существует число* $\overline{c}_2>0$ *такое, что если* $c_2\in(0,\overline{c}_2]$, *то окрестность* U *точки* $(\overline{x},\overline{\lambda})$ *и число* $\overline{c}_1(c_2)\geqslant 0$ *могут быть выбраны таким образом, чтобы для любого* $c_1>\overline{c}_1(c_2)$ *задача* (31) *имела в* U *единственную стационарную точку* $(\overline{x},\overline{\lambda})$.

Доказательство. Для доказательства а) нужно вычислить второй дифференциал функции φ_{c_1,c_2} в точке $(\overline{x},\overline{\lambda})$ и непосредственно убедиться в его положительной определенности при произвольном $c_2>0$ и достаточно большом c_1.

Справедливость б) следует из леммы 1 и изолированности решения $(\overline{x},\overline{\lambda})$ системы Лагранжа задачи (1), (2). Последнее же является очевидным следствием классической теоремы о неявной функции (теорема 1.3.1) и установленной при доказательстве теоремы 1 невырожденности матрицы $L''(\overline{x},\overline{\lambda})$ в сделанных предположениях. □

Задача 11. Доказать утверждение а) теоремы 6.

Недостатком введенной гладкой точной штрафной функции является необходимость согласовывать значения параметров штрафов c_1 и c_2. Существует ряд возможностей избавиться от этого недостатка за счет некоторого усложнения штрафной функции. Например, зафиксируем гладкое отображение $C\colon \mathbf{R}^n\to\mathbf{R}(l,n)$, введем семейство

функций

$$\varphi_c\colon \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R},$$

$$\varphi_c(x,\,\lambda) = L_c(x,\,\lambda) + \frac{1}{2}\left|C(x)\,\frac{\partial L}{\partial x}\,(x,\,\lambda)\right|^2 =$$

$$= L(x,\,\lambda) + \frac{c}{2}\,|F(x)|^2 + \frac{1}{2}\left|C(x)\,\frac{\partial L}{\partial x}\,(x,\,\lambda)\right|^2, \quad (32)$$

и для $c > 0$ будем рассматривать задачу

$$\varphi_c(x,\,\lambda) \to \min, \quad (x,\,\lambda) \in \mathbf{R}^n \times \mathbf{R}^l. \quad (33)$$

Как и для рассмотренной выше штрафной функции $\varphi_{c_1,\,c_2}$, любое решение системы Лагранжа задачи (1), (2) является стационарной точкой задачи (33).

Лемма 2. *Пусть в дополнение к условиям леммы* 1 *отображение* $C\colon \mathbf{R}^n \to \mathbf{R}(l, n)$ *дифференцируемо в некоторой окрестности точки* $\widehat{x} \in \mathbf{R}^n$, *причем его производная непрерывна в этой точке и* $\det\left(C(\widehat{x})(F'(\widehat{x}))^{\mathrm{T}}\right) \neq 0$.

Тогда для произвольного $\widehat{\lambda} \in \mathbf{R}^l$ *существуют окрестность* U *точки* $(\widehat{x}, \widehat{\lambda})$ *и число* $\overline{c} \geqslant 0$ *такие, что для любого* $c > \overline{c}$ *пара* $(\overline{x}, \overline{\lambda}) \in U$ *является стационарной точкой задачи* (33) *с целевой функцией, задаваемой формулой* (32), *в том, и только том случае, когда* $\overline{x}$ *— стационарная точка задачи* (1), (2), *а* $\overline{\lambda}$ *— отвечающий ей множитель Лагранжа.*

Доказательство. Схема доказательства та же, что и для леммы 1. Все сводится к проверке того, что для любого достаточно большого c матрица

$$\Lambda_c(\widehat{x},\,\widehat{\lambda}) = \begin{pmatrix} E^n + A(\widehat{x},\,\widehat{\lambda})C(x) & c(F'(\widehat{x}))^{\mathrm{T}} \\ F'(\widehat{x})(C(\widehat{x}))^{\mathrm{T}}C(\widehat{x}) & E^l \end{pmatrix} \in \mathbf{R}(l,\,n)$$

невырождена, где

$$A(\widehat{x},\,\widehat{\lambda}) = \left(\left(C(x)\,\frac{\partial L}{\partial x}\,(x,\,\widehat{\lambda})\right)'\Big|_{x=\widehat{x}}\right)^{\mathrm{T}} \in \mathbf{R}(n,\,l).$$

Рассмотрим произвольный элемент $\widetilde{u} = (\widetilde{x},\,\widetilde{\lambda}) \in \ker \Lambda_c(\widehat{x},\,\widehat{\lambda})$; тогда

$$(E^n + A(\widehat{x},\,\widehat{\lambda})C(x))\widetilde{x} + c(F'(\widehat{x}))^{\mathrm{T}}\widetilde{\lambda} = 0, \quad (34)$$

$$F'(\widehat{x})(C(\widehat{x}))^{\mathrm{T}}C(\widehat{x})\widetilde{x} + \widetilde{\lambda} = 0.$$

Из последнего равенства выразим $C(\widehat{x})\widetilde{x}$:

$$C(\widehat{x})\widetilde{x} = -(F'(\widehat{x})(C(\widehat{x}))^{\mathrm{T}})^{-1}\widetilde{\lambda}. \quad (35)$$

Тогда из (34) имеем

$$(-(E^l + C(\widehat{x})A(\widehat{x}, \widehat{\lambda})C(\widehat{x}))(F'(\widehat{x})(C(\widehat{x}))^{\mathrm{T}})^{-1} + cC(\widehat{x})(F'(\widehat{x}))^{\mathrm{T}})\widetilde{\lambda} =$$
$$= C(\widehat{x})(\widetilde{x} + A(\widehat{x}, \widehat{\lambda})C(x)\widetilde{x} + c(F'(\widehat{x}))^{\mathrm{T}}\widetilde{\lambda}) = 0.$$

Выберем число $\overline{c} \geqslant 0$ настолько большим, что матрица в левой части последнего равенства невырождена $\forall\, c > \overline{c}$. Для таких c имеем $\widetilde{\lambda} = 0$, и в силу (35) $C(\widehat{x})\widetilde{x} = 0$. Возвращаясь к (34), приходим к равенству $\widetilde{x} = 0$, т. е. $\widetilde{u} = 0$, что и требовалось доказать. □

Теорема 7. *Пусть в дополнение к условиям теоремы* 1 *отображение* $C\colon \mathbf{R}^n \to \mathbf{R}(l, n)$ *дифференцируемо в некоторой окрестности точки* $\overline{x}$, *причем его производная непрерывна в этой точке и* $\det(C(\overline{x})(F'(\overline{x}))^{\mathrm{T}}) \neq 0$.

Тогда существуют окрестность U *точки* $(\overline{x}, \overline{\lambda})$ *и число* $\overline{c} \geqslant 0$ *такие, что для любого* $c > \overline{c}$ *задача* (33) *с целевой функцией, задаваемой формулой* (32), *имеет в* U *единственную стационарную точку* $(\overline{x}, \overline{\lambda})$, *причем если функция* f *и отображение* F *трижды дифференцируемы в точке* $\overline{x}$, *то матрица* $\varphi_c''(\overline{x}, \overline{\lambda})$ *положительно определена, т.е. в стационарной точке* $(\overline{x}, \overline{\lambda})$ *выполнено сформулированное в теореме* 1.2.5 *достаточное условие второго порядка оптимальности.*

Доказательство этой теоремы аналогично доказательству теоремы 6 (см. также задачу 11), только вместо леммы 1 нужно сослаться на лемму 2. Подчеркнем, что условие невырожденности матрицы $C(\overline{x})(F'(\overline{x}))^{\mathrm{T}}$ подразумевает выполнение условия регулярности ограничений в точке $\overline{x}$.

Зависимость рассмотренных точных гладких штрафных функций не только от прямой переменной, но и от двойственной, является недостатком такого подхода, впрочем, вполне устранимым за счет дальнейшего усложнения штрафной функции. Дело в том, что функции, введенные в (30) и (32), являются квадратичными по λ при фиксированном x, поэтому их можно пытаться явно минимизировать по λ при каждом x, тем самым избавляясь от «лишних» переменных.

Задача 12. Зададим множество $\mathcal{R} \subset \mathbf{R}^n$ так же, как в задаче 1, и положим

$$C(x) = (F'(x)(F'(x))^{\mathrm{T}})^{-1}F'(x), \quad \lambda(x) = -C(x)f'(x), \quad x \in \mathcal{R}.$$

Показать, что для введенного в (32) семейства функций справедливо

$$\min_{\lambda \in \mathbf{R}^l} \varphi_{c+1}(x, \lambda) = \varphi_{c+1}(x, \lambda(x)) = L_c(x, \lambda(x)) \quad \forall\, x \in \mathcal{R}$$

для любого c.

Задача 13. Для задач безусловной минимизации (31) и (33) с целевыми функциями, задаваемыми формулами (30) и (32) соответственно, а также для задачи

$$L_c(x,\, \lambda(x)) \to \min, \quad x \in \mathcal{R}$$

(в обозначениях задачи 12), построить и обосновать неточные методы Ньютона, получаемые отбрасыванием слагаемых, содержащих третьи производные функции f и отображения F.

§ 4.4. Последовательное квадратичное программирование

Как следует из их названия, методы последовательного квадратичного программирования (общепринятая аббревиатура — SQP, от английского *Sequential quadratic programming*) состоят в последовательном решении задач квадратичного программирования, аппроксимирующих данную задачу оптимизации. Правильно выбранная задача квадратичного программирования оказывается достаточно адекватной локальной аппроксимацией исходной задачи. В то же время квадратичная задача достаточно проста, и для нее существуют эффективные (в том числе конечные) методы решения; см. § 7.3. Таким образом, указанный подход имеет практический смысл, в отличие, например, от использования аппроксимаций более высокого порядка.

С другой стороны, методы SQP могут рассматриваться как результат распространения фундаментальной идеи ньютоновских методов на задачи оптимизации с функциональными ограничениями. Например, в случае задачи с чистыми ограничениями-равенствами методы SQP суть не более чем специальные реализации ньютоновских методов решения системы Лагранжа, что и будет продемонстрировано ниже. Затем эта идея распространяется на задачи со смешанными ограничениями.

На сегодняшний день методы SQP входят в число наиболее эффективных оптимизационных методов общего назначения. При выборе метода решения той или иной прикладной задачи оптимизации, не обладающей какой-то специальной структурой, часто останавливаются на методах именно этого типа.

4.4.1. Ограничения-равенства. Предполагая двукратную дифференцируемость функции $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображения $F\colon \mathbf{R}^n \to \mathbf{R}^l$ на всем $\mathbf{R}^n$, будем рассматривать задачу

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in \mathbf{R}^n \mid F(x) = 0\}. \tag{2}$$

Пусть $x^k \in \mathbf{R}^n$ — текущее приближение к искомой стационарной точке $\overline{x}$ задачи (1), (2). Тогда вблизи x^k эту задачу можно аппроксимировать задачей квадратичного программирования

$$f(x^k)+\langle f'(x^k),\, x-x^k\rangle+\frac{1}{2}\,\langle H_k(x-x^k),\, x-x^k\rangle \to \min, \quad x \in D_k, \tag{3}$$

$$D_k = \{x \in \mathbf{R}^n \,\big|\, F(x^k) + F'(x^k)(x - x^k) = 0\}, \tag{4}$$

где $H_k \in \mathbf{R}(n,\, n)$ — некоторая симметрическая матрица.

На первый взгляд может показаться, что следует выбирать $H_k = f''(x^k)$, т. е. использовать в качестве целевой функции квадратичной задачи чистую аппроксимацию второго порядка функции f (ср. с условным методом Ньютона из п. 4.1.3). Однако такой интуитивно разумный выбор на самом деле неадекватен. Дело в том, что используемая линейная аппроксимация ограничений не может передать важную информацию о кривизне соответствующих поверхностей. С другой стороны, очевидно, что включение членов второго порядка в аппроксимацию ограничений крайне нежелательно, поскольку соответствующая аппроксимирующая задача перестанет быть задачей квадратичного программирования, будет немногим проще, чем исходная задача, и метод, требующий решения на каждом шаге подобной подзадачи, вряд ли может быть эффективен, пока не разработаны специальные методы решения таких подзадач. В последнее время в оптимизационной литературе стали появляться публикации по методам, подзадачи которых имеют квадратичные ограничения, но эти разработки пока далеки от завершения. К счастью, необходимую информацию о членах второго порядка в аппроксимации ограничений можно передать не через ограничения вспомогательной задачи, а через ее целевую функцию, полагая

$$H_k = \frac{\partial^2 L}{\partial x^2}\,(x^k,\, \lambda^k), \tag{5}$$

где $\lambda^k \in \mathbf{R}^l$ — некоторая аппроксимация множителя Лагранжа $\overline{\lambda}$, отвечающего искомой стационарной точке $\overline{x}$, а

$$L\colon \mathbf{R}^n \times \mathbf{R}^l \to \mathbf{R}, \quad L(x,\, \lambda) = f(x) + \langle \lambda,\, F(x)\rangle,$$

— функция Лагранжа задачи (1), (2). Заметим, что при таком выборе целевая функция задачи (3), (4) есть чистая квадратичная аппроксимация функции $L(\cdot,\, \lambda^k)$ вблизи текущей точки x^k.

Другой (более формальной) мотивировкой такого выбора служит интерпретация вспомогательной задачи (3), (4) как реализации рассмотренной в п. 4.3.1 ньютоновской итерации для системы Лагранжа исходной задачи (1), (2). Напомним, что система Лагранжа имеет вид

$$L'(x,\, \lambda) = 0, \tag{6}$$

а ньютоновская итерация для нее состоит в решении линейной системы

$$\frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x - x^k) + (F'(x^k))^{\mathrm{T}}(\lambda - \lambda^k) = \\ = -f'(x^k) - (F'(x^k))^{\mathrm{T}}\lambda^k, \quad (7)$$

$$F'(x^k)(x - x^k) = -F(x^k) \quad (8)$$

относительно $(x, \lambda) \in \mathbf{R}^n \times \mathbf{R}^l$, где $(x^k, \lambda^k) \in \mathbf{R}^n \times \mathbf{R}^l$ — текущее приближение к решению (6).

Действительно, пусть $x^{k+1} \in \mathbf{R}^n$ — стационарная точка задачи (3), (4), а $y^{k+1} \in \mathbf{R}^l$ — отвечающий ей множитель Лагранжа. Это значит, что пара (x^{k+1}, y^{k+1}) является решением системы уравнений

$$f'(x^k) + H_k(x - x^k) + (F'(x^k))^{\mathrm{T}}y = 0,$$

$$F(x^k) + F'(x^k)(x - x^k) = 0$$

относительно $(x, y) \in \mathbf{R}^n \times \mathbf{R}^l$. Сравнивая последнюю систему с системой (7), (8), убеждаемся, что они совпадают, если матрица H_k выбрана согласно (5). Таким образом, этот выбор согласуется с фундаментальной идеей ньютоновских методов.

Опишем *метод последовательного квадратичного программирования* (SQP) для задачи с ограничениями-равенствами.

Алгоритм 1. Выбираем $(x^0, \lambda^0) \in \mathbf{R}^n \times \mathbf{R}^l$ и полагаем $k = 0$.

1. Вычисляем $x^{k+1} \in \mathbf{R}^n$ как стационарную точку задачи (3), (4), где H_k задается согласно (5). Вычисляем $y^{k+1} \in \mathbf{R}^l$ как отвечающий стационарной точке x^{k+1} задачи (3), (4) множитель Лагранжа.

2. Полагаем $\lambda^{k+1} = y^{k+1}$.

3. Увеличиваем номер шага k на 1 и переходим к п. 1.

В силу сказанного выше при любом выборе начального приближения (x^0, λ^0) алгоритм 1 генерирует ту же траекторию $\{(x^k, \lambda^k)\}$, что и метод Ньютона для системы Лагранжа (6), рассмотренный в п. 4.3.1, поэтому отдельно исследовать (локальную) сходимость метода SQP в данном случае нет необходимости. В частности, если вторые производные функции f и отображения F непрерывны в точке $\overline{x}$ и в этой точке выполнены условие регулярности ограничений и сформулированное в теореме 1.3.7 достаточное условие второго порядка оптимальности, то в соответствии с теоремой 4.3.1 метод SQP локально сходится к $(\overline{x}, \overline{\lambda})$ со сверхлинейной скоростью. Более того, если

вторые производные f и F непрерывны по Липшицу в окрестности $\overline{x}$, то скорость сходимости квадратичная.

Заметим, однако, что как реализация метода Ньютона для системы Лагранжа метод SQP вряд ли особенно полезен, во всяком случае при локальных рассмотрениях. Эта реализация аналогична рассмотренной в п. 3.2.2 оптимизационной трактовке метода Ньютона для задачи безусловной оптимизации и может быть полезна при глобализации сходимости метода. Однако главное значение метода SQP для задачи с ограничениями-равенствами состоит в том, что он естественным образом приводит к идее соответствующего метода для задачи со смешанными ограничениями, что и будет продемонстрировано ниже.

Задача 1. Установить взаимосвязь задачи (3), (4), в которой матрица H_k задается согласно (5), с аналогичной задачей, в которой

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k) + c_k(F'(x^k))^{\mathrm{T}} F'(x^k),$$

где $c_k > 0$ — заданное число. Каковы возможные преимущества второго выбора?

Задача 2. Рассмотреть для задачи (1), (2) *модифицированный метод проекции градиента*

$$x^{k+1} = \pi_{D_k}(x^k - \alpha_k f'(x^k)), \quad k = 0, 1, \ldots,$$

где $\alpha_k > 0$ — параметр метода, а D_k введено в (4). Установить взаимосвязь метода с SQP (при соответствующем выборе H_k). Исследовать локальную сходимость метода.

4.4.2. Смешанные ограничения. Пусть теперь дополнительно отображение $G\colon \mathbf{R}^n \to \mathbf{R}^m$ дважды дифференцируемо на $\mathbf{R}^n$. Переходим к рассмотрению задачи оптимизации со смешанными ограничениями (равенствами и неравенствами):

$$f(x) \to \min, \quad x \in D, \tag{9}$$

$$D = \{x \in \mathbf{R}^n \mid F(x) = 0,\ G(x) \leqslant 0\}. \tag{10}$$

Пусть $x^k \in \mathbf{R}^n$ — текущее приближение к искомой стационарной точке $\overline{x}$ задачи (9), (10). Естественным аналогом аппроксимирующей задачи (3), (4) здесь является следующая задача квадратичного программирования:

$$\langle f'(x^k),\ d\rangle + \frac{1}{2}\langle H_k d,\ d\rangle \to \min, \quad d \in D_k, \tag{11}$$

$$D_k = \{d \in \mathbf{R}^n \mid F(x^k) + F'(x^k)d = 0,\ G(x^k) + G'(x^k)d \leqslant 0\}. \tag{12}$$

Для задачи со смешанными ограничениями *метод последовательного квадратичного программирования* (SQP) состоит в следующем.

Алгоритм 2. Выбираем $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ и полагаем $k = 0$.

1. Вычисляем $d^k \in \mathbf{R}^n$ как стационарную точку задачи (11), (12), где $H_k \in \mathbf{R}(n, n)$ — симметрическая матрица. Вычисляем $y^k \in \mathbf{R}^l$ и $z^k \in \mathbf{R}^m_+$ как отвечающие стационарной точке d^k задачи (11), (12) множители Лагранжа.

2. Полагаем $x^{k+1} = x^k + d^k$, $\lambda^{k+1} = y^k$, $\mu^{k+1} = z^k$.

3. Увеличиваем номер шага k на 1 и переходим к п. 1.

В соответствии со сказанным выше можно ожидать, что к высокой скорости сходимости приведет следующий выбор матрицы H_k:

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k), \tag{13}$$

где

$$L\colon \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}, \quad L(x, \lambda, \mu) = f(x) + \langle \lambda, F(x)\rangle + \langle \mu, G(x)\rangle,$$

— функция Лагранжа задачи (9), (10).

Заметим, что если $D_k \neq \varnothing$ и матрица H_k положительно определена, то задача (11), (12) имеет единственное решение, и это решение совпадает с ее единственной стационарной точкой (поскольку целевая функция этой задачи сильно выпукла). Однако в естественных предположениях, таких, как вводившиеся выше условия регулярности ограничений задачи (9), (10) в точке $\overline{x}$ и достаточное условие второго порядка оптимальности этой точки с множителями $\overline{\lambda}$ и $\overline{\mu}$, положительную определенность матрицы H_k, выбираемой согласно (13), гарантировать нельзя даже при x^k, близком к $\overline{x}$, а λ^k и μ^k, близких к $\overline{\lambda}$ и $\overline{\mu}$ соответственно (ср. с мотивировкой введения модифицированных функций Лагранжа в п. 4.3.3, а также задачей 1). Поэтому существование у задачи (11), (12) стационарной точки не является автоматическим и должно проверяться в рамках локального анализа алгоритма. Более того, даже если такая стационарная точка существует, она вовсе не обязательно единственна. В связи с последним при проведении локального анализа предполагается, что в качестве d^k берется стационарная точка задачи (11), (12), имеющая минимальную норму.

При практической реализации алгоритма 1 последнее требование часто просто игнорируют. А именно, к задаче (11), (12) часто применяют общие методы решения задач квадратичного программирования, позволяющие, наряду с решением d^k такой задачи, вычислить отвечающие этому решению множители Лагранжа y^k и z^k; см. § 7.3. Например, пользователи пакета `Matlab` часто используют те средства решения квадратичных задач, которые имеются в этом пакете, и во многих случаях с успехом. Заметим, однако, что, в силу

указанных выше причин, если матрица H_k не предполагается положительно определенной, то более правильно использовать специальные прямодвойственные методы, ориентированные на поиск стационарных точек задач квадратичного программирования, а не глобальных решений таких задач. О методах такого рода см. [45]. При такой реализации итерации метода SQP можно пытаться влиять на близость искомого d^k к 0, например, выбирая 0 в качестве начальной точки внутреннего процесса.

Теорема 1. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображения* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *и* $G\colon \mathbf{R}^n \to \mathbf{R}^m$ *дважды дифференцируемы в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$, *причем их вторые производные непрерывны в этой точке. Пусть в точке* $\overline{x}$ *выполнено условие линейной независимости, причем* $\overline{x}$ *— стационарная точка задачи* (9), (10), *а* $\overline{\lambda} \in \mathbf{R}^l$ *и* $\overline{\mu} \in \mathbf{R}^m_+$ *— однозначно отвечающие ей множители Лагранжа. Пусть, наконец, в точке* $\overline{x}$ *выполнено сформулированное в теореме* 1.4.5 *достаточное условие второго порядка оптимальности, а также условие строгой дополнительности.*

Тогда любое начальное приближение $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, *достаточно близкое к точке* $(\overline{x}, \overline{\lambda}, \overline{\mu})$, *корректно определяет сходящуюся к* $(\overline{x}, \overline{\lambda}, \overline{\mu})$ *траекторию алгоритма* 2, *в котором для каждого* $k = 0, 1, \ldots$ *матрица* H_k *выбирается согласно* (13), *а в качестве* d^k *берется стационарная точка задачи* (11), (12), *имеющая минимальную норму. Скорость сходимости сверхлинейная, а если вторые производные* f, F *и* G *непрерывны по Липшицу в окрестности точки* $\overline{x}$, *то квадратичная.*

Доказательство. Как обычно, через $g_i(\cdot)$, $i = 1, \ldots, m$, будем обозначать компоненты отображения G, а через $I(\overline{x}) = \{i = 1, \ldots, m \mid g_i(\overline{x}) = 0\}$ — множество индексов активных в точке $\overline{x}$ ограничений-неравенств задачи (9), (10). Положим $\Sigma = \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$. Прежде всего необходимо показать, что задача (11), (12) при $(x^k, \lambda^k, \mu^k) \in \Sigma$, достаточно близком к $(\overline{x}, \overline{\lambda}, \overline{\mu})$, имеет единственную стационарную точку d^k минимальной нормы, и ей отвечает единственная пара множителей Лагранжа (y^k, z^k).

Рассмотрим систему Каруша–Куна–Таккера для задачи (11), (12):

$$f'(x^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)d + (F'(x^k))^{\mathrm{T}} y + (G'(x^k))^{\mathrm{T}} z = 0, \tag{14}$$

$$F(x^k) + F'(x^k)d = 0, \tag{15}$$

$$G(x^k) + G'(x^k)d \leqslant 0, \quad z \geqslant 0, \tag{16}$$

$$z_i(g_i(x^k) + \langle g_i'(x^k), d\rangle) = 0, \quad i = 1, \ldots, m, \tag{17}$$

относительно $(d, y, z) \in \Sigma$. Сразу заметим, что если точка x^k достаточно близка к $\overline{x}$, а $d \in \mathbf{R}^n$ имеет достаточно малую норму, то

может существовать не более одной пары множителей $(y,\, z) \in \mathbf{R}^l \times \mathbf{R}^m$, таких, что точка $(d,\, y,\, z)$ удовлетворяет (14)–(17). Это следует из (14), условия линейной независимости, а также того факта, что из (17) при таких x^k и d вытекает

$$z_i = 0 \quad \forall\, i \in \{1, \ldots, m\} \setminus I(\overline{x}).$$

Введем отображение

$$\Phi\colon \Sigma \times \Sigma \to \Sigma,$$

$$\Phi(\sigma,\, u) = \begin{pmatrix} f'(x) + \dfrac{\partial^2 L}{\partial x^2}(x,\, \lambda, \mu)\, d + (F'(x))^{\mathrm{T}} y + (G'(x))^{\mathrm{T}} z \\ F(x) + F'(x) d \\ z_1 (g_1(x) + \langle g_1'(x),\, d \rangle) \\ \ldots \\ z_m (g_m(x) + \langle g_m'(x),\, d \rangle) \end{pmatrix},$$

$$\sigma = (x,\, \lambda,\, \mu), \qquad u = (d,\, y,\, z).$$

Система

$$\Phi(\sigma,\, u) = 0,$$

в которой $\sigma \in \Sigma$ играет роль параметра, отвечает уравнениям (14), (15), (17).

Согласно определению стационарной точки задачи (9), (10) и отвечающих этой точке множителей Лагранжа имеем $\Phi(\overline{\sigma},\, \overline{u}) = 0$, где $\overline{\sigma} = (\overline{x},\, \overline{\lambda},\, \overline{\mu})$, $\overline{u} = (0,\, \overline{\lambda},\, \overline{\mu})$. Кроме того,

$$\frac{\partial \Phi}{\partial u}(\overline{\sigma},\, \overline{u}) =$$

$$= \begin{pmatrix} \dfrac{\partial^2 L}{\partial x^2}(\overline{x},\, \overline{\lambda},\, \overline{\mu}) & (F'(\overline{x}))^{\mathrm{T}} & & (G'(\overline{x}))^{\mathrm{T}} & & \\ F'(\overline{x}) & 0 & & 0 & & \\ \overline{\mu}_1 g_1'(\overline{x}) & 0 & g_1(\overline{x}) & 0 & \ldots & 0 \\ \overline{\mu}_2 g_2'(\overline{x}) & 0 & 0 & g_2(\overline{x}) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \overline{\mu}_m g_m'(\overline{x}) & 0 & 0 & 0 & \ldots & g_m(\overline{x}) \end{pmatrix}.$$

Рассмотрим произвольный элемент $\widetilde{u} = (\widetilde{d},\, \widetilde{y},\, \widetilde{z}) \in \ker \dfrac{\partial \Phi}{\partial u}(\overline{\sigma},\, \overline{u})$, т. е.

$$\frac{\partial^2 L}{\partial x^2}(\overline{x},\, \overline{\lambda},\, \overline{\mu})\, \widetilde{d} + (F'(\overline{x}))^{\mathrm{T}} \widetilde{y} + (G'(\overline{x}))^{\mathrm{T}} \widetilde{z} = 0, \tag{18}$$

$$F'(\overline{x})\, \widetilde{d} = 0, \tag{19}$$

$$\overline{\mu}_i \langle g_i'(\overline{x}),\, \widetilde{d} \rangle + g_i(\overline{x}) \widetilde{z}_i = 0, \quad i = 1, \ldots, m. \tag{20}$$

Из (20) и условия строгой дополнительности вытекает, что

$$\widetilde{z}_i = 0 \quad \forall\, i \in \{1, \ldots, m\} \setminus I(\overline{x}), \quad \langle g_i'(\overline{x}),\, \widetilde{d}\rangle = 0 \quad \forall\, i \in I(\overline{x}). \tag{21}$$

Кроме того, условие строгой дополнительности в точке $\overline{x}$ влечет следующую формулу для критического конуса задачи (9), (10) в точке $\overline{x}$:

$$K(\overline{x}) = \{h \in \ker F'(\overline{x}) \,\big|\, \langle g_i'(\overline{x}),\, h\rangle = 0 \quad \forall\, i \in I(\overline{x})\} \tag{22}$$

(см. лемму 1.4.4). Поэтому из (19) и (21) следует, что $\widetilde{d} \in K(\overline{x})$. Умножая левую и правую части (18) скалярно на $\widetilde{d}$ и используя (19) и (21), получим

$$0 = \left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\, \overline{\lambda},\, \overline{\mu})\widetilde{d},\, \widetilde{d}\right\rangle + \langle \widetilde{y},\, F'(\overline{x})\widetilde{d}\rangle + \langle \widetilde{z},\, G'(\overline{x})\widetilde{d}\rangle =$$

$$= \left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\, \overline{\lambda},\, \overline{\mu})\widetilde{d},\, \widetilde{d}\right\rangle + \sum_{i=1}^{m} \widetilde{z}_i\, \langle g_i'(\overline{x}),\, \widetilde{d}\rangle = \left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\, \overline{\lambda},\, \overline{\mu})\widetilde{d},\, \widetilde{d}\right\rangle.$$

Это означает, что $\widetilde{d} = 0$, поскольку иначе было бы нарушено достаточное условие второго порядка оптимальности в точке $\overline{x}$. Но тогда из (18) и (21) вытекает, что

$$(F'(\overline{x}))^{\mathrm{T}}\widetilde{y} + \sum_{i \in I(\overline{x})} \widetilde{z}_i\, g_i'(\overline{x}) = 0.$$

Условие линейной независимости в точке $\overline{x}$ означает, что последнее равенство возможно лишь при $\widetilde{y} = 0$ и $\widetilde{z}_i = 0 \quad \forall\, i \in I(\overline{x})$. Тем самым показано, что по необходимости $\widetilde{u} = (\widetilde{d},\, \widetilde{y},\, \widetilde{z}) = 0$, т. е. матрица $\dfrac{\partial \Phi}{\partial u}\,(\overline{\sigma},\, \overline{u})$ невырождена.

Применяя к отображению Φ в точке $(\overline{\sigma},\, \overline{u})$ классическую теорему о неявной функции (теорема 1.3.1), получаем существование таких окрестностей U точки $\overline{\sigma}$ и V точки $\overline{u}$ в Σ, для которых существует единственное непрерывное в точке $\overline{\sigma}$ отображение $\chi(\cdot) = (d(\cdot),\ y(\cdot),\ z(\cdot))\colon\ U \to \Sigma$, такое, что $\chi(U) \subset V$ и

$$\Phi(\sigma,\, \chi(\sigma)) = 0 \quad \forall\, \sigma \in U. \tag{23}$$

В частности, $d(\overline{\sigma}) = 0$, $\quad y(\overline{\sigma}) = \overline{\lambda}$, $\quad z(\overline{\sigma}) = \overline{\mu}$. Отсюда, из непрерывности $\chi(\cdot)$ в точке $\overline{\sigma}$ и условия строгой дополнительности вытекает, что если окрестность U достаточно мала, то для любого $\sigma = (x,\, \lambda,\, \mu) \in U$ имеет место

$$g_i(x) + \langle g_i'(x),\, d(\sigma)\rangle < 0 \quad \forall\, i \in \{1, \ldots, m\} \setminus I(\overline{x}), \tag{24}$$

$$z_i(\sigma) > 0 \quad \forall\, i \in I(\overline{x}). \tag{25}$$

С учетом определения отображения Φ соотношения (23)–(25) означают, что при $\sigma^k = (x^k,\, \lambda^k,\, \mu^k) \in U$ система (14)–(17) имеет

на V единственное решение $(d(\sigma^k),\, y(\sigma^k),\, z(\sigma^k))$. В силу сказанного выше о единственности множителей можно утверждать, что если окрестность U достаточно мала, то это решение единственно не только на V, но и на множестве $\widetilde{V} \times \mathbf{R}^l \times \mathbf{R}^m$, где $\widetilde{V}$ — некоторая окрестность точки 0 в $\mathbf{R}^n$. Отсюда следует, что $d^k = d(\sigma^k)$ — единственная стационарная точка задачи (11), (12), имеющая минимальную норму, а $(y^k,\, z^k) = (y(\sigma^k),\, z(\sigma^k))$ — единственная пара множителей Лагранжа, отвечающая этой стационарной точке.

Заметим, далее, что согласно (24), (25) при $(x^k,\, \lambda^k,\, \mu^k) \in U$ условия (16), (17) в системе (14)–(17), определяющей $(d^k,\, y^k,\, z^k)$, можно заменить парой условий

$$g_i(x^k) + \langle g_i'(x^k),\, d\rangle = 0, \quad i \in I(\overline{x}), \tag{26}$$

$$z_i = 0, \quad i \in \{1, \ldots, m\} \setminus I(\overline{x}). \tag{27}$$

Итерация алгоритма равносильна отысканию решения $(d^k,\, y^k,\, z^k)$ системы линейных уравнений (14), (15), (26), (27).

Рассмотрим задачу оптимизации с ограничениями-равенствами:

$$f(x) \to \min, \quad x \in \widetilde{D},$$

$$\widetilde{D} = \{x \in \mathbf{R}^n \,\big|\, F(x) = 0,\, g_i(x) = 0,\, i \in I(\overline{x})\}.$$

Легко видеть, что в условиях доказываемой теоремы точка $\overline{x}$ является локальным решением такой задачи, и в этой точке выполнено условие регулярности ограничений и сформулированное в теореме 1.3.7 достаточное условие второго порядка оптимальности. Отсюда следует, что если соответствующую систему Лагранжа формально дополнить уравнениями

$$\mu_i = 0, \quad i \in \{1, \ldots, m\} \setminus I(\overline{x}),$$

то полученная система уравнений относительно $(x,\, \lambda,\, \mu) \in \Sigma$ будет иметь регулярное решение $(\overline{x},\, \overline{\lambda},\, \overline{\mu})$ (в том смысле, что производная оператора такой системы в этом решении невырождена). Соответственно в силу теоремы 3.2.1 метод Ньютона для такой системы локально сходится к этому решению со сверхлинейной скоростью, а если вторые производные функции f и отображений F и G непрерывны по Липшицу в окрестности точки $\overline{x}$, то с квадратичной скоростью.

Нетрудно убедиться, что для текущей точки $(x^k,\, \lambda^k,\, \mu^k)$ итерация метода Ньютона состоит в переходе к точке $(x^k + d^k,\, y^k,\, z^k)$, где $(d^k,\, y^k,\, z^k)$ определяется как решение системы, состоящей из уравнений (15), (26), (27) и уравнения

$$\frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)d - \sum_{\substack{i=1 \\ i \notin I(\overline{x})}}^{m} \mu_i^k g_i''(x^k)d +$$

$$+(F'(x^k))^{\mathrm{T}}y+\sum_{i\in I(\overline{x})}z_i^k g_i'(x^k)=-f'(x^k), \tag{28}$$

которое отличается от (14) лишь наличием второго слагаемого в левой части. Но поскольку $\overline{\mu}_i=0 \quad \forall\, i\in\{1,\dots,m\}\setminus I(\overline{x})$, найдется число $M>0$, такое, что для любого $x\in\mathbf{R}^n$, достаточно близкого к $\overline{x}$, и любого $\mu\in\mathbf{R}^m$, достаточно близкого к $\overline{\mu}$,

$$\left\|\sum_{\substack{i=1\\ i\notin I(\overline{x})}}^{m}\mu_i g_i''(x)\right\|\leqslant\sum_{\substack{i=1\\ i\notin I(\overline{x})}}^{m}|\mu_i-\overline{\mu}_i|\,\|g_i''(x)\|\leqslant M|\mu-\overline{\mu}|.$$

Это значит, что алгоритм, итерация которого задается уравнениями (14), (15), (26), (27), может рассматриваться как неточный метод Ньютона, итерация которого задается уравнениями (15), (26)–(28). Согласно следствию 3.2.1 локальные свойства сходимости и скорость сходимости алгоритма полностью аналогичны соответствующим характеристикам метода Ньютона [1]. $\square$

Заметим, далее, что даже из квадратичной скорости сходимости траектории $\{(x^k,\,\lambda^k,\,\mu^k)\}$ к $(\overline{x},\,\overline{\lambda},\,\overline{\mu})$, вообще говоря, не вытекает даже сверхлинейная скорость сходимости последовательности $\{x^k\}$ к $\overline{x}$ (см. задачу 2.1.1). Вместе с тем на практике обычно наибольший интерес представляет быстрая сходимость метода именно в прямых переменных. Соответствующий результат и содержится в следующей теореме. Одновременно будет рассмотрена возможность выбирать матрицы H_k, удовлетворяющие (13) не точно, а приближенно. Тем самым будет рассмотрено расширение изложенной выше схемы SQP в духе теоремы Дэнниса–Морэ (теорема 3.2.2). О практических правилах построения H_k с нужными свойствами см. [50].

Теорема 2. *Пусть выполнены условия теоремы* 1, *кроме, возможно, условия строгой дополнительности. Пусть, кроме того, траектория* $\{(x^k,\,\lambda^k,\,\mu^k)\}$ *алгоритма* 2 *сходится к* $(\overline{x},\,\overline{\lambda},\,\overline{\mu})$.

Тогда скорость сходимости последовательности $\{x^k\}$ *к* $\overline{x}$ *является сверхлинейной в том и только том случае, когда*

$$\pi_{K(\overline{x})}\left(\left(H_k-\frac{\partial^2 L}{\partial x^2}\,(\overline{x},\,\overline{\lambda},\,\overline{\mu})\right)(x^{k+1}-x^k)\right)=o(|x^{k+1}-x^k|), \tag{29}$$

где

$$K(\overline{x})=\{h\in\ker F'(\overline{x})\,\big|\,\langle g_i'(\overline{x}),\,h\rangle\leqslant 0 \quad \forall\, i\in I(\overline{x}),\ \langle f'(\overline{x}),\,h\rangle\leqslant 0\}$$

— критический конус задачи (9), (10).

[1] Из (27) следует, что на самом деле для траектории $\{(x^k,\,\lambda^k,\,\mu^k)\}$, генерируемой алгоритмом, будет выполнено $\mu_i^k=0 \quad \forall\, i\in\{1,\dots,m\}\setminus I(\overline{x})$, $\forall\, k=1,2,\dots$, т. е., уже начиная со второго шага, алгоритм будет работать как чистый, а не неточный метод Ньютона.

Доказательство. Напомним, что в соответствии с алгоритмом 2 для каждого k точка $\{(d^k,\ \lambda^{k+1},\ \mu^{k+1})\}$ (где $d^k = x^{k+1} - x^k$) удовлетворяет системе Каруша–Куна–Таккера:

$$f'(x^k) + H_k d^k + (F'(x^k))^{\mathrm{T}}\lambda^{k+1} + (G'(x^k))^{\mathrm{T}}\mu^{k+1} = 0, \tag{30}$$

$$F(x^k) + F'(x^k)d^k = 0, \tag{31}$$

$$G(x^k) + G'(x^k)d^k \leqslant 0, \quad \mu^{k+1} \geqslant 0, \tag{32}$$

$$\mu_i^{k+1}(g_i(x^k) + \langle g_i'(x^k),\ d^k\rangle) = 0, \quad i = 1, \ldots, m, \tag{33}$$

для задачи (11), (12).

Из (30) следует, что

$$\begin{aligned}
-H_k d^k &= f'(x^k) + (F'(x^k))^{\mathrm{T}}\lambda^{k+1} + (G'(x^k))^{\mathrm{T}}\mu^{k+1} = \\
&= \frac{\partial L}{\partial x}(x^k,\ \lambda^{k+1},\ \mu^{k+1}) = \\
&= \frac{\partial L}{\partial x}(\overline{x}, \lambda^{k+1}, \mu^{k+1}) + \frac{\partial^2 L}{\partial x^2}(\overline{x}, \lambda^{k+1}, \mu^{k+1})(x^k - \overline{x}) + o(|x^k - \overline{x}|) = \\
&= (F'(\overline{x}))^{\mathrm{T}}(\lambda^{k+1} - \overline{\lambda}) + (G'(\overline{x}))^{\mathrm{T}}(\mu^{k+1} - \overline{\mu}) + \\
&\qquad + \frac{\partial^2 L}{\partial x^2}(\overline{x},\ \overline{\lambda},\ \overline{\mu})(x^k - \overline{x}) + o(|x^k - \overline{x}|),
\end{aligned} \tag{34}$$

где учтено определение стационарной точки задачи (9), (10) и отвечающих этой точке множителей Лагранжа, а также сходимость $\{(\lambda^k,\ \mu^k)\}$ к $(\overline{\lambda},\ \overline{\mu})$. Из (34) имеем

$$\begin{aligned}
&\left(\frac{\partial^2 L}{\partial x^2}(\overline{x},\ \overline{\lambda},\ \overline{\mu}) - H_k\right) d^k = (F'(\overline{x}))^{\mathrm{T}}(\lambda^{k+1} - \overline{\lambda}) + \\
&+ (G'(\overline{x}))^{\mathrm{T}}(\mu^{k+1} - \overline{\mu}) + \frac{\partial^2 L}{\partial x^2}(\overline{x},\ \overline{\lambda},\ \overline{\mu})(x^{k+1} - \overline{x}) + o(|x^k - \overline{x}|).
\end{aligned} \tag{35}$$

Предположим сначала, что скорость сходимости $\{x^k\}$ к $\overline{x}$ сверхлинейная, т. е. $|x^{k+1} - \overline{x}| = o(|x^k - \overline{x}|)$. Тогда (35) принимает вид

$$\begin{aligned}
&\left(\frac{\partial^2 L}{\partial x^2}(\overline{x},\ \overline{\lambda},\ \overline{\mu}) - H_k\right) d^k = \\
&\qquad = (F'(\overline{x}))^{\mathrm{T}}(\lambda^{k+1} - \overline{\lambda}) + (G'(\overline{x}))^{\mathrm{T}}(\mu^{k+1} - \overline{\mu}) + o(|x^k - \overline{x}|).
\end{aligned} \tag{36}$$

Поскольку $\{d^k\} \to 0$, $\{\mu^k\} \to \overline{\mu}$ $(k \to \infty)$, из (33) вытекает, что если k достаточно велико, то $\mu_i^{k+1} = 0 \ \ \forall\, i \in \{1, \ldots, m\} \setminus I(\overline{x})$. Поэтому согласно лемме 1.4.4, второму неравенству в (32) и условию дополняющей нежесткости $\forall\, h \in K(\overline{x})$ имеем

$$\begin{aligned}
&\langle (F'(\overline{x}))^{\mathrm{T}}(\lambda^{k+1} - \overline{\lambda}) + (G'(\overline{x}))^{\mathrm{T}}(\mu^{k+1} - \overline{\mu}),\ h\rangle = \\
&\qquad = \langle \mu^{k+1},\ G'(\overline{x})h\rangle = \sum_{i \in I(\overline{x})} \mu_i^{k+1}\langle g_i'(\overline{x}),\ h\rangle \leqslant 0,
\end{aligned} \tag{37}$$

а это означает, что

$$\pi_{K(\overline{x})}\left((F'(\overline{x}))^{\mathrm{T}}(\lambda^{k+1}-\overline{\lambda})+(G'(\overline{x}))^{\mathrm{T}}(\mu^{k+1}-\overline{\mu})\right)=0$$

(см. задачу 1.1.10). Но тогда из (36) имеем

$$\pi_{K(\overline{x})}\left(\left(\frac{\partial^2 L}{\partial x^2}(\overline{x},\overline{\lambda},\overline{\mu})-H_k\right)d^k\right)=o(|x^k-\overline{x}|).$$

С учетом того, что $|d^k|\geqslant|x^k-\overline{x}|-|x^{k+1}-\overline{x}|=|x^k-\overline{x}|+o(|x^k-\overline{x}|)$, последнее соотношение влечет (29), поскольку

$$\frac{o(|x^k-\overline{x}|)}{|d^k|}\leqslant\frac{o(|x^k-\overline{x}|)}{|x^k-\overline{x}|+o(|x^k-\overline{x}|)}=$$

$$=\frac{o(|x^k-\overline{x}|)/|x^k-\overline{x}|}{1+o(|x^k-\overline{x}|)/|x^k-\overline{x}|}\to 0\quad(k\to\infty).$$

Пусть теперь выполнено (29). Прежде всего заметим, что согласно (31)

$$0=F(x^k)+F'(x^k)d^k=F'(\overline{x})(x^k-\overline{x})+F'(x^k)d^k+o(|x^k-\overline{x}|)=$$

$$=F'(\overline{x})(x^{k+1}-\overline{x})+\eta^k,\quad(38)$$

где, с учетом сходимости $\{x^k\}$ к $\overline{x}$,

$$\eta^k=(F'(x^k)-F'(\overline{x}))d^k+o(|x^k-\overline{x}|)=o(|d^k|)+o(|x^k-\overline{x}|).\quad(39)$$

Аналогичным образом с помощью первого неравенства в (32), (33) и сходимости $\{\mu^k\}$ к $\overline{\mu}$ устанавливается, что

$$0=\langle g_i'(\overline{x}),\,x^{k+1}-\overline{x}\rangle+\zeta_i^k\quad\forall\,i\in I_+(\overline{x}),\quad(40)$$

$$0\geqslant\langle g_i'(\overline{x}),\,x^{k+1}-\overline{x}\rangle+\zeta_i^k\quad\forall\,i\in I_0(\overline{x}),\quad(41)$$

$$0=\mu_i^{k+1}(\langle g_i'(\overline{x}),\,x^{k+1}-\overline{x}\rangle+\zeta_i^k)\quad\forall\,i\in I_0(\overline{x}),\quad(42)$$

где $I_+(\overline{x})=\{i\in I(\overline{x})\,|\,\overline{\mu}_i>0\}$, $I_0(\overline{x})=\{i\in I(\overline{x})\,|\,\overline{\mu}_i=0\}$ и, кроме того,

$$\zeta_i^k=o(|d^k|)+o(|x^k-\overline{x}|),\quad i\in I(\overline{x}).\quad(43)$$

Из условия линейной независимости и соотношений (39), (43) следует, что существует элемент $\xi^k\in\mathbf{R}^n$ такой, что

$$F'(\overline{x})\xi^k=\eta^k,\quad\langle g_i'(\overline{x}),\,\xi^k\rangle=\zeta_i^k\quad\forall\,i\in I(\overline{x}),\quad(44)$$

$$|\xi^k|=o(|d^k|)+o(|x^k-\overline{x}|).\quad(45)$$

Положим $h^k=x^{k+1}-\overline{x}+\xi^k$. Из (38)–(41), (44) следует, что $h^k\in K(\overline{x})$, и, в частности, аналогично (37) выводится соотношение

$$\langle(F'(\overline{x}))^{\mathrm{T}}(\lambda^{k+1}-\overline{\lambda})+(G'(\overline{x}))^{\mathrm{T}}(\mu^{k+1}-\overline{\mu}),\,h^k\rangle=$$

$$=\sum_{i\in I(\overline{x})}\mu_i^{k+1}\langle g_i'(\overline{x}),\,h\rangle=0,$$

где последнее равенство следует из (40) и (42). Домножая левую и правую части (35) скалярно на h^k и используя последнее соотношение, получим

$$\left\langle \left(\frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu}) - H_k \right) d^k, h^k \right\rangle =$$

$$= \left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})(x^{k+1} - \overline{x}), h^k \right\rangle + o(|x^k - \overline{x}||h^k|). \quad (46)$$

Далее, согласно достаточному условию второго порядка оптимальности найдется число $\gamma > 0$ такое, что

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})h, h \right\rangle \geqslant \gamma |h|^2 \quad \forall\, h \in K(\overline{x})$$

(это следует из замкнутости критического конуса). Отсюда и из (46) имеем

$$\begin{aligned}
\gamma |h^k|^2 &\leqslant \left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})h^k, h^k \right\rangle = \\
&= \left\langle \left(\frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu}) - H_k \right) d^k, h^k \right\rangle + \\
&\qquad + \left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})\xi^k, h^k \right\rangle + o(|x^k - \overline{x}||h^k|) = \\
&= \left\langle \left(\frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu}) - H_k \right) d^k, h^k \right\rangle + \\
&\qquad + o(|d^k||h^k|) + o(|x^k - \overline{x}||h^k|) \leqslant \\
&\leqslant \left\langle \pi_{K(\overline{x})} \left(\left(\frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu}) - H_k \right) d^k \right), h^k \right\rangle + \\
&\qquad + o(|d^k||h^k|) + o(|x^k - \overline{x}||h^k|) = \\
&= o(|d^k||h^k|) + o(|x^k - \overline{x}||h^k|),
\end{aligned}$$

где второе равенство следует из (45), второе неравенство — из результата задачи 1.1.7, а последнее равенство — из (29). Таким образом, $|h^k| = o(|d^k|) + o(|x^k - \overline{x}|)$. Но тогда в силу (45)

$$|x^{k+1} - \overline{x}| = |h^k - \xi^k| = o(|d^k|) + o(|x^k - \overline{x}|).$$

Это означает существование последовательности $\{t_k\} \subset \mathbf{R}_+$ такой, что $t_k \to 0 \quad (k \to \infty)$ и для любого k

$$|x^{k+1} - \overline{x}| \leqslant t_k(|d^k| + |x^k - \overline{x}|) \leqslant t_k(|x^{k+1} - \overline{x}| + 2|x^k - \overline{x}|).$$

Но тогда для любого достаточно большого k

$$|x^{k+1} - \overline{x}| \leqslant \frac{2t_k}{1 - t_k} |x^k - \overline{x}|,$$

что и дает равенство $|x^{k+1} - \overline{x}| = o(|x^k - \overline{x}|)$. $\square$

Сформулированные в теореме 1 условия, гарантирующие локальную сходимость метода SQP со сверхлинейной скоростью (в прямодвойственных переменных), могут быть ослаблены. Например, можно отказаться от условия строгой дополнительности, которое в современной литературе принято считать слишком обременительным, если вместо достаточного условия второго порядка потребовать выполнения следующего сильного достаточного условия второго порядка [1]:

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x},\, \overline{\lambda},\, \overline{\mu})h,\, h \right\rangle > 0 \quad \forall\, h \in K_+(\overline{x}) \setminus \{0\},$$

где

$$K_+(\overline{x}) = \{h \in \ker F'(\overline{x}) \,\big|\, \langle g_i'(\overline{x}),\, h\rangle = 0 \quad \forall\, i \in I_+(\overline{x})\}.$$

Напомним, что $I_+(\overline{x}) = \{i \in I(\overline{x}) \,\big|\, \overline{\mu}_i > 0\}$ (зависимость от $\overline{\mu}$ опущена, поскольку при выполнении условия линейной независимости множители $\overline{\lambda}$ и $\overline{\mu}$ однозначно определяются точкой $\overline{x}$). Заметим, что в случае выполнения условия строгой дополнительности, т. е. при $I_+(\overline{x}) = I(\overline{x})$, имеет место равенство $K_+(\overline{x}) = K(\overline{x})$, и сильное достаточное условие второго порядка совпадает с обычным.

Однако более тонкий анализ делает возможным дальнейшее ослабление условий локальной сходимости метода SQP со сверхлинейной скоростью. По-видимому, наиболее слабая известная на сегодняшний день комбинация условий такого рода — это строгое условие регулярности Мангасариана–Фромовица и обычное достаточное условие второго порядка. Обоснование этого факта, которое можно найти, например, в [20], использует тонкие результаты о чувствительности для систем Каруша–Куна–Таккера и выходит за рамки данной книги.

В заключение отметим одно важное обстоятельство. Оптимизационная природа методов SQP создает готовую базу для глобализации их сходимости (в отличие, например, от методов, обсуждаемых в следующем параграфе). Вопросы глобализации сходимости методов SQP рассматриваются в § 5.4.

§ 4.5. Методы решения системы Каруша–Куна–Таккера

В этом параграфе речь пойдет о методах ньютоновского типа, адаптированных для решения негладких систем уравнений, и их применении к системам Каруша–Куна–Таккера (ККТ) для задач оптимизации со смешанными ограничениями. Дело в том, что такие системы, изначально содержащие как уравнения, так и неравенства, могут быть

[1] Общепринятая аббревиатура для этого условия — SSOSC (от английского Strong second-order sufficient condition).

многими способами сведены к чистым системам уравнений. Однако возможность применения в данном контексте обычных ньютоновских методов решения гладких уравнений ограничена тем обстоятельством, что гладкие переформулировки систем ККТ могут удовлетворять условиям приведенных в п. 3.2.1 утверждений о локальной сходимости ньютоновских методов лишь при выполнении в искомом решении условия строгой дополнительности; в ином случае производная оператора такого уравнения в соответствующем его решении неизбежно вырождена. Именно это обстоятельство привело к тому, что в настоящее время значительно более популярны негладкие переформулировки систем ККТ: они не обладают неизбежной вырожденностью в случае отсутствия строгой дополнительности. Разумеется, это не компенсировало бы те сложности, которые связаны с негладкостью, если бы не наличие у рассматриваемых негладких переформулировок систем ККТ специальной структуры, которая позволяет строить на их основе вполне эффективные методы.

Читателю может показаться, что данный параграф перегружен терминологией, причем не всегда удачной. Дело, однако, в том, что эта терминология является общепринятой в оптимизационном сообществе, и знакомство с ней полезно, поскольку существенно облегчает чтение современной литературы по оптимизации.

4.5.1. Эквивалентные переформулировки системы Каруша–Куна–Таккера. Пусть $f\colon \mathbf{R}^n \to \mathbf{R}$ — гладкая функция, $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ — гладкие отображения. Рассматривается задача

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in \mathbf{R}^n \mid F(x) = 0,\ G(x) \leqslant 0\}. \tag{2}$$

Стационарные точки этой задачи и отвечающие им множители Лагранжа описываются системой ККТ

$$\frac{\partial L}{\partial x}(x,\ \lambda,\ \mu) = 0, \quad F(x) = 0, \tag{3}$$

$$\mu \geqslant 0, \quad G(x) \leqslant 0, \quad \langle \mu,\ G(x) \rangle = 0, \tag{4}$$

где

$$L\colon \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R},$$

$$L(x,\ \lambda,\ \mu) = f(x) + \langle \lambda,\ F(x) \rangle + \langle \mu,\ G(x) \rangle,$$

— функция Лагранжа.

Пусть $(\overline{x},\ \overline{\lambda},\ \overline{\mu})$ — решение системы (3), (4). Как обычно, будем обозначать через $g_i(\cdot)$, $i = 1, \ldots, m$, компоненты отображения G, а через $I(\overline{x}) = \{i = 1, \ldots, m \mid g_i(\overline{x}) = 0\}$ — множество индексов активных в точке $\overline{x}$ ограничений-неравенств задачи (1), (2).

Если предполагать выполнение условия строгой дополнительности, т. е.

$$\overline{\mu}_i > 0 \quad \forall\, i \in I(\overline{x}),$$

то вблизи точки $(\overline{x},\, \overline{\lambda},\, \overline{\mu})$ группа соотношений (4) эквивалентна системе уравнений

$$\mu_i = 0, \quad i \in \{1,\, \ldots,\, m\} \setminus I(\overline{x}), \quad g_i(x) = 0, \quad i \in I(\overline{x}). \tag{5}$$

Таким образом, система ККТ (3), (4) локально эквивалентна системе (3), (5), состоящей из $n + l + m$ уравнений относительно $(x,\, \lambda,\, \mu) \in \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$. Если f, F и G дважды дифференцируемы, то к последней системе применим обычный метод Ньютона, аналогично тому, как это делалось в п. 4.3.1 в случае задачи с чистыми ограничениями-равенствами. Разумеется, практическая применимость такого подхода ограничена тем, что множество $I(\overline{x})$ обычно заранее не известно. Способы идентификации этого множества, в том числе и без предположения о строгой дополнительности, а также основанные на такой идентификации методы будут рассмотрены в § 4.6.

Здесь же речь пойдет о методах решения системы ККТ, не связанных ни с предположением строгой дополнительности, ни с явной идентификацией $I(\overline{x})$. В случае выполнения условия строгой дополнительности эти методы редуцируются к обычному методу Ньютона для системы (3), (5). Идея по-прежнему состоит в том, чтобы переписать соотношения (4) в виде системы уравнений, но эта система не должна содержать неизвестных априори элементов. Разумеется, за удобство иметь дело с чистыми системами уравнений (без неравенств) приходится платить определенную цену. Как будет показано ниже, в данном случае цена состоит либо в потере гладкости, либо в неизбежной вырожденности производной оператора переформулированной системы в решении. В определенном смысле первое предпочтительнее, поскольку этот путь в сочетании с применением современного негладкого анализа естественным образом приводит к вполне реализуемым алгоритмам с привлекательными локальными свойствами сходимости.

Определение 1. Функция $\psi\colon \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ называется *функцией дополнительности*, если множество решений уравнения

$$\psi(a,\, b) = 0$$

совпадает с множеством решений системы

$$a \geqslant 0, \quad b \geqslant 0, \quad ab = 0.$$

Два важных примера функций дополнительности доставляют *функция естественной невязки*

$$\psi(a,\, b) = \min\{a,\, b\}, \quad a,\, b \in \mathbf{R}, \tag{6}$$

и *функция Фишера–Бурмейстера*

$$\psi(a,\, b) = \sqrt{a^2 + b^2} - a - b, \quad a,\, b \in \mathbf{R}. \tag{7}$$

Задача 1. Проверить, что заданные формулами (6) и (7) функции действительно являются функциями дополнительности.

Задача 2. Пусть $\omega\colon \mathbf{R} \to \mathbf{R}$ — любая монотонно возрастающая функция, $\omega(0) = 0$. Доказать, что функция

$$\psi\colon \mathbf{R} \times \mathbf{R} \to \mathbf{R}, \quad \psi(a,\, b) = \omega(|a - b|) - \omega(a) - \omega(b),$$

является функцией дополнительности.

Совершенно ясно, что, выбрав функцию дополнительности ψ, соотношения (4) можно эквивалентным образом записать в виде

$$\psi(\mu_i,\, -g_i(x)) = 0, \quad i = 1, \ldots, m. \tag{8}$$

Соответственно всю систему ККТ (3), (4) можно заменить системой (3), (8) из $n + l + m$ скалярных уравнений с теми же неизвестными.

Очевидно, функция ψ, определенная в (6) или (7), не является дифференцируемой в точке $(0,\, 0)$ (функция ψ из (6) не дифференцируема при $a = b$). Отсюда следует, что если условие строгой дополнительности не выполнено, то оператор соответствующей системы (3), (8) недифференцируем в ее решении $(\overline{x},\, \overline{\lambda},\, \overline{\mu})$, какими бы гладкими ни были f, F и G. Это обстоятельство является важнейшей мотивировкой развития обобщенных вариантов метода Ньютона, применимых к негладким уравнениям; см. п. 4.5.2.

Существуют также и всюду дифференцируемые функции дополнительности, например,

$$\psi(a,\, b) = 2ab - (\min\{0,\, a + b\})^2, \quad a,\, b \in \mathbf{R}. \tag{9}$$

Задача 3. Проверить, что заданная в (9) функция действительно является функцией дополнительности.

Переформулировки, основанные на гладких функциях дополнительности, неизбежно вырождены в следующем смысле: легко видеть, что при невыполнении условия строгой дополнительности производная оператора системы (3), (8) в точке $(\overline{x},\, \overline{\lambda},\, \overline{\mu})$ вырождена. А именно, строки матрицы этой производной, отвечающие тем индексам $i \in I(\overline{x})$, для которых $\overline{\mu}_i = 0$, являются нулевыми. Подчеркнем, что это свойство присуще любым гладким переформулировкам системы ККТ, т. е. не является следствием неудачного выбора функции дополнительности, например, введенной в (9). Поэтому обоснование стандартного метода Ньютона и применение его к системе ККТ связано с принципиальными трудностями.

Негладкие переформулировки могут быть невырождены (в поясняемом ниже смысле) и при невыполнении условия строгой дополнительности. Кроме того, негладкость этих переформулировок обладает весьма специальной структурой, допускающей построение эффективных алгоритмов.

4.5.2. Элементы негладкого анализа и обобщенный метод Ньютона. В этом пункте приводится краткая сводка понятий и фактов негладкого анализа, необходимых для дальнейшего изложения. Важнейшую роль в современном негладком анализе играет следующая теорема Радемахера [25, 28].

Теорема 1. *Пусть отображение* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *непрерывно по Липшицу на открытом множестве* $V \subset \mathbf{R}^n$.

Тогда для множества $\mathcal{D}_F$ *точек дифференцируемости* F *мера Лебега множества* $V \setminus \mathcal{D}_F$ *равна нулю.*

Значение теоремы Радемахера состоит в том, что она позволяет вводить содержательные обобщенные понятия производной для отображений, не дифференцируемых в обычном смысле. Так, *B-дифференциалом* отображения $F\colon \mathbf{R}^n \to \mathbf{R}^l$ в точке $x \in \mathbf{R}^n$ называется множество

$$\partial_B F(x) = \{A \in \mathbf{R}(l, n) \mid \exists\, \{x^k\} \subset \mathcal{D}_F\colon \{x^k\} \to x,\ \{F'(x^k)\} \to A\ \ (k \to \infty)\}$$

(индекс B — в честь Булигана). *Дифференциалом Кларка* отображения F в точке x называется множество

$$\partial F(x) = \operatorname{conv} \partial_B F(x).$$

Задача 4. Доказать, что если отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ непрерывно по Липшицу в некоторой окрестности точки $x \in \mathbf{R}^n$ с константой $L > 0$, то $\partial F(x)$ — непустой компакт, причем $\|A\| \leqslant L$ $\forall\, A \in \partial F(x)$.

Разумеется, практическое значение введенных понятий, и в том числе при построении методов оптимизации, проявляется в связи с тем, что B-дифференциал и/или дифференциал Кларка либо их отдельные элементы могут быть вычислены во многих важных случаях. Сразу отметим следующие очевидные факты. Для отображения $F\colon \mathbf{R}^n \to \mathbf{R}^{l_1} \times \mathbf{R}^{l_2}$, $F(x) = (F_1(x), F_2(x))$, справедливо включение $\partial_B F(x) \subset \partial_B F_1(x) \times \partial_B F_2(x)$ $\forall x \in \mathbf{R}^n$. Если при этом F_1 дифференцируемо в окрестности точки x, то $\{F_1'(x)\} \times \partial_B F_2(x) \subset \subset \partial_B F(x)$, а если производная F_1 непрерывна в точке x, то последнее включение выполнено как равенство.

Задача 5. Пусть $f_i\colon \mathbf{R}^n \to \mathbf{R}$ — дифференцируемые на $\mathbf{R}^n$ функции, $i = 1, \ldots, s$. Положим

$$f(x) = \min_{i=1,\ldots,s} f_i(x), \qquad I(x) = \{i = 1, \ldots, s \mid f_i(x) = f(x)\}, \quad x \in \mathbf{R}^n.$$

Показать, что для всякого $x \in \mathbf{R}^n$ такого, что производная f_i непрерывна в точке x для каждого $i \in I(x)$, справедливо

$$\partial_B f(x) \subset \{f_i'(x) \mid i \in I(x)\},$$

причем это включение выполнено как равенство, если для каждого $i \in I(x)$ найдется сходящаяся к x последовательность $\{x^{i,k}\} \subset \subset \mathbf{R}^n$ такая, что $I(x^{i,k}) = \{i\} \quad \forall\, k$.

Задача 6. Пусть $G\colon \mathbf{R}^n \to \mathbf{R}^m$ — дифференцируемое на $\mathbf{R}^n$ отображение с компонентами $g_i(\cdot)$, $i = 1, \ldots, m$. Введем отображение $\Phi\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^m$ с компонентами

$$\varphi_i(x, \mu) = \psi(\mu_i, -g_i(x)), \quad x \in \mathbf{R}^n, \quad \mu \in \mathbf{R}^m, \quad i = 1, \ldots, m,$$

где ψ — введенная в (6) функция естественной невязки. Показать, что для всякого $x \in \mathbf{R}^n$ такого, что производная G непрерывна в точке x, и $\forall\, \mu \in \mathbf{R}^m$ B-дифференциал $\partial_B \Phi(x, \mu)$ состоит из всех матриц вида $(-A(x, \mu)G'(x), E^m - A(x, \mu))$, где $A(x, \mu)$ — диагональная $m \times m$-матрица с диагональными элементами

$$a_i(x, \mu) = \begin{cases} 1, & \text{если } \mu_i > -g_i(x), \\ 0 \text{ или } 1, & \text{если } \mu_i = -g_i(x), \\ 0, & \text{если } \mu_i < -g_i(x), \end{cases} \qquad i = 1, \ldots, m. \tag{10}$$

Задача 7. В условиях задачи 6 рассмотреть случай, когда ψ — введенная в (7) функция Фишера–Бурмейстера. Показать, что для всякого $x \in \mathbf{R}^n$ такого, что производная G непрерывна в точке x, и всякого $\mu \in \mathbf{R}^m$ $\partial_B \Phi(x, \mu)$ состоит из матриц вида $(A(x, \mu)G'(x), B(x, \mu))$, где $A(x, \mu)$ и $B(x, \mu)$ — диагональные $m \times m$-матрицы с диагональными элементами

$$a_i(x, \mu) = \begin{cases} 1 + \dfrac{g_i(x)}{\sqrt{\mu_i^2 + (g_i(x))^2}}, & \text{если } |\mu_i| + |g_i(x)| \neq 0, \\ 1 + \alpha_i, & \text{если } |\mu_i| + |g_i(x)| = 0, \end{cases} \tag{11}$$

$$b_i(x, \mu) = \begin{cases} \dfrac{\mu_i}{\sqrt{\mu_i^2 + (g_i(x))^2}} - 1, & \text{если } |\mu_i| + |g_i(x)| \neq 0, \\ \beta_i - 1, & \text{если } |\mu_i| + |g_i(x)| = 0, \end{cases} \tag{12}$$

$$\alpha_i, \beta_i \in \mathbf{R}, \quad \alpha_i^2 + \beta_i^2 = 1, \quad i = 1, \ldots, m.$$

Показать, что если $g_i'(x)$, $i \in I_0(x, \mu) = \{i = 1, \ldots, m \mid g_i(x) = \mu_i = 0\}$ линейно независимы, то $\partial_B \Phi(x, \mu)$ состоит из всех матриц указанного вида.

На основе дифференциала Кларка вводится следующее условие, которое будет играть центральную роль при обосновании вводимого

ниже обобщенного метода Ньютона. Будем говорить, что отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ удовлетворяет в точке $x \in \mathbf{R}^n$ *условию гладкости Куммера*, если

$$\sup_{A \in \partial F(x+h)} |F(x+h) - F(x) - Ah| = o(|h|),$$

и *сильному условию гладкости Куммера*, если

$$\sup_{A \in \partial F(x+h)} |F(x+h) - F(x) - Ah| = O(|h|^2).$$

Некоторые другие важные концепции обобщенного дифференцирования основаны на обычном понятии производной по направлению.

Напомним, что отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ называется *дифференцируемым* в точке $x \in \mathbf{R}^n$ *по направлению* $h \in \mathbf{R}^n$, если существует конечный предел $\lim\limits_{t \to 0+} (F(x+th) - F(x))/t$. Сам предел при этом называется *производной* отображения F в точке x *по направлению* h и обозначается $F'(x;\, h)$. Разумеется, если F дифференцируемо в точке x, то оно дифференцируемо в этой точке по любому направлению, причем $F'(x;\, h) = F'(x)h \quad \forall\, h \in \mathbf{R}^n$.

Задача 8. Доказать, что если отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ непрерывно по Липшицу в некоторой окрестности точки $x \in \mathbf{R}^n$ и дифференцируемо в этой точке по любому направлению, то:

а) положительно однородное отображение $h \to F'(x;\, h)\colon \mathbf{R}^n \to \mathbf{R}^l$ непрерывно по Липшицу на $\mathbf{R}^n$;

б) для каждого $h \in \mathbf{R}^n$ найдется матрица $A \in \partial_B F(x)$ такая, что $F'(x;\, h) = Ah$;

в) отображение F *B-дифференцируемо* в точке x, т. е.

$$F(x+h) - F(x) - F'(x;\, h) = o(|h|).$$

Концепцией, объединяющей в себе непрерывность отображения по Липшицу в окрестности данной точки, (сильное) условие гладкости Куммера, а также дифференцируемость по любому направлению, служит часто используемое в последнее время условие полугладкости: отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$, удовлетворяющее указанным трем условиям в точке $x \in \mathbf{R}^n$, называется (*сильно*) *полугладким* в этой точке. Популярность этого условия связана, видимо, с тем, что составляющие его три условия часто оказываются выполненными одновременно. Существует ряд эквивалентных определений полугладкости.

Задача 9. Доказать, что если отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ непрерывно по Липшицу в некоторой окрестности точки $x \in \mathbf{R}^n$, то оно полугладко в точке x тогда и только тогда, когда для любого $h \in \mathbf{R}^n$ существует конечный предел $\lim\limits_{\xi \to h,\, t \to 0+} A\xi$, не зависящий от

выбора $A \in \partial F(x + t\xi)$. Показать, что при этом указанный предел совпадает с $F'(x; h)$.

Задача 10. Доказать, что если отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ непрерывно по Липшицу на некоторой окрестности точки $x \in \mathbf{R}^n$ и дифференцируемо в этой точке по любому направлению, то оно удовлетворяет в точке x условию гладкости Куммера тогда и только тогда, когда

$$\sup_{A \in \partial F(x+h)} |Ah - F'(x; h)| = o(|h|),$$

и сильному условию гладкости Куммера тогда и только тогда, когда

$$\sup_{A \in \partial F(x+h)} |Ah - F'(x; h)| = O(|h|^2).$$

Из теоремы о среднем легко выводится, что непрерывно дифференцируемое в окрестности данной точки отображение полугладко в этой точке. Если дополнительно производная отображения непрерывна по Липшицу вблизи рассматриваемой точки, то отображение сильно полугладко в этой точке.

Задача 11. Доказать, что если функция $F\colon \mathbf{R}^n \to \mathbf{R}$ выпукла на некоторой (выпуклой) окрестности точки $x \in \mathbf{R}^n$, то F полугладка в точке x.

Задача 12. Доказать, что если отображения $F_1, F_2\colon \mathbf{R}^n \to \mathbf{R}^l$ полугладки в точке $x \in \mathbf{R}^n$, то скалярное произведение, сумма, покомпонентный максимум и покомпонентный минимум F_1 и F_2 также полугладки в точке x.

Задача 13. Доказать, что если отображения $F_1\colon \mathbf{R}^n \to \mathbf{R}^{l_1}$ и $F_2\colon \mathbf{R}^n \to \mathbf{R}^{l_2}$ полугладки в точке $x \in \mathbf{R}^n$, то отображение $F\colon \mathbf{R}^n \to \mathbf{R}^{l_1} \times \mathbf{R}^{l_2}$, $F(x) = (F_1(x), F_2(x))$, полугладко в точке x.

Задача 14. Доказать, что если отображение $F_1\colon \mathbf{R}^n \to \mathbf{R}^{l_1}$ полугладко в точке $x \in \mathbf{R}^n$, а отображение $F_2\colon \mathbf{R}^{l_1} \to \mathbf{R}^{l_2}$ полугладко в точке $F_1(x)$, то отображение $F\colon \mathbf{R}^n \to \mathbf{R}^{l_2}$, $F(x) = F_2(F_1(x))$, полугладко в точке x.

Теперь рассмотрим отображение $\Phi\colon \mathbf{R}^n \to \mathbf{R}^n$. Естественным обобщением метода Ньютона для уравнения

$$\Phi(x) = 0 \tag{13}$$

является следующая итерационная схема:

$$x^{k+1} = x^k - A_k^{-1}\Phi(x^k), \quad k = 0, 1, \ldots, \tag{14}$$

где $A_k \in \partial\Phi(x^k)$ либо $A_k \in \partial_B\Phi(x^k)$.

Таким образом, *обобщенный метод Ньютона* состоит в следующем.

Алгоритм 1. Выбираем вариант B или C алгоритма. Выбираем $x^0 \in \mathbf{R}^n$ и полагаем $k = 0$.

1. Вычисляем некоторую матрицу $A_k \in \partial_B \Phi(x^k)$ в случае варианта B и $A_k \in \partial\Phi(x^k)$ в случае варианта C. Вычисляем $x^{k+1} \in \in \mathbf{R}^n$ как решение линейной системы
$$A_k(x - x^k) = -\Phi(x^k).$$

2. Увеличиваем номер шага k на 1 и переходим к п. 1.

Условие невырожденности, которое понадобится для обоснования локальной сходимости обобщенного метода Ньютона, вводится следующим образом. Отображение $\Phi\colon \mathbf{R}^n \to \mathbf{R}^n$ называется *BD-регулярным* (*CD-регулярным*) в точке $\overline{x} \in \mathbf{R}^n$, если любая матрица $A \in \in \partial_B \Phi(\overline{x})$ $(A \in \partial\Phi(\overline{x}))$ невырождена.

Теорема 2. *Пусть отображение* $\Phi\colon \mathbf{R}^n \to \mathbf{R}^n$ *непрерывно по Липшицу в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$ *и удовлетворяет в этой точке условию гладкости Куммера. Пусть* $\overline{x}$ *является решением уравнения* (13), *причем отображение* Φ *BD-регулярно в точке* $\overline{x} \in \mathbf{R}^n$ *в случае использования варианта* B *алгоритма* 1 *и CD-регулярно в случае использования варианта* C.

Тогда любое начальное приближение $x^0 \in \mathbf{R}^n$, *достаточно близкое к* $\overline{x}$, *корректно определяет траекторию алгоритма* 1, *которая сходится к* $\overline{x}$. *Скорость сходимости сверхлинейная, а если* Φ *удовлетворяет в точке* $\overline{x}$ *сильному условию гладкости Куммера, то квадратичная.*

Доказательство. Схема доказательства та же, что и схема доказательства теоремы 3.2.1. Ограничимся рассмотрением варианта B алгоритма (вариант C рассматривается аналогично). Из результата задачи 4, определения B-дифференциала и дифференциала Кларка, а также теоремы о малом возмущении невырожденной матрицы рассуждением от противного выводится существование окрестности U точки $\overline{x}$ и числа $M > 0$ таких, что
$$\det A \neq 0, \quad \|A^{-1}\| \leqslant M \quad \forall\, A \in \partial_B \Phi(x), \ \forall\, x \in U. \tag{15}$$
В частности, итерация алгоритма 1 из любой точки $x^k \in U$ корректно определена.

Далее, в силу (14), (15) и условия гладкости Куммера в точке $\overline{x}$ окрестность U можно выбрать так, что если $x^k \in U$ и $A_k \in \partial_B \Phi(x^k)$, то
$$|x^{k+1} - \overline{x}| = |x^k - \overline{x} - A_k^{-1}\Phi(x^k)| \leqslant$$
$$\leqslant \|A_k^{-1}\|\, |\Phi(x^k) - \Phi(\overline{x}) - A_k(x^k - \overline{x})| = o(|x^k - \overline{x}|). \tag{16}$$

Отсюда следует, что, во-первых, всякое начальное приближение x^0, достаточно близкое к $\overline{x}$, корректно определяет траекторию алгоритма 1, а во-вторых, эта траектория сходится к $\overline{x}$ со сверхлинейной скоростью.

Наконец, если в точке $\overline{x}$ выполнено сильное условие гладкости Куммера, то оценка (16) принимает вид

$$|x^{k+1} - \overline{x}| = O(|x^k - \overline{x}|^2),$$

что и дает квадратичную скорость сходимости. $\square$

В завершение этого пункта приведем результат об оценке расстояния до решения уравнения (13), который понадобится в § 4.6. Будем говорить, что дифференцируемое в точке $\overline{x} \in \mathbf{R}^n$ по любому направлению отображение $\Phi\colon \mathbf{R}^n \to \mathbf{R}^n$ обладает в этой точке R_0-*свойством*, если

$$\{h \in \mathbf{R}^n \mid \Phi'(\overline{x};\, h) = 0\} = \{0\}.$$

Подчеркнем, что согласно утверждению б) задачи 8, если отображение Φ вдобавок непрерывно по Липшицу в некоторой окрестности точки $\overline{x}$, то BD-регулярность Φ в точке $\overline{x}$ влечет R_0-свойство в этой точке.

Теорема 3. *Пусть отображение* $\Phi\colon \mathbf{R}^n \to \mathbf{R}^n$ *непрерывно по Липшицу в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$ *с константой* $L > 0$ *и дифференцируемо в этой точке по любому направлению, причем* $\Phi(\overline{x}) = 0$.

Тогда R_0-*свойство отображения* Φ *в точке* $\overline{x}$ *является необходимым и достаточным условием существования окрестности* U *точки* $\overline{x}$ *и числа* $M > 0$ *таких, что*

$$|x - \overline{x}| \leqslant M|\Phi(x)| \quad \forall\, x \in U.$$

Доказательство. Докажем достаточность. От противного: предположим, что найдется последовательность $\{x^k\} \subset \mathbf{R}^n$ такая, что $\{x^k\} \to \overline{x}$ $(k \to \infty)$, $x^k \neq \overline{x}$ $\forall\, k$ и, кроме того,

$$\frac{|\Phi(x^k)|}{|x^k - \overline{x}|} \to 0 \quad (k \to \infty).$$

Для каждого k положим $h^k = (x^k - \overline{x})/|x^k - \overline{x}|$, $t_k = |x^k - \overline{x}|$. Без ограничения общности можем считать, что последовательность $\{h^k\}$ сходится к некоторому $h \in \mathbf{R}^n$, $|h| = 1$. Тогда $\forall\, k$ имеет место

$$\frac{|\Phi(\overline{x} + t_k h)|}{t_k} \leqslant \frac{|\Phi(\overline{x} + t_k h) - \Phi(\overline{x} + t_k h^k)|}{t_k} + \frac{|\Phi(\overline{x} + t_k h^k)|}{t_k} \leqslant$$

$$\leqslant L|h^k - h| + \frac{|\Phi(x^k)|}{|x^k - \overline{x}|} \to 0 \quad (k \to \infty),$$

откуда следует равенство $\Phi'(\overline{x};\, h) = 0$, противоречащее R_0-свойству.

Для доказательства необходимости рассмотрим произвольный элемент $h \in \mathbf{R}^n$ такой, что $\Phi'(\overline{x};\, h) = 0$. Если окрестность U точки $\overline{x}$ и число $M > 0$, обладающие нужным свойством, существуют, то для любого достаточно малого $t > 0$

$$t|h| = |\overline{x} + th - \overline{x}| \leqslant M|\Phi(\overline{x} + th)| = \\ = M|\Phi(\overline{x} + th) - \Phi(\overline{x})| = M|t\Phi'(\overline{x};\, h)| + o(t) = o(t),$$

что возможно лишь при $h = 0$. $\square$

4.5.3. Обобщенный метод Ньютона для системы Каруша–Куна–Таккера. Согласно сказанному в п. 4.5.1, если выбрана функция дополнительности ψ (см. определение 1), система KKT для задачи (1), (2) сводится к уравнению

$$\Phi(x,\, \lambda,\, \mu) = 0, \tag{17}$$

где

$$\Phi\colon \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m,$$

$$\Phi(x,\, \lambda,\, \mu) = \begin{pmatrix} \dfrac{\partial L}{\partial x}(x,\, \lambda,\, \mu) \\ F(x) \\ \psi(\mu_1,\, -g_1(x)) \\ \dots \\ \psi(\mu_m,\, -g_m(x)) \end{pmatrix}. \tag{18}$$

Согласно теореме 2 для обоснования применимости обобщенного метода Ньютона к уравнению (17) достаточно указать условия, гарантирующие (сильную) полугладкость Φ и его CD-регулярность (или BD-регулярность) в искомом решении.

Предложение 1. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображения $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ дифференцируемы на $\mathbf{R}^n$, причем их производные полугладки в точке $\overline{x} \in \mathbf{R}^n$.*

Тогда при любых $\lambda \in \mathbf{R}^l$ и $\mu \in \mathbf{R}^m$ отображение Φ, введенное в (18) с использованием функции ψ, заданной в (6), полугладко в точке $(\overline{x},\, \lambda,\, \mu)$. Если же f, F и G дважды дифференцируемы в некоторой окрестности $\overline{x}$, причем их вторые производные непрерывны по Липшицу на этой окрестности, то Φ сильно полугладко в точке $(\overline{x},\, \lambda,\, \mu)$.

Задача 15. Используя результаты задач 12–14, доказать предложение 1.

Задача 16. Доказать аналог предложения 1 для случая использования функции ψ, заданной в (7).

Предложение 2. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображения* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *и* $G\colon \mathbf{R}^n \to \mathbf{R}^m$ *дифференцируемы на* $\mathbf{R}^n$ *и дважды дифференцируемы в точке* $\overline{x} \in \mathbf{R}^n$, *причем их первые производные непрерывны по Липшицу в некоторой окрестности* $\overline{x}$. *Пусть в точке* $\overline{x}$ *выполнено условие линейной независимости, причем* $\overline{x}$ — *стационарная точка задачи* (1), (2), *а* $\overline{\lambda} \in \mathbf{R}^l$ *и* $\overline{\mu} \in \mathbf{R}^m_+$ — *однозначно отвечающие ей множители Лагранжа. Пусть, наконец, в точке* $\overline{x}$ *выполнено сильное достаточное условие второго порядка, т. е.*

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})h,\ h \right\rangle > 0 \quad \forall\, h \in K_+(\overline{x}) \setminus \{0\}, \tag{19}$$

где

$$K_+(\overline{x}) = \{h \in \ker F'(\overline{x}) \,\big|\, \langle g_i'(\overline{x}),\ h\rangle = 0 \quad \forall\, i \in I_+(\overline{x})\},$$

$$I_+(\overline{x}) = \{i \in I(\overline{x}) \,\big|\, \overline{\mu}_i > 0\}.$$

Тогда отображение Φ, *введенное в* (18) *с использованием функции* ψ, *заданной в* (6), *CD-регулярно* (*а значит, и BD-регулярно*) *в точке* $(\overline{x}, \overline{\lambda}, \overline{\mu})$.

Доказательство. Для произвольной матрицы $H \in \partial\Phi(\overline{x}, \overline{\lambda}, \overline{\mu})$, согласно результату задачи 6, найдется множество индексов $J_0 \subset I_0(\overline{x}) = \{i \in I(\overline{x}) \,\big|\, \overline{\mu}_i = 0\}$ такое, что, полагая $J_+ = I(\overline{x}) \setminus J_0$, $J_- = \{1, \ldots, m\} \setminus I(\overline{x})$, при соответствующей нумерации последних m компонент отображения Φ будем иметь

$$H = \begin{pmatrix} \dfrac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu}) & (F'(\overline{x}))^{\mathrm{T}} & (G'_{J_+}(\overline{x}))^{\mathrm{T}} & (G'_{J_0}(\overline{x}))^{\mathrm{T}} & (G'_{J_-}(\overline{x}))^{\mathrm{T}} \\ F'(\overline{x}) & 0 & 0 & 0 & 0 \\ -G'_{J_+}(\overline{x}) & 0 & 0 & 0 & 0 \\ -AG'_{J_0}(\overline{x}) & 0 & 0 & E_{J_0} - A & 0 \\ 0 & 0 & 0 & 0 & E_{J_-} \end{pmatrix},$$

где A — диагональная $|J_0| \times |J_0|$-матрица с диагональными элементами $a_i \in [0, 1]$, $i = 1, \ldots, |J_0|$, и для множества индексов $J \subset \{1, \ldots, m\}$ E_J обозначает единичную $|J| \times |J|$-матрицу, а $G'_J(\overline{x})$ — подматрицу матрицы $G'(\overline{x})$, составленную из строк, номера которых принадлежат J.

Пусть

$$\widetilde{u} = (\widetilde{x}, \widetilde{\lambda}, \widetilde{\mu}^+, \widetilde{\mu}^0, \widetilde{\mu}^-) \in \ker H,$$

где $\widetilde{x} \in \mathbf{R}^n$, $\widetilde{\lambda} \in \mathbf{R}^l$, $\widetilde{\mu}^+ \in \mathbf{R}^{|J_+|}$, $\widetilde{\mu}^0 \in \mathbf{R}^{|J_0|}$, $\widetilde{\mu}^- \in \mathbf{R}^{|J_-|}$. Тогда

$$\frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})\,\widetilde{x} + (F'(\overline{x}))^{\mathrm{T}}\widetilde{\lambda} + \\ + (G'_{J_+}(\overline{x}))^{\mathrm{T}}\widetilde{\mu}^+ + (G'_{J_0}(\overline{x}))^{\mathrm{T}}\widetilde{\mu}^0 + (G'_{J_-}(\overline{x}))^{\mathrm{T}}\widetilde{\mu}^- = 0, \tag{20}$$

$$F'(\overline{x})\widetilde{x} = 0, \tag{21}$$

$$\langle g_i'(\overline{x}),\, \widetilde{x}\rangle = 0, \quad i \in J_+, \tag{22}$$

$$-a_i\langle g_i'(\overline{x}),\, \widetilde{x}\rangle + (1-a_i)\widetilde{\mu}_i^0 = 0, \quad i \in J_0, \tag{23}$$

$$\widetilde{\mu}^- = 0. \tag{24}$$

Но $I_+(\overline{x}) = I(\overline{x}) \setminus I_0(\overline{x}) \subset I(\overline{x}) \setminus J_0 = J_+$, поэтому условия (21) и (22) влекут включение $\widetilde{x} \in K_+(\overline{x})$. Умножая обе части (20) скалярно на $\widetilde{x}$ и используя соотношения (21)–(24), получим

$$\left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\, \overline{\lambda},\, \overline{\mu})\,\widetilde{x},\, \widetilde{x}\right\rangle + \sum_{i\in J_0} \widetilde{\mu}_i^0\langle g_i'(\overline{x}),\, \widetilde{x}\rangle = 0. \tag{25}$$

Равенство (23) подразумевает, что $\widetilde{\mu}_i^0\langle g_i'(\overline{x}),\ \widetilde{x}\rangle \geqslant 0 \ \ \forall\, i \in J_0$, поскольку $a_i \in [0,1]$. Но тогда из условия (19) и соотношения (25) следует, что $\widetilde{x} = 0$. Кроме того, используя равенство (24), из (20) выводим

$$(F'(\overline{x}))^{\mathrm{T}}\widetilde{\lambda} + (G_{J_+}'(\overline{x}))^{\mathrm{T}}\widetilde{\mu}^+ + (G_{J_0}'(\overline{x}))^{\mathrm{T}}\widetilde{\mu}^0 = 0,$$

где $J_+ \cup J_0 = I(\overline{x})$. Отсюда и из условия линейной независимости следует, что $\widetilde{\lambda} = 0$, $\widetilde{\mu}^+ = 0$, $\widetilde{\mu}^0 = 0$, т. е. $\widetilde{u} = 0$. Тем самым показано, что $\ker H = \{0\} \ \ \forall\, H \in \partial\Phi(\overline{x},\, \overline{\lambda},\, \overline{\mu})$. $\square$

Предложение 3. *В условиях предложения* 2 *отображение* Φ, *введенное в* (18) *с использованием функции* ψ, *заданной в* (7), *CD-регулярно* (*а значит, и BD-регулярно*) *в точке* $(\overline{x},\, \overline{\lambda},\, \overline{\mu})$.

Доказательство. Для произвольной матрицы $H \in \partial\Phi(\overline{x}, \overline{\lambda}, \overline{\mu})$ согласно результату задачи 7 имеем

$$H = \begin{pmatrix} \dfrac{\partial^2 L}{\partial x^2}\,(\overline{x},\, \overline{\lambda},\, \overline{\mu}) & (F'(\overline{x}))^{\mathrm{T}} & (G'(\overline{x}))^{\mathrm{T}} \\ F'(\overline{x}) & 0 & 0 \\ A(\overline{x},\, \overline{\mu})G'(\overline{x}) & 0 & B(\overline{x},\, \overline{\mu}) \end{pmatrix},$$

где $A(\overline{x},\, \overline{\mu})$ и $B(\overline{x},\, \overline{\mu})$ — диагональные $m \times m$-матрицы с определенными согласно (11), (12) диагональными элементами $a_i = a_i(\overline{x},\, \overline{\mu})$ и $b_i = b_i(\overline{x},\, \overline{\mu})$ соответственно, $i = 1, \ldots, m$.

Пусть $\widetilde{u} = (\widetilde{x},\, \widetilde{\lambda},\, \widetilde{\mu}) \in \ker H$; тогда

$$\frac{\partial^2 L}{\partial x^2}\,(\overline{x},\, \overline{\lambda},\, \overline{\mu})\,\widetilde{x} + (F'(\overline{x}))^{\mathrm{T}}\widetilde{\lambda} + (G'(\overline{x}))^{\mathrm{T}}\widetilde{\mu} = 0, \tag{26}$$

$$F'(\overline{x})\widetilde{x} = 0, \tag{27}$$

$$a_i\langle g_i'(\overline{x}),\, \widetilde{x}\rangle + b_i\widetilde{\mu}_i = 0, \quad i = 1, \ldots, m. \tag{28}$$

Заметим, что согласно (11), (12) для $i \in I_+(\overline{x})$ справедливо $a_i = 1$, $b_i = 0$. Поэтому из (28) следует, что

$$\langle g_i'(\overline{x}), \widetilde{x}\rangle = 0 \quad \forall\, i \in I_+(\overline{x}). \tag{29}$$

В частности, из (27) и (29) вытекает, что $\widetilde{x} \in K_+(\overline{x})$.

Далее, согласно (11), (12) для $i \in \{1, \ldots, m\} \setminus I(\overline{x})$ справедливы равенства $a_i = 0$, $b_i = -1$, а для $i \in I_0(\overline{x}) = \{i \in I(\overline{x}) \,\big|\, \overline{\mu}_i = 0\}$ справедливы неравенства $a_i \geqslant 1$, $b_i \leqslant 0$. Поэтому из (28) следует

$$\widetilde{\mu}_i = 0 \quad \forall\, i \in \{1, \ldots, m\} \setminus I(\overline{x}), \tag{30}$$

$$\widetilde{\mu}_i \langle g_i'(\overline{x}), \widetilde{x}\rangle \geqslant 0 \quad \forall\, i \in I_0(\overline{x}). \tag{31}$$

Умножая обе части (26) скалярно на $\widetilde{x}$ и принимая во внимание (27), (29), (30), получим

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})\, \widetilde{x}, \widetilde{x} \right\rangle + \sum_{i \in I_0(\overline{x})} \widetilde{\mu}_i \langle g_i'(\overline{x}), \widetilde{x}\rangle = 0.$$

Но тогда из (19) и (31) следует, что $\widetilde{x} = 0$. Кроме того, используя (30), из (26) выводим

$$(F'(\overline{x}))^{\mathrm{T}} \widetilde{\lambda} + \sum_{i \in I(\overline{x})} \widetilde{\mu}_i g_i'(\overline{x}) = 0.$$

Отсюда и из условия линейной независимости следует, что $\widetilde{\lambda} = 0$, $\widetilde{\mu}_i = 0$ $\forall\, i \in I(\overline{x})$, а значит, $\widetilde{u} = 0$. Таким образом, $\ker H = \{0\}$ $\forall\, H \in \partial\Phi(\overline{x}, \overline{\lambda}, \overline{\mu})$. $\square$

Алгоритм 2. Выбираем функцию дополнительности $\psi\colon \mathbf{R} \times \times \mathbf{R} \to \mathbf{R}$ согласно (6) или (7). Выбираем $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ и полагаем $k = 0$.

1. Вычисляем некоторую матрицу $H_k \in \mathbf{R}(n + l + m, n + l + m)$. Вычисляем $(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ как решение линейной системы
$$H_k((x, \lambda, \mu) - (x^k, \lambda^k, \mu^k)) = -\Phi(x^k, \lambda^k, \mu^k),$$
где Φ введено в (18).

2. Увеличиваем номер шага k на 1 и переходим к п. 1.

Если для каждого k в приведенном алгоритме выбирается $H_k \in \partial_B \Phi(x^k, \lambda^k, \mu^k)$ либо $H_k \in \partial\Phi(x^k, \lambda^k, \mu^k)$, то приходим к *обобщенному методу Ньютона для системы ККТ*. Для вычисления таких матриц могут использоваться формулы из задач 6 и 7, что делает метод вполне реализуемым (о выборе H_k еще будет сказано ниже). Результат о локальной сходимости и скорости сходимости обобщенного метода Ньютона для системы ККТ следует немедленно из теоремы 2, предложений 1–3 и результата задачи 16.

Теорема 4. *Пусть в дополнение к условиям предложения* 2 *функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображения* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *и* $G\colon \mathbf{R}^n \to \mathbf{R}^m$ *дважды дифференцируемы в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$, *причем их вторые производные непрерывны по Липшицу на этой окрестности.*

Тогда любое начальное приближение $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, *достаточно близкое к точке* $(\overline{x}, \overline{\lambda}, \overline{\mu})$, *корректно определяет сходящуюся к* $(\overline{x}, \overline{\lambda}, \overline{\mu})$ *траекторию алгоритма* 2, *в котором используются матрицы* $H_k \in \partial\Phi(x^k, \lambda^k, \mu^k)$ *для каждого* $k = 0, 1, \ldots$. *Скорость сходимости квадратичная.*

Поскольку из теоремы 4 не следует автоматически не только квадратичная, но даже сверхлинейная скорость сходимости алгоритма 2 в прямых переменных, рассмотрим отдельно вопрос о скорости сходимости в прямых переменных. Одновременно, как и в п. 4.4.2 для методов последовательного квадратичного программирования (SQP), рассмотрим квазиньютоновский вариант алгоритма 2, не требующий вычисления вторых производных функции f и отображений F и G. Для простоты ограничимся случаем использования функции естественной невязки. Будем выбирать H_k в соответствии со следующим правилом:

$$H_k = \begin{pmatrix} \Lambda_k & (F'(x^k))^{\mathrm{T}} & (G'_{J_+^k}(x^k))^{\mathrm{T}} & (G'_{J_-^k}(x^k))^{\mathrm{T}} \\ F'(x^k) & 0 & 0 & 0 \\ -G'_{J_+^k}(x^k) & 0 & 0 & 0 \\ 0 & 0 & 0 & E_{J_-^k} \end{pmatrix}, \qquad (32)$$

где $\Lambda_k \in \mathbf{R}(n, n)$ — некоторое приближение к $\dfrac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)$,

$$J_+^k = J_+(x^k, \mu^k) = \{i = 1, \ldots, m \mid \mu_i^k > -g_i(x^k)\},$$

$$J_-^k = J_-(x^k, \mu^k) = \{i = 1, \ldots, m \mid \mu_i^k \leqslant -g_i(x^k)\},$$

а $E_{J_-^k}$ — единичная $|J_-^k| \times |J_-^k|$-матрица.

Заметим, что если $\Lambda_k = \dfrac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)$, то, как следует из результата задачи 6, имеет место включение $H_k \in \partial\Phi(x^k, \lambda^k, \mu^k)$ (при соответствующей нумерации последних m компонент отображения Φ).

Теорема 5. *Пусть выполнены условия теоремы* 4. *Пусть, кроме того, траектория* $\{(x^k, \lambda^k, \mu^k)\}$ *алгоритма* 2, *в котором отображение* Φ *введено в* (18) *с использованием функции* ψ, *заданной в* (6), *и для каждого* $k = 0, 1, \ldots$ *матрица* H_k *выбирается согласно* (32), *сходится к* $(\overline{x}, \overline{\lambda}, \overline{\mu})$.

Тогда если скорость сходимости последовательности $\{x^k\}$ *к* $\overline{x}$ *является сверхлинейной, то*

$$\pi_{N(\overline{x})}\left(\left(\Lambda_k - \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})\right)(x^{k+1} - x^k)\right) = o(|x^{k+1} - x^k|), \quad (33)$$

где

$$N(\overline{x}) = \{h \in \ker F'(\overline{x}) \mid \langle g_i'(\overline{x}), h\rangle = 0 \quad \forall\, i \in I(\overline{x})\}.$$

Наоборот, если

$$\pi_{K_+(\overline{x})}\left(\left(\Lambda_k - \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})\right)(x^{k+1} - x^k)\right) = o(|x^{k+1} - x^k|), \quad (34)$$

то скорость сходимости $\{x^k\}$ *к* $\overline{x}$ *является сверхлинейной.*

Доказательство. В соответствии с алгоритмом 2 и (32) для каждого k

$$\Lambda_k(x^{k+1} - x^k) + (F'(x^k))^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) + (G'(x^k))^{\mathrm{T}}(\mu^{k+1} - \mu^k) = -\frac{\partial L}{\partial x}(x^k, \lambda^k, \mu^k), \quad (35)$$

$$F'(x^k)(x^{k+1} - x^k) = -F(x^k), \quad (36)$$

$$-\langle g_i'(x^k), x^{k+1} - x^k\rangle = -\min\{\mu_i^k, -g_i(x^k)\}, \quad i \in J_+^k, \quad (37)$$

$$\mu_i^{k+1} - \mu_i^k = -\min\{\mu_i^k, -g_i(x^k)\}, \quad i \in J_-^k. \quad (38)$$

Согласно определению J_+^k и J_-^k соотношения (37) и (38) переписываются в виде

$$g_i(x^k) + \langle g_i'(x^k), x^{k+1} - x^k\rangle = 0, \quad i \in J_+^k, \quad (39)$$

$$\mu_i^{k+1} = 0, \quad i \in J_-^k. \quad (40)$$

Кроме того, сходимость $\{(x^k, \mu^k)\}$ к $(\overline{x}, \overline{\mu})$ влечет включения

$$I_+(\overline{x}) \subset J_+^k, \quad \{1, \ldots, m\} \setminus I(\overline{x}) \subset J_-^k \quad (41)$$

для любого достаточно большого k.

Из (35) имеем

$$\begin{aligned} &-\Lambda_k(x^{k+1} - x^k) = f'(x^k) + (F'(x^k))^{\mathrm{T}}\lambda^{k+1} + (G'(x^k))^{\mathrm{T}}\mu^{k+1} = \\ &= \frac{\partial L}{\partial x}(x^k, \lambda^{k+1}, \mu^{k+1}) = \\ &= \frac{\partial L}{\partial x}(\overline{x}, \lambda^{k+1}, \mu^{k+1}) + \frac{\partial^2 L}{\partial x^2}(\overline{x}, \lambda^{k+1}, \mu^{k+1})(x^k - \overline{x}) + o(|x^k - \overline{x}|) = \\ &= (F'(\overline{x}))^{\mathrm{T}}(\lambda^{k+1} - \overline{\lambda}) + (G'(\overline{x}))^{\mathrm{T}}(\mu^{k+1} - \overline{\mu}) + \\ &\quad + \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})(x^k - \overline{x}) + o(|x^k - \overline{x}|), \end{aligned} \quad (42)$$

где учтено определение стационарной точки задачи (1), (2) и отвечающих этой точке множителей Лагранжа, а также сходимость

$\{(\lambda^k, \mu^k)\}$ к $(\overline{\lambda}, \overline{\mu})$. Из (42) имеем

$$\left(\frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu}) - \Lambda_k\right)(x^{k+1} - x^k) = (F'(\overline{x}))^{\mathrm{T}}(\lambda^{k+1} - \overline{\lambda}) +$$

$$+ (G'(\overline{x}))^{\mathrm{T}}(\mu^{k+1} - \overline{\mu}) + \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})(x^{k+1} - \overline{x}) + o(|x^k - \overline{x}|). \quad (43)$$

Предположим сначала, что скорость сходимости $\{x^k\}$ к $\overline{x}$ сверхлинейная, т. е. $|x^{k+1} - \overline{x}| = o(|x^k - \overline{x}|)$. Тогда (43) принимает вид

$$\left(\frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu}) - \Lambda_k\right)(x^{k+1} - x^k) =$$

$$= (F'(\overline{x}))^{\mathrm{T}}(\lambda^{k+1} - \overline{\lambda}) + (G'(\overline{x}))^{\mathrm{T}}(\mu^{k+1} - \overline{\mu}) + o(|x^k - \overline{x}|). \quad (44)$$

Согласно (40), (41) и условию дополняющей нежесткости $\forall\, h \in N(\overline{x})$ выполняется

$$\langle (F'(\overline{x}))^{\mathrm{T}}(\lambda^{k+1} - \overline{\lambda}) + (G'(\overline{x}))^{\mathrm{T}}(\mu^{k+1} - \overline{\mu}),\, h\rangle =$$

$$= \sum_{i \in I(\overline{x})} (\mu_i^{k+1} - \overline{\mu}_i)\langle g_i'(\overline{x}),\, h\rangle = 0, \quad (45)$$

а это означает, что

$$\pi_{N(\overline{x})}(F'(\overline{x}))^{\mathrm{T}}(\lambda^{k+1} - \overline{\lambda}) + (G'(\overline{x}))^{\mathrm{T}}(\mu^{k+1} - \overline{\mu})) = 0$$

(в данном случае $N(\overline{x})$ — линейное подпространство в $\mathbf{R}^n$, а $\pi_{N(\overline{x})}(\cdot)$ — оператор ортогонального проектирования на это подпространство). Но тогда из (44) имеем

$$\pi_{N(\overline{x})}\left(\left(\frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu}) - \Lambda_k\right)(x^{k+1} - x^k)\right) = o(|x^k - \overline{x}|).$$

Отсюда так же, как в соответствующей части доказательства теоремы 4.4.2, следует выполнение (33).

Пусть теперь выполнено (34). Прежде всего заметим, что согласно (36)

$$0 = F(x^k) + F'(x^k)(x^{k+1} - x^k) =$$

$$= F'(\overline{x})(x^k - \overline{x}) + F'(x^k)(x^{k+1} - x^k) + o(|x^k - \overline{x}|) =$$

$$= F'(\overline{x})(x^{k+1} - \overline{x}) + \eta^k, \quad (46)$$

где с учетом сходимости $\{x^k\}$ к $\overline{x}$

$$\eta^k = (F'(x^k) - F'(\overline{x}))(x^{k+1} - x^k) + o(|x^k - \overline{x}|) =$$

$$= o(|x^{k+1} - x^k|) + o(|x^k - \overline{x}|). \quad (47)$$

Аналогичным образом с помощью (39) устанавливается, что

$$0 = \langle g_i'(\overline{x}),\, x^{k+1} - \overline{x}\rangle + \zeta_i^k \quad \forall\, i \in J_+^k, \quad (48)$$

где

$$\zeta_i^k = o(|x^{k+1} - x^k|) + o(|x^k - \overline{x}|) \quad \forall\, i \in J_+^k. \tag{49}$$

Из условия линейной независимости и соотношений (46) и (48) следует, что существует элемент $\xi^k \in \mathbf{R}^n$ такой, что

$$F'(\overline{x})\xi^k = \eta^k, \quad \langle g_i'(\overline{x}),\, \xi^k\rangle = \zeta_i^k \quad \forall\, i \in J_+^k, \tag{50}$$

$$|\xi^k| = o(|x^{k+1} - x^k|) + o(|x^k - \overline{x}|). \tag{51}$$

Положим $h^k = x^{k+1} - \overline{x} + \xi^k$. Из (41) и (46)–(51) следует, что $h^k \in K_+(\overline{x})$, и аналогично равенству (45) выводится соотношение

$$\langle (F'(\overline{x}))^{\mathrm{T}}(\lambda^{k+1} - \overline{\lambda}),\, h^k\rangle + \langle (G'(\overline{x}))^{\mathrm{T}}(\mu^{k+1} - \overline{\mu}),\, h^k\rangle =$$

$$= \sum_{i \in I_0(\overline{x}) \cap J_+^k} (\mu_i^{k+1} - \overline{\mu}_i)\langle g_i'(\overline{x}),\, h^k\rangle + \sum_{i \in I_0(\overline{x}) \cap J_-^k} \mu_i^{k+1}\langle g_i'(\overline{x}),\, h^k\rangle = 0,$$

где последнее равенство следует из (40), (48) и (50).

Оставшаяся часть доказательства совершенно аналогична соответствующей части доказательства теоремы 4.4.2. $\square$

Кратко остановимся на сравнении достоинств и недостатков методов SQP и обобщенного метода Ньютона для системы KKT. С одной стороны, локальная сходимость со сверхлинейной скоростью (в прямодвойственных переменных) для обобщенного метода Ньютона доказана в теореме 4 лишь при выполнении условия линейной независимости и сильного достаточного условия второго порядка. В то же время, как отмечено в § 4.4, аналогичные свойства для метода SQP устанавливаются в существенно более слабых предположениях, а именно, при выполнении строгого условия регулярности Мангасариана–Фромовица и обычного достаточного условия второго порядка оптимальности. Кроме того, приведенное в теореме 5 условие (34), достаточное для наличия у обобщенного метода Ньютона сверхлинейной скорости сходимости в прямых переменных, несколько слабее соответствующего условия для методов SQP; см. теорему 4.4.2. С другой стороны, итерация обобщенного метода Ньютона менее трудоемка, чем у метода SQP, поскольку состоит в решении одной линейной системы, а не задачи квадратичного программирования.

Заметим также, что согласно результату задачи 6 отображение Φ, введенное в (18) с использованием функции ψ, заданной в (6), BD-регулярно в точке $(\overline{x},\, \overline{\lambda},\, \overline{\mu})$ тогда и только тогда, когда в этой точке выполнено так называемое *условие квазирегулярности*, состоящее

в следующем: матрица

$$\begin{pmatrix} \dfrac{\partial^2 L}{\partial x^2}(\overline{x},\,\overline{\lambda},\,\overline{\mu}) & (F'(\overline{x}))^{\mathrm{T}} & (G'_{I_+(\overline{x})}(\overline{x}))^{\mathrm{T}} & (G'_J(\overline{x}))^{\mathrm{T}} \\ F'(\overline{x}) & 0 & 0 & 0 \\ G'_{I_+(\overline{x})}(\overline{x}) & 0 & 0 & 0 \\ G'_J(\overline{x}) & 0 & 0 & 0 \end{pmatrix}$$

невырождена для любого множества индексов $J \subset I_0(\overline{x})$. Условие квазирегулярности содержит в себе условие линейной независимости, но при выполнении последнего является более слабым, чем сильное достаточное условие второго порядка. Тем не менее, как следует из теоремы 2, квазирегулярности достаточно для обоснования локальной сходимости соответствующей версии обобщенного метода Ньютона с квадратичной скоростью.

Что же касается глобальных свойств обсуждаемых методов, то, как неоднократно отмечалось выше, структура методов SQP допускает различные стратегии глобализации сходимости, естественным образом связанные с природой локального метода (см. § 5.4). Глобализация сходимости обобщенного метода Ньютона для системы ККТ также возможна, и в последнее время были найдены весьма эффективные схемы такой глобализации. Однако общим недостатком получаемых на этом пути алгоритмов является то, что они, в принципе, ориентированы на поиск стационарных точек задачи (1), (2), а не ее решений.

Требования гладкости в теореме 4 можно понизить в духе предложения 1, если говорить о сверхлинейной скорости сходимости вместо квадратичной. Это обобщение не вызывает затруднений и может быть реализовано читателем самостоятельно.

§ 4.6. Идентификация активных ограничений

В этом параграфе будем рассматривать задачу оптимизации с ограничениями-неравенствами

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in \mathbf{R}^n \,\big|\, G(x) \leqslant 0\}, \tag{2}$$

где $f\colon \mathbf{R}^n \to \mathbf{R}$ — гладкая функция, $G\colon \mathbf{R}^n \to \mathbf{R}^m$ — гладкое отображение с компонентами $g_i(\cdot)$, $i = 1, \ldots, m$. Проводимые ниже построения распространяются на случай смешанных ограничений очевидным образом; все принципиальные моменты здесь касаются именно ограничений-неравенств, поэтому ограничения-равенства не рассматриваются.

Пусть $\overline{x} \in \mathbf{R}^n$ — стационарная точка задачи (1), (2). Цель состоит в идентификации множества $I(\overline{x}) = \{i = 1, \ldots, m \mid g_i(\overline{x}) = 0\}$ индексов активных ограничений в точке $\overline{x}$. Такая идентификация может быть полезна в ряде случаев. Например, она позволяет сводить задачу с ограничениями-неравенствами к задаче с ограничениями-равенствами в следующем смысле: стационарная точка $\overline{x}$ задачи (1), (2) будет стационарной и в задаче

$$f(x) \to \min, \quad x \in \overline{D},$$

$$\overline{D} = \{x \in \mathbf{R}^n \mid g_i(x) = 0,\ i \in I(\overline{x})\}.$$

Задача с ограничениями-равенствами в принципе проще; например, к ней напрямую применимы методы, рассмотренные в § 4.3.

Приведенное соображение используется во многих алгоритмах решения задачи (1), (2). Методы, использующие оценки множества индексов активных ограничений, принято называть *методами активного множества* [1]. Типичными представителями этого класса методов являются рассматриваемые ниже симплекс-метод для задач линейного программирования и метод особых точек для задач квадратичного программирования (см. § 7.2 и § 7.3). Эти два метода являются конечными: после правильной идентификации множества $I(\overline{x})$ решение находится за конечное число шагов (или вообще за один шаг), поскольку соответствующая задача с ограничениями-равенствами очень проста. Для более общих (нелинейных и неквадратичных) задач оптимизации методы активного множества носят бесконечношаговый характер.

Разумеется, речь идет о возможности идентификации $I(\overline{x})$ без точного знания $\overline{x}$. Точнее, пусть разрешено использовать информацию, доступную в точке $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m$, где x — некоторое приближение к $\overline{x}$, а μ — приближение к (также неизвестному) множителю Лагранжа $\overline{\mu}$, отвечающему $\overline{x}$.

4.6.1. Идентификация, основанная на оценках расстояния. Стационарные точки задачи (1), (2) и отвечающие им множители Лагранжа описываются системой Каруша–Куна–Таккера

$$\frac{\partial L}{\partial x}(x, \mu) = 0, \tag{3}$$

$$\mu \geqslant 0, \quad G(x) \leqslant 0, \quad \langle \mu, G(x) \rangle = 0, \tag{4}$$

где

$$L\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}, \quad L(x, \mu) = f(x) + \langle \mu, G(x) \rangle,$$

— функция Лагранжа.

[1] Общепринятый английский термин — Active-set methods.

Пусть $\mathcal{M}(\overline{x})$ — множество множителей Лагранжа, отвечающих изолированной стационарной точке $\overline{x}$ задачи (1), (2), т. е. множество таких $\overline{\mu} \in \mathbf{R}^m$, что пара $(\overline{x}, \overline{\mu})$ является решением системы (3), (4) (множество $\mathcal{M}(\overline{x})$ не обязательно является одноточечным). Кроме того, введем множество

$$\Omega(\overline{x}) = \{\overline{x}\} \times \mathcal{M}(\overline{x}).$$

Сразу заметим, что задача идентификации активных ограничений не является сложной, если предполагать, что для некоторого $\overline{\mu} \in \mathcal{M}(\overline{x})$ выполнено условие строгой дополнительности

$$\overline{\mu}_i > 0 \quad \forall\, i \in I(\overline{x})$$

и что известна точка, достаточно близкая к $(\overline{x}, \overline{\mu})$ с таким $\overline{\mu}$. Действительно, легко видеть, что при этом найдется окрестность U точки $(\overline{x}, \overline{\mu})$ такая, что

$$I(x, \mu) = I(\overline{x}) \quad \forall\, (x, \mu) \in U,$$

где

$$I(x, \mu) = \{i = 1, \ldots, m \mid \mu_i \geqslant -g_i(x)\}, \quad x \in \mathbf{R}^n, \ \mu \in \mathbf{R}^m.$$

В общем случае можно гарантировать лишь выполнение включения

$$I(x, \mu) \subset I(\overline{x}) \quad \forall\, (x, \mu) \in U.$$

Заметим, наконец, что если $\mathcal{M}(\overline{x}) = \{\overline{\mu}\}$ при некотором $\overline{\mu} \in \mathbf{R}^m$, то окрестность U всегда можно выбрать так, чтобы выполнялось

$$I_+(\overline{x}) \subset I(x, \mu) \subset I(\overline{x}) \quad \forall\, (x, \mu) \in U,$$

где

$$I_+(\overline{x}) = \{i \in I(\overline{x}) \mid \overline{\mu}_i > 0\}.$$

Теперь покажем, как активные ограничения могут быть идентифицированы при разумных предположениях, не включающих в себя ни условие строгой дополнительности, ни единственность множителя Лагранжа, отвечающего искомой стационарной точке. Основная идея состоит в том, чтобы сравнивать значения $-g_i(x)$ не с μ_i, $i = 1, \ldots, m$, а со значениями некоторой «идентифицирующей функции», поведение которой вблизи множества $\Omega(\overline{x})$ известно. Такие функции обычно возникают при получении оценок расстояния до множества $\Omega(\overline{x})$. Несмотря на то, что эта техника чрезвычайно проста, она в действительности приводит к наиболее сильным из известных результатов об идентификации активных ограничений. Кроме того, эта техника не связана ни с каким конкретным алгоритмом и может быть использована в различных алгоритмах решения задачи (1), (2).

Будем говорить, что функция $r\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}_+$ *оценивает расстояние* до множества $\Omega(\overline{x})$, если $r(\overline{x}, \overline{\mu}) = 0 \quad \forall\, \overline{\mu} \in \mathcal{M}(\overline{x})$,

функция r непрерывна на $\Omega(\overline{x})$ и существуют числа $\delta > 0$, $M > 0$ и $\nu > 0$ такие, что

$$\operatorname{dist}((x, \mu), \Omega(\overline{x})) \leqslant M(r(x, \mu))^{\nu} \quad \forall\, (x, \mu) \in U(\Omega(\overline{x}), \delta) = \\ = \{(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m \,\big|\, \operatorname{dist}((x, \mu), \Omega(\overline{x})) \leqslant \delta\}. \quad (5)$$

Функции с такими свойствами часто могут быть указаны (см. п. 4.6.2). Здесь существенно, что $\overline{x}$ — изолированная стационарная точка; в противном случае следует оценивать расстояние до множества решений системы (3), (4). Фигурирующие в (5) константы M и/или ν на практике часто бывают не известны, однако их знание и не предполагается в описываемой ниже процедуре идентификации.

Во многих случаях можно использовать локальный вариант понятия оценки расстояния, заменив $U(\Omega(\overline{x}), \delta)$ в (5) на некоторую окрестность точки $(\overline{x}, \overline{\mu}) \in \Omega(\overline{x})$ при фиксированном $\overline{\mu} \in \mathcal{M}(\overline{x})$ (см. доказательства предложений 1 и 2).

Оценки расстояния (до различных множеств) возникают и используются в этом курсе неоднократно; см., например: теорему 1.3.3 об оценке расстояния до множества нулей гладкого отображения (обобщением этого результата на случай смешанных ограничений является теорема 4.7.6); теорему 4.7.5, где более общая оценка такого рода будет использована при анализе скорости сходимости методов штрафов; теорему 4.5.3 об оценке расстояния до изолированного нуля негладкого отображения; условие квадратичного роста и связанные с ним условия в п. 3.1.2, которые являются оценками расстояния до множества решений или стационарных точек задачи безусловной оптимизации.

Положим

$$I(x, \mu) = \{i = 1, \ldots, m \,\big|\, \rho(r(x, \mu)) \geqslant -g_i(x)\}, \quad x \in \mathbf{R}^n, \quad \mu \in \mathbf{R}^m. \quad (6)$$

Здесь используется функция

$$\rho\colon \mathbf{R}_+ \to \mathbf{R}, \quad \rho(t) = \begin{cases} -1/\ln \widehat{t}, & \text{если } t \geqslant \widehat{t}, \\ -1/\ln t, & \text{если } t \in (0, \widehat{t}), \\ 0, & \text{если } t = 0, \end{cases} \quad (7)$$

где $\widehat{t} \in (0, 1)$ — заданный параметр.

Заметим, что функция ρ непрерывна на $\mathbf{R}_+$. Утверждается, что введенное множество индексов $I(x, \mu)$ правильно идентифицирует активные ограничения для всякой точки (x, μ), достаточно близкой к $\Omega(\overline{x})$.

Предложение 1. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображение* $G\colon \mathbf{R}^n \to \mathbf{R}^m$ *дифференцируемы в точке* $\overline{x} \in \mathbf{R}^n$, *причем существуют окрестность* V *точки* $\overline{x}$ *и число* $L > 0$ *такие, что*

$$|G(x) - G(\overline{x})| \leqslant L|x - \overline{x}| \quad \forall\, x \in V. \tag{8}$$

Пусть $\overline{x}$ — *стационарная точка задачи* (1), (2), *причем функция* $r\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}_+$ *оценивает расстояние до множества* $\Omega(\overline{x})$.

Тогда для каждого $\overline{\mu} \in \mathcal{M}(\overline{x})$ *найдется окрестность* U *точки* $(\overline{x}, \overline{\mu})$ *такая, что для заданного формулами* (6), (7) *множества индексов* $I(x, \mu)$ *справедливо*

$$I(x, \mu) = I(\overline{x}) \quad \forall\, (x, \mu) \in U. \tag{9}$$

Доказательство. Пусть сначала $i \in I(\overline{x})$. Тогда если точка $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m$ достаточно близка к $(\overline{x}, \overline{\mu})$, то в силу (5), (8) имеем

$$-g_i(x) = g_i(\overline{x}) - g_i(x) \leqslant L|x - \overline{x}| \leqslant LM(r(x, \mu))^\nu \leqslant \rho(r(x, \mu)),$$

где приняты во внимание непрерывность и равенство нулю r на $\Omega(\overline{x})$, а также то обстоятельство, что, каким бы ни было число $\nu > 0$, имеет место предельное соотношение $\lim\limits_{t \to 0+} t^\nu \ln t = 0$. Таким образом, $i \in I(x, \mu)$, т. е. $I(\overline{x}) \subset I(x, \mu)$.

С другой стороны, если $i \in \{1, \ldots, m\} \setminus I(\overline{x})$, то найдутся окрестность U точки $(\overline{x}, \overline{\mu})$ и число $\gamma > 0$ такие, что $\forall\, (x, \mu) \in U$ имеет место

$$\rho(r(x, \mu)) < \gamma, \quad -g_i(x) \geqslant \gamma,$$

т. е. $i \notin I(x, \mu)$, а значит, $I(x, \mu) \subset I(\overline{x})$. $\square$

Легко видеть, что вместо функции ρ, введенной в (7), в определении $I(x, \mu)$ можно использовать, например, функцию

$$\rho\colon \mathbf{R}_+ \to \mathbf{R}, \quad \rho(t) = t^\theta, \tag{10}$$

где $\theta \in (0, \nu)$. Такой выбор может иметь свои преимущества, однако он предполагает знание показателя ν в (5).

Задача 1. Доказать, что если в дополнение к условиям предложения 1 в точке $\overline{x}$ выполнено условие регулярности Мангасариана–Фромовица, то найдется число $\delta > 0$ такое, что (9) имеет место при $U = U(\Omega(\overline{x}), \delta)$.

Если $\mathcal{M}(\overline{x}) = \{\overline{\mu}\}$ при некотором $\overline{\mu} \in \mathbf{R}^m$, т. е. стационарной точке $\overline{x}$ отвечает единственный множитель, то можно идентифицировать и множество $I_+(\overline{x})$ (а значит, и $I_0(\overline{x}) = I(\overline{x}) \setminus I_+(\overline{x})$). Эта информация также может быть полезна; например, в системе (3), (4) можно сразу положить $\mu_i = 0$, $i \in I_0(\overline{x})$. Положим

$$I_+(x, \mu) = \{i = 1, \ldots, m \mid \mu_i \geqslant \rho(r(x, \mu))\}, \quad x \in \mathbf{R}^n, \quad \mu \in \mathbf{R}^m. \tag{11}$$

Предложение 2. *В условиях предложения* 1, *если стационарной точке* $\overline{x}$ *задачи* (1), (2) *отвечает единственный множитель Лагранжа* $\overline{\mu}$, *то найдется окрестность* U *точки* $(\overline{x}, \overline{\mu})$ *такая, что для заданного формулами* (7), (11) *множества индексов* $I_+(x, \mu)$ *справедливо*

$$I_+(x, \mu) = I_+(\overline{x}) \quad \forall\, (x, \mu) \in U.$$

Доказательство. Включение $I_+(\overline{x}) \subset I_+(x, \mu)$ для точек $(x, \mu) \in \mathbf{R}^n \times \mathbf{R}^m$, достаточно близких к $(\overline{x}, \overline{\mu})$, очевидно. С другой стороны, для таких (x, μ), если $i \in \{1, \ldots, m\} \setminus I_+(\overline{x})$, то

$$\mu_i \leqslant |\mu_i - \overline{\mu}_i| \leqslant M(r(x, \mu))^\nu < \rho(r(x, \mu)),$$

т. е. $i \notin I_+(x, \mu)$, а значит, $I_+(x, \mu) \subset I_+(\overline{x})$. $\square$

Предложение 2 останется верным, если вместо (7) использовать формулу (10) при $\theta \in (0, \nu)$.

4.6.2. Оценки расстояния. Известно, что при подходящем выборе функции r оценки расстояния типа (5) имеют место во многих важных случаях. Не претендуя на детальный анализ этого вопроса и не пытаясь указать все относящиеся к нему результаты, ограничимся лишь некоторыми примерами.

Обычно функция r определяется через ту или иную невязку системы (3), (4), т. е.

$$r(x, \mu) = |\Phi(x, \mu)|, \quad x \in \mathbf{R}^n, \quad \mu \in \mathbf{R}^m, \tag{12}$$

где отображение $\Phi\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^s$ (при некотором s) выбирается так, чтобы множество решений системы (3), (4) совпадало с множеством решений уравнения

$$\Phi(x, \mu) = 0.$$

Например, при соответствующих требованиях гладкости на f и G оценка (5) будет иметь место при $\nu = 1$, если r выбрано в соответствии с (12),

$$\Phi\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^m, \quad \Phi(x, \mu) = \begin{pmatrix} \dfrac{\partial L}{\partial x}(x, \mu) \\ \min\{\mu_1, -g_1(x)\} \\ \ldots \\ \min\{\mu_m, -g_m(x)\} \end{pmatrix}, \tag{13}$$

и в точке $\overline{x}$ выполнено условие линейной независимости (гарантирующее единственность отвечающего $\overline{x}$ множителя Лагранжа $\overline{\mu}$), а также сильное достаточное условие второго порядка. Действительно, Φ полугладко в точке $(\overline{x}, \overline{\mu})$ согласно предложению 4.5.1 и *BD*-

регулярно в этой точке согласно предложению 4.5.2 [1]. Но тогда требуемый результат следует из теоремы 4.5.3, определения полугладкости и утверждения б) задачи 4.5.8. Заметим, что условие линейной независимости здесь можно опустить (отказавшись тем самым от единственности множителя и BD-регулярности Φ), а сильное достаточное условие заменить обычным, однако необходимый для этого анализ требует привлечения результатов о чувствительности, выходящих за рамки данного курса.

Указанная оценка расстояния использует функцию естественной невязки, однако, разумеется, при подходящих предположениях можно использовать и другие функции дополнительности $\psi\colon \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ (см. п. 4.5.1), полагая

$$\Phi(x,\, \mu) = \begin{pmatrix} \dfrac{\partial L}{\partial x}(x,\, \mu) \\ \psi(\mu_1,\, -g_1(x)) \\ \dots \\ \psi(\mu_m,\, -g_m(x)) \end{pmatrix}, \quad x \in \mathbf{R}^n, \quad \mu \in \mathbf{R}^m.$$

Оценка (5) будет иметь место при $\nu = 1$ и для такого Φ, если r выбрано в соответствии с (12), ψ — полугладкая в точке $(0,\, 0)$ функция и Φ BD-регулярно в точке $(\overline{x},\, \overline{\mu})$.

Оценки расстояния другого важного типа могут быть получены в случае, когда функция f и отображение G являются вещественно аналитическими в некоторой окрестности точки $\overline{x}$, т. е. представляются в этой окрестности абсолютно и равномерно сходящимися степенными рядами. В этом случае (5) имеет место, если r выбрано в соответствии с (12), $\Phi\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^m \times \mathbf{R}$,

$$\Phi(x,\, \mu) = \begin{pmatrix} \dfrac{\partial L}{\partial x}\,(x,\, \mu) \\ \min\{\mu_1,\, 0\} \\ \dots \\ \min\{\mu_m,\, 0\} \\ \max\{g_1(x),\, 0\} \\ \dots \\ \max\{g_m(x),\, 0\} \\ \langle \mu,\, G(x) \rangle \end{pmatrix},$$

однако показатель ν здесь, вообще говоря, неизвестен.

[1] Согласно комментариям в конце п. 4.5.3, комбинацию условия линейной независимости и сильного достаточного условия второго порядка здесь можно заменить несколько более слабым предположением, а именно, условием квазирегулярности.

§ 4.7. Штрафы и модифицированные функции Лагранжа для задачи со смешанными ограничениями

Будем рассматривать задачу

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in \mathbf{R}^n \mid F(x) = 0,\ G(x) \leqslant 0\}, \tag{2}$$

где $f\colon \mathbf{R}^n \to \mathbf{R}$ — гладкая функция, $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ — гладкие отображения. Этот параграф посвящен распространению результатов о штрафах и модифицированных функциях Лагранжа (и соответствующих методах), полученных в пп. 4.3.2, 4.3.3 для задачи с чистыми ограничениями-равенствами, на задачу (1), (2), содержащую также ограничения-неравенства. При этом будет использоваться следующий прием, уже упоминавшийся в § 1.4.

От наличия ограничений-неравенств всегда можно избавиться введением дополнительных переменных. Пусть отображение G имеет компоненты $g_i(\cdot)$, $i = 1, \ldots, m$. Задаче (1), (2) сопоставим задачу

$$f(x) \to \min, \quad (x,\, \sigma) \in \widetilde{D}, \tag{3}$$

$$\widetilde{D} = \{(x,\, \sigma) \in \mathbf{R}^n \times \mathbf{R}^m \mid F(x) = 0,\ g_i(x) + \sigma_i^2 = 0,\ i = 1, \ldots, m\}. \tag{4}$$

Заметим, что, во-первых, для любого $x \in \mathbf{R}^n$ условием $(x,\, \sigma) \in \widetilde{D}$ вектор σ определяется с точностью до знаков своих компонент, а во-вторых, выбор этих знаков не влияет на значение целевой функции задачи (3), (4) в точке $(x,\, \sigma)$. Поэтому все дальнейшие рассуждения проводятся «по модулю» этих знаков, которые не играют никакой роли: точки $(x,\, \sigma^1)$ и $(x,\, \sigma^2)$, отличающиеся лишь знаками компонент σ^1, $\sigma^2 \in \mathbf{R}^m$, можно отождествлять. В этом смысле локальные решения задач (1), (2) и (3), (4) находятся во взаимно однозначном соответствии: точка $\overline{x} \in \mathbf{R}^n$ является локальным решением задачи (1), (2) в том и только том случае, когда точка $(\overline{x},\, \overline{\sigma})$, где

$$\overline{\sigma} = (\sqrt{-g_1(\overline{x})}, \ldots, \sqrt{-g_m(\overline{x})}), \tag{5}$$

корректно определена и является локальным решением задачи (3), (4) (см. задачу 2 ниже), причем других локальных решений вида $(\overline{x},\, \sigma)$, $\sigma \in \mathbf{R}^m_+$, у этой задачи быть не может.

Впрочем, в силу причин, указанных в § 1.4 и § 4.3, злоупотребления этим приемом желательно избегать.

4.7.1. Штрафы. *Штрафом* для произвольного множества $D \subset \mathbf{R}^n$ называется любая функция $\psi\colon \mathbf{R}^n \to \mathbf{R}$, удовлетворяющая условиям

$$\psi(x) = 0 \quad \forall\, x \in D, \quad \psi(x) > 0 \quad \forall\, x \in \mathbf{R}^n \setminus D. \tag{6}$$

Соответствующее семейство *штрафных функций* имеет вид

$$\varphi_c\colon \mathbf{R}^n \to \mathbf{R}, \quad \varphi_c(x) = f(x) + c\psi(x), \tag{7}$$

а соответствующий *метод штрафов* для задачи (1) состоит в последовательном решении вспомогательных задач безусловной оптимизации

$$\varphi_c(x) \to \min, \quad x \in \mathbf{R}^n, \tag{8}$$

при все возрастающем значении *параметра штрафа* $c \geqslant 0$.

Алгоритм 1. Фиксируем функцию $\psi\colon \mathbf{R}^n \to \mathbf{R}$, удовлетворяющую (6). Выбираем последовательность $\{c_k\} \subset \mathbf{R}_+$ такую, что $c_k \to \infty$ $(k \to \infty)$, и полагаем $k = 0$.

1. Вычисляем $x^k \in \mathbf{R}^n$ как стационарную точку задачи (8) с целевой функцией, задаваемой формулой (7) при $c = c_k$.

2. Увеличиваем номер шага k на 1 и переходим к п. 1.

Начнем со следующего обобщения теоремы 4.3.3. Как указано в задаче 4.3.3, ни специфика ограничений, задающих допустимое множество, ни конкретный вид штрафа на самом деле не играют в этой теореме никакой роли. Поэтому справедлива

Теорема 1. *Пусть функция* $f\colon X \to \mathbf{R}$ *непрерывна в окрестности точки* $\overline{x} \in X$, $D \subset X$ *— заданное множество. Пусть точка* $\overline{x}$ *является строгим локальным решением задачи* (1).

Тогда если функция $\psi\colon \mathbf{R}^n \to \mathbf{R}$ *удовлетворяет* (6) *и непрерывна в окрестности точки* $\overline{x}$, *то найдется такое число* $\delta > 0$, *что для* $c \geqslant 0$ *и для любого (глобального) решения* $x(c)$ *задачи*

$$\varphi_c(x) \to \min, \quad x \in \overline{B}(\overline{x}, \delta),$$

целевая функция которой задается формулой (7), *имеет место*

$$\|x(c) - \overline{x}\| \to 0 \quad (c \to +\infty).$$

В частности, для любого достаточно большого числа c *точка* $x(c)$ *является локальным решением задачи* (8).

Задача 1. Предполагая дифференцируемость функции $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображений $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ в некоторой окрестности локального решения $\overline{x} \in \mathbf{R}^n$ задачи (1), (2) и непрерывность их производных в точке $\overline{x}$, вывести из теорем 1 и 1.2.3 сформулированное в теореме 1.4.3 необходимое условие оптимальности Ф. Джона. Для этого рассмотреть вспомогательную задачу оптимизации

$$f(x) + |x - \overline{x}|^2 \to \min, \quad x \in D,$$

где множество D задано в (2).

Важнейший класс штрафов для заданного в (2) множества D вводится следующим образом. Зададим отображение

$$\Psi\colon \mathbf{R}^n \to \mathbf{R}^l \times \mathbf{R}^m,$$

$$\Psi(x) = (F(x),\ (\max\{0,\ g_1(x)\}, \ldots, \max\{0,\ g_m(x)\})), \tag{9}$$

и положим

$$\psi(x) = |\Psi(x)|^p, \quad x \in \mathbf{R}^n, \tag{10}$$

где $p > 0$ — фиксированный показатель степени. В (10) удобно заменить евклидову норму $|\cdot| = |\cdot|_2$ на $|\cdot|_p$ либо $|\cdot|_\infty$ (напомним, что $|\cdot|_p$ является нормой лишь при $p \geqslant 1$, но здесь эта функция, формально определяемая как корень степени p из суммы возведенных в степень p модулей компонент ее аргумента, будет использоваться и при $p \in \in (0,\ 1)$). Получаемые таким образом функции

$$\psi(x) = |\Psi(x)|_p^p = |F(x)|_p^p + \sum_{i=1}^{m} (\max\{0,\ g_i(x)\})^p, \quad x \in \mathbf{R}^n, \tag{11}$$

и

$$\psi(x) = |\Psi(x)|_\infty^p = \\ = (\max\{|F(x)|_\infty,\ 0,\ g_1(x), \ldots, g_m(x)\})^p, \quad x \in \mathbf{R}^n, \tag{12}$$

естественно называть *степенными штрафами*. При $p = 2$ первая из этих формул приводит к *квадратичному штрафу*

$$\psi(x) = \frac{1}{2}\,|\Psi(x)|^2 = \frac{1}{2}\left(|F(x)|^2 + \sum_{i=1}^{m} (\max\{0,\ g_i(x)\})^2\right), \quad x \in \mathbf{R}^n \tag{13}$$

(множитель $1/2$ введен для удобства). Алгоритм 1 с таким выбором функции ψ называют *методом квадратичного штрафа* для задачи (1), (2). Именно об этом методе и пойдет речь в оставшейся части этого пункта.

Комментарии, которыми следовало бы сопроводить метод квадратичного штрафа, по сути дела те же, что и в п. 4.3.2. Важное отличие состоит в следующем. Функция φ_c, вводимая в соответствии с (7), (13), дифференцируема при дифференцируемых f, F и G, но может не иметь второй производной в точках $x \in \mathbf{R}^n$, в которых непусто множество индексов активных ограничений-неравенств $I(x) = \\ = \{i = 1, \ldots, m \mid g_i(x) = 0\}$ задачи (1), (2), какими бы гладкими ни были f, F и G. Это обстоятельство должно учитываться при выборе методов безусловной оптимизации для решения задачи (8).

Приводимые ниже теоремы 2 и 3 обобщают теоремы 4.3.2 и 4.3.4 соответственно на случай наличия ограничений-неравенств. Такие обобщения проще всего получить с помощью приема, упомянутого в начале этого параграфа. Для этого потребуются некоторые вспомогательные факты о связи между задачами (1), (2) и (3), (4).

Задача 2. Доказать, что для любой функции $f\colon \mathbf{R}^n \to \mathbf{R}$ и любых отображений $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ точка $\overline{x} \in \mathbf{R}^n$ является (строгим) локальным решением задачи (1), (2) в том и только том случае, когда точка $(\overline{x}, \overline{\sigma})$, где $\overline{\sigma} \in \mathbf{R}^m$ введено в (5), корректно определена и является (строгим) локальным решением задачи (3), (4).

Пусть

$$L\colon \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R},$$

$$L(x, \lambda, \mu) = f(x) + \langle \lambda, F(x) \rangle + \langle \mu, G(x) \rangle,$$

— функция Лагранжа задачи (1), (2). Введем также функцию Лагранжа задачи (3), (4):

$$\widetilde{L}\colon (\mathbf{R}^n \times \mathbf{R}^m) \times (\mathbf{R}^l \times \mathbf{R}^m) \to \mathbf{R}, \quad \widetilde{L}((x, \sigma), (\lambda, \mu)) =$$

$$= f(x) + \langle \lambda, F(x) \rangle + \sum_{i=1}^{m} \mu_i (g_i(x) + \sigma_i^2) = L(x, \lambda, \mu) + \sum_{i=1}^{m} \mu_i \sigma_i^2.$$

Лемма 1. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображения* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *и* $G\colon \mathbf{R}^n \to \mathbf{R}^m$ *дифференцируемы в точке* $\overline{x} \in \mathbf{R}^n$.

Тогда:

а) *если точка* $\overline{x}$ *является стационарной в задаче* (1), (2) *и ей отвечают множители Лагранжа* $\overline{\lambda} \in \mathbf{R}^l$ *и* $\overline{\mu} \in \mathbf{R}^m_+$, *то точка* $(\overline{x}, \overline{\sigma})$, *где* $\overline{\sigma} \in \mathbf{R}^m$ *введено в* (5), *корректно определена, является стационарной в задаче* (3), (4) *и ей отвечает множитель Лагранжа* $(\overline{\lambda}, \overline{\mu})$, *т. е.* $(\overline{x}, \overline{\sigma}) \in \widetilde{D}$ *и*

$$\frac{\partial \widetilde{L}}{\partial (x, \sigma)} ((\overline{x}, \overline{\sigma}), (\overline{\lambda}, \overline{\mu})) = 0;$$

б) *если для некоторого* $\overline{\sigma} \in \mathbf{R}^m$ *точка* $(\overline{x}, \overline{\sigma})$ *является стационарной в задаче* (3), (4) *и ей отвечает множитель Лагранжа* $(\overline{\lambda}, \overline{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$, *то* $\overline{x} \in D$ *и*

$$\frac{\partial L}{\partial x} (\overline{x}, \overline{\lambda}, \overline{\mu}) = 0, \tag{14}$$

$$\overline{\mu}_i g_i(\overline{x}) = 0 \quad \forall\, i = 1, \ldots, m. \tag{15}$$

Подчеркнем, что в б) не утверждается стационарность точки $\overline{x}$ в задаче (1), (2) в полном смысле: множитель $\overline{\mu}$ может иметь отрицательные компоненты.

Доказательство. Стационарность точки $\overline{x}$ в задаче (1), (2) с множителями $\overline{\lambda} \in \mathbf{R}^l$ и $\overline{\mu} \in \mathbf{R}^m_+$ означает выполнение условий допустимости

$$F(\overline{x}) = 0, \quad G(\overline{x}) \leqslant 0, \tag{16}$$

равенства (14) и условия дополняющей нежесткости (15).

Из (5) и (16) следует допустимость точки $(\overline{x}, \overline{\sigma})$ в задаче (3), (4). Кроме того, в силу (14) и (15) имеем

$$\frac{\partial \widetilde{L}}{\partial x}((\overline{x}, \overline{\sigma}), (\overline{\lambda}, \overline{\mu})) = \frac{\partial L}{\partial x}(\overline{x}, \overline{\lambda}, \overline{\mu}) = 0,$$

$$\frac{\partial \widetilde{L}}{\partial \sigma_i}((\overline{x}, \overline{\sigma}), (\overline{\lambda}, \overline{\mu})) = 2\overline{\mu}_i\overline{\sigma}_i = 0 \quad \forall\, i = 1, \ldots, m.$$

Утверждение а) доказано. Утверждение б) получается обратным рассуждением. □

Для произвольной точки $x \in \mathbf{R}^n$ введем множество $I^+(x) = \{i = 1, \ldots, m \mid g_i(x) > 0\}$ индексов тех ограничений-неравенств задачи (1), (2), которые нарушаются в точке x. Будем говорить, что в точке x выполнено *условие линейной независимости*, если строки матрицы $F'(x)$ и векторы $g_i'(x)$, $i \in I(x) \cup I^+(x)$, образуют линейно независимую систему в $\mathbf{R}^n$. Если точка x допустима в задаче (1), (2), то $I^+(x) = \varnothing$, и введенное условие линейной независимости совпадает с использовавшимся выше.

Задача 3. Доказать, что для любых отображений $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$, дифференцируемых в точке $x \in \mathbf{R}^n$, выполнение условия линейной независимости в точке x влечет выполнение условия регулярности ограничений задачи (3), (4) в точке (x, σ) при любом $\sigma \in \mathbf{R}^m$ таком, что $\sigma_i \neq 0 \quad \forall\, i \in \{1, \ldots, m\} \setminus (I(x) \cup I^+(x))$.

Лемма 2. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображения $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ дважды дифференцируемы в точке $\overline{x} \in \mathbf{R}^n$. Пусть в точке $\overline{x}$ выполнено условие линейной независимости, причем $\overline{x}$ — стационарная точка задачи (1), (2), а $\overline{\lambda} \in \mathbf{R}^l$ и $\overline{\mu} \in \mathbf{R}^m_+$ — однозначно отвечающие ей множители Лагранжа. Пусть, наконец, в точке $\overline{x}$ выполнено сформулированное в теореме 1.4.5 достаточное условие второго порядка оптимальности, а также условие строгой дополнительности.*

Тогда точка $(\overline{x}, \overline{\sigma})$, где $\overline{\sigma} \in \mathbf{R}^m$ введено в (5), корректно определена, в ней выполнено условие регулярности ограничений задачи (3), (4), $(\overline{x}, \overline{\sigma})$ является стационарной точкой этой задачи и ей однозначно отвечает множитель Лагранжа $(\overline{\lambda}, \overline{\mu})$. Кроме того, в этой точке для задачи (3), (4) выполнено сформулированное в теореме 1.3.7 достаточное условие второго порядка оптимальности,

т. е.

$$\left\langle \frac{\partial^2 \widetilde{L}}{\partial (x,\sigma)^2}\,((\overline{x},\overline{\sigma}),(\overline{\lambda},\overline{\mu}))(h,\tau),(h,\tau)\right\rangle > 0 \tag{17}$$

для всех $(h,\tau)\in(\mathbf{R}^n\times\mathbf{R}^m)\setminus\{0\}$ *таких, что*

$$F'(\overline{x})h=0,\quad \langle g_i'(\overline{x}),h\rangle+2\overline{\sigma}_i\tau_i=0\quad \forall\, i=1,\ldots,m. \tag{18}$$

Доказательство. Стационарность точки $(\overline{x},\,\overline{\sigma})$ с множителем Лагранжа $(\overline{\lambda},\,\overline{\mu})$ установлена в утверждении а) леммы 1, а выполнение в этой точке условия регулярности ограничений вытекает из результата задачи 3.

Пусть пара $(h,\,\tau)\in(\mathbf{R}^n\times\mathbf{R}^m)\setminus\{0\}$ удовлетворяет (18). Прямым вычислением убеждаемся, что

$$\left\langle \frac{\partial^2 \widetilde{L}}{\partial (x,\,\sigma)^2}\,((\overline{x},\,\overline{\sigma}),\,(\overline{\lambda},\,\overline{\mu}))(h,\,\tau),\,(h,\,\tau)\right\rangle = \left\langle \frac{\partial^2 L}{\partial x^2}\,(\overline{x},\,\overline{\lambda},\,\overline{\mu})h,\,h\right\rangle + 2\sum_{i=1}^{m}\overline{\mu}_i\tau_i^2. \tag{19}$$

Из (5), (18) и условия строгой дополнительности следует, что $h\in K(\overline{x})$, где, как обычно, через

$$K(\overline{x})=\{h\in\ker F'(\overline{x})\,\big|\,\langle g_i'(\overline{x}),\,h\rangle\leqslant 0\ \ \forall\, i\in I(\overline{x}),\ \langle f'(\overline{x}),\,h\rangle\leqslant 0\}$$

обозначен критический конус задачи (1), (2) в точке $\overline{x}$. Если $h\neq 0$, то неравенство (17) вытекает немедленно из достаточного условия второго порядка оптимальности для задачи (1), (2) в точке $\overline{x}$ и равенства (19), в котором второе слагаемое в правой части всегда неотрицательно в силу неотрицательности $\overline{\mu}$.

Пусть $h=0$; тогда $\tau\neq 0$, причем из (5) и (18) следует, что $\tau_i=0\ \ \forall\, i\in\{1,\ldots,m\}\setminus I(\overline{x})$. Таким образом, среди чисел $\tau_i^2,\ i\in I(\overline{x})$, есть положительные. Отсюда, из условия строгой дополнительности и (19) вновь получаем неравенство (17). □

Для задачи (3), (4) введем семейство штрафных функций

$$\widetilde{\varphi}_c\colon \mathbf{R}^n\times\mathbf{R}^m\to\mathbf{R},\quad \widetilde{\varphi}_c(x,\,\sigma)=f(x)+c\widetilde{\psi}(x,\,\sigma), \tag{20}$$

порождаемое квадратичным штрафом

$$\widetilde{\psi}\colon \mathbf{R}^n\times\mathbf{R}^m\to\mathbf{R},\quad \widetilde{\psi}(x,\,\sigma)=\frac{1}{2}\Big(|F(x)|^2+\sum_{i=1}^{m}(g_i(x)+\sigma_i^2)^2\Big), \tag{21}$$

а также соответствующее семейство задач безусловной оптимизации

$$\widetilde{\varphi}_c(x,\,\sigma)\to\min,\quad (x,\,\sigma)\in\mathbf{R}^n\times\mathbf{R}^m. \tag{22}$$

Заметим, при каждом фиксированном $x\in\mathbf{R}^n$ минимизацию по $\sigma\in\mathbf{R}^m$ в (22) можно произвести явно, тем самым избавившись

от вспомогательных переменных: глобальный минимум достигается в точке $\sigma(x) \in \mathbf{R}^m$, где

$$\sigma_i(x) = \begin{cases} 0, & \text{если } g_i(x) > 0, \\ \sqrt{-g_i(x)}, & \text{если } g_i(x) \leqslant 0, \end{cases} \qquad i = 1, \ldots, m, \tag{23}$$

причем этот минимум единственный с точностью до знаков компонент. Тогда, как нетрудно видеть,

$$\min_{\sigma \in \mathbf{R}^m} \widetilde{\varphi}_c(x, \sigma) = \widetilde{\varphi}_c(x, \sigma(x)) =$$

$$= f(x) + \frac{c}{2}\Big(|F(x)|^2 + \sum_{i=1}^{m} (\max\{0, g_i(x)\})^2\Big) = \varphi_c(x) \quad \forall\, x \in \mathbf{R}^n,$$

где функция φ_c вводится в соответствии с (7), (13).

Задача 4. Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ произвольны, а отображение $G\colon \mathbf{R}^n \to \mathbf{R}^m$ непрерывно в точке $\widetilde{x} \in \mathbf{R}^n$. Доказать, что для любого $c \geqslant 0$ точка $\widetilde{x}$ является (строгим) локальным решением задачи (8) с целевой функцией, задаваемой формулами (7), (13), в том и только том случае, когда точка $(\widetilde{x}, \sigma(\widetilde{x}))$, где $\sigma(\widetilde{x}) \in \mathbf{R}^m$ определяется в соответствии с (23), является (строгим) локальным решением задачи (22) с целевой функцией, задаваемой формулами (20), (21).

Лемма 3. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображения $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ дифференцируемы в точке $\widetilde{x} \in \mathbf{R}^n$.*

Тогда для любого $c \geqslant 0$ точка $\widetilde{x}$ является стационарной в задаче (8) с целевой функцией, задаваемой формулами (7), (13), в том и только том случае, когда точка $(\widetilde{x}, \sigma(\widetilde{x}))$, где $\sigma(\widetilde{x}) \in \mathbf{R}^m$ определяется в соответствии с (23), является стационарной в задаче (22) с целевой функцией, задаваемой формулами (20), (21).

Доказательство. Пусть, например, $\widetilde{x}$ — стационарная точка в задаче (8). Тогда из (23) имеем

$$\frac{\partial \widetilde{\varphi}_c}{\partial x}(\widetilde{x}, \sigma(\widetilde{x})) = f'(\widetilde{x}) + c\Big((F'(\widetilde{x}))^{\mathrm{T}} F(\widetilde{x}) + \sum_{i=1}^{m} (g_i(\widetilde{x}) + (\sigma_i(\widetilde{x}))^2) g_i'(\widetilde{x})\Big) =$$

$$= f'(\widetilde{x}) + c\Big((F'(\widetilde{x}))^{\mathrm{T}} F(\widetilde{x}) + \sum_{i=1}^{m} \max\{0, g_i(\widetilde{x})\} g_i'(\widetilde{x})\Big) =$$

$$= \varphi_c'(\widetilde{x}) = 0, \tag{24}$$

$$\frac{\partial \widetilde{\varphi}_c}{\partial \sigma_i}(\widetilde{x}, \sigma(\widetilde{x})) = 2c(g_i(\widetilde{x}) + (\sigma_i(\widetilde{x}))^2)\sigma_i(\widetilde{x}) = 0 \quad \forall\, i = 1, \ldots, m. \tag{25}$$

Равенства (24) и (25) и означают стационарность точки $(\widetilde{x}, \sigma(\widetilde{x}))$ в задаче (22). Обратное утверждение получается выкладкой, обратной к (24). $\square$

Заметим, что задача (22) может иметь стационарные точки $(\widetilde{x}, \widetilde{\sigma})$ такие, что не все модули компонент $\widetilde{\sigma} \in \mathbf{R}^m$ совпадают с соответствующими компонентами $\sigma(\widetilde{x})$, и при этом $\widetilde{x} \in \mathbf{R}^n$ может не быть стационарной точкой в задаче (8).

Пример 1. Пусть $n = m = 1$, $l = 0$ и

$$f(x) = x, \quad G(x) = x^2 - 1, \quad x \in \mathbf{R}.$$

Критические точки функции $\widetilde{\varphi}_c$ описываются уравнениями

$$\frac{\partial \widetilde{\varphi}_c}{\partial x}(x, \sigma) = 1 + 2c(x^2 + \sigma^2 - 1)x = 0,$$

$$\frac{\partial \widetilde{\varphi}_c}{\partial \sigma}(x, \sigma) = 2c(x^2 + \sigma^2 - 1)\sigma = 0.$$

Легко видеть, что при любом достаточно большом c эта система уравнений имеет решение вида $(\widetilde{x}, 0)$, где $\widetilde{x} \in [-1, 1]$. Но тогда

$$\varphi_c'(\widetilde{x}) = 1 + 2c \max\{0, \widetilde{x}^2 - 1\} \cdot \widetilde{x} = 1,$$

т. е. $\widetilde{x}$ не является критической точкой функции φ_c.

Однако указанное обстоятельство обычно не вызывает серьезных затруднений, поскольку такие «лишние» стационарные точки задачи (22) не могут быть ее локальными решениями.

Лемма 4. *Для любой функции* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и любых отображений* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *и* $G\colon \mathbf{R}^n \to \mathbf{R}^m$, *если для некоторого* $c > 0$ *точка* $(\widetilde{x}, \widetilde{\sigma}) \in \mathbf{R}^n \times \mathbf{R}^m$ *является локальным решением задачи* (22) *с целевой функцией, задаваемой формулами* (20), (21), *то* $\widetilde{\sigma}$ *совпадает с точкой* $\sigma(\widetilde{x})$, *определяемой в соответствии с* (23), *с точностью до знаков компонент.*

Доказательство. Введем функцию

$$\widehat{\varphi}_c\colon \mathbf{R}^m \to \mathbf{R},$$

$$\widehat{\varphi}_c(\sigma) = \widetilde{\varphi}_c(\widetilde{x}, \sigma) = f(\widetilde{x}) + \frac{c}{2}\left(|F(\widetilde{x})|^2 + \sum_{i=1}^{m}(g_i(\widetilde{x}) + \sigma_i^2)^2\right).$$

Эта функция бесконечно дифференцируема на $\mathbf{R}^m$ и имеет в точке $\widetilde{\sigma}$ безусловный минимум. Тогда согласно теореме 1.2.3

$$\frac{\partial \widehat{\varphi}_c}{\partial \sigma_i}(\widetilde{\sigma}) = 2c(g_i(\widetilde{x}) + \widetilde{\sigma}_i^2)\widetilde{\sigma}_i = 0 \quad \forall\, i = 1, \ldots, m. \tag{26}$$

Отсюда следует, что для тех индексов $i = 1, \ldots, m$, для которых $g_i(\widetilde{x}) \geqslant 0$, должно выполняться $\widetilde{\sigma}_i = 0 = \sigma_i(\widetilde{x})$.

Далее, в силу теоремы 1.2.4 матрица $\widehat{\varphi}_c''(\widetilde{\sigma})$ неотрицательно определена. Если предположить, что для некоторого $i = 1, \ldots, m$ такого,

что $g_i(\widetilde{x}) < 0$, выполнено $\widetilde{\sigma}_i^2 \neq (\sigma_i(\widetilde{x}))^2 = -g_i(\widetilde{x})$, то из (26) следует, что $\widetilde{\sigma}_i = 0$. Но тогда

$$\frac{\partial^2 \widehat{\varphi}_c}{\partial \sigma_i^2}(\widetilde{\sigma}) = 2c(g_i(\widetilde{x}) + 3\widetilde{\sigma}_i^2) = 2cg_i(\widetilde{x}) < 0,$$

что противоречит неотрицательной определенности диагональной матрицы $\widehat{\varphi}_c''(\widetilde{\sigma})$. $\square$

Таким образом, в отличие от стационарных точек, локальные решения задач (8) и (22) находятся во взаимно однозначном соответствии в том же смысле, в каком и локальные решения задач (1), (2) и (3), (4) (см. начало параграфа).

Теорема 2. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображения* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *и* $G\colon \mathbf{R}^n \to \mathbf{R}^m$ *непрерывно дифференцируемы на* $\mathbf{R}^n$.

Тогда если в алгоритме 1 *используется функция* $\psi\colon \mathbf{R}^n \to \mathbf{R}$, *введенная в* (13), *и последовательность* $\{x^k\}$ *сгенерирована этим алгоритмом, то любая ее предельная точка* $\overline{x} \in \mathbf{R}^n$, *в которой выполнено условие линейной независимости, является стационарной точкой задачи* (1), (2). *Более того, для любой сходящейся к* $\overline{x}$ *подпоследовательности* $\{x^{k_j}\}$ *справедливо*

$$\{c_{k_j} F(x^{k_j})\} \to \overline{\lambda} \quad (j \to \infty), \tag{27}$$

$$c_{k_j} \max\{0,\, g_i(x^{k_j})\} \to \overline{\mu}_i \quad (j \to \infty) \quad \forall\, i = 1, \ldots, m, \tag{28}$$

где $\overline{\lambda} \in \mathbf{R}^l$ *и* $\overline{\mu} \in \mathbf{R}^m_+$ *— (единственные) отвечающие* $\overline{x}$ *множители Лагранжа.*

Из (28) вытекает следующее интересное наблюдение: для любого достаточно большого j и для любого $i \in I(\overline{x})$ такого, что $\overline{\mu}_i > 0$, имеет место неравенство $g_i(x^{k_j}) > 0$.

Доказательство. Рассмотрим такую последовательность $\{(x^k,\, \sigma(x^k))\}$, где для каждого k точка $\sigma(x^k) \in \mathbf{R}^m$ определяется в соответствии с (23). Согласно лемме 3 последовательность $\{(x^k,\, \sigma(x^k))\}$ может быть сгенерирована методом квадратичного штрафа для задачи (3), (4). Из непрерывности отображения $\sigma(\cdot)$ вытекает, что $(\overline{x},\, \sigma(\overline{x}))$ — предельная точка этой последовательности, причем из результата задачи 3 следует регулярность ограничений задачи (3), (4) в этой точке.

Применяя теорему 4.3.2, получаем, что $(\overline{x},\, \sigma(\overline{x}))$ — стационарная точка задачи (3), (4), причем для сходящейся к $\overline{x}$ подпоследовательности $\{x^{k_j}\}$ справедливы предельные соотношения (27) и

$$c_{k_j}(g_i(x^{k_j}) + (\sigma_i(x^{k_j}))^2) \to \overline{\mu}_i \quad \forall\, i = 1, \ldots, m \quad (j \to \infty), \tag{29}$$

где $(\overline{\lambda},\, \overline{\mu}) \in \mathbf{R}^l \times \mathbf{R}^m$ — соответствующий множитель Лагранжа. Из равенства (23) следует, что (29) в точности совпадает с (28), а из (28),

в частности, вытекает, что $\overline{\mu} \geqslant 0$. Остается воспользоваться утверждением б) леммы 1. $\square$

Таким образом, любая предельная точка последовательности, генерируемой методом квадратичного штрафа, либо является стационарной в задаче (1), (2), либо в ней нарушается условие линейной независимости. В последнем случае предельная точка может быть даже недопустимой в задаче (1), (2).

Пример 2. Пусть $n = m = 1$, $l = 0$ и

$$f(x) = x, \quad G(x) = x^2 + 1, \quad x \in \mathbf{R}.$$

Тогда

$$\varphi_c(x) = x + \frac{c}{2}\,(\max\{0,\, x^2 + 1\})^2 = x + \frac{c}{2}\,(x^2 + 1)^2, \quad x \in \mathbf{R},$$

и стационарные точки задачи (8) описываются уравнением

$$\varphi_c'(x) = 1 + 2c(x^2 + 1)x = 0.$$

Для любого c это уравнение имеет единственное решение, которое стремится к нулю при $c \to \infty$. Но точка 0 недопустима в задаче (1), (2). Заметим, что $G'(0) = 0$, т. е. условие линейной независимости в точке 0 нарушено.

В приведенном примере исходная задача имеет пустое допустимое множество, однако недопустимость предельной точки последовательности метода квадратичного штрафа может иметь место и в случае непустого допустимого множества (разумеется, при нарушении в этой точке условия линейной независимости).

Пример 3. Пусть $n = 2$, $l = m = 1$ и

$$F(x) = x_1^2 - 1, \quad G(x) = -x_1 + x_2^2, \quad x \in \mathbf{R}^2,$$

а функция f постоянна на $\mathbf{R}^2$. Тогда точка 0 является стационарной в задаче (8) при любом c, будучи при этом недопустимой в задаче (1), (2). Здесь условие линейной независимости в точке 0 нарушено, поскольку $F'(0) = 0$.

Теорема 3. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображения* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *и* $G\colon \mathbf{R}^n \to \mathbf{R}^m$ *дважды дифференцируемы в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$, *причем их вторые производные непрерывны в этой точке. Пусть в точке* $\overline{x}$ *выполнено условие линейной независимости, причем* $\overline{x}$ *— стационарная точка задачи* (1), (2), *а* $\overline{\lambda} \in \mathbf{R}^l$ *и* $\overline{\mu} \in \mathbf{R}^m_+$ *— однозначно отвечающие ей множители Лагранжа. Пусть, наконец, в точке* $\overline{x}$ *выполнено сформулированное в теореме* 1.4.5 *достаточное условие второго порядка оптимальности, а также условие строгой дополнительности.*

Тогда существуют окрестность U *точки* $\overline{x}$ *и числа* $\overline{c} \geqslant 0$ *и* $M > 0$ *такие, что для каждого* $c > \overline{c}$ *задача* (8) *с целевой функцией,*

задаваемой формулами (7), (13), *имеет в* U *единственную стационарную точку* $x(c)$, *причем*

$$|x(c)-\overline{x}| \leqslant M\,\frac{|\overline{\lambda}|+|\overline{\mu}|}{c}\,, \quad |cF(x(c))-\overline{\lambda}| \leqslant M\,\frac{|\overline{\lambda}|+|\overline{\mu}|}{c}\,,$$

$$|c\max\{0,\, g_i(x(c))\}-\overline{\mu}_i| \leqslant M\,\frac{|\overline{\lambda}|+|\overline{\mu}|}{c} \quad \forall\, i=1,\ldots,m. \tag{30}$$

Задача 5. Используя леммы 2–4, теорему 4.3.4 и утверждение б) из задачи 4.3.4, доказать теорему 3.

Как следует из условия строгой дополнительности и оценки (30), при достаточно больших значениях параметра штрафа все активные в точке $\overline{x}$ ограничения-неравенства задачи (1), (2) нарушаются в точках определенной в теореме 3 траектории $x(\cdot)$: $g_i(x(c))>0$ $\forall\, i \in I(\overline{x})$ для любого достаточно большого c (ср. с замечанием, сопровождавшим формулировку теоремы 2).

Задача 6. Решить задачу

$$x^2-4x \to \min, \quad x \in D=\{x \in \mathbf{R}\big|\ x \leqslant 1\},$$

методом квадратичного штрафа.

В специальной литературе можно встретить множество различных штрафов, помимо степенных; более того, сама концепция семейства штрафных функций допускает определенные обобщения (по этим вопросам см. [6, 10, 15, 34, 37, 38]). Следует, однако, отметить, что использование штрафных функций, отличных от рассматривавшихся выше, если и бывает оправдано, то лишь в весьма специальных случаях (при наличии у задачи свойств выпуклости, при отсутствии ограничений-равенств и т. п.).

На практике иногда бывает полезно снимать за счет штрафования не все ограничения задачи (1), (2), а только часть, учитывая «простые» ограничения непосредственно. Эти простые ограничения удобно с самого начала трактовать как прямое ограничение, рассматривая допустимое множество вида

$$D=\{x \in P\big|\ F(x)=0,\, G(x) \leqslant 0\}, \tag{31}$$

где $P \subset \mathbf{R}^n$ — заданное множество «простой» структуры.

Штрафом для заданного в (31) множества D называется любая функция $\psi\colon P \to \mathbf{R}$, удовлетворяющая условиям $\psi(x)=0 \quad \forall\, x \in D$, $\psi(x)>0 \quad \forall\, x \in P \setminus D$.

Соответствующее семейство *штрафных функций* имеет вид

$$\varphi_c\colon P \to \mathbf{R}, \quad \varphi_c(x)=f(x)+c\psi(x),$$

а вспомогательную задачу безусловной оптимизации (8) следует заменить задачей

$$\varphi_c(x) \to \min, \quad x \in P,$$

к которой применимы методы оптимизации при простых ограничениях, рассмотренные в § 4.1.

4.7.2. Модифицированные функции Лагранжа. Для каждого $c > > 0$ определим для задачи (3), (4) модифицированную функцию Лагранжа (см. п. 4.3.3):

$$\widetilde{L}_c\colon (\mathbf{R}^n \times \mathbf{R}^m) \times (\mathbf{R}^l \times \mathbf{R}^m) \to \mathbf{R},$$

$$\widetilde{L}_c((x,\, \sigma),\, (\lambda,\, \mu)) = \\ = \widetilde{L}((x,\, \sigma),\, (\lambda,\, \mu)) + \frac{c}{2}\Big(|F(x)|^2 + \sum_{i=1}^{m}(g_i(x) + \sigma_i^2)^2\Big). \quad (32)$$

При заданных $\lambda \in \mathbf{R}^l$ и $\mu \in \mathbf{R}^m$ в задаче

$$\widetilde{L}_c((x,\, \sigma),\, (\lambda,\, \mu)) \to \min, \quad (x,\, \sigma) \in \mathbf{R}^n \times \mathbf{R}^m, \quad (33)$$

минимизация по $\sigma \in \mathbf{R}^m$ может быть выполнена явно. А именно, при каждом фиксированном $x \in \mathbf{R}^n$ глобальный минимум достигается в точке $\sigma(x,\, \mu,\, c) \in \mathbf{R}^m$, где

$$\sigma_i(x,\, \mu,\, c) = \begin{cases} 0, & \text{если } g_i(x) + \dfrac{\mu_i}{c} > 0, \\ \sqrt{-\Big(g_i(x) + \dfrac{\mu_i}{c}\Big)}, & \text{если } g_i(x) + \dfrac{\mu_i}{c} \leqslant 0, \end{cases} \quad (34)$$

$$i = 1, \ldots, m,$$

причем этот минимум единственный с точностью до знаков компонент. Заметим, что

$$g_i(x) + (\sigma_i(x,\, \mu,\, c))^2 = \\ = g_i(x) + \max\Big\{0,\, -g_i(x) - \frac{\mu_i}{c}\Big\} = \max\Big\{g_i(x),\, -\frac{\mu_i}{c}\Big\}, \quad (35)$$

откуда после несложных преобразований получаем окончательный вид *модифицированной функции Лагранжа* задачи (1), (2):

$$L_c\colon \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R},$$

$$L_c(x,\, \lambda,\, \mu) = \min_{\sigma \in \mathbf{R}^m} \widetilde{L}_c((x,\, \sigma),\, (\lambda,\, \mu)) =$$

$$= \widetilde{L}_c((x,\, \sigma(x,\, \mu,\, c)),\, (\lambda,\, \mu)) =$$

$$= f(x) + \langle \lambda,\, F(x)\rangle + \frac{c}{2}\,|F(x)|^2 + \\ + \frac{1}{2c}\sum_{i=1}^{m}((\max\{0,\, cg_i(x) + \mu_i\})^2 - \mu_i^2). \qquad (36)$$

Заметим, что эта функция дифференцируема, но, вообще говоря, только один раз.

Вместо (33) будем рассматривать задачу

$$L_c(x,\, \lambda,\, \mu) \to \min, \quad x \in \mathbf{R}^n. \qquad (37)$$

Задача 7. Убедиться, что если функция $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображения $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ дифференцируемы в точке $\overline{x} \in \mathbf{R}^n$, то для любых $\overline{\lambda} \in \mathbf{R}^l$ и $\overline{\mu} \in \mathbf{R}^m$ и любого $c > 0$ тройка $(\overline{x},\, \overline{\lambda},\, \overline{\mu})$ является решением системы Каруша–Куна–Таккера задачи (1), (2) в том и только том случае, когда

$$L_c'(\overline{x},\, \overline{\lambda},\, \overline{\mu}) = 0.$$

Следующие две леммы являются аналогами лемм 3 и 4 соответственно.

Лемма 5. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображения $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ дифференцируемы в точке $\widetilde{x} \in \mathbf{R}^n$.*

Тогда для любого $c > 0$ и любых $\lambda \in \mathbf{R}^l$ и $\mu \in \mathbf{R}^m$ точка $\widetilde{x}$ является стационарной в задаче (37) *с целевой функцией, задаваемой формулой* (36), *в том и только том случае, когда точка $(\widetilde{x}, \sigma(\widetilde{x}))$, где $\sigma(\widetilde{x}) \in \mathbf{R}^m$ определяется в соответствии с* (34), *является стационарной в задаче* (33) *с целевой функцией, задаваемой формулой* (32).

Задача 8. Доказать лемму 5.

Лемма 6. *Для любой функции $f\colon \mathbf{R}^n \to \mathbf{R}$ и любых отображений $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$, если для некоторого $c > 0$ и некоторых $\lambda \in \mathbf{R}^l$ и $\mu \in \mathbf{R}^m$ точка $(\widetilde{x}, \widetilde{\sigma}) \in \mathbf{R}^n \times \mathbf{R}^m$ является локальным решением задачи* (33) *с целевой функцией, задаваемой формулой* (32), *то $\widetilde{\sigma}$ совпадает с точкой $\sigma(\widetilde{x})$, определяемой в соответствии с* (34), *с точностью до знаков компонент.*

Задача 9. Доказать лемму 6.

Опишем *метод модифицированных функций Лагранжа* для задачи (1), (2).

Алгоритм 2. Выбираем монотонно неубывающую последовательность $\{c_k\} \subset \mathbf{R}_+ \setminus \{0\}$ и точки $\lambda^0 \in \mathbf{R}^l$ и $\mu^0 \in \mathbf{R}^m$, и полагаем $k = 0$.

1. Вычисляем $x^k \in \mathbf{R}^n$ как стационарную точку задачи (37) с целевой функцией, задаваемой формулой (36) при $c = c_k$, $\lambda = \lambda^k$ и $\mu = \mu^k$.

2. Полагаем

$$\lambda^{k+1} = \lambda^k + c_k F(x^k),$$

$$\mu_i^{k+1} = \max\{0,\, c_k g_i(x^k) + \mu_i^k\}, \quad i = 1, \ldots, m. \tag{38}$$

3. Увеличиваем номер шага k на 1 и переходим к п. 1.

Поясним происхождение формулы (38), предназначенной для пересчета приближений к множителю Лагранжа, отвечающему ограничениям-неравенствам. Согласно схеме метода модифицированных функций Лагранжа для задачи (3), (4) (см. алгоритм 4.3.3) естественно предложить следующий вид такой формулы:

$$\mu_i^{k+1} = \mu_i^k + c_k(g_i(x^k) + (\sigma(x^k,\, \mu^k,\, c_k))^2), \quad i = 1, \ldots, m,$$

а это и есть (38), если учесть (35).

Обобщением теоремы 4.3.5 на случай наличия ограничений-неравенств служит

Теорема 4. *Пусть выполнены условия теоремы* 3.

Тогда существуют окрестность U *точки* $\overline{x}$ *и числа* $\overline{c} \geqslant 0$, $\delta > 0$ *и* $M > 0$ *такие, что*:

а) *для любой тройки* $(\lambda,\, \mu,\, c) \in \Delta(\overline{c},\, \delta)$, *где*

$$\Delta(\overline{c},\, \delta) = \{(\lambda,\, \mu,\, c) \in \mathbf{R}^l \times \mathbf{R}^m \times \mathbf{R} \mid\ |\lambda - \overline{\lambda}| < \delta c,\ |\mu - \overline{\mu}| < \delta c,\ c > \overline{c}\},$$

задача (37) *с целевой функцией, задаваемой формулой* (36), *имеет в* U *единственную стационарную точку* $x(\lambda,\, \mu,\, c)$, *причем*

$$|x(\lambda,\, \mu,\, c) - \overline{x}| \leqslant M\, \frac{|\lambda - \overline{\lambda}| + |\mu - \overline{\mu}|}{c},$$

$$|\lambda + cF(x(\lambda,\, \mu,\, c)) - \overline{\lambda}| \leqslant M\, \frac{|\lambda - \overline{\lambda}| + |\mu - \overline{\mu}|}{c},$$

$$|\max\{0,\, cg_i(x(\lambda,\, \mu,\, c)) + \mu_i\} - \overline{\mu}_i| \leqslant$$

$$\leqslant M\, \frac{|\lambda - \overline{\lambda}| + |\mu - \overline{\mu}|}{c} \quad \forall\, i = 1, \ldots, m;$$

б) *существует число* $\overline{\delta} \in (0,\, \delta]$ *такое, что для любой монотонно неубывающей последовательности* $\{c_k\} \subset \mathbf{R}_+ \setminus \{0\}$ *и любых точек* $\lambda^0 \in \mathbf{R}^l$ *и* $\mu^0 \in \mathbf{R}^m$, *для которых выполнено*

$$(\lambda^0,\, \mu^0,\, c_0) \in \Delta(\overline{c},\, \overline{\delta}),$$

формулы

$$\lambda^{k+1} = \lambda^k + c_k F(x(\lambda^k,\, \mu^k,\, c_k)),$$

$$\mu_i^{k+1} = \max\{0,\, c_k g_i(x(\lambda^k,\, \mu^k,\, c_k)) + \mu_i^k\}, \quad i = 1, \ldots, m, \tag{39}$$

где $k = 0, 1, \ldots$, *корректно определяют последовательности* $\{\lambda^k\}$ *и* $\{\mu^k\}$ (*в том смысле, что* $\{(\lambda^k,\, \mu^k,\, c_k)\} \subset \Delta(\overline{c},\, \delta)$, *т.е. для всякого* k *точка* $x(\lambda^k,\, \mu^k,\, c_k)$ *корректно определена*), *причем* $\{\lambda^k\} \to \overline{\lambda}$, $\{\mu^k\} \to \overline{\mu}$, $\{x(\lambda^k,\, \mu^k,\, c_k)\} \to \overline{x}$ $(k \to \infty)$. *Скорость сходимости последовательностей* $\{\lambda^k\}$ *и* $\{\mu^k\}$ *линейная, а если* $\{c_k\} \to \infty$ $(k \to \infty)$, *то сверхлинейная.*

Задача 10. Используя леммы 2, 5 и 6, теорему 4.3.5 и результат задачи 4.3.10, доказать теорему 4.

В дополнение к утверждению б) теоремы 4 можно отметить следующий факт. Поскольку $\{\mu^k\} \to \overline{\mu}$, $\{x(\lambda^k,\, \mu^k,\, c_k)\} \to \overline{x}$ $(k \to \infty)$, то для любого индекса $i \in \{1, \ldots, m\} \setminus I(\overline{x})$ и любого достаточно большого k имеет место неравенство

$$c_k g_i(x(\lambda^k,\, \mu^k,\, c_k)) + \mu_i^k < 0.$$

Поэтому из (39) следует, что $\mu_i^{k+1} = 0$, т.е. приближения к множителям, отвечающим неактивным ограничениям-неравенствам, обнуляются после конечного числа итераций.

4.7.3. Точные штрафы. Одним из важнейших средств современной оптимизации являются точные штрафы, использование которых не предполагает бесконечного увеличения параметра штрафа. Начнем со следующей теоремы об оценке скорости сходимости по функции методов штрафов в предположении о том, что сходимость по аргументу имеет место. При этом специфика ограничений, задающих множество D, а также конкретный вид используемого штрафа не будут играть роли, как это было в теореме 1.

А именно, предположим, что функция $\psi\colon \mathbf{R}^n \to \mathbf{R}_+$ локально оценивает расстояние до допустимого множества D в следующем смысле: существуют окрестность U искомого решения $\overline{x}$ задачи (1) и числа $M > 0$ и $\nu > 0$ такие, что

$$\psi(x) = 0 \quad \forall\, x \in D \cap U \tag{40}$$

и

$$\operatorname{dist}(x,\, D) \leqslant M(\psi(x))^\nu \quad \forall\, x \in U \tag{41}$$

(непрерывность ψ не предполагается). Очевидно, такая функция ψ локально удовлетворяет (6) и может использоваться в качестве штрафа для множества D, по крайней мере вблизи точки $\overline{x}$.

Теорема 5. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *непрерывна по Липшицу с константой* $N > 0$ *в некоторой окрестности* U *точки* $\overline{x} \in \mathbf{R}^n$, *а множество* $D \subset \mathbf{R}^n$ *замкнуто. Пусть* $\overline{x}$ *— локальное решение задачи* (1). *Пусть, наконец, для функции* $\psi\colon \mathbf{R}^n \to \mathbf{R}_+$ *выполнены условия* (40) *и* (41) *при некоторых* $M > 0$ *и* $\nu > 0$.

Тогда если для последовательностей $\{c_k\} \subset \mathbf{R}$ *и* $\{x^k\} \subset \mathbf{R}^n$ *имеют место соотношения*

$$\varphi_{c_k}(x^k) \leqslant \varphi_{c_k}(\overline{x}) = f(\overline{x}) \quad \forall\, k \tag{42}$$

(*где функция* φ_{c_k} *задана формулой* (7)) *и*

$$c_k \to +\infty, \quad \{x^k\} \to \overline{x} \quad (k \to \infty), \tag{43}$$

то для любого достаточно большого k *справедливо следующее*:

а) *если* $\nu < 1$, *то*

$$0 \leqslant f(\overline{x}) - f(x^k) \leqslant \left(\frac{(MN)^{1/\nu}}{c_k} \right)^{\nu/(1-\nu)}, \tag{44}$$

$$\operatorname{dist}(x^k, D) \leqslant M \left(\frac{MN}{c_k} \right)^{\nu/(1-\nu)}; \tag{45}$$

б) *если* $\nu \geqslant 1$, *то*

$$f(x^k) = f(\overline{x}), \quad x^k \in D;$$

в частности, если $\overline{x}$ *является строгим локальным решением задачи* (1), (2), *то* $x^k = \overline{x}$.

Доказательство. В силу (40)–(42) и второго соотношения в (43) для любого достаточно большого k имеем

$$f(\overline{x}) = f(\overline{x}) + c_k\psi(\overline{x}) = \varphi_{c_k}(\overline{x}) \geqslant \varphi_{c_k}(x^k) =$$

$$= f(x^k) + c_k\psi(x^k) \geqslant f(x^k) + c_k \left(\frac{\operatorname{dist}(x^k, D)}{M} \right)^{1/\nu}. \tag{46}$$

Для каждого достаточно большого k обозначим через $\widetilde{x}^k$ произвольную проекцию точки x^k на D (существование проекции вытекает из замкнутости D и следствия 1.1.2). Из второго соотношения в (43) и того, что $\overline{x} \in D$, следует, что

$$\{\widetilde{x}^k\} \to \overline{x} \quad (k \to \infty).$$

Но $\overline{x}$ является локальным решением задачи (1), поэтому для любого достаточно большого k

$$f(\overline{x}) - f(x^k) \leqslant f(\widetilde{x}^k) - f(x^k) \leqslant N \operatorname{dist}(x^k, D). \tag{47}$$

Объединяя (46) и (47), приходим к неравенству

$$f(\overline{x}) - f(x^k) \geqslant c_k \left(\frac{f(\overline{x}) - f(x^k)}{MN} \right)^{1/\nu}. \tag{48}$$

Введем множество

$$\mathcal{K} = \{k = 0, 1, \ldots \mid f(x^k) \neq f(\overline{x})\}.$$

Соотношения (46) и (48) влекут неравенства

$$0 \leqslant (f(\overline{x}) - f(x^k))^{(1-\nu)/\nu} \leqslant \frac{(MN)^{1/\nu}}{c_k} \tag{49}$$

для любого достаточно большого $k \in \mathcal{K}$. Утверждение а) выполняется тривиальным образом при $k \notin \mathcal{K}$ и следует немедленно из (49) при $k \in \mathcal{K}$ ((45) следует из (44) и (46)). Если же $\nu \geqslant 1$, то согласно (43) средняя часть (49) не стремится к нулю при $k \to \infty$, в то время как правая стремится. Это означает, что множество $\mathcal{K}$ по необходимости ограничено, т. е. $f(\overline{x}) = f(x^k)$ для любого достаточно большого k. Кроме того, для таких k в силу (46) выполнено $x^k \in D$, что завершает доказательство утверждения б). $\square$

В доказанной теореме вместо (40) на самом деле достаточно предполагать, что $\psi(\overline{x}) = 0$.

Свойство (42) всегда имеет место, если, например, для каждого k точка x^k является глобальным минимумом функции φ_{c_k} на некотором множестве, содержащем $\overline{x}$. Выполнение подобного свойства вполне типично для любого метода штрафов. Второе предельное соотношение в (43) выражает предположение о сходимости рассматриваемого метода по аргументу. Оба указанных свойства будут выполнены, если $\overline{x}$ — строгое локальное решение задачи (1), а точка $x^k = x(c_k)$ определяется, скажем, в соответствии с тем, как это делалось в доказательстве теоремы 4.3.3 (см. также задачу 4.3.3 и теорему 1).

Следствие 1. *Пусть в условиях теоремы* 5 *точка* $\overline{x}$ *— строгое локальное решение задачи* (1) *и пусть* $\nu \geqslant 1$.

Тогда существует число $\overline{c} \geqslant 0$ *такое, что для каждого* $c > \overline{c}$ *точка* $\overline{x}$ *является строгим локальным решением задачи* (8) *с целевой функцией, задаваемой формулой* (7).

Доказательство. Предположим противное. Тогда существуют последовательности $\{c_i\} \subset \mathbf{R}$ и $\{x^i\} \subset \mathbf{R}^n$ такие, что

$$\varphi_{c_i}(x^i) \leqslant \varphi_{c_i}(\overline{x}), \quad x^i \neq \overline{x} \quad \forall\, i, \tag{50}$$

$$c_i \to +\infty, \quad \{x^i\} \to \overline{x} \quad (i \to \infty). \tag{51}$$

Но одновременное выполнение (50) и (51) противоречит утверждению б) теоремы 5. $\square$

Один из важнейших вопросов, возникающих в связи с полученными результатами, состоит в отыскании условий, при которых можно указать функцию ψ, для которой справедлива локальная оценка вида (41) расстояния до заданного в (2) множества D. Сколько-нибудь полное обсуждение этого вопроса выходит за рамки настоящего курса. Ограничимся тем, что приведем без доказательства следующую

теорему Робинсона, обобщающую теорему 1.3.3 (доказательство см., например, в [20, 26]).

Теорема 6. *Пусть отображения* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *и* $G\colon \mathbf{R}^n \to \mathbf{R}^m$ *дифференцируемы в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$, *причем их производные непрерывны в точке* $\overline{x}$. *Пусть* $\overline{x} \in D$, *где множество* D *введено в* (2), *и в точке* $\overline{x}$ *выполнено условие регулярности Мангасариана–Фромовица.*

Тогда найдутся окрестность U *точки* $\overline{x}$ *и число* $M > 0$ *такие, что*

$$\operatorname{dist}(x,\, D) \leqslant M|\Psi(x)| \quad \forall\, x \in U, \tag{52}$$

где отображение $\Psi\colon \mathbf{R}^n \to \mathbf{R}^l \times \mathbf{R}^m$ *введено в* (9).

Таким образом, в условиях теоремы 6 функция ψ, определяемая при некотором $p > 0$ соотношением (10), удовлетворяет (41) при $\nu = 1/p$. Разумеется, вместо $|\cdot|$ здесь может использоваться любая другая норма в $\mathbf{R}^l \times \mathbf{R}^m$. Если использовать $|\cdot|_p$ либо $|\cdot|_\infty$, то приходим к степенным штрафам, введенным в п. 4.7.1 (см. (11) и (12)). Если $p = 2$ (квадратичный штраф), то оценки (44), (45) принимают вид

$$0 \leqslant f(\overline{x}) - f(x^k) \leqslant \frac{(NM)^2}{c_k}, \quad \operatorname{dist}(x^k,\, D) \leqslant \frac{M^2 N}{c_k}$$

(ср. с теоремой 4). Согласно теореме 5 уменьшение показателя степени p приводит к увеличению скорости сходимости соответствующих методов штрафов. Более того, согласно следствию 1 при достаточно малом p (а именно, при $p \leqslant 1$) степенной штраф является *точным* в том смысле, что искомое (строгое) локальное решение $\overline{x}$ задачи (1), (2) является локальным решением вспомогательной задачи безусловной оптимизации (8) при любом достаточно большом c. Иными словами, если при реализации соответствующего метода степенного штрафа обеспечивается отыскание «удачных» локальных решений задач вида (8), то увеличивать значение параметра штрафа до бесконечности нет необходимости.

Разумеется, установленная «точность» штрафа не является точностью в полном смысле. Очень желательно доказать еще и следующее утверждение: в естественных условиях в некоторой окрестности точки $\overline{x}$ задача (8) не имеет других локальных решений, кроме $\overline{x}$ (иначе «удачный» выбор такой стационарной точки будет задачей трудно выполнимой). Этот важный вопрос здесь тоже подробно не обсуждается. В [6] для степенного штрафа, порождаемого нормой $|\cdot|_\infty$ при $p = 1$, показано, что сформулированное утверждение будет справедливо при выполнении в точке $\overline{x}$ условия линейной независимости.

Заметим, что уменьшение p негативно влияет на гладкость штрафной функции φ_c, что приводит к сужению арсенала методов, пригодных для решения задачи (8). В частности, при $p \leqslant 1$ вводимый формулой (10) степенной штраф, вообще говоря, не является гладким, какой бы высокой гладкостью не обладали F и G. В частности, само понятие стационарной точки задачи (8) при этом должно уточняться. Другая трудность, с которой приходится сталкиваться при непосредственном использовании точных штрафов, состоит в том, что обычно заранее не известно, насколько большим следует брать c (об эвристических приемах такого выбора см. [6]).

Некоторые методы решения негладких задач безусловной оптимизации обсуждаются в гл. 6. Кроме того, необходимо отметить, что на самом деле точные штрафы обычно используются не непосредственно, а в сочетании с другими методами, например, в целях глобализации сходимости последних (см. § 5.4, где также приводятся некоторые дополнительные сведения о негладких точных штрафах, и [6]). Поэтому, в частности, предлагаемый здесь материал по точным штрафам не содержит никаких строго сформулированных алгоритмов.

В завершение разговора о негладких точных штрафах заметим, что в более слабых предположениях, чем условие регулярности Мангасариана–Фромовица, можно рассчитывать на оценку, более слабую, чем (52). Во всяком случае при отсутствии ограничений-неравенств оценку вида

$$\operatorname{dist}(x,\, D) \leqslant M|F(x)|^{1/q} \quad \forall\, x \in U \tag{53}$$

при $q > 1$ удается получить в более слабых предположениях, чем условие регулярности $\operatorname{rank} F'(\overline{x}) = l$ (см. [2, 21, 22]).

Задача 11. Пусть отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ дифференцируемо в каждой точке множества

$$D = \{x \in \mathbf{R}^n \mid F(x) = 0\},$$

достаточно близкой к точке $\overline{x} \in D$. Доказать, что оценка (53) может иметь место для некоторой окрестности U точки $\overline{x}$ и некоторого $M > 0$ лишь при $q \geqslant 1$.

Весьма тонкие результаты о скорости сходимости методов штрафов (в том числе и по аргументу) можно получить с помощью средств теории чувствительности; см. [20]. По сравнению с приведенными выше, это результаты совершенно иной природы: они используют достаточные условия оптимальности, но зато не предполагают выполнение каких-либо условий регулярности ограничений или оценок расстояния до допустимого множества.

Коротко остановимся на *точных гладких штрафных функциях* для задачи (1), (2). Сначала введем семейство таких функций для задачи (3), (4), например,

$$\widetilde{\varphi}_{c_1,\,c_2}\colon (\mathbf{R}^n \times \mathbf{R}^m) \times (\mathbf{R}^l \times \mathbf{R}^m) \to \mathbf{R},$$

$$\begin{aligned}
\widetilde{\varphi}_{c_1,\,c_2}((x,\,\sigma),\,(\lambda,\,\mu)) &= \\
&= \widetilde{L}_{c_1}((x,\,\sigma),\,(\lambda,\,\mu)) + \frac{c_2}{2}\left|\frac{\partial \widetilde{L}}{\partial (x,\,\sigma)}\,((x,\,\sigma),\,(\lambda,\,\mu))\right|^2 = \\
&= L(x,\,\lambda,\,\mu) + \sum_{i=1}^{m}\mu_i\sigma_i^2 + \frac{c_1}{2}\Big(|F(x)|^2 + \sum_{i=1}^{m}(g_i(x)+\sigma_i^2)^2\Big) + \\
&\quad + \frac{c_2}{2}\Big(\Big|\frac{\partial L}{\partial x}(x,\,\lambda,\,\mu)\Big|^2 + 4\sum_{i=1}^{m}(\mu_i\sigma_i)^2\Big)
\end{aligned} \tag{54}$$

(см. п. 4.3.3), и для $c_1,\,c_2 > 0$ будем рассматривать задачу

$$\begin{gathered}
\widetilde{\varphi}_{c_1,\,c_2}((x,\,\sigma),\,(\lambda,\,\mu)) \to \min, \\
((x,\,\sigma),\,(\lambda,\,\mu)) \in (\mathbf{R}^n \times \mathbf{R}^m) \times (\mathbf{R}^l \times \mathbf{R}^m).
\end{gathered} \tag{55}$$

Несложно убедиться, что при любых фиксированных $x \in \mathbf{R}^n$, $\lambda \in \mathbf{R}^l$ и $\mu \in \mathbf{R}^m$ глобальный минимум по $\sigma \in \mathbf{R}^m$ в (55) достигается в единственной с точностью до знаков компонент точке $\sigma(x,\,\mu,\,c_1,\,c_2)$, где

$$\begin{aligned}
&\sigma_i(x,\,\mu,\,c_1,\,c_2) = \\
&= \begin{cases}
0, & \text{если } g_i(x) + \dfrac{\mu_i(1+2c_2\mu_i)}{c_1} > 0, \\
\sqrt{-\Big(g_i(x) + \dfrac{\mu_i(1+2c_2\mu_i)}{c_1}\Big)}, & \text{если } g_i(x) + \dfrac{\mu_i(1+2c_2\mu_i)}{c_1} \leqslant 0,
\end{cases} \\
&\qquad i = 1,\,\ldots,\,m.
\end{aligned} \tag{56}$$

После несложных преобразований приходим к следующему виду точной гладкой штрафной функции для задачи (1), (2):

$$\varphi_{c_1,\,c_2}\colon \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R},$$

$$\begin{aligned}
\varphi_{c_1,\,c_2}(x,\,\lambda,\,\mu) &= \min_{\sigma \in \mathbf{R}^m} \widetilde{\varphi}_{c_1,\,c_2}((x,\,\sigma),\,(\lambda,\,\mu)) = \\
&= \widetilde{\varphi}_{c_1,\,c_2}((x,\,\sigma(x,\,\mu,\,c_1,\,c_2)),\,(\lambda,\,\mu)) = \\
&= f(x) + \langle \lambda,\,F(x)\rangle + \frac{c_1}{2}\,|F(x)|^2 + \frac{c_2}{2}\left|\frac{\partial L}{\partial x}(x,\,\lambda,\,\mu)\right|^2 + \\
&\quad + \frac{1}{2c_1}\sum_{i=1}^{m}((\max\{0,\,c_1 g_i(x) + \mu_i(1+2c_2\mu_i)\})^2 - \\
&\quad - \mu_i^2(1+2c_2\mu_i)^2 - 4c_1c_2\mu_i^2 g_i(x)).
\end{aligned} \tag{57}$$

Будем рассматривать задачу

$$\varphi_{c_1,\,c_2}(x,\,\lambda,\,\mu)\to\min,\quad (x,\,\lambda,\,\mu)\in\mathbf{R}^n\times\mathbf{R}^l\times\mathbf{R}^m. \tag{58}$$

Задача 12. Пусть функция $f\colon\mathbf{R}^n\to\mathbf{R}$ и отображения $F\colon\mathbf{R}^n\to\mathbf{R}^l$ и $G\colon\mathbf{R}^n\to\mathbf{R}^m$ дважды дифференцируемы в точке $\overline{x}\in\mathbf{R}^n$. Показать, что если точка $\overline{x}$ является стационарной в задаче (1), (2) и ей отвечают множители Лагранжа $\overline{\lambda}\in\mathbf{R}^l$ и $\overline{\mu}\in\mathbf{R}^m_+$, то для любых $c_1>0$ и $c_2\geqslant 0$ точка $(\overline{x},\,\overline{\lambda},\,\overline{\mu})$ является стационарной в задаче (58) с целевой функцией, задаваемой формулой (57).

Аналогом леммы 3 является

Лемма 7. *Пусть функция $f\colon\mathbf{R}^n\to\mathbf{R}$ и отображения $F\colon\mathbf{R}^n\to\mathbf{R}^l$ и $G\colon\mathbf{R}^n\to\mathbf{R}^m$ дважды дифференцируемы в точке $\widetilde{x}\in\mathbf{R}^n$. Тогда для любых $c_1>0$ и $c_2\geqslant 0$ и любых $\widetilde{\lambda}\in\mathbf{R}^l$ и $\widetilde{\mu}\in\mathbf{R}^n$ точка $(\widetilde{x},\widetilde{\lambda},\widetilde{\mu})$ является стационарной в задаче* (58) *с целевой функцией, задаваемой формулой* (57), *в том и только том случае, когда точка $((\widetilde{x},\sigma(\widetilde{x},\widetilde{\mu},c_1,c_2)),(\widetilde{\lambda},\widetilde{\mu}))$, где $\sigma(\widetilde{x},\widetilde{\mu},c_1,c_2)\in\mathbf{R}^m$ определяется в соответствии с* (56), *является стационарной в задаче* (55) *с целевой функцией, задаваемой формулой* (54).

Задача 13. Доказать лемму 7.

Следующая теорема обобщает утверждение б) теоремы 4.3.6.

Теорема 7. *Пусть выполнены условия теоремы* 3. *Тогда существует число $\overline{c}_2>0$ такое, что если $c_2\in(0,\,\overline{c}_2]$, то окрестность U точки $(\overline{x},\,\overline{\lambda},\,\overline{\mu})$ и число $\overline{c}_1(c_2)\geqslant 0$ могут быть выбраны таким образом, чтобы для любого $c_1>\overline{c}_1(c_2)$ задача* (58) *с целевой функцией, задаваемой формулой* (57), *имела в U единственную стационарную точку $(\overline{x},\,\overline{\lambda},\,\overline{\mu})$.*

Задача 14. Используя леммы 2 и 7, теорему 4.3.6 и утверждение из задачи 12, доказать теорему 7.

Разумеется, для задачи (3), (4) могут использоваться и другие точные гладкие штрафные функции, помимо введенной в (54) (см. п. 4.3.3), что приводит к соответствующим точным гладким штрафным функциям для задачи (1), (2). Заметим, правда, что легко избавиться от присутствия одновременно как дополнительной переменной σ, так и двойственной переменной μ в данном случае не удается. Тем не менее семейства точных гладких штрафных функций для задачи (1), (2), зависящих только от прямой переменной этой задачи, известны, но их введение требует привлечения иных соображений.

Глава 5
СТРАТЕГИИ ГЛОБАЛИЗАЦИИ СХОДИМОСТИ

Многие обсуждавшиеся выше методы можно отнести к методам ньютоновского типа (см. § 3.2, п. 4.3.1, § 4.4, п. 4.5.3). Для таких методов обычно удается установить локальную сходимость со сверхлинейной скоростью, но глобальная сходимость, вообще говоря, не имеет места. Совершенно естественным является стремление так модифицировать методы ньютоновского типа, чтобы обеспечить их глобальную сходимость, сохранив при этом высокую скорость локальной сходимости. Иными словами, локально сходящиеся методы обретают практический смысл лишь после того, как процесс поиска достаточно хорошего начального приближения включен в окончательный алгоритм. При этом очень желательно, чтобы глобальный алгоритм был оптимизационным, т. е. ориентированным на поиск решений (а не, скажем, стационарных точек) рассматриваемой задачи, а также чтобы он был естественным образом связан с глобализуемым локальным алгоритмом (гибридные схемы, объединяющие в себе методы совершенно разной природы, обычно оказываются менее эффективными). Некоторым предназначенным для этого подходам и посвящена настоящая глава. Более детально с этими подходами можно познакомиться по специальной литературе (см., например, [40, 44]).

§ 5.1. Одномерный поиск

Напомним, что важную группу методов оптимизации составляют методы спуска (см. § 3.1, пп. 4.1.2, 4.1.3, § 4.2), для которых характерно наличие глобальной сходимости, а вот скорость их локальной сходимости обычно бывает не выше линейной. Это приводит к мысли объединить привлекательные глобальные свойства методов спуска и локальные свойства методов ньютоновского типа в рамках одного алгоритма. Для этого в последние вводят параметр длины шага, выбираемый на каждой итерации посредством процедуры одно-

мерного поиска для соответствующей *функции качества* [1]. В идеале эта функция должна вводиться так, чтобы любая ее критическая точка в естественных предположениях оказывалась стационарной в исходной задаче (функция качества может быть негладкой, и тогда понятие ее критической точки должно уточняться). Если организованный таким образом метод является методом спуска для задачи безусловной минимизации функции качества, то от него можно ожидать глобальной сходимости (в соответствующем смысле) к стационарным точкам исходной задачи. Более того, если удается показать, что вблизи стационарной точки, к которой сходится траектория метода, он принимает единичный параметр длины шага, т. е. превращается в исходный метод ньютоновского типа, то будет обеспечена и высокая скорость сходимости.

Такова общая идея глобализации сходимости методов посредством одномерного поиска для соответствующей функции качества. Если речь идет о задаче безусловной минимизации

$$f(x) \to \min, \quad x \in \mathbf{R}^n, \tag{1}$$

функции $f\colon \mathbf{R}^n \to \mathbf{R}$, то в роли функции качества может выступать сама целевая функция f. Предполагая достаточную гладкость f, продемонстрируем реализацию общей идеи в рамках алгоритма 3.2.3, т. е. схемы

$$x^{k+1} = x^k + \alpha_k d^k, \quad d^k = -Q_k f'(x^k), \quad k = 0, 1, \ldots, \tag{2}$$

где для каждого k симметрическая матрица $Q_k \in \mathbf{R}(n, n)$ положительно определена, а $\alpha_k > 0$ — параметр длины шага. Будем считать, что этот параметр выбирается в соответствии с правилом Армихо с начальным пробным значением $\widehat{\alpha} = 1$.

Согласно теореме 3.2.2 сверхлинейной скорости сходимости такого алгоритма к стационарной точке $\overline{x} \in \mathbf{R}^n$ задачи (1) (при наличии сходимости) можно ожидать, если Q_k «аппроксимирует» $(f''(\overline{x}))^{-1}$ при $k \to \infty$ в смысле соотношения (3.2.17). Более того, при этом $\alpha_k = 1$ для любого достаточно большого k. Вместе с тем, если просто положить $\alpha_k = 1$ для любого k, то глобальная сходимость метода не будет гарантирована: вдали от стационарных точек соответствующий шаг метода может оказываться слишком длинным, чтобы обеспечить хотя бы просто монотонное убывание последовательности $\{f(x^k)\}$.

С другой стороны, как отмечено в п. 3.2.3, алгоритм 3.2.3, в котором α_k не полагается «насильно» равным 1, а определяется посредством правила Армихо, является методом спуска для задачи (1). Более того, если предположить, что

$$\|Q_k\| \leqslant \Gamma, \quad \langle Q_k h, h\rangle \geqslant \gamma |h|^2 \quad \forall\, h \in \mathbf{R}^n, \quad \forall\, k \tag{3}$$

[1] Общепринятый английский термин — Merit function.

при некоторых $\Gamma > 0$ и $\gamma > 0$, то в силу (2) для произвольного k

$$\frac{\langle f'(x^k), d^k\rangle}{|d^k|^2} = -\frac{\langle f'(x^k), Q_k f'(x^k)\rangle}{|Q_k f'(x^k)|^2} \leqslant -\frac{\gamma}{\Gamma^2} < 0.$$

Таким образом, выполнено (3.1.9) при $\delta = -\gamma/\Gamma^2$, откуда вытекает существование не зависящей от k константы $\breve{\alpha}$, для которой

$$\alpha_k \geqslant \breve{\alpha} > 0 \tag{4}$$

(см. п. 3.1.1).

Следующая теорема является обобщением теоремы 3.1.1. Случаи использования правила одномерной минимизации и правила постоянного параметра сводятся к случаю использования правила Армихо так же, как при доказательстве теоремы 3.1.1.

Теорема 1. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ дифференцируема на $\mathbf{R}^n$ и ее производная непрерывна по Липшицу на $\mathbf{R}^n$. Пусть в алгоритме* 3.2.3 *используется правило Армихо. Пусть, наконец, для выбираемой в алгоритме последовательности матриц $\{Q_k\}$ выполнено* (3) *при некоторых $\Gamma > 0$ и $\gamma > 0$.*

Тогда любая предельная точка любой траектории $\{x^k\}$ алгоритма 3.2.3 *является стационарной точкой задачи* (1). *Если предельная точка существует или если функция f ограничена снизу на $\mathbf{R}^n$, то*

$$\{f'(x^k)\} \to 0 \quad (k \to \infty).$$

Задача 1. Используя (4) и повторяя с очевидными изменениями доказательство теоремы 3.1.1, доказать теорему 1.

Выполнения (3) обычно удается добиться, если генерировать последовательность матриц $\{Q_k\}$ с помощью тех или иных квазиньютоновских формул, а параметры длины шага α_k выбирать согласно правилу Вулфа (роль правила Вулфа в контексте квазиньютоновских методов уже отмечалась в п. 3.2.3).

Часто для каждого k в качестве Q_k берут обратную матрицу к положительно определенной модификации $f''(x^k)$ (см. п. 3.2.2). Если же, предполагая положительную определенность $f''(x^k)$, использовать матрицу $Q_k = (f''(x^k))^{-1}$ (что привлекательно с точки зрения скорости локальной сходимости), то предположение о выполнении второго условия в (3) будет немногим слабее требования сильной выпуклости функции f на $\mathbf{R}^n$. В случае отсутствия сильной выпуклости и, в частности, в случае наличия точек вырождения матрицы Гессе функции f разумное поведение траекторий метода существующей теорией не гарантируется. Очевидно, в случае попадания траектории метода в такую точку (либо в ее малую окрестность) сама итерация метода будет (численно) некорректной. В [46] поднимался вопрос о возможном «застревании» траекторий метода Ньютона

с регулировкой длины шага вблизи нестационарных точек вырождения матрицы Гессе целевой функции. Этот вопрос до сих пор до конца не решен: соответствующий пример не найден и в то же время не доказана невозможность такого примера. Недавние исследования показывают, что такой эффект крайне маловероятен, если вообще может иметь место.

Для задач условной оптимизации выбор функции качества не столь очевиден: этот выбор должен отражать стремление не только к уменьшению значения целевой функции, но и к удовлетворению ограничений. В § 5.4 при глобализации методов последовательного квадратичного программирования для этой цели используется негладкая точная штрафная функция. В роли функции качества часто выступают модифицированные функции Лагранжа и точные штрафные функции [41, 42], обычно негладкие, хотя, скажем, в работе [6] для глобализации сходимости некоторых ньютоновских методов активного множества использовались и точные гладкие штрафные функции. В методах, ориентированных на поиск стационарных точек, естественными функциями качества являются различные невязки систем условий оптимальности (Лагранжа, Каруша–Куна–Таккера). Иногда на разных итерациях используются различные функции качества (например, штрафная функция при различных значениях параметра штрафа). Это создает дополнительные трудности при теоретическом анализе, но может быть практически оправдано (см. § 5.4).

При подходящем выборе функции качества процедуры одномерного поиска могут использоваться для глобализации сходимости методов ньютоновского типа применительно не только к задачам оптимизации и связанным с ними специальным системам уравнений, но и к произвольному уравнению

$$\Phi(x) = 0, \tag{5}$$

где $\Phi\colon \mathbf{R}^n \to \mathbf{R}^n$ — гладкое отображение (см. [31]). В этом случае функцию качества следует выбирать так, чтобы в естественных предположениях ее критические точки оказывались решениями исходного уравнения. Один популярный выбор функции качества предлагается методом наименьших квадратов:

$$\varphi\colon \mathbf{R}^n \to \mathbf{R}, \quad \varphi(x) = \frac{1}{2}\,|\Phi(x)|^2.$$

Тогда

$$\varphi'(x) = (\Phi'(x))^{\mathrm{T}}\Phi(x), \quad x \in \mathbf{R}^n,$$

и если $\overline{x} \in \mathbf{R}^n$ — критическая точка функции φ, причем матрица $\Phi'(\overline{x})$ невырождена, то $\overline{x}$ — решение уравнения (5). Кроме того, в произвольной точке $x \in \mathbf{R}^n$, в которой матрица $\Phi'(x)$ невырождена, ньютоновское направление $d = -(\Phi'(x))^{-1}\Phi(x)$ является направлением убывания функции φ, если только x не является решением

уравнения (5). Это следует из леммы 3.1.1, поскольку

$$\langle \varphi'(x),\, d\rangle = -\langle (\Phi'(x))^{\mathrm{T}}\Phi(x),\, (\Phi'(x))^{-1}\Phi(x)\rangle = -|\Phi(x)|^2 < 0.$$

Заметим, однако, что соответствующий метод Ньютона с регулировкой длины шага вполне может «застревать» вблизи точек вырождения производной отображения Φ, не являющихся ни решениями уравнения (5), ни даже критическими точками функции φ, либо осциллировать между несколькими такими точками; примеры такого поведения известны.

§ 5.2. Методы доверительной области

Другой естественный подход к глобализации сходимости методов ньютоновского типа основан на следующем соображении. Вспомогательная задача всякого такого метода, решаемая на каждой итерации, является локальной аппроксимацией (в том или ином смысле) исходной задачи либо системы условий первого порядка оптимальности для исходной задачи. Разумеется, такой аппроксимации можно «доверять» лишь в достаточно малой окрестности текущего приближения. Представляется естественным решать вспомогательную задачу не на всем пространстве, а именно на такой окрестности, называемой *доверительной областью* [1]).

Например, для задачи безусловной оптимизации

$$f(x) \to \min, \quad x \in \mathbf{R}^n, \tag{1}$$

гладкой функции $f\colon \mathbf{R}^n \to \mathbf{R}$, если $x^k \in \mathbf{R}^n$ — текущее приближение, то итерационная вспомогательная задача ньютоновского метода имеет вид

$$\psi_k(x) \to \min, \quad x \in \mathbf{R}^n, \tag{2}$$

где

$$\psi_k\colon \mathbf{R}^n \to \mathbf{R},$$

$$\psi_k(x) = f(x^k) + \langle f'(x^k),\, x - x^k\rangle + \frac{1}{2}\,\langle H_k(x - x^k),\, x - x^k\rangle; \tag{3}$$

здесь $H_k \in \mathbf{R}(n,\, n)$ — некоторая симметрическая матрица (ср. с (3.2.12)). Итерация соответствующего метода доверительной области состоит в отыскании глобального решения задачи

$$\psi_k(x) \to \min, \quad x \in \overline{B}(x^k,\, \delta_k), \tag{4}$$

где $\delta_k > 0$ — радиус доверительной области, адекватное управление которым как раз и составляет главный вопрос для данного подхода.

Давняя дискуссия о том, какой из двух основных подходов к глобализации сходимости — одномерный поиск или стратегия доверительной области — естественнее и эффективнее, по-видимому, может

[1]) Общепринятый английский термин — Trust-region.

продолжаться бесконечно. Сторонники каждой точки зрения имеют свои аргументы. Не вдаваясь в детали этого спора, отметим лишь следующее.

Очевидно, отыскание глобального решения вспомогательной задачи (4) значительно более трудоемко, чем простой одномерный поиск. Более того, в зависимости от реализации метода доверительной области может возникать необходимость решения на текущей итерации не одной, а нескольких задач вида (4) при различных значениях $\delta_k > 0$ с целью подбора такого значения, которое обеспечило бы достаточное убывание функции f. Правда, хорошие реализации позволяют избежать необходимости многократного решения вспомогательных задач (см. ниже). В любом случае существуют эффективные специальные алгоритмы минимизации квадратичной функции на шаре [50], т. е. решение задачи (4) не требует привлечения методов глобальной оптимизации общего назначения.

Кроме того, интуитивно ясно, что (4) является куда лучшей аппроксимацией исходной задачи, чем задача минимизации f вдоль любого фиксированного направления. Следовательно, можно ожидать, что, несмотря на более высокую стоимость решения вспомогательной задачи метода доверительной области, решение этой задачи обеспечит больший прогресс в минимизации f, чем одномерный поиск, и итоговый алгоритм может быть более эффективен. Из практического опыта известно, что хорошие реализации методов доверительной области вполне могут конкурировать с аналогичными методами, использующими одномерный поиск, и даже превосходят последние в плане робастности в том смысле, что их применение реже заканчивается неудачей.

Наконец, с точки зрения теории методы доверительной области имеют то преимущество, что они не требуют положительной определенности матриц H_k и, как будет показано ниже, для них удается доказать глобальную сходимость не к любым стационарным точкам, а лишь к тем, в которых выполнено необходимое условие второго порядка оптимальности.

Еще раз подчеркнем, что эффективность всего метода доверительной области во многом определяется эффективностью используемого алгоритма решения вспомогательных задач вида (4). Хорошее описание эффективных алгоритмов решения (4) имеется, например, в [16]. Здесь ограничимся лишь кратким обсуждением основных подходов к этой проблеме.

Задача 1. Пусть заданы симметрическая матрица $A \in \mathbf{R}(n, n)$, элемент $b \in \mathbf{R}^n$ и число $\delta > 0$. Показать, что если $\widetilde{x} \in \mathbf{R}^n$ является глобальным решением задачи

$$\langle Ax, x\rangle + \langle b, x\rangle \to \min, \quad x \in \overline{B}(0, \delta), \tag{5}$$

то существует единственное число $\mu \geqslant 0$ такое, что

$$2(A+\mu E^n)\widetilde{x}+b=0, \quad \mu(|\widetilde{x}|-\delta)=0,$$

причем матрица $A+\mu E^n$ неотрицательно определена. Если матрица A положительно определена, то

$$\widetilde{x}=\begin{cases} -\frac{1}{2}A^{-1}b, & \text{если } \frac{1}{2}|A^{-1}b| \leqslant \delta, \\ -\frac{1}{2}(A+\mu E^n)^{-1}b, & \text{если } \frac{1}{2}|A^{-1}b| > \delta, \end{cases}$$

где $\mu \geqslant 0$ однозначно определяется равенством $|(A+\mu E^n)^{-1}b|/2=\delta$.

Задача 2. В обозначениях задачи 1 пусть элемент $\widetilde{x} \in \mathbf{R}^n$ и число μ таковы, что $2(A+\mu E^n)\widetilde{x}+b=0$ и матрица $A+\mu E^n$ неотрицательно определена. Доказать следующие утверждения:

а) если $|\widetilde{x}| \leqslant \delta$ и $\mu=0$, то $\widetilde{x}$ является глобальным решением задачи (5);

б) если $|\widetilde{x}|=\delta$, то $\widetilde{x}$ является глобальным решением задачи

$$\langle Ax, x\rangle + \langle b, x\rangle \to \min, \quad x \in \{x \in \mathbf{R}^n \,\big|\, |x|=\delta\};$$

в) если $|\widetilde{x}|=\delta$ и $\mu \geqslant 0$, то $\widetilde{x}$ является глобальным решением задачи (5).

Более того, если матрица $A+\mu E^n$ положительно определена, то в каждом из утверждений а)–в) $\widetilde{x}$ является единственным глобальным решением соответствующей задачи.

Пусть $\widetilde{x}^k \in \mathbf{R}^n$ — решение задачи (4) с положительно определенной матрицей H_k. Положим

$$d(\mu)=-(H_k+\mu E^n)^{-1}f'(x^k), \quad \mu \in \mathbf{R}_+. \tag{6}$$

Как указано в задаче 1,

$$\widetilde{x}^k = x^k + d^k, \quad d^k = d(\mu_k),$$

где $\mu_k=0$ при $|H_k^{-1}f'(x^k)| \leqslant \delta_k$ (при этом $\widetilde{x}^k$ совпадает с результатом чистой ньютоновской итерации, т. е. является решением задачи (2)), а в противном случае число $\mu_k \geqslant 0$ однозначно определяется скалярным уравнением

$$|d(\mu)|=\delta_k \tag{7}$$

относительно $\mu \in \mathbf{R}$. Это уравнение обычно решают (приближенно) с помошью специальных реализаций метода Ньютона. При этом в силу некоторых вычислительных причин предпочтительнее иметь дело с эквивалентным уравнением

$$\frac{1}{|d(\mu)|} = \frac{1}{\delta_k}.$$

После отыскания μ_k элемент d^k находится как решение линейной системы

$$(H_k + \mu_k E^n) d = -f'(x^k)$$

относительно $d \in \mathbf{R}^n$.

Другой подход основан на следующих соображениях. С изменением $\mu \in \mathbf{R}_+$ элемент $d(\mu)$ описывает в пространстве $\mathbf{R}^n$ непрерывную кривую — от ньютоновского шага $d(0) = -H_k^{-1} f'(x^k)$ до «почти градиентного» шага $d(\mu) \approx -f'(x^k)/\mu$ при больших значениях μ. Более того, как видно из (6) и (7), изменения μ обратны изменениям δ_k. Поэтому вместо решения вспомогательной задачи (4) для пробных значений радиуса δ_k доверительной области можно пытаться непосредственно отследить кривую значений $d(\cdot)$, аппроксимируя ее теми или иными методами [16, 50] (см. также § 5.3).

Перейдем к описанию базового *метода доверительной области*. Для простоты ограничимся случаем, когда $H_k = f''(x^k)$ для каждого k, оставляя рассмотрение квазиньютоновских вариантов метода читателю.

Алгоритм 1. Выбираем $x^0 \in \mathbf{R}^n$ и полагаем $k = 0$. Выбираем параметры $\overline{\delta} > 0$, $\varepsilon, \theta_1, \theta_2 \in (0, 1)$, $\theta_1 < \theta_2$.

1. Выбираем $\delta \geqslant \overline{\delta}$.

2. Полагаем $\delta_k = \delta$. Вычисляем $\widetilde{x}^k \in \mathbf{R}^n$ как глобальное решение задачи (4) с целевой функцией, задаваемой формулой (3) при $H_k = f''(x^k)$.

3. Если $\widetilde{x}^k = x^k$, то переходим к п. 4. В противном случае проверяем выполнение неравенства

$$f(\widetilde{x}^k) - f(x^k) \leqslant \varepsilon(\psi_k(\widetilde{x}^k) - f(x^k)). \tag{8}$$

   Если оно выполнено, то переходим к п. 4. В противном случае выбираем $\delta \in [\theta_1 |\widetilde{x}^k - x^k|, \theta_2 \delta_k]$ и переходим к п. 2.

4. Полагаем $x^{k+1} = \widetilde{x}^k$, увеличиваем номер шага k на 1 и переходим к п. 1.

Легко видеть, что если для некоторого k в п. 3 алгоритма реализуется равенство $\widetilde{x}^k = x^k$, то точка x^k является стационарной в задаче (1), причем в этой точке выполнено необходимое условие второго порядка оптимальности.

Величина $\psi_k(\widetilde{x}^k) - f(x^k)$ в правой части (8)) имеет смысл «предсказанного» вспомогательной задачей (4) убывания значения целевой функции. Таким образом, неравенство (8) означает, что реальное убывание значения целевой функции должно составлять как минимум

заданную (определяемую выбором параметра $\varepsilon \in (0, 1)$) долю от «предсказанного»убывания.

Разумеется, алгоритм 1 довольно схематичен; способы его реализации могут быть различны и здесь не конкретизируются. Например, как отмечено выше, отыскание нового (приближенного) решения вспомогательной задачи (4) для нового значения δ_k вовсе не обязательно подразумевает необходимость непосредственного решения новой оптимизационной задачи. Также не обсуждается вопрос о выборе начального значения δ в п. 1 алгоритма. На практике разумно брать большое δ, если предшествующая итерация была «очень успешной», т. е. привела к существенному убыванию значения целевой функции, и ограничиваться небольшим значением в противном случае.

Сначала покажем, что итерация алгоритма 1 определена корректно.

Предложение 1. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *дважды дифференцируема на* $\mathbf{R}^n$.

Тогда если для некоторого $k = 0, 1, \ldots$ *алгоритмом* 1 *сгенерирована точка* $x^k \in \mathbf{R}^n$, *которая либо не является стационарной в задаче* (1), *либо в этой точке нарушено сформулированное в теореме* 1.2.4 *необходимое условие второго порядка оптимальности, то алгоритм корректно определяет и точку* x^{k+1}, *причем* $f(x^{k+1}) < f(x^k)$.

Доказательство. Для каждого $\delta > 0$ обозначим через $\widetilde{x}(\delta)$ некоторое глобальное решение задачи (4) при $\delta_k = \delta$ (всюду далее считаем, что целевая функция задачи (4) задается формулой (3) при $H_k = f''(x^k)$) и положим

$$\rho(\delta) = \frac{f(\widetilde{x}(\delta)) - f(x^k)}{\psi_k(\widetilde{x}(\delta)) - f(x^k)}. \tag{9}$$

В сделанных предположениях относительно x^k имеет место следующее: либо $f'(x^k) \neq 0$, либо $f'(x^k) = 0$, но матрица $f''(x^k)$ не является неотрицательно определенной. Значит, $\widetilde{x}(\delta) \neq x^k$; более того, $\psi_k(\widetilde{x}(\delta)) < f(x^k)$. Поэтому достаточно показать, что

$$\lim_{\delta \to 0+} \rho(\delta) = 1 \tag{10}$$

(это будет означать, что после конечного числа изменений δ в п. 3 алгоритма неравенство (8) неизбежно выполнится).

Сначала предположим, что $f'(x^k) \neq 0$. Возьмем произвольный элемент $h \in \mathbf{R}^n$ такой, что $\langle f'(x^k), h\rangle < 0$, $|h| = 1$. Тогда, поскольку $x^k + \delta h \in \overline{B}(x^k, \delta_k)$, имеем

$$\psi_k(\widetilde{x}(\delta)) \leqslant \psi_k(x^k + \delta h) = f(x^k) + \delta\langle f'(x^k), h\rangle + \frac{\delta^2}{2}\langle f''(x^k)h, h\rangle,$$

поэтому

$$\psi_k(\widetilde{x}(\delta)) - f(x^k) \leqslant \delta\left(\langle f'(x^k), h\rangle + \frac{\delta}{2}\|f''(x^k)\|\right) \leqslant \frac{\delta}{2}\langle f'(x^k), h\rangle,$$

где последнее неравенство справедливо для любого достаточно малого $\delta > 0$. Далее,

$$f(\widetilde{x}(\delta)) - \psi_k(\widetilde{x}(\delta)) = f(\widetilde{x}(\delta)) - f(x^k) - \langle f'(x^k), \widetilde{x}(\delta) - x^k\rangle - \\ - \frac{1}{2}\langle f''(x^k)(\widetilde{x}(\delta) - x^k), \widetilde{x}(\delta) - x^k\rangle = o(\delta^2). \quad (11)$$

Из двух последних соотношений имеем

$$|\rho(\delta) - 1| = \left| \frac{f(\widetilde{x}(\delta)) - \psi_k(\widetilde{x}(\delta))}{\psi_k(\widetilde{x}(\delta)) - f(x^k)} \right| = o(\delta),$$

что и дает (10).

Пусть теперь $f'(x^k) = 0$, но существует элемент $h \in \mathbf{R}^n$ такой, что $\langle f''(x^k)h, h\rangle < 0$, $|h| = 1$. Тогда

$$\psi_k(\widetilde{x}(\delta)) \leqslant \psi_k(x^k + \delta h) = f(x^k) + \frac{\delta^2}{2}\langle f''(x^k)h, h\rangle,$$

поэтому

$$\psi_k(\widetilde{x}(\delta)) - f(x^k) \leqslant \frac{\delta^2}{2}\langle f''(x^k)h, h\rangle.$$

Отсюда и из (11) имеем

$$|\rho(\delta) - 1| = \frac{o(\delta^2)}{\delta^2},$$

что опять дает (10). □

Теорема 1. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ дважды непрерывно дифференцируема на $\mathbf{R}^n$.*

Тогда любая предельная точка любой траектории $\{x^k\}$ алгоритма 1 *является стационарной точкой задачи* (1), *причем в этой точке выполнено сформулированное в теореме* 1.2.4 *необходимое условие второго порядка оптимальности.*

Доказательство. В силу предложения 1 можно считать, что последовательность $\{f(x^k)\}$ монотонно убывает. Пусть последовательность $\{x^k\}$ имеет предельную точку $\overline{x} \in \mathbf{R}^n$. Тогда из монотонного убывания последовательности $\{f(x^k)\}$ следует ее ограниченность снизу, а значит, и сходимость. В частности, $f(x^{k+1}) - f(x^k) \to 0$ $(k \to \infty)$, откуда и из (8) следует, что

$$\psi_k(x^{k+1}) - f(x^k) \to 0 \quad (k \to \infty). \quad (12)$$

Пусть $\{x^{k_j}\}$ — сходящаяся к $\overline{x}$ подпоследовательность последовательности $\{x^k\}$. Возможны два случая:

$$\liminf_{j\to\infty} \delta_{k_j} > 0 \quad (13)$$

либо

$$\liminf_{j\to\infty} \delta_{k_j} = 0. \quad (14)$$

Рассмотрим сначала случай выполнения (13). Положим

$$\breve{\delta} = \liminf_{j\to\infty} \delta_{k_j} > 0,$$

и пусть $\overline{y} \in \mathbf{R}^n$ — глобальное решение задачи

$$\langle f'(\overline{x}),\, x-\overline{x}\rangle + \frac{1}{2}\,\langle f''(\overline{x})(x-\overline{x}),\, x-\overline{x}\rangle \to \min, \quad x \in \overline{D}, \tag{15}$$

$$\overline{D} = \left\{ x \in \mathbf{R}^n \,\middle|\, |x-\overline{x}| \leqslant \frac{\breve{\delta}}{4} \right\}. \tag{16}$$

Тогда для любого достаточно большого j

$$|\overline{y} - x^{k_j}| \leqslant |\overline{y}-\overline{x}| + |x^{k_j}-\overline{x}| \leqslant \frac{\breve{\delta}}{2} \leqslant \delta_{k_j}.$$

В частности, точка $\overline{y}$ допустима в задаче (4) при $k = k_j$, поэтому

$$\psi_{k_j}(x^{k_j+1}) \leqslant \psi_{k_j}(\overline{y}) =$$
$$= f(x^{k_j}) + \langle f'(x^{k_j}),\, \overline{y}-x^{k_j}\rangle + \frac{1}{2}\,\langle f''(x^{k_j})(\overline{y}-x^{k_j}),\, \overline{y}-x^{k_j}\rangle.$$

Переходя к пределу при $j\to\infty$ и используя (12), получаем

$$0 \leqslant \langle f'(\overline{x}),\, \overline{y}-\overline{x}\rangle + \frac{1}{2}\,\langle f''(\overline{x})(\overline{y}-\overline{x}),\, \overline{y}-\overline{x}\rangle,$$

откуда следует, что $\overline{x}$ также является глобальным решением задачи (15), (16) (а значение этой задачи равно нулю). Но $\overline{x} \in \operatorname{int}\overline{D}$, откуда, привлекая теоремы 1.2.3 и 1.2.4, легко получаем требуемое.

Пусть теперь выполнено (14). Без ограничения общности можем считать, что вся последовательность $\{\delta_{k_j}\}$ сходится к нулю. Но тогда для любого достаточно большого j выполняется $\delta_{k_j} < \overline{\delta}$, т. е. на шаге k_j в п. 3 алгоритма радиус доверительной области был уменьшен по крайней мере один раз. Поэтому найдется число $\widehat{\delta}_{k_j}$, для которого существует такое глобальное решение $\widetilde{x}(\delta)$ задачи (4) при $x^k = x^{k_j}$, $\delta_k = \widehat{\delta}_{k_j}$, что

$$f(\widetilde{x}(\widehat{\delta}_{k_j})) > f(x^{k_j}) + \varepsilon(\psi_{k_j}(\widetilde{x}(\widehat{\delta}_{k_j})) - f(x^{k_j})), \tag{17}$$

$$\delta_{k_j} \geqslant \theta_1 |\widetilde{x}(\widehat{\delta}_{k_j}) - x^{k_j}|.$$

В силу выбора подпоследовательности $\{\delta_{k_j}\}$ последнее, в частности, означает, что $\{\widetilde{x}(\widehat{\delta}_{k_j}) - x^{k_j}\} \to 0$ $(j\to\infty)$.

Дальнейшие рассуждения аналогичны тем, которые использовались при доказательстве предложения 1. Предположим, что существует элемент $h \in \mathbf{R}^n$ такой, что $|h| = 1$ и либо

$$\langle f'(\overline{x}),\, h\rangle < 0, \tag{18}$$

либо

$$f'(\overline{x}) = 0, \quad \langle f''(\overline{x})h,\, h\rangle < 0. \tag{19}$$

Тогда

$$\psi_{k_j}(\widetilde{x}(\widehat{\delta}_{k_j})) \leqslant \psi_{k_j}(x^{k_j} + \widehat{\delta}_{k_j} h) =$$
$$= f(x^{k_j}) + \widehat{\delta}_{k_j}\langle f'(x^{k_j}),\, h\rangle + \frac{\widehat{\delta}^2_{k_j}}{2}\langle f''(x^{k_j})h,\, h\rangle,$$

откуда в случае выполнения (18) следует, что для любого достаточно большого j

$$\psi_{k_j}(\widetilde{x}(\widehat{\delta}_{k_j})) - f(x^{k_j}) \leqslant$$
$$\leqslant \widehat{\delta}_{k_j}\left(\langle f'(x^{k_j}),\, h\rangle + \frac{\widehat{\delta}_{k_j}}{2}\|f''(x^{k_j})\|\right) \leqslant \frac{\widehat{\delta}^2_{k_j}}{2}\langle f'(\overline{x}),\, h\rangle.$$

Отсюда, полагая

$$\rho_{k_j} = \frac{f(\widetilde{x}(\widehat{\delta}_{k_j}) - f(x^{k_j})}{\psi_{k_j}(\widetilde{x}(\widehat{\delta}_{k_j})) - f(x^{k_j})},$$

легко приходим к равенству $|\rho_{k_j} - 1| = o(\widehat{\delta}^2_{k_j})/\widehat{\delta}_{k_j}$, означающему, что

$$\rho_{k_j} \to 1 \quad (j \to \infty). \tag{20}$$

Но это противоречит (17).

В случае выполнения (19) аналогичным образом приходим к (20) (см. доказательство предложения 1), а значит, и к противоречию с неравенством (17). $\square$

Последний вопрос, который требует ответа, состоит в следующем: будет ли в естественных предположениях итерация алгоритма 1 превращаться в чистую итерацию метода Ньютона вблизи решения? Этот вопрос легко решается положительно с помощью утверждения из задачи 1.

Задача 3. Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ дважды дифференцируема в некоторой окрестности точки $\overline{x} \in \mathbf{R}^n$, причем ее вторая производная непрерывна в этой точке. Пусть $\overline{x}$ является стационарной точкой задачи (1), причем в этой точке выполнено достаточное условие второго порядка оптимальности. Показать, что если для некоторого $k = 0, 1, \ldots$ алгоритмом 1 сгенерирована точка $x^k \in \mathbf{R}^n \setminus \{\overline{x}\}$, достаточно близкая к $\overline{x}$, то алгоритм генерирует

$$x^{k+1} = x^k - (f''(x^k))^{-1} f'(x^k).$$

Отметим, что вместо задачи (4) можно рассматривать вспомогательную задачу вида

$$\psi_k(x) \to \min, \quad x \in D_k,$$
$$D_k = \{x \in \mathbf{R}^n \,\big|\, |x - x^k|_\infty \leqslant \delta_k\},$$

т. е. вместо евклидовой нормы $|\cdot| = |\cdot|_2$ использовать $|\cdot|_\infty$ (использование других норм здесь едва ли может быть практически оправдано). Такая замена превращает вспомогательную задачу в задачу квадратичного программирования на параллелепипеде (о методах решения задач квадратичного программирования см. § 7.3), а свойства сходимости метода доверительной области по существу не изменяются.

Разумеется, как и для алгоритмов, использующих одномерный поиск, область применения методов доверительной области вовсе не ограничивается рамками безусловной оптимизации и методами ньютоновского типа. Для условных задач радиус доверительной области подбирается так, чтобы он обеспечивал достаточное убывание той или иной функции качества.

§ 5.3. Продолжение по параметру

Еще один продуктивный подход к получению хороших начальных приближений для локально сходящихся методов основан на следующей идее. Исходную оптимизационную задачу либо систему условий оптимальности для нее всегда можно включить в семейство задач, параметризованное (естественным или искусственным) скалярным параметром. Предположим, что для некоторого значения параметра решение либо хорошее приближение к решению соответствующей задачи семейства может быть легко найдено. Это решение (приближение) можно пытаться продолжить по параметру до того значения последнего, которому отвечает исходная задача. Разумеется, такое продолжение, если оно вообще возможно, обычно приходится осуществлять приближенно. Тем не менее при определенном согласовании погрешностей можно ожидать, что получаемая указанным способом точка будет достаточно хорошим приближением к решению исходной задачи.

Вновь ограничимся рассмотрением задачи безусловной оптимизации

$$f(x) \to \min, \quad x \in \mathbf{R}^n, \tag{1}$$

достаточно гладкой функции $f\colon \mathbf{R}^n \to \mathbf{R}$. Пусть достаточно гладкое отображение $\Phi\colon [0,\,1] \times \mathbf{R}^n \to \mathbf{R}^n$ обладает тем свойством, что

$$\Phi(1,\,x) = f'(x), \quad x \in \mathbf{R}^n,$$

причем некоторое решение $x^0 \in \mathbf{R}^n$ уравнения

$$\Phi(0,\,x) = 0 \tag{2}$$

может быть найдено с любой наперед заданной точностью. Например, если для некоторого отображения $\Phi_0\colon \mathbf{R}^n \to \mathbf{R}^n$ уравнение

$$\Phi_0(x) = 0$$

имеет известное решение, то можно взять

$$\Phi(t,\, x) = tf'(x) + (1-t)\Phi_0(x), \quad t \in [0,\, 1], \quad x \in \mathbf{R}^n.$$

В частности, всегда можно фиксировать $x^0 \in \mathbf{R}^n$ и положить

$$\Phi_0(x) = f'(x) - f'(x^0), \quad x \in \mathbf{R}^n,$$

тогда

$$\Phi(t,\, x) = f'(x) - (1-t)f'(x^0), \quad t \in [0,\, 1], \quad x \in \mathbf{R}^n.$$

Заметим, однако, что во многих случаях (и, в частности, при решении задач условной оптимизации) при введении отображения Φ используются значительно более изощренные подходы.

Далее, предположим, что существует отображение $\chi\colon [0,\, 1] \to \mathbf{R}^n$, непрерывное на $[0,\, 1]$ и такое, что

$$\chi(0) = x^0, \quad \Phi(t,\, \chi(t)) = 0 \quad \forall\, t \in [0,\, 1]. \tag{3}$$

Тогда точка $\overline{x} = \chi(1)$ является стационарной в задаче (1), и ее можно пытаться аппроксимировать, двигаясь вдоль задаваемой отображением χ кривой в $\mathbf{R}^n$ от $\chi(0)$ до $\chi(1)$.

Достаточные условия локальной продолжаемости решения даются классической теоремой о неявной функции (теорема 1.3.1). А именно, если матрица $\dfrac{\partial \Phi}{\partial x}\,(0,\, x^0)$ невырождена, то найдутся окрестность U точки 0 в $\mathbf{R}$ и отображение $\chi\colon U \to \mathbf{R}^n$, непрерывное (и даже гладкое) на U и такое, что

$$\chi(0) = x^0, \quad \Phi(t,\, \chi(t)) = 0 \quad \forall\, t \in U.$$

Однако гарантировать глобальную продолжаемость решения, т. е. выполнение включения $U \supset [0,\, 1]$, можно лишь в весьма сильных дополнительных предположениях. Некоторые утверждения такого рода имеются в [31] (см. также задачу 1 ниже). Весьма продуктивным может быть отказ от естественной параметризации: в значительно более слабых предположениях на Φ существует кривая решений в пространстве $\mathbf{R} \times \mathbf{R}^n$, задаваемая при некотором $\overline{s} > 0$ непрерывным отображением вида $(\tau(\cdot),\, \chi(\cdot))\colon [0,\, \overline{s}] \to \mathbf{R} \times \mathbf{R}^n$, $(\tau(0),\, \chi(0)) = (0,\, x^0)$, $\tau(\overline{s}) = 1$. В качестве нового параметра разумно использовать длину дуги такой кривой [40, 50]. В этом случае искомая кривая характеризуется начальной задачей

$$\frac{\partial \Phi}{\partial t}\,(t,\, x)\dot{t} + \frac{\partial \Phi}{\partial x}\,(t,\, x)\dot{x} = 0, \quad |(\dot{t},\, \dot{x})| = 1, \quad (t(0),\, x(0)) = (0,\, x^0),$$

где точка над t и над x означает производную по новому параметру.

Тем не менее для простоты далее ограничимся случаем глобальной продолжаемости по естественному параметру.

5.3.1. Конечные алгоритмы продолжения. Для произвольного натурального N введем на отрезке $[0,\, 1]$ сетку $\{t_0^N,\, t_1^N, \ldots, t_N^N\}$

$(0 = t_0^N < t_1^N < \ldots < t_N^N = 1)$ и положим

$$\Delta_N = \max_{i=0,1,\ldots,N-1} (t_{i+1}^N - t_i^N). \tag{4}$$

Пусть $x^{0,0} \in \mathbf{R}^n$ — имеющееся приближение к решению x^0 уравнения (2). Стартуя из этой точки, делаем k_0 шагов метода Ньютона для уравнения

$$\Phi(t_i^N, x) = 0 \tag{5}$$

при $i = 0$ (т. е. для уравнения (2)). Стартуя из полученной таким образом точки, делаем k_1 шагов метода Ньютона для уравнения (5) при $i = 1$. Повторяя эту процедуру для каждого $i = 0, 1, \ldots, N - 1$, приходим к формулам

$$x^{i,k+1} = x^{i,k} - \left(\frac{\partial \Phi}{\partial x} (t_i^N, x^{i,k}) \right)^{-1} \Phi(t_i^N, x^{i,k}), \tag{6}$$

$$k = 0, 1, \ldots, k_i - 1, \quad i = 0, 1, \ldots, N - 1,$$

где

$$x^{i,0} = x^{i-1,k_{i-1}}, \quad i = 1, \ldots, N - 1. \tag{7}$$

Полученную в результате этих вычислений точку $\widetilde{x}^N = x^{N-1,k_{N-1}}$ и предлагается использовать в качестве начального приближения к стационарной точке $\overline{x} = \chi(1)$ задачи (1).

В приводимом ниже анализе ограничимся случаем, когда $k_i = 1$ $\forall\, i = 0, 1, \ldots, N - 1$, т. е. когда для каждого уравнения вида (5) делается всего один шаг метода Ньютона.

Алгоритм 1. На отрезке $[0, 1]$ задаем сетку $\{t_0^N, t_1^N, \ldots, t_N^N\}$ $(0 = t_0^N < t_1^N < \ldots < t_N^N = 1)$. Выбираем $x^{0,0} \in \mathbf{R}^n$. Полагаем $\widetilde{x}^0 = x^{0,0}$ и вычисляем точку $\widetilde{x}^N \in \mathbf{R}^n$ по формулам

$$\widetilde{x}^{i+1} = \widetilde{x}^i - \left(\frac{\partial \Phi}{\partial x} (t_i^N, \widetilde{x}^i) \right)^{-1} \Phi(t_i^N, \widetilde{x}^i), \quad i = 0, 1, \ldots, N - 1. \tag{8}$$

Оказывается, в естественных предположениях можно гарантировать любую нужную близость генерируемой алгоритмом точки $\widetilde{x}^N$ к $\overline{x}$, если шаг сетки достаточно мал, а точка $x^{0,0}$ достаточно близка к x^0.

Напомним, что *графиком* отображения $\chi\colon [0, 1] \to \mathbf{R}^n$ называется множество

$$\operatorname{graph} \chi = \{(t, x) \in [0, 1] \times \mathbf{R}^n \,\big|\, \chi(t) = x\}.$$

Теорема 1. *Пусть отображение* $\Phi\colon [0, 1] \times \mathbf{R}^n \to \mathbf{R}^n$ *и точка* $x^0 \in \mathbf{R}^n$ *таковы, что существует отображение* $\chi\colon [0, 1] \to \mathbf{R}^n$,

непрерывное на $[0, 1]$ *и удовлетворяющее* (3). *Пусть* Φ *дифференцируемо по* x *в некоторой окрестности* $\operatorname{graph}\chi$, *причем частная производная* Φ *по* x *непрерывна на* $\operatorname{graph}\chi$. *Пусть, наконец,*

$$\det \frac{\partial \Phi}{\partial x}(t, \chi(t)) \neq 0 \quad \forall\, t \in [0, 1]. \tag{9}$$

Тогда для любого достаточно малого числа $\delta > 0$ *найдется число* $\Delta > 0$ *такое, что если сетка* $\{t_0^N, t_1^N, \ldots, t_N^N\}$ *на отрезке* $[0, 1]$ *и начальное приближение* $x^{0,0} \in \mathbf{R}^n$ *удовлетворяют условиям*

$$\Delta_N < \Delta, \quad |x^{0,0} - x^0| < \delta, \tag{10}$$

где Δ_N *введено в* (4), *то алгоритм* 1 *корректно определяет точку* $\widetilde{x}^N \in \mathbf{R}^n$, *причем*

$$|\widetilde{x}^N - \chi(1)| < \delta. \tag{11}$$

Доказательство. Используя (9), непрерывность частной производной Φ по x на $\operatorname{graph}\chi$, теорему о малом возмущении невырожденной матрицы и компактность отрезка $[0, 1]$, легко убедиться, что найдутся окрестность $\mathcal{U}$ множества $\operatorname{graph}\chi$ и число $M > 0$ такие, что

$$\det \frac{\partial \Phi}{\partial x}(t, x) \neq 0, \quad \left\| \left(\frac{\partial \Phi}{\partial x}(t, x) \right)^{-1} \right\| \leqslant M \quad \forall\, (t, x) \in \mathcal{U}. \tag{12}$$

С помощью рассуждений, аналогичных использованным при доказательстве теоремы 3.2.1, отсюда выводится, что для всякого $q \in (0, 1)$ найдется число $\widehat{\delta} > 0$ такое, что вне зависимости от выбора сетки, если для некоторого $i \in \{0, 1, \ldots, N-1\}$ точка $\widetilde{x}^i$ принадлежит $B(\chi(t_i^N), \widehat{\delta})$, то формула (8) корректно определяет точку $\widetilde{x}^{i+1}$, причем

$$|\widetilde{x}^{i+1} - \chi(t_i^N)| \leqslant q|\widetilde{x}^i - \chi(t_i^N)|. \tag{13}$$

Обозначим

$$\widetilde{\Delta}_N = \max_{i=0,1,\ldots,N-1} |\chi(t_{i+1}^N) - \chi(t_i^N)|. \tag{14}$$

Из (4) и равномерной непрерывности непрерывной функции на компакте следует, что для любого $\delta \in (0, \widehat{\delta})$ найдется $\Delta > 0$ такое, что при выполнении первого неравенства в (10) будем иметь

$$\frac{\widetilde{\Delta}_N}{1 - q} < \delta. \tag{15}$$

Предположим, далее, что алгоритмом корректно определены точки $\widetilde{x}^0, \widetilde{x}^1, \ldots, \widetilde{x}^i$, $i \in \{0, 1, \ldots, N-1\}$, причем $|\widetilde{x}^j - \chi(t_j^N)| < \widehat{\delta}$ $\forall\, j = 0, 1, \ldots, i$. Тогда алгоритм корректно определяет точку $\widetilde{x}^{i+1}$, причем в силу второго неравенства в (10) и (13)–(15)

$$|\widetilde{x}^{i+1} - \chi(t_{i+1}^N)| \leqslant |\widetilde{x}^{i+1} - \chi(t_i^N)| + |\chi(t_{i+1}^N) - \chi(t_i^N)| \leqslant$$

$$\leqslant q|\widetilde{x}^i - \chi(t_i^N)| + \widetilde{\Delta}_N \leqslant q(q|\widetilde{x}^{i-1} - \chi(t_{i-1}^N)| + \widetilde{\Delta}_N) + \widetilde{\Delta}_N \leqslant \ldots$$

$$\ldots \leqslant q^{i+1}|x^{0,0} - \chi(t_0^N)| + \widetilde{\Delta}_N \sum_{j=0}^{i} q^j \leqslant q^{i+1}\delta + \widetilde{\Delta}_N \frac{1-q^{i+1}}{1-q} =$$

$$= q^{i+1}\left(\delta - \frac{\widetilde{\Delta}_N}{1-q}\right) + \frac{\widetilde{\Delta}_N}{1-q} \leqslant q\left(\delta - \frac{\widetilde{\Delta}_N}{1-q}\right) + \frac{\widetilde{\Delta}_N}{1-q} = q\delta + \widetilde{\Delta}_N < \delta.$$

Отсюда следует требуемое. □

Разумеется, конечные методы продолжения можно строить не только на базе метода Ньютона и методов ньютоновского типа. Из приведенного доказательства видно, что существенным является наличие, как минимум, линейной скорости локальной сходимости базового метода к решению $\chi(t)$ уравнения

$$\Phi(t, x) = 0$$

для всех $t \in [0, 1]$. В связи с этим отметим важность условия (9): при его нарушении отслеживание кривой, задаваемой отображением χ, становится весьма трудной задачей. Во многих случаях преодолеть эти трудности позволяет упоминавшаяся выше смена параметризации [40].

Иногда параметризуют не условия оптимальности для задачи (1), а саму эту задачу, вводя функцию $\varphi\colon [0, 1] \times \mathbf{R}^n \to \mathbf{R}$ такую, что

$$\varphi(1, x) = f(x), \quad x \in \mathbf{R}^n,$$

причем некоторое (локальное или глобальное) решение x^0 задачи

$$\varphi(0, x) \to \min, \quad x \in \mathbf{R}^n,$$

может быть найдено с любой наперед заданной точностью. Если это решение продолжаемо в том смысле, что существует непрерывное на $[0, 1]$ отображение $\chi\colon [0, 1] \to \mathbf{R}^n$ такое, что $\chi(0) = x^0$ и $\chi(t)$ является решением задачи

$$\varphi(t, x) \to \min, \quad x \in \mathbf{R}^n,$$

для каждого $t \in [0, 1]$, то при построении конечных алгоритмов продолжения наряду с методами ньютоновского типа к последней задаче могут применяться, например, методы спуска, метод сопряженных градиентов и другие.

Задача 1. Пусть функция $\varphi\colon [0, 1] \times \mathbf{R}^n \to \mathbf{R}$ обладает следующим свойством: существует компакт $D \subset \mathbf{R}^n$ такой, что эта функция непрерывна на $[0, 1] \times D$ и для любого $t \in [0, 1]$ задача

$$\varphi(t, x) \to \min, \quad x \in D,$$

имеет единственное глобальное решение $\chi(t)$. Доказать, что отображение $\chi\colon [0, 1] \to \mathbf{R}^n$ непрерывно на $[0, 1]$.

5.3.2. Продолжение посредством решения начальной задачи. Если для каждого $t \in [0, 1]$ матрица $\dfrac{\partial \Phi}{\partial x}(t, \chi(t))$ невырождена, то при соответствующей гладкости отображения Φ теорема 1.3.1 гарантирует гладкость отображения χ, причем, дифференцируя второе равенство в (3) по t как композицию гладких отображений (либо просто ссылаясь на соответствующую формулу из теоремы 1.3.1), получаем, что $\chi(\cdot)$ является решением начальной задачи

$$\dot{x} = -\left(\frac{\partial \Phi}{\partial x}(t, x)\right)^{-1} \frac{\partial \Phi}{\partial t}(t, x), \quad x(0) = x^0. \tag{16}$$

Возникает естественная мысль аппроксимировать $\overline{x} = \chi(1)$, решая эту начальную задачу с помощью тех или иных приближенных схем, например, схемы Эйлера либо более точных схем Рунге–Кутта [5]. В случае использования схемы Эйлера приходим к следующему алгоритму (ср. с алгоритмом 1).

Алгоритм 2. На отрезке $[0, 1]$ задаем сетку $\{t_0^N, t_1^N, \ldots, t_N^N\}$ $(0 = t_0^N < t_1^N < \ldots < t_N^N = 1)$. Выбираем $x^{0,0} \in \mathbf{R}^n$. Полагаем $\widetilde{x}^0 = x^{0,0}$ и вычисляем точку $\widetilde{x}^N \in \mathbf{R}^n$ по формулам

$$\widetilde{x}^{i+1} = \widetilde{x}^i - (t_{i+1}^N - t_i^N) \times \left(\frac{\partial \Phi}{\partial x}(t_i^N, \widetilde{x}^i)\right)^{-1} \frac{\partial \Phi}{\partial t}(t_i^N, \widetilde{x}^i), \quad i = 0, 1, \ldots, N-1. \tag{17}$$

Теорема 2. *Пусть отображение* $\Phi\colon [0, 1] \times \mathbf{R}^n \to \mathbf{R}^n$ *и точка* $x^0 \in \mathbf{R}^n$ *таковы, что существует отображение* $\chi\colon [0, 1] \to \mathbf{R}^n$, *непрерывное на* $[0, 1]$ *и удовлетворяющее* (3). *Пусть* Φ *дифференцируемо в некоторой окрестности* $\mathcal{V}$ *графика* χ, *причем его производная непрерывна на* $\operatorname{graph}\chi$. *Пусть существует число* $L > 0$ *такое, что*

$$\|\Phi'(t, x) - \Phi'(t, \chi(t))\| \leqslant L|x - \chi(t)| \quad \forall\, (t, x) \in \mathcal{V}. \tag{18}$$

Пусть, наконец, выполнено (9).

Тогда для любых чисел $\varepsilon > 0$ *и* $C > 0$ *найдутся натуральное число* $\overline{N}$ *и число* $\delta > 0$ *такие, что если сетка* $\{t_0^N, t_1^N, \ldots, t_N^N\}$ *на отрезке* $[0, 1]$ *и начальное приближение* $x^{0,0} \in \mathbf{R}^n$ *удовлетворяют условиям*

$$\Delta_N \leqslant \frac{C}{N}, \quad N \geqslant \overline{N}, \quad |x^{0,0} - x^0| < \delta, \tag{19}$$

где Δ_N *введено в* (4), *то алгоритм* 2 *корректно определяет точку* $\widetilde{x}^N \in \mathbf{R}^n$, *причем*

$$|\widetilde{x}^N - \chi(1)| < \varepsilon.$$

Для доказательства потребуется следующий факт, называемый дискретным аналогом леммы Гронуолла.

Лемма 1. *Пусть числа* $\theta_0, \theta_1, \ldots, \theta_N$, a *и* b *удовлетворяют неравенствам*

$$0 \leqslant \theta_0 \leqslant a, \quad 0 \leqslant \theta_{i+1} \leqslant a + b\sum_{j=0}^{i} \theta_j \quad \forall\, i = 0,1,\ldots,N-1.$$

Тогда

$$\theta_i \leqslant a(1+b)^i \quad \forall\, i = 0,1,\ldots,N.$$

Задача 2. Используя индукцию, доказать лемму 1.

Доказательство теоремы 2. Используя (9), непрерывность $\chi(\cdot)$ на $[0, 1]$ и производной Φ на $\operatorname{graph}\chi$, (18), теорему о малом возмущении невырожденной матрицы и компактность отрезка $[0, 1]$, легко убедиться в следующем: найдутся окрестность $\mathcal{U} \subset \mathcal{V}$ множества $\operatorname{graph}\chi$ и число $\widetilde{L} > 0$ такие, что отображение

$$\Psi\colon \mathcal{U} \to \mathbf{R}^n, \quad \Psi(t, x) = -\left(\frac{\partial \Phi}{\partial x}(t, x)\right)^{-1} \frac{\partial \Phi}{\partial t}(t, x),$$

корректно определено, отображение $\Psi(\cdot, \chi(\cdot))\colon [0, 1] \to \mathbf{R}^n$ непрерывно на $[0, 1]$ и

$$\|\Psi(t, x) - \Psi(t, \chi(t))\| \leqslant \widetilde{L}|x - \chi(t)| \quad \forall\, (t, x) \in \mathcal{U}. \tag{20}$$

Выберем число $\widehat{\delta} > 0$ из условия

$$\{t\} \times B(\chi(t), \widehat{\delta}) \subset \mathcal{U} \quad \forall\, t \in [0, 1]$$

(возможность такого выбора следует из компактности $[0, 1]$).

Обозначим

$$\widetilde{\Delta}_N = \max_{i=0,1,\ldots,N-1} \max_{t \in [t_i^N,\, t_{i+1}^N]} |\Psi(t_i^N, \chi(t_i^N)) - \Psi(t, \chi(t))|. \tag{21}$$

Из (4), (19), непрерывности отображения $\Psi(\cdot, \chi(\cdot))$ на $[0,1]$ и равномерной непрерывности непрерывной функции на компакте легко следует, что найдется число $\delta > 0$ такое, что при выполнении первого неравенства в (10)

$$(\delta + \widetilde{\Delta}_N)(1 + \widetilde{L}\Delta_N)^N \leqslant$$

$$\leqslant (\delta + \widetilde{\Delta}_N)\left(1 + \frac{\widetilde{L}C}{N}\right)^N \leqslant (\delta + \widetilde{\Delta}_N)e^{\widetilde{L}C} < \min\{\varepsilon, \widehat{\delta}\} \tag{22}$$

при любом достаточно большом N.

Предположим, далее, что алгоритмом корректно определены точки $\widetilde{x}^0, \widetilde{x}^1, \ldots, \widetilde{x}^i$, $i \in \{0, 1, \ldots, N-1\}$, причем $|\widetilde{x}^j - \chi(t_j^N)| < \widehat{\delta}$ $\forall\, j = 0, 1, \ldots, i$. Тогда алгоритм корректно определяет точку $\widetilde{x}^{i+1}$, причем в силу (17)

$$\widetilde{x}^{i+1} = \widetilde{x}^i + (t_{i+1}^N - t_i^N)\Psi(t_i^N, \widetilde{x}^i) =$$

$$= \widetilde{x}^{i-1} + (t_i^N - t_{i-1}^N)\Psi(t_{i-1}^N, \widetilde{x}^{i-1}) + (t_{i+1}^N - t_i^N)\Psi(t_i^N, \widetilde{x}^i) = \dots$$

$$\dots = \widetilde{x}^0 + \sum_{j=0}^{i}(t_{j+1}^N - t_j^N)\Psi(t_j^N, \widetilde{x}^j) = \widetilde{x}^0 + \sum_{j=0}^{i}\int\limits_{t_j^N}^{t_{j+1}^N}\Psi(t_j^N, \widetilde{x}^j)\,dt.$$

Кроме того, в силу (16) и формулы Ньютона–Лейбница имеем

$$\chi(t_{i+1}^N) = x^0 + \int\limits_0^{t_{i+1}^N}\Psi(t, \chi(t))\,dt = x^0 + \sum_{j=0}^{i}\int\limits_{t_j^N}^{t_{j+1}^N}\Psi(t, \chi(t))\,dt.$$

Из двух последних соотношений, равенства (4), последнего неравенства в (19) и выражений (20) и (21) выводим

$$|\widetilde{x}^{i+1} - \chi(t^{i+1})| \leqslant |\widetilde{x}^0 - x^0| + \sum_{j=0}^{i}\int\limits_{t_j^N}^{t_{j+1}^N}|\Psi(t_j^N, \widetilde{x}^j) - \Psi(t, \chi(t))|\,dt \leqslant$$

$$\leqslant |x^{0,0} - x^0| + \sum_{j=0}^{i}\int\limits_{t_j^N}^{t_{j+1}^N}(|\Psi(t_j^N, \widetilde{x}^j) - \Psi(t_j^N, \chi(t_j^N))| +$$

$$+ |\Psi(t_j^N, \chi(t_j^N)) - \Psi(t, \chi(t))|)\,dt \leqslant \delta + \widetilde{\Delta}_N + \widetilde{L}\Delta_N\sum_{j=0}^{i}|\widetilde{x}^j - \chi(t_j^N)|.$$

Поэтому в силу леммы 1 и неравенства (22) имеем

$$|\widetilde{x}^{i+1} - \chi(t^{i+1})| \leqslant (\delta + \widetilde{\Delta}_N)(1 + \widetilde{L}\Delta_N)^{i+1} < \min\{\varepsilon, \widehat{\delta}\}.$$

Отсюда следует требуемое. $\square$

Если судить по теоремам 1 и 2, алгоритм 1 имеет очевидные преимущества перед алгоритмом 2 как в плане качества результата его применения, так и в плане условий гладкости, гарантирующих это качество. В то же время алгоритм 2 может быть менее требовательным к степени мелкости используемой сетки, поэтому оба рассмотренных метода продолжения находят применение на практике, причем особенно часто они применяются в комбинации друг с другом. Например, приближение, получаемое в очередной точке сетки посредством схемы Эйлера, может уточняться за счет одной или нескольких ньютоновских итераций. Для таких комбинированных алгоритмов используется название *предиктор-корректор*. Сетка в них обычно не фиксируется заранее, а конструируется адаптивно, по ходу процесса.

Более подробно о методах продолжения (в том числе и совершенно иной природы, чем рассмотренные выше) см., например, [40].

§ 5.4. Глобализация сходимости методов последовательного квадратичного программирования

Привлекательные свойства локальной сходимости методов последовательного квадратичного программирования (SQP) делают весьма желательным построение глобально сходящихся модификаций этих методов. Подчеркнем, что в данном контексте применяются обе основные стратегии глобализации сходимости, обсуждавшиеся в этой главе: как одномерный поиск для подходящей функции качества, так и методы доверительной области. Здесь подробно рассматривается только первая стратегия, реализация которой для методов SQP несколько проще, чем реализация второй (о которой см. [50]).

5.4.1. Глобализация сходимости. Предполагая гладкость функции $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображений $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ на $\mathbf{R}^n$, будем рассматривать задачу

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in \mathbf{R}^n \,\big|\, F(x) = 0,\ G(x) \leqslant 0\}. \tag{2}$$

Пусть, как обычно,

$$L\colon \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R},$$

$$L(x,\ \lambda,\ \mu) = f(x) + \langle \lambda,\ F(x)\rangle + \langle \mu,\ G(x)\rangle$$

— функция Лагранжа задачи (1), (2).

В роли функции качества, используемой при одномерном поиске для методов SQP, обычно выступает некоторая точная штрафная функция задачи (1), (2) при соответствующем значении параметра штрафа $c \geqslant 0$, например (см. п. 4.7.2),

$$\varphi_c\colon \mathbf{R}^n \to \mathbf{R}, \quad \varphi_c(x) = f(x) + c\psi(x), \tag{3}$$

где

$$\psi(x) = |F(x)|_1 + \sum_{i=1}^{m} \max\{0,\ g_i(x)\}, \quad x \in \mathbf{R}^n, \tag{4}$$

а $g_i(\cdot)$ — компоненты отображения G, $i = 1, \ldots, m$.

Подчеркнем, что в этом пункте вопрос о скорости сходимости методов SQP вообще не обсуждается: цель состоит лишь в обеспечении их глобальной сходимости. Получаемые на этом пути методы можно отнести к так называемым методам линеаризации [6, 34, 37]. В п. 5.4.2 будет показано, как следует видоизменить функцию φ_c

для того, чтобы не только обеспечить глобальную сходимость соответствующих модификаций методов SQP, но и сохранить сверхлинейную скорость локальной сходимости.

Алгоритм 1. Выбираем $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ и симметрическую матрицу $H_0 \in \mathbf{R}(n, n)$ и полагаем $k = 0$. Фиксируем параметры $\overline{c} > 0$ и $\theta, \varepsilon \in (0, 1)$.

1. Вычисляем $d^k \in \mathbf{R}^n$ как стационарную точку задачи квадратичного программирования

$$\langle f'(x^k), d\rangle + \frac{1}{2}\langle H_k d, d\rangle \to \min, \quad d \in D_k, \tag{5}$$

$$D_k = \{d \in \mathbf{R}^n \mid F(x^k) + F'(x^k)d = 0,\ G(x^k) + G'(x^k)d \leqslant 0\}. \tag{6}$$

Вычисляем $y^k \in \mathbf{R}^l$ и $z^k \in \mathbf{R}^m_+$ как отвечающие стационарной точке d^k задачи (5), (6) множители Лагранжа.

2. Если $d^k = 0$, то полагаем α_k равным любому числу и переходим к п. 3. В противном случае выбираем число [1])

$$c_k \geqslant |(y^k, z^k)|_\infty + \overline{c} \tag{7}$$

и полагаем $\alpha = 1$.

2.1. Проверяем выполнение неравенства

$$\varphi_{c_k}(x^k + \alpha d^k) \leqslant \varphi_{c_k}(x^k) + \varepsilon\alpha\Delta_k, \tag{8}$$

где

$$\Delta_k = \langle f'(x^k), d^k\rangle - c_k\psi(x^k), \tag{9}$$

а функции φ_{c_k} и ψ вводятся в соответствии с (3), (4).

2.2. Если (8) не выполнено, то заменяем α на $\theta\alpha$ и переходим к п. 2.1. В противном случае полагаем $\alpha_k = \alpha$.

3. Полагаем $x^{k+1} = x^k + \alpha_k d^k$, $\lambda^{k+1} = y^k$, $\mu^{k+1} = z^k$.

4. Увеличиваем номер шага k на 1, выбираем новую симметрическую матрицу $H_k \in \mathbf{R}(n, n)$ и переходим к п. 1.

Реализация описанного алгоритма подразумевает существование у задачи квадратичного программирования (5), (6) стационарной точки для каждого k. Вопрос о разрешимости вспомогательных задач

[1]) Здесь существенно следующее обстоятельство: нормы $|\cdot|_1$ и $|\cdot|_\infty$ являются *взаимодвойственными* в том смысле, что

$$|x|_1 = \max_{\xi \in \mathbf{R}^n:\, |\xi|_\infty = 1} \langle x, \xi\rangle, \quad |x|_\infty = \max_{\xi \in \mathbf{R}^n:\, |\xi|_1 = 1} \langle x, \xi\rangle \quad \forall\, x \in \mathbf{R}^n.$$

в контексте методов SQP не имеет автоматического положительного ответа. Прежде всего, система ограничений задачи (5), (6) может вообще быть несовместной.

Задача 1. Показать, что если отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ аффинно, функции g_i выпуклы и дифференцируемы в точке $x^k \in \mathbf{R}^n$, $i = 1, \ldots, m$, а заданное в (2) множество D непусто, то непусто и множество D_k, заданное в (6).

Другой пример условия, гарантирующего непустоту D_k, доставляет линейная независимость совокупности строк матриц $F'(x^k)$ и $G'(x^k)$, но предположение о выполнении этого условия в любой точке генерируемой методом траектории $\{x^k\}$ весьма ограничительно. В любом случае непустоты допустимых множеств вспомогательных задач всегда можно добиться за счет довольно несложных модификаций самих этих задач (см. задачу 3 ниже).

Другая возможная трудность состоит том, что целевая функция задачи (5), (6) может быть не ограничена снизу на допустимом множестве, и тогда существование у этой задачи стационарной точки тоже не гарантировано. Такая ситуация исключена в случае положительно определенной матрицы H_k, и в этом смысле предпочтительнее выбирать матрицы H_k именно положительно определенными. Согласно доказываемой ниже теореме 1 для обеспечения глобальной сходимости рассматриваемого алгоритма достаточно для каждого k в качестве H_k брать одну и ту же произвольную положительно определенную симметрическую матрицу, например $H_k = E^n$. В то же время с точки зрения скорости локальной сходимости наиболее привлекательным является выбор $H_k = \dfrac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)$ (см. п. 4.4.2), но положительная определенность таких матриц не гарантируется никакими естественными в этом контексте предположениями. Здесь может помочь замена L модифицированной функцией Лагранжа L_c при достаточно большом значении параметра штрафа c (см. задачи 4.3.8 и 4.4.1) либо непосредственная модификация матриц $\dfrac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)$ (см. п. 3.2.2). Кроме того, известны формулы пересчета H_k в духе квазиньютоновских методов (возможно, в сочетании с несколько видоизмененным правилом одномерного поиска; см. комментарии о правиле Вулфа в п. 3.2.3), которые обеспечивают положительную определенность H_k для каждого k. Помимо прочего такие формулы не требуют вычисления вторых производных функции f и отображений F и G.

Напомним, что через $\psi'(x; d)$ обозначается производная функции ψ в точке $x \in \mathbf{R}^n$ по направлению $d \in \mathbf{R}^n$ (см. п. 4.5.2), а че-

рез $I(x) = \{i = 1, \ldots, m \mid g_i(x) = 0\}$ — множество индексов ограничений-неравенств задачи (1), (2), активных в точке x.

Задача 2. Показать, что если отображения $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ дифференцируемы в точке $x \in \mathbf{R}^n$, то введенная в (4) функция ψ дифференцируема в точке x по любому направлению $d \in \mathbf{R}^n$, причем

$$\psi'(x;\, d) = \sum_{j \in J^+(x)} \langle f_j'(x),\, d\rangle + \sum_{j \in J^0(x)} |\langle f_j'(x),\, d\rangle| -$$

$$- \sum_{j \in J^-(x)} \langle f_j'(x),\, d\rangle + \sum_{i \in I^+(x)} \langle g_i'(x),\, d\rangle + \sum_{i \in I(x)} \max\{0,\, \langle g_i'(x),\, d\rangle\},$$

где $f_j(\cdot)$ — компоненты отображения F, $j = 1, \ldots, l$, $J^+(x) = \{j = 1, \ldots, l \mid f_j(x) > 0\}$, $J^0(x) = \{j = 1, \ldots, l \mid f_j(x) = 0\}$, $J^-(x) = \{j = 1, \ldots, l \mid f_j(x) < 0\}$, $I^+(x) = \{i = 1, \ldots, m \mid g_i(x) > 0\}$.

Лемма 1. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображения $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ дифференцируемы в точке $x^k \in \mathbf{R}^n$. Пусть $H_k \in \mathbf{R}(n, n)$ — симметрическая матрица, $d^k \in \mathbf{R}^n$ — стационарная точка задачи (5), (6), $y^k \in \mathbf{R}^l$ и $z^k \in \mathbf{R}^m_+$ — отвечающие ей множители Лагранжа, а число c_k удовлетворяет (7). Пусть функции φ_{c_k} и ψ введены в соответствии с (3), (4).*

Тогда функция φ_{c_k} дифференцируема в точке x^k по любому направлению, причем

$$\varphi'_{c_k}(x^k;\, d^k) \leqslant \Delta_k \leqslant -\langle H_k d^k,\, d^k\rangle - \overline{c}\psi(x^k), \tag{10}$$

где число Δ_k определено в (9).

Доказательство. По условию тройка $(d^k,\, y^k,\, z^k)$ удовлетворяет системе Каруша–Куна–Таккера задачи (5), (6), т. е.

$$f'(x^k) + H_k d^k + (F'(x^k))^{\mathrm{T}} y^k + (G'(x^k))^{\mathrm{T}} z^k = 0, \tag{11}$$

$$F(x^k) + F'(x^k) d^k = 0, \tag{12}$$

$$G(x^k) + G'(x^k) d^k \leqslant 0, \quad z^k \geqslant 0, \tag{13}$$

$$z_i^k (g_i(x^k) + \langle g_i'(x^k),\, d^k\rangle) = 0, \quad i = 1, \ldots, m. \tag{14}$$

Из (12) следует, что $\langle f_j'(x^k),\, d^k\rangle = -f_j(x^k) \quad \forall\, j = 1, \ldots, l$, и поэтому

$$-|f_j(x^k)| = \begin{cases} \langle f_j'(x^k),\, d^k\rangle, & \text{если } j \in J^+(x^k), \\ |\langle f_j'(x^k),\, d^k\rangle|, & \text{если } j \in J^0(x^k), \\ -\langle f_j'(x^k),\, d^k\rangle, & \text{если } j \in J^-(x^k). \end{cases}$$

Кроме того, согласно первому неравенству в (13)

$$-\max\{0,\, g_i(x^k)\} = \begin{cases} -g_i(x^k) \geqslant \langle g_i'(x^k),\, d^k\rangle, & \text{если } i \in I^+(x^k), \\ 0 = \max\{0,\, \langle g_i'(x^k),\, d^k\rangle\}, & \text{если } i \in I(x^k). \end{cases}$$

Используя два последних соотношения, (9), а также утверждение из задачи 2, получаем первое неравенство в (10).

Далее, умножая скалярно обе части (11) на d^k и используя (12), (14), получаем

$$\begin{aligned} \langle f'(x^k), d^k\rangle &= -\langle H_k d^k, d^k\rangle - \langle y^k, F'(x^k)d^k\rangle - \langle z^k, G'(x^k)d^k\rangle = \\ &= -\langle H_k d^k, d^k\rangle + \langle y^k, F(x^k)\rangle + \langle z^k,\, G(x^k)\rangle \leqslant \\ &\leqslant -\langle H_k d^k,\, d^k\rangle + \sum_{j=1}^{l} |y_j^k||f_j(x^k)| + \sum_{i=1}^{m} z_i^k \max\{0, g_i(x^k)\} \leqslant \\ &\leqslant -\langle H_k d^k,\, d^k\rangle + |y^k|_\infty |F(x^k)|_1 + |z^k|_\infty \sum_{i=1}^{m} \max\{0, g_i(x^k)\}. \end{aligned}$$

Отсюда в силу (4), (7) и (9) имеем

$$\Delta_k = \langle f'(x^k),\, d^k\rangle - c_k|F(x^k)|_1 - c_k \sum_{i=1}^{m} \max\{0,\, g_i(x^k)\} \leqslant$$
$$\leqslant -\langle H_k d^k,\, d^k\rangle - \overline{c}\psi(x^k),$$

а это и есть второе неравенство в (10). □

Из второго неравенства в (10) следует, что для каждого k, если матрица H_k положительно определена и генерируемое в п. 1 алгоритма 1 направление d^k отлично от нуля, то

$$\Delta_k < 0. \tag{15}$$

Отсюда и из первого неравенства в (10) несложно вывести следующее: $d^k \in \mathcal{D}_{\varphi_{c_k}}(x^k)$ (ср. с леммой 3.1.1), и процедура одномерного поиска в п. 2 алгоритма корректно определена в том смысле, что она принимает некоторое значение параметра длины шага $\alpha_k > 0$ после конечного числа дроблений (ср. с леммой 3.1.2). На самом деле в (8) вместо величины Δ_k можно было бы непосредственно использовать производную по направлению $\varphi_{c_k}'(x^k;\, d^k)$, и тогда неравенство (8) стало бы очевидным обобщением неравенства Армихо на случай негладкой функции. Однако формула (9) для Δ_k проще, чем явное выражение для $\varphi_{c_k}'(x^k;\, d^k)$ (см. задачу 2), и, в частности, не использует производных отображений F и G.

Вообще говоря, на разных шагах алгоритма 1 используются различные значения параметра штрафа, причем они определяются не

заранее, а в ходе вычислительного процесса. Однако при доказательстве глобальной сходимости удобно предполагать, что для всех достаточно больших k

$$c_k = c, \tag{16}$$

где c — некоторая константа. Заметим, что при этом, начиная с некоторого шага, алгоритм 1 можно трактовать как метод спуска для задачи безусловной оптимизации

$$\varphi_c(x) \to \min, \quad x \in \mathbf{R}^n.$$

Возможность такого выбора параметров штрафов подразумевает, что генерируемая алгоритмом последовательность $\{(y^k,\, z^k)\}$ является ограниченной (см. (7)). С практической точки зрения последнее предположение вполне естественно. Кроме того, ограниченность последовательности $\{(y^k,\, z^k)\}$ может быть гарантирована при некоторых разумных условиях, однако здесь, дабы не отвлекаться на второстепенные вопросы, будем ограниченность этой последовательности просто предполагать. В таком случае выполнения (16) при некотором c для всех достаточно больших k можно добиться, если, например, использовать следующую очевидную процедуру выбора c_k. На нулевом шаге полагаем $c_0 = |(y^0,\, z^0)|_\infty + 2\overline{c}$. На каждом шаге с номером $k = 1, 2, \ldots$ проверяем выполнение неравенства (7) при $c_k = c_{k-1}$. Если оно выполнено, то принимаем такое c_k; в противном случае полагаем $c_k = |(y^k,\, z^k)|_\infty + 2\overline{c}$. Более изощренные способы выбора c_k можно найти в [42].

Теорема 1. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображения* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *и* $G\colon \mathbf{R}^n \to \mathbf{R}^m$ *дифференцируемы на* $\mathbf{R}^n$ *и их производные непрерывны по Липшицу на* $\mathbf{R}^n$. *Пусть в алгоритме* 1 *матрицы* H_k *выбираются так, что последовательность* $\{H_k\}$ *ограничена и*

$$\langle H_k h,\, h\rangle \geqslant \gamma |h|^2 \quad \forall\, h \in \mathbf{R}^n, \quad \forall\, k \tag{17}$$

при некотором $\gamma > 0$, *и пусть траектория* $\{(x^k,\, \lambda^k, \mu^k)\}$ *сгенерирована этим алгоритмом, причем для любого достаточно большого* k *выполнено равенство* (16), *где* c — *некоторая константа.*

Тогда при $k \to \infty$ *либо*

$$\varphi_{c_k}(x^k) \to -\infty, \tag{18}$$

либо

$$\{d^k\} \to 0,$$

$$\frac{\partial L}{\partial x}\,(x^k,\, \lambda^{k+1},\, \mu^{k+1}) \to 0,$$

$$F(x^k) \to 0,$$

$$\max\{0,\, g_i(x^k)\} \to 0 \qquad \forall\, i = 1, \ldots, m,$$

$$\langle \mu^{k+1}, G(x^k)\rangle \to 0.$$

В частности, если $(\overline{x}, \overline{\lambda}, \overline{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ *является предельной точкой траектории* $\{(x^k, \lambda^k, \mu^k)\}$, *то* $\overline{x}$ *— стационарная точка задачи* (5), (6), *а* $\overline{\lambda}$ *и* $\overline{\mu}$ *— отвечающие ей множители Лагранжа.*

Заметим, что требование (17) является весьма обременительным (см. комментарии в п. 4.4.2 о возможном отсутствии положительной определенности матриц H_k). С другой стороны, это требование существенно упрощает анализ глобальной сходимости алгоритма 1.

Доказательство. Напомним, что тройка (d^k, y^k, z^k) удовлетворяет условиям (11)–(14) для каждого k. Если $d^k = 0$ для некоторого k, то эти условия означают, что точка (x^k, y^k, z^k) удовлетворяет системе Каруша–Куна–Таккера задачи (5), (6), что и дает требуемый результат. Всюду далее предполагаем, что $d^k \neq 0 \quad \forall k$.

Докажем, что генерируемая в п. 2 алгоритма последовательность $\{\alpha_k\}$ параметров длины шага отделена от нуля. Возьмем произвольное число $\alpha \in (0, 1]$. Пусть $M > 0$ — максимум из констант Липшица для производных f, F и G на $\mathbf{R}^n$. Тогда согласно лемме 3.1.4 $\forall k, \ \forall i = 1, \ldots, m$

$$\begin{aligned}\max\{0, g_i(x^k + \alpha d^k)\} - \max\{0,\ g_i(x^k) + \alpha\langle g_i'(x^k),\ d^k\rangle\} \leqslant \\ \leqslant \max\{0,\ g_i(x^k + \alpha d^k) - g_i(x^k) - \alpha\langle g_i'(x^k),\ d^k\rangle\} \leqslant \\ \leqslant |g_i(x^k + \alpha d^k) - g_i(x^k) - \alpha\langle g_i'(x^k),\ d^k\rangle| \leqslant \frac{M}{2}\,\alpha^2|d^k|^2. \quad (19)\end{aligned}$$

Кроме того,

$$\begin{aligned}\max\{0, g_i(x^k) + \alpha\langle g_i'(x^k),\ d^k\rangle\} = \\ = \max\{0,\ \alpha(g_i(x^k) + \langle g_i'(x^k),\ d^k\rangle) + (1-\alpha)g_i(x^k)\} \leqslant \\ \leqslant \alpha\max\{0,\ g_i(x^k) + \langle g_i'(x^k),\ d^k\rangle\} + (1-\alpha)\max\{0,\ g_i(x^k)\} = \\ = (1-\alpha)\max\{0,\ g_i(x^k)\},\end{aligned}$$

где последнее равенство следует из (13). Но тогда из (19) имеем

$$\max\{0,\ g_i(x^k + \alpha d^k)\} \leqslant (1-\alpha)\max\{0,\ g_i(x^k)\} + \frac{M}{2}\,\alpha^2|d^k|^2.$$

Далее, в силу (12) и леммы 3.1.4 $\forall j = 1, \ldots, l$

$$\begin{aligned}|f_j(x^k + \alpha d^k)| = |f_j(x^k + \alpha d^k) - \alpha f_j(x^k) - \alpha\langle f_j'(x^k),\ d^k\rangle| \leqslant \\ \leqslant (1-\alpha)|f_j(x^k)| + \frac{M}{2}\,\alpha^2|d^k|^2.\end{aligned}$$

Отсюда, используя (3), (4) и вновь привлекая лемму 3.1.4, выводим

$$\begin{aligned}\varphi_{c_k}(x^k + \alpha d^k) = f(x^k + \alpha d^k) + \\ + c_k\Big(|F(x^k + \alpha d^k)|_1 + \sum_{i=1}^{m}\max\{g_i(0,\ x^k + \alpha d^k)\}\Big) \leqslant\end{aligned}$$

$$\leqslant f(x^k) + \alpha\langle f'(x^k),\, d^k\rangle + \\ + c_k(1-\alpha)\Big(|F(x^k)|_1 + \sum_{i=1}^{m}\max\{0,\, g_i(x^k)\}\Big) + C_k\alpha^2|d^k|^2 = \\ = \varphi_{c_k}(x^k) + \alpha\Delta_k + C_k\alpha^2|d^k|^2,$$

где

$$C_k = \frac{M}{2}\,(1 + (l+m)c_k) > 0. \tag{20}$$

Но тогда неравенство (8) имеет место для любого $\alpha \in (0,\, \overline{\alpha}_k]$, где

$$\overline{\alpha}_k = \frac{(\varepsilon-1)\Delta_k}{C_k|d^k|^2}\,.$$

В силу второго неравенства в (10) и (17)

$$\overline{\alpha}_k \geqslant \frac{(1-\varepsilon)\gamma}{C_k}\,,$$

причем последовательность $\{C_k\}$ ограничена в силу (20) и равенства (16), справедливого при больших k. Но тогда существует такое не зависящее от k число $\breve{\alpha} > 0$, что

$$\alpha_k \geqslant \breve{\alpha} > 0. \tag{21}$$

Далее, из (8), (16) и (21) следует, что

$$\varphi_{c_{k+1}}(x^{k+1}) \leqslant \varphi_{c_k}(x^k) + \varepsilon\breve{\alpha}\Delta_k \tag{22}$$

для любого достаточно большого k, причем, напомним, имеет место неравенство (15). В частности, последовательность $\{\varphi_{c_k}(x^k)\}$ монотонно убывает. Из предположения о неограниченности этой последовательности снизу следует предельное соотношение (18) при $k \to \infty$. Если же эта последовательность ограничена, то она сходится, и из (22) в этом случае следует, что

$$\Delta_k \to 0 \quad (k \to \infty).$$

Но тогда, в очередной раз используя второе неравенство в (10), а также (17), имеем

$$\{d^k\} \to 0 \quad (k \to \infty),$$

откуда и из (11) немедленно следует, что $\frac{\partial L}{\partial x}(x^k, \lambda^{k+1}, \mu^{k+1}) \to 0$ $k \to \infty$.

Кроме того, из полученных предельных соотношений и из второго неравенства в (10) следует, что $\psi(x^k) \to 0$, откуда в силу (4) вытекают предельные соотношения $F(x^k) \to 0$ и $\max\{0,\, g_i(x^k)\} \to 0$ $\forall\, i = 1, \ldots m$ $(k \to \infty)$.

Далее, из (9) и доказанных предельных соотношений $\Delta_k \to 0$ и $\psi(x^k) \to 0$ вытекает, что $\langle f'(x^k), d^k\rangle \to 0$ $(k \to \infty)$. Поэтому, в силу (11) и доказанного предельного соотношения $\{d^k\} \to 0$, имеем $\langle y^k, F'(x^k)d^k\rangle + \langle z^k, G'(x^k)d^k\rangle \to 0$ $(k \to \infty)$. Отсюда, используя (12)

и доказанное соотношение $\{F(x^k)\} \to 0$, выводим $\langle y^k, F'(x^k)d^k\rangle = -\langle y^k, F(x^k)\rangle \to 0$, и поэтому $\langle z^k, G'(x^k)d^k\rangle \to 0$ $(k \to \infty)$. Значит, согласно (14),

$$\langle \mu^{k+1}, G(x^k)\rangle = \langle z^k, G(x^k)\rangle = -\langle z^k, G'(x^k)d^k\rangle \to 0 \quad (k \to \infty),$$

что и завершает доказательство. □

Задача 3. Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображения $F\colon \mathbf{R}^n \to \mathbf{R}^l$ и $G\colon \mathbf{R}^n \to \mathbf{R}^m$ дифференцируемы в точке $x^k \in \mathbf{R}^n$. Пусть $H_k \in \mathbf{R}(n, n)$ — симметрическая матрица, $(\sigma_k, d^k) \in \mathbf{R} \times \mathbf{R}^n$ — стационарная точка задачи

$$c_k\sigma + \langle f'(x^k), d\rangle + \frac{1}{2}\langle H_k d, d\rangle \to \min, \quad (\sigma, d) \in U_k, \tag{23}$$

$$U_k = \{u = (\sigma, d) \in \mathbf{R}_+ \times \mathbf{R}^n \mid |F(x^k) + F'(x^k)d|_\infty \leqslant \sigma,$$
$$g_i(x^k) + \langle g_i'(x^k), d\rangle \leqslant \sigma,\ i = 1, \ldots, m\}, \tag{24}$$

где $c_k > 0$. Пусть функция φ_{c_k} введена в соответствии с (3), где

$$\psi(x) = \max\{|F(x)|_\infty, 0, g_1(x), \ldots, g_m(x)\}, \quad x \in \mathbf{R}^n. \tag{25}$$

Показать, что если $d^k \neq 0$, то $d^k \in \mathcal{D}_{\varphi_{c_k}}(x^k)$.

На основе задачи (23), (24) и функции φ_{c_k}, вводимой в соответствии с (3), (25), можно построить аналог алгоритма 1 [1]. Читатель может разработать этот аналог и исследовать его сходимость самостоятельно либо познакомиться с ним, например, в [6, 41]. Преимуществом такого подхода является то обстоятельство, что допустимое множество задачи (23), (24) всегда непусто. Недостаток же состоит в том, что для каждого k подходящее (достаточно большое) значение параметра штрафа c_k должно быть указано до отыскания стационарной точки задачи (23), (24) и отвечающих ей множителей Лагранжа. Однако с практической точки зрения этот недостаток не слишком серьезен, поскольку подходящее значение c_k обычно удается найти с помощью известных эвристических приемов.

5.4.2. Восстановление сверхлинейной скорости сходимости. Как показывает теорема 1, в разумных предположениях сходимость методов SQP может быть успешно глобализована. Возникает естественный вопрос: сохранится ли при такой модификации присущая методам SQP высокая скорость локальной сходимости? Согласно теореме 4.4.2 сверхлинейной скорости сходимости метода к решению $(\overline{x}, \overline{\lambda}, \overline{\mu}) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ системы Каруша–Куна–Таккера задачи (1), (2) (при наличии сходимости) можно ожидать, если

[1] Такую версию алгоритма SQP по-английски называют Elastic mode.

матрица H_k «аппроксимирует» матрицу $\dfrac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})$ при $k \to \infty$ в смысле соотношения (4.4.29). (Подчеркнем, что такая «аппроксимация» вовсе не означает, что обязательно $\{H_k\} \to \dfrac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})$ $(k \to \infty)$. Более того, выполнение этого предельного соотношения может быть нежелательным, поскольку матрица $\dfrac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}, \overline{\mu})$ может не быть положительно определенной.) Однако этого мало: нужно еще, чтобы при больших k в п. 2 алгоритма 1 принимался предлагаемый в качестве начального значения для дробления единичный параметр длины шага. Будет ли это так, если текущая точка $(x^k, \lambda^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ близка к $(\overline{x}, \overline{\lambda}, \overline{\mu})$ и в точке $(\overline{x}, \overline{\lambda}, \overline{\mu})$ выполнены все условия сверхлинейной локальной сходимости методов SQP, сформулированные в теореме 4.4.2? К сожалению, ответ на этот вопрос отрицателен, как показывает следующий пример, иллюстрирующий так называемый *эффект Маратоса*. (Подчеркнем, что этот пример ни в коей мере не является патологическим.)

Пример 1. Пусть $n = 2$, $l = 1$, $m = 0$ и

$$f(x) = x_1, \quad F(x) = x_1^2 + x_2^2 - 1, \quad x \in \mathbf{R}^2.$$

Глобальным решением задачи (1), (2) является точка $\overline{x} = (-1, 0)$, ей отвечает множитель Лагранжа $\overline{\lambda} = 1/2$, причем в точке $\overline{x}$ выполнено как условие регулярности ограничений, так и достаточное условие второго порядка оптимальности. Для произвольной точки $x^k \in \mathbf{R}^2$ пусть $(d^k, y^k) \in \mathbf{R}^2 \times \mathbf{R}$ — решение системы Лагранжа вспомогательной задачи (5), (6) с «идеальным» с точки зрения скорости сходимости выбором $H_k = \dfrac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}) = E^2$, т. е.

$$1 + d_1^k + 2y^k x_1^k = 0, \quad d_2^k + 2y^k x_2^k = 0, \tag{26}$$

$$(x_1^k)^2 + (x_2^k)^2 - 1 + 2x_1^k d_1^k + 2x_2^k d_2^k = 0 \tag{27}$$

(см. (11), (12)). Подчеркнем, что матрица H_k положительно определена, и при $x^k \neq 0$ соотношениями (26), (27) пара (d^k, y^k) определена однозначно и может быть указана явно, однако явное выражение не потребуется. Потребуются лишь следующие свойства. Во-первых, домножив левую и правую части первого равенства в (26) на d_1^k, а второго равенства на d_2^k, сложив соответствующие части этих равенств и использовав (27), получим

$$d_1^k = y^k((x_1^k)^2 + (x_2^k)^2 - 1) - (d_1^k)^2 - (d_2^k)^2,$$

поэтому

$$f(x^k + d^k) - f(x^k) = d_1^k = y^k F(x^k) - |d^k|^2.$$

Во-вторых, из (27) следует равенство

$$F(x^k + d^k) = (x_1^k + d_1^k)^2 + (x_2^k + d_2^k)^2 - 1 = |d^k|^2.$$

Используя два последних равенства, имеем

$$\varphi_c(x^k + d^k) - \varphi_c(x^k) = y^k F(x^k) - c|F(x^k)| + (c-1)|d^k|^2 > 0$$

для любого $c > 1$, если, например, $F(x^k) = 0$ и точка x^k не является стационарной в задаче (1), (2) (последнее гарантирует, что $d^k \neq 0$; см. начало доказательства теоремы 1). При этом $\alpha = 1$ не может удовлетворять неравенству (8) в силу выполнения (15). Но в любой окрестности точки $\overline{x}$ имеется сколько угодно допустимых точек, не являющихся стационарными в задаче (1), (2).

Заметим, что в приведенном примере для $c \in (1/2, 1)$ единичный параметр длины шага приводит к уменьшению значения функции φ_c и противоречия между требованиями, гарантирующими глобальную сходимость, и требованиями, гарантирующими сверхлинейную скорость локальной сходимости, для таких значений параметра штрафа может и не быть. Однако этот пример несложно модифицировать так, чтобы «подходящие» значения параметра штрафа перестали существовать.

Пример 2. Пусть все определено так же, как в примере 1, но

$$f(x) = x_1 + x_1^2 + x_2^2, \quad x \in \mathbf{R}^2.$$

Глобальным решением задачи (1), (2) по-прежнему является точка $\overline{x} = (-1, 0)$, ей отвечает множитель Лагранжа $\overline{\lambda} = -1/2$, причем в точке $\overline{x}$ по-прежнему выполнено как условие регулярности ограничений, так и достаточное условие второго порядка оптимальности. Рассмотрим произвольную точку $x^k \in \mathbf{R}^2$ такую, что $F(x^k) = 0$ и x^k не является стационарной точкой задачи (1), (2). Пусть $(d^k, y^k) \in \mathbf{R}^2 \times \mathbf{R}$ — решение системы Лагранжа вспомогательной задачи (5), (6) при $H_k = \dfrac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}) = E^2$, т. е.

$$1 + 2x_1^k + d_1^k + 2y^k x_1^k = 0, \quad 2x_2^k + d_2^k + 2y^k x_2^k = 0, \tag{28}$$

$$x_1^k d_1^k + x_2^k d_2^k = 0 \tag{29}$$

(см. (11), (12)). Домножив левую и правую части первого равенства в (28) на d_1^k, а второго на d_2^k, сложив соответствующие части этих равенств и использовав (29), получим

$$d_1^k + (d_1^k)^2 + (d_2^k)^2 = 0,$$

поэтому

$$f(x^k + d^k) - f(x^k) = d_1^k + (x_1^k + d_1^k)^2 + (x_2^k + d_2^k)^2 - (x_1^k)^2 - (x_2^k)^2 = 0,$$

$$F(x^k + d^k) = (x_1^k + d_1^k)^2 + (x_2^k + d_2^k)^2 - 1 = |d^k|^2$$

(здесь вновь учтено (29)). Из двух последних равенств имеем

$$\varphi_c(x^k + d^k) - \varphi_c(x^k) = c|d^k|^2 > 0 \quad \forall\, c > 0.$$

Эффект Маратоса можно на качественном уровне объяснить так. Если текущая точка x^k оказывается близкой к допустимому множеству D, то при переходе от x^k к $x^k + d^k$ значение негладкого штрафа может вырасти на величину того же порядка, что и величина уменьшения значения целевой функции f. В результате при достаточно большом c происходит увеличение значения штрафной функции φ_c, и единичный параметр длины шага не принимается. В этом смысле ситуация с сохранением сверхлинейной скорости локальной сходимости методов SQP при глобализации их сходимости за счет одномерного поиска существенно сложнее, чем, например, для методов ньютоновского типа применительно к задачам безусловной оптимизации (см. теорему 3.2.2 и § 5.1). Чтобы избежать возможности возникновения эффекта Маратоса, алгоритм 1 следует модифицировать. Известны несколько способов преодоления эффекта Маратоса, и ниже излагаются два из них. Первый основан на модификации используемой функции качества в духе модифицированных функций Лагранжа, а второй (наиболее часто применяемый в настоящее время) — на идее криволинейного одномерного поиска «вдоль границы» допустимого множества.

Будем рассматривать задачу с чистыми ограничениями-равенствами: будем полагать, что вместо (2) множество D задается формулой

$$D = \{x \in \mathbf{R}^n \,\big|\, F(x) = 0\}. \tag{30}$$

Введем семейство функций

$$\varphi_{c,\,\eta}\colon \mathbf{R}^n \to \mathbf{R}, \quad \varphi_{c,\,\eta}(x) = f(x) + \langle \eta,\, F(x)\rangle + c|F(x)|_1, \tag{31}$$

где $c \geqslant 0$ и $\eta \in \mathbf{R}^l$ — параметры. Имеет место следующий факт (ср. со следствием 4.7.1).

Предложение 1. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ и отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ дважды дифференцируемы в точке $\overline{x} \in \mathbf{R}^n$. Пусть $\overline{x}$ — стационарная точка задачи* (1), (30) *и в ней выполнено сформулированное в теореме* 1.3.7 *достаточное условие второго порядка оптимальности с множителем Лагранжа $\overline{\lambda} \in \mathbf{R}^l$.*

Тогда существует число $\overline{c} \geqslant 0$ такое, что для любых $c > \overline{c}$ и $\eta \in \mathbf{R}^l$, удовлетворяющих неравенству

$$c > |\eta - \overline{\lambda}|_\infty, \tag{32}$$

точка $\overline{x}$ является строгим локальным решением задачи

$$\varphi_{c,\,\eta}(x) \to \min, \quad x \in \mathbf{R}^n,$$

с целевой функцией, задаваемой формулой (31).

Доказательство. Вспоминая вид модифицированной функции Лагранжа задачи (1), (30) и используя утверждение из задачи 4.3.8, для любого достаточно большого $c > 0$ имеем

$$\varphi_{c,\eta}(\overline{x}) = f(\overline{x}) = L(\overline{x}, \overline{\lambda}) + \frac{c}{2}|F(\overline{x})|^2 < L(x, \overline{\lambda}) + \frac{c}{2}|F(x)|^2 \qquad (33)$$

для любого $x \in \mathbf{R}^n \setminus \{\overline{x}\}$, достаточно близкого к $\overline{x}$. С другой стороны, для таких x

$$\varphi_{c,\eta}(x) \geqslant L(x, \overline{\lambda}) + (c - |\eta - \overline{\lambda}|_\infty)|F(x)|_1 \geqslant L(x, \overline{\lambda}) + \frac{c}{2}|F(x)|^2, \qquad (34)$$

где принято во внимание неравенство (32). Объединяя (33) и (34), получаем требуемое. □

Для текущего приближения $x^k \in \mathbf{R}^n$ положим

$$D_k = \{d \in \mathbf{R}^n \mid F(x^k) + F'(x^k)d = 0\}. \qquad (35)$$

Пусть $d^k \in \mathbf{R}^n$ — стационарная точка задачи квадратичного программирования (5), (35), а $y^k \in \mathbf{R}^l$ — отвечающий ей множитель Лагранжа. Таким образом,

$$f'(x^k) + H_k d^k + (F'(x^k))^{\mathrm{T}} y^k = 0, \quad F(x^k) + F'(x^k)d^k = 0. \qquad (36)$$

Используя результат задачи 2 и соотношение (36) и рассуждая так же, как при доказательстве леммы 1, легко убедиться, что при любых c и η

$$\varphi'_{c,\eta}(x^k; d^k) = -\langle H_k d^k, d^k\rangle + \langle y^k - \eta, F(x^k)\rangle - c|F(x^k)|_1, \qquad (37)$$

откуда, в частности, следует, что если $c \geqslant |y^k - \eta|_\infty$, то

$$\varphi'_{c,\eta}(x^k; d^k) \leqslant -\langle H_k d^k, d^k\rangle. \qquad (38)$$

Поэтому если матрица H_k положительно определена, а $d^k \neq 0$, то $d^k \in \mathcal{D}_{\varphi_{c,\eta}}(x^k)$. Теперь ясно, что можно построить аналог алгоритма 1, использующий новую штрафную функцию $\varphi_{c,\eta}$. Условие (7) следует заменить условием

$$c_k \geqslant |y^k - \eta^k|_\infty, \qquad (39)$$

а неравенство (8) — неравенством

$$\varphi_{c_k,\eta^k}(x^k + \alpha d^k) \leqslant \varphi_{c_k,\eta^k}(x^k) + \varepsilon\alpha\varphi'_{c_k,\eta^k}(x^k; d^k). \qquad (40)$$

Следующая теорема в совокупности с теоремой 4.4.2 показывает, что при соответствующем выборе матриц H_k модифицированный указанным способом алгоритм будет обладать не только глобальной сходимостью, но и сверхлинейной скоростью сходимости.

Теорема 2. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображение* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *дважды дифференцируемы в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$, *причем их вторые производные непрерывны в этой*

точке. Пусть в точке $\overline{x}$ *выполнено условие регулярности ограничений, причем* $\overline{x}$ *— стационарная точка задачи* (1), (30), *а* $\overline{\lambda} \in \mathbf{R}^l$ *— однозначно отвечающий ей множитель Лагранжа. Пусть в точке* $\overline{x}$ *выполнено сформулированное в теореме* 1.3.7 *достаточное условие второго порядка оптимальности. Пусть, кроме того, последовательность* $\{x^k\} \subset \mathbf{R}^n$ *сходится к* $\overline{x}$ *и для любого* k *симметрическая матрица* $H_k \in \mathbf{R}(n, n)$ *удовлетворяет условию*

$$\left\langle \left(H_k - \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}) - c(F'(\overline{x}))^{\mathrm{T}} F'(\overline{x}) \right) d^k, d^k \right\rangle \geqslant o(|d^k|^2), \qquad (41)$$

где число $c \geqslant 0$ *таково, что матрица* $\dfrac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}) + c(F'(\overline{x}))^{\mathrm{T}} F'(\overline{x})$ *положительно определена, пара* $(d^k, y^k) \in \mathbf{R}^n \times \mathbf{R}^l$, *число* c_k *и вектор* $\eta^k \in \mathbf{R}^l$ *удовлетворяют* (36), (39), *причем* $\{d^k\} \to 0$ $(k \to \infty)$.

Тогда для любого числа $\varepsilon \in (0, 1/2)$ *найдется число* $\delta > 0$ *такое, что для любого достаточно большого* k *из выполнения неравенств*

$$c_k \leqslant \delta, \quad |\eta^k - \overline{\lambda}| \leqslant \delta \qquad (42)$$

следует выполнение неравенства (40) *при* $\alpha = 1$ *для вводимой в соответствии с* (31) *функции* φ_{c_k, η^k}.

В условиях теоремы для любого достаточно большого k функция φ_{c_k, η^k} дифференцируема в точке x^k по любому направлению, и поэтому неравенство (40) имеет смысл.

Напомним, что при выполнении в точке $\overline{x}$ достаточного условия второго порядка оптимальности матрица $\dfrac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}) + c(F'(\overline{x}))^{\mathrm{T}} F'(\overline{x})$ положительно определена при любом достаточно большом c (см. задачу 4.3.8). Условие (41) на выбор H_k по существу означает, что H_k выбирается как «оценка сверху» для матрицы, вводимой в соответствии с задачей 4.4.1 при достаточно большом c_k.

Доказательство теоремы 2. Из (41) и выбора числа c немедленно следует, что существует число $\gamma > 0$ такое, что для любого достаточно большого k

$$\langle H_k d^k, d^k \rangle \geqslant \gamma |d^k|^2. \qquad (43)$$

Кроме того, для любого k из (41) и неотрицательной определенности матрицы $(F'(\overline{x}))^{\mathrm{T}} F'(\overline{x})$ следует, что

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}) d^k, d^k \right\rangle \leqslant \langle H_k d^k, d^k \rangle + o(|d^k|^2). \qquad (44)$$

Из (36) имеем

$$\langle f'(x^k), d^k \rangle = -\langle H_k d^k, d^k \rangle - \langle y^k, F'(x^k) d^k \rangle = \\ = -\langle H_k d^k, d^k \rangle + \langle y^k, F(x^k) \rangle.$$

Отсюда в силу классической теоремы о среднем (для скалярнозначных функций) и сходимости $\{x^k\}$ к $\overline{x}$, а $\{d^k\}$ к 0 выводим

$$f(x^k+d^k)=f(x^k)+\langle f'(x^k),\,d^k\rangle+\frac{1}{2}\,\langle f''(x^k+\widetilde{t}_k d^k)d^k,\,d^k\rangle=$$
$$=f(x^k)-\langle H_k d^k,\,d^k\rangle+\langle y^k,\,F(x^k)\rangle+\frac{1}{2}\,\langle f''(\overline{x})d^k,\,d^k\rangle+o(|d^k|^2),\tag{45}$$

где $\widetilde{t}_k\in[0,\,1]$. Аналогично, но с учетом второго равенства в (36) и теоремы о среднем (для отображений), получаем

$$\left|F(x^k+d^k)-\frac{1}{2}\,F''(\overline{x})[d^k,\,d^k]\right|=$$
$$=\left|F(x^k+d^k)-F(x^k)-F'(x^k)d^k-\frac{1}{2}\,F''(\overline{x})[d^k,\,d^k]\right|\leqslant$$
$$\leqslant\sup_{t\in[0,\,1]}|F'(x^k+td^k)-F'(x^k)-F''(\overline{x})[d^k]|\,|d^k|\leqslant$$
$$\leqslant\sup_{t_1\in[0,\,1]}\sup_{t_2\in[0,\,1]}|F''(x^k+t_1t_2d^k)-F''(\overline{x})|\,|d^k|^2=o(|d^k|^2).\tag{46}$$

Следующая выкладка использует (31), (37), (38), (42)–(46):

$$\varphi_{c_k,\,\eta^k}(x^k+d^k)-\varphi_{c_k,\,\eta^k}(x^k)=$$
$$=-\langle H_k d^k,\,d^k\rangle+\frac{1}{2}\,\langle f''(\overline{x})d^k,\,d^k\rangle+$$
$$+\langle y^k-\eta^k,\,F(x^k)\rangle-c_k|F(x^k)|_1+$$
$$+\frac{1}{2}\,\langle\eta^k,\,F''(\overline{x})[d^k,\,d^k]\rangle+O(c_k|d^k|^2)+o(|d^k|^2)=$$
$$=\varphi'_{c_k,\,\eta^k}(x^k;\,d^k)+\frac{1}{2}\left\langle\frac{\partial^2L}{\partial x^2}(\overline{x},\,\overline{\lambda})d^k,\,d^k\right\rangle+$$
$$+O((c_k+|\eta^k-\overline{\lambda}|)|d^k|^2)+o(|d^k|^2)\leqslant$$
$$\leqslant\varphi'_{c_k,\,\eta^k}(x^k;\,d^k)+\frac{1}{2}\,\langle H_k d^k,\,d^k\rangle+O(\delta|d^k|^2)\leqslant$$
$$\leqslant\varepsilon\varphi'_{c_k,\,\eta^k}(x^k;\,d^k)-\frac{1-\varepsilon-1/2}{2}\,\langle H_k d^k,\,d^k\rangle+O(\delta|d^k|^2)\leqslant$$
$$\leqslant\varepsilon\varphi'_{c_k,\,\eta^k}(x^k;\,d^k)-\left(\frac{1}{2}-\varepsilon\right)\gamma|d^k|^2+O(\delta|d^k|^2)\leqslant$$
$$\leqslant\varepsilon\varphi'_{c_k,\,\eta^k}(x^k;\,d^k)$$

для любого достаточно большого k, если число $\delta>0$ достаточно мало́. $\square$

Практически реализуемые процедуры выбора параметров c_k и η^k, обеспечивающие как глобальную сходимость методов SQP, модифицированных указанным способом, так и сверхлинейную скорость локальной сходимости, известны и могут быть найдены в специальной

литературе. Как построение, так и обоснование таких процедур далеко не очевидно, весьма громоздко и здесь не обсуждается.

Обратимся к другому способу преодоления эффекта Маратоса, основанному на так называемых *поправках второго порядка*. Эти поправки имеют целью приближение следующей точки к допустимому множеству задачи и, тем самым, к уменьшению значения штрафа. Направление поправки $\widetilde{d}^k$ вычисляется как решение линейного уравнения

$$F(x^k + d^k) + F'(x^k)\widetilde{d} = 0,$$

обладающее минимальной нормой. Иными словами, $\widetilde{d}^k$ является (глобальным) решением задачи

$$|\widetilde{d}| \to \min, \quad \widetilde{d} \in \widetilde{D}_k, \tag{47}$$

$$\widetilde{D}_k = \{\widetilde{d} \in \mathbf{R}^n \mid F(x^k + d^k) + F'(x^k)\widetilde{d} = 0\}. \tag{48}$$

Задача 4. Пусть отображение $F\colon \mathbf{R}^n \to \mathbf{R}^l$ дифференцируемо в точке $x^k \in \mathbf{R}^n$, причем

$$\operatorname{rank} F'(x^k) = l.$$

Показать, что для всякого $d^k \in \mathbf{R}^n$ единственным решением задачи (47), (48) является точка

$$\widetilde{d}^k = -(F'(x^k))^{\mathrm{T}}(F'(x^k)(F'(x^k))^{\mathrm{T}})^{-1}F(x^k + d^k). \tag{49}$$

Заметим, что определение $\widetilde{d}^k$ требует лишь одного дополнительного вычисления значения отображения F (в точке $x^k + d^k$). При этом, если предполагать, что траектория $\{x^k\}$ сходится к точке $\overline{x}$, в некоторой окрестности которой отображение F имеет ограниченную вторую производную, и что $\{d^k\} \to 0$ $(k \to \infty)$, то с учетом (36) имеем

$$F(x^k + d^k) = F(x^k) + F'(x^k)d^k + O(|d^k|^2) = O(|d^k|^2). \tag{50}$$

Тогда если $\operatorname{rank} F'(\overline{x}) = l$, то из (49) вытекает оценка

$$\widetilde{d}^k = O(|F(x^k + d^k)|) = O(|d^k|^2) \tag{51}$$

(отсюда название «поправка второго порядка»). Из (51) следует, что если шаг в направлении d^k приводит к сверхлинейному убыванию расстояния до решения, т. е. $x^k + d^k - \overline{x} = o(|x^k - \overline{x}|)$, то

$$x^k + d^k + \widetilde{d}^k - \overline{x} = x^k + d^k - \overline{x} + O(|d^k|^2) = o(|x^k - \overline{x}|),$$

т. е. шаг в точку $x^k + d^k + \widetilde{d}^k$ сохраняет сверхлинейное убывание расстояния до решения. С другой стороны, из (48) и (51) вытекает оценка

$$F(x^k + d^k + \widetilde{d}^k) = F(x^k + d^k) + F'(x^k + d^k)\widetilde{d}^k + o(|\widetilde{d}^k|) =$$

$$= F(x^k + d^k) + F'(x^k)\widetilde{d}^k + o(|\widetilde{d}^k|) = o(|d^k|^2), \tag{52}$$

т. е. шаг в точку $x^k + d^k + \widetilde{d}^k$ приводит к большему по порядку убыванию невязки ограничений, чем шаг в точку $x^k + d^k$. Это позволяет рассчитывать на то, что вблизи решения шаг в точку $x^k + d^k + \widetilde{d}^k$ будет обеспечивать достаточное убывание значения функции φ_{c_k}.

Для обеспечения разумного глобального поведения метода предлагается искать следующее приближение x^{k+1} в виде $x^k + \alpha d^k + \alpha^2\widetilde{d}^k$, $\alpha \geqslant 0$, осуществляя одномерный поиск вдоль указанной дуги с начальным пробным значением $\alpha = 1$, которое дробится (последовательно умножается на $\theta \in (0, 1)$) до тех пор, пока не выполнится неравенство

$$\varphi_{c_k}(x^k + \alpha d^k + \alpha^2\widetilde{d}^k) \leqslant \varphi_{c_k}(x^k) + \varepsilon\alpha\Delta_k, \tag{53}$$

где

$$\varphi_c\colon \mathbf{R}^n \to \mathbf{R}, \quad \varphi_c(x) = f(x) + c|F(x)|_1. \tag{54}$$

Если начальное пробное значение принимается, то переход осуществляется в точку $x^k + d^k + \widetilde{d}^k$. С другой стороны, из (10) вытекает, что в условиях леммы 1 неравенство (53) выполняется для всех достаточно малых $\alpha > 0$.

Приведем результат, который вместе с теоремой 4.4.2 дает условия, гарантирующие сверхлинейную скорость сходимости описанной модификации алгоритма.

Теорема 3. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и отображение* $F\colon \mathbf{R}^n \to \mathbf{R}^l$ *дважды дифференцируемы в некоторой окрестности точки* $\overline{x} \in \mathbf{R}^n$, *причем их вторые производные непрерывны в этой точке. Пусть в точке* $\overline{x}$ *выполнено условие регулярности ограничений, причем* $\overline{x}$ *— стационарная точка задачи* (1), (30), *а* $\overline{\lambda} \in \mathbf{R}^l$ *— однозначно отвечающий ей множитель Лагранжа. Пусть в точке* $\overline{x}$ *выполнено сформулированное в теореме* 1.3.7 *достаточное условие второго порядка оптимальности. Пусть, кроме того, последовательность* $\{x^k\} \subset \mathbf{R}^n$ *сходится к* $\overline{x}$, *и для любого* k *симметрическая матрица* $H_k \in \mathbf{R}(n, n)$ *удовлетворяет условию* (41), *где число* $c \geqslant 0$ *таково, что матрица* $\frac{\partial^2 L}{\partial x^2}(\overline{x}, \overline{\lambda}) + c(F'(\overline{x}))^{\mathrm{T}}F'(\overline{x})$ *положительно определена, пара* $(d^k, y^k) \in \mathbf{R}^n \times \mathbf{R}^l$ *и число* c_k *удовлетворяют* (36) *и неравенству*

$$c_k \geqslant |y^k|_\infty, \tag{55}$$

причем $\{d^k\} \to 0$ $(k \to \infty)$, *а последовательность* $\{c_k\}$ *ограничена.*

Тогда для любого числа $\varepsilon \in (0, 1/2)$ *и любого достаточно большого* k *неравенство* (53) *выполняется при* $\alpha = 1$ *для вводимой в соответствии с* (54) *функции* φ_{c_k}.

Доказательство. Как и при доказательстве теоремы 2, из (41) и выбора числа c следует существование такого числа $\gamma > 0$,

что для любого достаточно большого k имеют место соотношения (43), (44). Используя (9), (36), (41), (43), (44), (49)–(52), (55), ограниченность $\{c_k\}$ и сходимость $\{x^k\}$ к $\overline{x}$, выводим

$$\varphi_{c_k}(x^k + d^k + \widetilde{d}^k) - \varphi_{c_k}(x^k) - \varepsilon\Delta_k =$$

$$= f(x^k + d^k + \widetilde{d}^k) - f(x^k) + c_k|F(x^k + d^k + \widetilde{d}^k)|_1 - c_k|F(x^k)|_1 - \varepsilon\Delta_k =$$

$$= \langle f'(x^k),\, d^k + \widetilde{d}^k\rangle + \frac{1}{2}\langle f''(x^k)d^k,\, d^k\rangle - c_k|F(x^k)|_1 - \varepsilon\Delta_k + o(|d^k|^2) =$$

$$= \langle f'(x^k),\, \widetilde{d}^k\rangle + \frac{1}{2}\langle f''(x^k)d^k,\, d^k\rangle + (1-\varepsilon)\Delta_k + o(|d^k|^2) =$$

$$= -\langle f'(x^k),\, (F'(x^k))^{\mathrm{T}}(F'(x^k)(F'(x^k))^{\mathrm{T}})^{-1}F(x^k + d^k)\rangle +$$

$$+ \frac{1}{2}\langle f''(x^k)d^k,\, d^k\rangle + (1-\varepsilon)\Delta_k + o(|d^k|^2) =$$

$$= -\langle (F'(x^k)(F'(x^k))^{\mathrm{T}})^{-1}F'(x^k)f'(x^k),\, F(x^k + d^k)\rangle +$$

$$+ \frac{1}{2}\langle f''(x^k)d^k,\, d^k\rangle + (1-\varepsilon)\Delta_k + o(|d^k|^2) =$$

$$= \langle (F'(x^k)(F'(x^k))^{\mathrm{T}})^{-1}F'(x^k)(F'(x^k))^{\mathrm{T}}\overline{\lambda},\, F(x^k + d^k)\rangle -$$

$$- \left\langle (F'(x^k)(F'(x^k))^{\mathrm{T}})^{-1}F'(x^k)\frac{\partial L}{\partial x}(x^k,\, \overline{\lambda}),\, F(x^k + d^k)\right\rangle +$$

$$+ \frac{1}{2}\langle f''(x^k)d^k,\, d^k\rangle + (1-\varepsilon)\Delta_k + o(|d^k|^2) =$$

$$= \langle \overline{\lambda},\, F(x^k + d^k)\rangle + \frac{1}{2}\langle f''(x^k)d^k,\, d^k\rangle + (1-\varepsilon)\Delta_k + o(|d^k|^2) =$$

$$= \frac{1}{2}\langle f''(x^k)d^k,\, d^k\rangle + \left\langle \overline{\lambda},\, F(x^k) + F'(x^k)d^k + \frac{1}{2}F''(x^k)[d^k,\, d^k]\right\rangle +$$

$$+ (1-\varepsilon)\Delta_k + o(|d^k|^2) =$$

$$= \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x^k,\, \overline{\lambda})d^k,\, d^k\right\rangle + (1-\varepsilon)\Delta_k + o(|d^k|^2) =$$

$$= (1-\varepsilon)(\langle f'(x^k),\, d^k\rangle - c_k|F(x^k)|_1) + \frac{1}{2}\left\langle \frac{\partial L}{\partial x}(\overline{x},\, \overline{\lambda})d^k,\, d^k\right\rangle + o(|d^k|^2) =$$

$$= (1-\varepsilon)(-\langle H_k d^k,\, d^k\rangle - \langle y^k,\, F'(x^k)d^k\rangle - c_k|F(x^k)|_1) +$$

$$+ \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x},\, \overline{\lambda})d^k,\, d^k\right\rangle + o(|d^k|^2) =$$

$$= -\left(\frac{1}{2} - \varepsilon\right)\langle H_k d^k,\, d^k\rangle + (1-\varepsilon)(\langle y^k,\, F(x^k)\rangle - c_k|F(x^k)|_1) -$$

$$- \frac{1}{2}\left(\langle H_k d^k,\, d^k\rangle - \left\langle \frac{\partial^2 L}{\partial x^2}(\overline{x},\, \overline{\lambda})d^k,\, d^k \right\rangle\right) + o(|d^k|^2) \leqslant$$

$$\leqslant -\left(\frac{1}{2} - \varepsilon\right)\langle H_k d^k,\, d^k\rangle + (1-\varepsilon)(|y^k|_\infty - c_k)|F(x^k)|_1 + o(|d^k|^2) \leqslant$$

$$\leqslant -\left(\frac{1}{2} - \varepsilon\right)\langle H_k d^k,\, d^k\rangle + o(|d^k|^2) \leqslant -\left(\frac{1}{2} - \varepsilon\right)\gamma|d^k|^2 + o(|d^k|^2) < 0$$

для любого достаточно большого k. $\square$

В завершение этого параграфа очень коротко остановимся на других стратегиях глобализации сходимости методов SQP. Пусть $x^k \in \in \mathbf{R}^n$ — текущее приближение. Тогда итерационная вспомогательная задача метода доверительной области на базе SQP будет иметь вид (5), где множество D_k задается формулой

$$D_k = \{d \in \overline{B}(0,\, \delta_k) \mid\; F(x^k) + F'(x^k)d = 0,\; G(x^k) + G'(x^k)d \leqslant 0\} \quad (56)$$

(ср. с (6)), а $\delta_k > 0$ — радиус доверительной области. Пусть d^k — глобальное решение такой задачи. Точка $\widetilde{x}^k = x^k + d^k$ принимается в качестве очередного приближения x^{k+1} в том случае, когда переход от x^k к $\widetilde{x}^k$ обеспечивает достаточное относительное убывание значения функции качества, в роли которой обычно выступает некоторая штрафная функция. В противном случае задача (5), (56) решается вновь при меньшем значении δ_k.

Один из недостатков указанного подхода, как и рассмотренного выше подхода с одномерным поиском (и вообще любого подхода, использующего некоторую функцию качества для отбраковки слишком длинных шагов вдали от решения), связан с тем, что, как отмечалось выше, практический выбор конкретных значений параметра штрафа в функции качества часто оказывается весьма трудным делом. Нередко этот выбор вообще оставляется пользователю, которого в лучшем случае снабжают эвристическими рекомендациями, основанными на вычислительном опыте авторов. Совсем недавно был предложен следующий (довольно неожиданный, хотя и вполне естественный) способ, позволяющий избавиться от указанного недостатка. Идея этого способа состоит в отказе от использования единой функции качества и может применяться не только к методам SQP.

Будем трактовать невязку ограничений (например, введенный формулой (4) штраф ψ) как дополнительный критерий, который, как и целевую функцию f, желательно минимизировать. Более того, в определенном смысле минимизацию этого дополнительного критерия можно считать приоритетной целью: необходимо добиться равенства его значения минимально возможному значению (нулю). Штрафная функция выступает в роли агрегированного критерия, в котором тем большее значение отводится критерию ψ, чем больше значение параметра штрафа. С другой стороны, исходную

задачу (1), (2) можно трактовать в духе задачи двукритериальной оптимизации с критериями f и ψ на всем $\mathbf{R}^n$. Для принятия или отклонения пробного приближения $\widetilde{x}^k$ теперь используется не функция качества, а известный в многокритериальной оптимизации принцип доминирования стратегий. После k итераций предполагается известным так называемый *фильтр* $\mathcal{F}_k$, т. е. конечное множество в $\mathbf{R} \times \mathbf{R}$ такое, что ни одна пара чисел $(a_1,\, b_1) \in \mathcal{F}_k$ не доминируется никакой другой парой $(a_2,\, b_2) \in \mathcal{F}_k$ в том смысле, что неравенства $a_1 \geqslant a_2,\ b_1 \geqslant b_2$ одновременно выполняться не могут. Пробная точка $\widetilde{x}^k$ принимается в качестве очередного приближения x^{k+1} в том случае, если пара $(f(\widetilde{x}^k),\, \psi(\widetilde{x}^k))$ не доминируется ни одной парой из $\mathcal{F}_k$. При этом новый фильтр $\mathcal{F}_{k+1}$ генерируется так: пара $(f(\widetilde{x}^k),\, \psi(\widetilde{x}^k))$ включается в фильтр, а все доминируемые ею пары из фильтра удаляются. В противном случае уменьшается либо параметр α_k длины шага, либо радиус δ_k доверительной области.

Разумеется, сказанное отражает лишь основную идею данного подхода. Например, каждая итерация на самом деле обычно состоит из двух фаз: *нормальной фазы*, имеющей целью приблизить текущую точку к допустимому множеству, т. е. уменьшить значение ψ, и *касательной фазы*, направленной на уменьшение значения f. Ряд необходимых усовершенствований позволяет обеспечить как глобальную сходимость (в некотором смысле) соответствующего алгоритма, так и сверхлинейную скорость локальной сходимости. Высокая эффективность этого нового подхода подтверждается опубликованными результатами вычислительных экспериментов, которые весьма многообещающи.

Глава 6
МЕТОДЫ НЕГЛАДКОЙ ВЫПУКЛОЙ ОПТИМИЗАЦИИ

Эта глава начинается с изложения общей концепции двойственности для (возможно, невыпуклых) задач оптимизации, естественным образом приводящей к негладким выпуклым задачам. Такой порядок изложения избран по нескольким причинам. Например, он согласуется с общей «невыпуклой идеологией» настоящей книги в том смысле, что изначальная задача, которую предполагается решить, не обязательно выпукла. Но более важно то, что двойственная релаксация [1] действительно является весьма полезным средством в оптимизации, распространяемым даже на такие области, как целочисленное программирование, и авторы считают необходимым дать этому кругу вопросов хотя бы некоторое освещение.

Что же касается собственно алгоритмов, сначала рассматриваются базовые субградиентные методы. Помимо прочего это сделано в силу исторических причин. Кроме того, приближенные версии субградиентных методов дают удобную схему для анализа более совершенных алгоритмов. Следует подчеркнуть, что субградиентные методы на практике обычно крайне неэффективны. Они по-прежнему используются, но главным образом тогда, когда достаточно найти очень грубое приближение к решению. В тех случаях, когда отыскание точного приближения к решению негладкой задачи составляет суть проблемы, субградиентные методы в настоящее время используются только «любителями» (скажем, прикладниками, которые просто не знают о существовании лучших методов). Именно появление современных многошаговых алгоритмов решения негладких задач сделало негладкую выпуклую оптимизацию по-настоящему плодотворной областью с вычислительной точки зрения. Наибольшее значение имеют методы такого рода, использующие на каждом шаге квадратичную вспомогательную задачу, конструируемую по накопленной к этому шагу информации. В некотором (очень ограниченном) понимании такие

[1] Общепринятый английский термин — Lagrangian relaxation.

алгоритмы могут рассматриваться как результат регуляризации методов кусочно линейной аппроксимации, занимающих промежуточное место между субградиентными методами и многошаговыми алгоритмами с квадратичными подзадачами.

§ 6.1. Элементы выпуклого анализа и двойственные методы

6.1.1. Элементы субдифференциального исчисления. Настоящий пункт написан для того, чтобы не загромождать текст этой главы многочисленными сносками: здесь содержится краткое (без доказательств) изложение некоторых сведений из выпуклого анализа, необходимых для работы с выпуклыми задачами оптимизации, т. е. с задачами минимизации выпуклой функции либо максимизации вогнутой функции на выпуклом множестве. Напомним, что такие задачи обладают рядом полезных свойств, которые могут не иметь места в невыпуклом случае. В частности, любое локальное решение выпуклой задачи является глобальным, т. е. имеет смысл говорить просто о решении, причем множество решений всегда выпукло.

Главным образом речь здесь пойдет о фактах, связанных с фундаментальным понятием субдифференциала выпуклой функции. Доказательство всех приводимых утверждений можно найти в обширной литературе по выпуклому анализу, например в [10, 24, 25, 33, 35, 37].

Определение 1. Пусть $X \subset \mathbf{R}^n$ — произвольное множество. Вектор $y \in \mathbf{R}^n$ называется *субградиентом* функции $f\colon X \to \mathbf{R}$ в точке $x \in X$, если

$$f(\xi) \geqslant f(x) + \langle y, \xi - x\rangle \quad \forall\, \xi \in X.$$

Множество всех субградиентов функции f в точке x называется ее *субдифференциалом* в этой точке и обозначается $\partial f(x)$.

Элементарно проверяется, что субдифференциал — всегда замкнутое выпуклое множество.

Далее в этом пункте будем рассматривать функцию $f\colon \mathbf{R}^n \to \mathbf{R}$, предполагая ее выпуклость на всем $\mathbf{R}^n$. Понятие субдифференциала содержательно именно для выпуклых функций (см. теорему 1 ниже). Если функция f вогнута, то для нее вводится понятие *супердифференциала* как субдифференциала выпуклой функции $-f$. Супердифференциал вогнутой функции будем обозначать так же, как и субдифференциал выпуклой функции, поскольку это не приводит к путанице. Отметим, наконец, что для выпуклых и вогнутых функций $\partial f(x)$ совпадает с дифференциалом Кларка функции f в точке x, так что и в этом плане путаницы с обозначениями не возникает.

Теорема 1. *Функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *выпукла на* $\mathbf{R}^n$ *тогда и только тогда, когда* $\partial f(x) \neq \varnothing \ \ \forall\, x \in \mathbf{R}^n$.

Теорема 2. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *выпукла на* $\mathbf{R}^n$.
Тогда:
а) *функция* f *непрерывна на* $\mathbf{R}^n$;
б) *функция* f *дифференцируема в любой точке* $x \in \mathbf{R}^n$ *по любому направлению* $h \in \mathbf{R}^n$, *причем*

$$f'(x;\, h) = \max_{y \in \partial f(x)} \langle y,\, h \rangle;$$

в) *функция* f *дифференцируема в точке* $x \in \mathbf{R}^n$ *тогда и только тогда, когда* $\partial f(x)$ *состоит из единственного элемента, которым по необходимости является* $f'(x)$.

Теорема 3. *Пусть функции* $f_1,\, f_2\colon \mathbf{R}^n \to \mathbf{R}$ *выпуклы на* $\mathbf{R}^n$.
Тогда функция $f(\cdot) = f_1(\cdot) + f_2(\cdot)$ *выпукла на* $\mathbf{R}^n$, *причем*

$$\partial f(x) = \partial f_1(x) + \partial f_2(x) \quad \forall\, x \in \mathbf{R}^n.$$

Теорема 4. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *выпукла на* $\mathbf{R}^n$.
Тогда точка $\overline{x} \in \mathbf{R}^n$ *является решением задачи безусловной оптимизации*

$$f(x) \to \min, \quad x \in \mathbf{R}^n,$$

если и только если $0 \in \partial f(\overline{x})$.

Теорема 5. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *выпукла на* $\mathbf{R}^n$, $X \subset \mathbf{R}^n$ *— ограниченное множество.*
Тогда величина $L = \sup\limits_{y \in \cup_{x \in X} \partial f(x)} |y|$ *конечна, причем функция* f *непрерывна по Липшицу на* X *с константой* L.

6.1.2. Двойственная релаксация. Будем рассматривать задачу с прямым и функциональными ограничениями

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in P \,\big|\, F(x) = 0,\, G(x) \leqslant 0\}, \tag{2}$$

где $P \subset \mathbf{R}^n$ — некоторое множество, $f\colon P \to \mathbf{R}$ — заданная функция, $F\colon P \to \mathbf{R}^l$ и $G\colon P \to \mathbf{R}^m$ — заданные отображения.
Пусть $L\colon P \times \mathbf{R}^l \times \mathbf{R}^m \to \mathbf{R}$ — функция Лагранжа задачи (1), (2). Поскольку

$$\sup_{(\lambda,\, \mu) \in \mathcal{P}} L(x,\, \lambda,\, \mu) = \begin{cases} f(x), \text{ если } x \in D, \\ +\infty, \text{ если } x \in P \setminus D, \end{cases}$$

где $\mathcal{P} = \mathbf{R}^l \times \mathbf{R}^m_+$, то задача (1), (2) может быть записана в виде

$$\sup_{(\lambda,\, \mu) \in \mathcal{P}} L(x,\, \lambda,\, \mu) \to \min, \quad x \in D, \tag{3}$$

причем

$$D = \left\{ x \in P \,\middle|\, \sup_{(\lambda,\,\mu)\in\mathcal{P}} L(x,\,\lambda,\,\mu) < +\infty \right\}. \tag{4}$$

Идея *двойственной релаксации* по существу состоит в изменении порядка минимизации и максимизации в (3). В результате приходим к *двойственной задаче*

$$\varphi(\lambda,\,\mu) \to \max, \quad (\lambda,\,\mu) \in \mathcal{D}, \tag{5}$$

$$\mathcal{D} = \left\{ (\lambda,\,\mu) \in \mathcal{P} \;\middle|\; \inf_{x\in P} L(x,\,\lambda,\,\mu) > -\infty \right\}, \tag{6}$$

где

$$\varphi\colon \mathcal{D} \to \mathbf{R}, \quad \varphi(\lambda,\,\mu) = \inf_{x\in P} L(x,\,\lambda,\,\mu). \tag{7}$$

Задачу (1), (2), или, что то же самое, (3), (4), будем называть *прямой* (по отношению к задаче (5), (6)). Таким образом, вместо прямой задачи (1), (2) предлагается решать двойственную задачу (5), (6). Термин «релаксация» выражает тот факт, что двойственная задача часто оказывается проще прямой.

Задача 1. Рассмотрим задачу линейного программирования

$$\langle c_1,\,x_1\rangle + \langle c_2,\,x_2\rangle \to \min, \quad (x_1,\,x_2) \in D,$$

$$D = \{(x_1,\,x_2) \in P \,\big|\, A_{11}x_1 + A_{12}x_2 = b_1,\; A_{21}x_1 + A_{22}x_2 \geqslant b_2\},$$

где $c_1 \in \mathbf{R}^{n_1}$, $c_2 \in \mathbf{R}^{n_2}$, $A_{11} \in \mathbf{R}(l,\,n_1)$, $A_{12} \in \mathbf{R}(l,\,n_2)$, $A_{21} \in \in \mathbf{R}(m,\,n_1)$, $A_{22} \in \mathbf{R}(m,\,n_2)$, $b_1 \in \mathbf{R}^l$, $b_2 \in \mathbf{R}^m$, $P = \mathbf{R}^{n_1} \times \mathbf{R}^{n_2}_+$. Показать, что двойственная к этой задача имеет вид

$$\langle b_1,\,\lambda\rangle + \langle b_2,\,\mu\rangle \to \max, \quad (\lambda,\,\mu) \in \mathcal{D},$$

$$\mathcal{D} = \{(\lambda,\,\mu) \in \mathcal{P} \,\big|\, A_{11}^{\mathrm{T}}\lambda + A_{21}^{\mathrm{T}}\mu = c_1,\; A_{12}^{\mathrm{T}}\lambda + A_{22}^{\mathrm{T}}\mu \leqslant c_2\},$$

где $\mathcal{P} = \mathbf{R}^l \times \mathbf{R}^m_+$.

Задача 2. Рассмотрим задачу квадратичного программирования

$$\frac{1}{2}\,\langle Cx,\,x\rangle + \langle c,\,x\rangle \to \min, \quad x \in D,$$

$$D = \{x \in \mathbf{R}^n \,\big|\, Ax \leqslant b\},$$

где $C \in \mathbf{R}(n,\,n)$ — симметрическая положительно определенная матрица, $c \in \mathbf{R}^n$, $A \in \mathbf{R}(m,\,n)$, $b \in \mathbf{R}^m$. Показать, что двойственная к этой задача имеет вид

$$-\frac{1}{2}\,\langle AC^{-1}A^{\mathrm{T}}\mu,\,\mu\rangle - \langle AC^{-1}c + b,\,\mu\rangle - \frac{1}{2}\,\langle C^{-1}c,\,c\rangle \to \max, \quad \mu \in \mathbf{R}^m_+.$$

Показать, что единственное решение $\overline{x}$ прямой задачи выражается через любое решение $\overline{\mu}$ двойственной задачи равенством

$$\overline{x} = -C^{-1}(c + A^{\mathrm{T}}\overline{\mu}).$$

Разумеется, без дальнейших разъяснений переход от прямой задачи к двойственной не более чем формальный прием. Когда и почему использование этого приема может иметь смысл? Прежде всего, речь идет о ситуациях, в которых отыскание решения двойственной задачи в некотором смысле проще, чем прямой, причем решения этих задач определенным образом связаны друг с другом. В любом случае подразумевается, что имеется эффективный способ вычисления значений функции φ, т. е. *оракул* («черный ящик»), который для любой подаваемой на его вход пары $(\lambda, \mu) \in \mathcal{P}$ относительно легко находит значение точной нижней грани в (7). В некоторых важных случаях внутренняя задача минимизации в (7) допускает декомпозицию, а значит, легко решается.

Задача 3. Пусть $n = \sum\limits_{j=1}^{s} n_j$, $P_j \subset \mathbf{R}^{n_j}$, $f_j\colon P_j \to \mathbf{R}$, $F_j\colon P_j \to \mathbf{R}^l$, $G_j\colon P_j \to \mathbf{R}^m$, $j = 1, \ldots, s$, $P = \prod\limits_{j=1}^{s} P_j$,

$$f(x) = \sum_{j=1}^{s} f_j(x_j), \quad F(x) = \sum_{j=1}^{s} F_j(x_j),$$

$$G(x) = \sum_{j=1}^{s} G_j(x_j), \quad x = (x_1, \ldots, x_s) \in P$$

(функции и отображения такого вида называют *сепарабельными*). Показать, что для множества $\mathcal{D}$ и функции φ, вводимых согласно (6) и (7), и для всякой пары $(\lambda, \mu) \in \mathcal{D}$ справедливо равенство

$$\varphi(\lambda, \mu) = \sum_{j=1}^{s} \varphi_j(\lambda, \mu),$$

где

$$\varphi_j(\lambda, \mu) = \inf_{x_j \in P_j} (f_j(x_j) + \langle \lambda, F_j(x_j)\rangle + \langle \mu, G_j(x_j)\rangle), \quad j = 1, \ldots, s.$$

В частности, внутренняя задача минимизации в (7) распадается на s задач меньшей размерности, которые могут быть решены независимо (например, параллельно).

Обсудим вопрос о том, какую пользу можно извлечь из двойственной релаксации. По построению

$$f(x) \geqslant L(x, \lambda, \mu) \geqslant \varphi(\lambda, \mu) \quad \forall\, x \in D, \quad \forall\, (\lambda, \mu) \in \mathcal{D}, \tag{8}$$

т. е. значение целевой функции двойственной задачи в любой ее допустимой точке служит оценкой снизу для значения целевой функции прямой задачи в любой ее допустимой точке, а значит, оценкой снизу для значения $\overline{v} = \inf\limits_{x \in D} f(x)$ прямой задачи. (Такие оценки особенно полезны в «трудных» задачах, например, в целочисленном программировании, когда множество P дискретно.) В частности,

$$\overline{v} \geqslant \overline{\omega}, \tag{9}$$

где

$$\overline{\omega} = \sup_{(\lambda,\, \mu) \in \mathcal{D}} \varphi(\lambda,\, \mu)$$

— значение двойственной задачи. Соотношение (9) называется *слабым соотношением двойственности.*

Теорема 6. *Для любого множества* $P \subset \mathbf{R}^n$, *любой функции* $f\colon P \to \mathbf{R}$ *и любых отображений* $F\colon P \to \mathbf{R}^l$ *и* $G\colon P \to \mathbf{R}^m$ *множество* $\mathcal{D}$, *вводимое согласно* (6), *выпукло, а функция* φ, *вводимая согласно* (7), *вогнута и полунепрерывна сверху на* $\mathcal{D}$. *Кроме того, если для некоторой пары* $(\lambda,\, \mu) \in \mathcal{D}$ *точная нижняя грань в* (7) *достигается в точке* $x_{\lambda,\, \mu} \in P$, *то* $(F(x_{\lambda,\, \mu}),\, G(x_{\lambda,\, \mu})) \in \partial\varphi(\lambda,\, \mu)$.

Задача 4. Доказать теорему 6.

Выводы, которые можно сделать из теоремы 6, весьма примечательны. Двойственная задача должна решаться относительно легко, если только значения ее целевой функции достаточно легко вычислимы. Действительно, двойственная задача всегда является задачей максимизации вогнутой функции на выпуклом множестве, причем эта задача «корректно поставлена» в смысле полунепрерывности сверху ее целевой функции на допустимом множестве. И это — без каких-либо предположений относительно сложности и других свойств прямой задачи (в частности, множество P может иметь сколь угодно плохую структуру, например, дискретную). Далее, размерность пространства переменных двойственной задачи равна числу (скалярных) функциональных ограничений прямой, а это число часто бывает значительно меньше размерности пространства переменных прямой задачи. Кроме того, ограничения двойственной задачи часто оказываются проще ограничений прямой; иллюстрацией этого факта служит, например, задача 2. Наконец, если для данной пары $(\lambda,\, \mu) \in \mathcal{D}$ найдено решение $x_{\lambda,\, \mu}$ внутренней задачи минимизации в (7), то тем самым определено не только значение $\varphi(\lambda,\, \mu)$, но и (без дополнительных затрат) один суперградиент $(F(x_{\lambda,\, \mu}),\, G(x_{\lambda,\, \mu})) \in \partial\varphi(\lambda,\, \mu)$.

Разумеется, чудес не бывает (см. ниже комментарии о возможном разрыве двойственности и восстановлении решений прямой задачи).

Сделаем следующее важное замечание. В общем случае функция φ не является гладкой. Ее гладкость можно гарантировать, если точная нижняя грань в (7) достигается на единственном элементе для любой пары $(\lambda, \mu) \in \mathcal{D}$; но это не так во многих прикладных задачах. Поэтому в контексте двойственной релаксации на первый план выходят методы негладкой оптимизации. Кроме того, если внутренняя задача минимизации в (7) имеет более одного решения, то типичной является ситуация, когда удается вычислить лишь одно из этих решений, а значит, только один суперградиент. Поэтому любой практический метод должен подразумевать использование оракула, выдающего в лучшем случае только ограниченную информацию — значение функции и один суперградиент. Имеется в виду, что, скажем, схема, предполагающая вычисление всего супердифференциала целевой функции двойственной задачи, вряд ли будет практически полезна.

Под *разрывом двойственности* можно понимать разность $\overline{v} - \overline{\omega}$ (см. неравенство (9)). Если выполнено *соотношение двойственности*

$$\overline{v} = \overline{\omega},$$

т. е.

$$\inf_{x \in P} \sup_{(\lambda, \mu) \in \mathcal{P}} L(x, \lambda, \mu) = \sup_{(\lambda, \mu) \in \mathcal{P}} \inf_{x \in P} L(x, \lambda, \mu),$$

то говорят, что разрыв двойственности отсутствует. В этом случае, как следует из (8), все глобальные решения прямой задачи (1), (2) можно найти, решая задачу

$$L(x, \overline{\lambda}, \overline{\mu}) \to \min, \quad x \in P, \tag{10}$$

где $(\overline{\lambda}, \overline{\mu}) \in \mathcal{D}$ — любое решение двойственной задачи (5), (6). Не вдаваясь в подробности, заметим, что отсутствие разрыва двойственности можно гарантировать в случае, когда (1), (2) — задача выпуклого программирования, ограничения которой удовлетворяют некоторому условию регулярности. Задача (1), (2) называется *задачей выпуклого программирования*, если множество P выпукло, отображение F аффинно, а компоненты $g_i(\cdot)$ отображения G выпуклы на P, $i = 1, \ldots, m$. Одним из наиболее важных в этом контексте условий регулярности является так называемое *условие Слейтера*, состоящее в том, что $l = 0$ (ограничения-равенства отсутствуют) и существует точка $\widetilde{x} \in P$ такая, что $G(\widetilde{x}) < 0$, а также условие линейности (двойственность для линейных задач изучается в п. 7.1.2).

Необходимо, однако, иметь в виду, что даже при отсутствии разрыва двойственности задача (10) может иметь «лишние» решения, не являющиеся решениями задачи (1), (2) (такие решения заведомо не являются допустимыми точками в (1), (2)).

Вопрос о разрыве двойственности и восстановлении решений прямой задачи при отсутствии предположения о ее выпуклости весьма

важен, но здесь не рассматривается. Ограничимся лишь следующим фактом.

Задача 5. Пусть для множества $P \subset \mathbf{R}^n$, функции $f\colon P \to \mathbf{R}$, отображений $F\colon P \to \mathbf{R}^l$ и $G\colon P \to \mathbf{R}^m$ и некоторой пары $(\lambda, \mu) \in \in \mathcal{D}$ точная нижняя грань в (7) достигается в точке $x_{\lambda,\mu} \in P$. Показать, что $x_{\lambda,\mu}$ является глобальным решением следующего возмущения задачи (1), (2):

$$f(x) \to \min, \quad x \in D_{\lambda,\mu},$$

$$D_{\lambda,\mu} = \Bigg\{ x \in P \Big| \; F(x) = F(x_{\lambda,\mu}),$$

$$g_i(x) \leqslant \begin{cases} g_i(x_{\lambda,\mu}), & \text{если } \mu_i > 0, \\ \max\{0,\, g_i(x_{\lambda,\mu})\}, & \text{если } \mu_i = 0, \end{cases} \qquad i = 1, \ldots, m \Bigg\}.$$

В частности, если точка $x_{\lambda,\mu}$ допустима в прямой задаче и удовлетворяет равенствам $g_i(x_{\lambda,\mu}) = 0$ для всех $i = 1, \ldots, m$ таких, что $\mu_i > 0$, то $x_{\lambda,\mu}$ является глобальным решением задачи (1), (2).

§ 6.2. Субградиентные методы. Кусочно линейная аппроксимация

6.2.1. Субградиентные методы. Обратимся к задаче безусловной оптимизации

$$f(x) \to \min, \quad x \in \mathbf{R}^n, \tag{1}$$

где функция $f\colon \mathbf{R}^n \to \mathbf{R}$ предполагается выпуклой на $\mathbf{R}^n$, но не предполагается гладкой.

Начнем с некоторых комментариев, имеющих целью продемонстрировать, что идеи рассматривавшихся выше методов гладкой оптимизации в данном контексте неприменимы, что приводит к необходимости развития специальных методов негладкой оптимизации.

Согласно утверждению б) теоремы 6.1.2 $\forall\, x,\, d \in \mathbf{R}^n$

$$f'(x;\, d) = \max_{y \in \partial f(x)} \langle y,\, d \rangle,$$

откуда легко следует, что $d \in \mathcal{D}_f(x)$ тогда и только тогда, когда $\langle y,\, d \rangle < 0 \;\; \forall\, y \in \partial f(x)$. Последствием этого является тот неутешительный факт, что для отыскания направления убывания функции f в точке x нужно знать весь субдифференциал $\partial f(x)$ (напомним, что субдифференциал выпуклой функции всегда непуст; см. теорему 6.1.1). Как уже отмечалось выше в связи с двойственной релаксацией, в практических задачах это обычно невозможно, если только функция f не является дифференцируемой в точке x (относительно последнего случая см. утверждение в) теоремы 6.1.2). Сказанное свидетельствует о том, что использование идеи методов спуска в контексте негладкой оптимизации может быть весьма проблематичным, во всяком случае в ее оригинальной форме.

Задача 1. Рассмотрим функцию $f\colon \mathbf{R}^2 \to \mathbf{R}$, $f(x) = |x_1| + 2|x_2|$. Показать, что для любого $x_1 \in \mathbf{R}$ в точке $x = (x_1, 0)$ справедливо следующее утверждение: $y = (1, 2) \in \partial f(x)$, но $-y \notin \mathcal{D}_f(x)$.

Другую очень серьезную проблему представляет получение осмысленных правил остановки вычислительного процесса. Например, весьма популярный в гладкой безусловной оптимизации тест на достаточную малость $|f'(x^k)|$ в текущей точке $x^k \in \mathbf{R}^n$ не имеет удовлетворительного аналога в негладком случае. Среди прочего это связано с «недостатком непрерывности» субградиентного отображения $x \to \partial f(x)\colon \mathbf{R}^n \to \mathbf{R}^n$. Рассмотрим простейший пример, в котором $n = 1$, $f(x) = |x|$, $x \in \mathbf{R}$. Как бы близка ни была точка $x^k \neq 0$ к решению $\overline{x} = 0$ задачи (1), имеет место равенство $|y| = 1$ $\forall\, y \in \partial f(x^k)$. Более того, даже в случае попадания в точное решение, т. е. при $x^k = 0$, нет никакой гарантии, что этот факт будет обнаружен, а вычислительный процесс остановлен. Действительно, если оракул в каждой точке выдает лишь один субградиент, то он может выдать любой вектор $y \in \partial f(0) = [-1, 1]$, и величина $|y|$ совершенно не обязательно будет мала. Не вдаваясь в подробности, констатируем, что и другие правила остановки, естественные для гладкой оптимизации, оказываются неадекватными в негладком случае.

В *субградиентных методах* сложности, связанные с вычислением направления убывания, преодолеваются (в формальном, чисто теоретическом смысле) очень просто за счет снятия самого требования монотонного убывания значений целевой функции и за счет априорного назначения параметров длины шага. Имея в виду связь между субградиентными и более изощренными методами, рассматриваемыми ниже, будем считать, что вместо субдифференциала используется его ε-расширение при $\varepsilon \geqslant 0$, а именно множество

$$\partial_\varepsilon f(x) = \{y \in \mathbf{R}^n \mid f(\xi) \geqslant f(x) + \langle y, \xi - x\rangle - \varepsilon \quad \forall\, \xi \in \mathbf{R}^n\}, \tag{2}$$

называемое *ε-субдифференциалом* функции f в точке $x \in \mathbf{R}^n$.

Алгоритм 1. Выбираем $x^0 \in \mathbf{R}^n$ и полагаем $k = 0$. Выбираем последовательности $\{\varepsilon_k\}, \{\alpha_k\} \subset \mathbf{R}_+$.

1. Вычисляем $d^k \in \partial_{\varepsilon_k} f(x^k)$.

2. Полагаем

$$x^{k+1} = x^k - \alpha_k d^k. \tag{3}$$

3. Увеличиваем номер шага k на 1 и переходим к п. 1.

Теорема 1. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ выпукла на $\mathbf{R}^n$. Пусть в алгоритме 1 последовательности $\{\varepsilon_k\}$ и $\{\alpha_k\}$ удовлетворяют условиям*

$$\sum_{k=0}^{\infty} \alpha_k = +\infty, \tag{4}$$

$$\lim_{k\to\infty} \varepsilon_k = 0. \tag{5}$$

Тогда для любой траектории $\{x^k\}$ *алгоритма* 1 *такой, что для соответствующей последовательности* $\{d^k\}$ *выполнено условие*

$$\lim_{k\to\infty} \alpha_k |d^k|^2 = 0, \tag{6}$$

имеет место

$$\liminf_{k\to\infty} f(x^k) = \overline{v},$$

где $\overline{v} = \inf\limits_{x\in\mathbf{R}^n} f(x)$.

Доказательство. Из (2) и (3) для произвольного $x \in \mathbf{R}^n$ имеем

$$|x^{k+1} - x|^2 = |x^k - x|^2 + 2\langle x^{k+1} - x^k,\, x^k - x\rangle + |x^{k+1} - x^k|^2 \leqslant$$
$$\leqslant |x^k - x|^2 + 2\alpha_k(f(x) - f(x^k) + \varepsilon_k) + \alpha_k^2 |d^k|^2. \tag{7}$$

Предположим, что $\liminf\limits_{k\to\infty} f(x^k) > \overline{v}$. Тогда найдутся элемент $x \in \mathbf{R}^n$, число $\delta > 0$ и индекс $\overline{k}$ такие, что

$$f(x) < f(x^k) - \delta \quad \forall\, k \geqslant \overline{k}. \tag{8}$$

Из (5) и (6) следует, что если $\overline{k}$ достаточно велико, то

$$\alpha_k |d^k|^2 + 2\varepsilon_k \leqslant \delta \quad \forall\, k \geqslant \overline{k}.$$

Но тогда согласно (7) и (8)

$$|x^{k+1} - x|^2 \leqslant |x^k - x|^2 + \alpha_k(2\varepsilon_k + \alpha_k |d^k|^2 - 2\delta) \leqslant$$
$$\leqslant |x^k - x|^2 - \delta\alpha_k \quad \forall\, k \geqslant \overline{k}.$$

Поэтому

$$\delta \sum_{i=\overline{k}}^{k} \alpha_i \leqslant \sum_{i=\overline{k}}^{k} (|x^i - x|^2 - |x^{i+1} - x|^2) = |x^{\overline{k}} - x|^2 - |x^{k+1} - x|^2 \leqslant |x^{\overline{k}} - x|^2,$$

что противоречит (4). □

Субградиентные направления в алгоритме 1 удобно нормировать; в этом случае можно предложить условия на параметр длины шага, позволяющие усилить утверждение о сходимости. Для простоты ограничимся случаем точных субдифференциалов. В итоге приходим к схеме

$$x^{k+1} = x^k - \beta_k \frac{d^k}{|d^k|}, \quad d^k \in \partial f(x^k), \quad k = 0, 1, \ldots, \tag{9}$$

где $\beta_k > 0$ — параметры длины шага (подразумевается, что если в текущей точке x^k вычислен субградиент $d^k = 0$, то процесс останавливают).

Теорема 2. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *выпукла на* $\mathbf{R}^n$, *причем задача* (1) *имеет решение. Пусть в алгоритме* 1

$$\varepsilon_k = 0, \quad \alpha_k = \frac{\beta_k}{|d^k|} \quad \forall\, k, \tag{10}$$

где последовательность $\{\beta_k\}$ *удовлетворяет условиям*

$$\sum_{k=0}^{\infty} \beta_k = +\infty, \quad \sum_{k=0}^{\infty} \beta_k^2 < +\infty. \tag{11}$$

Тогда любая траектория $\{x^k\}$ *алгоритма* 1 *сходится к решению задачи* (1).

Задача 2. Доказать, что если последовательности $\{a_k\}, \{b_k\} \subset \subset \mathbf{R}_+$ удовлетворяют условию $a_{k+1} \leqslant a_k + b_k \quad \forall\, k$, причем $\sum\limits_{k=0}^{\infty} b_k < < +\infty$, то последовательность $\{a_k\}$ сходится.

Доказательство теоремы 2. Пусть $\overline{x} \in \mathbf{R}^n$ — некоторое решение задачи (1). Тогда $f(\overline{x}) \leqslant f(x^k) \quad \forall\, k$, и, взяв в (7) $x = \overline{x}$, с учетом (10) получим

$$|x^{k+1} - \overline{x}|^2 \leqslant |x^k - \overline{x}|^2 + \beta_k^2.$$

Отсюда, из второго соотношения в (11) и утверждения из задачи 2 следует, что последовательность $\{|x^k - \overline{x}|\}$ сходится. В частности, последовательность $\{x^k\}$ ограничена, поэтому согласно теореме 6.1.5 последовательность $\{d^k\}$ также ограничена. Но тогда из (10) и (11) вытекают соотношения (4)–(6). Поэтому, применяя теорему 1, получаем, что $\liminf\limits_{k\to\infty} f(x^k) = f(\overline{x})$. Отсюда, из непрерывности функции f на $\mathbf{R}^n$ (см. утверждение а) теоремы 6.1.2) и из ограниченности последовательности $\{x^k\}$ следует, что $\{x^k\}$ имеет предельную точку $\widetilde{x} \in \mathbf{R}^n$, являющуюся решением задачи (1). Заменяя в проведенном выше рассуждении $\overline{x}$ на $\widetilde{x}$, получим сходимость последовательности $\{|x^k - \widetilde{x}|\}$, причем сходимость к нулю, поскольку $\widetilde{x}$ является предельной точкой. $\square$

С помощью операции проектирования субградиентные методы несложно распространить на случай наличия (простых) ограничений, т. е. на задачу

$$f(x) \to \min, \quad x \in P, \tag{12}$$

где $P \subset \mathbf{R}^n$ — замкнутое выпуклое множество.

Задача 3. Пусть $P \subset \mathbf{R}^n$ — замкнутое выпуклое множество, а функция $f\colon \mathbf{R}^n \to \mathbf{R}$ выпукла на $\mathbf{R}^n$. Модифицируем алгоритм 1, заменив в нем формулу (3) формулой

$$x^{k+1} = \pi_P(x^k - \alpha_k d^k).$$

Получить аналоги теорем 1 и 2 для такой модификации алгоритма 1 применительно к задаче (12).

По-видимому, единственным привлекательным свойством субградиентных методов является их чрезвычайная простота, если известен простой способ вычисления субградиентов. К сожалению, на практике эти методы часто оказываются бесполезными. Априорное задание параметров длины шага неизбежно приводит к низкой (сублинейной) скорости сходимости, какие бы (разумные) требования не накладывались на решаемую задачу, и это, разумеется, совершенно неприемлемо. Кроме того, в рамках субградиентных методов никак не решается фундаментальная проблема выбора сколько-нибудь надежного правила остановки, обсуждавшаяся выше.

Можно даже сказать, что два условия в (11) на выбор параметров длины шага несовместимы с вычислительной точки зрения: второе условие подразумевает, что $\beta_k \to 0 \quad (k \to \infty)$, т. е. в представлении компьютера β_k становится нулем при больших k, и это делает невозможным выполнение первого условия. С другой стороны, теоретически существует бесконечное множество подходящих последовательностей $\{\beta_k\}$, и совершенно не ясно, почему один выбор этой последовательности предпочтительнее другого. Это, конечно же, свидетельствует о том, что метод едва ли может быть эффективен.

Наконец, известны некоторые ситуации, в которых субградиентный метод можно несколько улучшить. Одна из таких ситуаций возникает, если заранее известно значение $\overline{v}$ задачи (1).

Задача 4. Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ выпукла на $\mathbf{R}^n$, причем задача (1) имеет решение. Доказать, что любая траектория $\{x^k\}$, генерируемая по схеме (9), в которой $\beta_k = f(x^k) - \overline{v}$, $\overline{v} = \inf\limits_{x \in \mathbf{R}^n} f(x)$, сходится к некоторому решению задачи (1), причем

$$\liminf_{k \to \infty} \sqrt{k}(f(x^k) - \overline{v}) = 0.$$

Если дополнительно выполнено так называемое *условие остроты (линейного роста)*, т. е. существуют окрестность U множества решений S задачи (1) и число $\gamma > 0$ такие, что

$$f(x) - \overline{v} \geqslant \gamma \operatorname{dist}(x, S) \quad \forall\, x \in U,$$

то скорость сходимости по аргументу геометрическая.

Если значение $\overline{v}$ не известно, можно попытаться получить модификацию метода из задачи 4, использующую динамический выбор оценок снизу величины $\overline{v}$.

6.2.2. Методы кусочно линейной аппроксимации. Будем рассматривать задачу (12), предполагая, что $P \subset \mathbf{R}^n$ — выпуклый компакт, а функция $f\colon \mathbf{R}^n \to \mathbf{R}$ выпукла на $\mathbf{R}^n$. Компактность P

предполагается для того, чтобы гарантировать существование решений у вводимых ниже вспомогательных задач, что иначе не является автоматическим.

Базовая идея методов, рассматриваемых в этом пункте, состоит в использовании информации, накапливаемой по ходу итераций, для построения все более точной кусочно линейной аппроксимации снизу целевой функции. В частности, такие методы являются многошаговыми. Будем считать, что к $(k+1)$-му шагу получена следующая *выборка* [1]:

$$x^i \in \mathbf{R}^n, \quad y^i \in \partial f(x^i), \quad i = 0, 1, \ldots, k.$$

Тогда $(k+1)$-й шаг *метода кусочно линейной аппроксимации* состоит в решении задачи

$$\psi_k(x) \to \min, \quad x \in P, \tag{13}$$

где

$$\psi_k\colon \mathbf{R}^n \to \mathbf{R}, \quad \psi_k(x) = \max_{i=0,\,1,\ldots,k} \{f(x^i) + \langle y^i,\, x - x^i\rangle\}. \tag{14}$$

Заметим, что если множество P — полиэдр, то задача (13) сводится к задаче линейного программирования

$$\sigma \to \min, \quad (\sigma,\, x) \in U_k, \tag{15}$$

$$U_k = \{(\sigma,\, x) \in \mathbf{R} \times P \,\big|\, f(x^i) + \langle y^i,\, x - x^i\rangle \leqslant \sigma,\ i = 0, 1, \ldots, k\}. \tag{16}$$

Из определения субдифференциала и (14) следует, что

$$f(x) \geqslant \psi_{k+1}(x) \geqslant \psi_k(x) \quad \forall\, x \in \mathbf{R}^n. \tag{17}$$

По мере добавления информации в выборку последняя становится все богаче, а аппроксимация (14) функции f — все точнее. Это дает основания ожидать, что решения задач (13) будут все лучше приближать решения задачи (12).

Алгоритм 2. Выбираем $x^0 \in \mathbf{R}^n$ и полагаем $k = 0$.

1. Вычисляем $f(x^k)$ и $y^k \in \partial f(x^k)$.
2. Вычисляем x^{k+1} как решение задачи (13) с целевой функцией, задаваемой формулой (14).
3. Увеличиваем номер шага k на 1 и переходим к п. 1.

[1] В западной литературе для подобных выборок информации обычно используется термин Bundle, что можно перевести как «связка» или «пучок». В связи с этим методы, рассматриваемые в § 6.3, обычно называют Bundle methods. Однако, с точки зрения авторов, такая терминология при всей ее выразительности не слишком информативна. Методы, рассматриваемые в этом параграфе, известны в западной литературе под названием Cutting plane methods.

Наиболее важным достоинством описанного метода по сравнению с субградиентными методами является, видимо, возможность ввести разумное правило остановки. Для каждого k положим

$$\Delta_k = f(x^{k+1}) - \psi_k(x^{k+1}).$$

Из (17) следует, что $\Delta_k \geqslant 0$ и

$$f(x) \geqslant \psi_k(x) \geqslant \psi_k(x^{k+1}) = f(x^{k+1}) - \Delta_k \quad \forall\, x \in P,$$

поэтому

$$f(x^{k+1}) \geqslant \overline{v} \geqslant f(x^{k+1}) - \Delta_k,$$

где $\overline{v}$ — значение задачи (12). Это значит, что разумно останавливать процесс после $(k+1)$-го шага, если Δ_k не превосходит заданной величины, выражающей допустимую погрешность аппроксимации значения задачи. Заметим, что для некоторого k указанное событие неминуемо произойдет (см. теорему 3 ниже).

Какие недостатки имеют методы кусочно линейной аппроксимации? Требование компактности P едва ли следует считать слишком ограничительным. Более серьезную проблему представляет все возрастающая сложность используемой аппроксимации по мере расширения выборки, т. е. все возрастающее количество (скалярных) функциональных ограничений в (15), (16). И хотя некоторые стратегии удаления «лишней» информации из выборки известны, для методов кусочно линейной аппроксимации в их примитивном виде реализация этих стратегий весьма проблематична (ср. с методами, рассматриваемыми в § 6.3).

Другой очень серьезный недостаток методов кусочно линейной аппроксимации состоит в их неизбежной неустойчивости. В частности, последовательности $\{x^k\}$ и $\{f(x^k)\}$ обычно ведут себя очень хаотично. Чтобы убедиться в этом, достаточно применить алгоритм 2 к задаче (12), в которой $n = 1$, $P = [-1, 1]$, $f(x) = x^2$, $x \in \mathbf{R}$. Разумеется, этот пример весьма специфичен в том смысле, что функция f в нем гладкая; тем не менее он хорошо иллюстрирует смысл происходящего. Можно утверждать, что методы кусочно линейной аппроксимации хорошо работают лишь в том случае, когда задача имеет единственное решение, которое является острым (см. задачу 4).

Теорема 3. *Пусть $P \subset \mathbf{R}^n$ — выпуклый компакт, а функция $f\colon \mathbf{R}^n \to \mathbf{R}$ выпукла на $\mathbf{R}^n$.*

Тогда для любой траектории $\{x^k\}$ алгоритма 2 имеет место

$$\lim_{k\to\infty} \psi_k(x^{k+1}) = \overline{v} = \liminf_{k\to\infty} f(x^k),$$

где $\overline{v} = \inf\limits_{x\in P} f(x)$, а функции ψ_k задаются формулой (14).

Доказательство. Из (17) следует монотонное неубывание последовательности $\{\psi_k(x^{k+1})\}$ и ее ограниченность сверху величи-

ной $\overline{v}$. Но тогда эта последовательность сходится. Предположим, что она сходится не к $\overline{v}$, т. е. найдется число $\delta > 0$ такое, что

$$\psi_k(x^{k+1}) \leqslant \overline{v} - \delta \quad \forall\, k. \tag{18}$$

Последовательность $\{x^k\}$ лежит в компакте P, поэтому, при необходимости переходя к подпоследовательности, можем считать, что эта последовательность сходится. Но тогда из теоремы 6.1.5 следует существование числа $M > 0$ такого, что для любого достаточно большого k

$$|y^k| \leqslant M, \quad |x^{k+1} - x^k| \leqslant \frac{\delta}{2M}\,.$$

Отсюда, из (14), (18) и неравенства Коши–Буняковского–Шварца получаем

$$\overline{v} - \delta \geqslant \psi_k(x^{k+1}) \geqslant f(x^k) + \langle y^k,\, x^{k+1} - x^k\rangle \geqslant$$

$$\geqslant \overline{v} - M|x^{k+1} - x^k| \geqslant \overline{v} - \frac{\delta}{2},$$

что невозможно.

Теперь предположим, что существует число $\delta > 0$ такое, что для любого достаточно большого k

$$f(x^k) \geqslant \overline{v} + \delta. \tag{19}$$

Вновь, при необходимости переходя к подпоследовательности, можем считать, что

$$|x^{k+1} - x^k| \leqslant \frac{\delta}{2L}\,,$$

где $L = \sup\limits_{y \in \cup_{x \in P} \partial f(x)} |y|$. Напомним, что в силу теоремы 6.1.5 величина L конечна и является константой Липшица для функции f, а значит, и для ψ_{k+1} (см. задачу 2.2.3) на P. Отсюда и из (14), (17), (19) получаем

$$\overline{v} + \delta \leqslant f(x^{k+1}) \leqslant \max\{\psi_k(x^{k+1}),\, f(x^{k+1})\} = \psi_{k+1}(x^{k+1}) =$$

$$= \psi_{k+1}(x^k) + \psi_{k+1}(x^{k+1}) - \psi_{k+1}(x^k) \leqslant \overline{v} + L|x^{k+1} - x^k| \leqslant \overline{v} + \frac{\delta}{2},$$

что опять же невозможно. □

§ 6.3. Многошаговые методы с квадратичными подзадачами

Вновь будем рассматривать задачу безусловной оптимизации

$$f(x) \to \min, \quad x \in \mathbf{R}^n, \tag{1}$$

где функция $f\colon \mathbf{R}^n \to \mathbf{R}$ выпукла на $\mathbf{R}^n$, но не обязательно является гладкой.

Рассматриваемые в этом параграфе методы строятся на основе идеи кусочно линейной аппроксимации (см. п. 6.2.2), которая дополняется необходимым стабилизирующим механизмом и удовлетворительными правилами остановки и управления размерами выборки информации, определяющей используемую аппроксимацию. Подчеркнем, что, в отличие от субградиентных методов и методов кусочно линейной аппроксимации, эти методы обеспечивают убывание значения целевой функции от итерации к итерации.

Чтобы объяснить основную идею, рассмотрим сначала упрощенный вариант метода. Пусть $x^k \in \mathbf{R}^n$ — текущее приближение к решению задачи (1), а $z^i \in \mathbf{R}^n$, $y^i \in \partial f(z^i)$, $i = 0, 1, \ldots, k$, — полученная к $(k+1)$-му шагу выборка; подчеркнем, что точка z^k может не совпадать с x^k. Точка $z^{k+1} \in \mathbf{R}^n$ ищется как решение задачи

$$\psi_k(x) + \frac{\gamma_k}{2}|x - x^k|^2 \to \min, \quad x \in \mathbf{R}^n, \tag{2}$$

где

$$\psi_k : \mathbf{R}^n \to \mathbf{R}, \quad \psi_k(x) = \max_{i=0,\,1,\,\ldots,\,k} \{f(z^i) + \langle y^i,\, x - z^i\rangle\}, \tag{3}$$

а $\gamma_k > 0$ — так называемый *проксимальный параметр*. Точку x^k можно рассматривать как центр стабилизации, а роль квадратичного члена состоит в том, чтобы не допустить слишком большого отклонения точки z^{k+1} от этого центра, что может иметь место, если использовать чистую кусочно линейную аппроксимацию (см. п. 6.2.2). Идея состоит в том, чтобы менять центр стабилизации лишь в том случае, когда это действительно оправдано, а именно, когда переход от x^k к z^{k+1} приводит к достаточному уменьшению значения функции f (*серьезный шаг* [1]). Если это так, то полагают $x^{k+1} = z^{k+1}$. В противном случае центр стабилизации (и текущее приближение) оставляют без изменений, полагая $x^{k+1} = x^k$ (*нулевой шаг* [2]). Заметим, что информация, полученная за счет решения задачи (2), при этом не пропадает: она обогащает выборку, а значит, улучшает аппроксимацию целевой функции.

Однако, для того, чтобы описанная базовая схема приводила к практическим методам, ее следует дополнить рядом важных деталей.

Прежде всего, заметим, что введенную в (3) функцию ψ_k можно представить в следующем «центрированном в точке x^k» виде:

$$\psi_k(x) = f(x^k) + \max_{i=0,\,1,\,\ldots,\,k} \{-e_i^k + \langle y^i,\, x - x^k\rangle\}, \tag{4}$$

где через e_i^k обозначена ошибка «линеаризации» функции f в точке z^i, вычисленная в точке x^k, т. е.

$$e_i^k = f(x^k) - f(z^i) - \langle y^i,\, x^k - z^i\rangle \geqslant 0, \quad i = 0, 1, \ldots, k \tag{5}$$

[1] От общепринятого английского термина Serious step.

[2] От общепринятого английского термина Null step.

(неотрицательность следует из определения субградиента). Отметим, что такое представление ψ_k может иметь преимущества перед (3) при практической реализации метода, например, из-за более низких требований к используемой памяти: для каждого $i = 0, 1, \ldots, k$ вместо вектора $z^i \in \mathbf{R}^n$ достаточно хранить число e_i^k, считая, что выборка состоит именно из пар вида (y^i, e_i^k) (а не (z^i, y^i)). Разумеется, ошибки «линеаризации», определяемые в (5), нужно правильно пересчитывать всякий раз, когда меняется центр x^k.

Из (5) и определения ε-субградиента очевидно следует, что

$$y^i \in \partial_{e_i^k} f(x^k) \quad \forall\, i = 0, 1, \ldots, k. \tag{6}$$

Далее, заметим, что трудоемкость решения вспомогательной задачи (2) зависит от количества $k+1$ линейных «кусков», определяющих функцию ψ_k в (4). Как будет показано ниже, количество ограничений в задаче квадратичного программирования, к которой сводится (2), равно $k+1$. Ясно, что количество таких ограничений не может увеличиваться неограниченно и должно оставаться приемлимым с вычислительной точки зрения. Это значит, что когда количество элементов в выборке достигает определенного (реалистичного) предела, то выборка должна некоторым образом сжиматься.

Сжатие выборки означает, что функция ψ_k заменяется другой кусочно линейной функцией, в определении которой участвует меньшее количество линейных «кусков», и которая будет по-прежнему обозначаться ψ_k. Имея ввиду представление (4), формально положим

$$\psi_k(x) = f(x^k) + \max_{i \in B_k} \{-e_i^k + \langle y^i,\, x - x^k \rangle\}, \tag{7}$$

где $B_k \subset \{0, 1, \ldots, k\}$ — некоторое множество индексов, определяющее текущую выборку, причем

$$y^i \in \partial_{e_i^k} f(x^k) \quad \forall\, i \in B_k \tag{8}$$

(ср. с (6)), считая, что теперь текущая выборка определяется как множество пар (y^i, e_i^k), $i \in B_k$, удовлетворяющих (8). Заметим, что из (7), (8) и определения ε-субдифференциала следует, что

$$f(x) \geqslant \psi_k(x) \quad \forall\, x \in \mathbf{R}^n, \tag{9}$$

т. е. функция ψ_k по-прежнему оценивает f снизу (ср. с 6.2.17).

Еще раз подчеркнем, что B_k не обязательно совпадает с $\{0, 1, \ldots, k\}$: идея состоит как раз в том, что при больши́х k в B_k должно входить значительно меньшее количество индексов. Кроме того, по причинам, которые станут ясны ниже, здесь уже не предполагается, что если $i \in B_k$, то обязательно $y^i \in \partial f(z^i)$ для некоторого $i \in \{0, 1, \ldots, k\}$ (хотя, конечно, это может быть и так). Естественно, если используются не все «куски», определяемые предыдущими $k+1$ точками, то ε-субградиенты в (8) должны

выбираться умным образом, чтобы не была потеряна информация, существенная для аппроксимации целевой функции f, а значит, и для сходимости алгоритма. В связи с этим, фундаментальную роль играет обсуждаемая ниже процедура агрегирования, которая основана на использовании информации, получаемой при решении вспомогательной задачи (2).

Поскольку целевая функция задачи (2) сильно выпукла [1], решение этой задачи всегда существует и единственно. Кроме того, задача (2) с целевой функцией, определяемой согласно (7), сводится к выпуклой задаче квадратичного программирования

$$\sigma + \frac{\gamma_k}{2}|x - x^k|^2 \to \min, \quad (\sigma,\, x) \in U_k, \tag{10}$$

$$U_k = \{(\sigma,\, x) \in \mathbf{R} \times \mathbf{R}^n \,\big|\, f(x^k) - e_i^k + \langle y^i,\, x - x^k\rangle \leqslant \sigma,\ i \in B_k\}. \tag{11}$$

С вычислительной точки зрения (выпуклые) задачи квадратичного программирования сравнимы по сложности с задачами линейного программирования. Эти два класса задач допускают сходные эффективные численные стратегии (конечные методы активного множества, методы внутренней точки; см. гл. 7).

Более того, при реализации рассматриваемых здесь методов вместо (10), (11) обычно решают следующую задачу, возникающую из соображений, связанных с двойственностью (см. задачу 1 ниже):

$$\frac{1}{2}\Big|\sum_{i\in B_k} \nu_i y^i\Big|^2 + \gamma_k \sum_{i\in B_k} \nu_i e_i^k \to \min, \quad \nu \in \mathcal{U}_k, \tag{12}$$

$$\mathcal{U}_k = \Big\{\nu \in \mathbf{R}_+^{|B_k|} \,\Big|\, \sum_{i\in B_k} \nu_i = 1\Big\}. \tag{13}$$

Заметим, что задача (12), (13) всегда имеет решение в силу теоремы Вейерштрасса (теорема 1.1.1). Более того, ограничения этой задачи имеют очень специальную структуру: $\mathcal{U}_k$ есть стандартный симплекс в $\mathbf{R}^{|B_k|}$, и этот факт может быть эффективно использован при численном решении таких задач.

Установим некоторые свойства решения z^{k+1} вспомогательной задачи (2).

Лемма 1. *Пусть функция f:$\mathbf{R}^n \to \mathbf{R}$ выпукла на $\mathbf{R}^n$. Для произвольных $x^k \in \mathbf{R}^n$, $e_i^k \geqslant 0$, $y^i \in \partial_{e_i^k} f(x^k)$, $i \in B_k$, и $\gamma_k > 0$ положим*

[1] Сумма выпуклой и сильно выпуклой функции сильно выпукла. Максимум семейства выпуклых функций — выпуклая функция.

$$d^k = \sum_{i \in B_k} \nu_i^k y^i, \tag{14}$$

где $\nu^k \in \mathbf{R}^{|B_k|}$ *— произвольное решение задачи* (12), (13).

Тогда единственное решение z^{k+1} *задачи* (2) *с определяемой согласно* (7) *целевой функцией имеет вид*

$$z^{k+1} = x^k - \frac{1}{\gamma_k} d^k, \tag{15}$$

причем

$$d^k \in \partial_{\varepsilon_k} f(x^k), \tag{16}$$

где

$$\varepsilon_k = \sum_{i \in B_k} \nu_i^k e_i^k \geqslant 0. \tag{17}$$

Задача 1. Показать, что в условиях леммы 1 двойственная задача к (10), (11) имеет вид

$$-\frac{1}{2\gamma_k} \left| \sum_{i \in B_k} \nu_i y^i \right|^2 - \sum_{i \in B_k} \nu_i e_i^k + f(x^k) \to \max, \quad \nu \in \mathcal{U}_k, \tag{18}$$

где множество $\mathcal{U}_k$ введено в (13).

Таким образом, элемент ν^k в лемме 1 играет роль множителя Лагранжа, отвечающего решению z^{k+1} задачи (10), (11).

Доказательство леммы 1. Очевидно, $z^{k+1} \in \mathbf{R}^n$ является решением задачи (2) в том, и только том случае, когда $(\sigma_{k+1},\, z^{k+1})$ является решением задачи (10), (11), где $\sigma_{k+1} = \psi_k(z^{k+1})$. Согласно результату задачи 1 ν^k является решением двойственной задачи к (10), (11) (очевидно, решения задач (12), (13) и (18), (13) совпадают), а значит, в силу сказанного в конце п. 6.1.2 о восстановлении решения прямой задачи при отсутствии разрыва двойственности, $(\sigma_{k+1},\, z^{k+1})$ является решением задачи

$$\sigma + \frac{\gamma_k}{2}|x - x^k|^2 + \sum_{i \in B_k} \nu_i^k (f(x^k) - e_i^k + \langle y^i,\, x - x^k \rangle - \sigma) \to \min, \quad (\sigma,\, x) \in \mathbf{R} \times \mathbf{R}^n.$$

Применяя к этой задаче принцип Ферма (теорема 1.2.3), приходим к равенству

$$\gamma_k (z^{k+1} - x^k) + \sum_{i \in B_k} \nu_i^k y^i = 0,$$

а это с учетом (14) и есть (15).

Далее, в силу утверждения в) теоремы 6.1.2 и теорем 6.1.3, 6.1.4 имеем

$$0 \in \partial \psi_k(z^{k+1}) + \gamma_k (z^{k+1} - x^k),$$

что с учетом доказанного равенства (15) влечет включение

$$d^k \in \partial\psi_k(z^{k+1}). \tag{19}$$

Отсюда, из (9) и определения субдифференциала имеем

$$f(x) \geqslant \psi_k(x) \geqslant \psi_k(z^{k+1}) + \langle d^k,\, x - z^{k+1}\rangle \quad \forall\, x \in \mathbf{R}^n. \tag{20}$$

Используя соотношение двойственности (см. п. 6.1.2) применительно к взаимодвойственным задачам (10), (11) и (18), (13) и учитывая (14) приходим к равенству

$$\psi_k(z^{k+1}) + \frac{\gamma_k}{2}|z^{k+1} - x^k|^2 = f(x^k) - \frac{|d^k|^2}{2\gamma_k} - \sum_{i\in B_k} \nu_i^k e_i^k.$$

Отсюда и из (15) следует, что

$$\psi_k(z^{k+1}) = f(x^k) - \frac{|d^k|^2}{\gamma_k} - \varepsilon_k, \tag{21}$$

где ε_k определено в (17).

Объединяя (20) и (21) и используя (15), получаем

$$f(x) \geqslant f(x^k) + \langle d^k,\, x - x^k\rangle + \langle d^k,\, x^k - z^{k+1}\rangle - \frac{|d^k|^2}{\gamma_k} - \varepsilon_k =$$

$$= f(x^k) + \langle d^k,\, x - x^k\rangle - \varepsilon_k \quad \forall\, x \in \mathbf{R}^n,$$

что и дает (16). □

Обратимся к упомянутой выше процедуре агрегирования и введем функцию

$$\psi_k^a\colon \mathbf{R}^n \to \mathbf{R}, \quad \psi_k^a(x) = f(x^k) + \langle d^k,\, x - x^k\rangle - \varepsilon_k, \tag{22}$$

где $d^k \in \mathbf{R}^n$ и $\varepsilon_k \geqslant 0$ определяются также, как в лемме 1 (см. (14), (15), (17)). Эта функция полностью определяется решением $\nu^k \in \mathbf{R}^{|B_k|}$ задачи (12), (13). Согласно (16) и определению ε-субдифференциала

$$f(x) \geqslant \psi_k^a(x) \quad \forall\, x \in \mathbf{R}^n, \tag{23}$$

т. е. ψ_k^a, как и ψ_k, оценивает f снизу.

Чтобы иметь основания надеяться на сходимость конструируемого алгоритма, нужно обеспечить выполнение следующего свойства: если x^k не является решением задачи (1), то следующий серьезный шаг должен реализоваться через конечное число нулевых шагов. Это свойство нетрудно обосновать, если продолжать добавлять новые линейные «куски» в аппроксимацию целевой функции на каждом шаге, и при этом не избавляться ни от каких из уже имеющихся «кусков». Однако в силу указанных выше причин такая стратегия с практической точки зрения не удовлетворительна. С другой стороны, удалять «куски» нужно аккуратно, так, чтобы новая функция ψ_{k+1} в некотором смысле лучше аппроксимировала f, чем ψ_k, причем даже в том

случае, когда ψ_{k+1} определяется не бóльшим количеством «кусков», чем ψ_k.

Оказывается, что, грубо говоря, имеющиеся «куски» можно удалять произвольным образом, если добавить в аппроксимацию в качестве нового «куска» агрегированную функцию ψ_k^a. Более точно, для обеспечения работоспособности метода достаточно использовать кусочно линейные аппроксимации ψ_k, которые для всякого $k = 0, 1, \ldots$ удовлетворяют условию (9), причем если $x^{k+1} = x^k$, то

$$\psi_{k+1}(x) \geqslant \max\{\psi_k^a(x),\ f(z^{k+1}) + \langle y^{k+1},\ x - z^{k+1}\rangle\} \quad \forall\, x \in \mathbf{R}^n, \quad (24)$$

где $y^{k+1} \in \partial f(z^{k+1})$.

Условие (24) означает, что при определении новой выборки после нулевого шага, достаточно включить в нее пару $(y^{k+1},\ e_{k+1}^{k+1})$, где

$$e_{k+1}^{k+1} = f(x^k) - f(z^{k+1}) - \langle y^{k+1},\ x^k - z^{k+1}\rangle \geqslant 0, \quad (25)$$

а также пару вида $(d^k,\ \varepsilon_k)$, заменив ей некоторую имеющуюся пару $(y^i,\ e_i^k)$, $i \in B_k$ (напомним, что теперь не предполагается, что первый элемент пары, входящей в выборку, должен быть субградиентом функции f в какой-либо из вычисленных к данному моменту точек, а условие (6) при $y^i = d^k$ и $e_i^k = \varepsilon_k$ выполняется в силу (16)). Иными словами, новая кусочно линейная аппроксимация должна включать в себя линейный «кусок», определяемый новой парой $(z^{k+1},\ y^{k+1})$, а также «агрегированный кусок» ψ_k^a. В принципе, на каждой итерации выборка может содержать как угодно мало элементов, лишь бы она содержала указанные два. В частности, можно вообще удалить из выборки все предыдущие пары и определять точку z^{k+2} как решение задачи

$$\begin{aligned} &\max\{\psi_k^a(x),\ f(z^{k+1}) + \langle y^{k+1},\ x - z^{k+1}\rangle\} + \\ &\qquad + \frac{\gamma_{k+1}}{2}|x - x^{k+1}|^2 \to \min, \quad x \in \mathbf{R}^n. \end{aligned} \quad (26)$$

Разумеется, столь экстремальное сжатие выборки на практике обычно не применяется, но принципиальная возможность удалить из выборки любое количество имеющихся пар чрезвычайно важна для эффективного управления ее размером, а значит, для контроля трудоемкости решения вспомогательных задач (10), (11). Обычно чем беднее используемые выборки, тем хуже ψ_k аппроксимирует f и тем ниже скорость сходимости метода, но и тем дешевле обходится решение вспомогательных задач.

Заметим, что если не удалять из выборки имеющиеся пары, то для обеспечения выполнения (24) нет нужды вводить в новую аппроксимацию ψ_{k+1} «агрегированный кусок» ψ_k^a; достаточно ввe-

сти «кусок», определяемый новой парой (z^{k+1}, y^{k+1}). Действительно, при этом для всякого $x \in \mathbf{R}^n$ в силу (15), (20)–(22) имеем

$$\begin{aligned}\psi_{k+1}(x) &= \max\{\psi_k(x),\, f(z^{k+1}) + \langle y^{k+1},\, x - z^{k+1}\rangle\} \geqslant \\ &\geqslant \psi_k(x) \geqslant \psi_k(z^{k+1}) + \langle d^k,\, x - z^{k+1}\rangle = \\ &= f(x^k) - \frac{|d^k|^2}{\gamma_k} - \varepsilon_k + \langle d^k,\, x - z^{k+1}\rangle = \psi_k^a(x).\end{aligned}$$

Наконец, если очередной шаг оказывается серьезным, т. е. $x^{k+1} = z^{k+1} \neq x^k$, то единственное требование на функцию ψ_{k+1} состоит в том, что она должна оценивать функцию f снизу. Если это так, то конечное число последующих нулевых шагов (с выполнением (24)) позволит получить аппроксимацию f достаточно хорошую для реализации нового серьезного шага. Ясно однако, что разумно и начинать это построение с хорошей аппроксимации (а не «с нуля», как на первой итерации). Чтобы использовать для этого пары, уже хранящиеся в выборке, нужно, как уже отмечалось выше, правильно пересчитывать соответствующие ошибки «линеаризации», чтобы, в частности, обеспечить выполнение ключевого соотношения (8) с $k+1$ вместо k. Легко проверить, что можно использовать следующую формулу пересчета:

$$e_i^{k+1} = e_i^k + f(x^{k+1}) - f(x^k) + \langle y^i,\, x^k - x^{k+1}\rangle, \quad i \in B_{k+1} \setminus \{k+1\}. \tag{27}$$

Перейдем к формальному описанию *многошагового метода с квадратичными подзадачами*.

Алгоритм 1. Фиксируем целое число $M \geqslant 2$ (максимальный размер выборки), выбираем $x^0 \in \mathbf{R}^n$ и полагаем $k = 0$, $\mathcal{K} = \varnothing$. Выбираем $\varepsilon \in (0, 1)$ и последовательность $\{\gamma_k\} \in \mathbf{R}_+ \setminus \{0\}$. Вычисляем $f(x^0)$ и $y^0 \in \partial f(z^0)$, и полагаем $z^0 = x^0$, $e_0^0 = 0$. Полагаем $B_0 = \{0\}$ и определяем соответствующую начальную выборку, состоящую из одной пары (y^0, e_0^0).

1. Вычисляем z^{k+1} как решение задачи (2), целевая функция которой определяется согласно (7).

2. Вычисляем $f(z^{k+1})$, $y^{k+1} \in \partial f(z^{k+1})$ и
$$\Delta_k = f(x^k) - \psi_k(z^{k+1}) - \frac{\gamma_k}{2}|z^{k+1} - x^k|^2. \tag{28}$$

3. Если
$$f(x^k) - f(z^{k+1}) \geqslant \varepsilon \Delta_k, \tag{29}$$
то полагаем $x^{k+1} = z^{k+1}$ и включаем индекс k в множество $\mathcal{K}$ (серьезный шаг). В противном случае полагаем $x^{k+1} = x^k$ (нулевой шаг).

4. Если $|B_k| < M$, то полагаем $B_{k+1} = B_k \cup \{k+1\}$. Если же $|B_k| = M$, то выбираем $i_1, i_2 \in B_k$ и полагаем $B_{k+1} = (B_k \setminus \{i_2\}) \cup \{k+1\}$, $y^{i_1} = d^k$, $e^k_{i_1} = \varepsilon_k$, где $d^k \in \mathbf{R}^n$ и $\varepsilon_k \geqslant 0$ определяются согласно лемме 1 (см. (14), (15), (17)).

5. Если $k \in \mathcal{K}$, то вычисляем e^{k+1}_i, $i \in B_{k+1} \setminus \{k+1\}$, по формуле (27). В противном случае полагаем $e^{k+1}_i = e^k_i$, $i \in B_{k+1} \setminus \{k+1\}$. Вычисляем e^{k+1}_{k+1} по формуле (25). Определяем новую выборку как множество пар (y^i, e^{k+1}_i), $i \in B_{k+1}$.

6. Увеличиваем номер шага k на 1 и переходим к п. 1.

На шаге 4 алгоритма в качестве i_1 и i_2 разумно выбирать те индексы, для которых соответствующие компоненты вектора ν^k равны нулю (если такие компоненты есть). Данная рекомендация основана на следующем соображении. Удаление из выборки пар, отвечающих неактивным в решении z^{k+1} ограничениям задачи (10), (11), не меняет решения z^{k+1} этой задачи, а неактивным ограничениям отвечают именно нулевые компоненты множителя Лагранжа ν^k.

Что касается выбора последовательности $\{\gamma_k\}$, то теоретические ограничения на этот выбор весьма необременительны (см. теоремы 1 и 2 ниже). Вместе с тем, на практике выбор $\{\gamma_k\}$ очень серьезно влияет на эффективность алгоритма и относится к области искусства профессионалов-вычислителей. Обычно эта последовательность не выбирается заранее, а генерируется в ходе итерационного процесса. Ограничимся следующей формулой, имеющей квазиньютоновское происхождение и часто используемой на практике: для каждого k полагаем

$$\gamma_{k+1} = \left(\gamma_k^{-1} + \frac{\langle x^{k+1} - x^k,\, y^{k+1} - y^k\rangle}{|y^{k+1} - y^k|^2}\right)^{-1}$$

(предполагается, что $y^{k+1} \neq y^k$; детали см. [42]). Заметим, что в случае нулевого шага эта формула дает $\gamma_{k+1} = \gamma_k$.

Обратимся к вопросу о правиле остановки. Из (16) и неравенства Коши–Буняковкого–Шварца имеем

$$f(x) \geqslant f(x^k) + \langle d^k,\, x - x^k\rangle - \varepsilon_k \geqslant f(x^k) - |d^k||x - x^k| - \varepsilon_k \quad \forall\, x \in \mathbf{R}^n,$$

поэтому имеет смысл останавливать процесс после $(k+1)$-го шага, если $|d^k|$ и ε_k и не превосходят требуемой точности. Из (15), (21) и (28) вытекает равенство

$$\Delta_k = \varepsilon_k + \frac{|d^k|^2}{2\gamma_k}. \tag{30}$$

Поэтому $\Delta_k \geqslant 0$, и если эта величина мала, то малы и $|d^k|$ и ε_k (по крайней мере, в случае ограниченности последовательности

$\{\gamma_k\}$). Ниже будет показано, что в случае бесконечного числа серьезных шагов подпоследовательность последовательности $\{\Delta_k\}$, отвечающая $k \in \mathcal{K}$, стремится к нулю. В случае же конечного числа серьезных шагов вся последовательность $\{\Delta_k\}$ стремится к нулю. Значит, критерий остановки, состоящий в достаточной малости Δ_k, будет необходимым образом выполнен после конечного числа шагов. Подчеркнем, что Δ_k можно рассматривать и как меру качества аппроксимации функции f функцией ψ_k.

Приведенное правило остановки допускает дальнейшую интерпретацию и обоснование через свойства непрерывности ε-субдифференциала, но этот анализ выходит за рамки данной книги. Отметим только следующий ключевой факт: если $d^k \in \partial_{\varepsilon_k} f(x^k)$ $\forall k$ и $\{x^k\} \to \overline{x}$, $\varepsilon_k \to \overline{\varepsilon}$, $\{d^k\} \to \overline{d}$ $(k \to \infty)$ при некоторых $\overline{\varepsilon} \in \mathbf{R}_+$ и $\overline{d} \in \mathbf{R}^n$, то $\overline{d} \in \partial_{\overline{\varepsilon}} f(\overline{x})$. В частности, если считать, что в правиле остановки используется нулевая точность, то $\varepsilon_k \to 0$, $\{d^k\} \to 0$ $(k \to \infty)$, причем $0 \in \partial_0 f(\overline{x}) = \partial f(\overline{x})$.

Доказательство сходимости алгоритма 1 распадается на два случая в зависимости от конечности или бесконечности количества серьезных шагов.

Теорема 1. *Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ выпукла на $\mathbf{R}^n$. Пусть в алгоритме 1 реализуется бесконечное множество $\mathcal{K}$, причем последовательность $\{\gamma_k\}$ удовлетворяет условию*

$$\sum_{k \in \mathcal{K}} \frac{1}{\gamma_k} = +\infty. \tag{31}$$

Тогда для траектории $\{x^k\}$ алгоритма 1 имеет место

$$\lim_{k \to \infty} f(x^k) = \overline{v}, \tag{32}$$

где $\overline{v} = \inf_{x \in \mathbf{R}^n} f(x)$, и для $k \in \mathcal{K}$

$$\Delta_k \to 0 \quad (k \to \infty). \tag{33}$$

Если же задача (1) имеет решение, причем

$$\gamma_k \geqslant \breve{\gamma} > 0 \quad \forall\, k \in \mathcal{K}, \tag{34}$$

где число $\breve{\gamma}$ не зависит от k, то $\{x^k\}$ сходится к некоторому решению задачи (1).

Доказательство. В силу леммы 1 для всякого $k \in \mathcal{K}$ имеем

$$x^{k+1} = x^k - \frac{1}{\gamma_k} d^k, \quad d^k \in \partial_{\varepsilon_k} f(x^k), \tag{35}$$

$$f(x^k) - f(x^{k+1}) \geqslant \varepsilon \Delta_k,$$

где числа $\gamma_k > 0$, $|d^k|$, $\Delta_k \geqslant 0$ и $\varepsilon_k \geqslant 0$ связаны соотношением (30).

Если $\overline{v} > -\infty$, то

$$\sum_{k\in\mathcal{K}} \Delta_k \leqslant \sum_{k\in\mathcal{K}} \frac{f(x^k) - f(x^{k+1})}{\varepsilon} \leqslant \frac{f(x^0) - \overline{v}}{\varepsilon} < +\infty. \tag{36}$$

Отсюда и из (30) следует, что

$$\sum_{k\in\mathcal{K}} \varepsilon_k < +\infty, \quad \sum_{k\in\mathcal{K}} \frac{|d^k|^2}{\gamma_k} < +\infty. \tag{37}$$

Соотношения (31), (35), (37) и теорема 6.2.1 влекут равенство

$$\liminf_{k\to\infty} f(x^k) = \overline{v}.$$

С учетом того, что по построению последовательность $\{f(x^k)\}$ монотонно невозрастает (свойство, отсутствующее у субградиентных методов и методов кусочно линейной аппроксимации!), отсюда следует (32). Предельное соотношение (33) вытекает немедленно из (36).

Пусть теперь $\overline{x} \in \mathbf{R}^n$ — решение задачи (1). Тогда $\forall\, k \in \mathcal{K}$

$$|x^{k+1} - \overline{x}|^2 \leqslant |x^k - \overline{x}|^2 + \frac{|d^k|^2}{\gamma_k^2} + \frac{2\varepsilon_k}{\gamma_k} \tag{38}$$

(ср. с 6.2.7). Из (34) и (37) вытекает неравенство

$$\sum_{k\in\mathcal{K}} \left(\frac{|d^k|^2}{\gamma_k^2} + \frac{2\varepsilon_k}{\gamma_k} \right) < +\infty,$$

которое в совокупности с (38) и утверждением из задачи 6.2.2 гарантирует сходимость последовательности $\{|x^k - \overline{x}|\}$. В частности, последовательность $\{x^k\}$ ограничена. Пусть $\widetilde{x}$ — некоторая предельная точка этой последовательности, тогда в силу доказанного предельного соотношения (32) $\widetilde{x}$ является решением задачи (1). Заменяя в проведенном выше рассуждении $\overline{x}$ на $\widetilde{x}$ получим сходимость последовательности $\{|x^k - \widetilde{x}|\}$, причем, разумеется, сходимость к нулю. $\square$

При некоторых дополнительных условиях можно показать, что на серьезных шагах алгоритма 1 имеет место линейная оценка скорости сходимости, но обоснование этого потребовало бы привлечения дальнейших средств выпуклого анализа. Кроме того, как будет указано ниже, более изощренный выбор квадратичного члена для задач более конкретной структуры может обеспечить даже сверхлинейную скорость сходимости.

Следующий результат показывает, что при правильном управлении выборкой, а именно, при выполнении условий (9) и (в случае $x^{k+1} = x^k$) (24) для всех k, справедливо следующее. Если текущий центр x^k не является решением задачи (1), то тест на убывание (29) (или критерий остановки) обязательно выполнится после конечного

числа нулевых шагов. Если же x^k является решением задачи (1), то критерий остановки обязательно выполнится после конечного числа нулевых шагов.

Теорема 2. *Пусть функция* $f:\mathbf{R}^n \to \mathbf{R}$ *выпукла на* $\mathbf{R}^n$. *Пусть в алгоритме* 1 *реализуется конечное множество* $\mathcal{K}$, *т.е. для траектории* $\{x^k\}$ *справедливо* $x^k = x^{\overline{k}}\ \forall\, k \geqslant \overline{k}$ *для некоторого индекса* $\overline{k}$, *причем последовательность* $\{\gamma_k\}$ *ограничена и удовлетворяет условию*

$$\gamma_{k+1} \geqslant \gamma_k \quad \forall\, k \geqslant \overline{k}. \tag{39}$$

Тогда $x^{\overline{k}}$ *— решение задачи* (1), *причем последовательность* $\{z^k\}$ *сходится к* $x^{\overline{k}}$ *и выполняется* (33).

Доказательство. Всюду считаем, что $k \geqslant \overline{k}$. Сначала заметим, что, согласно (15), (21), (22), для всякого $x \in \mathbf{R}^n$

$$\psi_k^a(x) = f(x^{\overline{k}}) + \langle d^k,\, x - x^{\overline{k}}\rangle - \varepsilon_k = \psi_k(z^{k+1}) + \frac{|d^k|^2}{\gamma_k} + \langle d^k,\, x - x^{\overline{k}}\rangle =$$

$$= \psi_k(z^{k+1}) + \gamma_k\langle x^{\overline{k}} - z^{k+1},\, x - z^{k+1}\rangle.$$

В частности, $\psi_k^a(z^{k+1}) = \psi_k(z^{k+1})$, и для всякого $x \in \mathbf{R}^n$

$$\psi_k^a(x) + \frac{\gamma_k}{2}|x - x^{\overline{k}}|^2 = \psi_k(z^{k+1}) + \frac{\gamma_k}{2}|z^{k+1} - x^{\overline{k}}|^2 + \frac{\gamma_k}{2}|x - z^{k+1}|^2. \tag{40}$$

Используя (9), определение z^{k+2}, (24), (39), а также (40) (при $x = z^{k+2}$), выводим

$$\begin{aligned} f(x^{\overline{k}}) \geqslant \psi_{k+1}(x^{\overline{k}}) &\geqslant \psi_{k+1}(z^{k+2}) + \frac{\gamma_{k+1}}{2}|z^{k+2} - x^{\overline{k}}|^2 \geqslant \\ &\geqslant \psi_k^a(z^{k+2}) + \frac{\gamma_k}{2}|z^{k+2} - x^{\overline{k}}|^2 = \\ &= \psi_k(z^{k+1}) + \frac{\gamma_k}{2}|z^{k+1} - x^{\overline{k}}|^2 + \frac{\gamma_k}{2}|z^{k+2} - z^{k+1}|^2. \end{aligned} \tag{41}$$

Из промежуточных соотношений в (41) следует, что последовательность $\{\psi_k(z^{k+1}) + \gamma_k|z^{k+1} - x^{\overline{k}}|^2/2\}$ монотонно неубывает и ограничена, а значит сходится. Используя первое неравенство в (41), а также (24) и (40) (при $x = x^{\overline{k}}$), получаем

$$f(x^{\overline{k}}) \geqslant \psi_k^a(x^{\overline{k}}) = \psi_k(z^{k+1}) + \frac{\gamma_k}{2}|z^{k+1} - x^{\overline{k}}|^2 + \frac{\gamma_k}{2}|z^{k+1} - x^{\overline{k}}|^2.$$

Отсюда, в силу сходимости $\{\psi_k(z^{k+1}) + \gamma_k|z^{k+1} - x^{\overline{k}}|^2/2\}$ и ограниченности последовательности $\{\gamma_k\}$ снизу (величиной $\gamma_{\overline{k}} > 0$; см. (39)), следует ограниченность последовательности $\{z^k\}$. По аналогичным причинам, из (41) следует предельное соотношение

$$\{z^{k+2} - z^{k+1}\} \to 0 \quad (k \to \infty). \tag{42}$$

В силу ограниченности последовательности $\{z^k\}$ и теоремы 6.1.5 величина $L = \sup_{y \in \cup_{k=0}^{\infty} \partial f(z^k)} |y|$ конечна и является константой Липшица для функции f на множестве, состоящем из точек этой последовательности. Поэтому с учетом (9), (24) получаем

$$L|z^{k+2} - z^{k+1}| \geqslant f(z^{k+2}) - f(z^{k+1}) \geqslant \psi_{k+1}(z^{k+2}) - f(z^{k+1}) \geqslant$$
$$\geqslant \langle y^{k+1},\, z^{k+2} - z^{k+1} \rangle \geqslant -L|z^{k+2} - z^{k+1}|.$$

Отсюда и из (42) следует, что

$$\{f(z^k) - \psi_k(z^{k+1})\} \to 0 \quad (k \to \infty). \tag{43}$$

Пусть $\overline{z}$ — произвольная предельная точка ограниченной последовательности $\{z^k\}$, а $\{z^{k_j}\}$ — такая подпоследовательность этой последовательности, что $\{z^{k_j}\} \to \overline{z}$ $(j \to \infty)$. Тогда из (43) и из непрерывности f (см. утверждение а) теоремы 6.1.2) имеем

$$\{\psi_{k_j}(z^{k_j+1})\} \to f(\overline{z}) \quad (j \to \infty). \tag{44}$$

Кроме того, в силу (9), (15) и (19), для любого $x \in \mathbf{R}^n$ справедливо

$$f(x) \geqslant \psi_k(x) \geqslant \psi_k(z^{k+1}) + \langle d^k,\, x - z^{k+1} \rangle =$$
$$= \psi_k(z^{k+1}) + \gamma_k \langle x^{\overline{k}} - z^{k+1},\, x - z^{k+1} \rangle. \tag{45}$$

Заметим, что согласно (42) $\{z^{k_j+1}\} \to \overline{z}$ $(j \to \infty)$. Поэтому, переходя в (45) к пределу вдоль подпоследовательности $\{z^{k_j+1}\}$ и используя (44), получаем неравенство

$$f(x) \geqslant f(\overline{z}) + \overline{\gamma} \langle x^{\overline{k}} - \overline{z},\, x - \overline{z} \rangle \quad \forall\, x \in \mathbf{R}^n,$$

где $\overline{\gamma}$ — предел ограниченной монотонно неубывающей последовательности $\{\gamma_k\}$. Поэтому

$$\overline{\gamma}(x^{\overline{k}} - \overline{z}) \in \partial f(\overline{z}), \tag{46}$$

откуда следует, что $\overline{z}$ является решением задачи

$$f(x) + \frac{\overline{\gamma}}{2}|x - x^{\overline{k}}|^2 \to \min, \quad x \in \mathbf{R}^n$$

(см. теоремы 6.1.2–6.1.4). В частности,

$$f(\overline{z}) + \frac{\overline{\gamma}}{2}|\overline{z} - x^{\overline{k}}|^2 \leqslant f(x^{\overline{k}}). \tag{47}$$

С другой стороны, поскольку тест на убывание (29) не выполняется ни для одного $k \geqslant \overline{k}$, в силу (28) имеем

$$f(z^{k+1}) - f(x^{\overline{k}}) > -\varepsilon \Delta_k =$$
$$= -\varepsilon \left(f(x^{\overline{k}}) - \psi_k(z^{k+1}) - \frac{\gamma_k}{2}|z^{k+1} - x^{\overline{k}}|^2 \right) \geqslant -\varepsilon \left(f(x^{\overline{k}}) - \psi_k(z^{k+1}) \right).$$

Переходя к пределу вдоль подпоследовательности $\{z^{k_j+1}\}$ и используя (44), получаем

$$(1-\varepsilon)\left(f(x^{\overline{k}})-f(\overline{z})\right)\leqslant 0.$$

Но $\varepsilon<1$, поэтому отсюда следует, что $f(x^{\overline{k}})\leqslant f(\overline{z})$. Тогда из (47) вытекает равенство $\overline{z}=x^{\overline{k}}$, т.е. последовательность $\{z^k\}$ сходится к $x^{\overline{k}}$. Отсюда и из (28) и (43) следует (33). Кроме того, вспоминая (46), приходим к включению $0\in\partial f(x^{\overline{k}})$, которое, в силу теоремы 6.1.4, и означает, что $x^{\overline{k}}$ является решением задачи (1). $\square$

Другая мотивировка введения в (2) стабилизирующего квадратичного члена состоит в следующем. Общая теория численных методов оптимизации, а также здравый смысл свидетельствуют о том, что на высокую скорость сходимости можно расчитывать лишь при использовании аппроксимации второго порядка функции f. Разумеется, целевая функция задачи (2) едва ли может претендовать на роль такой аппроксимации. Однако, не вдаваясь в детали заметим, что проведенный выше анализ легко распространить на случай использования вместо (2) более общей вспомогательной задачи вида

$$\psi_k(x)+\frac{1}{2}\langle H_k(x-x^k),\,x-x^k\rangle\to\min,\quad x\in\mathbf{R}^n, \tag{48}$$

где $H_k\in\mathbf{R}(n,n)$ — равномерно положительно определенные симметрические матрицы. Схожая ситуация имеет место в гладкой условной оптимизации в связи с методами линеаризации и последовательного квадратичного программирования (см. § 5.4). Разумеется, здесь адекватный выбор H_k — весьма непростая проблема, поскольку H_k должна аппроксимировать некий «объект второго порядка», в то время как целевая функция задачи (1) может не иметь даже первой производной. Попытки предложить единый способ такого выбора (для произвольной выпуклой функции f), видимо, безнадежны. Тем не менее, достигнутые в последнее время успехи в анализе второго порядка для некоторых важных классов негладких выпуклых функций позволяют получать для соответствующих задач сверхлинейно сходящиеся методы с подзадачами вида (48).

Идеи, лежащие в основе обсуждавшихся в этом параграфе методов, могут реализовываться и в других формах. Например, вместо введения стабилизирующего квадратичного члена в целевую функцию вспомогательной задачи можно использовать доверительную область. В этом случае вспомогательная задача принимает вид

$$\psi_k(x)\to\min,\quad x\in D_k,$$

$$D_k=\{x\in\mathbf{R}^n\,\big|\,|x-x^k|\leqslant\delta_k\},$$

где $\delta_k > 0$ — радиус доверительной области. Если в определении D_k вместо $|\cdot|$ использовать $|\cdot|_\infty$, то эта задача сводится к задаче линейного программирования. Другая возможность состоит в введении вспомогательной задачи вида

$$|x - x^k|^2 \to \min, \quad x \in L_{\psi_k, \mathbf{R}^n}(c_k),$$

где параметр c_k назначается по определенным правилам.

Может сложиться впечатление, что вариант, использующий доверительные области (и, соответственно, линейные подзадачи), предпочтительнее варианта, использующего квадратичные подзадачи. Однако, на самом деле это не так. Во-первых, как уже отмечалось выше, трудоемкость решения задач линейного программирования и выпуклых задач квадратичного программирования посредством современных численных методов сравнима. Во-вторых, методы, использующие квадратичные подзадачи, могут выигрывать в некоторых аспектах реализации, например, связанных с управлением размером выборки. Можно показать, что описанные разные формы рассматриваемых методов в некотором теоретическом смысле эквивалентны: решение вспомогательной задачи для одной из этих форм будет решением вспомогательной задачи для любой другой формы при соответствующем (явно указываемом) выборе параметров. Вместе с тем реализации и практические свойства этих форм различны.

В завершение этого параграфа коротко обсудим многошаговые методы с квадратичными подзадачами для задач условной оптимизации. Если задача имеет вид

$$f(x) \to \min, \quad x \in P,$$

где $P \subset \mathbf{R}^n$ — полиэдр, то разумная стратегия состоит в прямом учете дополнительных линейных ограничений, задающих P, во вспомогательных задачах квадратичного программирования. Это значит, что вместо (2) следует использовать вспомогательные задачи вида

$$\psi_k(x) + \frac{\gamma_k}{2}|x - x^k|^2 \to \min, \quad x \in P,$$

которые по-прежнему очевидным образом сводятся к задачам квадратичного программирования. Проведенный выше анализ может быть распространен и на этот случай, но не без определенных изменений (достаточно заметить, что двойственная задача к новой вспомогательной задаче квадратичного программирования будет отлична от (18)).

В случае задачи с нелинейными ограничениями

$$f(x) \to \min, \quad x \in D, \tag{49}$$

$$D = \{x \in \mathbf{R}^n \mid G(x) \leqslant 0\}, \tag{50}$$

где $G\colon \mathbf{R}^n \to \mathbf{R}^m$ — отображение с выпуклыми компонентами $g_i(\cdot)$, $i = 1, \ldots, m$, можно использовать описанный выше метод применительно к задаче безусловной оптимизации

$$\varphi_c(x) \to \min, \quad x \in \mathbf{R}^n, \tag{51}$$

где используется семейство штрафных функций

$$\varphi_c\colon \mathbf{R}^n \to \mathbf{R}, \quad \varphi_c(x) = f(x) + c\sum_{i=1}^{m} \max\{0,\, g_i(x)\},$$

с параметром штрафа $c \geqslant 0$. Для фиксированного значения $c > 0$ параметра штрафа все построения этого параграфа проходят без изменений. Вместе с тем, если предполагать выполненным условие Слейтера, то можно показать, что данный штраф является точным, т. е. существует число $\overline{c} \geqslant 0$ такое, что для всякого $c > \overline{c}$ любое решение задачи (51) является решением задачи (49), (50) (ср. со следствием 4.7.1). Подходящее значение $\overline{c}$ заранее неизвестно, и его следует подбирать динамически в процессе вычислений. Эта процедура и другие детали должны быть хорошо проработаны для получения практически эффективного алгоритма.

Заметим, наконец, что существуют и более изощренные возможности, чем использование точных штрафных функций, но их обсуждение выходит за рамки данной книги.

Глава 7
СПЕЦИАЛЬНЫЕ ЗАДАЧИ ОПТИМИЗАЦИИ

Заключительная глава посвящена главным образом задачам линейного программирования (ЗЛП) и задачам квадратичного программирования (ЗКП). Важность этих классов задач определяется в первую очередь тем неоднократно отмечавшимся выше фактом, что многие методы общего назначения подразумевают многократное решение вспомогательных ЗЛП либо ЗКП, и итоговая эффективность таких методов в значительной степени определяется тем, насколько эффективно решаются вспомогательные задачи.

§ 7.1. Элементы теории линейного программирования

Можно, видимо, утверждать, что наиболее полно изученный класс задач условной оптимизации составляют ЗЛП, для которых разработаны как законченная теория, так и эффективные численные методы (хотя и в этой области до сих пор появляются новые результаты). Изложению важнейших вопросов теории линейного программирования и посвящен настоящий параграф.

7.1.1. Общие свойства линейных задач. Напомним, что под ЗЛП понимается задача оптимизации, целевая функция которой линейна, а допустимое множество является полиэдром. ЗЛП является выпуклой задачей оптимизации, а значит, обладает всеми свойствами, присущими выпуклым задачам. В частности, любое локальное решение ЗЛП является глобальным (поэтому далее будем говорить просто о решении), а любая стационарная (в смысле определения 1.2.2) точка является решением.

Общую ЗЛП будем записывать в виде

$$\langle c_1,\, x_1\rangle + \langle c_2,\, x_2\rangle \to \min, \quad (x_1,\, x_2) \in D, \tag{1}$$

$$D = \{(x_1,\, x_2) \in P \,\big|\, A_{11}x_1 + A_{12}x_2 = b_1,\ A_{21}x_1 + A_{22}x_2 \geqslant b_2\}, \tag{2}$$

где $c_1 \in \mathbf{R}^{n_1}$, $c_2 \in \mathbf{R}^{n_2}$, $A_{11} \in \mathbf{R}(l,\, n_1)$, $A_{12} \in \mathbf{R}(l,\, n_2)$, $A_{21} \in \in \mathbf{R}(m,\, n_1)$, $A_{22} \in \mathbf{R}(m,\, n_2)$, $b_1 \in \mathbf{R}^l$, $b_2 \in \mathbf{R}^m$, $P = \mathbf{R}^{n_1} \times \mathbf{R}^{n_2}_+$.

Выделение условия неотрицательности части переменных в качестве прямого ограничения оказывается удобным в этом контексте.

Вектор $c = (c_1, c_2)$ называют *вектором коэффициентов целевой функции*, матрицу

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

— *матрицей ограничений*, а вектор $b = (b_1, b_2)$ — *вектором правых частей ограничений* задачи (1), (2).

Задача 1. Пусть полиэдр $X = \{x \in \mathbf{R}^n \mid Ax \geqslant b\}$, где $A \in \mathbf{R}(m, n)$, $b \in \mathbf{R}^m$, ограничен, причем $\operatorname{int} X \neq \varnothing$. Требуется вписать в X шар наибольшего радиуса. Сформулировать эту задачу как ЗЛП.

Особую роль играет так называемая *каноническая ЗЛП*, которая получается из (1), (2) при $n_1 = m = 0$. Таким образом, полагая $n = n_2$, каноническую ЗЛП можно записать в виде

$$\langle c, x\rangle \to \min, \quad x \in D, \tag{3}$$

$$D = \{x \in P \mid Ax = b\}, \tag{4}$$

где $c = c_2 \in \mathbf{R}^n$, $A = A_{12} \in \mathbf{R}(l, n)$, $b = b_1 \in \mathbf{R}^l$, $P = \mathbf{R}^n_+$. Заметим, что общую ЗЛП (1), (2) за счет введения вспомогательных переменных всегда можно свести к канонической

$$\langle c_1, u^1 - u^2\rangle + \langle c_2, x_2\rangle \to \min, \quad (u^1, u^2, x_2, y) \in \widetilde{D},$$

$$\widetilde{D} = \{(u^1, u^2, x_2, y) \in \widetilde{P} \mid A_{11}(u^1 - u^2) + A_{12}x_2 = b_1,$$
$$A_{21}(u^1 - u^2) + A_{22}x_2 - y = b_2\},$$

где $\widetilde{P} = \mathbf{R}^{n_1}_+ \times \mathbf{R}^{n_1}_+ \times \mathbf{R}^{n_2}_+ \times \mathbf{R}^m_+$. Допустимые множества и решения этих двух задач связаны очевидным образом.

Теорема 1. *Если допустимое множество ЗЛП непусто, а значение конечно, то эта ЗЛП имеет решение.*

Доказательство. В силу сказанного выше о сведении общей ЗЛП (1), (2) к канонической доказательство достаточно провести только для последней. Пусть допустимое множество D задачи (3), (4) непусто, причем $\overline{v} = \inf\limits_{x \in D} \langle c, x\rangle < +\infty$. Предположим, что утверждение теоремы неверно, т. е. система

$$Ax = b, \quad \langle c, x\rangle = \overline{v}, \quad x \geqslant 0$$

относительно $x \in \mathbf{R}^n$ несовместна. Очевидно, это равносильно несовместности однородной системы

$$Ax - tb = 0, \quad \langle c, x\rangle - t\overline{v} = 0, \quad x \geqslant 0, \quad t > 0$$

относительно $(x, t) \in \mathbf{R}^n \times \mathbf{R}$, применяя к которой теорему Моцкина (лемму 1.4.2), получим существование пары $(y, \tau) \in \mathbf{R}^l \times \mathbf{R}$ такой,

что

$$A^{\mathrm{T}}y + \tau c \geqslant 0, \quad \langle b,\, y\rangle + \tau\overline{v} < 0.$$

Но тогда $\forall\, x \in D$ из первого неравенства имеем

$$\langle A^{\mathrm{T}}y,\, x\rangle + \tau\langle c,\, x\rangle \geqslant 0,$$

что вместе со вторым неравенством дает

$$\tau\langle c,\, x\rangle \geqslant -\langle y,\, Ax\rangle = -\langle y,\, b\rangle > \tau\overline{v}.$$

Поэтому

$$\inf_{x\in D} \tau\langle c,\, x\rangle > \tau\overline{v} = \tau \inf_{x\in D}\langle c,\, x\rangle,$$

что не может иметь места ни при каком τ. $\square$

Доказанный результат означает, что в ЗЛП решения может не быть лишь в том случае, когда либо пусто ее допустимое множество, либо бесконечно значение. Напомним, что нелинейные задачи могут не иметь решения и в других случаях.

Приведем критерий оптимальности для ЗЛП. Необходимость в этом критерии вытекает из теоремы 1.4.3 и того факта, что условие линейности является условием регулярности ограничений (см. задачу 1.4.4; ограничение $x_2 \geqslant 0$ здесь удобно трактовать как функциональное). Достаточность же следует из сказанного выше о выпуклости ЗЛП (впрочем, достаточность легко устанавливается и непосредственной проверкой).

Теорема 2. *Пусть* $c_1 \in \mathbf{R}^{n_1}$, $c_2 \in \mathbf{R}^{n_2}$, $A_{11} \in \mathbf{R}(l,\, n_1)$, $A_{12} \in \in \mathbf{R}(l,\, n_2)$, $A_{21} \in \mathbf{R}(m,\, n_1)$, $A_{22} \in \mathbf{R}(m,\, n_2)$, $b_1 \in \mathbf{R}^l$, $b_2 \in \mathbf{R}^m$, $P = \mathbf{R}^{n_1} \times \mathbf{R}^{n_2}_+$.

Точка $(\overline{x}_1,\, \overline{x}_2) \in D$ *является решением задачи* (1), (2) *тогда и только тогда, когда найдутся элементы* $\overline{\lambda} \in \mathbf{R}^l$ *и* $\overline{\mu} \in \mathbf{R}^m_+$ *такие, что*

$$A_{11}^{\mathrm{T}}\overline{\lambda} + A_{21}^{\mathrm{T}}\overline{\mu} = c_1, \quad A_{12}^{\mathrm{T}}\overline{\lambda} + A_{22}^{\mathrm{T}}\overline{\mu} \leqslant c_2, \tag{5}$$

$$\langle \overline{\mu},\, A_{21}\overline{x}_1 + A_{22}\overline{x}_2 - b_2\rangle = 0, \quad \langle A_{12}^{\mathrm{T}}\overline{\lambda} + A_{22}^{\mathrm{T}}\overline{\mu} - c_2,\, \overline{x}_2\rangle = 0. \tag{6}$$

Задача 2. Используя геометрические построения, высказать гипотезу о решении задачи

$$-2x_1 - x_2 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^2_+ \,\big|\, x_1 - x_2 \geqslant -2,\ -x_1 - 2x_2 \geqslant -7,\ -4x_1 + 3x_2 \geqslant -6\}.$$

Проверить эту гипотезу с помощью теоремы 2.

Задача 3. То же задание для задачи

$$-8x_1 + 2x_2 + 3x_3 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^3_+ \,\big|\, 2x_1 - x_2 - x_3 = 2,\ x_1 - 3x_2 - x_3 \geqslant -4,\ -7x_1 + 2x_3 \geqslant -16\}.$$

Займемся фундаментальным понятием вершины полиэдра D. Здесь будет удобно считать, что $n_2 = 0$, т. е. прямых ограничений нет (они отнесены к функциональным ограничениям-неравенствам). Тогда, полагая $n = n_1$, можем переписать (2) в виде

$$D = \{x \in \mathbf{R}^n \mid A_1 x = b_1,\ A_2 x \geqslant b_2\}, \tag{7}$$

где $A_1 = A_{11} \in \mathbf{R}(l, n)$, $A_2 = A_{21} \in \mathbf{R}(l, m)$. Пусть a_i, $i = 1, \ldots, m$, — строки матрицы A_2. Как обычно, для всякой точки $x \in \mathbf{R}^n$ через $I(x) = \{i = 1, \ldots, m \mid \langle a_i, x\rangle = (b_2)_i\}$ будем обозначать множество номеров активных ограничений-неравенств в этой точке. Следующее определение вполне соответствует интуитивному представлению о вершинах.

Определение 1. Точка $\widehat{x} \in D$ называется *вершиной* заданного в (7) полиэдра D, если строки матрицы A_1 и векторы a_i, $i \in I(\widehat{x})$, образуют в $\mathbf{R}^n$ систему ранга n. Если число элементов в этой системе равно n, то вершина называется *невырожденной*, а если больше n, то *вырожденной*.

Иными словами, $\widehat{x} \in D$ является вершиной полиэдра D вида (7), если $\widehat{x}$ является единственным решением линейной системы

$$A_1 x = b_1, \quad \langle a_i, x\rangle = (b_2)_i, \quad i \in I(\widehat{x}).$$

Отсюда немедленно вытекает следующий важный факт.

Предложение 1. *Любой полиэдр имеет не более конечного числа вершин.*

Теорема 3. *Для любых* $A_1 \in \mathbf{R}(l, n)$, $A_2 \in \mathbf{R}(m, n)$, $b_1 \in \mathbf{R}^l$ *и* $b_2 \in \mathbf{R}^m$ *заданный в* (7) *полиэдр* D *имеет по крайней мере одну вершину тогда и только тогда, когда* $D \neq \varnothing$ *и для матрицы*

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$$

справедливо $\operatorname{rank} A = n$. *Более того, если* $\operatorname{rank} A = n$, *то для любой точки* $x \in D$ *существует такая вершина* $\widehat{x}$, *что* $I(x) \subset I(\widehat{x})$.

Доказательство. Необходимость очевидна. Для доказательства достаточности возьмем произвольную точку $x \in D$ и положим $I = I(x)$, $J = \{1, \ldots, m\} \setminus I$. Если строки матрицы A_1 и векторы a_i, $i \in I$, образуют в $\mathbf{R}^n$ систему ранга n, то x является вершиной по определению, и все доказано. Пусть ранг этой системы равен $r < n$. Тогда найдется вектор $h \in \mathbf{R}^n \setminus \{0\}$ такой, что

$$A_1 h = 0, \quad \langle a_i, h\rangle = 0 \quad \forall\, i \in I. \tag{8}$$

Если предположить, что $\langle a_i, h\rangle = 0$ $\forall\, i \in J$, то $Ah = 0$, что противоречит условию $\operatorname{rank} A = n$. Поэтому найдется номер $i \in J$

такой, что $\langle a_i,\, h\rangle \neq 0$; без ограничения общности будем считать, что $\langle a_i,\, h\rangle < 0$.

Для произвольного числа $t \geqslant 0$, положив $x(t) = x + th$, в силу (8) имеем

$$A_1 x(t) = A_1 x + t A_1 h = b_1,$$

$$\langle a_i,\, x(t)\rangle = \langle a_i,\, x\rangle + t\langle a_i,\, h\rangle = (b_2)_i \quad \forall\, i \in I. \tag{9}$$

Поэтому $x(t) \in D$ тогда и только тогда, когда

$$\langle a_i,\, x(t)\rangle = \langle a_i,\, x\rangle + t\langle a_i,\, h\rangle \geqslant (b_2)_i \quad \forall\, i \in J.$$

Положим

$$t_1 = \max_{t\geqslant 0:\ x(t)\in D} t.$$

Легко видеть, что существует номер $i_1 \in J$ такой, что $\langle a_{i_1},\, h\rangle < 0$ и

$$0 < t_1 = \frac{(b_2)_{i_1} - \langle a_{i_1},\, x\rangle}{\langle a_{i_1},\, h\rangle} \tag{10}$$

(ср. с задачей 4.1.6; напомним, что $(b_2)_i - \langle a_i,\, x\rangle < 0 \quad \forall\, i \in J$).

Положим $x^1 = x(t_1)$, $I_1 = I(x^1)$; тогда из (9) следует включение

$$I \subset I_1. \tag{11}$$

Из (10) имеем

$$\langle a_{i_1},\, x^1\rangle = \langle a_{i_1},\, x\rangle + \frac{(b_2)_{i_1} - \langle a_{i_1},\, x\rangle}{\langle a_{i_1},\, h\rangle}\,\langle a_{i_1},\, h\rangle = (b_2)_{i_1},$$

т. е. $i_1 \in I_1$. Если предположить, что

$$a_{i_1} = A_1^{\mathrm{T}} y + \sum_{i\in I} \beta_i a_i$$

при некоторых $y \in \mathbf{R}^l$ и числах β_i, $i \in I$, то в силу (8)

$$0 > \langle a_{i_1},\, h\rangle = \langle y,\, A_1 h\rangle + \sum_{i\in I} \beta_i \langle a_i,\, h\rangle = 0,$$

что невозможно. Это значит, что строки матрицы A_1 и векторы a_i, $i \in I_1$, образуют в $\mathbf{R}^n$ систему ранга $r_1 > r$.

Если $r_1 = n$, то x^1 является вершиной, и в силу (11) все доказано. В противном случае нужно повторить всю процедуру, заменив точку x на x^1. В результате s-кратного повторения получим такой набор точек $x,\, x^1,\, x^2, \ldots, x^s \in D$, что

$$I(x) \subset I(x^1) \subset I(x^2) \subset \ldots \subset I(x^s),$$

$$r < r_1 < r_2 < \ldots < r_s,$$

где r_k — ранг системы строк матрицы A_1, дополненных векторами a_i, $i \in I(x^k)$, $k = 1, \ldots, s$. Ясно, что для какого-то номера s выполнится $r_s = n$, а значит, $\widehat{x} = x^s$ — вершина, причем $I(x) \subset \subset I(\widehat{x})$. $\square$

Если нужно найти все вершины полиэдра D вида (7), то можно действовать по следующей схеме. Для каждого $(n-l)$-элементного множества $I \subset \{1, \ldots, m\}$ решаем линейную систему

$$A_1 x = b_1, \quad \langle a_i, x\rangle = (b_2)_i, \quad i \in I.$$

Если решение этой системы существует, единственно и принадлежит D, то оно является вершиной D. В противном случае переходим к следующему множеству I.

Заданное в (4) допустимое множество канонической ЗЛП называется *каноническим полиэдром* (напомним, что в (4) $P = \mathbf{R}^n_+$, $A \in \in \mathbf{R}(l, n)$, $b \in \mathbf{R}^l$). Для всякой точки $x \in \mathbf{R}^n$ положим $J(x) = \{j = = 1, \ldots, n \mid x_j > 0\}$. Пусть a^j, $j = 1, \ldots, n$, — столбцы матрицы A.

Теорема 4. *Для любых $A \in \mathbf{R}(l, n)$ и $b \in \mathbf{R}^l$ точка $\widehat{x} \in D$ является вершиной заданного в (4) (при $P = \mathbf{R}^n_+$) канонического полиэдра D тогда и только тогда, когда векторы a^j, $j \in J(\widehat{x})$, линейно независимы. Если множество $J(\widehat{x})$ состоит из l элементов, то вершина $\widehat{x}$ невырожденная, а если из менее чем l элементов, то вырожденная.*

Задача 4. Доказать теорему 4.

Важнейшее свойство канонических полиэдров содержится в следующей теореме.

Теорема 5. *Для любых $A \in \mathbf{R}(l, n)$ и $b \in \mathbf{R}^l$ заданный в (4) (при $P = \mathbf{R}^n_+$) канонический полиэдр D имеет по крайней мере одну вершину. Более того, для любой точки $x \in D$ существует такая вершина $\widehat{x}$ полиэдра D, что $J(x) \supset J(\widehat{x})$.*

Задача 5. Используя теорему 3, доказать теорему 5.

Для канонического полиэдра применяется следующая специальная схема отыскания его вершин. Для каждого множества $J \subset \{1, \ldots \ldots, n\}$ такого, что векторы a^j, $j \in J$, линейно независимы, решаем линейную систему

$$\sum_{j \in J} a^j x_j = b.$$

Если эта система имеет решение с компонентами $\widehat{x}_j \geqslant 0 \quad \forall\, j \in J$, то, положив $\widehat{x}_j = 0$, $j \in \{1, \ldots, n\} \setminus J$, получим вершину $\widehat{x} = (\widehat{x}_1, \ldots \ldots, \widehat{x}_n)$. В противном случае переходим к очередному множеству J.

Следующая теорема, справедливость которой вытекает немедленно из теорем 2 и 3, важна в плане практического отыскания решений ЗЛП. Геометрическая интуиция подсказывает, что если ЗЛП имеет решение, то решением обязательно является одна из вершин допустимого множества. Учитывая, что число вершин полиэдра конечно (предложение 1), для отыскания решения достаточно просто перебрать все вершины (например, указанными выше способами), выбрав те, где значение целевой функции максимально; по крайней мере одно решение таким образом будет найдено.

Теорема 6. *Если ЗЛП имеет решение, причем ее допустимое множество имеет по крайней мере одну вершину, то по крайней мере одна из вершин допустимого множества является решением ЗЛП.*

7.1.2. Теория двойственности для линейных задач. Двойственная задача к общей ЗЛП (1), (2) (где $P = \mathbf{R}^{n_1} \times \mathbf{R}^{n_2}_+$) имеет вид

$$\langle b_1,\, \lambda\rangle + \langle b_2,\, \mu\rangle \to \max, \quad (\lambda,\, \mu) \in \mathcal{D}, \tag{12}$$

$$\mathcal{D} = \{(\lambda,\, \mu) \in \mathcal{P} \,\big|\, A_{11}^{\mathrm{T}}\lambda + A_{21}^{\mathrm{T}}\mu = c_1,\ A_{12}^{\mathrm{T}}\lambda + A_{22}^{\mathrm{T}}\mu \leqslant c_2\}, \tag{13}$$

где $\mathcal{P} = \mathbf{R}^l \times \mathbf{R}^m_+$ (см. задачу 6.1.1). Подчеркнем, что указанный вид двойственная задача имеет именно для задачи вида (1), (2) (т. е. для ЗЛП на минимум и с нужной направленностью знаков ограничений-неравенств). Заметим, что двойственная к ЗЛП сама является ЗЛП.

Например, двойственная к канонической ЗЛП (3), (4) (где $P = \mathbf{R}^n_+$) имеет вид

$$\langle b,\, \lambda\rangle \to \max, \quad \lambda \in \mathcal{D},$$

$$\mathcal{D} = \{\lambda \in \mathbf{R}^l \,\big|\, A^{\mathrm{T}}\lambda \leqslant c\};$$

такая задача называется *основной ЗЛП на максимум*. Если очевидными преобразованиями привести последнюю к виду (1), (2) и в соответствии с (12), (13) выписать двойственную к ней, то полученная задача будет совпадать с (3), (4) (для формального совпадения нужно будет поменять знак у целевой функции полученной задачи, чтобы перейти к задаче на минимум). Аналогичный факт имеет место и для общей ЗЛП. А именно, на классе ЗЛП переход к двойственной задаче есть, как говорят, *двойственная операция*: ее двукратное применение дает исходную задачу. Напомним, что для нелинейных задач это, вообще говоря, неверно: двойственной к любой (даже невыпуклой) задаче оптимизации является выпуклая (при соответствующем выборе знака целевой функции) задача (см. теорему 6.1.6).

Теорема 7. *ЗЛП* (1), (2) *и* (12), (13) *взаимодвойственны*: *двойственной к* (12), (13) *является задача* (1), (2) (*с точностью до выбора знака целевой функции и знаков ограничений-неравенств*).

Задача 6. Доказать теорему 7.

Таким образом, множество всех ЗЛП разбивается на пары взаимодвойственных задач: при необходимости задачу (12), (13) можно считать прямой, а задачу (1), (2) двойственной.

Напомним (см. (6.1.8), (6.1.9)), что

$$\langle c_1,\, x_1\rangle + \langle c_2,\, x_2\rangle \geqslant \\ \geqslant \langle b_1,\, \lambda\rangle + \langle b_2,\, \mu\rangle \quad \forall\,(x_1,\, x_2) \in D, \quad \forall\,(\lambda,\, \mu) \in \mathcal{D}, \quad (14)$$

и, в частности,

$$\overline{v} \geqslant \overline{\omega},$$

где $\overline{v} = \inf\limits_{(x_1,\, x_2)\in D} (\langle c_1,\, x_1\rangle + \langle c_2,\, x_2\rangle)$ — значение прямой задачи, а $\overline{\omega} = \sup\limits_{(\lambda,\, \mu)\in\mathcal{D}} (\langle b_1,\, \lambda\rangle + \langle b_2,\, \mu\rangle)$ — значение двойственной (здесь важно, что прямая задача является задачей на минимум, иначе знаки неравенств были бы направлены в другую сторону). Особое значение имеет случай, когда в (14) реализуется равенство, т. е. для некоторых пар $(\overline{x}_1,\, \overline{x}_2) \in D$ и $(\overline{\lambda},\, \overline{\mu}) \in \mathcal{D}$ имеет место

$$\langle c_1,\, \overline{x}_1\rangle + \langle c_2,\, \overline{x}_2\rangle = \langle b_1,\, \overline{\lambda}\rangle + \langle b_2,\, \overline{\mu}\rangle. \quad (15)$$

Из (14) следует, что при этом $(\overline{x}_1,\, \overline{x}_2)$ — решение задачи (1), (2), а $(\overline{\lambda},\, \overline{\mu})$ — решение задачи (12), (13), причем выполнено соотношение двойственности

$$\overline{v} = \overline{\omega}. \quad (16)$$

Лемма 1. *Если в условиях теоремы* 2 $\mathcal{P} = \mathbf{R}^l \times \mathbf{R}^m_+$, *а множества* D *и* $\mathcal{D}$ *заданы в* (2) *и* (13) *соответственно, то для произвольных пар* $(\overline{x}_1, \overline{x}_2) \in D$ *и* $(\overline{\lambda}, \overline{\mu}) \in \mathcal{D}$ *равенство* (15) *имеет место тогда и только тогда, когда выполнено* (6).

Доказательство. Из условий $(\overline{x}_1,\, \overline{x}_2) \in D$ и $(\overline{\lambda},\, \overline{\mu}) \in \mathcal{D}$ выводим

$$\langle c_1,\, \overline{x}_1\rangle + \langle c_2,\, \overline{x}_2\rangle \geqslant \langle A_{11}^{\mathrm{T}}\overline{\lambda} + A_{21}^{\mathrm{T}}\overline{\mu},\, \overline{x}_1\rangle + \langle A_{12}^{\mathrm{T}}\overline{\lambda} + A_{22}^{\mathrm{T}}\overline{\mu},\, \overline{x}_2\rangle = \\ = \langle \overline{\lambda},\, A_{11}\overline{x}_1 + A_{12}\overline{x}_2\rangle + \langle \overline{\mu},\, A_{21}\overline{x}_1 + A_{22}\overline{x}_2\rangle \geqslant \langle \overline{\lambda},\, b_1\rangle + \langle \overline{\mu},\, b_2\rangle.$$

Если выполнено (15), то левая часть равна правой, а значит, промежуточные неравенства выполняются как равенства. Легко видеть, что это возможно только при выполнении (6). Обратно, если выполнено (6), то оба промежуточных неравенства выполняются как равенства, что и дает (15). □

Соотношение (6) было получено как условие дополняющей нежесткости для прямой задачи. Таким образом, соотношение (15), или, что то же самое, условие дополняющей нежесткости (6), является достаточным условием оптимальности допустимых точек $(\overline{x}_1,\, \overline{x}_2)$

в прямой задаче и $(\overline{\lambda}, \overline{\mu})$ в двойственной. Чтобы показать, что это условие является и необходимым, заметим, что включение $(\overline{\lambda}, \overline{\mu}) \in \mathcal{D}$ в точности совпадает с условием (5), дополненным неравенством $\overline{\mu} \geqslant 0$. Отсюда, из леммы 1 и теоремы 2 вытекает следующий критерий оптимальности в ЗЛП.

Теорема 8. *Пусть в условиях теоремы* 2 $\mathcal{P} = \mathbf{R}^l \times \mathbf{R}^m_+$ *и множества* D *и* $\mathcal{D}$ *заданы в* (2) *и* (13) *соответственно.*

Точка $(\overline{x}_1, \overline{x}_2) \in D$ *является решением задачи* (1), (2) *тогда и только тогда, когда существует пара* $(\overline{\lambda}, \overline{\mu}) \in \mathcal{D}$, *удовлетворяющая равенству* (15), *или, что то же самое,* (6). *При этом* $(\overline{\lambda}, \overline{\mu})$ *является решением задачи* (12), (13).

Если принять во внимание теорему 7, становится ясно, что теорема 8 содержит в себе следующие две классические формулировки, которые называют теоремами двойственности для линейных задач.

Теорема 9. *Прямая ЗЛП имеет решение тогда и только тогда, когда двойственная к ней ЗЛП имеет решение. При этом значения* $\overline{v}$ *и* $\overline{\omega}$ *этих задач совпадают, т. е. выполнено соотношение двойственности* (16).

Теорема 10. *В условиях теоремы* 8 *пары* $(\overline{x}_1, \overline{x}_2) \in D$ *и* $(\overline{\lambda}, \overline{\mu}) \in \mathcal{D}$ *являются решениями взаимодвойственных задач* (1), (2) *и* (12), (13) *соответственно тогда и только тогда, когда выполнено условие дополняющей нежесткости* (6).

Наконец, из теорем 1 и 9 вытекает следующая теорема существования, позволяющая судить о существовании решения у одной из взаимодвойственных задач по виду допустимого множества другой.

Теорема 11. *Если допустимые множества взаимодвойственных ЗЛП непусты, то обе они имеют решения. Если же непусто допустимое множество лишь одной из этих задач, то ее значение бесконечно.*

Как отмечалось в п. 6.1.2, использование перехода к двойственной задаче может приносить непосредственную пользу, если двойственная задача оказывается проще прямой и если нет разрыва двойственности. Этот тезис становится особенно наглядным в контексте ЗЛП. Решив сначала двойственную ЗЛП, можно по ее решению восстановить решения прямой, например с помощью теоремы 10.

Задача 7. Используя геометрические построения, высказать гипотезу о решении задачи, двойственной к задаче

$$-7x_1 - x_3 + 4x_4 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^4_+ \,|\, -x_1 + x_2 - 2x_3 + x_4 \geqslant -6,\ -2x_1 - x_2 + x_3 \geqslant 1\}.$$

Проверить эту гипотезу и восстановить решение прямой задачи с помощью теоремы 10.

Задача 8. То же задание для задачи

$$2x_1 + 3x_2 + 2x_3 + 3x_4 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^4_+ \mid 2x_1 - x_2 - x_3 + x_4 \geqslant 2,\ -x_1 - 3x_2 + x_3 + x_4 \geqslant 1\}.$$

Задача 9. Найти все решения задачи

$$x_1 + x_2 + x_3 + x_4 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^4_+ \mid x_1 - x_2 \geqslant 0,\ x_1 + x_2 - x_3 + x_4 \geqslant 1\}.$$

Задача 10. Построить задачу, двойственную к так называемой *блочной ЗЛП*:

$$\sum_{j=1}^{s} \langle c_j,\, x_j \rangle \to \min, \quad (x_1, \ldots, x_s) \in D,$$

$$D = \left\{(x_1, \ldots, x_s) \in P \;\middle|\; A_j x_j \geqslant b_j,\ j = 1, \ldots, s,\ \sum_{j=1}^{s} B_j x_j \geqslant d\right\},$$

где $c_j \in \mathbf{R}^{n_j}$, $A \in \mathbf{R}(m_j, n_j)$, $b_j \in \mathbf{R}^{m_j}$, $B \in \mathbf{R}(m, n_j)$, $j = 1, \ldots, s$, $d \in \mathbf{R}^m$, $P = \prod_{j=1}^{s} \mathbf{R}^{n_j}_+$.

Задача 11. Построить задачу, двойственную к так называемой *динамической задаче планирования производства*:

$$\sum_{j=1}^{s} \langle c_j,\, x_j \rangle \to \min, \quad (x_1, \ldots, x_s) \in D,$$

$$D = \{(x_1, \ldots, x_s) \in P \mid Ax_j \geqslant x_{j-1},\ j = 1, \ldots, s\},$$

где $c_j \in \mathbf{R}^n$, $j = 1, \ldots, s$, $A \in \mathbf{R}(m, n)$, $x_0 \in \mathbf{R}^n$ задано, $P = \prod_{j=1}^{s} \mathbf{R}^n_+$.

Задача 12. Решить задачу

$$\langle c, x \rangle \to \min, \quad x \in D = \{x \in P \mid \langle a, x \rangle = b\},$$

где $c, a \in \mathbf{R}^n$, $b \in \mathbf{R}$, $a > 0$, $b > 0$, $P = \mathbf{R}^n_+$.

Задача 13. Решить задачу

$$\sum_{j=1}^{n} j x_j \to \min, \quad x \in D = \left\{x \in P \;\middle|\; \sum_{j=1}^{i} x_j \geqslant i,\ i = 1, \ldots, n\right\},$$

где $P = \mathbf{R}^n_+$.

Задача 14. Множество решений ЗЛП может:
а) состоять из одной точки;
б) состоять более чем из одной точки, но быть ограниченным;
в) быть неограниченным.
Привести примеры, показывающие, что для множеств решений взаимодвойственных ЗЛП возможно любое сочетание этих случаев.

Задача 15. Доказать, что допустимые множества основной ЗЛП и двойственной к ней не могут быть одновременно непустыми и ограниченными.

§ 7.2. Симплекс-метод

Симплекс-метод (СМ), о котором пойдет речь в этом параграфе, можно назвать основным численным методом решения ЗЛП, хотя это вовсе не означает, что именно его следует всегда выбирать на практике (см. § 7.4). Общепринятое название метода не слишком удачно, поскольку выражает лишь тот факт, что исторически первые задачи, к которым он применялся, имели в качестве допустимого множества стандартный симплекс.

СМ вводится для канонической ЗЛП

$$\langle c,\, x\rangle \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in P \,\big|\, Ax = b\}, \tag{2}$$

где $c \in \mathbf{R}^n$, $A \in \mathbf{R}(l,\, n)$, $b \in \mathbf{R}^l$, $P = \mathbf{R}^n_+$. Разумеется, это нисколько не ограничивает область применения метода, поскольку, как отмечалось в п. 7.1.1, любую ЗЛП можно свести к канонической.

Всюду в этом параграфе используется обозначение $J(x) = \{j = 1, \ldots, n \,\big|\, x_j > 0\}$ для множества номеров положительных компонент точки $x \in \mathbf{R}^n$. Через a^j, $j = 1, \ldots, n$, обозначаются столбцы матрицы A.

7.2.1. Общая схема симплекс-метода. В предложении 7.1.1 и теоремах 7.1.5 и 7.1.6 обоснован простейший численный метод решения ЗЛП, а именно метод полного перебора вершин допустимого множества (способы такого перебора обсуждались в п. 7.1.1): если задача (1), (2) имеет решение, то по крайней мере одно решение будет найдено таким методом за конечное число шагов.

Теоретически метод полного перебора вершин применим к любой ЗЛП, однако уже при n и l порядка нескольких десятков реализация метода требует огромного количества вычислений (см. п. 7.4.2). Таким образом, практической ценности этот метод не имеет; его значение в том, что он естественным образом приводит к основной идее СМ: полный перебор нужно заменить упорядоченным, «разумным»,

в процессе которого заведомо неоптимальные вершины исключались бы из рассмотрения.

Напомним, что по теореме 7.1.4 точка $\widehat{x} \in D$ является вершиной заданного в (2) канонического полиэдра D в том и только том случае, когда столбцы a^j, $j \in J(\widehat{x})$, линейно независимы. При этом если число элементов в множестве $J(\widehat{x})$ равно l, то $\widehat{x}$ — невырожденная вершина, а если меньше l, то вырожденная.

Заметим, что для любого $x \in D$

$$b = Ax = \sum_{j=1}^{n} a^j x_j = \sum_{j \in J(x)} a^j x_j,$$

т. е. вектор правых частей ограничений b линейно выражается через столбцы матрицы A, отвечающие положительным компонентам допустимой точки x, причем коэффициентами такого выражения являются сами эти положительные компоненты.

Определение 1. *Базисом вершины* $\widehat{x}$ заданного в (2) канонического полиэдра D называется любая система из l линейно независимых столбцов матрицы A, включающая в себя все столбцы a^j, $j \in J(\widehat{x})$.

Иными словами, система $\mathcal{B} = \{a^j,\, j \in J\}$, где $J \subset \{1, \ldots, n\}$, является базисом вершины $\widehat{x}$, если, во-первых, $\mathcal{B}$ — базис в $\mathbf{R}^l$ (это предполагает, что J состоит из l номеров), а во-вторых, $J \supset J(\widehat{x})$. Ясно, что невырожденная вершина имеет единственный базис $\mathcal{B} = \{a^j,\, j \in J(\widehat{x})\}$, а вырожденная вершина может иметь несколько базисов либо вовсе не иметь базиса.

Наличие у канонического полиэдра вырожденных вершин связано с нетипичной ситуацией, когда вектор b лежит в линейном подпространстве пространства $\mathbf{R}^l$, натянутом менее чем на l столбцов матрицы A. Почти любое малое возмущения b приводит к исчезновению вырожденных вершин.

Напомним, что в силу теоремы 7.1.5 непустой канонический полиэдр всегда имеет хотя бы одну вершину. Очевидно, произвольная вершина имеет базис в том и только том случае, когда $\operatorname{rank} A = l$. Далее считаем это предположение выполненным, чего всегда можно добиться, исключая из системы ограничений-равенств, определяющих D, «лишние» уравнения.

Симплекс-метод представляет собой вычислительную процедуру, которая специальным образом генерирует вершины x^k допустимого множества D и их базисы $\mathcal{B}_k$, $k = 0, 1, \ldots$, причем для каждого k в зависимости от знаков определенных параметров делается один из следующих трех выводов:

1) x^k — решение задачи (1), (2);

2) задача (1), (2) не имеет решений;

3) существует (конструктивно указываемая) вершина x^{k+1} множества D с базисом $\mathcal{B}_{k+1}$, причем либо

$$x^{k+1} \neq x^k, \quad \langle c,\, x^{k+1}\rangle < \langle c,\, x^k\rangle, \tag{3}$$

либо

$$x^{k+1} = x^k, \quad \mathcal{B}_{k+1} \neq \mathcal{B}_k. \tag{4}$$

Если делается вывод 1) либо 2), то алгоритм, разумеется, останавливают. Если вершина x^k невырожденная и делается вывод 3), то всегда обязательно реализуется (3). Если же вершина x^k вырождена, то может реализовываться и (4), однако процедура допускает такую организацию, которая исключает возможность циклического перебора базисов. При этом рано или поздно реализуется (3), т. е. осуществляется переход к новой («лучшей», чем текущая) вершине. Теперь ясно, почему в литературе СМ иногда называют *методом последовательного улучшения плана*; это название значительно более информативно. Ясно также, что СМ является методом активного множества: его можно интерпретировать как метод определения множества $I(\overline{x}) = \{1, \ldots, n\} \setminus J(\overline{x})$ нулевых компонент искомой оптимальной вершины $\overline{x}$ (т. е. множества индексов активных в этой точке ограничений-неравенств задачи (1), (2)), после чего ненулевые компоненты этой вершины могут быть определены из линейной системы

$$\sum_{j \in J(\overline{x})} a^j x_j = b.$$

Так как число вершин допустимого множества конечно, и при разрешимости задачи (1), (2) среди вершин обязательно есть решение, то такой процесс за конечное число шагов приведет либо к выводу 1), либо к 2). Иными словами, если задача (1), (2) разрешима, то решение будет найдено за конечное число шагов; если же задача (1), (2) не имеет решения, то этот факт будет установлен, причем также за конечное число шагов.

Подчеркнем, что теоретически на пути к решению СМ может «посетить» все вершины допустимого множества, т. е. свестись к полному перебору вершин. Известны искусственно построенные примеры, в которых это явление действительно имеет место. Однако, как показывает практический опыт, обычно число итераций СМ находится в пределах от l до $2l$, и метод позволяет за приемлемое время решать ЗЛП с размерностями n и l порядка нескольких сотен и даже тысяч.

7.2.2. Итерация симплекс-метода. Пусть вершины $x^0, x^1, \ldots, x^k$ полиэдра D уже получены (о методах поиска начальной вершины см. ниже). Пусть $\mathcal{B}_k = \{a^j,\ j \in J_k\}$ — базис вершины x^k, $J_k \subset \subset \{1, \ldots, n\}$. В этом пункте будет описана одна очередная итерация, поэтому для краткости будем опускать индекс k: $x = x^k$, $\mathcal{B} = \mathcal{B}_k$, $J = J_k$.

Поскольку $\mathcal{B}$ — базис в $\mathbf{R}^l$, столбцы матрицы A однозначно раскладываются по этой системе векторов:

$$a^i = \sum_{j \in J} \gamma_{ji} a^j, \quad i = 1, \ldots, n, \tag{5}$$

где числа γ_{ji}, $j \in J$, $i = 1, \ldots, n$, называются *коэффициентами замещения*. Определим так называемые *оценки замещения*

$$\Delta_i = c_i - \sum_{j \in J} c_j \gamma_{ji}, \quad i = 1, \ldots, n. \tag{6}$$

Заметим, что

$$\gamma_{ji} = \begin{cases} 1, & \text{если } j = i, \\ 0, & \text{если } j \in J \setminus \{i\}, \end{cases} \qquad \Delta_i = 0 \quad \forall\, i \in J,$$

и для $i \in J$ коэффициенты и оценки замещения интереса не представляют. А вот анализ знаков Δ_i и γ_{ji} при $j \in J$, $i \notin J$ как раз и позволяет сделать один из трех выводов, о которых говорилось в п. 7.2.1 (здесь и далее запись $i \notin J$ означает, что $i \in \{1, \ldots, n\} \setminus J$).

Теорема 1. *Пусть для $A \in \mathbf{R}(l, n)$ и $b \in \mathbf{R}^l$ точка $x \in D$ является вершиной заданного в* (2) (*при $P = \mathbf{R}^n_+$*) *полиэдра D, а $\mathcal{B} = \{a^j,\ j \in J\}$ — базис этой вершины, $J \subset \{1, \ldots, n\}$.*

Тогда если для вводимых согласно (5) *и* (6) *оценок замещения справедливо*

$$\Delta_i \geqslant 0 \quad \forall\, i \notin J, \tag{7}$$

то x — решение задачи (1), (2).

Доказательство. Рассмотрим основную ЗЛП, являющуюся двойственной к задаче (1), (2):

$$\langle b, \lambda \rangle \to \max, \quad \lambda \in \mathcal{D}, \tag{8}$$

$$\mathcal{D} = \{\lambda \in \mathbf{R}^l \mid A^{\mathrm{T}} \lambda \leqslant c\}. \tag{9}$$

Согласно теореме 7.1.8 достаточно указать такую точку $\lambda \in \mathcal{D}$, что

$$\langle c, x \rangle = \langle b, \lambda \rangle. \tag{10}$$

Так как $\mathcal{B}$ — базис в $\mathbf{R}^l$, то существует единственный элемент $\lambda \in \mathbf{R}^l$, удовлетворяющий линейной системе

$$\langle a^j, \lambda \rangle = c_j, \quad j \in J. \tag{11}$$

Для этого λ с учетом (5)–(7) получаем: $\forall\, i \notin J$ имеет место

$$\langle a^i,\, \lambda\rangle = \sum_{j\in J}\langle a^j,\, \lambda\rangle\gamma_{ji} = \sum_{j\in J} c_j\gamma_{ji} = c_i - \Delta_i \leqslant c_i,$$

что в совокупности с (11) дает включение $\lambda \in \mathcal{D}$.

Далее, так как $\mathcal{B}$ — базис вершины x, то $J(x) \subset J$, т. е. $x_j = 0$ $\forall\, j \notin J$. Отсюда, из (11) и включения $x \in D$ получаем

$$\langle c,\, x\rangle = \sum_{j=1}^{n}\langle a^j,\, \lambda\rangle x_j = \Big\langle \sum_{j=1}^{n} a^j x_j,\, \lambda\Big\rangle = \langle Ax,\, \lambda\rangle = \langle b,\, \lambda\rangle,$$

т. е. выполнено (10). □

Из приведенного доказательства следует, что если выполнено (7), то решение двойственной задачи (8), (9) может быть найдено как решение линейной системы (11).

Лемма 1. *В условиях теоремы* 1 *для произвольных номера* $s \notin J$ *и числа* $t \geqslant 0$ *точка* $x(t) \in \mathbf{R}^n$ *с компонентами*

$$x_j(t) = \begin{cases} x_j - t\gamma_{js}, & \text{если } j \in J, \\ t, & \text{если } j = s, \\ 0, & \text{если } j \notin J, j \neq s, \end{cases} \tag{12}$$

удовлетворяет равенствам

$$Ax(t) = b, \tag{13}$$

$$\langle c, x(t)\rangle = \langle c, x\rangle + t\Delta_s. \tag{14}$$

Доказательство. В силу (5), (6) и (12) имеем

$$Ax(t) = \sum_{j=1}^{n} a^j x_j(t) = \sum_{j\in J} a^j(x_j - t\gamma_{js}) + ta^s =$$

$$= \sum_{j\in J} a^j x_j + t\Big(a^s - \sum_{j\in J} a^j\gamma_{js}\Big) = \sum_{j=1}^{n} a^j x_j = Ax = b,$$

$$\langle c,\, x(t)\rangle = \sum_{j=1}^{n} c_j x_j(t) = \sum_{j\in J} c_j(x_j - t\gamma_{js}) + tc_s =$$

$$= \sum_{j\in J} c_j x_j + t\Big(c_s - \sum_{j\in J} c_j\gamma_{js}\Big) = \sum_{j=1}^{n} c_j x_j + t\Delta_s = \langle c,\, x\rangle + t\Delta_s,$$

что и дает (13), (14). □

Теорема 2. *Пусть выполнены условия теоремы* 1.

Тогда если для вводимых согласно (5) *и* (6) *коэффициентов и оценок замещения существует номер* $s \notin J$ *такой, что*

$$\Delta_s < 0, \quad \gamma_{js} \leqslant 0 \quad \forall\, j \in J, \tag{15}$$

то задача (1), (2) *не имеет решения.*

Доказательство. Из (14) и (15) для точки $x(t)$, компоненты которой задаются формулой (12), имеем

$$x(t) \geqslant 0 \quad \forall\, t \geqslant 0,$$

$$\langle c,\, x(t)\rangle + t\Delta_s \to -\infty \quad (t \to +\infty).$$

Вместе с (13) эти два соотношения означают, что $\overline{v} = -\infty$. $\square$

Задача 1. Доказать теорему 2 без использования леммы 1, опираясь на теорию двойственности.

Теперь предположим, что существуют номера $s \notin J$ и $j \in J$ такие, что

$$\Delta_s < 0, \quad \gamma_{js} > 0. \tag{16}$$

Общая идея итерации СМ такова: нужно так изменить базис текущей вершины, чтобы в отвечающей новому базису вершине значение целевой функции задачи было меньше, чем в текущей. Положим

$$\widetilde{t} = \max_{t \geqslant 0:\, x(t) \geqslant 0} t. \tag{17}$$

Из (12) ясно, что

$$\widetilde{t} = \min_{j \in J:\, \gamma_{js} > 0} \frac{x_j}{\gamma_{js}}\,. \tag{18}$$

Будем считать, что минимум в последнем равенстве достигается на номере r, т. е.

$$r \in J, \quad \gamma_{rs} > 0, \quad \frac{x_r}{\gamma_{rs}} = \widetilde{t}. \tag{19}$$

Введем множество

$$\widetilde{J} = (J \setminus \{r\}) \cup \{s\}. \tag{20}$$

Теорема 3. *Пусть выполнены условия теоремы* 1.

Тогда если для вводимых согласно (5) *и* (6) *коэффициентов и оценок замещения существуют номера* $s \notin J$ *и* $j \in J$ *такие, что выполнено* (16), *то точка* $\widetilde{x} = x(\widetilde{t})$, *вычисленная по формулам* (12), (18), *является вершиной полиэдра* D, *а система* $\widetilde{\mathcal{B}} = \{a^j,\, j \in \widetilde{J}\}$, *где множество* $\widetilde{J}$ *задано в* (19), (20), *является ее базисом. При этом если* $\widetilde{t} > 0$, *то* $\langle c, \widetilde{x}\rangle < \langle c, x\rangle$; *если же* $\widetilde{t} = 0$, *то* $\widetilde{x} = x$, *но* $\widetilde{\mathcal{B}} \neq \mathcal{B}$.

Доказательство. Из (12), (17) и равенства в (19) следует, что $\widetilde{x} \geqslant 0$ и $\widetilde{x}_r = x_r - \widetilde{t}\gamma_{rs} = 0$. Следовательно, привлекая лемму 1

и (20), получаем: $\widetilde{x} \in D$ и

$$J(\widetilde{x}) \subset \widetilde{J}. \tag{21}$$

Кроме того, из (5) имеем

$$a^s = \sum_{j \in J} \gamma_{js} a^j = \sum_{j \in J \setminus \{r\}} \gamma_{js} a^j + \gamma_{rs} a^r,$$

причем согласно (19) $\gamma_{rs} > 0$, поэтому

$$a^r = \frac{1}{\gamma_{rs}}\, a^s - \sum_{j \in J \setminus \{r\}} \frac{\gamma_{js}}{\gamma_{rs}}\, a^j. \tag{22}$$

Отсюда, из (20) и того, что $\mathcal{B}$ — базис в $\mathbf{R}^l$, следует, что $\widetilde{\mathcal{B}}$ — тоже базис в $\mathbf{R}^l$. Но тогда согласно (21) и теореме 7.1.4 точка $\widetilde{x}$ является вершиной D, а $\widetilde{\mathcal{B}}$ — ее базисом.

Последнее утверждение теоремы следует немедленно из (12), (14) и (16). $\square$

Таким образом, новый базис $\widetilde{\mathcal{B}}$ отличается от старого $\mathcal{B}$ лишь тем, что в нем вместо столбца a^r присутствует a^s. Говорят, что столбец a^r «выводится из базиса», а столбец a^s «вводится в базис». Коэффициент замещения γ_{rs} иногда называют *ведущим элементом*.

Если x — невырожденная вершина, то $J(x) = J$, т. е. $x_j > 0$ $\forall\, j \in J$. Но тогда из (18) следует, что $\widetilde{t} > 0$, поэтому на такой итерации СМ обязательно реализуется (3), т. е. происходит переход к новой, «лучшей» вершине.

Если же x — вырожденная вершина, то, возможно, найдется номер $j \in J$ такой, что $x_j = 0$ и $\gamma_{js} > 0$. Тогда из (12) и (18) следует, что $\widetilde{t} = 0$, $\widetilde{x} = x$, т. е. перехода к новой вершине не произойдет, и результатом итерации будет лишь смена базиса $\mathcal{B}$ вершины x на $\widetilde{\mathcal{B}}$. Подчеркнем, что это тоже нетривиальная операция, поскольку для нового базиса пересчитываются коэффициенты и оценки замещения, и, возможно, на следующей итерации поменяется и сама вершина. Однако может возникнуть ситуация, когда итерации СМ сведутся к перебору базисов одной вершины, которые, в силу того, что их конечное число, с определенного момента начнут повторяться, т. е. произойдет *зацикливание* метода. Известны примеры, в которых это явление действительно имеет место. На практике же зацикливание возникает крайне редко. Напомним, что само наличие у допустимого множества задачи вырожденных вершин — «редкая» ситуация, однако и для вырожденной вершины обычно рано или поздно находится базис, который позволяет перейти к новой вершине.

Кроме того, итерацию СМ можно организовать так, чтобы полностью исключить возможность зацикливания, и такие меры предосторожности принято предпринимать, поскольку СМ должен быть абсолютно надежен, как средство многократного решения вспомогательных ЗЛП, возникающих при реализации численных методов решения более общих задач оптимизации. Суть таких реализаций СМ состоит в специальных правилах выбора столбцов для вывода из базиса и для ввода в базис в тех случаях, когда «претендентов» на эти роли несколько. Одним из простейших правил такого рода является так называемое *правило Блэнда*, согласно которому при выполнении неравенств (16) следует брать наименьшие подходящие s и r:

$$s = \min_{i \notin J:\, \Delta_i < 0} i, \qquad r = \min_{j \in J:\, \gamma_{js} > 0,\, x_j/\gamma_{js} = \tilde{t}} j$$

(см. [3]).

Таким образом, описана итерация СМ. На следующей итерации процедура повторяется для вершины $\widetilde{x}$ и ее базиса $\widetilde{\mathcal{B}}$. Для этого сначала необходимо вычислить соответствующие коэффициенты $\widetilde{\gamma}_{ji}$ и оценки $\widetilde{\Delta}_i$ замещения, $j \in \widetilde{J}$, $i = 1, \ldots, n$, что, в принципе, требует решения соответствующих линейных систем (см. (5)). Однако последнего можно избежать, поскольку коэффициенты и оценки замещения на двух последовательных итерациях связаны друг с другом явными формулами. А именно, используя равенства (5), (6) и (22), легко убедиться, что $\forall\, i = 1, \ldots, n$

$$\widetilde{\gamma}_{ji} = \begin{cases} \gamma_{ji} - \gamma_{js}\gamma_{ri}/\gamma_{rs}, & \text{если } j \in J \setminus \{r\}, \\ \gamma_{ri}/\gamma_{rs}, & \text{если } j = s, \end{cases} \qquad \widetilde{\Delta}_i = \Delta_i - \Delta_s \frac{\gamma_{ri}}{\gamma_{rs}}.$$

Задача 2. Вывести формулы, связывающие коэффициенты и оценки замещения на двух последовательных итерациях СМ.

Задача 3. Решить задачу

$$x_1 + 2x_2 + 4x_3 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^3_+ \,\big|\, 2x_1 + 3x_2 + 4x_3 = 10,\ x_1 + x_2 - x_3 = 4\},$$

методом полного перебора вершин.

Задача 4. Из вершины $x = (0,\, 1,\, 3)$ сделать для задачи

$$-x_1 - x_2 - x_3 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^3_+ \,\big|\, -2x_1 + x_2 + x_3 = 4,\ x_1 - 2x_2 + x_3 = 1\},$$

одну итерацию СМ.

Задача 5. То же задание для задачи

$$-x_1 + 2x_2 - x_3 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^3_+ \mid x_1 - x_2 + 2x_3 = 0,\ x_1 + x_2 + 5x_3 = 2\},$$

и вершины $x = (1,\ 1,\ 0)$.

Задача 6. Стартуя из вершины $x = (1,\ 0,\ 1,\ 0)$, решить задачу

$$-x_1 + 3x_2 + 5x_3 + x_4 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^4_+ \mid x_1 + 4x_2 + 4x_3 + x_4 = 5,\ x_1 + 7x_2 + 8x_3 + 2x_4 = 9\},$$

с помощью СМ.

Задача 7. То же задание для задачи

$$-3x_1 - x_2 - x_3 + x_4 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^4_+ \mid x_1 - x_2 + x_3 = 1,\ 2x_1 + x_2 + x_4 = 3\},$$

и вершины $x = (0,\ 0,\ 1,\ 3)$.

В заключение обсудим вопрос о поиске начальной вершины, которая вместе с некоторым ее базисом должна быть указана для начала работы СМ. Иногда сделать это нетрудно. Рассмотрим, например, так называемую *стандартную ЗЛП*

$$\langle c,\ x\rangle \to \min, \quad x \in D, \quad D = \{x \in P \mid Ax \geqslant b\},$$

где, напомним, $c \in \mathbf{R}^n$, $A \in \mathbf{R}(l,\ n)$, $b \in \mathbf{R}^l$, $P = \mathbf{R}^n_+$, причем будем предполагать, что $b \geqslant 0$. Эта задача сводится к следующей канонической ЗЛП:

$$\langle c,\ x\rangle \to \min, \quad (x,\ y) \in \widetilde{D},$$

$$\widetilde{D} = \{z = (x,\ y) \in \widetilde{P} \mid Ax - y = b\}, \tag{23}$$

где $\widetilde{P} = \mathbf{R}^n_+ \times \mathbf{R}^m_+$. Легко видеть, что точка $z = (0,\ b)$ является вершиной канонического полиэдра $\widetilde{D}$, а стандартный базис пространства $\mathbf{R}^m$ является базисом этой вершины, причем очень удобным.

В общем же случае поиск начальной вершины оказывается сравнимым по трудоемкости с самой процедурой СМ. Рассмотрим один метод построения начальной вершины, называемый *методом искусственного базиса*, который напрямую использует СМ.

Без ограничения общности считаем, что в задаче (1), (2) $b \geqslant 0$; этого всегда можно добиться, меняя знаки у ограничений-равенств. Введем вспомогательную ЗЛП

$$\sum_{i=1}^{l} y_i \to \min, \quad (x,\ y) \in \widetilde{D}, \tag{24}$$

где множество $\widetilde{D}$ введено в (23). Тогда, как отмечено выше, известна вершина $z = (0,\ b)$ полиэдра $\widetilde{D}$ и ее базис. Стартуя из этой вершины, будем решать задачу (23) посредством СМ. Очевидно, целевая

функция задачи (24) ограничена снизу (числом 0) на непустом множестве $\widetilde{D}$, поэтому согласно теореме 7.1.1 задача (24) имеет решение. Поэтому СМ остановится в решении $z^0 = (x^0, y^0)$, которое является вершиной полиэдра $\widetilde{D}$.

Возможны два случая. Если $y^0 \neq 0$, то значение задачи (24) положительно, и допустимое множество D задачи (1), (2) пусто. Действительно, если предположить, что существует точка $x \in D$, то $z = (x, 0) \in \widetilde{D}$, причем значение целевой функции задачи (24) в точке z равно нулю. Но это противоречит тому, что z^0 является решением задачи (24). Если же $y^0 = 0$, то x^0 — вершина множества D, так как это равносильно тому, что $(x^0, 0)$ — вершина множества $\widetilde{D}$.

Метод искусственного базиса можно снабдить процедурой, позволяющей одновременно указывать и базис начальной вершины x^0.

Процедуры поиска вершин полиэдров находят различные применения, в том числе и не связанные с ЗЛП. Например, с помощью таких процедур можно определять множители Лагранжа, отвечающие известному решению задачи оптимизации, поскольку множество множителей есть полиэдр (а при отсутствии ограничений-равенств — канонический полиэдр). В контексте ЗКП этот вопрос будет обсуждаться в § 7.3.

З а д а ч а 8. Доказать, что если в условиях теоремы 1 вершина x невырожденная, то справедливо утверждение, обратное утверждению этой теоремы: если x является решением задачи (1), (2), то выполнено условие (7), причем если x — единственное решение, то $\Delta_i > 0 \ \ \forall\, i \notin J$. Привести пример, показывающий, что для вырожденной вершины такое утверждение может не иметь места.

§ 7.3. Методы решения задач квадратичного программирования

Основной интерес к ЗКП объясняется теми же причинами, что и интерес к ЗЛП: такие задачи возникают в качестве вспомогательных при реализации многих численных методов решения более общих задач. Яркими примерами являются обсуждавшиеся в § 4.4 и § 5.4 методы последовательного квадратичного программирования, которые, безусловно, относятся к числу наиболее эффективных современных оптимизационных алгоритмов общего назначения, а также многошаговые методы негладкой выпуклой оптимизации из § 6.3. Эффективность таких методов существенно зависит от того, насколько эффективно решаются вспомогательные ЗКП.

Как и ЗЛП, выпуклая ЗКП допускает конечные методы решения. Ниже излагается один такой метод активного множества, иногда называемый методом особых точек. Подчеркнем, что для выпуклых

ЗКП существуют также и эффективные бесконечношаговые алгоритмы, например методы внутренней точки (см. § 7.4). Об этих и других методах решения ЗКП см. [50].

Будем рассматривать ЗКП вида

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in \mathbf{R}^n \mid Ax \leqslant b\}, \tag{2}$$

где

$$f\colon \mathbf{R}^n \to \mathbf{R}, \quad f(x) = \frac{1}{2}\langle Cx, x\rangle + \langle c, x\rangle,$$

$C \in \mathbf{R}(n, n)$ — симметрическая положительно определенная матрица, $c \in \mathbf{R}^n$, $A \in \mathbf{R}(m, n)$, $b \in \mathbf{R}^m$. Специальный вид ограничений не снижает общности, так как любую ЗКП можно свести к задаче с ограничениями такого вида. Кроме того, описываемый ниже метод можно модифицировать так, чтобы он был применим и в случае неотрицательно, но не обязательно положительно определенной матрицы C (о таких модификациях см., например, [34]).

Будем предполагать, что $D \neq \varnothing$. Тогда задача (1), (2) имеет единственное решение, поскольку целевая функция этой задачи сильно выпукла на выпуклом множестве D.

7.3.1. Особые точки. Пусть a_i, $i = 1, \ldots, m$, — строки матрицы A. Центральную роль в излагаемом ниже методе играет следующее понятие.

Определение 1. Точка $\widehat{x} \in D$ называется *особой точкой* задачи (1), (2), если существует множество $I \subset \{1, \ldots, m\}$ такое, что $\widehat{x}$ является решением задачи

$$f(x) \to \min, \quad x \in D_I, \tag{3}$$

$$D_I = \{x \in \mathbf{R}^n \mid \langle a_i, x\rangle = b_i,\ i \in I\} \tag{4}$$

(случаи $I = \varnothing$ и $I = \{1, \ldots, m\}$ не исключаются).

Теорема 1. *Для всякой симметрической положительно определенной матрицы $C \in \mathbf{R}(n, n)$ и любых $c \in \mathbf{R}^n$, $A \in \mathbf{R}(m, n)$ и $b \in \mathbf{R}^m$ число особых точек задачи* (1), (2) *конечно, причем решение этой задачи является ее особой точкой.*

Доказательство. Число задач вида (3), (4) конечно, причем каждая из них имеет сильно выпуклую целевую функцию, а значит, не более одного решения (решений нет для тех I, для которых $D_I = \varnothing$). Это дает конечность числа особых точек.

Пусть $\overline{x} \in \mathbf{R}^n$ — решение задачи (1), (2). Очевидно, $\overline{x}$ является локальным решением задачи

$$f(x) \to \min, \quad x \in \overline{D},$$

$$\overline{D} = \{x \in \mathbf{R}^n \mid \langle a_i, x\rangle \leqslant b_i,\ i \in I(\overline{x})\},$$

где $I(\overline{x}) = \{i = 1, \ldots, m \mid \langle a_i, \overline{x}\rangle = b_i\}$ — множество номеров активных ограничений в точке $\overline{x}$. Но эта задача выпукла, поэтому $\overline{x}$ является ее глобальным решением, а значит, и решением задачи (3), (4) при $I = I(\overline{x})$. $\square$

Из доказанной теоремы вытекает, что для отыскания решения задачи (1), (2) достаточно перебрать все особые точки этой задачи, выбрав ту, где значение целевой функции минимально. Для отыскания всех особых точек задачи (1), (2) достаточно для каждого множества $I \subset \{1, \ldots, m\}$, для которого $D_I \neq \varnothing$, решить задачу (3), (4) и проверить допустимость полученной точки в задаче (1), (2).

Что касается отыскания решений задач вида (3), (4) (с ограничениями-равенствами), то всякая такая задача может быть многими способами сведена к задаче безусловной минимизации сильно выпуклой квадратичной функции. Для этого можно использовать операцию проектирования на множество D_I (см. утверждение е) из задачи 4.1.3), представление точек множества D_I через некоторый базис линейного подпространства $\{x \in \mathbf{R}^n \mid \langle a_i, x\rangle = 0,\ i \in I\}$, а также двойственную релаксацию (подробности см. в [37, 50]). Задача безусловной минимизации сильно выпуклой квадратичной функции решается за конечное число шагов, например, методом сопряженных градиентов (п. 3.3.1) либо методами вычислительной линейной алгебры.

Описанный метод полного перебора особых точек находит решение задачи (1), (2) за конечное число шагов. Однако, как и обсуждавшийся в пп. 7.1.1, 7.2.1 метод полного перебора вершин для ЗЛП, метод полного перебора особых точек для ЗКП не имеет практической ценности: число различных задач вида (3), (4) равно 2^m, и реализация метода требует огромного количества вычислений уже при небольших n и m. Поэтому, как и при введении симплекс-метода, возникает идея заменить полный перебор упорядоченным, что и будет сделано ниже.

Задача 1. Решить задачу

$$x_1^2 + x_2^2 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^2 \mid -x_1 - x_2 \leqslant -4,\ x_1 \leqslant 5,\ x_2 \leqslant 3\},$$

методом полного перебора особых точек.

Задача 2. То же задание для задачи

$$(x_1 - 1)^2 + (x_2 + 1)^2 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^2 \mid -x_1 + x_2 \leqslant 1,\ x_1 + x_2 \leqslant 1,\ x_2 \geqslant 0\}.$$

7.3.2. Метод особых точек. *Метод особых точек* генерирует особые точки x^k, $k = 0, 1, \ldots$, задачи (1), (2), причем так, что значение функции f в каждой последующей точке меньше, чем в предыдущей. Согласно теореме 1, этот процесс не может продолжаться бесконечно и неминуемо прервется в решении задачи (1), (2) после конечного числа шагов.

Пусть уже получены особые точки $x^0, x^1, \ldots, x^k$. Требуется проверить, является ли точка x^k решением, и если нет, то найти такую особую точку x^{k+1} задачи (1), (2), что

$$f(x^{k+1}) < f(x^k). \tag{5}$$

Это делается в два этапа.

Цель первого этапа — найти любую (не обязательно особую) точку $\widetilde{x}^k \in \mathbf{R}^n$ такую, что

$$f(\widetilde{x}^k) < f(x^k), \quad \widetilde{x}^k \in D, \tag{6}$$

либо установить отсутствие такой точки (последнее будет означать, что x^k — решение задачи (1), (2)). Это построение можно осуществить многими способами. Например, достаточно сделать из точки x^k для задачи (1), (2) один шаг простейшего метода возможных направлений (см. § 4.2).

С учетом линейности ограничений задачи (1), (2) вместо (4.2.5), (4.2.6) будем рассматривать упрощенную вспомогательную ЗЛП

$$\langle f'(x^k),\, d\rangle = \langle Cx^k + c,\, d\rangle \to \min, \quad d \in D_k, \tag{7}$$

$$D_k = \{d \in \mathbf{R}^n \,\big|\, \langle a_i,\, d\rangle \leqslant 0,\ i \in I(x^k),\ -1 \leqslant d_j \leqslant 1,\ j = 1, \ldots, n\}. \tag{8}$$

Пусть d^k — решение этой задачи, а $v_k = \langle f'(x^k),\, d\rangle$ — ее значение. Если $v_k = 0$, то процесс останавливают.

Задача 3. Пусть $C \in \mathbf{R}(n,\, n)$ — симметрическая неотрицательно определенная матрица, $c \in \mathbf{R}^n$, $A \in \mathbf{R}(m,\, n)$, $b \in \mathbf{R}^m$.

Показать, что если в точке $x^k \in D$, где множество D введено в (2), значение задачи (7), (8) равно нулю, то x^k — решение задачи (1), (2).

Если $v_k < 0$, то, как следует из лемм 3.1.1, 4.1.3 и утверждения из задачи 4.1.5, вектор d^k удовлетворяет условию $d^k \in \mathcal{D}_f(x^k) \cap \cap\, \mathcal{F}_D(x^k)$. Поэтому условия (6) будут выполнены при $\widetilde{x}^k = x^k + \alpha d^k$ для любого достаточно малого числа $\alpha > 0$. Конкретное значение $\alpha = \alpha_k$ может выбираться различными способами, скажем, дроблением заданного числа $\widehat{\alpha} > 0$ до тех пор, пока не выполнится (6). Представляется разумным брать в качестве α_k решение одномерной задачи

$$f(x^k + \alpha d^k) \to \min, \quad \alpha \in [0,\, \widehat{\alpha}_k], \tag{9}$$

где

$$\widehat{\alpha}_k = \max_{\alpha \geqslant 0:\, x^k + \alpha d^k \in D} \alpha = \min_{i=1,\ldots,m:\, \langle a_i,\, d^k\rangle > 0} \frac{b_i - \langle a_i,\, x^k\rangle}{\langle a_i,\, d^k\rangle} > 0 \qquad (10)$$

(см. задачу 4.1.6; имеется в виду, что если $\langle a_i,\, d^k\rangle \leqslant 0 \quad \forall\, i = 1, \ldots$ $\ldots, m$, то $\widehat{\alpha}_k = +\infty$). Используя квадратичность функции f и положительную определенность C, получаем явную формулу для решения задачи (9):

$$\alpha_k = \min\left\{ -\frac{\langle Cx^k + c,\, d^k\rangle}{\langle Cd^k,\, d^k\rangle},\ \widehat{\alpha}_k \right\}.$$

На втором этапе (если не произошла остановка на первом) ищется особая точка x^{k+1} задачи (1), (2), удовлетворяющая неравенству

$$f(x^{k+1}) \leqslant f(\widetilde{x}^k). \qquad (11)$$

Для этого строится вспомогательная последовательность точек

$$\xi^0,\ \widetilde{\xi}^0,\ \xi^1,\ \widetilde{\xi}^1,\ \ldots,\ \xi^s,\ \widetilde{\xi}^s,\ \ldots \qquad (12)$$

в $\mathbf{R}^n$ по следующему правилу.

Полагаем $\xi^0 = \widetilde{x}^k$. Пусть уже получена точка $\xi^s \in D$; тогда в качестве $\widetilde{\xi}^s$ берется решение задачи (3), (4) при $I = I(\xi^s)$. Точка ξ^s допустима в такой задаче, поэтому

$$f(\widetilde{\xi}^s) \leqslant f(\xi^s). \qquad (13)$$

Возможны два случая.

Если $\widetilde{\xi}^s \in D$, то $\widetilde{\xi}^s$ — особая точка задачи (1), (2) по определению. В этом случае процесс построения последовательности (12) прерывается: полагаем $x^{k+1} = \widetilde{\xi}^s$.

Если же $\widetilde{\xi}^s \notin D$, то в качестве ξ^{s+1} берется точка отрезка, соединяющего ξ^s и $\widetilde{\xi}^s$, ближайшая к $\widetilde{\xi}^s$ среди всех точек этого отрезка, допустимых в задаче (1), (2):

$$\xi^{s+1} = \xi^s + t_s(\widetilde{\xi}^s - \xi^s), \qquad (14)$$

$$t_s = \max_{t \in [0,\,1]:\, \xi^s + t_s(\widetilde{\xi}^s - \xi^s) \in D} t = \min_{i=1,\ldots,m:\, \langle a_i,\, \widetilde{\xi}^s\rangle > b_i} \frac{b_i - \langle a_i,\, \xi^s\rangle}{\langle a_i,\, \widetilde{\xi}^s - \xi^s\rangle} \qquad (15)$$

(последнее равенство легко проверяется; ср. с (10)). При этом из соотношений (13)–(15) и выпуклости функции f имеем

$$f(\xi^{s+1}) \leqslant t_s f(\widetilde{\xi}^s) + (1 - t_s) f(\xi^s) \leqslant f(\xi^s). \qquad (16)$$

Далее, очевидно, что $I(\xi^{s+1}) \subset I(\xi^s)$. Кроме того, пусть минимум в правой части (15) достигается на индексе $i = i_s$. Тогда $i_s \notin$ $\notin I(\xi^s)$ и

$$\langle a_{i_s},\, \xi^{s+1}\rangle = \langle a_{i_s},\, \xi^s\rangle + \frac{b_{i_s} - \langle a_{i_s},\, \xi^s\rangle}{\langle a_{i_s},\, \widetilde{\xi}^s - \xi^s\rangle} \langle a_{i_s},\, \widetilde{\xi}^s - \xi^s\rangle = b_{i_s},$$

т. е. $i_s \in I(\xi^{s+1})$, и, в частности, множество $I(\xi^{s+1})$ содержит по крайней мере на один элемент больше, чем $I(\xi^s)$. Но все эти множества содержатся в $\{1, \ldots, m\}$, поэтому при некотором s неминуемо реализуется случай $\widetilde{\xi}^s \in D$, т. е. процесс построения последовательности (12) будет прерван после конечного числа шагов. Для полученной особой точки $x^{k+1} = \widetilde{\xi}^s$ в силу (13), (16) имеем

$$f(x^{k+1}) = f(\widetilde{\xi}^s) \leqslant f(\xi^s) \leqslant f(\xi^{s-1}) \leqslant \ldots \leqslant f(\xi^0) = f(\widetilde{x}^k),$$

т. е. выполнено (11). Из неравенства в (6) и (11) получаем (5). Итерация метода полностью описана.

Как и для симплекс-метода, нужно еще указать способ построения начальной особой точки x^0. Для этого достаточно найти произвольную точку $\widetilde{x}^{-1} \in D$ (например, с помощью обсуждавшегося в п. 7.2.2 метода искусственного базиса), а потом воспользоваться рассмотренным выше алгоритмом второго этапа.

В некоторых случаях оказывается полезным применять описанный метод в сочетании с двойственной релаксацией. Согласно результату задачи 6.1.2 двойственной к (1), (2) является задача

$$-\frac{1}{2}\langle AC^{-1}A^{\mathrm{T}}\mu,\, \mu\rangle - \langle AC^{-1}c + b,\, \mu\rangle - \frac{1}{2}\langle C^{-1}c,\, c\rangle \to \max, \quad \mu \in \mathbf{R}^m_+, \tag{17}$$

причем если $\overline{\mu}$ — решение двойственной задачи, то единственное решение прямой дается формулой

$$\overline{x} = -C^{-1}(c + A^{\mathrm{T}}\overline{\mu}). \tag{18}$$

Таким образом, достаточно решить двойственную ЗКП (17), а затем восстановить решение прямой ЗКП (1), (2) по явной формуле (18). Заметим, что ограничения задачи (17) очень просты, и конструкция метода особых точек для такой ЗКП существенно упрощается [37]. В то же время матрица $AC^{-1}A^{\mathrm{T}}$, будучи неотрицательно определенной, является положительно определенной лишь при дополнительном условии $\operatorname{rank} A = m$. С другой стороны, как отмечалось выше, метод особых точек можно распространить и на ЗКП с выпуклыми, но не обязательно сильно выпуклыми целевыми функциями.

Задача 4. Для допустимой точки $\widetilde{x} = (0,\, 3)$ задачи

$$(x_1 - 6)^2 + x_2^2 \to \min,$$

$$x \in D = \{x \in \mathbf{R}^2 \,\big|\, x_1 + x_2 \leqslant 4,\ 3x_1 + 2x_2 \leqslant 8,\ -4x_1 + x_2 \leqslant 4\},$$

применить алгоритм построения особой точки без увеличения значения целевой функции (второй этап итерации метода особых точек).

В заключение отметим, что часто (например, при реализации методов последовательного квадратичного программирования; см. § 4.4) требуется найти не только само решение $\overline{x} \in \mathbf{R}^n$ ЗКП (1) и (2),

но и отвечающие этому решению множители Лагранжа. Множество таких множителей имеет вид

$$\mathcal{M}(\overline{x}) = \Big\{\overline{\mu} \in \mathbf{R}^m_+ \Big| \sum_{i=1}^{m} \overline{\mu}_i a_i = -C\overline{x} - c,\ \overline{\mu}_i = 0,\ i \in \{1, \ldots, m\} \setminus I(\overline{x})\Big\}$$

и является каноническим полиэдром. Точку (вершину) этого полиэдра можно найти, например, методом искусственного базиса; см. п. 7.2.2.

§ 7.4. Методы внутренней точки

Методы внутренней точки (МВТ) привлекают большое внимание специалистов с середины 80-х годов прошлого века. Первоначальный энтузиазм был обусловлен главным образом тем, что эти методы обладают так называемым свойством полиномиальной трудоемкости, в отличие, скажем, от симплекс-метода (СМ), трудоемкость которого экспоненциальна (см. ниже). Обоснованность этого интереса была подтверждена развитием весьма совершенных алгоритмов внутренней точки, которые оказались на практике не только конкурентоспособными, но и более эффективными, чем СМ, при решении ЗЛП большой размерности (скажем, начиная с размерностей порядка 10^4). Кроме того, МВТ естественным образом распространяются на нелинейные задачи выпуклого программирования, в частности на выпуклые ЗКП. В этом параграфе предлагается очень краткое изложение основных идей МВТ; сколько-нибудь полное изложение потребовало бы написания отдельной книги.

7.4.1. Барьеры. В определенной степени МВТ (особенно прямые) могут рассматриваться как реализация идеи использования так называемых барьерных функций, которую и предлагается обсудить в первую очередь.

Рассмотрим задачу

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$x \in D = \{x \in P \,|\, G(x) \leqslant 0\}, \tag{2}$$

где множество $P \subset \mathbf{R}^n$, функция $f\colon P \to \mathbf{R}$ и отображение $G\colon P \to \mathbf{R}^m$ с компонентами $g_i(\cdot)$, $i = 1, \ldots, m$, заданы. Определим множество

$$D^0 = \{x \in P \,|\, G(x) < 0\}, \tag{3}$$

и будем предполагать, что

$$D^0 \neq \varnothing, \quad \inf_{x \in D^0} f(x) = \overline{v} > -\infty, \tag{4}$$

где $\overline{v} = \inf\limits_{x \in D} f(x)$ — значение задачи (1), (2).

Идея методов барьеров, как и методов штрафов, состоит в замене задачи (1), (2) последовательностью задач, в которых функциональные ограничения отсутствуют (или могут игнорироваться). В данном случае эта идея реализуется за счет добавления к целевой функции слагаемого, которое бесконечно растет при приближении к $D \setminus D^0$.

А именно, функция $\psi\colon D^0 \to \mathbf{R}$ называется *барьером* (*внутренним штрафом*) для заданного в (2) множества D, если она непрерывна на D^0, причем для любой последовательности $\{x^k\} \in D^0$ такой, что $g_i(x^k) \to 0$ хотя бы для одного $i \in \{1, \ldots, m\}$, выполнено $\psi(x^k) \to +\infty \quad (k \to \infty)$. Соответствующее семейство *барьерных функций* имеет вид

$$\varphi_\sigma\colon D^0 \to \mathbf{R}, \quad \varphi_\sigma(x) = f(x) + \sigma\psi(x), \tag{5}$$

где $\sigma > 0$ — *барьерный параметр*. Задача (1), (2) аппроксимируется задачами вида

$$\varphi_\sigma(x) \to \min, \quad x \in D^0. \tag{6}$$

Наиболее часто используются *логарифмический барьер*

$$\psi(x) = -\sum_{i=1}^{m} \ln(-g_i(x)), \quad x \in D^0, \tag{7}$$

и *обратный барьер*

$$\psi(x) = -\sum_{i=1}^{m} \frac{1}{g_i(x)}, \quad x \in D^0. \tag{8}$$

Задачу (6) можно считать задачей без функциональных ограничений в следующем смысле. Ясно, что если просто игнорировать формальные функциональные ограничения (строгие неравенства) в определении D_0 и решать задачу

$$\varphi_\sigma(x) \to \min, \quad x \in P,$$

любым подходящим итерационным методом, использующим начальную точку из D^0, то при любом $\sigma > 0$ можно ожидать, что вся траектория метода, как и ее предельные точки, останется в D^0. Вместе с тем, решения задачи (1), (2) обычно лежат в $D \setminus D^0$, поэтому для обеспечения итоговой аппроксимации решения барьерный параметр σ нужно устремить к нулю. Опишем *метод барьеров*.

Алгоритм 1. Для заданного в (3) множества D^0 фиксируем барьер $\psi\colon D^0 \to \mathbf{R}$. Выбираем монотонно убывающую последовательность $\{\sigma_k\} \subset \mathbf{R}_+ \setminus \{0\}$ и полагаем $k = 0$.

1. Вычисляем $x^k \in D^0$ как глобальное решение задачи (6) с целевой функцией, задаваемой формулой (5) при $\sigma = \sigma_k$.

2. Увеличиваем номер шага k на 1 и переходим к п. 1.

Последовательность $\{\sigma_k\}$ может выбираться не заранее, а адаптивно, в ходе процесса.

Теорема 1. *Пусть функция* $f\colon P \to \mathbf{R}$ *полунепрерывна снизу на* P, *а отображение* $G\colon P \to \mathbf{R}^m$ *непрерывно на* P, $P \subset \mathbf{R}^n$. *Пусть для заданного в* (3) *множества* D^0 *выполнено* (4), *где* $\overline{v} = \inf\limits_{x \in D} f(x)$.

Тогда если траектория $\{x^k\}$ *сгенерирована алгоритмом* 1, *то*

$$f(x^{k+1}) \leqslant f(x^k), \quad \psi(x^{k+1}) \geqslant \psi(x^k) \quad \forall\, k, \tag{9}$$

и если $\sigma_k \to 0$ $(k \to \infty)$, *то любая предельная точка последовательности* $\{x^k\}$ *является глобальным решением задачи* (1), (2).

Доказательство. Для всякого k из (5) и определения x^{k+1} имеем

$$f(x^{k+1}) + \sigma_{k+1}\psi(x^{k+1}) = \varphi_{\sigma_{k+1}}(x^{k+1}) \leqslant$$
$$\leqslant \varphi_{\sigma_{k+1}}(x^k) = f(x^k) + \sigma_{k+1}\psi(x^k). \tag{10}$$

С другой стороны, из (5) и определения x^k имеем

$$f(x^k) + \sigma_k\psi(x^k) = \varphi_{\sigma_k}(x^k) \leqslant \varphi_{\sigma_k}(x^{k+1}) = f(x^{k+1}) + \sigma_k\psi(x^{k+1}).$$

Складывая соответственно левые и правые части двух последних формул и группируя слагаемые, приходим к неравенству

$$(\sigma_k - \sigma_{k+1})\psi(x^k) \leqslant (\sigma_k - \sigma_{k+1})\psi(x^{k+1}),$$

которое в силу монотонного убывания последовательности $\{\sigma_k\}$ дает второе неравенство в (9). Но тогда, вновь используя (10), получаем

$$f(x^{k+1}) - f(x^k) \leqslant \sigma_{k+1}(\psi(x^k) - \psi(x^{k+1})) \leqslant 0,$$

а это — первое неравенство в (9).

Теперь предположим, что $\overline{x}$ — предельная точка последовательности $\{x^k\}$, $\{x^{k_j}\} \to \overline{x}$ $(j \to \infty)$. Из (2), (3), включения $\{x^k\} \subset D^0$ и непрерывности G следует, что $\overline{x} \in D$.

Если $f(\overline{x}) = \overline{v}$, то все доказано. Пусть $f(\overline{x}) > \overline{v}$. Введем число

$$\delta = \frac{1}{2}\,(f(\overline{x}) - \overline{v}) > 0. \tag{11}$$

Тогда из второго соотношения в (4) следует существование такого элемента $\widetilde{x} \in D^0$, что $f(\widetilde{x}) < \overline{v} + \delta$. Последовательность $\{f(x^k)\}$ невозрастающая (в силу первого неравенства в (9)) и ограничена снизу (в силу включения $\{x^k\} \subset D^0$ и второго соотношения в (4)), а значит, сходится, причем из полунепрерывности снизу функции f вытекает, что для всякого j

$$f(x^{k_j}) \geqslant \lim_{k \to \infty} f(x^k) \geqslant f(\overline{x}) = \overline{v} + 2\delta,$$

где также принято во внимание (11). Поэтому в силу выбора $\widetilde{x}$

$$f(x^{k_j}) - f(\widetilde{x}) \geqslant \overline{v} - f(\widetilde{x}) + 2\delta > \delta.$$

Отсюда, из определения x^{k_j}, включения $\widetilde{x} \in D^0$, формулы (5) и второго соотношения в (9) следует, что если $\sigma_k \to 0 \quad (k \to \infty)$, то для любого достаточно большого j

$$0 \geqslant \varphi_{\sigma_{k_j}}(x^{k_j}) - \varphi_{\sigma_{k_j}}(\widetilde{x}) = f(x^{k_j}) + \sigma_{k_j}\psi(x^{k_j}) - f(\widetilde{x}) - \sigma_{k_j}\psi(\widetilde{x}) >$$

$$> \delta + \sigma_{k_j}(\psi(x^{k_0}) - \psi(\widetilde{x})) > 0,$$

что невозможно. □

Задача 1. Пусть $P \subset \mathbf{R}^n$ — выпуклое множество, функция $f\colon P \to \mathbf{R}$ и компоненты g_i отображения $G\colon P \to \mathbf{R}^m$ выпуклы и дифференцируемы на P. Для каждого $\sigma > 0$ обозначим через $x(\sigma)$ произвольное решение задачи (6) с целевой функцией, задаваемой формулой (5), где барьер ψ задан в (7). Доказать оценку

$$f(x(\sigma)) \leqslant \overline{v} + m\sigma.$$

Для случая барьера ψ, заданного в (8), доказать оценку

$$f(x(\sigma)) \leqslant \overline{v} - \sigma \sum_{i=1}^{m} \frac{1}{g_i(x(\sigma))}.$$

Для всякого $k = 1, 2, \ldots$ в качестве начального приближения для метода решения задачи (6) при $\sigma = \sigma_k$ естественно брать точку x^{k-1}, как и в методе штрафов (см. п. 4.3.2). Однако с практической точки зрения перспектива решать точно (или во всяком случае достаточно точно) бесконечную последовательность вспомогательных задач (6) выглядит не слишком реалистично, даже несмотря на то, что вспомогательные задачи существенно проще исходной. Более разумным представляется воспользоваться идеей конечных алгоритмов продолжения по параметру (см. п. 5.3.1), т. е. решать задачу (6) для данного значения параметра $\sigma = \sigma_k$ приближенно, ограничившись лишь несколькими шагами некоторого эффективного итерационного метода из точки x^{k-1}. Сказанное выражает сущность прямых МВТ.

Подчеркнем, что, в отличие от методов продолжения, обсуждавшихся в § 5.3, здесь речь идет не о построении достаточно хорошего приближения к решению исходной задачи, а об аппроксимации самого решения. Это приводит к необходимости бесконечно измельчать «сетку» по мере приближения σ к нулю. В результате возникает схема, которую можно трактовать как обычный итерационный метод и, в частности, говорить о его сходимости и скорости сходимости.

7.4.2. Некоторые замечания о трудоемкости алгоритмов. Под *трудоемкостью* алгоритма понимается зависимость наибольшего возможного количества вычислений (итераций, арифметических операций, битовых операций), необходимого алгоритму для отыскания точного решения задачи рассматриваемого класса, от размерности задачи. Если эта зависимость оценивается сверху полиномом (показательной функцией), то говорят о *полиномиальной* (соответственно *экспоненциальной*) *трудоемкости*. Подчеркнем, что речь идет об оценках в наихудшем возможном случае, т. е. в терминологии исследования операций, о «гарантированном результате».

Оказывается, МВТ обладают полиномиальной трудоемкостью. Вообще, эти методы близки по своей природе к типичным алгоритмам нелинейного программирования. Напротив, СМ имеет ярко выраженную комбинаторную природу, и его трудоемкость экспоненциальна. Не вдаваясь в формальный анализ, ограничимся некоторыми замечаниями о понятии трудоемкости.

Рассмотрим ЗЛП, например, каноническую:

$$\langle c,\, x\rangle \to \min, \quad x \in D, \tag{12}$$

$$D = \{x \in \mathbf{R}^n_+ \,\big|\, Ax = b\}, \tag{13}$$

где $c \in \mathbf{R}^n$, $A \in \mathbf{R}(l,\, n)$, $b \in \mathbf{R}^l$ заданы. Количество вычислений обычно оценивают в терминах *размера задачи* (12), (13); так называют число

$$s = q + n + 1,$$

где q — общее число бит, используемое для представления элементов c, A и b в предположении о целочисленности этих данных (рациональные данные можно сводить к целочисленным за счет масштабирования). Смысл введения числа s состоит в следующем: 2^s — очень большое число, превосходящее результат любой арифметической операции с данными задачи. Соответственно 2^{-s} — очень маленькое число, обладающее следующими свойствами, которые можно доказать формально. Если $\widehat{x}$ является вершиной полиэдра D, то, во-первых, $\widehat{x}_j \notin (0,\, 2^{-s})$ $\forall\, j = 1, \ldots, n$, а во-вторых, $\langle c, \widehat{x}\rangle - \overline{v} \notin (0, 2^{-2s})$, где $\overline{v} = \inf\limits_{x \in D} \langle c, x\rangle$ — значение задачи (12), (13).

Напомним, что СМ конечен, т. е. находит точное решение за конечное число шагов. Как будет видно из излагаемого ниже, МВТ этим свойством не обладают: они генерируют приближения, которые аппроксимируют решение «изнутри» допустимого множества, в то время как точное решение обычно находится на «границе». Однако по достаточно хорошему приближению точное решение может быть

восстановлено с помощью так называемой *процедуры уточнения* [1)], которая для произвольного $x \in D$ вычисляет такую вершину $\widehat{x}$ полиэдра D, что $\langle c, \widehat{x}\rangle \leqslant \langle c, x\rangle$, причем не более чем за $O(n^3)$ арифметических операций. Разумеется, эта процедура играет чисто теоретическую роль и не используется в практических вычислениях; она служит для того, чтобы сделать СМ и МВТ сравнимыми.

Таким образом, при выводе оценок трудоемкости МВТ считают, что метод останавливается при выполнении в текущей точке x неравенства

$$\langle c, x\rangle - \overline{v} < 2^{-2s}. \tag{14}$$

Когда говорят о полиномиальной трудоемкости МВТ, то имеют в виду следующее: количество вычислений, необходимое методу для достижения (14), оценивается сверху полиномом относительно n и s. В силу сказанного выше после достижения (14) процедура уточнения позволяет найти вершину, являющуюся точным решением, и оценка количества вычислений при этом по-прежнему будет полиномиальна. Разумеется, для разных вариантов МВТ получаются разные оценки. Для количества итераций известны оценки порядка $O(\sqrt{n}s)$, а для количества арифметических операций — порядка $O(n^{3.5}s)$.

В то же время ясно, что количество вершин полиэдра D может экспоненциально зависеть от n (например, параллелепипед в $\mathbf{R}^n$ имеет 2^n вершин), и, как отмечалось в п. 7.2.1, СМ может посетить все вершины D прежде, чем будет найдено решение. Существуют примеры, в которых это действительно имеет место, и количество итераций СМ растет экспоненциально с ростом n.

Необходимо, однако, подчеркнуть, что хотя благоприятные оценки трудоемкости весьма желательны, они никоим образом не являются критерием качества практического поведения алгоритма на разумных (а не придуманных искусственно) задачах. (В этой связи обращаем внимание на вероятностный анализ трудоемкости СМ [43], в рамках которого уже удается получить полиномиальные оценки.) Нередко методы с лучшими свойствами трудоемкости проигрывают по эффективности теоретически более трудоемким методам. Это одна из причин, по которым вопросы трудоемкости не обсуждаются здесь более детально. Существуют МВТ с полиномиальной трудоемкостью, которые совершенно непрактичны в сравнении с СМ. В то же время удачные МВТ (особенно прямодвойственные) действительно эффективны, и в настоящее время именно они часто выбираются для решения ЗЛП большой размерности.

[1)] Общепринятый английский термин — Purification procedure.

7.4.3. Прямые методы внутренней точки для линейных задач. Опишем МВТ для канонической ЗЛП (12), (13), что, напомним, не ограничивает общности. Здесь будет удобно положить $P = \{x \in \mathbf{R}^n \mid Ax = b\}$, а условия неотрицательности переменных рассматривать как функциональные ограничения. Будем предполагать, что заданное в (12) множество D ограничено, причем $D^0 \neq \varnothing$, где в данном случае

$$D^0 = \{x \in P \mid x > 0\}. \tag{15}$$

Пусть, кроме того, $\operatorname{rank} A = l$, чего, напомним, всегда можно добиться.

Определим (*прямую*) *центральную траекторию* (ЦТ) как отображение $\chi\colon (\mathbf{R}_+ \setminus \{0\}) \to D^0$, сопоставляющее каждому значению $\sigma > 0$ барьерного параметра решение задачи (6), где

$$\varphi_\sigma\colon \{x \in \mathbf{R}^n \mid x > 0\} \to \mathbf{R}, \quad \varphi_\sigma(x) = \langle c,\, x\rangle - \sigma \sum_{j=1}^{n} \ln x_j. \tag{16}$$

Привлекая утверждение из задачи 1.1.5, а также тот факт, что в сделанных предположениях введенная в (16) функция φ_σ сильно выпукла на D^0, легко видеть, что для всякого $\sigma > 0$ решение задачи (6) существует и единственно, а значит, ЦТ корректно определена.

Подчеркнем, что естественно двигаться вдоль ЦТ от $\sigma = +\infty$ к $\sigma = 0$. В связи с этим ЦТ обычно доопределяют при $\sigma = +\infty$ следующим образом. Задаче (6) с целевой функцией из (16) сопоставим задачу

$$\sum_{j=1}^{n} \ln x_j - \frac{1}{\sigma} \langle c,\, x\rangle \to \max, \quad x \in D^0,$$

которая при $\sigma > 0$ имеет то же решение, что и (6). Переходя к пределу при $\sigma \to +\infty$, получаем задачу

$$\sum_{j=1}^{n} \ln x_j \to \max, \quad x \in D^0,$$

решение которой и берут в качестве $\chi(+\infty)$. Заметим, что эта начальная точка ЦТ, называемая *аналитическим центром* заданного в (13) множества D, зависит только от A и b, но не от c.

Заметим также, что для $\sigma > 0$ точку $\chi(\sigma)$ можно рассматривать как аналитический центр пересечения множества D с гиперплоскостью, на которой целевая функция задачи (12), (13) принимает значение $\langle c,\, \chi(\sigma)\rangle$: $\chi(\sigma)$ является решением задачи

$$\sum_{j=1}^{n} \ln x_j \to \max, \quad x \in D^0_\sigma,$$

$$D^0_\sigma = \{x \in D^0 \mid \langle c, x\rangle = \langle c, \chi(\sigma)\rangle\}.$$

Можно показать, что $\lim\limits_{\sigma\to 0+} \chi(\sigma)$ существует и является аналитическим центром множества решений задачи (12), (13). Этот факт здесь не доказывается, поскольку уже теорема 1 убеждает, что движение вдоль ЦТ — вполне разумная стратегия, особенно если будет предложен эффективный способ такого движения, не связанный с многократным вычислением точных (или «почти» точных) значений отображения χ.

Опишем базовую схему (*прямых*) *методов внутренней точки.*

Алгоритм 2. Выбираем число $\sigma_0 > 0$. Для заданного в (15) множества D^0 выбираем $x^{-1} \in D^0$, и полагаем $k = 0$.

1. Стартуя из точки x^{k-1}, определяем $x^k \in D^0$ как результат нескольких шагов некоторого внутреннего (обычно ньютоновского) метода для задачи (6) с целевой функцией, задаваемой формулой (16) при $\sigma = \sigma_k$.

2. Выбираем число $\sigma_{k+1} \in (0, \sigma_k)$.

3. Увеличиваем номер шага k на 1 и переходим к п. 1.

Подразумевается, что точка x^{-1} должна быть достаточно близка к $\chi(\sigma_0)$, а назначаемое количество шагов внутреннего метода обеспечивает близость очередной точки x^k к $\chi(\sigma_k)$. Подчеркнем, что существуют алгоритмы с полиномиальной трудоемкостью, которые позволяют для любого $\sigma_0 > 0$ аппроксимировать $\chi(\sigma_0)$ с любой требуемой точностью.

Вообще, алгоритм 2 не является алгоритмом в полном смысле: он представляет собой лишь базовую схему МВТ. Для получения конкретного алгоритма в рамках этой схемы необходимо конкретизировать ряд важных позиций. Например, большое значение имеет скорость убывания σ_k с ростом k. Если последовательность барьерных параметров будет убывать медленно, то для каждого k точка $\chi(\sigma_k)$ будет близка к $\chi(\sigma_{k-1})$, и тогда можно надеяться, что небольшое количество шагов внутреннего метода (или даже один шаг) из точки x^{k-1} приведет в точку x^k, близость которой к $\chi(\sigma_k)$ достаточна для обеспечения общей сходимости процесса. Ниже будет показано, что это действительно имеет место. Вместе с тем, такая стратегия может потребовать большого количества внешних итераций для того, чтобы сама точка $\chi(\sigma_k)$ оказалась достаточно близка к решению задачи (12), (13). Альтернативой является быстрое уменьшение барьерного параметра, за что приходится платить увеличением количеством

внутренних итераций. В первом случае говорят о *методах с коротким шагом* [1], а во втором — о *методах с длинным шагом* [2]. Ниже обсуждаются именно методы с коротким шагом, которые обладают лучшими оценками трудоемкости. Заметим, однако, что удачные реализации методов с длинным шагом на практике обычно выигрывают в сравнении с методами с коротким шагом.

Конкретный МВТ определяется также выбором той или иной параметризации центральной траектории, инициализацией (т. е. выбором σ_0 и способом определения x^{-1}), способом управления погрешностью аппроксимации точек центральной траектории (т. е. количеством шагов внутреннего метода), масштабированием (см. ниже) и правилами изменения всех параметров, участвующих в реализации метода. Кроме того, возможны другие определения центральной траектории в зависимости от того, какой барьер используется.

Пусть $x \in D^0$ — текущее приближение к решению задачи (6). Ньютоновский шаг для этой задачи определим так: следующее приближение $\widetilde{x} \in \mathbf{R}^n$ ищется как решение задачи

$$\langle \varphi'_\sigma(x),\, \xi - x\rangle + \frac{1}{2}\,\langle \varphi''_\sigma(x)(\xi - x),\, \xi - x\rangle \to \min, \quad \xi \in P. \tag{17}$$

Если отбросить в задаче (6) формальные функциональные ограничения, т. е. условия положительности переменных, то, с одной стороны, это итерация условного метода Ньютона с единичным параметром длины шага (см. п. 4.1.3), а с другой стороны (с учетом линейности ограничений, описывающих множество P) — итерация метода последовательного квадратичного программирования (см. п. 4.4.1).

Всюду далее $e = (1, \ldots, 1) \in \mathbf{R}^n$, и для заданного $x \in \mathbf{R}^n$ через $X \in \mathbf{R}(n, n)$ обозначается диагональная матрица с диагональю $(x_1, \ldots, x_n)$. Введем отображения

$$\lambda\colon (\mathbf{R}_+ \setminus \{0\}) \times D^0 \to \mathbf{R}^l, \quad \lambda(\sigma, x) = (AX^2A^{\mathrm{T}})^{-1}AX\,(Xc - \sigma e), \tag{18}$$

$$d\colon (\mathbf{R}_+ \setminus \{0\}) \times D^0 \to \mathbf{R}^n, \quad d(\sigma, x) = \frac{1}{\sigma}\,X\,(c - A^{\mathrm{T}}\lambda(\sigma, x)) - e. \tag{19}$$

Задача 2. Пусть для заданных числа $\sigma > 0$ и $x \in D^0$ точка $\widetilde{x} \in \mathbf{R}^n$ является решением задачи (17), где функция φ_σ задана в (16), $c \in \mathbf{R}^n$, $A \in \mathbf{R}(l, n)$, $b \in \mathbf{R}^l$, $P = \{x \in \mathbf{R}^n \mid Ax = b\}$, а множество D^0 задано в (15). Пусть векторы $d(\sigma, x) \in \mathbf{R}^n$ и $\lambda(\sigma, x) \in \mathbf{R}^l$ определены в соответствии с (18), (19). Показать, что

$$\widetilde{x} = x - X\,d(\sigma, x),$$

[1] Общепринятый английский термин — Short-step methods.

[2] Общепринятый английский термин — Long-step methods.

причем $\lambda(\sigma, x)$ является решением задачи

$$\left| \frac{1}{\sigma} X(c - A^{\mathrm{T}}y) - e \right| \to \min, \quad y \in \mathbf{R}^l.$$

Показать, что $d(\sigma, x) = 0$ тогда и только тогда, когда x — решение задачи (6).

Таким образом, вектор $-d(\sigma, x) = X^{-1}(\widetilde{x} - x)$ можно рассматривать как масштабированный (матрицей X^{-1}) ньютоновский шаг для задачи (6) из точки x. Как показывается ниже, величина $|d(x, \sigma)|$ может служить мерой близости x к $\chi(\sigma)$.

Следующий результат относится скорее не к МВТ, а к основной схеме методов барьеров (т. е. к алгоритму 1) применительно к ЗЛП (12), (13). С другой стороны, он позволяет установить, с какой точностью нужно решать подзадачи МВТ для обеспечения его общей сходимости.

Предложение 1. *Пусть* $c \in \mathbf{R}^n$, $A \in \mathbf{R}(l, n)$, $b \in \mathbf{R}^l$, $P = \{x \in \mathbf{R}^n \mid Ax = b\}$, *а множество* D^0 *задано в* (15). *Пусть отображения* λ *и* d *определяются формулами* (18), (19).

Тогда, если для заданных числа $\sigma > 0$ *и точки* $x \in D^0$ *выполнено* $|d(\sigma, x)| < 1$, *то*

$$\langle c, x\rangle - \overline{v} \leqslant \langle c, x\rangle - \langle b, \lambda(\sigma, x)\rangle \leqslant \sigma(n + \sqrt{n}\,|d(\sigma, x)|) < \sigma(n + \sqrt{n}), \tag{20}$$

где $\overline{v} = \inf\limits_{x \in D} \langle c, x\rangle$.

Доказательство. Согласно (19)

$$\left| \frac{X(c - A^{\mathrm{T}}\lambda(\sigma, x))}{\sigma} - e \right| < 1,$$

откуда следует, что $X(c - A^{\mathrm{T}}\lambda(\sigma, x)) > 0$. Отсюда и из условия $x \in D^0$ вытекает, что

$$c - A^{\mathrm{T}}\lambda(\sigma, x) > 0. \tag{21}$$

Тогда для любого решения $\overline{x}$ задачи (12), (13)

$$\overline{v} = \langle c, \overline{x}\rangle \geqslant \langle A^{\mathrm{T}}\lambda(\sigma, x), \overline{x}\rangle = \langle \lambda(\sigma, x), A\overline{x}\rangle = \langle \lambda(\sigma, x), b\rangle,$$

где учтено включение $\overline{x} \in D$. Тем самым доказано первое неравенство в (20).

Далее, в силу неравенства Коши–Буняковского–Шварца

$$\left\langle e, \frac{1}{\sigma} X(c - A^{\mathrm{T}}\lambda(\sigma, x)) - e \right\rangle \leqslant$$

$$\leqslant |e| \left| \frac{1}{\sigma} X(c - A^{\mathrm{T}}\lambda(\sigma, x)) - e \right| = \sqrt{n}|d(\sigma, x)| < \sqrt{n},$$

а с другой стороны,

$$\left\langle e,\ \frac{1}{\sigma}\,X(c - A^{\mathrm{T}}\lambda(\sigma,\ x)) - e\right\rangle = \frac{1}{\sigma}\,\langle x,\ c - A^{\mathrm{T}}\lambda(\sigma,\ x)\rangle - |e|^2 =$$

$$= \frac{1}{\sigma}\,(\langle c,\ x\rangle - \langle b,\ \lambda(\sigma,\ x)\rangle) - n,$$

где учтено включение $x \in P$. Объединяя последние два соотношения, получаем два последних неравенства в (20). $\square$

Напомним, что рассматриваемая точка x допустима в прямой задаче (12), (13). В то же время в условиях предложения 1 выполнено (21), а значит, $\lambda(\sigma,\ x)$ — допустимая точка двойственной к (12), (13) задачи

$$\langle b,\ \lambda\rangle \to \max, \quad \lambda \in \mathcal{D}, \tag{22}$$

$$\mathcal{D} = \{\lambda \in \mathbf{R}^l \,\big|\, A^{\mathrm{T}}\lambda \leqslant c\}. \tag{23}$$

При этом (20) содержит готовую оценку близости $\langle c, x\rangle$ и $\langle b, \lambda(\sigma, x)\rangle$, причем разность этих величин стремится к нулю при $\sigma \to 0$ (ср. с теоремой 7.1.8).

Теперь убедимся, что множество $\{x \in D^0 \,\big|\, |d(\sigma,\ x)| < 1\}$ содержится в области квадратичной сходимости описанного ньютоновского метода для вспомогательной задачи (6). После этого будет показано, как следует изменять барьерный параметр, чтобы на следующем внешнем шаге остаться в области быстрой сходимости для очередной подзадачи.

Предложение 2. *В условиях предложения* 1, *если для заданных числа* $\sigma > 0$ *и точки* $x \in D^0$ *выполнено* $|d(\sigma,\ x)| < 1$, *то для точки* $\widetilde{x} = x - X\,d(\sigma,\ x)$ *справедливо*

$$\widetilde{x} \in D^0, \quad |d(\sigma,\ \widetilde{x})| \leqslant |d(\sigma,\ x)|^2. \tag{24}$$

Доказательство. Положим $z = X(c - A^{\mathrm{T}}\lambda(\sigma,\ x))/\sigma$, тогда согласно (19) $d(\sigma,\ x) = z - e$. Из условия $|d(\sigma,\ x)| < 1$ вытекает, что $0 < z_j < 2 \ \ \forall\, j = 1, \ldots, n$. Кроме того, $\widetilde{x} = x - X(z - e) = 2x - X\,z$, поэтому $\widetilde{x}_j = (2 - z_j)x_j > 0 \ \ \forall\, j = 1, \ldots, n$. В силу первого утверждения из задачи 2 $\widetilde{x} \in P$, а значит, выполнено включение в (24).

Далее, $\lambda(\sigma,\ \widetilde{x})$ является решением задачи

$$\left|\frac{1}{\sigma}\,\widetilde{X}(c - A^{\mathrm{T}}y) - e\right| \to \min, \quad y \in \mathbf{R}^l$$

(см. задачу 2), поэтому

$$|d(\sigma,\ \widetilde{x})| = \left|\frac{1}{\sigma}\,\widetilde{X}(c - A^{\mathrm{T}}\lambda(\sigma,\ \widetilde{x})) - e\right| \leqslant$$

$$\leqslant \left|\frac{1}{\sigma}\,\widetilde{X}(c - A^{\mathrm{T}}\lambda(\sigma,\ x)) - e\right| = |\widetilde{X}X^{-1}z - e|.$$

Вспоминая, что $\widetilde{x} = 2x - X\,z$, имеем

$$\widetilde{X}X^{-1}z = (2X - Z\,X)X^{-1}z = 2z - Z\,z.$$

Последние два соотношения дают

$$|d(\sigma,\,\widetilde{x})|^2 \leqslant |2z - Zz - e|^2 = \sum_{j=1}^{n}(2z_j - z_j^2 - 1)^2 =$$

$$= \sum_{j=1}^{n}(z_j - 1)^4 \leqslant \left(\sum_{j=1}^{n}(z_j - 1)^2\right)^2 = |z - e|^4 = |d(\sigma,\,x)|^4,$$

а это и есть неравенство в (24). $\square$

Этот результат показывает, что если величина $|d(\sigma,\,x)|$ достаточно мала, то после ньютоновского шага величина $|d(\sigma,\,\widetilde{x})|$ будет еще (значительно) меньше. Поэтому, если новое значение $\widetilde{\sigma}$ барьерного параметра не слишком отличается от σ, то можно ожидать, что величина $|d(\widetilde{\sigma},\,\widetilde{x})|$ также будет мала. Точнее, справедливо следующее

Предложение 3. *В условиях предложения* 1, *если для заданных числа* $\sigma > 0$ *и точки* $x \in D^0$ *выполнено* $|d(\sigma,\,x)| = \delta < 1$, *то при любом* $\theta \in (0,\,\sqrt{n})$ *для числа*

$$\widetilde{\sigma} = \left(1 - \frac{\theta}{\sqrt{n}}\right)\sigma \in (0,\,\sigma) \tag{25}$$

и точки $\widetilde{x} = x - X\,d(\sigma,\,x)$ *справедливо*

$$|d(\widetilde{\sigma},\,\widetilde{x})| \leqslant \frac{\delta^2 + \theta}{1 - \theta/\sqrt{n}}\,.$$

В частности, если $\theta \leqslant \delta(1-\delta)/(1+\delta)$, *то*

$$|d(\widetilde{\sigma},\,\widetilde{x})| \leqslant \delta.$$

Доказательство. Из (19), неравенства в (24), (25) и того факта, что $\lambda(\widetilde{\sigma},\,\widetilde{x})$ является решением задачи

$$\left|\frac{1}{\widetilde{\sigma}}\,\widetilde{X}(c - A^{\mathrm{T}}y) - e\right| \to \min, \quad y \in \mathbf{R}^l$$

(см. задачу 2), имеем

$$|d(\widetilde{\sigma},\,\widetilde{x})| = \left|\frac{1}{\widetilde{\sigma}}\,\widetilde{X}(c - A^{\mathrm{T}}\lambda(\widetilde{\sigma},\,\widetilde{x})) - e\right| \leqslant$$

$$\leqslant \left|\frac{1}{\widetilde{\sigma}}\,\widetilde{X}(c - A^{\mathrm{T}}\lambda(\sigma,\,\widetilde{x})) - e\right| = \left|\frac{\widetilde{X}(c - A^{\mathrm{T}}\lambda)(\sigma,\,\widetilde{x})}{(1 - \theta/\sqrt{n})\sigma} - e\right| =$$

$$= \left|\frac{d(\sigma,\,\widetilde{x}) + e}{1 - \theta/\sqrt{n}} - e\right| = \frac{|d(\sigma,\,\widetilde{x}) + \theta e/\sqrt{n}|}{1 - \theta/\sqrt{n}} \leqslant$$

$$\leqslant \frac{|d(\sigma,\,\widetilde{x})| + \theta|e|/\sqrt{n}}{1 - \theta/\sqrt{n}} \leqslant \frac{|d(\sigma,\,x)|^2 + \theta}{1 - \theta/\sqrt{n}} = \frac{\delta^2 + \theta}{1 - \theta/\sqrt{n}},$$

что и требовалось. $\square$

Итак, если выбрать точку x^{-1} достаточно близкой к $\chi(\sigma_0)$ и адекватно управлять барьерными параметрами σ^k, то при каждом k одного шага ньютоновского метода будет достаточно для аппроксимации $\chi(\sigma_k)$. В частности, всегда можно заставить генерируемые точки x^k быть настолько близкими к центральной траектории, насколько требуется. Как уже говорилось выше, платой за это является необходимость уменьшать барьерный параметр достаточно медленно (см. формулу (25) в предложении 3). Тем не менее приведенные факты концептуально важны и дают средства для разработки различных практических реализаций метода. Еще раз отметим, что аккуратно сконструированные методы с длинным шагом обычно оказываются предпочтительнее.

Важный момент, не отраженный в алгоритме 2, связан с масштабированием. Близость текущей точки $x^{k-1} \in D^0$ к границе неотрицательного ортанта может приводить к необходимости делать очень короткие шаги, а также к проблемам с вычислительной устойчивостью, поскольку матрица X^{k-1} будет близка к врожденной (см. формулы (18), (19)). Эти проблемы снимаются, если масштабировать задачу, например, следующим образом. Пространство $\mathbf{R}^n$ формально преобразуем линейным оператором $(X^{k-1})^{-1}$; элемент x^{k-1} при этом переходит в $\xi^{k-1} = e$. Вычисляем $-d(\sigma_k, \xi^{k-1})$ как ньютоновский шаг для масштабированной задачи (6) при $\sigma = \sigma_k$ (заметим, что $\Xi^{k-1} = E^n$). Полагаем $\xi^k = \xi^{k-1} - d(\sigma_k, \xi^{k-1})$, после чего обращаем масштабирование: $x^k = X^k \xi^k$. Суть этой процедуры состоит в том, чтобы «отодвинуть» текущую точку так далеко от границы $\mathbf{R}^n_+$, чтобы получить достаточно «пространства для действия». Кроме того, некоторые формулы для барьеров и их производных после масштабирования упрощаются.

7.4.4. Прямодвойственные методы внутренней точки для линейных задач. Выпишем даваемую теоремой 7.1.2 систему необходимых и достаточных условий оптимальности для задачи (12), (13), считая все ее ограничения функциональными:

$$A^{\mathrm{T}}\lambda + z = c, \quad Ax = b, \tag{26}$$

$$x \geqslant 0, \quad z \geqslant 0, \quad \langle z, x\rangle = 0 \tag{27}$$

относительно $(x, \lambda, z) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$; двойственная переменная z отвечает условию неотрицательности переменных прямой задачи. Идея прямодвойственных методов внутренней точки состоит в таком преобразовании приведенной системы, которое позволило бы освободиться от необходимости иметь дело с негладкой (комбинаторной) природой условия дополняющей нежесткости в (27). Для этого заменим (27) условиями

$$x_j > 0, \quad z_j > 0, \quad x_j z_j = \sigma, \quad j = 1, \ldots, n,$$

где $\sigma > 0$ — параметр (подразумевается, что этот параметр будет стремиться к нулю). Тогда, отбрасывая строгие неравенства, вместо (26), (27) имеем систему

$$A^{\mathrm{T}}\lambda + z = c, \quad Ax = b, \quad XZe = \sigma e. \tag{28}$$

Двойственная задача (22), (23) может быть записана в виде

$$\langle b,\, \lambda\rangle \to \max, \quad (\lambda,\, z) \in \widetilde{\mathcal{D}}, \tag{29}$$

$$\widetilde{\mathcal{D}} = \{(\lambda,\, z) \in \mathbf{R}^l \times \mathbf{R}^n_+ \,\big|\, A^{\mathrm{T}}\lambda + z = c\}. \tag{30}$$

Для фиксированного $\sigma > 0$ рассмотрим вспомогательную задачу

$$\varphi_\sigma(x,\, z) \to \min, \quad (x,\, (\lambda,\, z)) \in D^0 \times \widetilde{\mathcal{D}}^0, \tag{31}$$

где D^0 введено в (15) при $P = \{x \in \mathbf{R}^n \,\big|\, Ax = b\}$,

$$\widetilde{\mathcal{D}}^0 = \{(\lambda,\, z) \in \widetilde{\mathcal{P}} \,\big|\, z > 0\}, \tag{32}$$

$$\widetilde{\mathcal{P}} = \{(\lambda,\, z) \in \mathbf{R}^l \times \mathbf{R}^n \,\big|\, A^{\mathrm{T}}\lambda + z = c\}, \tag{33}$$

$\varphi_\sigma\colon \{x \in \mathbf{R}^n \,\big|\, x > 0\} \times \{z \in \mathbf{R}^n \,\big|\, z > 0\} \to \mathbf{R}$,

$$\varphi_\sigma(x,\, z) = \langle x,\, z\rangle - \sigma \sum_{j=1}^{n} \ln(z_j x_j). \tag{34}$$

Прямодвойственная центральная траектория определяется как отображение $(\chi(\cdot),\, (\nu(\cdot),\, \zeta(\cdot)))\colon (\mathbf{R}_+ \setminus \{0\}) \to D^0 \times \widetilde{\mathcal{D}}^0$, сопоставляющее каждому $\sigma > 0$ решение задачи (31).

Задача 3. Пусть $c \in \mathbf{R}^n$, $A \in \mathbf{R}(l,\, n)$, $b \in \mathbf{R}^l$. Пусть $P = \{x \in \mathbf{R}^n \,\big|\, Ax = b\}$, множество D^0 задано в (15), а множество $\widetilde{\mathcal{D}}^0$ — в (32), (33). Показать, что для всякого числа $\sigma > 0$ решения задачи (31) с целевой функцией, заданной в (34), характеризуются системой уравнений (28).

Задача 4. Показать, что для канонической ЗЛП (12), (13) прямая часть $\chi(\cdot)$ прямодвойственной центральной траектории совпадает с прямой центральной траекторией.

Итерация *прямодвойственных методов внутренней точки* состоит в следующем. В текущей точке $(x^k,\, \lambda^k,\, z^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ такой, что $x^k > 0$ и $(\lambda^k,\, z^k) \in \widetilde{\mathcal{D}}^0$, для текущего $\sigma_k > 0$ определяют ньютоновское направление для системы (28). По этому направлению организуют одномерный поиск для функции качества

$$\widetilde{\varphi}\colon \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}, \quad \widetilde{\varphi}(x,\, z) = \langle x,\, z\rangle + |Ax - b|, \tag{35}$$

одновременно обеспечивая выполнение для получаемого следующего приближения $(x^{k+1},\, \lambda^{k+1},\, z^{k+1})$ неравенств $x^{k+1} > 0$ и $z^{k+1} > 0$.

Заметим, что

$$0 < \langle x^k,\, z^k\rangle = \langle c,\, x^k\rangle - \langle b,\, \lambda^k\rangle + \langle b - Ax^k,\, \lambda^k\rangle.$$

В частности, если $x^k \in P$ (а значит, $x^k \in D^0$), то

$$\widetilde{\varphi}(x^k,\, z^k) = \langle c,\, x^k\rangle - \langle b,\, \lambda^k\rangle.$$

Вообще же стремление $\widetilde{\varphi}(x^k,\, z^k)$ к нулю означает, что

$$\langle c,\, x^k\rangle - \langle b,\, \lambda^k\rangle \to 0, \quad \{Ax^k\} \to b \quad (k \to \infty),$$

откуда и из теоремы 7.1.8 следует, что любая предельная точка последовательности $\{(x^k,\, \lambda^k)\}$ является прямодвойственным решением задачи (12), (13). Заметим, что, как будет показано ниже, допустимость точек генерируемой последовательности $\{x^k\}$ в прямой задаче для данного метода не является автоматической, в отличие от прямых МВТ, и введение в функцию качества $\widetilde{\varphi}$ второго слагаемого как раз и связано с необходимостью решить эту проблему, по крайней мере в пределе.

Алгоритм 3. Выбираем число $\sigma_0 > 0$. Для заданного в (32), (33) множества $\widetilde{\mathcal{D}}^0$ выбираем $(x^0,\, \lambda^0,\, z^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$, $x^0 > 0$, $(\lambda^0,\, z^0) \in \widetilde{\mathcal{D}}^0$, и полагаем $k = 0$.

1. Вычисляем $(d_x^k,\, d_\lambda^k,\, d_z^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ как ньютоновский шаг для системы (28) из точки $(x^k,\, \lambda^k,\, z^k)$ при $\sigma = \sigma_k$.

2. Полагаем

   $$(x^{k+1},\, \lambda^{k+1},\, z^{k+1}) = (x^k,\, \lambda^k,\, z^k) + \alpha_k(d_x^k,\, d_\lambda^k,\, d_z^k),$$

   где параметр длины шага $\alpha_k \in (0,\, 1]$ выбирается так, чтобы выполнялось

   $$x^{k+1} > 0, \quad z^{k+1} > 0, \quad \widetilde{\varphi}(x^{k+1},\, z^{k+1}) < \widetilde{\varphi}(x^k,\, z^k).$$

3. Выбираем число $\sigma_{k+1} \in (0,\, \sigma_k)$.

4. Увеличиваем номер шага k на 1 и переходим к п. 1.

Как и в случае прямых МВТ, алгоритм 3 обычно снабжают подходящими процедурами масштабирования. Вообще, сколько-нибудь полное описание и анализ прямодвойственных МВТ потребовали бы обсуждения массы технических деталей. Ниже приводится лишь набросок анализа для описанной базовой схемы.

Задача 5. Пусть $x \in \mathbf{R}^n$, $x > 0$, $(\lambda,\, z) \in \widetilde{\mathcal{D}}^0$, где множество $\widetilde{\mathcal{D}}^0$ введено в (32), (33), $c \in \mathbf{R}^n$, $A \in \mathbf{R}(l,\, n)$, $b \in \mathbf{R}^l$.

Показать, что для заданного числа $\sigma > 0$ ньютоновский шаг $(d_x,\, d_\lambda,\, d_z) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^n$ для системы (28) из точки $(x,\, \lambda,\, z)$ задается равенствами

$$d_\lambda = (AZ^{-1}XA^{\mathrm{T}})^{-1}(AZ^{-1}(XZe - \sigma e) + b - Ax),$$

$$d_z = -A^{\mathrm{T}}d_\lambda, \tag{36}$$

$$d_x = Z^{-1}(\sigma e - XZe) - Z^{-1}X\,d_z.$$

Показать, что точка $x + d_x$ допустима в прямой задаче (12), (13), а точка $(\lambda + d_\lambda,\, z + d_z)$ — в двойственной задаче (29), (30).

Если в задаче 5 в качестве параметра длины шага берется $\alpha \in \in (0,\, 1)$, то точка $x + \alpha d_x$ может уже не быть допустима в прямой задаче, хотя, как нетрудно видеть, точка $(\lambda + \alpha d_\lambda,\, z + \alpha d_z)$ по-прежнему будет допустима в двойственной задаче. Покажем, что определенное согласование значений σ и α приводит к уменьшению значения функции качества $\widetilde{\varphi}$.

Предложение 4. *Пусть* $c \in \mathbf{R}^n$, $A \in \mathbf{R}(l,\, n)$, $b \in \mathbf{R}^l$. *Пусть* $x \in \mathbf{R}^n$, $x > 0$, $(\lambda,\, z) \in \widetilde{\mathcal{D}}^0$, *где множество* $\widetilde{\mathcal{D}}^0$ *введено в* (32), (33), *и для заданного числа* $\sigma > 0$ *вектор* $(d_x,\, d_\lambda,\, d_z) \in \mathbf{R}^n \times \mathbf{R}^l \times \times \mathbf{R}^n$ *определен как ньютоновский шаг для системы* (28) *из точки* $(x,\, \lambda,\, z)$.

Тогда $\forall\, \alpha \in (0, 1]$

$$A(x + \alpha d_x) - b = (1 - \alpha)(Ax - b), \tag{37}$$

$$\widetilde{\varphi}(x + \alpha d_x,\, z + \alpha d_z) =$$

$$= \widetilde{\varphi}(x,\, z) - \alpha(\langle x,\, z\rangle - n\sigma + |Ax - b|) + \alpha^2\langle Ax - b,\, d_\lambda\rangle, \tag{38}$$

где функция $\widetilde{\varphi}$ *введена в* (35).

Доказательство. Ньютоновский шаг для системы (28) из точки $(x,\, \lambda,\, z)$ определяется равенствами

$$A^{\mathrm{T}}d_\lambda + d_z = 0, \quad Ad_x = b - Ax, \quad Z\,d_x + X\,d_z = \sigma e - XZe. \tag{39}$$

Из второго равенства имеем

$$A(x + \alpha d_x) - b = Ax + \alpha(b - Ax) - b = (1 - \alpha)(Ax - b),$$

что и дает (37).

Далее,

$$\langle x + \alpha d_x,\, z + \alpha d_z\rangle = \langle x,\, z\rangle + \alpha(\langle x,\, d_z\rangle + \langle z,\, d_x\rangle) + \alpha^2\langle d_x,\, d_z\rangle. \tag{40}$$

Умножая левую и правую части последнего равенства в (39) скалярно на e, получаем

$$\langle x,\, d_z\rangle + \langle z,\, d_x\rangle = \langle e,\, \sigma e - XZe\rangle = n\sigma - \langle x,\, z\rangle.$$

Кроме того, используя (36) и второе равенство в (39), имеем

$$\langle d_x,\, d_z\rangle = -\langle d_x,\, A^{\mathrm{T}}d_\lambda\rangle = \langle Ax - b,\, d_\lambda\rangle.$$

Последние два равенства позволяют преобразовать (40) к виду

$$\langle x + \alpha d_x,\, z + \alpha d_z\rangle = \langle x,\, z\rangle - \alpha(\langle x,\, z\rangle - n\sigma) + \alpha^2\langle Ax - b,\, d_\lambda\rangle,$$

откуда и из (35), (37) окончательно получаем

$$\widetilde{\varphi}(x + \alpha d_x,\, z + \alpha d_z) = \langle x + \alpha d_x,\, z + \alpha d_z\rangle + |A(x + \alpha d_x) - b| =$$
$$= \langle x,\, z\rangle - \alpha(\langle x,\, z\rangle - n\sigma) + \alpha^2\langle Ax - b,\, d_\lambda\rangle + (1 - \alpha)|Ax - b|,$$

а это и есть (38). □

Как следует из (37), невязка ограничений прямой задачи уменьшается, как бы ни выбирался параметр длины шага $\alpha \in (0,\, 1]$, если только точка x сама не является допустимой в задаче (12), (13) (если точка x допустима, то новая точка $x + \alpha d_x$ также допустима). Кроме того, как следует из (38), если $\sigma < \langle x,\, z\rangle/n$, то существует число $\overline{\alpha} > 0$ такое, что любой выбор $\alpha \in (0,\, \overline{\alpha})$ приводит к уменьшению значения функции $\widetilde{\varphi}$.

Правильное построение последовательностей $\{\sigma_k\}$ и $\{\alpha_k\}$ в сочетании с масштабированием, а также разумным выбором методов вычислительной линейной алгебры для решения вспомогательных линейных систем, позволяет построить такие прямодвойственные МВТ, которые оказываются чрезвычайно эффективными на практике [50].

Задача 6. Рассмотрим задачу

$$f(x) \to \min, \quad x \in D = \{x \in \mathbf{R}^n \,\big|\, Ax = b,\; G(x) \leqslant 0\}, \tag{41}$$

где функция $f\colon \mathbf{R}^n \to \mathbf{R}$ и компоненты g_i отображения $G\colon \mathbf{R}^n \to \mathbf{R}^m$ выпуклы и дважды дифференцируемы на $\mathbf{R}^n$, $i = 1, \ldots, m$, $A \in \mathbf{R}(l,\, n)$, $b \in \mathbf{R}^l$. Пусть $\operatorname{rank} A = l$, $\operatorname{rank} G'(x) = m$ $\forall\, x \in D$, причем существует точка $\widetilde{x} \in D$ такая, что $G(\widetilde{x}) < 0$. Определим прямодвойственную центральную траекторию как отображение

$$(\chi(\cdot), \nu(\cdot), \zeta(\cdot))\colon (\mathbf{R}_+ \setminus \{0\}) \to \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m,$$

сопоставляющее каждому $\sigma > 0$ решение системы

$$f'(x) + A^{\mathrm{T}}\lambda + (G'(x))^{\mathrm{T}}\mu = 0, \quad Ax = b, \tag{42}$$

$$g_i(x) < 0, \quad \mu_i > 0, \quad \mu_i g_i(x) = \sigma, \quad i = 1, \ldots, m, \tag{43}$$

относительно $(x,\, \lambda,\, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$. Доказать, что следующие утверждения эквивалентны:

а) множество решений задачи (41) непусто и ограничено;

б) при любом $\sigma > 0$ система (42), (43) имеет единственное решение и, в частности, прямодвойственная центральная траектория корректно определена;

в) система (42), (43) имеет единственное решение хотя бы при одном $\sigma > 0$;

г) система

$$f'(x) + A^{\mathrm{T}}\lambda + (G'(x))^{\mathrm{T}}\mu = 0, \quad G(x) < 0, \quad \mu > 0,$$

имеет решение;

д) система

$$f'(x) + A^{\mathrm{T}}\lambda + (G'(x))^{\mathrm{T}}\mu = 0, \quad \mu > 0,$$

имеет решение.

СПИСОК ЛИТЕРАТУРЫ

1. *Аоки М.* Введение в методы оптимизации. — М.: Наука, 1977.
2. *Арутюнов А.В.* Условия экстремума. Анормальные и вырожденные задачи. — М.: Факториал, 1997.
3. *Ашманов С.А.*, *Тимохов А.В.* Теория оптимизации в задачах и упражнениях. — М.: Наука, 1991.
4. *Базара М.*, *Шетти К.* Нелинейное программирование. Теория и алгоритмы. — М.: Мир, 1982.
5. *Бахвалов Н.С.*, *Жидков Н.П.*, *Кобельков Г.М.* Численные методы. — М.: Наука, 1987.
6. *Бертсекас Д.* Условная оптимизация и методы множителей Лагранжа. — М.: Радио и связь, 1987.
7. *Васильев Ф.П.* Лекции по методам решения экстремальных задач. — М.: Изд-во МГУ, 1974.
8. *Васильев Ф.П.* Методы оптимизации. — М.: Факториал Пресс, 2002.
9. *Васильев Ф.П.* Методы решения экстремальных задач. — М.: Наука, 1981.
10. *Васильев Ф.П.* Численные методы решения экстремальных задач. — М.: Наука, 1988.
11. *Васильев Ф.П.*, *Иваницкий А.Ю.* Линейное программирование. — М.: Факториал, 1998.
12. *Воеводин В.В.*, *Кузнецов Ю.А.* Матрицы и вычисления. — М.: Наука, 1984.
13. *Гилл Ф.*, *Мюррей У.*, *Райт М.* Практическая оптимизация. — М.: Мир, 1985.
14. *Гольштейн Е.Г.*, *Третьяков Н.В.* Модифицированные функции Лагранжа. — М.: Наука, 1989.
15. *Гроссман К.*, *Каплан А.А.* Нелинейное программирование на основе безусловной минимизации. — Новосибирск: Наука, 1981.
16. *Дэннис Дж.*, *Шнабель Р.* Численные методы безусловной оптимизации и решения нелинейных уравнений. — М.: Мир, 1988.
17. *Евтушенко Ю.Г.* Методы решения экстремальных задач и их применение в системах оптимизации. — М.: Наука, 1982.
18. *Жиглявский А.А.*, *Жилинскас А.Г.* Методы поиска глобального экстремума. — М.: Наука, 1991.

19. *Зангвилл У.И.* Нелинейное программирование. — М.: Сов. радио, 1973.
20. *Измаилов А.Ф.* Чувствительность в оптимизации — М.: Физматлит, 2006.
21. *Измаилов А.Ф.*, *Третьяков А.А.* Факторанализ нелинейных отображений. — М.: Наука, 1994.
22. *Измаилов А.Ф.*, *Третьяков А.А.* 2-регулярные решения нелинейных задач. Теория и численные методы. — М.: Физматлит, 1999.
23. *Ильин В.А.*, *Садовничий В.А.*, *Сендов Бл.Х.* Математический анализ. — М.: Наука, 1979.
24. *Карманов В.Г.* Математическое программирование. — М.: Физматлит, 2000.
25. *Кларк Ф.* Оптимизация и негладкий анализ. — М.: Наука, 1988.
26. *Левитин Е.С.* Теория возмущений в математическом программировании и ее приложения. — М.: Наука, 1992.
27. *Мину М.* Математическое программирование. Теория и алгоритмы. — М.: Наука, 1990.
28. *Михалевич В.С.*, *Гупал А.М.*, *Норкин В.И.* Методы невыпуклой оптимизации. — М.: Наука, 1987.
29. *Моисеев Н.Н.*, *Иванилов Ю.П.*, *Столярова Е.М.* Методы оптимизации. — М.: Наука, 1978.
30. *Нурминский Е.А.* Численные методы выпуклой оптимизации. — М.: Наука, 1991.
31. *Ортега Дж.*, *Рейнболдт В.* Итерационные методы решения нелинейных систем уравнений со многими неизвестными. — М.: Мир, 1975.
32. *Полак Э.* Численные методы оптимизации. Единый подход. — М.: Мир, 1974.
33. *Поляк Б.Т.* Введение в оптимизацию. — М.: Наука, 1983.
34. *Пшеничный Б.Н.*, *Данилин Ю.М.* Численные методы в экстремальных задачах. — М.: Наука, 1975.
35. *Рокафеллар Р.* Выпуклый анализ. — М.: Мир, 1973.
36. *Стронгин Р.Г.* Численные методы в многоэкстремальных задачах. — М.: Наука, 1978.
37. *Сухарев А.Г.*, *Тимохов А.В.*, *Федоров В.В.* Курс методов оптимизации. — М.: Наука, 1986.
38. *Фиакко А.*, *Мак-Кормик Г.* Нелинейное программирование. Методы последовательной безусловной минимизации. — М.: Мир, 1972.
39. *Химмельблау Д.* Прикладное нелинейное программирование. — М.: Мир, 1975.
40. *Allgower E.L.*, *Georg K.* Numerical continuation methods. An introduction. — Berlin, Heidelberg: Springer-Verlag, 1990.
41. *Bertsekas D.P.* Nonlinear programming. — Second Edition. — Belmont: Athena, 1999.

42. *Bonnans J.F.*, *Gilbert J.Ch.*, *Lemaréchal C.*, *Sagastizábal C.* Numerical optimization. Theoretical and practical aspects. — Berlin: Springer-Verlag, 2003.

43. *Borgwardt K.H.* The simplex methods. A probabilistic analysis. — Berlin, Heidelberg: Springer-Verlag, 1987.

44. *Conn A.R.*, *Gould N.I.M.*, *Toint Ph.L.* Trust-region methods. — Philadelphia: SIAM, 2000.

45. *Cottle R.W.*, *Pang J.-S.*, *Stone R.E.* The linear complimentarity problem. — New York: Academic Press, 1992.

46. *Fletcher R.* Practical methods of optimization. V. 1. Unconstrained optimization. — Chichester, New York, Brisbane, Toronto: John Wiley, 1980.

47. *Fletcher R.* Practical methods of optimization. V. 2. Constrained optimization. — Chichester, New York, Brisbane, Toronto: John Wiley, 1981.

48. *Mangasarian O.L.* Nonlinear Programming. — Philadelphia: SIAM, 1994.

49. *Moré J.J.*, *Wright S.J.* Optimization software guide. — Philadelphia: SIAM, 1993.

50. *Nocedal J.*, *Wright S.J.* Numerical optimization. — New York, Berlin, Heidelberg: Springer-Verlag, 2000.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ