ЭКОНОМИКО-
МАТЕМАТИЧЕСКАЯ
БИБЛИОТЕКА

В. С. МИХАЛЕВИЧ, А. М. ГУПАЛ, В. И. НОРКИН

МЕТОДЫ НЕВЫПУКЛОЙ ОПТИМИЗАЦИИ

МОСКВА «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1987

ББК 22.18
М69
УДК 519.6

Михалевич В. С., Гупал А. М., Норкин В. И. **Методы невыпуклой оптимизации.**— М. : Наука. Гл. ред. физ.-мат. лит., 1987.— 280 с.

В прикладных задачах часто приходится иметь дело с недифференцируемыми функциями и с невыпуклыми областями. Разработка методов оптимизации для таких задач представляет большую трудность. В книге рассмотрены конечномерные задачи невыпуклой негладкой оптимизации и численные методы их решения.

Для специалистов в области прикладной математики, информатики, экономики, кибернетики.

Библиогр. 188 назв.

Рецензент доктор физико-математических наук *В. В. Федоров*

М $\frac{\mathbf{1702070000\text{-}177}}{\mathbf{053(02)\text{-}87}}$ **25-87**

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Эта книга посвящена конечномерным задачам невыпуклой негладкой оптимизации и численным методам их решения. Теория сложности экстремальных задач утверждает, что невыпуклые задачи как объект исследования очень сложны, и трудоемкость получения гарантированного решения экспоненциально повышается с ростом размерности. Однако большая практическая важность этих задач убеждает исследователей развивать численные методы несмотря на столь обескураживающие предупреждения об их сложности.

Проблема невыпуклости в книге изучается на двух основных моделях невыпуклых зависимостей — это так называемые обобщенно дифференцируемые и локально липшицевы функции. Изучение негладких функций это не дань сложившейся моде. Негладкие функции естественным образом возникают в различных приложениях. Кроме того, они часто появляются и в самой теории экстремальных задач за счет операций взятия максимума и минимума, приемов декомпозиции, точных негладких штрафов, двойственности.

Рассматриваемые модели невыпуклости достаточно общи и охватывают большинство практически важных задач оптимизации; на них хорошо просматриваются все трудности невыпуклой оптимизации. Способ исследования обобщенно дифференцируемых функций заключается в том, что для этих функций вводится обобщение понятия градиента, строится исчисление и в терминах обобщенных градиентов исследуются различные свойства невыпуклых задач. Что касается численных методов, то на задачи с обобщенно дифференцируемыми функциями удается распространить теорию и алгоритмы субградиентного спуска выпуклой оптимизации.

Методы решения липшицевых задач характеризуются тем, что исходные функции аппроксимируются сглаженными и к ним применяются итерационные процедуры минимизации. При таком подходе удается аппроксимировать градиенты сглаженных функций стохастическими конечными разностями и, таким образом, построить методы без вычисления градиентов. Аналогичный подход можно обосновать в липшицевом стохастическом программировании. При этом получаются разнообразные обобщения классического метода стохастической аппроксимации для решения задач липшицева стохастического программирования с ограничениями.

К настоящему времени математическая теория выпуклой оптимизации приобрела законченные черты: разработана теория сложности задач выпуклого программирования и построены эффективные численные методы. Теория невыпуклой оптимизации разработана в гораздо меньшей степени, однако можно заметить, что отдельные компоненты теории (такие, как необходимые условия экстремума) продвинуты весьма далеко, а остальные компоненты нуждаются в дальнейшей разработке; в первую очередь это касается развития методов невыпуклой оптимизации, преодолевающих негладкость, невыпуклость и овражность функций, а также учитывающих многоэкстремальность и возможную стохастическую природу этих задач.

В теории оптимизации вводились и изучались самые разнообразные классы невыпуклых негладких функций. Это функции: максимума, псевдовыпуклые, квазивыпуклые, слабо выпуклые, полувыпуклые, локально выпуклые, полугладкие, квазидифференцируемые, локально липшицевы, разрывные и др. В гл. 1 также изучается новый класс недифференцируемых функций — обобщенно дифференцируемые функции. Возникает вопрос: для чего вводится еще один класс функций?

Обобщенно дифференцируемые функции являются непосредственным обобщением непрерывно дифференцируемых и выпуклых функций. Для них строятся псевдоградиенты, аналогичные субградиентам выпуклых функций, а также имеет место дифференциальное разложение, подобное разложению дифференцируемых функций. Вместе с тем этот класс функций достаточно широк: он включает в себя непрерывно дифференцируемые, выпуклые и вогнутые, слабо выпуклые и слабо вогнутые, полугладкие функции. Кроме того, класс обобщенно дифференцируемых функций замкнут относительно операций взятия максимума, минимума и суперпозиции и взятия математического ожидания. Имеют место простые формулы (исчисление) для вычисления псевдоградиентов сложных обобщенно дифференцируемых функций, аналогичные формулам вычисления обычных градиентов сложных дифференцируемых функций, в то время как для других классов негладких функций построение такого исчисления встречает значительные трудности.

Для обобщенно дифференцируемых функций легко выводятся необходимые условия экстремума и имеет место аналог теоремы о среднем. И, наконец, самый важный момент состоит в том, что для минимизации обобщенно дифференцируемых функций по аналогии с выпуклым случаем удается построить численные методы, использующие псевдоградиенты этих функций.

В работах [145, 146] введены понятия «обобщенный градиент» и «обобщенная производная по направлению» для довольно широкого класса локально липшицевых функций. Оказывается, что класс локально липшицевых функций настолько широк, что для него не применимы формулы подсчета обобщенных градиентов сложных функций (например, для функций максимума, суммы и суперпозиции). Это означает, что для минимизации липшицевых функций нельзя широко использовать методы обобщенного градиентного спуска.

В гл. 2 показано, что к успеху приводят стохастические конечно-разностные процедуры, которые не требуют вычислений обобщенных градиентов, в то время как обычные конечно-разностные аппроксимации не работают. Основная идея выбора направления спуска состоит в том, что конечная разность вычисляется не в текущей точке, а в близкой, случайным образом выбранной точке, причем вектор математических ожиданий построенных конечных разностей близок к множеству обобщенных градиентов. Показано, что построенные конечно-разностные методы находят стационарные точки липшицевых задач. С точки зрения количества вычислений важно отметить, что для функций максимума конечная разность строится по той функции, на которой достигается максимум. В том случае, когда одно вычисление функции цели занимает значительное время, построение конечных разностей с помощью случайных направлений приводит к значительному сокращению числа вычислений функции цели.

Анализ эффективности конечно-разностных методов на задачах выпуклого программирования показывает, что замена субградиентов введенными конечными разностями, грубо говоря, приводит к ситуации, когда субградиенты вычисляются с некоторыми случайными помехами.

Среди методов негладкой оптимизации метод обобщенного градиентного спуска [136] привлекает своей простотой, экономичностью, определенной оптимальностью для выпуклых задач большой размерности [73], сходимостью в невыпуклом случае [4, 35, 80, 88].

В гл. 3 метод обобщенного градиента распространен на задачи локальной минимизации обобщенно дифференцируемых функций при ограничениях. В негладкой выпуклой оптимизации получили распространение релаксационные ε-субградиентные методы. В гл. 3 предлагаются несколько отличные релаксационные методы, пригодные для минимизации невыпуклых негладких функций.

Обобщенные градиентные методы позволяют находить лишь локальные минимумы функций, поэтому отдельный параграф посвящен некоторым подходам, позволяющим преодолевать многоэкстремальность задач. В частности, изучаются схемы комбинированных алгоритмов, в которых разработанные локальные алгоритмы используются для приближенного нахождения локальных минимумов, а глобальные алгоритмы — для выхода из зон притяжения локальных экстремумов. Предлагается алгоритм многомерного глобального поиска при ограничениях, обобщающий метод (см. [96]). Исследованы глобальные свойства процедуры сглаживания функций, широко используемой в данной книге. Отметим также, что определенными глобальными свойствами обладают рассмотренные в гл. 4 методы тяжелого шарика и овражного шага.

Давно замечено, что (обобщенные) градиентные методы медленно сходятся в овражных ситуациях. Один из способов преодоления этого недостатка состоит в усреднении градиентов (или соответствующих конечных разностей), полученных на предыдущих итерациях, с целью нахождения направления вдоль оврага. В этой связи в гл. 4 изучаются немонотонные методы усредненных градиентов (конечных разностей) для минимизации невыпуклых негладких функций при ограничениях, которые сохраняют многие достоинства градиентных методов, но уже обладают противоовражными свойствами. Оказывается, в этот класс попадают соответствующим образом обобщенные известные методы тяжелого шарика [100] и овражного шага [14, 74]. Методы усредненных градиентов можно также трактовать как процедуры минимизации последовательности некоторых усредненных функций.

В гл. 5 исследуются численные методы минимизации липшицевых функций при ограничениях. Изучаются методы условного и приведенного градиентов. Для задач с нелинейными ограничениями строятся методы возможных направлений и штрафных функций. Для минимизации липшицевых функций предлагается весьма простой метод типа штрафных функций, в котором коэффициент штрафа определяется аналитическим выражением.

Обширной областью применения идей негладкой оптимизации является стохастическое программирование. От детерминированной задачи нелинейного программирования стохастическая задача отличается тем, что значения функций и их градиентов в общем случае не могут быть вычислены точно за конечное время и, таким образом, известны лишь с некоторой случайной ошибкой.

Первые методы стохастического программирования восходят к методам стохастической аппроксимации. Для решения задач выпуклого стохастического программирования были предложены методы стохастических квазиградиентов [41]; важные усовершенствования сделаны в [73].

Значительные трудности возникают на пути решения невыпуклых стохастических задач. Для того чтобы обосновать стохастические аналоги градиентных процедур минимизации, в первую очередь нужно показать перестановочность знаков обобщенного дифференцирования и взятия математического ожидания. Этот факт показан в гл. 6 для обобщенно дифференцируемых функций. На основе аппарата измеримых многозначных отображений построено исчисление случайных псевдоградиентов сложных обобщенно дифференцируемых функций, полностью аналогичное исчислению псевдоградиентов детерминированных функций. Поэтому для этих функций можно обосновать методы стохастических обобщенных градиентов.

Для липшицевых функций ситуация иная: градиентные процедуры обосновать не удается, так как обобщенный градиент функции математического ожидания лишь включен в математическое ожидание (случайного) обобщенного градиента функции, стоящей под знаком математического ожидания. В этой связи важное значение имеют разработанные конечно-разностные методы липшицева стохастического программирования, которые в значительной степени обобщают методы стохастической аппроксимации Кифера — Вольфовица.

В гл. 7 изучаются стохастические аналоги рассмотренных ранее методов решения детерминированных задач.

ЭЛЕМЕНТЫ ТЕОРИИ НЕВЫПУКЛОГО АНАЛИЗА

§ 1. Обобщенно дифференцируемые функции

В этом параграфе вводится один класс невыпуклых негладких функций, названных здесь обобщенно дифференцируемыми. Эти функции обладают многими хорошими с точки зрения оптимизации свойствами. Они являются естественным обобщением непрерывно дифференцируемых и выпуклых функций, для них вводятся псевдоградиенты, аналогичные субградиентам выпуклых функций, и справедливо дифференциальное разложение с участием псевдоградиентов.

Обобщенно дифференцируемые функции удовлетворяют локальному условию Липшица. Они, вообще говоря, не имеют производных по направлениям, но почти всюду непрерывно дифференцируемы. Псевдоградиенты функций определяются не однозначно. Для одной функции существует целое семейство псевдоградиентных отображений, минимальное по включению из которых совпадает с субдифференциальным отображением по Кларку [145] этой функции. В то же время любое псевдоградиентное отображение определяет обобщенно дифференцируемую функцию с точностью до константы.

Непрерывно дифференцируемые, выпуклые и вогнутые, слабо выпуклые и слабо вогнутые [91], полугладкие [167] функции являются и обобщенно дифференцируемыми, а их градиенты, субградиенты, квазиградиенты и обобщенные градиенты являются и псевдоградиентами.

Класс обобщенно дифференцируемых функций замкнут относительно конечных операций взятия максимума, минимума и суперпозиции, а также, как будет показано в § 23, относительно взятия математического ожидания. Для сложных обобщенно дифференцируемых функций существует исчисление псевдоградиентов.

Для обобщенно дифференцируемых функций выписываются необходимые условия экстремума (§ 3), справедлива обобщенная теорема Лагранжа о конечных приращениях. И наконец, как будет показано в последующих главах, для минимизации обобщенно дифференцируемых функций можно развить обобщенные градиентные методы, в том числе стохастические.

В этом параграфе без комментариев используются некоторые общие сведения из выпуклого анализа, теории многозначных отображений (§ 21) и локально липшицевых функций (§ 2).

Определение 1.1. Функция $f : E_n \to E_1$ называется *обобщенно дифференцируемой в точке* $x \in E_n$, если в некоторой окрестности точки x определено такое полунепрерывное сверху в x многозначное отображение G_f, что его значения $G_f(y)$ $(y \in E_n)$ являются непустыми ограниченными выпуклыми замкнутыми множествами и в окрестности точки x справедливо разложение

$$f(y) = f(x) + (g, y - x) + o(x, y, g), \tag{1.1}$$

где $g \in G_f(y)$, (,) — скалярное произведение, а функция $o(x, y, g)$ удовлетворяет условию

$$\lim_{k\to\infty} (o(x, y^k, g^k)/\| y^k - x \|) = 0 \tag{1.2}$$

для любых последовательностей $y^k \to x$ $(y^k \neq x)$, $g^k \in G_f(y^k)$ или, что эквивалентно,

$$\lim_{\varepsilon\to+0} \sup_{\{y \mid \|y-x\| \leqslant \varepsilon, y \neq x\}} \sup_{g \in G_f(y)} \frac{|o(x, y, g)|}{\| y - x \|} = 0. \tag{1.3}$$

Функция называется *обобщенно дифференцируемой в области*, если она обобщенно дифференцируема в каждой точке области. Векторы $g \in G_f(y)$ будем называть *псевдоградиентами* (*обобщенными градиентами*) функции f в точке y.

Если выполнено условие (1.3) на функцию $o(x, y, g)$, то очевидно, что выполнено и (1.2). Обратное легко доказывается от противного.

Определение 1.2. Функция f: $E_n \to E_1$ называется *локально липшицевой*, если для любого компакта $K \subset E_n$ существует константа Липшица L_K такая, что

$$| f(x') - f(x'') | \leqslant L_K \| x' - x'' \| \qquad \forall\, x', x'' \in K.$$

Теорема 1.1. *Обобщенно дифференцируемые функции локально липшицевы и, следовательно, непрерывны.*

Доказательство. Пусть $f(x)$ — обобщенно дифференцируемая функция, K — компакт в E_n, $\mathrm{co}\, K$ — его выпуклая оболочка. Для $y, x \in K$ представим

$$f(y) - f(x) = (g, y - x) + o(x, y, g) =$$
$$= \left(\left(g, \frac{y - x}{\| y - x \|}\right) + \frac{o(x, y, g)}{\| y - x \|}\right) \| y - x \|,$$

где $g \in G_f(y)$. В силу условия (1.2) для любого ε существует такое $\delta(x, \varepsilon)$, что

$$| o(x, y, g) | / \| y - x \| \leqslant \varepsilon \quad \text{при} \quad \| y - x \| \leqslant \delta(x, \varepsilon), \quad g \in G_f(y).$$

Так как отображение G_f полунепрерывно сверху, а его значения $G_f(x)$ являются компактными множествами, то оно локально ограничено; поэтому существует число

$$\Gamma_K = \sup \{ \| g \| \mid g \in G_f(y),\ y \in \mathrm{co}\, K \} < +\infty.$$

Таким образом, для каждого $x \in \mathrm{co}\, K$ найдена такая δ-окрестность точки x, что

$$| f(y) - f(x) | \leqslant (\Gamma_K + \varepsilon) \| y - x \| \quad \text{при} \quad \| y - x \| \leqslant \delta, \qquad y \in \mathrm{co}\, K.$$

Пусть теперь $x', x'' \in K$; рассмотрим отрезок

$$M = \{x = x' + t(x'' - x'),\ t \in [0, 1]\}, \qquad M \subset \mathrm{co}\, K.$$

Указанные выше δ-окрестности точек $x \in M$ покрывают M; поэтому существует конечное покрытие с центрами $x^1 = x', x^2, \ldots, x^k = x''$, причем соответственно $t_1 = 0 < t_2 < \ldots < t_k = 1$. Легко видеть, что для соседних точек x^j и x^{j+1} во всех случаях

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

откуда

$$|f(x') - f(x'')| \leqslant \sum_{j=1}^{k-1} |f(x^j) - f(x^{j+1})| \leqslant (\Gamma_K + \varepsilon) \| x' - x'' \|.$$

Теорема доказана.

Теорема 1.2. *Непрерывно дифференцируемые функции обобщенно дифференцируемы, а их градиенты можно взять в качестве псевдоградиентов.*

Доказательство. Пусть функция $f(y)$ $(y \in E_n)$ непрерывно дифференцируема, $g(y)$ — ее градиент. Тогда в каждой точке $x \in E_n$ имеет место разложение

$$f(y) = f(x) + (g(x), y - x) + o(\| y - x \|) =$$
$$= f(x) + (g(y), y - x) + o(x, y, g),$$

где остаточный член

$$o(x, y, g) = (g(x) - g(y), y - x) + o(\| y - x \|)$$

удовлетворяет требуемым условиям малости (1.2) и (1.3).

Определение 1.3 [88]. Непрерывная функция $f\colon E_n \to E_1$ называется *слабо выпуклой*, если для любого x существует непустое множество $G(x)$ таких квазиградиентов g, что для всех $g \in G(x)$ справедливо неравенство

$$f(y) \geqslant f(x) + (g, y - x) + r(x, y), \tag{1.4}$$

где $r(x, y)/\| y - x \| \to 0$, когда $\| y - x \| \to 0$, и x, y принадлежат замкнутому ограниченному множеству.

Очевидно, что выпуклые функции являются и слабо выпуклыми; в этом случае $r(x, y) \equiv 0$. Известно [91], что множество квазиградиентов $G(x)$ непусто, ограничено, выпукло и замкнуто, многозначное отображение $G\colon x \to G(x)$ полунепрерывно сверху. Слабо выпуклые функции имеют производные по направлениям.

Теорема 1.3. *Выпуклые и слабо выпуклые функции являются и обобщенно дифференцируемыми, а их субградиенты и квазиградиенты можно взять в качестве псевдоградиентов этих функций.*

Доказательство. Здесь роль G_f играет G. Докажем справедливость разложения (1.1) в окрестности точки x. По определению слабо выпуклой функции выполняется неравенство (1.4). Запишем это неравенство относительно точек y:

$$f(x) \geqslant f(y) + (g(y), x - y) + r(y, x). \tag{1.5}$$

Введем обозначение

$$o(x, y, g) = f(y) - f(x) - (g, y - x).$$

Из (1.4) и (1.5) следует

$$(g(x) - g(y), y - x) + r(x, y) \leqslant o(x, y, g(y)) \leqslant -r(y, x),$$

где $g(x) \in G(x)$, $g(y) \in G(y)$. Выберем $g(x)$ так, чтобы

$$\| g(x) - g(y) \| = \rho(g(y), G(x)) \equiv \inf\{\| g(y) - g \| \mid g \in G(x)\}.$$

Тогда

$$-\rho(g(y), G(x)) \| y - x \| + r(x, y) \leqslant o(x, y, g(y)) \leqslant -r(y, x). \quad (1.6)$$

В силу полунепрерывности сверху отображения G и определения слабо выпуклой функции $\rho(g(y), G(x)) \to 0$,

$$r(x, y)/\| y - x \| \to 0, \quad r(y, x)/\| y - x \| \to 0$$

при $\| y - x \| \to 0$ равномерно по $g(y) \in G(y)$. Тогда из (1.6) следует требуемое свойство $o(x, y, g)$.

Определение 1.4 [167]. Функция $f: E_n \to E_1$ называется *полугладкой*, если:

1) она локально липшицева;

2) для любых $x \in E_n$, любых последовательностей $y^k \to x$ $((y^k - x)/\| y^k - x \| \to d)$ и $g^k \in \partial f(y^k)$ существует предел скалярных произведений $\lim\limits_{k \to \infty} (g^k, d)$.

Здесь $\partial f(y)$ — субдифференциал (множество обобщенных градиентов) Кларка локально липшицевой функции $f(y)$. Субдифференциал ∂f подробно изучается в § 2.

Очевидно, что для одного и того же направления d все скалярные произведения (g^k, d) в определении 1.4 имеют один и тот же предел. Нетрудно показать, используя условие 2) и теорему о среднем 2.9, что полугладкие функции имеют производные $f'(x; d)$ по направлению d и

$$\lim_{k \to \infty} (g^k, d) = f'(x; d)$$

(см. теорему 1.14).

Теорема 1.4. *Полугладкие функции обобщенно дифференцируемы, а их обобщенные градиенты по Кларку можно взять в качестве псевдоградиентов.*

Доказательство. Здесь роль G_f играет замкнутозначное полунепрерывное сверху отображение ∂f (см. § 2). Докажем справедливость разложения (1.1). Предположим противное. Тогда найдутся такие $x \in E_n$, $\varepsilon > 0$, $y^k \to x$ $((y^k - x)/\| y^k - x \| \to d)$, $g^k \in \partial f(y^k)$, что

$$\left| \frac{f(y^k) - f(x)}{\| y^k - x \|} - \left(g^k, \frac{y^k - x}{\| y^k - x \|} \right) \right| \geqslant \varepsilon > 0.$$

Используя теорему 2.9 о конечных приращениях для локально липшицевых функций, перепишем предыдущее неравенство в виде

$$\left|\left(\bar{g}^k, \frac{y^k - x}{\| y^k - x \|}\right) - \left(g^k, \frac{y^k - x}{\| y^k - x \|}\right)\right| \geqslant \varepsilon > 0,$$

где $\bar{g}^k \in \partial f(z^k)$, $z^k = x + \theta^k (y^k - x)$, $0 < \theta^k < 1$. Тогда

$$(z^k - x)/\| z^k - x \| = (y^k - x)/\| y^k - x \| \to d,$$

$$\left|\left(\bar{g}^k, \frac{z^k - x}{\| z^k - x \|}\right) - \left(g^k, \frac{y^k - x}{\| y^k - x \|}\right)\right| \geqslant \varepsilon > 0. \qquad (1.7)$$

Оба скалярных произведения в (1.7) имеют в силу определения полугладкой функции один и тот же предел при $k \to +\infty$, откуда получаем противоречие: $0 \geqslant \varepsilon > 0$. Теорема доказана.

Итак, класс обобщенно дифференцируемых функций содержит непрерывно дифференцируемые, выпуклые и полугладкие функции. Покажем теперь, что он замкнут по отношению к конечным операциям взятия максимума и суперпозиции; отсюда будет следовать, что вогнутые и слабо вогнутые функции [91], а также функция минимума от конечного числа обобщенно дифференцируемых функций являются обобщенно дифференцируемыми.

Теорема 1.5. *Пусть* $f_i(y)$ $(y \in E_n,\ i = 1, \ldots, m)$ — *обобщенно дифференцируемые функции*, $G_{f_i}(y)$ — *их псевдоградиентные отображения. Тогда функция максимума*

$$f(y) = \max(f_1(y), \ldots, f_m(y))$$

обобщенно дифференцируема и

$$G_f(y) = \mathrm{co}\,\{g \in E_n \mid g \in G_{f_i}(y),\ f_i(y) = f(y)\}. \qquad (1.8)$$

Доказательство. Достаточно рассмотреть максимум от двух функций. Отображение G_f выпуклозначно. Покажем, что оно замкнуто и локально ограничено; отсюда будет следовать его полунепрерывность сверху. Замкнутость означает, что из $y^k \to y$, $g^k \in G_f(y^k)$ и $g^k \to g$ следует $g \in G_f(y)$. Представим

$$g^k = \sum_{i=1}^{m} \lambda_i^k g_i^k, \quad \lambda_i^k \geqslant 0, \quad \sum_{i=1}^{m} \lambda_i^k = 1,$$

где $\lambda_i^k = 0$, если

$$f_i(y^k) < f(y^k).$$

В силу ограниченности λ_i^k, замкнутости и локальной ограниченности отображений G_{f_i} и конечности числа m существуют подпоследовательности

$$\lambda_i^{k_s} \to \lambda_i, \quad g_i^{k_s} \to g_i, \quad s = 0, 1, \ldots,$$

причем

$$\lambda_i \geqslant 0, \qquad \sum_{i=1}^{m} \lambda_i = 1, \qquad g_i \in G_{f_i}(y),$$

$$g = \lim_{s\to\infty} g^{k_s} = \sum_{i=1}^{m} \lambda_i g_i.$$

Если для некоторого i

$$f_i(y) < f(y),$$

то это выполнено и в некоторой окрестности точки y, поэтому при всех достаточно больших k_s имеем $\lambda_i^{k_s} = 0$; следовательно,

$$\lambda_i = \lim_{s\to\infty} \lambda_i^{k_s} = 0.$$

Таким образом, $g \in G_f(y)$, и замкнутость G_f доказана. Каждое из отображений G_{f_i} компактнозначно и полунепрерывно сверху и, следовательно, локально ограничено. Очевидно, что тогда и G_f локально ограничено. А из замкнутости и локальной ограниченности следует полунепрерывность сверху отображения G_f.

Теперь необходимо установить справедливость разложения (1.1), т. е. необходимо показать, что для любого $\varepsilon > 0$ найдется такое $\delta(\varepsilon) > 0$, что

$$-\varepsilon \leqslant \frac{f(y) - f(x) - (g, y - x)}{\| y - x \|} \leqslant \varepsilon \tag{1.9}$$

при всех $\{y \mid \|y - x\| \leqslant \delta,\ y \neq x\}$ и всех $g \in G_f(y)$. В силу обобщенной дифференцируемости $f_1(y)$ и $f_2(y)$ для любого $\varepsilon > 0$ существуют такие $\delta_1(\varepsilon) > 0$ и $\delta_2(\varepsilon) > 0$, что

$$-\varepsilon \leqslant \frac{f_1(y) - f_1(x) - (g^1, y - x)}{\| y - x \|} \leqslant \varepsilon \tag{1.10}$$

при всех $\{y \mid \| y - x \| \leqslant \delta_1(\varepsilon),\ y \neq x\}$ и всех $g^1 \in G_{f_1}(y)$ и что

$$-\varepsilon \leqslant \frac{f_2(y) - f_2(x) - (g^2, y - x)}{\| y - x \|} \leqslant \varepsilon \tag{1.11}$$

при всех $\{y \mid \| y - x \| \leqslant \delta_2(\varepsilon),\ y \neq x\}$ и всех $g^2 \in G_{f_2}(y)$.

Если

$$f_1(x) \neq f_2(x),$$

например $f_1(x) > f_2(x)$, то это же неравенство сохраняется и в некоторой δ_3-окрестности точки x, и там $f(y) \equiv f_1(y)$. Для данного ε возьмем

$$\delta(\varepsilon) = \min(\delta_1(\varepsilon), \delta_3(\varepsilon)),$$

тогда требуемое неравенство (1.9) следует из (1.10).

Пусть теперь

$$f_1(x) = f_2(x) = f(x).$$

Для произвольного ε возьмем

$$\delta(\varepsilon) = \min(\delta_1(\varepsilon), \delta_2(\varepsilon))$$

и проверим выполнение неравенства (1.9) для всех $\{y \mid \|y - x\| \leqslant \delta, y \neq x\}$ и всех $g \in G_f(y)$, где $G_f(y)$ построены согласно (1.8).

Пусть y таково, что $\|y - x\| \leqslant \delta$ и, например, $f_1(y) > f_2(y)$. Тогда

$$f(y) = f_1(y), \quad g = g^1 \in G_{f_1}(y),$$

и (1.9) справедливо в силу (1.10).

Рассмотрим такие y, что $\|y - x\| \leqslant \delta$ и $f(y) = f_1(y) = f_2(y)$. Тогда, согласно (1.8),

$$g = \lambda_1 g^1 + \lambda_2 g^2,$$

где $\lambda_1 + \lambda_2 = 1$, $g^1 \in G_{f_1}(y)$, $g^2 \in G_{f_2}(y)$. Подставив эти y, g^1, g^2 в (1.10) и (1.11), умножив (1.10) и (1.11) на λ_1 и λ_2 и сложив, получим (1.9). Теорема доказана.

Теорема 1.6. *Пусть $f_0(z)$ $(z \in E_m)$, $f_i(y)$ $(y \in E_n,\ i = 1, \ldots, m)$ — обобщенно дифференцируемые функции. Тогда сложная функция*

$$f(y) = f_0(z(y)) = f_0(f_1(y), f_2(y), \ldots, f_m(y))$$

также обобщенно дифференцируема и

$$G_f(y) = \operatorname{co}\{g \in E_n \mid g = [g^1 \ldots g^m]\, g^0,\ g^0 \in G_{f_0}(z(y)),$$
$$g^i \in G_{f_i}(y),\ i = 1, \ldots, m\}, \qquad (1.12)$$

где $[g^1 \ldots g^m]$ — матрица размера $n \times m$, составленная из векторов (столбцов) $g^1, \ldots, g^m$, $z(y) = (f_1(y), \ldots, f_m(y))$.

Формула (1.12) является обобщением цепного правила дифференцирования сложных функций в математическом анализе. Таким образом, класс обобщенно дифференцируемых функций образует линейное пространство и замкнут относительно суперпозиции.

Доказательство. Отображение G_f выпуклозначно. Покажем, что оно локально ограничено и замкнуто; отсюда будет следовать его полунепрерывность сверху. Заметим, что отображения G_{f_i} $(i = 1, 2, \ldots, m)$ локально ограничены, так как они компактнозначны и полунепрерывны сверху. Это эквивалентно их ограниченности в каждом компакте. Отображение $G_{f_0}(z(y))$ локально ограничено как суперпозиция непрерывного отображения $z(y)$ и ограниченного в каждом компакте отображения $G_{f_0}(z)$. Тогда из (1.12) очевидно, что и G_f локально ограничено. Для замкнутости G_f нужно показать, что из $y^k \to y$, $g^k \in G_f(y^k)$ и $g^k \to g$ следует $g \in G_f(y)$. В силу теоремы Каратеодори g^k можно представить в виде выпуклой комбинации

$$g^k = \sum_{j=1}^{n+1} \lambda_{jk} [(g^1)^{jk} \ldots (g^m)^{jk}]\, (g^0)^{jk}, \qquad \lambda_{jk} \geqslant 0, \qquad \sum_{j=1}^{n+1} \lambda_{jk} = 1,$$
$$(g^i)^{jk} \in G_{f_i}(y^k), \quad i = 1, \ldots, m, \quad j = 1, \ldots, n+1,$$
$$(g^0)^{jk} \in G_{f_0}(z(y^k)).$$

Последовательности $\{(g^i)^{jk}\}_{k=1}^{\infty}$ $(i = 0, 1, \dots, m,\ \ j = 1, \dots, n+1)$ ограничены в силу локальной ограниченности отображений G_{f_i} $(i = 1, \dots$ $\dots, m)$ и $G_{f_0}(z(y))$. Последовательности $\{\lambda_{jk}\}_{k=1}^{\infty}$ $(j = 1, \dots, n+1)$ также ограничены. В силу конечности m и n существует такая подпоследовательность индексов k_s, что

$$\lambda_{jk_s} \to \lambda_j, \quad (g^i)^{jk_s} \to (g^i)^j, \quad i = 0, \dots, m, \quad j = 1, \dots, n+1.$$

При этом

$$\lambda_j \geqslant 0, \quad \sum_{j=1}^{n+1} \lambda_j = 1, \quad (g^i)^j \in G_{f_i}(y), \quad i = 1, \dots, m, \quad j = 1, \dots, n+1,$$

а также

$$(g^0)^j \in G_{f_0}(z(y))$$

в силу замкнутости отображений $G_{f_i}(y)$ $(i = 1, \dots, m)$ и $G_{f_0}(z(y))$. Поэтому

$$g = \lim_{s\to\infty} g^{k_s} = \sum_{j=1}^{n+1} \lambda_j [(g^1)^j \dots (g^m)^j] (g^0)^j \in G_f(y).$$

Замкнутость G_f доказана. Итак, G_f локально ограничено и замкнуто; следовательно, оно полунепрерывно сверху.

Докажем теперь разложение (1.1). Представим

$$\begin{aligned}
f(y) &= f_0(z(y)) = f_0(f_1(y), \dots, f_m(y)) = \\
&= f_0(z(x)) + (g^0(z(y)), z(y) - z(x)) + o_0(z(x), z(y), g^0(z(y))) = \\
&= f(x) + \sum_{i=1}^{m} g^0_{(i)}(z(y))(f_i(y) - f_i(x)) + o_0(z(x), z(y), g^0(z(y))) = \\
&= f(x) + \sum_{i=1}^{m} g^0_{(i)}(z(y))((g^i(y), y - x) + o_i(x, y, g^i)) + \\
&+ o_0(z(x), z(y), g^0(z(y))) = f(x) + ([g^1 \dots g^m] g^0(z(y)), y - x) + \\
&+ \sum_{i=1}^{m} g^0_{(i)}(z(y))\, o_i(x, y, g^i) + o_0(z(x), z(y), g^0(z(y))), \qquad (1.13)
\end{aligned}$$

где

$$g^0(z(y)) = (g^0_{(1)}(z(y)), \dots, g^0_{(m)}(z(y))) \in G_{f_0}(z(y)),$$

$$g^i(y) \in G_{f_i}(y), \quad i = 1, \dots, m,$$

$o_i(x, y, g^i)/\|y - x\| \to 0$ равномерно по $y \to x$ и по $g^i \in G_{f_i}(y)$.

Из представления (1.13) следует требуемое разложение (1.1)—(1.3). Действительно, зафиксируем некоторую δ-окрестность точки x. Так как $G_{f_0}(z(y))$ ограничено в каждом компакте, то существует такая константа $c < \infty$, что для всех $\{y \mid \|y - x\| \leqslant \delta\}$ и всех $g^0 \in G_{f_0}(z(y))$ будет $\|g\| \leqslant c$. Отображение $z(y)$ в силу теоремы 1.1 локально липшицево; поэтому для всех $\{y \mid \|y - x\| \leqslant \delta\}$ имеем

$$\|z(y) - z(x)\| / \|y - x\| \leqslant L < \infty.$$

Пусть теперь $g \in G_f(y)$, $\|y - x\| \leqslant \delta$. Обозначим

$$o(x, y, g) = f(y) - f(x) - (g, y - x). \tag{1.14}$$

В силу теоремы Каратеодори вектор g можно представить в виде выпуклой комбинации:

$$g = \sum_{j=1}^{n+1} \lambda_j [(g^1)^j \ldots (g^m)^j] (g^0)^j, \quad \lambda_j \geqslant 0, \quad \sum_{j=1}^{n+1} \lambda_j = 1,$$

$$(g^i)^j \in G_{f_i}(y), \quad i = 1, \ldots, m, \quad j = 1, \ldots, n+1; \quad (g^0)^j \in G_{f_0}(z(y)).$$

Записав представление (1.13) для каждого j и просуммировав их с весами λ_j, для функции (1.14) получим

$$o(x, y, g) = \sum_{j=1}^{n+1} \lambda_j \sum_{i=1}^{m} (g^0_{(i)})^j o_i(x, y, (g^i)^j) + \sum_{j=1}^{n+1} \lambda_j o_0(z(x), z(y), (g^0)^j),$$

$$\frac{|o(x, y, g)|}{\|y - x\|} \leqslant c \sum_{j=1}^{n+1} \lambda_j \sum_{i=1}^{m} \frac{|o_i(x, y, (g^i)^j)|}{\|y - x\|} + \sum_{j=1}^{n+1} \lambda_j \frac{|o_0(z(x), z(y), (g^0)^j)|}{\|y - x\|} \leqslant$$

$$\leqslant c \sum_{j=1}^{n+1} \lambda_j \sum_{i=1}^{m} \frac{|o_i(x, y, (g^i)^j)|}{\|y - x\|} + L \sum_{j=1}^{n+1} \lambda_j \frac{|o_0(z(x), z(y), (g^0)^j)|}{\|z(y) - z(x)\|}. \tag{1.15}$$

Используя полученную оценку для $|o(x, y, g)| / \|y - x\|$, проверим выполнение условия (1.3). Пусть задано произвольное $\varepsilon > 0$. В силу обобщенной дифференцируемости $f_i(y)$ $(i = 1, \ldots, m)$ существует такое $\delta_1 \leqslant \delta$, что

$$\frac{|o_i(x, y, g^i)|}{\|y - x\|} \leqslant \frac{\varepsilon}{2cm} \quad \text{при} \quad \|y - x\| \leqslant \delta_1, \quad g^i \in G_{f_i}(y). \tag{1.16}$$

В силу обобщенной дифференцируемости $f_0(z)$ существует такое ε_1, что

$$\frac{|o_0(z(x), z, g^0)|}{\|z - z(x)\|} \leqslant \frac{\varepsilon}{2L} \quad \text{при} \quad \|z - z(x)\| \leqslant \varepsilon_1, \quad g^0 \in G_{f_0}(z). \tag{1.17}$$

В силу непрерывности $z(y)$ для данного ε_1 существует такое $\delta_2 \leqslant \delta_1$, что

$$\|z(y) - z(x)\| \leqslant \varepsilon_1 \quad \text{при} \quad \|y - x\| \leqslant \delta_2.$$

Таким образом, при $\|y-x\| \leqslant \delta_2$ и $g \in G_f(y)$ из (1.15)—(1.17) следует, что

$$|o(x, y, g)|/\|y-x\| \leqslant \varepsilon.$$

Условие (1.3) проверено, и теорема доказана.

Следствие 1.1. *Вогнутые и слабо вогнутые* [91] *функции обобщенно дифференцируемы, а их субградиенты и квазиградиенты можно взять в качестве псевдоградиентов этих функций.*

Следствие 1.2. *Функция минимума от конечного числа обобщенно дифференцируемых функций*

$$f(y) = \min\{f_1(y), \ldots, f_m(y)\}$$

обобщенно дифференцируема, и

$$G_f(y) = \operatorname{co}\{g \in E_n \mid g \in G_{f_i}(y),\ f_i(y) = f(y)\}. \tag{1.18}$$

Следствие 1.3. *Функция модуля обобщенно дифференцируемой функции* $f(x)$

$$h(x) = |f(x)| = \max\{f(x), -f(x)\}$$

обобщенно дифференцируема.

Следствие 1.4. *Функция Лагранжа для задачи математического программирования с обобщенно дифференцируемыми функциями обобщенно дифференцируема по совокупности прямых и двойственных переменных.*

Согласно теореме 1.1, обобщенно дифференцируемые функции являются локально липшицевыми. Следующие примеры показывают, что обобщенно дифференцируемые функции, вообще говоря, не имеют производных по направлениям, но не все липшицевы функции (и даже не все дифференцируемые функции) являются обобщенно дифференцируемыми.

Пример 1.1. Рассмотрим нетривиальный пример обобщенно дифференцируемой функции. Покажем, что функция

$$f(x) = \begin{cases} (2k - \ln|\ln|x||)\,|x|, & e^{-e^{2k}} \leqslant |x| \leqslant e^{-e^{2k-1}}, \\ (-2k + \ln|\ln|x||)\,|x|, & e^{-e^{2k+1}} \leqslant |x| \leqslant e^{-e^{2k}}, \\ 0, & x = 0, \end{cases}$$

где $k = 0, 1, \ldots,$ $x \in E_1,$ $|x| \leqslant e^{-1/e}$, обобщенно дифференцируема, но не имеет производных по направлениям в точке $x = 0$.

Убедимся, что кусочно заданная функция $f(x)$ непрерывна. Введем обозначения

$$f^+(x) = \lim_{y \to x+0} f(y), \quad f^-(x) = \lim_{y \to x-0} f(y).$$

Необходимо проверить непрерывность $f(x)$ в точках $x = e^{-e^n}$ $(n = 0, 1, \ldots)$. При $n = 2k$ имеем

$$f^+(e^{-e^{2k}}) = f^-(e^{-e^{2k}}) = 0.$$

При $n = 2k+1$ имеем

$$f^+(e^{-e^{2k+1}}) = f^-(e^{-e^{2k+1}}) = e^{-e^{2k+1}}.$$

Осталось проверить точку $x = 0$. Имеет место $0 \leqslant f(x) \leqslant |x|$; поэтому

$$\lim_{x \to 0} f(x) = 0.$$

Итак, $f(x)$ непрерывна.

Ясно, что $f(x)$ кусочно непрерывно дифференцируема, причем $(x > 0)$

$$f'(x) = \begin{cases} 2k - \ln|\ln x| + \dfrac{1}{|\ln x|}, & e^{-e^{2k}} < x < e^{-e^{2k-1}}, \\ -2k + \ln|\ln x| - \dfrac{1}{|\ln x|}, & e^{-e^{2k+1}} < x < e^{-e^{2k}}. \end{cases}$$

Обозначим $(x > 0)$

$$f'_+(x) = \lim_{y \to x+0} f'(y), \qquad f'_-(x) = \lim_{y \to x-0} f'(y).$$

При $x = e^{-e^{2k}}$ имеем

$$f'_+(x) = e^{-2k} = -f'_-(x).$$

При $x = e^{-e^{2k+1}}$ имеем

$$f'_+(x) = 1 - e^{-2k-1}, \qquad f'_-(x) = 1 + e^{-2k-1}.$$

Определим

$$G_f(x) = \begin{cases} f'(x), & x \neq e^{-e^n}, \quad n = 0, 1, \ldots, \\ \mathrm{co}\{f'_+(x), f'_-(x)\}, & x = e^{-e^n}, \quad n = 0, 1, \ldots, \\ [0, 1], & x = 0. \end{cases}$$

Замкнутость, выпуклость, ограниченность $G_f(x)$ и полунепрерывность сверху G_f очевидны. Дифференциальное представление (1.1) для $f(x)$, очевидно, имеет место в точках излома $x = e^{-e^n}$; в этих точках $f(x)$ имеет и производные по направлениям. Осталось показать, что в точке $x = 0$ также имеет место представление (1.1), но нет производной по направлению. Функция

$$\left| \frac{o(x, y, g)}{y} \right| = \left| \frac{f(y) - f(0)}{y} - g(y) \right| = \left| \frac{f(y)}{y} - g(y) \right| \leqslant \frac{1}{|\ln y|}$$

стремится к нулю при $y \to +0$ равномерно по $g \in \mathrm{co}\{f'_+(y), f'_-(y)\}$. Следовательно, представление (1.1) имеет место в точке $x = 0$. Рассмотрим отношение

$$f(y)/y = (f(y) - f(0))/y$$

и покажем, что оно не имеет предела при $y \to 0$. Действительно,

$$\lim_{k \to \infty} f(e^{-e^{2k}})/e^{-e^{2k}} = 0, \quad \lim_{k \to \infty} f(e^{-e^{2k+1}})/e^{-e^{2k+1}} = 1.$$

Следовательно, функция $f(x)$ не имеет производных по направлениям в точке $x = 0$.

Пример 1.2. Дифференцируемая в каждой точке функция

$$f(y) = y^2 \sin(1/y)$$

не является обобщенно дифференцируемой в нуле. Отметим, что она локально липшицева, но не является непрерывно дифференцируемой в нуле.

Действительно, функция

$$\frac{o(0, y, f'(y))}{y} = \cos\frac{1}{y} - 2y \sin\frac{1}{y}$$

не имеет предела при $y \to 0$.

Пример 1.3. Локально липшицева функция максимума

$$f(y) = \max_k \left(\frac{1}{2 \cdot 3^k} - \left| x - \frac{1}{3^k} \right|\right), \qquad k = 0, 1, \ldots, \infty,$$

не является обобщенно дифференцируемой в нуле. Этот пример показывает, что счетная операция взятия максимума над обобщенно дифференцируемыми функциями выводит из класса обобщенно дифференцируемых функций.

Действительно, $o(0, 3^{-k}, +1)/3^{-k}$ равно -1 и не стремится к нулю при $k \to \infty$.

Теорема 1.7. *Для обобщенно дифференцируемых функций $f(x)$ $(x \in E_n)$ справедлива обобщенная теорема Лагранжа о конечных приращениях, а именно*

$$f(y) - f(x) = (g, y - x), \qquad g \in G_f(x + \theta(y - x)), \qquad 0 < \theta < 1.$$

Доказательство. Рассмотрим функцию

$$\varphi(t) = f(x + t(y - x)) - t(f(y) - f(x)) - f(x).$$

Обобщенные градиенты функции $\varphi(t)$, согласно (1.12), равны

$$G_\varphi(t) = \{(g, y - x) - f(y) + f(x) \mid g \in G_f(x + t(y - x))\}. \quad (1.19)$$

Отметим, что $\varphi(0) = \varphi(1) = 0$. Поэтому всегда существует внутренняя экстремальная точка $t = \theta \in (0, 1)$ функции $\varphi(t)$. Согласно теореме 3.1, $0 \in G_\varphi(\theta)$ независимо от того, максимальная это или минимальная экстремальная точка. Тогда из (1.19) следует, что существует такой псевдоградиент $g \in G_f(x + \theta(y - x))$, что

$$(g, y - x) - f(y) + f(x) = 0.$$

Теорема доказана.

Теперь рассмотрим вопрос: однозначно ли задано определением 1.1 псевдоградиентное отображение обобщенно дифференцируемой функции? Мы покажем, что для одной и той же обобщенно дифференцируемой функции существует целое семейство псевдоградиентных отображений, удовлетворяющих определению 1.1. Однако среди них есть минимальное по включению отображение; оно совпадает с субдифференциальным отображением Кларка этой функции. Ввиду этого следует сказать, что формулы (1.8), (1.12), (1.18) вычисления псевдоградиентов сложных функций задают лишь некоторое, не обязательно минимальное, псевдоградиентное отображение. Именно поэтому в этих формулах имеет место равенство, а не включение в отличие от соответствующих формул (2.9), (2.13), (2.17) для локально липшицевых функций. Однако неоднозначность задания псевдоградиентных отображений не является препятствием к их применению как в теоретических вопросах, так и при построении численных методов минимизации обобщенно дифференцируемых функций. Все псевдоградиентные отображения совершенно равноправны, поскольку от них требуется только, чтобы они удовлетворяли определению 1.1.

Теорема 1.8. *Если некоторое псевдоградиентное отображение обобщенно-дифференцируемой функции произвольно изменить в конечном числе точек, но так, чтобы оно осталось компактнозначным, выпуклозначным и замкнутым, то измененное отображение остается псевдоградиентным для этой функции.*

Доказательство. Действительно, измененное отображение по-прежнему локально ограничено, замкнуто и, следовательно, полунепрерывно сверху. В точках изменения отображения разложение (1.1) сохраняется, поскольку псевдоградиенты этих точек, по существу, в него не входят. Записанное в остальных точках разложение (1.1) остается справедливым, но, быть может, для меньших окрестностей этих точек.

Теорема 1.9. *Пусть $f(x)$ и $h(x)$ — обобщенно дифференцируемые функции. Тогда наряду с $G_f(x)$ множества*

$$G'_f(x) = \begin{cases} G_f(x), & f(x) \neq h(x), \\ \operatorname{co}\{G_f(x),\ G_h(x)\}, & f(x) = h(x), \end{cases}$$

определяют другое псевдоградиентное отображение функции $f(x)$.

Эта теорема доказывается аналогично теореме 1.5. Здесь расширение G_f оказывается возможным, поскольку оно не нарушает определяющего разложения (1.1) и других требований к псевдоградиентному отображению.

Теорема 1.10. *Минимальное по включению псевдоградиентное отображение обобщенно дифференцируемой функции совпадает с субдифференциальным (обобщенно градиентным) отображением по Кларку (§ 2) этой функции.*

Доказательство. Напомним, что, согласно теореме 1.1, обобщенно дифференцируемая функция $f(x)$ является локально липшицевой, поэтому для нее определен субдифференциал Кларка $\partial f(x)$, который представляет собой множество таких векторов $g \in E_n$, что для любого направления $d \in E_n$ имеет место

$$(g, d) \leqslant \lim_{\delta \to 0} \sup_{\substack{\|y - x\| \leqslant \delta \\ 0 < \lambda \leqslant \delta}} [f(y + \lambda d) - f(y)]\, \lambda^{-1}.$$

Для $g \in \partial f(x)$ в силу теоремы 1.7 и полунепрерывности сверху G_f можем написать

$$(g, d) \leqslant \lim_{\delta \to 0} \sup_{\substack{\|y-x\| \leqslant \delta \\ 0<\lambda \leqslant \delta}} (g(y+\bar{\lambda}d), d) \leqslant \sup_{g \in G_f(x)} (g, d),$$

где

$$g(y+\bar{\lambda}d) \in G_f(y+\bar{\lambda}d), \quad 0<\bar{\lambda}<\lambda \leqslant \delta.$$

В силу выпуклости и замкнутости $G_f(x)$ и произвольности d отсюда следует, что

$$\partial f(x) \subset G_f(x).$$

Тогда ∂f вместе с G_f удовлетворяет разложению (1.1). Кроме того, $\partial f(x)$ — непустые ограниченные выпуклые замкнутые множества, а отображение $x \to \partial f(x)$ полунепрерывно сверху (§ 2). Следовательно, ∂f — минимальное по включению псевдоградиентное отображение для $f(x)$. Теорема доказана.

В следующих далее утверждениях устанавливается связь между обычной и обобщенной дифференцируемостью функций.

Теорема 1.11. *Пусть f — обобщенно дифференцируемая функция, G_f — ее псевдоградиентное отображение. Если в точке x множество $G_f(x)$ состоит из единственной точки, то функция f дифференцируема в x (по Фреше и по Гато).*

Доказательство. Обозначим $g(x) = G_f(x)$. Для f справедливо разложение (1.1), которое представим в следующем виде:

$$f(y) = f(x) + (g(x), y-x) + (g-g(x), y-x) + o(x, y, g),$$

где $g \in G_f(y)$. Заметим, что слагаемое

$$r(x, y) \equiv (g-g(x), y-x) + o(x, y, g)$$

фактически не зависит от g.

Таким образом,

$$f(y) = f(x) + (g(x), y-x) + r(x, y).$$

Справедлива оценка

$$\frac{|r(x, y)|}{\|y-x\|} \leqslant \sup_{g \in G_f(y)} \|g(x)-g\| + \sup_{g \in G_f(y)} \frac{|o(x, y, g)|}{\|y-x\|},$$

откуда при фиксированном x в силу полунепрерывности сверху G_f и условия (1.2) получаем

$$\lim_{y \to x} \frac{r(x, y)}{\|y-x\|} = 0,$$

т. е. функция $f(x)$ дифференцируема в точке x по Фреше, а следовательно, и по Гато. Теорема доказана.

Лемма 1.1. *Пусть $f(x)$ $(x \in E_1)$ — обобщенно дифференцируемая функция от одной переменной. Введем функцию разрыва градиента*

$$d(x) = \sup_{g \in G_f(x)} g - \inf_{g \in G_f(x)} g.$$

Тогда для любого $\varepsilon > 0$ *и любого ограниченного множества* $K \subset E_1$ *множество точек* ε*-разрыва градиента* $\{x \in K \mid d(x) \geqslant \varepsilon > 0\}$ *конечно.*

Доказательство. Предположим противное. Тогда существует такая бесконечная последовательность точек $\{x^k\}$, что $x^k \in K$ и $d(x^k) \geqslant \varepsilon > 0$. Выберем подпоследовательность $x^{k_s} \to x$ $(s = 0, 1, \ldots)$. Пусть

$$g_1^{k_s} = \sup_{g \in G_f(x^{k_s})} g, \qquad g_2^{k_s} = \inf_{g \in G_f(x^{k_s})} g.$$

В силу обобщенной дифференцируемости $f(x)$ имеем

$$f(x^{k_s}) = f(x) + (g_1^{k_s}, x^{k_s} - x) + o(x, x^{k_s}, g_1^{k_s}),$$

$$f(x^{k_s}) = f(x) + (g_2^{k_s}, x^{k_s} - x) + o(x, x^{k_s}, g_2^{k_s}),$$

причем

$$\lim_{s \to \infty} \frac{o(x, x^{k_s}, g_1^{k_s})}{\| x^{k_s} - x \|} = \lim_{s \to \infty} \frac{o(x, x^{k_s}, g_2^{k_s})}{\| x^{k_s} - x \|} = 0.$$

Отсюда получаем

$$(g_1^{k_s} - g_2^{k_s}, \; x^{k_s} - x) = o(x, x^{k_s}, g_2^{k_s}) - o(x, x^{k_s}, g_1^{k_s})$$

и приходим к противоречию:

$$\varepsilon \leqslant g_1^{k_s} - g_2^{k_s} \leqslant \frac{| o(x, x^{k_s}, g_1^{k_s}) |}{\| x^{k_s} - x \|} + \frac{| o(x, x^{k_s}, g_2^{k_s}) |}{\| x^{k_s} - x \|} \to 0$$

при $s \to \infty$. Лемма доказана.

Следствие 1.5. *Для обобщенно дифференцируемой функции* $f(x)$ $(x \in E_1)$ *от одной переменной множество точек неоднозначности (разрыва) ее псевдоградиентного отображения* $G_f(x)$ *не более чем счетно.*

Действительно, множество A точек разрыва псевдоградиента можно представить в виде счетного объединения конечных наборов точек

$$A = \bigcup_{i,j=1}^{\infty} \left\{ x \in E_1 \,\middle|\, |x| \leqslant i, \quad d(x) \geqslant \frac{1}{j} \right\}.$$

Определение 1.5. *Диаметром множества* $G \subset E_n$ *в направлении* $l \in E_n$ называется величина

$$d_G(l) = \sup_{g \in G}(g, l) - \inf_{g \in G}(g, l).$$

Легко видеть, что функция $d_G(l)$ выпукла по l.

Определение 1.6. *Диаметром множества* $G \subset E_n$ называется величина

$$d_G = \sup \{ \| g_1 - g_2 \| \mid g_1, g_2 \in G \}.$$

Нетрудно убедиться, что

$$d_G = \sup \{ d_G(l) \mid \| l \| \leqslant 1 \}.$$

Лемма 1.2. *Для того чтобы множество $G \subset E_n$ состояло из единственной точки, необходимо и достаточно, чтобы для некоторой системы n линейно независимых направлений $\{l_i\}$ было $d_G(l_i) = 0$ ($i = 1, \ldots, n$).*

Доказательство. Необходимость очевидна. Докажем достаточность. Ясно, что

$$d_G(-l_i) = 0, \qquad i = 1, \ldots, n.$$

Для любого направления $l \in E_n$ справедливо представление

$$l = \sum_{i=1}^{n} \alpha_i l_i = \sum_{i=1}^{n} |\alpha_i| \, l_i \operatorname{sign} \alpha_i,$$

где

$$\operatorname{sign} \alpha = \begin{cases} 1, & \alpha > 0, \\ 0, & \alpha = 0, \\ -1, & \alpha < 0. \end{cases}$$

В силу выпуклости функции $d_G(l)$ имеем

$$d_G(l) \leqslant \sum_{i=1}^{n} |\alpha_i| \, d_G(l_i \operatorname{sign} \alpha_i) = 0.$$

Следовательно,

$$d_G = \sup \{d_G(l) \mid \|l\| \leqslant 1\} = 0,$$

т. е. G состоит из одной точки. Лемма доказана.

Теорема 1.12. *Множество псевдоградиентов $G_f(x)$ обобщенно дифференцируемой функции $f(x)$ ($x \in E_n$) почти всюду в E_n состоит из одной точки, и, следовательно, обобщенно дифференцируемая функция почти всюду дифференцируема (по Фреше и по Гато), а ее градиент непрерывен на множестве, где он определен* *).

Доказательство. Пусть векторы l_i ($i = 1, \ldots, n$) являются ортами координатных осей в E_n. Обозначим через $d(x)$ диаметр множества $G_f(x)$, а через $d(x, l)$ его диаметр в направлении $l \in E_n$. В силу леммы 1.2 множество

$$A = \{x \in E_n \mid d(x) > 0\}$$

точек неоднозначности псевдоградиентного отображения $G_f(x)$ представимо в виде

$$A = \bigcup_{i=1}^{n} A_i, \qquad A_i = \{x \in E_n \mid d(x, l_i) > 0\}.$$

Определим множества

$$A_{ij} = \{x \in E_n \mid d(x, l_i) \geqslant 1/j\}, \qquad j = 1, 2, \ldots$$

---

*) Отображение $g(x)$ называется непрерывным на множестве X, если для любых $x \in X$ и $\{x^k\} \to x$ ($x^k \in X$) имеет место $g(x^k) \to g(x)$.

Очевидно, что $A = \bigcup\limits_{i,j} A_{ij}$. Легко показать, что функция $d(\cdot, l_i)$ полунепрерывна сверху; поэтому множества A_{ij} замкнуты и, следовательно, измеримы по Лебегу в E_n. Введем множества

$$K_m = \{x \in E_n \mid \max_{1 \leqslant s \leqslant n} |x_{(s)}| \leqslant m\}, \qquad A_{ijm} = A_{ij} \cap K_m.$$

Они также замкнуты и измеримы. Имеем $A = \bigcup\limits_{i,j,m} A_{ijm}$.

Покажем, что мера Лебега множеств A_{ijm} равна нулю. Определим гиперплоскость

$$H_i = \{x \in E_n \mid (x, l_i) = 0\}.$$

Для точек

$$x = (x_{(1)}, \ldots, x_{(i-1)}, 0, x_{(i+1)}, \ldots, x_{(n)}) \in H_i$$

обозначим проходящие через x прямые

$$\Pi_i(x) = \{x + \lambda l_i \mid -\infty < \lambda < \infty\}.$$

Пересечения H_i и $\Pi_i(x)$ с кубом K_m обозначим через H_{im} и $\Pi_{im}(x)$ соответственно. Рассмотрим множества

$$A_{ijm}(x) = A_{ijm} \cap \Pi_{im}(x).$$

Легко видеть, что они являются множествами точек разрыва псевдоградиента для обобщенно дифференцируемой функции

$$\varphi(\lambda) = f(x + \lambda l_i), \qquad |\lambda| \leqslant m.$$

Согласно лемме 1.1, множества $A_{ijm}(x)$ конечны или пусты; следовательно, их одномерная мера Лебега равна

$$\mu_1 A_{ijm}(x) = 0.$$

По теореме Фубини n-мерная мера Лебега имеет вид

$$\mu_n A_{ijm} = \int\limits_{H_{im}} \mu_1 A_{ijm}(x)\, \mu_{n-1} dx = 0.$$

В результате получаем

$$\mu_n A = \sum_{i,j,m} \mu_n A_{ijm} = 0.$$

Таким образом, почти всюду в E_n отображение $G_f(x)$ однозначно и непрерывно (т. е. полунепрерывно сверху и снизу). Отсюда в силу теоремы 1.11 следует, что обобщенно дифференцируемые функции почти всюду в E_n дифференцируемы (по Фреше и по Гато), и их градиенты непрерывны там, где они определены. Теорема доказана.

Теперь докажем для обобщенно дифференцируемых функций обобщенную формулу Ньютона — Лейбница. Отсюда будет следовать, что обобщенно дифференцируемая функция определяется любым своим псевдоградиентным отображением с точностью до константы.

Теорема 1.13. *Пусть* $f(x)$ $(x \in E_n)$ — *обобщенно дифференцируемая функция,* $g(x)$ — *произвольное однозначное сечение псевдоградиентного отображения* $G_f(x)$, *т. е.*

$$g(x) \in G_f(x).$$

Тогда имеет место интегральная формула

$$f(x+h) - f(x) = \int_0^1 (g(x+th),\ h)\, dt,$$

где интеграл понимается в смысле Лебега.

Доказательство. Рассмотрим функцию

$$\varphi(x, h) = \int_0^1 f(x+th)\, dt$$

и вычислим ее производную $\varphi_h'(x, h)$ по направлению h двумя способами. С одной стороны,

$$\varphi_h'(x, h) = \lim_{\alpha \to +0} \frac{1}{\alpha} [\varphi(x+\alpha h, h) - \varphi(x, h)] =$$

$$= \lim_{\alpha \to +0} \frac{1}{\alpha} \left[\int_\alpha^{1+\alpha} f(x+t'h)\, dt' - \int_0^1 f(x+th)\, dt \right] = f(x+h) - f(x).$$

С другой стороны,

$$\varphi_h'(x, h) = \lim_{\alpha \to +0} \int_0^1 \frac{1}{\alpha} [f(x+th+\alpha h) - f(x+th)]\, dt. \qquad (1.20)$$

Здесь очевидно, что подынтегральная функция ограничена по модулю константой Липшица при $t \in [0, 1]$ равномерно по $\alpha \in [0, 1]$. В силу обобщенной дифференцируемости $f(x)$ справедливо представление

$$\frac{1}{\alpha} [f(x+th+\alpha h) - f(x+th)] =$$

$$= (g(x+th+\alpha h), h) + \frac{1}{\alpha} o(x+th,\ x+th+\alpha h,\ g(x+th+\alpha h)),$$

причем

$$\lim_{\alpha \to +0} \frac{1}{\alpha} o(x+th,\ x+th+\alpha h,\ g) = 0.$$

Функция

$$\psi(t) = f(x+th), \qquad t \in [0, 1],$$

согласно теореме 1.6, является обобщенно дифференцируемой; ее псевдоградиентное отображение задается формулой

$$G_\psi(t) = \{(g, h) \mid g \in G_f(x+th)\}.$$

Согласно следствию 1.5, отображение $G_\psi(t)$ почти всюду на $[0, 1]$ однозначно и, следовательно, непрерывно. Отсюда следует, что почти для всех $t \in [0, 1]$

$$\lim_{\alpha \to +0} (g(x + th + \alpha h), h) = (g(x + th), h).$$

Теперь, согласно теореме Лебега, предел в (1.20) можно внести под знак интеграла. Таким образом, получаем соотношение

$$f(x + h) - f(x) = \varphi'_h(x, h) = \int_0^1 (g(x + th), h)\, dt.$$

Теорема доказана.

Лемма 1.3. *Пусть $f_1(x)$, $f_2(x)$ — обобщенно дифференцируемые функции и $G_{f_1}(x)$, $G_{f_2}(x)$ — их псевдоградиентные отображения, причем*

$$G_{f_1}(x) \cap G_{f_2}(x) \neq \varnothing$$

для всех $x \in X$ (X — выпуклое множество). Тогда

$$f_1(x) - f_2(x) = \text{const}.$$

Доказательство. Согласно теореме 1.6, функция

$$f(x) = f_1(x) - f_2(x)$$

является обобщенно дифференцируемой, а ее псевдоградиентное отображение задается формулой

$$G_f(x) = \{g \in E_n \mid g = g_1 - g_2, \quad g_1 \in G_{f_1}(x), \quad g_2 \in G_{f_2}(x)\}.$$

Так как

$$G_{f_1}(x) \cap G_{f_2}(x) \neq \varnothing,$$

то отсюда следует, что $0 \in G_f(x)$ для $x \in X$. Покажем, что в таком случае $f(x) = \text{const}$ при $x \in X$. Действительно, по теореме 1.13 для любых $x, y \in X$ имеем

$$f(y) = f(x) + \int_0^1 (g(x + t(y - x)), y - x)\, dt,$$

где

$$g(x + t(y - x)) \in G_f(x + t(y - x)).$$

Полагая

$$g(x + t(y - x)) \equiv 0,$$

получаем $f(y) = f(x)$. Лемма доказана.

Следствие 1.6. *Обобщенно дифференцируемая функция определяется любым своим псевдоградиентным отображением с точностью до константы.*

Следствие 1.7. *Полугладкая функция определяется своим субдифференциалом с точностью до константы, поскольку она является обобщенно дифференцируемой, и ее субдифференциал можно взять в качестве псевдоградиентного отображения.*

В заключение параграфа уточним связь между обобщенно дифференцируемыми и полугладкими функциями (см. определение 1.4).

Определение 1.7. Функция называется *обобщенно дифференцируемой в узком смысле* в области, если она обобщенно дифференцируема в области и имеет производную по любому направлению в каждой точке области.

Теорема 1.14. *Как классы обобщенно дифференцируемые в узком смысле и полугладкие функции совпадают.*

Доказательство. Пусть $f(x)$ — полугладкая функция. Согласно теореме 1.4, она является обобщенно дифференцируемой. Покажем, что $f(x)$ имеет производные по направлениям

$$f'(x;\ d) = \lim_{\lambda_k \to +0} [f(x + \lambda_k d) - f(x)]/\lambda_k.$$

В силу теоремы 2.9 можно представить

$$[f(x + \lambda_k d) - f(x)]/\lambda_k = (\bar{g}^k,\ d), \qquad \bar{g}^k \in \partial f(x + \theta_k \lambda_k d),$$
$$0 < \theta_k < 1.$$

Положим

$$y^k = x + \theta_k \lambda_k d, \quad (y^k - x)/\| y^k - x \| = d/\| d \|.$$

По определению полугладких функций существует $\lim\limits_{k\to\infty} (\bar{g}^k, d)$, и, следовательно, существует производная по направлению $f'(x;\ d)$. Итак, полугладкая функция является обобщенно дифференцируемой в узком смысле.

Пусть теперь $f(x)$ обобщенно дифференцируема и имеет производные по направлениям в каждой точке. Покажем, что она является полугладкой. Согласно теореме 1.1, $f(x)$ локально липшицева, а, согласно теореме 1.10, $\partial f(x)$ есть ее псевдоградиентное отображение. Пусть

$$y^k \to x, \qquad g^k \in \partial f(y^k),$$

причем

$$(y^k - x)/\| y^k - x \| \to d.$$

Справедливо разложение

$$f(y^k) = f(x) + (g^k,\ y^k - x) + o(x,\ y^k,\ g^k),$$

откуда

$$\lim_{k\to\infty}\left(g^k,\ \frac{y^k - x}{\| y^k - x \|}\right) = \lim_{k\to\infty} (g^k,\ d) = \lim_{k\to\infty} \frac{f(y^k) - f(x)}{\| y^k - x \|}.$$

Последний предел существует и равен производной функции f по направлению d. Действительно, примем

$$\lambda_k = (y^k - x,\ d)$$

и представим

$$y^k = x + \lambda_k d + r_k.$$

Нетрудно проверить, что

$$\lambda_k \to 0, \ r^k/\lambda_k \to 0, \quad \| y^k - x \|/\lambda_k \to 1.$$

Тогда

$$\lim_{k\to\infty} \frac{f(y^k) - f(x)}{\| y^k - x \|} = \lim_{k\to\infty} \left[\frac{f(x + \lambda_k d) - f(x)}{\lambda_k} \frac{\lambda_k}{\| y^k - x \|} + \right.$$

$$\left. + \frac{f(x + \lambda_k d + r_k) - f(x + \lambda_k d)}{\lambda_k} \frac{\lambda_k}{\| y^k - x \|} \right] = \lim_{k\to\infty} \frac{f(x + \lambda_k d) - f(x)}{\lambda_k}.$$

Мы показали, что предел скалярных произведений $\lim\limits_{k\to\infty} (g^k, d)$ всегда существует; следовательно, $f(x)$ — полугладкая функция. Теорема доказана.

Итак, как классы полугладкие и обобщенно дифференцируемые в узком смысле функции совпадают. Разница между ними заключается в способе определения обобщенных градиентов. Для полугладких функций используются готовые обобщенные градиенты Кларка, или, если смотреть с позиций теоремы 1.10, минимальные в определенном смысле обобщенные градиенты. Однако, как видно из результатов этого параграфа и будет подтверждено в дальнейшем, полезно рассматривать не только минимальные, но и более широкие обобщенно градиентные отображения. Например, это позволяет построить исчисление псевдоградиентов, которого нет для минимальных обобщенных градиентов Кларка. Кроме того, переход от полугладких к обобщенно дифференцируемым функциям позволяет без какого-либо ущерба ослабить требования к дифференцируемости рассматриваемых функций. И наконец, важным моментом является переход к разложению (1.1), из которого, собственно, и следуют все результаты.

§ 2. Сглаживание функций и основные свойства локально липшицевых функций

Успешный прием, с помощью которого исследуются локально липшицевы функции и развиваются численные методы их решения, связан с рассмотрением предельных экстремальных задач такого типа: имеется последовательность функций $f(x, \alpha)$, которая сходится при $\alpha \to 0$ к $f(x)$; требуется, обращаясь только к функциям последовательности $f(x, \alpha)$, найти минимум $f(x)$ в области X. Операция предельного перехода может портить некоторые хорошие свойства функций $f(x, \alpha)$, что характерно для задач аппроксимации, когда «плохая» по каким-то причинам функция приближается последовательностью «хороших». В таких случаях возникает интересная задача оптимизации предельной функции $f(x)$ с использованием информации о членах последовательности $f(x, \alpha)$.

В этом параграфе изучение ряда свойств локально липшицевых функций проводится на основе исследования свойств сглаженных функций $f(x, \alpha)$, сходящихся при $\alpha \to 0$ к $f(x)$. Искусственное усреднение функции $f(x)$ позволяет во многих случаях сгладить и уменьшить ее колебательный характер. На практике наиболее часто применяются линейные операторы усреднения

$$f(x, \alpha) = \int\limits_{E_n} h_\alpha(u) f(x+u)\, du,$$

где $h_\alpha(u)$ — весовая функция (ядро оператора), $u \in E_n$, α — неотрицательный скалярный параметр.

В теории обобщенных функций рассматривается бесконечно дифференцируемая функция

$$f(x, \alpha) = \int\limits_{-\alpha}^{\alpha} \omega_\alpha(u) f(x+u)\, du,$$

где

$$\omega_\alpha(u) = \begin{cases} C_\alpha \exp\left\{-\dfrac{\alpha^2}{\alpha^2 - |u|^2}\right\}, & |u| \leqslant \alpha, \\ 0, & |u| > \alpha; \end{cases}$$

C_α выбирается из условия

$$\int\limits_{-\alpha}^{\alpha} \omega_\alpha(u)\, du = 1.$$

В настоящей книге изучаются сглаженные функции следующего вида:

$$f(x, \alpha) = \frac{1}{(2\alpha)^n} \int\limits_{-\alpha}^{\alpha} \cdots \int\limits_{-\alpha}^{\alpha} f(x+y)\, dy_1 \ldots dy_n. \tag{2.1}$$

Класс локально липшицевых функций чрезвычайно широк; он охватывает практически все изучаемые в математическом программировании классы непрерывных функций. Вместе с тем локально липшицевы функции сохраняют ряд полезных свойств, которые позволяют ввести понятие «обобщенный градиент» и построить численные процедуры определения стационарных точек.

Лемма 2.1. *У локально липшицевой функции частная производная*

$$\partial f(x)/\partial x_i, \qquad i = 1, \ldots, n,$$

существует всюду, за исключением множества меры нуль.

Доказательство. Из того, что функция удовлетворяет условию Липшица, следует, что она абсолютно непрерывна по каждой переменной x_i ($i = 1, \ldots, n$) при фиксированных остальных переменных. Напомним определение абсолютно непрерывной функции.

Пусть на отрезке $[a, b]$ задана конечная функция $g(t)$. Если для любого $\varepsilon > 0$ существует такое $\delta > 0$, что для любой конечной системы взаимно непересекающихся интервалов $(a_1, b_1), \ldots, (a_m, b_m)$

$$\sum_{i=1}^{m} (b_i - a_i) < \delta$$

и выполняется неравенство

$$\sum_{i=1}^{m} |g(b_i) - g(a_i)| < \varepsilon,$$

то $g(t)$ *абсолютно непрерывна.*

Абсолютно непрерывная функция обладает следующим свойством: почти всюду в отрезке $[a, b]$ (т. е. за исключением множества меры нуль) существует конечная производная. Таким образом, у локально липшицевых функций почти всюду по переменной x_i ($i = 1, \ldots, n$) существует частная производная $\partial f(x)/\partial x_i$. Отсюда по теореме Фубини о произведении мер [58] $f(x)$ имеет почти всюду в E_n обычную частную производную $\partial f(x)/\partial x_i$ ($i = 1, \ldots, n$).

Оказывается, что результат леммы 2.1 можно усилить.

Теорема 2.1. *Локально липшицева функция дифференцируема почти всюду, т. е. $\nabla f(x)$ существует всюду, за исключением множества меры нуль.*

Этот факт известен как теорема Радемахера [173]; другое доказательство теоремы содержится в работе [55]. На основе этого свойства понятие «обобщенный градиент $\partial f(x)$ локально липшицевой функции $f(x)$ в точке x» первоначально было введено как выпуклое замыкание множества точек y вида

$$y = \lim_{k\to\infty} \nabla f(x^k),$$

где $\{x^k\}$ — последовательность, сходящаяся к x и такая, что $f(x)$ дифференцируема в каждой точке x^k.

В дальнейшем мы будем пользоваться другим, во многих случаях более удобным эквивалентным определением множества $\partial f(x)$.

Локально липшицева функция может не иметь обычной производной по направлению

$$f'(x; e) = \lim_{\lambda\to 0} [f(x + \lambda e) - f(x)]/\lambda.$$

Обобщенная производная функции $f(x)$ в точке x по направлению e, обозначаемая $f^0(x; e)$, определяется по формуле

$$f^0(x; e) = \lim_{\delta\downarrow 0} \sup_{\substack{\|v\|\leqslant\delta \\ 0<\lambda\leqslant\delta}} [f(x + v + \lambda e) - f(x + v)]/\lambda. \tag{2.2}$$

Лемма 2.2. *Функция $e \to f^0(x; e)$ выпукла.*

Доказательство. В силу теоремы 4.7 [109] функция $f(x)$ выпукла, если она положительно однородна и полуаддитивна, т. е.

$$f(\lambda x) = \lambda f(x), \qquad 0 < \lambda < \infty,$$
$$f(x + y) \leqslant f(x) + f(y) \qquad \forall x, y.$$

Положительная однородность очевидна, а полуаддитивность вытекает из цепочки соотношений

$$f^0(x;\ e_1+e_2)=\lim_{\delta\downarrow 0}\sup_{\substack{\|v\|\leqslant\delta\\ 0<\lambda\leqslant\delta}}[f(x+v+\lambda e_1+\lambda e_2)-f(x+v)]/\lambda\leqslant$$

$$\leqslant\lim_{\delta\downarrow 0}\sup_{\substack{\|v\|\leqslant\delta\\ 0<\lambda\leqslant\delta}}[f(x+v+\lambda e_1+\lambda e_2)-f(x+v+\lambda e_2)]/\lambda+$$

$$+\lim_{\delta\downarrow 0}\sup_{\substack{\|v\|\leqslant\delta\\ 0<\lambda\leqslant\delta}}[f(x+v+\lambda e_2)-f(x+v)]/\lambda=f^0(x;\ e_1)+f^0(x;\ e_2).$$

Определение 2.1. *Обобщенным градиентом* функции $f(x)$ в точке x, обозначаемым $\partial f(x)$, называется субдифференциал выпуклой функции $f^0(x;\ \cdot)$ в нуле, т. е. $y\in\partial f(x)$, если

$$f^0(x;\ e)\geqslant(y,\ e)\qquad\forall e. \tag{2.3}$$

Следовательно, множество $\partial f(x)$ выпукло и замкнуто, а в силу неравенства $|f^0(x;\ e)|\leqslant L\|e\|$ и ограничено.

Из неравенства (2.3) следует, что если $f(x)$ дифференцируема в точке x, то

$$\nabla f(x)\in\partial f(x).$$

Теорема 2.2. *Справедлива формула*

$$f^0(x;\ e)=\max_{y\in\partial f(x)}(y,\ e). \tag{2.4}$$

Доказательство. Согласно теореме 13.1 монографии [109], каждое выпуклое множество M определяется своей опорной функцией $\sup\limits_{y\in M}(y,\ e)$, где e — произвольный вектор. Это означает, что $h\in\bar{M}$ тогда и только тогда, когда

$$(h,\ e)\leqslant\sup_{y\in M}(y,\ e)\qquad\forall e. \tag{2.5}$$

Помимо этого, если на множествах M_1 и M_2

$$\sup_{y\in M_1}(y,\ e)=\sup_{y\in M_2}(y,\ e), \tag{2.6}$$

то $\bar{M}_1=\bar{M}_2$. Поэтому формула (2.4) следует из соотношений (2.3) и (2.5).

Теорема 2.3. *Пусть A — выпуклое замкнутое множество. Тогда*

$$\partial f(x)\subset A,$$

если

$$f^0(x;\ e)\leqslant\max_{u\in A}(u,\ e)\qquad\forall e. \tag{2.7}$$

Доказательство. Если существует некоторый вектор $y\notin A$, $y\in\partial f(x)$, то по теореме отделимости найдется такой вектор e, что

$$\max_{u\in A}(e,\ u)<(e,\ y).$$

Но это неравенство противоречит (2.7).

Другое эквивалентное определение множества $\partial f(x)$ заключается в следующем.

Теорема 2.4. *Множество $\partial f(x)$ есть выпуклое замыкание множества всех точек вида*

$$G = \lim_{k\to\infty} \nabla f(x^k),$$

где $\{x^k\}$ — последовательность, сходящаяся к x и такая, что $f(x)$ дифференцируема в каждой точке x^k.

Доказательство. Обозначим $M = \mathrm{co}\, G$. Легко заметить, что множество M выпукло и ограничено. Его замкнутость следует из замкнутости множества G. Пусть $\{y^k\} \in G$ и $y^k \to y$. Нужно показать, что $y \in G$. Для каждого номера k найдется такая точка x^k, что

$$\| y^k - \nabla f(x^k) \| \leqslant 1/k, \qquad \| x^k - x \| \leqslant 1/k,$$

причем в точках $x = x^k$ функция $f(x)$ дифференцируема. Поэтому

$$\| y - \nabla f(x^k) \| \leqslant \| y - y^k \| + \| y^k - \nabla f(x^k) \| \leqslant \| y - y^k \| + 1/k,$$

т. е.

$$\| y - \nabla f(x^k) \| \to 0 \quad \text{при} \quad k \to \infty,$$

откуда следует, что $y \in G$.

Первоначально в работе [145] множество обобщенных градиентов локально липшицевой функции определялось как $\partial f(x) = \mathrm{co}\, G$. При таком определении $\partial f(x)$ было показано, что

$$f^0(x, e) = \max_{y \in \mathrm{co}\, G} (y, e). \tag{2.8}$$

Отсюда, сопоставляя соотношения (2.4), (2.6) и (2.8), получаем утверждение теоремы.

Важное свойство заключается в замкнутости точечно-множественного отображения $x \to \partial f(x)$.

Теорема 2.5. *Пусть*

$$g(x^k) \in \partial f(x^k), \quad x^k \to x, \quad g(x^k) \to g.$$

Тогда $g \in \partial f(x)$.

Доказательство. По определению имеем

$$f^0(x^k, e) \geqslant (g(x^k), e).$$

Поэтому существуют такие $v^k \in E_n$ и $\lambda_k > 0$, что

$$\frac{1}{\lambda_k} [f(x^k + v^k + \lambda_k e) - f(x^k + v^k)] \geqslant (g(x^k), e) - \frac{1}{k},$$

где

$$\| v^k \| + \lambda_k < 1/k.$$

Так как $(g(x^k), e)$ сходится к (g, e), то из последнего неравенства вытекает неравенство

$$f^0(x; e) \geqslant (g, e).$$

Поэтому по определению обобщенного градиента $g \in \partial f(x)$, что и требовалось доказать.

Теорема 2.6. *Если $f(x)$ — выпуклая функция, то множество $\partial f(x)$ совпадает с обычным субдифференциалом для $f(x)$.*

Доказательство. Известно, что у выпуклой функции существует производная по направлению $f'(x; e)$, для которой

$$f'(x; e) = \max_{y \in \hat{\partial} f(x)} (y, e),$$

где $\hat{\partial} f(x)$ — субдифференциал. Из условий (2.4), (2.6) вытекает, что для доказательства достаточно показать равенство

$$f'(x; e) = f^0(x; e).$$

Неравенство

$$f'(x; e) \leqslant f^0(x; e)$$

очевидно. Для выпуклой функции выполняется теорема о среднем

$$\frac{f(x + v + \lambda e) - f(x + v)}{\lambda} = (g(x + v + t\lambda e), e), \quad 0 \leqslant t \leqslant 1,$$

$$g(x + v + t\lambda e) \in \hat{\partial} f(x + v + t\lambda e),$$

откуда в силу замкнутости отображения $x \to \hat{\partial} f(x)$ выполняется неравенство

$$f^0(x; e) \leqslant \max_{y \in \hat{\partial} f(x)} (y, e) = f'(x; e).$$

Рассмотрим теперь вопрос, связанный с применением обычных операций сложения и умножения локально липшицевых функций.

Теорема 2.7. *Если функции $f_1(x)$ и $f_2(x)$ удовлетворяют локальному условию Липшица, то*

$$f_3(x) = f_1(x) \pm f_2(x), \quad f_4(x) = f_1(x) f_2(x)$$

также принадлежит этому классу.

Доказательство. Пусть L_1, L_2 — константы Липшица функций f_1, f_2 относительно произвольного ограниченного множества S. Мы должны показать, что функции f_3, f_4 удовлетворяют в S условию Липшица.

Для f_3 справедливо выражение

$$\begin{aligned} |f_3(x) - f_3(y)| &= |f_1(x) \pm f_2(x) - f_1(y) \mp f_2(y)| \leqslant \\ &\leqslant |f_1(x) - f_1(y)| + |\pm f_2(x) \mp f_2(y)| \leqslant \\ &\leqslant L_1 \|x - y\| + L_2 \|x - y\| = L \|x - y\|, \end{aligned}$$

где $L = L_1 + L_2$.

Пусть a_1 и a_2 — верхние границы функций $|f_1(x)|$, $|f_2(x)|$ на множестве $\bar{S}$, где $\bar{S}$ — замыкание S; $L = \max\{L_1, L_2\}$. Тогда

$$|f_4(x) - f_4(y)| = |f_1(x) f_2(x) - f_1(y) f_2(y)| =$$
$$= |f_1(x) f_2(x) - f_1(x) f_2(y) + f_1(x) f_2(y) - f_1(y) f_2(y)| \leqslant$$
$$\leqslant |f_1(x)| \times |f_2(x) - f_2(y)| + |f_2(y)| \times |f_1(x) - f_1(y)| \leqslant$$
$$\leqslant a_1 L_2 \| x - y \| + a_2 L_1 \| x - y \| \leqslant (a_1 + a_2) L \| x - y \|,$$

где $x, y \in S$.

Теорема 2.8. *Справедлива формула*

$$\partial (f_1(x) + f_2(x)) \subset \partial f_1(x) + \partial f_2(x). \tag{2.9}$$

Доказательство. Воспользуемся теоремой 2.3. Пусть

$$A = \partial f_1(x) + \partial f_2(x).$$

Тогда

$$\max_{u \in \partial f_1(x) + \partial f_2(x)} (u, e) = f_1^0(x; e) + f_2^0(x; e).$$

Осталось заметить, что

$$(f_1 + f_2)^0 (x; e) \leqslant f_1^0(x; e) + f_2^0(x; e).$$

Легко видеть, что в соотношении (2.9) может иметь место строгое включение.

Весьма успешный прием, с помощью которого удается вывести ряд других интересных свойств локально липшицевых функций, состоит в аппроксимации функции $f(x)$ последовательностью сглаженных функций $f(x, \alpha)$, сходящихся к $f(x)$ при $\alpha \to 0$. Исследуем функцию $f(x, \alpha)$ вида (2.1):

$$f(x, \alpha) = \frac{1}{(2\alpha)^n} \int_{-\alpha}^{\alpha} \ldots \int_{-\alpha}^{\alpha} f(x + y)\, dy_1 \ldots dy_n,$$

где $\alpha > 0$.

Ниже мы часто будем пользоваться следующим фактом.

Лемма 2.2*. *Последовательность функций $f(x, \alpha)$ в любой ограниченной области пространства E_n равномерно сходится к $f(x)$ при $\alpha \to 0$.*

Доказательство леммы естественным образом следует из соотношений

$$f(x, \alpha) = \frac{1}{(2\alpha)^n} \int_{-\alpha}^{\alpha} \ldots \int_{-\alpha}^{\alpha} [f(x + y) - f(x)]\, dy + f(x) \leqslant f(x) + \sqrt{n} L\alpha,$$

где L — константа Липшица функции $f(x)$.

Сформулируем важные дифференциальные свойства функции $f(x, \alpha)$.

Л е м м а 2.3. *Градиент функции $f(x, \alpha)$ в точке x определяется по формуле*

$$\nabla f(x, \alpha) = \frac{1}{(2\alpha)^n} \int_{-\alpha}^{\alpha} \dots \int_{-\alpha}^{\alpha} \nabla f(x+y)\, dy. \tag{2.10}$$

Доказательство вытекает из теоремы Лебега о предельном переходе под знаком интеграла.

Л е м м а 2.4. *Градиент функции $f(x, \alpha)$ в любой ограниченной области пространства E_n удовлетворяет условию Липшица с константой $L = C/\alpha$, где C — некоторая постоянная.*

Д о к а з а т е л ь с т в о. Частная производная $\partial f(x, \alpha)/\partial x_1$ определяется выражением

$$\frac{\partial f(x, \alpha)}{\partial x_1} = \frac{\partial}{\partial x_1} \frac{1}{(2\alpha)^n} \int_{-\alpha}^{\alpha} \dots \int_{-\alpha}^{\alpha} f(x+y)\, dy =$$

$$= \frac{\partial}{\partial x_1} \frac{1}{(2\alpha)^n} \int_{x_1-\alpha}^{x_1+\alpha} \dots \int_{x_n-\alpha}^{x_n+\alpha} f(y)\, dy =$$

$$= \frac{1}{(2\alpha)^n} \int_{x_2-\alpha}^{x_2+\alpha} \dots \int_{x_n-\alpha}^{x_n+\alpha} [f(x_1+\alpha, y_2, \dots, y_n) - f(x_1-\alpha, y_2, \dots, y_n)]\, dy_2 \dots dy_n =$$

$$= \frac{1}{(2\alpha)^n} \int_{-\alpha}^{\alpha} \dots \int_{-\alpha}^{\alpha} [f(x_1+\alpha, x_2+z_2, \dots, x_n+z_n) -$$

$$- f(x_1-\alpha, x_2+z_2, \dots, x_n+z_n)]\, dz_2 \dots dz_n, \tag{2.11}$$

которое следует из формулы дифференцирования интеграла по верхнему пределу. Аналогично вычисляются остальные частные производные.

Согласно формуле (2.11), частная производная $\partial f(x, \alpha)/\partial x_1$ удовлетворяет условию Липшица с константой $L = C/\alpha$, т. е.

$$\left| \frac{\partial f(x, \alpha)}{\partial x_1} - \frac{\partial f(y, \alpha)}{\partial y_1} \right| \leqslant \frac{C}{\alpha} \| x - y \|.$$

Отсюда в свою очередь вытекает, что

$$\| \nabla f(x, \alpha) - \nabla f(y, \alpha) \| \leqslant \frac{C\sqrt{n}}{\alpha} \| x - y \|.$$

Формулы (2.10), (2.11) показывают, что существуют два способа вычисления градиента сглаженной функции: при одном $\nabla f(x, \alpha)$ определяется градиентами функции $f(x)$, при другом — через конечные разности $f(x)$. Эти формулы окажутся полезными при построении численных методов минимизации липшицевых функций. В соответствии с ними будут обоснованы два типа методов: конечно-разностные методы и методы обобщенного градиентного спуска.

В дальнейшем часто используется следующая характеристика градиента функции $f(x, \alpha)$.

Лемма 2.5. *Пусть $\alpha_k \to 0$. Для любого $\varepsilon > 0$ существуют такие $\delta(x) > 0$ и номер $\bar{k}$, что для всех точек y таких, что $\|y - x\| \leqslant \delta$, при $k \geqslant \bar{k}$ выполняется неравенство*

$$\rho(\nabla f(y, \alpha_k), \partial f(x)) \equiv \min_{g \in \partial f(x)} \|\nabla f(y, \alpha_k) - g\| \leqslant \varepsilon.$$

Доказательство. Предварительно покажем следующее: для любого $\varepsilon > 0$ найдется такое $\bar{\delta} > 0$, что для всех точек y, в которых функция $f(x)$ дифференцируема и таких, что $\|y - x\| \leqslant \bar{\delta}$, имеет место неравенство

$$\rho(\nabla f(y), \partial f(x)) \leqslant \varepsilon.$$

Предположим противное: существует такое ε, что для любых $\bar{\delta}$ найдутся точки $\tilde{y}$, $\|\tilde{y} - x\| \leqslant \bar{\delta}$, для которых $\rho(\nabla f(\tilde{y}), \partial f(x)) > \varepsilon$. Тогда предельные точки последовательности $\nabla f(\tilde{y})$ при $\bar{\delta} \to 0$ не принадлежат $\partial f(x)$. Таким образом, получаем противоречие с определением множества $\partial f(x)$. Выберем требуемое в условии леммы значение $\delta = \bar{\delta}/2$, а номер $\bar{k}$ таким, чтобы $\alpha_k \leqslant \bar{\delta}/(2\sqrt{n})$ при $k \geqslant \bar{k}$. Применив лемму 2.3, получим утверждение леммы.

Следствие 2.1.

$$\lim_{k \to \infty} \rho(\nabla f(x, \alpha_k), \partial f(x)) = 0.$$

Лемма 2.5 означает, что при $\alpha_k \to 0$ градиенты сглаженной функции $f(x, \alpha_k)$, вычисленные в достаточно малой окрестности точки x, при достаточно больших k близки к множеству обобщенных градиентов функции $f(x)$ в точке x. Это свойство играет большую роль при исследовании сходимости численных методов минимизации липшицевых функций.

Лемма 2.6. *Пусть*

$$\{x^k\} \to x, \quad \nabla f(x^k, \alpha_k) \to g, \quad \alpha_k \to 0, \quad k \to \infty.$$

Тогда $g \in \partial f(x)$.

Доказательство этого утверждения непосредственно вытекает из леммы 2.5.

Продолжим изучение свойств локально липшицевых функций. Для них справедлив аналог теоремы Лагранжа о среднем.

Теорема 2.9. *Пусть точки x, y принадлежат выпуклому множеству из E_n. Тогда существуют такие $g \in \partial f(x + t(y - x))$, $t \in [0, 1]$, что*

$$f(y) - f(x) = (g, y - x). \tag{2.12}$$

(Можно даже утверждать [159] существование $t \in (0, 1)$.)

Доказательство следует из формулы о среднем для $f(x, \alpha_k)$:

$$f(y, \alpha_k) - f(x, \alpha_k) = (\nabla f(x + t_k(y - x), \alpha_k), y - x), \quad 0 \leqslant t_k \leqslant 1.$$

Перейдя к пределу при $k \to \infty$ в этом неравенстве и применив лемму 2.6, получим соотношение (2.12).

Многозначное отображение $x \to \partial f(x)$ обладает интересной особенностью.

Теорема 2.10. *Следующие утверждения эквивалентны:*

а) *Множество $\partial f(x)$ состоит из единственного вектора $g(x)$.*

б) *В точке x существует градиент $\nabla f(x) = g(x)$, который непрерывен в x относительно множества точек, где он определен.*

Доказательство. Из условия б) и теоремы 2.4 непосредственно вытекает а). Покажем, что из а) следует б). Существование $\nabla f(x)$ в точке x следует из соотношения (2.12) и замкнутости отображения $x \to \partial f(x)$, а непрерывность $\nabla f(x)$ в точке x непосредственно вытекает из условия а).

Следующее свойство характеризует множество обобщенных градиентов функции максимума.

Теорема 2.11. *Пусть $f_i(x)$ $(i = 1, \ldots, m)$ удовлетворяют локальному условию Липшица. Тогда*

$$\varphi(x) = \max_{1 \leqslant i \leqslant m} f_i(x)$$

есть функция из этого класса и справедлива формула

$$\partial\varphi(x) \subset \operatorname{co}\{\partial f_i(x) \mid i \in I(x) = \{i \mid \varphi(x) = f_i(x)\}\}. \tag{2.13}$$

Доказательство. Пусть x, y — произвольные точки некоторого ограниченного множества, для которого константа Липшица функции $f_i(x)$ есть L_i $(i = 1, \ldots, m)$. Тогда

$$\varphi(x) = f_\nu(x) \leqslant f_\nu(y) + L_\nu \|x - y\| \leqslant \varphi(y) + L \|x - y\|,$$

где

$$\nu \in I(x), \qquad L = \max_{1 \leqslant i \leqslant m} L_i.$$

Так как x, y произвольны, справедливо и обратное неравенство. Отсюда следует, что $\varphi(x)$ удовлетворяет локальному условию Липшица.

Из формулы (2.12) вытекает, что

$$\begin{gathered} \frac{f_i(x + \lambda e) - f_i(x)}{\lambda} = (g^i(x + t\lambda e), e), \quad 0 \leqslant t \leqslant 1, \\ g^i(x + t\lambda e) \in \partial f_i(x + t\lambda e), \quad i = 1, \ldots, m. \end{gathered} \tag{2.14}$$

Так как множество $\{1, \ldots, m\}$ конечно, то существует такая подпоследовательность $\{\lambda_s\} \to 0$, $s \to \infty$, что

$$i_{\lambda_s} \in I(x + \lambda_s e), \quad i_{\lambda_s} = \nu. \tag{2.15}$$

Легко заметить, что $\nu \in I(x)$. Если $\nu \notin I(x)$, то $f_\nu(x) < \varphi(x)$, и из непрерывности f_ν, φ для достаточно малых λ_s следует

$$f_\nu(x + \lambda_s e) < \varphi(x + \lambda_s e),$$

что противоречит (2.15).

Пусть в точке x существует $\nabla\varphi(x)$. Поскольку отображение $x \to \partial f(x)$ замкнуто, то, перейдя к пределу при $\lambda_s \to 0$ в равенстве

$$\frac{\varphi(x+\lambda_s e)-\varphi(x)}{\lambda_s} = \frac{f_\nu(x+\lambda_s e)-f_\nu(x)}{\lambda_s},$$

из соотношения (2.14) получаем

$$\nabla\varphi(x) = g^\nu(x) \in \partial f_\nu(x). \qquad (2.16)$$

Пусть имеется последовательность $x^k \to x$ точек, в которых существует $\nabla\varphi(x)$ и $\lim\limits_{k\to\infty}\nabla\varphi(x^k) = g(x)$. Из соотношений (2.16) следует, что

$$\nabla\varphi(x^k) \in \partial f_{\nu_k}(x^k), \quad \nu_k \in I(x^k).$$

Поэтому

$$g(x) \in \partial f_i(x), \quad i \in I(x).$$

Следовательно,

$$\partial\varphi(x) \subset \mathrm{co}\,\{\partial f_i(x) \mid i \in I(x)\},$$

что и требовалось доказать.

Покажем, что в соотношении (2.13) может выполняться строгое включение. Пусть $f_1(x) = -|x|$, $f_2(x) = x$. Тогда

$$f(x) = \max\{f_1(x), f_2(x)\} = x;$$

$\partial f(0) = \{1\}$, в то время как $\partial f_1(0)$ состоит из отрезка $[-1, 1]$.

Аналогичная ситуация складывается и для сложных функций. Обозначим через $[v^1 \ldots v^m]$ матрицу размера $n\times m$, i-й столбец которой есть вектор $v^i \in E_n$ для $i = 1, \ldots, m$.

Теорема 2.12. *Пусть f_i: $E_n \to E_1$, $i = 1, \ldots, m$, и пусть F: $E_m \to E_1$ — локально липшицевы функции. Для $x \in E_n$ определим*

$$Y(x) = (f_1(x), \ldots, f_m(x)), \quad \Phi(x) = F(Y(x)),$$
$$H(x) = \mathrm{co}\,\{g \in E_n \mid g = [g^1 \ldots g^m]\,\omega\},$$
$$g^i \in \partial f_i(x), \quad i = 1, \ldots, m, \quad \omega \in \partial F(Y(x)).$$

Тогда функция $\Phi(x)$ удовлетворяет локальному условию Липшица и имеет место формула

$$\partial\Phi(x) \subset H(x). \qquad (2.17)$$

Доказательство. По теореме 2.9

$$\frac{F(Y(x+v+\lambda e)) - F(Y(x+v))}{\lambda} = (\bar\omega, Y(x+v+\lambda e) - Y(x+v)),$$

где

$$\bar\omega \in \partial F(Y(x+v) + t(Y(x+v+\lambda e) - Y(x+v))).$$

Точно так же

$$f_i(x+v+\lambda e) - f_i(x+v) = \lambda(\bar g^i, e),$$
$$\bar g^i \in \partial f_i(x+v+t\lambda e), \quad i = 1, \ldots, m.$$

Следовательно, из определения $f^0(x; e)$ и из теоремы 2.5 вытекает, что

$$f^0(x; e) \leqslant \max_{u \in H(x)} (u, e).$$

Поэтому, применив теорему 2.3, получаем требуемое соотношение (2.17).

Формулы (2.9), (2.13), (2.17) показывают, что для липшицевых функций нельзя построить исчисления градиентов, как это имело место для обобщенно дифференцируемых функций. Таким образом, в липшицевом программировании нельзя широко применять методы обобщенного градиентного спуска. Поэтому важное значение приобретают конечно-разностные методы, которые исследуются в последующих главах.

§ 3. Необходимые условия экстремума

Выяснение необходимых условий экстремума важно по нескольким причинам. Во многих случаях эти условия помогают предсказать структуру оптимизационных методов. Они играют большую роль при исследовании сходимости алгоритмов. Целесообразно при этом вывести такие условия, которые были бы применимы к широкому классу задач, чтобы не строить в каждом отдельном случае свои соотношения.

Необходимые условия экстремума в задачах минимизации обобщенно дифференцируемых и липшицевых функций в отличие от выпуклых функций характеризуют только локальные свойства. Иными словами, если в точке $x = x^*$ выполняются необходимые условия экстремума, то, не делая никаких дополнительных предположений о поведении функции, можно выдвинуть гипотезу о том, что в точке $x = x^*$ достигается локальный экстремум $f(x)$. В настоящем параграфе необходимые условия для указанных классов формулируются в такой форме, которая наиболее удобна для исследования ходимости методов минимизации, рассматриваемых в дальнейшем.

Сначала мы кратко рассмотрим необходимые условия экстремума в задачах математического программирования с обобщенно дифференцируемыми функциями. Изложение будет носить, скорее, иллюстративный характер с целью показать принципиальную возможность вывода необходимых условий экстремума только из определения 1.1. В детальном исследовании здесь нет необходимости, поскольку оно будет проведено далее для более общего класса локально липшицевых функций. Дело в том, что обобщенно дифференцируемые функции являются локально липшицевыми. Более того, как показано в теореме 1.10, обобщенные градиенты Кларка этих функций всегда являются и псевдоградиентами этих функций; поэтому необходимые условия экстремума для задач с обобщенно дифференцируемыми функциями в терминах обобщенных градиентов Кларка совпадают с необходимыми условиями экстремума в терминах псевдоградиентов.

Рассмотрим следующую задачу:

$$f(x) \to \min, \qquad h(x) \leqslant 0, \tag{3.1}$$

где $f(x)$, $h(x)$ — обобщенно дифференцируемые функции. Определим многозначное отображение G: $x \to G(x)$:

$$G(x) = \begin{cases} G_f(x), & h(x) < 0, \\ \mathrm{co}\,\{G_f(x) \cup G_h(x)\}, & h(x) = 0, \\ G_h(x), & h(x) > 0. \end{cases} \tag{3.2}$$

Теорема 3.1. *Пусть x^* — точка минимума $f(x)$ при ограничении $h(x) \leqslant 0$. Тогда*

$$0 \in G(x) \tag{3.3}$$

(условие типа Джона). Если выполнено условие регулярности

$$\inf \{\| g \| \mid g \in G_h(x),\ h(x) = 0\} = \gamma > 0,$$

то существуют такие $\lambda^* \geqslant 0$, $g_f^* \in G_f(x^*)$, $g_h^* \in G_h(x^*)$, что

$$g_f^* + \lambda^* g_h^* = 0, \qquad \lambda^* h(x^*) = 0 \tag{3.4}$$

(условие типа Куна — Таккера).

Доказательство. Условие (3.4) очевидным образом следует из (3.3). Условие регулярности вводится для того, чтобы коэффициент при градиенте функции $f(x)$, удовлетворяющей (3.3), не обращался в нуль. Ситуация, когда он равен нулю, как правило, не интересна, так как в необходимых условиях экстремума не присутствует минимизируемая функция.

Пусть $h(x^*) = 0$ (случай $h(x^*) < 0$ доказывается аналогично). Предположим противное утверждению теоремы. Тогда в силу выпуклости и замкнутости $G(x^*)$

$$\rho(0, G(x^*)) \equiv \min_{g \in G(x^*)} \| g \| = \| d \| > 0,$$

причем $d \in G(x^*)$. В окрестности точки $x = x^*$ справедливы разложения

$$f(y) = f(x^*) + (g_f, y - x^*) + o_f(x^*, y, g_f),$$

$$h(y) = (g_h, y - x^*) + o_h(x^*, y, g_h),$$

где

$$o_f(x^*, y, g_f)/\| y - x^* \| \to 0,$$

$$o_h(x^*, y, g_h)/\| y - x^* \| \to 0$$

равномерно по $y \to x^*$, по $g_f \in G_f(y)$ и по $g_h \in G_h(y^*)$. Очевидно, что отображение G локально ограничено и замкнуто; поэтому оно полунепрерывно сверху. Рассмотрим $y = x^* - td$, $g \in G(y)$, $t \geqslant 0$. При достаточно малых t в силу свойств функций o_f и o_h и полунепрерывности сверху G имеем

$$o_f(x^*, y, g_f) \leqslant t \| d \|^2/3, \qquad o_h(x^*, y, g_h) \leqslant t \| d \|^2/3,$$

$$\rho(g, G(x^*)) = \min_{g' \in G(x^*)} \| g - g' \| \leqslant \| d \|/3,$$

$$(g, y - x^*) \leqslant -(2/3) \| d \|^2 t.$$

Тогда

$$f(x^* - td) \leqslant f(x^*) - t\|d\|^2/3, \qquad h(x^* - td) \leqslant -t\|d\|^2/3,$$

что противоречит локальной оптимальности x^*. Теорема доказана.

Рассмотрим теперь необходимые условия экстремума липшицевых функций.

Теорема 3.2. *Для того чтобы локально липшицева функция $f(x)$, $x \in E_n$, достигала в точке $x = x^*$ локального минимума, необходимо, чтобы*

$$0 \in \partial f(x^*).$$

Доказательство. Предположим, что $0 \notin \partial f(x^*)$. По теореме отделимости найдутся вектор a и число $\varepsilon > 0$, для которых $(a, g(x^*)) < -\varepsilon$, где $g(x^*)$ — произвольный вектор множества $\partial f(x^*)$. Применяя формулу Лагранжа, имеем

$$f(y, \alpha_k) = f(x^*, \alpha_k) + (\nabla f(x^* + \theta(y - x^*), \alpha_k), y - x^*),$$

$$0 \leqslant \theta \leqslant 1.$$

Рассмотрим точки $y = x^* + ta$, $t > 0$. Тогда в силу леммы 2.5, если t достаточно мало,

$$(a, \nabla f(x^* + \theta(y - x^*), \alpha_k)) \leqslant -\varepsilon/2, \qquad k \geqslant \bar{k}.$$

Следовательно,

$$f(y, \alpha_k) \leqslant f(x^*, \alpha_k) - \lambda\varepsilon/2.$$

Перейдя к пределу по $\alpha_k \to 0$ в этом неравенстве, получим

$$f(y) \leqslant f(x^*) - \lambda\varepsilon/2.$$

Это соотношение противоречит тому, что x^* — точка локального минимума функции $f(x)$.

Рассмотрим задачу

$$f_0(x) \to \min \tag{3.5}$$

при ограничениях

$$f_i(x) \leqslant 0, \qquad i = 1, \ldots, m, \tag{3.6}$$

где $f_\nu(x)$ ($\nu = 0, \ldots, m$) удовлетворяют локальному условию Липшица.

Теорема 3.3. *Если x^* — решение задачи* (3.5), (3.6), *то найдутся такие числа λ_0, $\lambda_i \geqslant 0$ $(i \in I(x^*))$, не все равные нулю, что*

$$\begin{gathered} \lambda_0 f_0^0(x^*; e) + \sum_{i \in I(x^*)} \lambda_i f_i^0(x^*; e) \geqslant 0 \qquad \forall e \in E_n, \\ I(x^*) = \{i \mid f_i(x^*) = 0\}. \end{gathered} \tag{3.7}$$

Доказательство. Предположим противное: существует такой вектор v, что для любых чисел λ_0, $\lambda_i \geqslant 0$, из которых не все равны нулю, выполняется неравенство

$$\lambda_0 f_0^0(x^*; v) + \sum_{i \in I(x^*)} \lambda_i f_i^0(x^*; v) < 0.$$

Отсюда следует, что

$$f_0^0(x^*; v) < 0, \quad f_i^0(x^*; v) < 0, \qquad i \in I(x^*).$$

Поэтому для достаточно малых $t > 0$

$$f_0(x^* + tv) < f_0(x^*), \quad f_i(x^* + tv) < f_i(x^*) = 0, \qquad i \in I(x^*).$$

Таким образом, при малых $t > 0$ точка $x^* + tv$ удовлетворяет всем ограничениям, так как в силу непрерывности $f_\nu(x)$ $(\nu = 0, \ldots, m)$ для достаточно малых t

$$f_i(x^* + tv) < 0, \qquad i \notin I(x^*),$$

а также

$$f_0(x^* + tv) < f_0(x^*).$$

Это противоречит тому, что x^* — решение задачи (3.5), (3.6). Теорема доказана.

Из соотношений (2.4) и (3.7) вытекает, что найдутся $\lambda_0, \lambda_i \geqslant 0$, не все равные нулю, для которых

$$\lambda_0 \max_{g^0 \in \partial f_0(x^*)} (g^0, e) + \sum_{i \in I(x^*)} \lambda_i \max_{g^i \in \partial f_i(x^*)} (g^i, e) \geqslant 0 \qquad \forall e \in E_n. \tag{3.8}$$

Лемма 3.1. *Если выполняется неравенство* (3.8), *то*

$$0 \in \lambda_0 \partial f_0(x^*) + \sum_{i \in I(x^*)} \lambda_i \partial f_i(x^*). \tag{3.9}$$

Доказательство. Рассмотрим выпуклое множество

$$P = \left\{\lambda_0 \partial f_0(x^*) + \sum_{i \in I(x^*)} \lambda_i \partial f_i(x^*)\right\}.$$

Пусть $0 \notin P$. Тогда по теореме отделимости найдутся такие вектор a и число $\varepsilon > 0$, что для любых

$$g^0 \in \partial f_0(x^*), \quad g^i \in \partial f_i(x^*), \quad i \in I(x^*),$$

выполняется неравенство

$$\lambda_0(a, g^0) + \sum_{i \in I(x^*)} \lambda_i(a, g^i) \leqslant -\varepsilon,$$

которое противоречит (3.9).

Теорема Куна — Таккера имеет место при выполнении дополнительного условия регулярности.

Теорема 3.4. *Пусть выполняется условие: существует такое* d, *что*

$$f_i^0(x^*; d) < 0 \qquad \forall i \in I(x^*). \tag{3.10}$$

Тогда существуют такие $\lambda_i \geqslant 0$ $(i = 1, \ldots, m)$, *что*

$$0 \in \partial f_0(x^*) + \sum_{i=1}^{m} \lambda_i \partial f_i(x^*),$$
$$\lambda_i f_i(x^*) = 0, \qquad i = 1, \ldots, m. \tag{3.11}$$

Доказательство очевидным образом следует из соотношения (3.7).

Обозначим через Λ множество векторов $\lambda = (\lambda_1, \ldots, \lambda_m) \geqslant 0$, удовлетворяющих соотношению (3.11). Из определения множеств $\partial f_0(x)$, $\partial f_i(x)$ $(i = 1, \ldots, m)$ и замкнутости отображений $x \to \partial f_\nu(x)$ $(\nu = 0, 1, \ldots$ $\ldots, m)$ вытекает, что Λ выпукло и замкнуто. Как будет показано ниже, (3.10) является необходимым и достаточным условием ограниченности Λ, т. е. в этом смысле является слабейшим из возможных условий регулярности.

Теорема 3.5. *Условие* (3.10) *эквивалентно ограниченности множества* Λ.

Доказательство. Пусть выполняется условие (3.10). Тогда из соотношения (3.11) и определения $\partial f_\nu(x)$ $(\nu = 0, 1, \ldots, m)$ вытекает, что

$$f_0^0(x^*; e) + \sum_{i=1}^{m} \lambda_i f_i^0(x^*; e) \geqslant 0 \qquad \forall e \in E_n,$$
$$\lambda_i \geqslant 0, \quad \lambda_i f_i(x^*) = 0. \qquad i = 1, \ldots, m,$$

откуда сразу следует ограниченность Λ.

Пусть теперь Λ непусто и ограничено. Образуем множество

$$M(x^*) = \operatorname{co} \bigcup_{i \in I(x^*)} \partial f_i(x^*)$$

и покажем, что $0 \notin M(x^*)$. В противном случае

$$0 = \sum_{i \in I(x^*)} \alpha_i h^i,$$
$$\alpha_i \geqslant 0, \quad h^i \in \partial f_i(x^*), \qquad \sum_{i \in I(x^*)} \alpha_i = 1. \tag{3.12}$$

Из соотношения (3.11) вытекает, что существуют такие $\lambda_i \geqslant 0$, $g^0 \in$ $\in \partial f_0(x^*)$, $g^i \in \partial f_i(x^*)$ $(i \in I(x^*))$, что

$$0 = g^0 + \sum_{i \in I(x^*)} \lambda_i g^i. \tag{3.13}$$

Умножив первое равенство в (3.12) на $\sigma > 0$ и сложив с (3.13), получим

$$0 = g^0 + \sum_{i \in I(x^*)} (\lambda_i g^i + \sigma \alpha_i h^i). \tag{3.14}$$

Положим

$$\lambda_i(\sigma) = \lambda_i + \sigma\alpha_i \geqslant 0$$

для $i \in I(x^*)$ и $\lambda_i(\sigma) = 0$ для $i \notin I(x^*)$. Если $\lambda_i(\sigma) > 0$, то

$$P^i(\sigma) = [\lambda_i(\sigma)]^{-1}(\lambda_i g^i + \sigma\alpha_i h^i) \in \partial f_i(x^*).$$

Поэтому из (3.14) имеем

$$0 = g^0 + \sum_{i=1}^{m} \lambda_i(\sigma) P^i(\sigma).$$

Так как найдется $\alpha_l > 0$ для $l \in I(x^*)$, то

$$\lambda_l(\sigma) = \lambda_l + \sigma\alpha_l \to \infty \quad \text{при} \quad \sigma \to \infty,$$

что противоречит ограниченности Λ.

Поскольку $0 \notin M(x^*)$, то по теореме отделимости найдутся такие вектор d и число $\varepsilon > 0$, что для $i \in I(x^*)$ имеем $(d, g^i) \leqslant -\varepsilon$ для всех $g^i \in \partial f_i(x^*)$. Следовательно, $f_i^0(x^*; d) < 0$ $(i \in I(x^*))$. Теорема доказана.

ГЛАВА 2

МИНИМИЗАЦИЯ ЛИПШИЦЕВЫХ ФУНКЦИЙ БЕЗ ВЫЧИСЛЕНИЯ ГРАДИЕНТОВ

§ 4. Конечно-разностный метод минимизации липшицевых функций

В случае, когда требуется минимизировать негладкую функцию, движение по обобщенному градиенту или в направлении разностной аппроксимации градиента не дает, вообще говоря, монотонного изменения ни самой функции, ни расстояния до точки экстремума. В связи с этим возникает интересный вопрос исследования сходимости методов, аналогичных конечно-разностным или градиентным, для минимизации невыпуклых негладких функций.

С практической точки зрения ситуация, когда информация о градиентах недоступна, наиболее типична в задачах математического программирования. При минимизации функций градиентными методами пользователю нужно подготовить n подпрограмм вычисления компонент градиента. Если размерность пространства велика, то эта вспомогательная процедура может занять значительное время. Кроме того, в задачах большой размерности весьма непросто составить формулы обобщенных градиентов. В гладкой оптимизации, если функция аналитически сложна, составляющие градиента аппроксимируются разностными соотношениями; например,

$$\nabla f(x) \approx \sum_{i=1}^{n} \frac{f(x+\alpha e_i) - f(x)}{\alpha} e_i. \qquad (4.1)$$

Для негладких функций разностная аппроксимация (4.1) неприменима, поскольку она приводит к несходящимся методам. В этом легко убедиться, рассмотрев следующий пример. Пусть каждая линия уровня выпуклой функции — граница квадрата. При вычислении (4.1) в левом нижнем углу квадрата получим нулевое значение, хотя все субградиенты в этой точке могут быть отличны от нуля. Вместе с тем небольшое изменение схемы (4.1) приводит к тому, что ее можно использовать в процессе поиска экстремума негладких функций.

Что касается градиентного метода

$$x^{k+1} = x^k - \rho_k g(x^k), \quad g(x^k) \in \partial f(x^k), \qquad (4.2)$$

то применение его для минимизации липшицевых функций также не гарантирует сходимости последовательности $\{x^k\}$ к стационарным точкам $f(x)$. Для функций этого класса направление $-g(x^k)$ не приводит к убыванию ни значений $f(x^k)$, ни расстояния до точки минимума. Поэтому вопрос выбора направления спуска в методах типа (4.2) решается иначе, чем для гладких и выпуклых функций.

Обоснование численных методов минимизации локально липшицевых функций проводится следующим образом. Рассмотрим последовательность сглаженных функций

$$f(x, \alpha) = \frac{1}{(2\alpha)^n} \int\limits_{x_1-\alpha}^{x_1+\alpha} \dots \int\limits_{x_n-\alpha}^{x_n+\alpha} f(y_1, \dots, y_n)\, dy_1 \dots dy_n.$$

Как отмечалось в § 2, функции $f(x, \alpha)$ равномерно в любой ограниченной области из E_n сходятся к $f(x)$ при $\alpha \to 0$. Возникает вопрос: можно ли в процессе минимизации использовать в качестве вектора спуска градиент функции $f(x, \alpha)$ и как при этом изменять α, устремляя его к нулю, чтобы в пределе получить решение исходной задачи?

Переход к усредненным функциям $f(x, \alpha)$ важен еще тем, что позволяет исключить влияние мелких локальных особенностей функции $f(x)$ и часто полезен для отыскания глобального экстремума при решении многоэкстремальных задач. Может оказаться, что функция $f(x)$ имеет множество локальных экстремумов, хотя есть и ярко выраженный глобальный экстремум. Минимизация функции $f(x)$ может привести к одному из локальных экстремумов, минимизация же сглаженной функции — дать хорошее приближение к глобальному минимуму, если в процессе вычислений изменять параметр α, устремляя его к нулю (см. § 12).

В § 2 было показано, что частная производная $\partial f(x, \alpha)/\partial x_1$ функции $f(x, \alpha)$ вычисляется по формуле

$$\frac{\partial f(x, \alpha)}{\partial x_1} = \frac{1}{(2\alpha)^n} \int\limits_{x_2-\alpha}^{x_2+\alpha} \dots \int\limits_{x_n-\alpha}^{x_n+\alpha} [f(x_1 + \alpha, y_2, \dots, y_n) - f(x_1 - \alpha, y_2, \dots, y_n)]\, dy_2 \dots dy_n. \tag{4.3}$$

Аналогично вычисляются $\partial f(x, \alpha)/\partial x_2, \dots, \partial f(x, \alpha)/\partial x_n$.

Градиент функции $f(x, \alpha)$ можно выразить иначе:

$$\nabla f(x, \alpha) = \frac{1}{(2\alpha)^n} \int\limits_{-\alpha}^{\alpha} \dots \int\limits_{-\alpha}^{\alpha} g(x + y)\, dy_1 \dots dy_n, \tag{4.4}$$

$$g(x + y) \in \partial f(x + y).$$

Сложность использования $\nabla f(x, \alpha)$ в процессе минимизации $f(x)$ связана с трудностью вычисления многомерных интегралов в формулах (4.3), (4.4). Однако эту задачу можно решить вероятностными методами.

Построим такой случайный вектор $H(x, \alpha)$, что

$$MH(x, \alpha) = \nabla f(x, \alpha), \tag{4.5}$$

где M — знак математического ожидания. Легко заметить, что случайный вектор, удовлетворяющий (4.5), определяется по формуле

$$H(x, \alpha) = \frac{1}{2\alpha} \sum_{i=1}^{n} [f(\tilde{x}_1, \ldots, x_i + \alpha, \ldots, \tilde{x}_n) - f(\tilde{x}_1, \ldots, x_i - \alpha, \ldots, \tilde{x}_n)]\, e_i, \tag{4.6}$$

где $\tilde{x}_i\ (i = 1, \ldots, n)$ — независимые случайные величины, равномерно распределенные на отрезках $[x_i - \alpha, x_i + \alpha]$ $(\alpha > 0)$.

Взяв конечно-разностную аппроксимацию градиента сглаженной функции $f(x, \alpha)$, получим

$$H(x, \alpha) = \sum_{i=1}^{n} \frac{f(\tilde{x} + \Delta e_i) - f(\tilde{x})}{\Delta}\, e_i. \tag{4.7}$$

Соотношение (4.7) отличается от (4.1) тем, что оно вычисляется в случайной точке $\tilde{x}$, выбранной (на основе равномерного распределения) в n-мерном кубе с центром в x и стороной 2α, причем

$$\mathsf{M} \sum_{i=1}^{n} \frac{f(\tilde{x} + \Delta e_i) - f(\tilde{x})}{\Delta}\, e_i = \nabla f(x, \alpha) + b,$$

где $\| b \| \leqslant \sqrt{n} L \Delta / \alpha$.

Если применить формулу (4.4), то вектор, удовлетворяющий (4.5), есть

$$H(x, \alpha) = g(\tilde{x}). \tag{4.8}$$

Формулы (4.6)—(4.8) приводят нас к рассмотрению конечно-разностных и градиентных методов минимизации локально липшицевых функций

$$x^{k+1} = x^k - \rho_k H(x^k, \alpha_k), \tag{4.9}$$

где вектор $H(x^k, \alpha_k)$ определяется по одной из формул (4.6) — (4.8).

Чтобы не вводить нормирующих множителей, полагаем, что множество стационарных точек $X^* = \{x^* \mid 0 \in \partial f(x^*)\}$, удовлетворяющих необходимому условию экстремума, ограничено. Тогда без ограничения общности можно сказать, что последовательность x^k, определяемая по формулам (4.6)—(4.9), принадлежит ограниченному множеству.

Нетрудно заметить, что последовательность точек x^k $(k = 0, 1, \ldots)$ имеет вероятностную природу. Как принято в теории вероятностей, случайные векторы $x^k(\omega)$ $(k = 0, 1, \ldots)$ определены на вероятностном пространстве $(\Omega, \Sigma, \mathsf{P})$ (являющемся произведением пространств, в которых выбрасываются точки $\tilde{x}$); ω обычно называют элементарным событием этого пространства. В дальнейшем зависимость x^k от ω, как правило, опускается.

Исследование сходимости $\{x^k\}$ к множеству $X^* \subset E_n$ проводится на основе следующих достаточных условий [91], незначительно отличающихся от более общих условий, описанных в § 8. Оказывается, что предельные точки последовательности $\{x^k\}$ с вероятностью 1 (почти наверное) принадлежат некоторому множеству X^*, если с вероятностью 1 выполнены следующие условия:

I. $$\lim_{k\to\infty} \| x^{k+1}(\omega) - x^k(\omega) \| = 0. \tag{4.10}$$

II. Существует такое ограниченное множество $X(\omega) \in E_n$, что

$$x^k(\omega) \in X(\omega), \qquad k = 0, 1, \ldots \tag{4.11}$$

III. Для любой подпоследовательности $x^s(\omega)$, для которой

$$\lim_{s\to\infty} x^s(\omega) = x'(\omega) \notin X^*, \quad s \in S \subset \{1, 2, \ldots\},$$

существует такое $\bar{\delta}(\omega)$, что для всех достаточно больших s и всех $\delta \in (0, \bar{\delta}(\omega)]$

$$k(s) = \min\{r \mid \| x^r - x^s \| > \varepsilon, r > s\} < \infty. \tag{4.12}$$

IV. Существует непрерывная функция $W(x)$, для которой

$$\overline{\lim_{s\to\infty}} W(x^{k(s)}(\omega)) < \lim_{s\to\infty} W(x^s(\omega)) = W(x'(\omega)). \tag{4.13}$$

V. Множество

$$W^* = \{W(x) \mid x \in X^*\}$$

не содержит интервалов.

Наиболее важная особенность условий I—V состоит в том, что они не требуют монотонного поведения всей последовательности $\{W(x^k)\}$. Условия III, IV означают, что определенная монотонность наблюдается только на некоторых специальным образом построенных подпоследовательностях.

Сформулируем теперь условия на величину шага ρ_k и выбор параметров Δ_k, α_k в формулах (4.6) — (4.8), при которых последовательность $\{x^k\}$ сходится в указанном смысле к множеству точек, удовлетворяющих необходимому условию экстремума.

Т е о р е м а 4.1. *Пусть выполняются условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < \infty,$$

$$\rho_k/\alpha_k \to 0, \quad \Delta_k/\alpha_k \to 0, \quad |\alpha_{k+1} - \alpha_k|/\rho_k \to 0, \quad \alpha_k \to 0.$$

Тогда все предельные точки последовательностей $\{x^k\}$, *определяемых по формулам* (4.6)—(4.9), *с вероятностью* 1 *принадлежат множеству* $X^* = \{x^* \mid 0 \in \partial f(x^*)\}$.

Д о к а з а т е л ь с т в о. Соотношение (4.10) непосредственно следует из условий теоремы и ограниченности $H(x^k, \alpha_k)$. Основную трудность представляет вывод соотношений (4.11), (4.12). Для простоты обозначений доказательство проводим для случая (4.5).

Для сглаженных функций $f(x, \alpha_k)$ справедлива формула конечных приращений Лагранжа

$$f(x^{k+1}, \alpha_k) = f(x^k, \alpha_k) + (\nabla f(x^k + \tau(x^{k+1} - x^k), \alpha_k), x^{k+1} - x^k),$$
$$0 \leqslant \tau \leqslant 1.$$

Ограниченные величины будем обозначать буквой C. Из леммы 2.4 следует

$$\begin{aligned} f(x^{k+1}, \alpha_k) &= f(x^k, \alpha_k) + \\ &+ (\nabla f(x^k + \tau(x^{k+1} - x^k), \alpha_k) - \nabla f(x^k, \alpha_k) + \nabla f(x^k, \alpha_k), x^{k+1} - x^k) \leqslant \\ &\leqslant f(x^k, \alpha_k) + C\rho_k^2/\alpha_k - \rho_k(\nabla f(x^k, \alpha_k), H(x^k, \alpha_k) - \nabla f(x^k, \alpha_k) + \\ &+ \nabla f(x^k, \alpha_k)) \leqslant f(x^k, \alpha_k) - \rho_k \| \nabla f(x^k, \alpha_k)\|^2 + C\rho_k^2/\alpha_k + \\ &+ \rho_k(\nabla f(x^k, \alpha_k) - H(x^k, \alpha_k), \nabla f(x^k, \alpha_k)). \end{aligned} \tag{4.14}$$

Оценим величину $|f(x, \alpha_{k+1}) - f(x, \alpha_k)|$. Сделав замену переменных, получим

$$f(x, \alpha_k) = \frac{1}{2^n} \int_{-1}^{1} \ldots \int_{-1}^{1} f(x + \alpha_k z)\, dz_1 \ldots dz_n,$$

$$f(x, \alpha_{k+1}) = \frac{1}{2^n} \int_{-1}^{1} \ldots \int_{-1}^{1} f(x + \alpha_{k+1} z)\, dz_1 \ldots dz_n.$$

Следовательно, в силу того, что $f(x)$ удовлетворяет условию Липшица, имеем

$$|f(x, \alpha_{k+1}) - f(x, \alpha_k)| \leqslant C |\alpha_k - \alpha_{k+1}|.$$

Из неравенства (4.14) вытекает

$$\begin{aligned} f(x^{k+1}, \alpha_{k+1}) &\leqslant f(x^k, \alpha_k) - \rho_k \| \nabla f(x^k, \alpha_k)\|^2 + \rho_k(\nabla f(x^k, \alpha_k) - \\ &- H(x^k, \alpha_k), \nabla f(x^k, \alpha_k)) + C|\alpha_k - \alpha_{k+1}| + C\rho_k^2/\alpha_k. \end{aligned} \tag{4.15}$$

Покажем, что выполняется условие (4.12). Пусть существует подпоследовательность $\{x^s(\omega)\} \to x' \notin X^*$. Тогда можно указать такие положительные числа $\bar{s}$ и $\bar{\delta}$, для которых $2\bar{\delta}$-окрестности точек x^s ($s \geqslant \bar{s}$, $s \in S$) не пересекаются с X^*.

Построим в точке x^s ($s \geqslant \bar{s}$) окрестность радиуса $\delta \leqslant \bar{\delta}/2$. Предположим, что условие (4.12) не выполняется. Тогда все точки x^k, начиная с некоторого $k \geqslant \bar{s}$, содержатся в δ-окрестности точки x^s ($s \in S$). Поэтому из леммы 2.5 следует, что для точек x^k, находящихся в довольно малой δ-окрестности точки x^s ($s \geqslant \bar{s}$), для достаточно больших номеров $\bar{s}$ выполняются неравенства

$$\| \nabla f(x^k, \alpha_k) \| \geqslant \sigma > 0.$$

Согласно условиям теоремы, начиная с некоторого номера $k \geqslant \bar{s}$ имеем

$$\frac{C\,|\,\alpha_k - \alpha_{k+1}|}{\rho_k} + \frac{C\rho_k}{\alpha_k} \leqslant \frac{\sigma^2}{2}.$$

Просуммировав неравенства (4.15), получим

$$f(x^{k+1}, \alpha_{k+1}) \leqslant$$

$$\leqslant f(x^s, \alpha_s) - \frac{\sigma^2}{2}\sum_{r=s}^{k} \rho_r - \sum_{r=s}^{k} \rho_r \left[\frac{\sigma^2}{2} - \frac{C\,|\,\alpha_r - \alpha_{r+1}|}{\rho_r} - \frac{C\rho_r}{\alpha_r}\right] +$$

$$+ \sum_{r=s}^{k} \rho_r (\nabla f(x^r, \alpha_r) - H(x^r, \alpha_r), \nabla f(x^r, \alpha_r)) \leqslant f(x^s, \alpha_s) - \frac{\sigma^2}{2}\sum_{r=s}^{k} \rho_r +$$

$$+ \sum_{r=s}^{k} \rho_r (\nabla f(x^r, \alpha_r) - H(x^r, \alpha_r), \nabla f(x^r, \alpha_r)). \quad (4.16)$$

Воспользуемся следующим известным фактом [9].

Лемма 4.1. *Если $\{\varphi_k\}$ — такая последовательность случайных величин, что*

$$\sum_{k=0}^{\infty} M\varphi_k^2 < \infty,$$

то последовательность случайных величин

$$\sum_{k=0}^{\infty} \varphi_k - M(\varphi_k \mid \varphi_0, \varphi_1, \ldots, \varphi_{k-1})$$

сходится с вероятностью 1.

В данном случае

$$\varphi_k = -\rho_k (H(x^k, \alpha_k), \nabla f(x^k, \alpha_k)),$$

$$M(\varphi_k \mid \varphi_0, \varphi_1, \ldots, \varphi_{k-1}) = -\rho_k \,\|\nabla f(x^k, \alpha_k)\|^2.$$

Следовательно, с вероятностью 1 сходится ряд

$$\sum_{k=0}^{\infty} \rho_k (\nabla f(x^k, \alpha_k) - H(x^k, \alpha_k), \nabla f(x^k, \alpha_k)).$$

По условию теоремы ряд $\sum_{k=0}^{\infty} \rho_k$ расходится, поэтому, переходя к пределу по $k \to \infty$ в неравенстве (4.16), получим противоречие с ограниченностью $\{f(x^k)\}$. Это означает, что справедливо соотношение (4.12): найдется такой номер $k(s) < \infty$, что $k(s) = \min\{r \mid \|x^r - x^s\| > \delta, r > s\}$.

Осталось вывести условие (4.13). Из соотношений

$$x^{k(s)} = x^s - \sum_{r=s}^{k(s)-1} \rho_r H(x^r, \alpha_r), \quad \|x^{k(s)} - x^s\| > \delta$$

получаем

$$\sum_{r=s}^{k(s)-1} \rho_r > \frac{\delta}{C}.$$

Поэтому из неравенства (4.16) для достаточно больших s имеем $f(x^{k(s)}, \alpha_{k(s)}) \leqslant f(x^s, \alpha_s) - \sigma^2\delta/(4C)$, откуда в силу равномерной сходимости $f(x, \alpha) \to f(x)$ следует

$$\overline{\lim_{s\to\infty}} f(x^{k(s)}) \leqslant \lim_{s\to\infty} f(x^s) - \sigma^2\delta/(4C), \tag{4.17}$$

что и доказывает (4.13). Таким образом, в качестве функции $W(x)$ выступает $f(x)$.

Покажем теперь на основе выведенных условий (4.12), (4.13) сходимость последовательности $\{x^k\}$ к множеству X^*.

Для любых таких чисел a и b, что

$$\overline{\lim_{s\to\infty}} f(x^{k(s)}) < a < b < \lim_{s\to\infty} f(x^s) = f(x'),$$

последовательность $\{f(x^k)\}$ пересекает интервал (a, b) слева направо бесконечное число раз. Тогда можно выбрать две последовательности точек $\{x^r\}$ $(r \in R)$, $\{x^p\}$ $(p \in P)$, $R, P \subset \{1, 2, \ldots, k, \ldots\}$, для которых

$$f(x^r) \leqslant a, \quad f(x^{r+1}) > a,$$

$$f(x^p) \geqslant b, \quad f(x^k) > a, \quad r < k < p.$$

Для простоты считаем, что последовательность $\{x^r\}$ $(r \in R)$ сходится к x''; в противном случае через $\{x^r\}$ $(r \in R)$ обозначаем сходящуюся подпоследовательность. Поскольку $\rho_k \to 0$, то

$$\lim_{r\to\infty} f(x^r) = a.$$

В силу условия V число a можно выбрать так, чтобы

$$a \notin \{f(x) \mid x \in X^*\}.$$

Для последовательности $\{x^r\}$ $(r \in R)$ выполним те же процедуры, которые уже проделаны для x^s $(s \in S)$. Пусть

$$k(r) = \min\{t \mid \|x^t - x^r\| > \delta, t > r\}.$$

В силу непрерывности функции $f(x)$ число δ подберем столь малым, чтобы для всех достаточно больших $r \in R$ получить $r < k(r) < p$, т. е.

$$f(x^r) \leqslant a < f(x^{k(r)}),$$

$$\overline{\lim_{r\to\infty}} f(x^{k(r)}) \geqslant \lim_{r\to\infty} f(x^r),$$

что противоречит неравенству (4.17), записанному для $\{x^r\}$ $(r \in R)$. Теорема доказана.

Следствие 4.1. *С вероятностью* 1 *существует подпоследовательность* $\{x^s\}$ $(s \in S)$, *для которой* $\nabla f(x^s, \alpha_s) \to 0$, $s \to \infty$.

Таким образом, если $\{x^s\} \to x$, то в силу леммы 2.6 $x \in X^*$.

Доказательство. Предположим, что существуют такие ε и $\bar{k}$, для которых

$$\|\nabla f(x^k, \alpha_k)\| \geqslant \varepsilon, \qquad k \geqslant \bar{k}.$$

Тогда для достаточно больших k из неравенства (4.15) следует

$$f(x^{k+1}, \alpha_{k+1}) \leqslant$$
$$\leqslant f(x^k, \alpha_k) - \varepsilon^2\rho_k/2 + \rho_k(\nabla f(x^k, \alpha_k) - H(x^k, \alpha_k), \nabla f(x^k, \alpha_k)).$$

Просуммировав эти неравенства по k от 0 до N и устремив N к ∞, получим противоречие с ограниченностью $\{f(x^k)\}$, поскольку с вероятностью 1

$$\sum_{k=0}^{\infty} \rho_k(\nabla f(x^k, \alpha_k) - H(x^k, \alpha_k), \nabla f(x^k, \alpha_k)) < \infty.$$

Сформулируем важное свойство множества $\tilde{X}^*$ предельных точек последовательности $\{x^k\}$, определяемой по формулам (4.6) — (4.9).

Следствие 4.2. *Пусть x^1 и x^2 — точки из множества $\tilde{X}^*$. Тогда $f(x^1) = f(x^2)$, т. е. последовательность $\{f(x^k)\}$ с вероятностью 1 сходится.*

Замечание. Последовательность $\{x^k\}$ можно искусственно сделать ограниченной, если рассматривать итеративный процесс

$$x^{k+1} = \begin{cases} x^k - \rho_k H(x_k, \alpha_k), & f(x^k) \leqslant f(x^0) + C, \\ x^0, & f(x^k) > f(x^0) + C, \end{cases}$$

где x^0 — начальное приближение, $C > 0$ — некоторая постоянная. Подобные алгоритмы исследовались в [91]. Применяя условия (4.12), (4.13), легко показать, что с вероятностью 1 последовательность $\{x^k\}$ лишь конечное число раз выходит за пределы множества $\{x \mid f(x) \leqslant \leqslant f(x^0) + C\}$.

§ 5. Построение конечных разностей для функции максимума

С точки зрения практических приложений большой интерес для минимизации функции максимума представляют численные методы, не требующие вычислений градиентов. С одной стороны, этот факт объясняется тем, что в задачах с большим числом переменных пользователю весьма сложно вычислять обобщенные градиенты, которые требуются в методах градиентного типа. С другой стороны, в задаче минимизации функции

$$f(x) = \max_{1 \leqslant i \leqslant r} f_i(x)$$

ему необходимо подготовить $n \cdot r$ подпрограмм для вычислений обобщенных градиентов $f_i(x)$ $(i = 1, \ldots, r)$. Поэтому в задачах большой размерности эта процедура может занять значительное время.

Построение конечно-разностных методов на основе функции $f(x)$ неэффективно, поскольку для них требуется порядка $n \cdot r$ вычислений значений функций $f_i(x)$ $(i = 1, \ldots, r)$. Рассматриваемые ниже методы используют $n + r + 1$ (или $2n + r$) вычислений $f_i(x)$; они применимы и в том случае, когда $f_i(x)$ недифференцируемые.

Рассмотрим задачу:

$$f(x) = \max_{1 \leqslant i \leqslant r} f_i(x) \to \min \tag{5.1}$$

при условии

$$x \in X, \tag{5.2}$$

где $f_i(x)$ — выпуклые функции, X — ограниченное выпуклое и замкнутое множество в E_n. Область X такова, что операция проектирования на X осуществляется достаточно просто.

Конечно-разностные методы решения задачи (5.1), (5.2) определяются соотношениями

$$x^{k+1} = \pi_X(x^k - H^j(x^k, \alpha_k)), \tag{5.3}$$

где вектор $H^j(x^k, \alpha_k)$ выбирается одним из двух способов:

$$H^j(x^k, \alpha_k) = \frac{1}{2\alpha_k} \sum_{i=1}^{n} [f_j(\tilde{x}_1^k, \ldots, x_i^k + \alpha_k, \ldots, \tilde{x}_n^k) - \\ - f_j(\tilde{x}_1^k, \ldots, x_i^k - \alpha_k, \ldots, \tilde{x}_n^k)]\, e_i, \tag{5.4}$$

$$H^j(x^k, \alpha_k) = \sum_{i=1}^{n} \frac{f_j(\tilde{x}^k + \Delta_k e_i) - f_j(\tilde{x}^k)}{\Delta_k}\, e_i; \tag{5.5}$$

$\pi_X(x)$ — оператор проектирования точки x на множество X; $\tilde{x}_i^k$ $(i = 1, \ldots, n)$ — независимые случайные величины, равномерно распределенные на отрезках $[x_i^k - \alpha_k, x_i^k + \alpha_k]$; индекс j таков, что

$$f_j(x^k) = f(x^k) = \max_{1 \leqslant i \leqslant r} f_i(x^k).$$

Таким образом, в методе (5.3)—(5.5), как и в алгоритмах субградиентного типа, конечная разность вычисляется по той функции, на которой достигается максимум.

Теорема 5.1. *Пусть выполняются условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < \infty, \quad \frac{\Delta_k}{\alpha_k} \to 0, \quad \alpha_k \to 0.$$

Тогда с вероятностью 1 *предельные точки последовательности* $\{x^k\}$ *принадлежат множеству решений задачи* (5.1), (5.2).

Доказательство. В § 4 показано, что случайные векторы $H^j(x^k, \alpha_k)$, определяющие направление спуска в методе (5.3), удовлетворяют условию

$$M H^j(x^k, \alpha_k) \mid x^k = \nabla f_j(x^k, \alpha_k) + b_k,$$

где $f_j(x, \alpha)$ — сглаженные функции, $\|b_k\| \leqslant C\Delta_k/\alpha_k$. Из неравенства

$$f_j(y, \alpha_k) - f_j(x, \alpha_k) \geqslant (\nabla f_j(x, \alpha_k), y - x)$$

вытекает, что $\nabla f_j(x, \alpha_k)$ равномерно сходится к $\partial f_j(x)$ при $\alpha_k \to 0$, где $\partial f_j(x)$ — субдифференциал функции $f_j(x)$. В силу свойств функции максимума для выпуклых функций справедливо соотношение

$$\partial f_j(x) \subset \partial f(x), \quad f_j(x) = f(x).$$

Пусть $x^* \in X^*$, X^* — множество решений задачи (5.1), (5.2). Применяя условия сходимости (4.10) — (4.13) с $W(x) = \min\limits_{x^* \in X^*} \|x - x^*\|$ утверждение теоремы получаем непосредственно из неравенств

$$\|x^{k+1} - x^*\|^2 \leqslant \|x^k - \rho_k H^j(x^k, \alpha_k) - x^*\|^2 =$$
$$= \|x^k - x^*\|^2 - 2\rho_k (\nabla f_j(x^k, \alpha_k) + b_k, x^k - x^*) +$$
$$+ 2\rho_k (\nabla f_j(x^k, \alpha_k) + b_k - H^j(x^k, \alpha^k), x^k - x^*) + \rho_k^2 \|H^j(x^k, \alpha_k)\|^2$$

в силу того, что с вероятностью 1

$$\sum_{k=0}^{\infty} \rho_k (\nabla f_j(x^k, \alpha_k) + b_k - H^j(x^k, \alpha_k), x^k - x^*) < \infty.$$

§ 6. Случайные конечно-разностные направления

Определение векторов $H(x^k, \alpha_k)$ в формулах (4.6), (4.7) требует $2n$ или $n+1$ вычислений значений целевой функции. Если одно вычисление функции $f(x)$ занимает значительное время, то разностные аппроксимации (4.6), (4.7) могут оказаться неприемлемыми. В таком случае целесообразно использовать идеи алгоритмов случайного поиска.

Рассмотрим разностную аппроксимацию

$$z(x^k) = \sum_{i=1}^{p} \frac{f(\tilde{x}^k + \Delta_k \mu_i^k) - f(\tilde{x}^k)}{\Delta_k} \mu_i^k. \tag{6.1}$$

Здесь $\tilde{x}^k$ — случайная точка, равномерно распределенная в n-мерном кубе с центром в x^k и с ребром $2\alpha_k$; компоненты вектора μ_i^k $(i = 1, \ldots, p)$ — независимые и равномерно распределенные на $[-1, 1]$ случайные величины.

По формуле Лагранжа имеем

$$(f(x^k + \Delta_k \mu_i^k, \alpha_k) - f(x^k, \alpha_k))/\Delta_k = (\nabla f(x^k + \tau_k \Delta_k \mu_i^k, \alpha_k), \mu_i^k) =$$
$$= (\nabla f(x^k, \alpha_k), \mu_i^k) + v^k, \qquad 0 \leqslant \tau_k \leqslant 1,$$

где $f(x, \alpha)$ — сглаженная функция, v^k — некоторый вектор такой, что

$$\|v^k\| \leqslant C\Delta_k/\alpha_k.$$

Так как $M(\mu_i^k)^2 = 2/3$, то условное математическое ожидание вектора $z(x^k)$ удовлетворяет соотношению

$$Mz(x^k)\,|\,x^k = \frac{2}{3}\,p\nabla f(x^k, \alpha_k) + b_k,$$

где

$$\|b_k\| \leqslant C\Delta_k/\alpha_k.$$

Следовательно, для минимизации локально липшицевых функций можно применить метод случайного поиска

$$x^{k+1} = x^k - \rho_k z(x^k), \tag{6.2}$$

в котором вектор $z(x^k)$ вычисляется по формуле (6.1).

Теорема 6.1. *Пусть выполняются условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \qquad \sum_{k=0}^{\infty} \rho_k^2 < \infty, \qquad \rho_k/\alpha_k \to 0,$$

$$\Delta_k/\alpha_k \to 0, \quad |\alpha_k - \alpha_{k+1}|/\rho_k \to 0, \quad \alpha_k \to 0.$$

Тогда все предельные точки последовательности $\{x^k\}$, *определенной по формулам* (6.1), (6.2), *с вероятностью* 1 *принадлежат множеству* $X^* = \{x^* \mid 0 \in \partial f(x^*)\}$; *последовательность* $\{f(x^k)\}$ *сходится почти наверное.*

Доказательство. Выведем неравенство, которое вместе с условиями (4.12), (4.13) непосредственно приводит к доказательству теоремы 6.1.

Для сглаженных функций $f(x, \alpha)$ имеем

$$f(x^{k+1}, \alpha_k) = f(x^k, \alpha_k) + (\nabla f(x^k + \tau_k(x^{k+1} - x^k), \alpha_k) - $$
$$- f(x^k, \alpha_k) + \nabla f(x^k, \alpha_k), x^{k+1} - x^k) \leqslant f(x^k, \alpha_k) + C\frac{\rho_k^2}{\alpha_k} -$$
$$- \rho_k(\nabla f(x^k, \alpha_k), z(x^k)), \qquad 0 \leqslant \tau_k \leqslant 1.$$

Тогда

$$f(x^{k+1}, \alpha_{k+1}) \leqslant$$
$$\leqslant f(x^k, \alpha_k) - \frac{2}{3}\,p\rho_k\|\nabla f(x^k, \alpha_k)\|^2 + C|\alpha_k - \alpha_{k+1}| + C\frac{\rho_k^2}{\alpha_k} +$$
$$+ C\rho_k\frac{\Delta_k}{\alpha_k} + \rho_k\left(\frac{2}{3}\,p\nabla f(x^k, \alpha_k) + b_k - z(x^k), \nabla f(x^k, \alpha_k)\right).$$

С учетом полученного неравенства доказательство теоремы 6.1 незначительно отличается от доказательства теоремы 4.1.

Метод минимизации выпуклой функции максимума

$$f(x) = \max_{1 \leqslant i \leqslant r} f_i(x), \qquad x \in X,$$

определяется последовательностью

$$x^{k+1} = \pi_X\left(x^k - \rho_k \sum_{i=1}^{p} \frac{f_j(\tilde{x}^k + \Delta_k \mu_i^k) - f_j(\tilde{x}^k)}{\Delta_k}\, \mu_i^k\right),$$

где

$$f_j(x^k) = f(x^k) = \max_{1\leqslant i\leqslant r} f_i(x^k).$$

§ 7. Эффективность конечно-разностных методов

В последнее время вопросы эффективности численных методов оптимизации приобретают все большее значение. В силу значительного расширения приложений и усложнения задач оптимизации вытекает необходимость иметь в арсенале вычислительных средств в некотором смысле оптимальные методы, которые обеспечивали бы решение всех задач определенного класса при минимально возможной трудоемкости.

В работе [73] была построена теория сложности экстремальных задач выпуклого программирования.

Рассмотрим задачу

$$\begin{gathered} f_0(x) \to \min, \\ f_i(x) \leqslant 0, \qquad i = 1, \dots, m, \\ x \in X, \end{gathered} \tag{7.1}$$

где $f_j(x)$ $(j = 0, 1, \dots, m)$ — выпуклые функции, X — ограниченное выпуклое замкнутое множество некоторого конечномерного банахова пространства. Задача (7.1) снабжена оракулом, который вычисляет значения функций $f_j(x)$ $(j = 0, 1, \dots, m)$ (и значения их градиентов) в указанной ему точке.

Численные методы дают приближенное решение задачи (7.1). Это решение может быть точкой множества X либо указанием того, что задача (7.1) несовместна (последний факт обозначается символом $*$). В качестве погрешности решения принимается относительная погрешность точки $x \in X$:

$$v(x) = \begin{cases} 1, & \text{задача (7.1) совместна, } x = *, \\ 0, & \text{задача (7.1) несовместна, } x = *, \\ \max\left\{\dfrac{f_0(x) - \bar{f}_0}{r_0}, \dfrac{f_1(x)}{r_1}, \dots, \dfrac{f_m(x)}{r_m}\right\}, & x \neq *, \end{cases}$$

где $\bar{f}_0$ — оптимальное значение $f_0(x)$. Здесь $r_j > 0$ $(0 \leqslant j \leqslant m)$ — нормирующие множители, которые выбираются, как правило, так, что относительные погрешности любой точки $x \in X$ не превышают 1.

Метод решения задачи (7.1) есть набор правил формирования очередных вопросов к оракулу, момента остановки и выдачи результата. Трудоемкость решения данной задачи данным методом определяется как число шагов его работы на ней. Помимо трудоемкости метод характеризуется своей погрешностью.

Сложность $N(\nu)$ данного класса задач в функции от ν равна минимально возможной трудоемкости метода, который решает любую задачу класса с погрешностью, не превосходящей ν.

В работе [73] показано, что сложность выпуклых задач допускает оценку

$$C_1 \leqslant \frac{N(\nu)}{1 + n \ln(1/\nu)} \leqslant C_2,$$

где C_1, C_2 — положительные константы. Правое неравенство верно при всех $\nu < 1$, а левое — в асимптотике по $\nu \to 0$.

В работе [56] был предложен метод центра тяжести (МЦТ), который имеет трудоемкость $O(n \ln(1/\nu))$ и является оптимальным методом при $\nu \to 0$. Этот метод основан на проведении опорных гиперплоскостей через центр тяжести области, где локализован минимум. Процедура уменьшения области локализации путем последовательных отсечений является очень эффективной, однако практически этот метод не может быть использован из-за чрезмерной сложности шага.

В работах [141, 135] были предложены модифицированные методы центра тяжести (иначе их еще называют методами эллипсоидов), в которых начальная область локализации минимума — полушар — погружается в эллипсоид минимального объема, найти центр тяжести которого весьма просто. Затем эллипсоид путем изменения метрики превращают в шар и операцию повторяют. Метод эллипсоидов по-другому можно интерпретировать как метод растяжения пространства в направлении субградиента. Этот метод уже практически реализуем; его трудоемкость равна $O(n^2 \ln(1/\nu))$ с числом $O(n^2)$ арифметических операций на каждом шаге и памятью $O(n^2)$. Таким образом, трудоемкости указанных методов не связаны ни с наличием (и числом) ограничений задачи, ни с ее гладкостью и степенью обусловленности.

Оказывается, что при росте размерности задачи n сложность $N(\nu)$ зависит от геометрических свойств X. Если X — параллелепипед, то $N(\nu) \sim n \ln(1/\nu)$ при всех $\nu < 1/4$ и всех n, так что сложность линейно растет с ростом n. Если же X — эллипсоид, то при увеличении n функция $N(\nu)$ ограничена сверху величиной $O(\nu^{-2})$ независимо от размерности и при данном ν есть $O(\nu^{-2})$ для всех достаточно больших n.

В работе [73] построены градиентные методы, которые были названы методами зеркального спуска для решения выпуклых задач (7.1) на выпуклых телах X типа L_p-шаров (в метрике пространства L_p, $1 \leqslant p < \infty$). Оценка трудоемкости соответствующего метода имеет вид

$$C(p)\,(1/\nu)^{\max(2,p)} \quad \text{при} \quad p > 1 \quad \text{для} \quad p = 1 - C(1)\,(1/\nu)^2 \ln n.$$

Следовательно, выпуклые задачи на эллипсоидах можно решать методами, трудоемкость которых не зависит, вообще говоря, от размерности задачи. Кроме того, в задачах большой размерности трудоемкость этих методов неулучшаема.

Если исходное пространство евклидово, то метод зеркального спуска превращается в обычный субградиентный метод со специальной регулировкой шага. Если в текущей точке x^k ($k=0,1,\ldots$) с точностью до величин порядка ν выполняются ограничения $f_i(x) \leqslant 0$ $(i = 1, \ldots \ldots, m)$, то спуск в точке x^k происходит по направлению вектора $-g_0(x^k)$, $g_0(x^k) \in \partial f_0(x^k)$, в противном случае — по антисубградиенту функции ограничения, которое не удовлетворяется в точке x^k. Методы зеркального спуска устойчивы по отношению к ошибкам в исходной информации.

В пространствах, отличных от евклидовых, градиентный спуск происходит в сопряженном пространстве, после чего производится отображение в исходное пространство. Например, если исходное пространство есть $L_p^{(n)}$ с нормой $\|x\|_p = \left(\sum\limits_{i=1}^{n} |x_i|^p\right)^{1/p}$, то основное движение происходит в сопряженном к $L_p^{(n)}$ пространстве $L_q^{(n)}$, $1/p + 1/q = 1$.

Методы зеркального спуска были распространены на задачи стохастического программирования, в которых функции цели и ограничений не заданы аналитически и имеют вероятностный характер. Источник информации о решаемой задаче (оракул) сообщает в нужный момент времени значения функций и их субградиенты в рассматриваемой точке. Оракул может ошибаться, но так, чтобы ошибка не превосходила заданной точности решения, а также его ответы могут искажаться случайными помехами.

К задаче предъявляются следующие требования А — В, которые характеризуют случайные помехи. Пусть $R > 0$ и ν_0 — точность оракула.

А. При всех $x \in X$, $0 \leqslant j \leqslant m$ функционалы

$$h_j(y) = M_\theta \Psi_j(x, \theta) + (M_\theta \xi_j(x, \theta), y - x) - \nu_0 R$$

удовлетворяют условиям

$$f_j(y) \geqslant h_j(y), \qquad y \in X, \qquad f_j(x) \leqslant h_j(x) + \nu_0 R.$$

Скалярные функции $\Psi(x, \theta)$ интерпретируются как наблюдения величин $f_j(x)$ в точке x при шуме оракула θ, $\xi_j(x, \theta)$ интерпретируются как наблюдения субградиентов функций $f_j(x)$.

Б. При всех $j = 0, \ldots, m$ и $x \in X$

$$M_\theta \| \xi_j(x, \theta) \|^r \leqslant (R/(2\rho(X)))^r, \qquad r > 1,$$

где $\rho(X)$ — радиус множества X.

В. При всех $x \in X$

$$M_\theta | \Psi_j(x, \theta) - f_j(x) |^r \leqslant R^r.$$

В качестве нормирующих множителей выбирают

$$r_j = \begin{cases} R, & j = 0, \\ R + \max[0, \min\limits_{x \in X} f(x)], & j \geqslant 1. \end{cases}$$

Требование А означает, что при вычислениях функций и их субградиентов допускается определенная систематическая ошибка. Требования Б и В ограничивают интенсивность помех оракула.

В евклидовом случае стохастический метод зеркального спуска для задачи $\min f(x)$, $x \in X$ превращается в известный метод

$$x^{k+1} = \pi_X (x^k - \rho_k \xi (x^k, \theta^k)) \tag{7.2}$$

с усреднением

$$\bar{x}^k = \left(\sum_{s=1}^{k} \rho_s\right)^{-1} \sum_{s=1}^{k} \rho_s x^s.$$

При условиях $\sum_{k=0}^{\infty} \rho_k = \infty$, $\rho_k \to 0$ траектория $\bar{x}^k$ в среднем по функционалу сходится к оптимуму и дается оценка скорости сходимости. Если выбирать шаги постоянными по формуле $\rho_k = 4(\nu - \nu_0)/R$, то число итераций, обеспечивающих точность $\nu > \nu_0$, равно $N =]1/(4(\nu - \nu_0)^2)[$ ($]t[$ — минимальное целое, не меньшее t).

Трудоемкость указанных стохастических методов при достижении заданной средней погрешности такая же, как и у методов зеркального спуска для детерминированных задач. Иными словами, градиентные методы недифференцируемой оптимизации почти не реагируют на шумы в поставляемой информации.

Наша цель состоит в том, чтобы показать, что стохастические конечные разности, рассмотренные в § 4—6, удовлетворяют А—В. Рассмотрим сглаженную функцию

$$f(x, \alpha) = \frac{1}{(2\alpha)^n} \int_{-\alpha}^{\alpha} \dots \int_{-\alpha}^{\alpha} f(x + y)\, dy_1 \dots dy_n.$$

Пусть функция $f(x)$ липшицева с константой L. Исследование проводится в пространстве E_n. Тогда

$$| f(x, \alpha) - f(x) | \leqslant \sqrt{n} L \alpha. \tag{7.3}$$

В § 4 показано, что для вектора $H(x, \alpha)$, вычисляемого по формуле

$$H(x, \alpha) = \\ = \frac{1}{2\alpha} \sum_{i=1}^{n} [f(\tilde{x}_1, \dots, x_i + \alpha, \dots, \tilde{x}_n) - f(\tilde{x}_1, \dots, x_i - \alpha, \dots, \tilde{x}_n)]\, e_i, \tag{7.4}$$

выполняется условие $\mathrm{M} H(x, \alpha) = \nabla f(x, \alpha)$, где $\tilde{x}_i$ $(i = 1, \dots, n)$ — независимые случайные величины, равномерно распределенные на отрезках $[x_i - \alpha,\ x_i + \alpha]$.

Из формулы (7.4) следует, что

$$\mathrm{M} \| H(x, \alpha) \|^2 \leqslant (\sqrt{n} L)^2. \tag{7.5}$$

Параметр сглаживания α в силу (7.3) и неравенства для выпуклых функций

$$f(y, \alpha) - f(x, \alpha) \geqslant (\nabla f(x, \alpha), y - x)$$

можно подобрать так, чтобы выполнялись требования A — B. Таким образом, с точностью до случайных помех, удовлетворяющих требованиям А—В, субградиент моделируется стохастической конечной разностью (7.4). Следовательно, в методе (7.2) вместо $\xi(x^k, \theta^k)$ можно использовать вектор конечных разностей $H(x, \alpha)$ (7.4):

$$x^{k+1} = \pi_X(x^k - \rho_k H(x^k, \alpha)). \tag{7.6}$$

Из неравенства (7.5) вытекает, что трудоемкость метода (7.6), настроенного на абсолютную погрешность $\nu > \nu_0$, равна $O(\sqrt{n}L/(\nu - \nu_0))^2$; при этом предполагается, что сложности вычисления субградиента функции $f(x)$ и вектора $H(x, \alpha)$ — величины одного порядка. Взяв конечно-разностную аппроксимацию градиента, получим случайный вектор

$$H(x, \alpha) = \sum_{i=1}^{n} \frac{f(\tilde{x} + \Delta e_i) - f(\tilde{x})}{\Delta} e_i. \tag{7.7}$$

Условное математическое ожидание вектора (7.7) совпадает с $\nabla f(x, \alpha)$ с точностью до помехи величиной $\sqrt{n}L\Delta/\alpha$, т. е. опять же параметры Δ и α можно подобрать такими, чтобы выполнялись условия А — В.

Если одно вычисление функции занимает значительное время, то используется конечно-разностная аппроксимация, рассмотренная в § 6:

$$H(x, \alpha) = \sum_{i=1}^{p} \frac{f(\tilde{x} + \Delta\mu_i) - f(\tilde{x})}{\Delta} \mu_i. \tag{7.8}$$

Здесь $\tilde{x}_i$ $(i = 1, \ldots, n)$ равномерно распределены на отрезках $[x_i - \alpha, x_i + \alpha]$, компоненты вектора μ_i — независимые и равномерно распределенные на $[-1, 1]$ случайные величины.

Из формулы (7.8) следует, что

$$\mathsf{M}\,\|H(x, \alpha)\|^2 \leqslant (\sqrt{n}pL)^2,$$

причем

$$\mathsf{M}H(x, \alpha) = \frac{2}{3} p\nabla f(x, \alpha) + b,$$

где $\|b\| \leqslant pL\Delta n^{3/2}/\alpha$. Поэтому в качестве направления спуска в методе (7.6) можно использовать вектор конечных разностей (7.8), однако относительно смещений Δ здесь предъявляются более жесткие требования.

Большой практический интерес представляет построение конечно-разностных методов минимизации функции максимума

$$f(x) = \max_{1 \leqslant s \leqslant r} f_s(x), \qquad x \in X, \tag{7.9}$$

поскольку операция максимума является одним из основных источников, порождающих негладкие функции.

В § 5 показано, что конечно-разностный метод минимизации функции (7.9) строится аналогично субградиентному методу:

$$x^{k+1} = \pi_X (x^k - \rho_k H_j (x^k, \alpha)),$$

где векторы $H_j (x^k, \alpha)$, определяемые по формулам (7.4), (7.7), (7.8), вычисляются по той функции, на которой достигается максимум в формуле (7.9). Таким образом, во многих важных случаях вычисление субградиента практически можно заменить вычислением конечных разностей (7.4), (7.7), (7.8).

Один из основных результатов выпуклого анализа заключается в том, что решение задачи (7.1) эквивалентно минимизации негладкой штрафной функции

$$\Phi (x) = f_0 (x) + \sum_{i=1}^{m} r_i \max (0, f_i (x)), \qquad x \in X, \tag{7.10}$$

где $r_i > 0$ — штрафные коэффициенты, $\Phi (x)$ — функция максимума специального вида; поэтому для ее минимизации используются конечно-разностные методы:

$$x^{k+1} = \pi_X \left(x^k - \rho_k \left(H_0 (x^k, \alpha) + \sum_{i=1}^{m} r_i^+ H_i (x^k, \alpha)\right)\right), \tag{7.11}$$

$$r_i^+ = \begin{cases} r_i, & f_i (x^k) > 0, \\ 0, & f_i (x^k) \leqslant 0; \end{cases}$$

векторы $H_i (x^k, \alpha)$ $(i = 0, 1, \ldots, m)$ определяются по одной из формул (7.4), (7.7), (7.8).

Следовательно, трудоемкость конечно-разностных методов решения выпуклых задач (7.1) оценивается величиной $O (m + 1) (\sqrt{n} L/(\nu - \nu_0))^2$. Исследование метода (7.11) приводится в § 18.

Таким образом, конечно-разностные методы, изучаемые в настоящей монографии, можно широко использовать при решении негладких экстремальных задач.

Рассмотрим скорость сходимости конечно-разностных методов. Предыдущие оценки трудоемкости были получены для конечно-разностных методов с постоянными параметрами ρ_k и α_k. Рассмотрим теперь ситуацию, когда ρ_k, $\alpha_k \to 0$ при $k \to \infty$, а минимизируемая функция $f (x)$ сильно выпукла, т. е. удовлетворяет неравенству

$$f (\varepsilon x + (1 - \varepsilon) y) \leqslant \varepsilon f (x) + (1 - \varepsilon) f (y) - \varepsilon (1 - \varepsilon) \lambda \| x - y \|^2$$

для любых x, y, $\varepsilon \in [0, 1]$; $\lambda > 0$ — параметр сильной выпуклости. Исследуем конечно-разностный метод минимизации $f (x)$ на выпуклом множестве X, определяемый формулой

$$x^{k+1} = \pi_X \left(x^k - \frac{\rho_k}{2\alpha_k} \sum_{i=1}^{n} [f(\tilde{x}_1^k, \dots, x_i^k + \alpha_k, \dots, \tilde{x}_n^k) - \right.$$

$$\left. - f(\tilde{x}_1^k, \dots, x_i^k - \alpha_k, \dots, \tilde{x}_n^k)] e_i \right).$$

Параметры ρ_k, α_k выбираются следующим образом:

$$\rho_k = bk^{-\beta}, \quad b > 0, \quad 0 < \beta \leqslant 1,$$

$$\alpha_k = mk^{-2\mu}, \quad m > 0, \quad \mu > 0.$$

Считается, что $x^* \in X$, где x^* — точка минимума $f(x)$. Из результатов работы [70] вытекает, что если $\beta = 1$, $\mu > 1$, $2b\lambda > 1$, то средне-квадратическая скорость и скорость сходимости почти наверное являются наибольшими, причем $M \| x^k - x^* \|^2$ имеет порядок k^{-1}.

ГЛАВА 3

МЕТОДЫ ОБОБЩЕННОГО ГРАДИЕНТНОГО СПУСКА

В данной главе изучаются простейшие методы минимизации невыпуклых негладких функций — метод обобщенного градиента и метод типа локального наискорейшего спуска. Они используют значения и обобщенные градиенты минимизируемых функций и сходятся к локальным экстремумам соответствующих задач оптимизации. Для доказательства сходимости привлекаются обобщенные функции Ляпунова. Обсуждаются проблемы поиска глобального экстремума.

§ 8. Условия сходимости итерационных алгоритмов нелинейного программирования

В этом параграфе детально рассматривается техника доказательства сходимости алгоритмов оптимизации, изучаемых в данной книге. Вводится некоторое общее понятие «итерационный алгоритм», определяется его сходимость и устанавливаются некоторые необходимые и достаточные, а также общие достаточные условия сходимости алгоритмов в терминах обобщенной функции Ляпунова.

Исследование сходимости численных методов невыпуклой негладкой оптимизации заметно сложнее, чем методов выпуклой или гладкой оптимизации, поскольку в них часто отсутствует какая бы то ни было релаксационность процесса. Для доказательства сходимости здесь используется специальная техника, основанная на обобщенной функции Ляпунова. Эта техника была разработана в [87] в виде достаточных условий сходимости алгоритмов нелинейного программирования и затем совершенствовалась в работах многих авторов. Мы приведем современный вариант достаточных условий, а также для сравнения некоторые необходимые и достаточные условия сходимости. Можно также сказать, что в этом параграфе будет собрана и подробно рассмотрена общая часть доказательств сходимости всех методов, изучаемых в книге. Поэтому для конкретных методов в соответствующих параграфах основное внимание будет уделено специфической части доказательства сходимости, т. е. проверке общих условий сходимости.

1. Определения. Пусть M — полное метрическое пространство, $\rho(x, y)$ — расстояние между точками $x, y \in M$, $M^n = M \times \ldots \times M$ — n-я декартова степень M, 2^M — множество подмножеств пространства M, $\bar{D}$ — замыкание множества $D \subset M$,

$$\rho(x, D) = \inf \{\rho(x, y) \mid y \in D\}$$

— расстояние от точки x до множества D.

Определение 8.1. *Итерационным алгоритмом* $A=(A_k, D^k)_{k\geqslant l}$ называется набор таких множеств $D^k \subset M^{k+1}$ и отображений $A_k\colon D^k \to 2^M$, $k = l, l+1, \ldots$, что для любых $(x^0, \ldots, x^k) \in D^k$ имеет место

$$(x^0, \ldots, x^k) \times A_k(x^0, \ldots, x^k) \subset D^{k+1}, \qquad k = l, l+1, \ldots$$

Определение 8.2. Всякий набор точек $\{x^k\}_{k=0}^s$ из M такой, что $(x^0, \ldots, x^s) \in D^s$, где $s \geqslant l$, будем называть *начальными данными для алгоритма*

$$A = (A_k, D^k)_{k\geqslant l}.$$

Всякую последовательность $\{x^k\}_{k=0}^\infty$ из M такую, что $(x^0, \ldots, x^k) \in D^k$ для всех $k \geqslant s$, где $s \geqslant l$, будем называть *последовательностью начальных данных* для A.

Определение 8.3. Последовательность точек $\{x^k\}_{k=0}^\infty$ называется *порожденной алгоритмом*

$$A = (A_k, D^k)_{k\geqslant l}$$

при начальных данных $(x^0, \ldots, x^s)$, если $(x^0, \ldots, x^s) \in D^s$ и

$$x^{k+1} \in A_k(x^0, \ldots, x^k), \qquad k = s, s+1, \ldots, \qquad s \geqslant l.$$

Очевидно, что всякая порожденная алгоритмом A последовательность является и последовательностью начальных данных для A.

Определение 8.4. *Областью определения алгоритма* $A = (A_k, D^k)_{k\geqslant l}$ называется множество

$$D_A = \bigcup_{k=l}^{\infty} \bigcup_{i=0}^{k} \{x^i \mid (x^0, \ldots, x^k) \in D^k\}.$$

В приведенных определениях подразумевается, что итерационный алгоритм вырабатывает бесконечные последовательности точек, т. е. работает бесконечно долго. Конечные алгоритмы формально можно рассматривать как бесконечно долго работающие, если считать, что, начиная с некоторого момента, они вырабатывают одну и ту же точку.

Определение 8.5. Последовательность точек $\{x^k\}_{k=0}^\infty$ называется *сходящейся к множеству* $D \subset M$ *по расстоянию*, если

$$\lim_{k\to\infty} \rho(x^k, D) = 0.$$

Определение 8.6. Алгоритм A называется *сходящимся к множеству* $X^* \subset M$ *по расстоянию*, если любая порожденная им последовательность сходится к X^* по расстоянию.

Часто используются несколько отличные определения сходимости последовательностей и алгоритмов к множеству.

Определение 8.7. Последовательность $\{x^k\} \subset M$ называется *компактной*, если она принадлежит некоторому компактному множеству.

Бесконечная последовательность компактна тогда и только тогда, когда любая ее подпоследовательность имеет предельные точки.

Определение 8.8. Компактная последовательность точек $\{x^k\}_{k=0}^{\infty}$ называется *сходящейся к множеству* $X^* \subset M$ *предельными точками*, если все ее предельные точки принадлежат X^*.

Определение 8.9. Алгоритм A называется *компактным*, если при любых начальных данных он порождает компактные последовательности.

Определение 8.10. Компактный алгоритм A называется *сходящимся к множеству* $X^* \subset M$ *предельными точками*, если любая порожденная им последовательность сходится к X^* предельными точками.

Для компактного алгоритма A и замкнутого множества X^* сходимость A к X^* по расстоянию эквивалентна сходимости предельными точками.

Замечание 8.1. Пусть алгоритм A, сходящийся к X^*, порождает последовательности $\{x^k\}$, в которых, начиная с некоторого момента, повторяется одна и та же точка. Это предположение соответствует как бы конечности алгоритма A. Тогда из определений сходимости следует, что

$$x = \lim_{k\to\infty} x^k \in \overline{X}^*,$$

в полном согласии с интуитивным пониманием сходимости конечных алгоритмов.

2. Необходимые и достаточные условия сходимости алгоритмов. В невыпуклой оптимизации сходимость итерационных алгоритмов часто устанавливается с помощью рассуждений от противного. В общей форме им можно придать следующий вид.

Теорема 8.1. *Для того чтобы алгоритм* $A = (A_k, D^k)_{k\geq l}$ *сходился к множеству* X^* (*по расстоянию*), *необходимо и достаточно, чтобы существовала непрерывная функция* $W(x)$ $(x \in \bar{D}_A)$ *такая, что для любой последовательности* $\{x^k\}$, *порожденной* A, *выполняются условия*:

1) *числовая последовательность* $\{W(x^k)\}_{k=0}^{\infty}$ *имеет предел* (*или вместо этого выполнено какое-либо условие, эквивалентное сходимости* $\{W(x^k)\}_{k=0}^{\infty}$);

2) *из формального предположения, что* $\{x^k\}$ *не сходится к* X^* (*по расстоянию*), *следует, что* $\{W(x^k)\}$ *предела не имеет* (*или выполнено какое-либо достаточное условие расходимости* $\{W(x^k)\}$).

Доказательство. Сначала докажем достаточность. Для всех последовательностей $\{x^k\}$, порожденных алгоритмом, числовая последовательность $\{W(x^k)\}$, согласно условию 1), имеет предел. Разобьем все множество последовательностей $\{x^k\}$ на подмножества сходящихся и не сходящихся к X^*. На не сходящихся к X^* последовательностях $\{x^k\}$ числовая последовательность $\{W(x^k)\}$ в силу условия 2) не имеет предела, и, следовательно, множество таких последовательностей пусто.

Докажем необходимость. Пусть алгоритм A сходится к X^*. Возьмем $W(x) = \rho(x, X^*)$; функция $\rho(x, X^*)$ непрерывна [46]. Для всех $\{x^k\}$, порожденных алгоритмом A, в силу определения сходимости A имеем

$$\lim_{k\to\infty} \rho(x^k, X^*) = \lim_{k\to\infty} W(x^k) = 0,$$

т. е. выполнено условие 1) теоремы. Условие 2) теоремы относится к не сходящимся к множеству X^* последовательностям $\{x^k\}$, множество которых пусто. Поэтому посылка условия 2) является ложной, и с точки зрения формальной логики оно (независимо от следствия) является истинным. Теорема доказана.

Теорема 8.1 не выделяет явно задачи минимизации или максимизации. Ей можно придать форму, ориентированную на задачи минимизации. Сначала докажем одно необходимое и достаточное условие сходимости числовых последовательностей.

Лемма 8.1. *Для того чтобы числовая последовательность $\{W^k\}_{k=0}^{\infty}$ имела конечный предел, необходимо и достаточно, чтобы:*

1) $\{W^k\}_{k=0}^{\infty}$ *была ограниченной;*

2) *если* $\lim\limits_{s\to\infty} W^{k_s} = W'$, *то для любого* $\varepsilon > 0$ *найдется такой номер* $K(\varepsilon)$, *что для всех* $k \geqslant K(\varepsilon)$ *будет* $W^k \leqslant W' + \varepsilon$.

Доказательство. Необходимость непосредственно следует из определения предела числовой последовательности. Докажем достаточность. Предположим противное:

$$\underline{W} = \varliminf_{k\to\infty} W^k < \varlimsup_{k\to\infty} W^k = \overline{W}.$$

Выберем такие числа c и d, что $\underline{W} < c < d < \overline{W}$. Последовательность $\{W^k\}$ бесконечно много раз пересекает отрезок $[c, d]$ от c к d; поэтому существуют такие индексы k_s и m_s $(s = 0, 1, \ldots)$, что $k_s < m_s$ и $W^{k_s} \leqslant c < W^k < d \leqslant W^{m_s}$ при $k \in (k_s, m_s)$ или $W^{k_s} \leqslant c < d \leqslant W^{k_s+1}$. Не теряя общности, можем считать, что

$$\lim_{s\to\infty} W^{k_s} = W' \leqslant c.$$

Возьмем $\varepsilon < d - c$. Согласно условию 2) леммы, для всех достаточно больших k имеем $W^k \leqslant W' + \varepsilon < d$, что противоречит сделанному ранее построению. Лемма доказана.

Теорема 8.2. *Для того чтобы алгоритм $A = (A_k, D^k)_{k\geqslant l}$ сходился к множеству X^* (по расстоянию), необходимо и достаточно, чтобы существовала непрерывная функция $W: \overline{D}_A \to E_1$ такая, что для любой последовательности $\{x^k\}$, порожденной алгоритмом, выполняются условия:*

0) *числовая последовательность $\{W(x^k)\}$ ограничена;*

1) *если некоторая подпоследовательность $\{x^{k_s}\}_{s=0}^{\infty}$ не сводится к*

X^* (по расстоянию) и $\lim\limits_{s\to\infty} W(x^{k_s}) = W'$, то существует другая такая подпоследовательность $\{x^{l_s}\}_{s=0}^{\infty}$, что

$$W' = \lim_{s\to\infty} W(x^{k_s}) > \overline{\lim_{s\to\infty}} W(x^{l_s});$$

2) если $\lim\limits_{s\to\infty} W(x^{k_s}) = W'$, то для любого $\varepsilon > 0$ существует такой номер $K(\varepsilon)$, что при всех $k \geqslant K(\varepsilon)$ имеет место неравенство

$$W(x^k) \leqslant W' + \varepsilon.$$

Доказательство. Условия 0), 2) теоремы, согласно лемме 8.1, эквивалентны сходимости последовательности $\{W(x^k)\}$, а условие 1) является достаточным для расходимости $\{W(x^k)\}$, поэтому теорема 8.2 следует из теоремы 8.1. Теорема доказана.

Теперь перегруппируем условия теоремы 8.2. Условие 2) разобьем на два: первое будет относиться к подпоследовательностям $\{x^{k_s}\}$, не сходящимся к X^* (его затем объединим с условием 1) теоремы), а второе будет относиться только к подпоследовательностям $\{x^{k_s}\}$, сходящимся к X^*. В результате придем к следующим необходимым и достаточным условиям сходимости.

Теорема 8.3. *Для того чтобы алгоритм $A = (A_k, D^k)_{k\geqslant l}$ сходился к X^* (по расстоянию), необходимо и достаточно, чтобы существовала такая непрерывная функция $W: \overline{D}_A \to E_1$, что для любой последовательности $\{x^k\}$, порожденной A, выполняются условия:*

0) числовая последовательность $\{W(x^k)\}$ ограничена;

1) если подпоследовательность $\{x^{k_s}\}$ не сходится к X^* (по расстоянию) и $\lim\limits_{s\to\infty} W(x^{k_s}) = W'$, то для любого $\varepsilon > 0$ найдутся такие индексы $l_s \geqslant k_s$, что $W(x^k) \leqslant W' + \varepsilon$ при $k \in [k_s, l_s)$ и

$$W' = \lim_{s\to\infty} W(x^{k_s}) > \overline{\lim_{s\to\infty}} W(x^{l_s});$$

2) если подпоследовательность $\{x^{k_s}\}$ сходится к X^* (по расстоянию) и $\lim\limits_{s\to\infty} W(x^{k_s}) = W'$, то для любого $\varepsilon > 0$ существует такой номер $K(\varepsilon)$, что для всех $k \geqslant K(\varepsilon)$ имеет место

$$W(x^k) \leqslant W' + \varepsilon.$$

Доказательство. Покажем, что теоремы 8.2 и 8.3 эквивалентны в том смысле, что если функция W и последовательность $\{x^k\}$, порожденная алгоритмом, удовлетворяют условиям 0) — 2) теоремы 8.2, то они удовлетворяют и условиям теоремы 8.3, и наоборот.

Введем для последовательности $\{x^k\}$ и такой ее подпоследовательности $\{x^{k_s}\}$, что

$$\lim_{s\to\infty} W(x^{k_s}) = W',$$

конечные или бесконечные числа

$$m_s(\varepsilon) = \sup\{m \mid |W(x^k) - W'| \leqslant \varepsilon \text{ при } k \in [k_s, m)\},$$

если $|W(x^{k_s}) - W'| \leqslant \varepsilon$, и $m_s(\varepsilon) = k_s$ в противном случае.

Пусть для $\{x^k\}$ и некоторой функции W выполнены условия 0) — 2) теоремы 8.2. Очевидно, что тогда для тех же $\{x^k\}$ и W выполнены условия 0), 2) теоремы 8.3. Покажем, что также выполнено условие 1) теоремы 8.3. Пусть существует подпоследовательность $\{x^{k_s}\}$, не сходящаяся к X^*, и

$$\lim_{s\to\infty} W(x^{k_s}) = W'.$$

В силу условия 1) теоремы 8.2 найдется $\varepsilon' > 0$, для которого $m_s(\varepsilon') < \infty$. Покажем, что для любого $\varepsilon \in (0, \varepsilon']$ имеет место неравенство

$$W' = \lim_{s\to\infty} W(x^{k_s}) > \overline{\lim_{s\to\infty}}\, W(x^{m_s(\varepsilon)});$$

отсюда будет следовать условие 1) теоремы 8.3. Предположим противное; тогда для некоторого $\varepsilon \leqslant \varepsilon'$ найдутся такие сколь угодно большие s, что

$$W' + \varepsilon \leqslant W(x^{m_s(\varepsilon)}).$$

Это противоречит условию 2) теоремы 8.2.

Пусть теперь для $\{x^k\}$, порожденной алгоритмом A, и некоторой функции $W(x)$ выполнены условия 0) — 2) теоремы 8.3. Очевидно, что тогда для тех же $\{x^k\}$ и W выполнены условия 0), 1) теоремы 8.2. Проверим справедливость условия 2) теоремы 8.2. Пусть

$$\lim_{s\to\infty} W(x^{k_s}) = W'.$$

Это условие с очевидностью выполняется на подпоследовательностях $\{x^{k_s}\}$, сходящихся к X^*. Пусть теперь подпоследовательность $\{x^{k_s}\}$ не сходится к X^*, и предположим, что в этом случае условие 2) теоремы 8.2 не выполняется. Тогда для некоторого $\varepsilon > 0$ и бесконечного числа номеров s существуют такие индексы $m_s \geqslant k_s$, что

$$W(x^{m_s}) > W' + \varepsilon.$$

Без ограничения общности можем считать, что это неравенство выполнено для всех s.

В силу условия 1) теоремы 8.3 найдутся индексы l_s, для которых $k_s \leqslant l_s \leqslant m_s$ и $W(x^{l_s}) < W'$. Числовые последовательности $\{W(x^{l_s}), \ldots$ $\ldots, W(x^{m_s})\}$ $(s = 0, 1, \ldots)$ пересекают отрезок $[W', W' + \varepsilon]$ от W' к

$W' + \varepsilon$; поэтому найдутся такие индексы p_s и q_s $(l_s \leqslant p_s < q_s \leqslant m_s)$, что

$$W(x^{p_s}) \leqslant W' < W(x^k) < W' + \varepsilon \leqslant W(x^{q_s}), \qquad k \in (p_s, q_s), \quad (8.1)$$

или

$$W(x^{p_s}) \leqslant W' < W' + \varepsilon \leqslant W(x^{q_s}), \qquad q_s = p_s + 1.$$

Без ограничения общности можем считать, что

$$\lim_{s \to \infty} W(x^{p_s}) = W'' \leqslant W'.$$

Теперь, если $\{x^{p_s}\}$ сходится к X^*, то неравенства (8.1) противоречат условию 2) теоремы 8.3, а если $\{x^{p_s}\}$ не сходится к X^*, то к $\{x^{p_s}\}$ применимо условие 1) теоремы 8.3, которое также противоречит построению (8.1). Теорема доказана.

3. **Достаточные условия сходимости.** Сначала несколько ослабим условие 2) теоремы 8.3 и таким образом получим необходимые условия сходимости алгоритмов. При ослабленных (необходимых) условиях уже не будет иметь места сходимость алгоритмов в смысле определения 8.6, но, оказывается, в этом случае имеет место ослабленная сходимость, т. е. в некотором смысле сходимость алгоритмов по функции (теорема 8.4). Затем несколько усилим условия теоремы 8.3 и таким образом получим достаточные условия сходимости алгоритмов (теорема 8.5).

Введем обозначения

$$\underline{W} = \varliminf_{k \to \infty} W(x^k), \qquad \overline{W} = \varlimsup_{k \to \infty} W(x^k). \qquad (8.2)$$

Через W^* обозначим множество предельных точек числовых последовательностей $\{W(x^k)\}$, где $\{x^k\}$ — произвольные бесконечные последовательности точек, сходящиеся к X^*. Если X^* — компакт в M, то

$$W^* = \{W(x) \mid x \in X^*\}.$$

Теорема 8.4. *Для того чтобы алгоритм $A = (A_k, D^k)_{k \geqslant l}$ сходился к множеству X^* (по расстоянию), необходимо, чтобы существовала такая непрерывная функция W: $\bar{D}_A \to E_1$, что для любой последовательности $\{x^k\}$, порожденной A, выполняются условия:*

0) *числовая последовательность $\{W(x^k)\}$ ограничена;*

1) *если подпоследовательность $\{x^{k_s}\}$ не сходится к X^* (по расстоянию) и $\lim_{s \to \infty} W(x^{k_s}) = W'$, то для любого $\varepsilon > 0$ найдутся такие индексы $l_s \geqslant k_s$, что $W(x^k) \leqslant W' + \varepsilon$ при $k \in [k_s, l_s)$, и имеет место неравенство*

$$W' = \lim_{s \to \infty} W(x^{k_s}) > \varlimsup_{s \to \infty} W(x^{l_s});$$

2) *если подпоследовательность* $\{x^{k_s}\}$ *сходится к* X^* (*по расстоянию*) *и* $\lim\limits_{s\to\infty} W(x^{k_s}) = W'$, *то*

$$\overline{\lim_{s\to\infty}}\,[W(x^{k_s+1}) - W(x^{k_s})] \leqslant 0.$$

Если существует непрерывная функция W, *удовлетворяющая условиям* 0) — 2), *то алгоритм* A *сходится по функции, т. е. для любой последовательности* $\{x^k\}$, *порожденной* A, *отрезок* $[\underline{W}, \overline{W}]$ *вложен в множество* W^*, *и все подпоследовательности* $\{x^{k_s}\}$ *такие, что* $\lim\limits_{s\to\infty} W(x^{k_s}) = \underline{W}$, *сходятся к* X^*.

Доказательство. Условие 2) теоремы 8.4 следует из условия 2) теоремы 8.3; поэтому условия теоремы 8.4 являются необходимыми для сходимости алгоритма A. Докажем их достаточность для сходимости алгоритма по функции.

Покажем, что если

$$\lim_{s\to\infty} W(x^{k_s}) = \underline{W},$$

то $\{x^{k_s}\}$ сходится к X^*. Предположим противное. Тогда к $\{x^{k_s}\}$ применимо условие 2) теоремы, которое утверждает, что существует такая подпоследовательность $\{x^{l_s}\}$, что

$$\overline{\lim_{s\to\infty}}\, W(x^{l_s}) \leqslant \underline{W}.$$

Это противоречит определению $\underline{W}$ как нижнего предела последовательности $\{W(x^k)\}$. Требуемое утверждение доказано. Одновременно мы показали, что $\underline{W} \in W^*$.

Докажем теперь, что $[\underline{W}, \overline{W}) \subset W^*$. Предположим противное. Тогда существует число $c \in [\underline{W}, \overline{W})$ и $c \notin W^*$. Так как $\underline{W} \in W^*$, то $c > \underline{W}$. Выберем число d так, что $\underline{W} < c < d < \overline{W}$; тогда отрезок $[c, d]$ вложен в $(\underline{W}, \overline{W})$ и $c \notin W^*$. Последовательность $\{W(x^k)\}$ бесконечное число раз пересекает отрезок $[c, d]$ от c к d; поэтому существуют части $\{x^{k_s}, \ldots$ $\ldots, x^{m_s}\}$ $(s = 0, 1, \ldots)$ последовательности $\{x^k\}$ такие, что либо

$$W(x^{k_s}) \leqslant c < W(x^k) < d \leqslant W(x^{m_s}), \qquad k \in (k_s, m_s), \tag{8.3}$$

либо

$$W(x^{k_s}) \leqslant c < d \leqslant W(x^{k_s+1}).$$

Поскольку $\{W(x^k)\}$ ограничена, то, не ограничивая общности, можно считать, что

$$\lim_{s\to\infty} W(x^{k_s}) = W' \leqslant c.$$

Покажем, что $\{x^{k_s}\}$ не сходится к X^*. Предположим противное. Тогда в силу (8.3) и условия 2) теоремы имеет место

$$\lim_{s\to\infty} W(x^{k_s}) = \lim_{s\to\infty} W(x^{k_s+1}) = c,$$

и так как, согласно предположению, $x^{k_s} \to X^*$, то $c \in W^*$, что противоречит выбору c. Следовательно, $\{x^{k_s}\}$ не сходится к X^*.

К подпоследовательности $\{x^{k_s}\}$ применимо условие 1) теоремы. Возьмем $\varepsilon < d - c$. Согласно условию 1) теоремы, найдутся такие индексы $l_s \geqslant k_s$, что $W(x^k) \leqslant W' + \varepsilon < d$ при $k \in [k_s, l_s)$, и имеет место

$$W' = \lim_{s \to \infty} W(x^{k_s}) > \overline{\lim_{s \to \infty}} W(x^{l_s}).$$

Но это означает, что $l_s \leqslant m_s$, и минимизирующее условие 1) теоремы противоречит сделанному построению (8.3). Итак, $[\underline{W}, \overline{W}) \subset W^*$.

Осталось заметить, что множество W^* замкнуто; поэтому из $[\underline{W}, \overline{W}) \subset W^*$ следует $[\underline{W}, \overline{W}] \subset W^*$. Теорема доказана.

Теперь несколько усилим необходимые условия теоремы 8.4, а именно добавим к ним одно дополнительное требование. При этом окажется, что усиленные условия будут достаточными для сходимости алгоритмов.

Теорема 8.5. *Чтобы алгоритм* $A = (A_k, D^k)_{k \geqslant l}$ *сходился к множеству* X^* *(по расстоянию), достаточно существования такой непрерывной функции* $W\colon \bar{D}_A \to E_1$, *чтобы для любой последовательности* $\{x^k\}$, *порожденной* A, *выполнялись условия* 1), 2) *теоремы* 8.3 *и условие*

3) *множество* W^* *не содержит отрезков.*

В этом случае $\{x^k\}$ *сходится к* X^* *(по расстоянию), и числовая последовательность* $\{W(x^k)\}$ *имеет предел.*

Доказательство. Согласно теореме 8.3 $[\underline{W}, \overline{W}] \subset W^*$. Но по условию теоремы линейное множество W^* не содержит отрезков; поэтому $\{W(x^k)\}$ имеет предел

$$\lim_{k \to \infty} W(x^k) = W'.$$

Предположим, что существует подпоследовательность $\{x^{k_s}\}$, не сходящаяся к X^*. Очевидно, что

$$\lim_{s \to \infty} W(x^{k_s}) = W'.$$

К подпоследовательности $\{x^{k_s}\}$ теперь применимо условие 1) теоремы 8.3, которое, в частности, означает расходимость $\{W(x^k)\}$. Полученное противоречие доказывает теорему. Заметим, что здесь мы фактически проверили условия общей теоремы 8.1.

Необходимые условия теоремы 8.3 и достаточные условия теоремы 8.4 дают вилку, содержащую необходимые и достаточные условия сходимости алгоритмов.

Конкретизируем теоремы 8.4, 8.5 для компактных алгоритмов.

Теорема 8.6. *Для сходимости компактного алгоритма* $A = (A_k, D^k)_{k \geqslant l}$ *к компактному множеству* X^* *(предельными точками) необходимо существование непрерывной функции* $W\colon \bar{D}_A \to E_1$ *такой,* *что для любой последовательности $\{x^k\}$, порожденной A, выполнены условия:*

1) *если* $\{x^{k_s}\} \to x \notin X^*$, *то для любого* $\varepsilon > 0$ *найдутся индексы* $l_s \geqslant k_s$ *такие, что*

$$W(x^k) \leqslant W(x) + \varepsilon \quad \text{при} \quad k \in [k_s, l_s),$$

$$W(x) = \lim_{s \to \infty} W(x^{k_s}) > \overline{\lim_{s \to \infty}}\, W(x^{l_s});$$

2) *если* $\{x^{k_s}\} \to x \in X^*$, *то*

$$\overline{\lim_{s \to \infty}}\, [W(x^{k_s+1}) - W(x^{k_s})] \leqslant 0.$$

Если существует функция W, удовлетворяющая условиям 1), 2), *то компактный алгоритм A сходится к X^* по функции, т. е. отрезок* $[\underline{W}, \overline{W}]$ *вложен в множество* $W^* = \{W(x) \mid x \in X^*\}$, *и минимальные по значению W предельные точки $\{x^k\}$ принадлежат X^*.*

Теорема 8.7. *Для сходимости компактного алгоритма* $A = (A_k, D^k)_{k \geqslant l}$ *к компактному множеству X^* (предельными точками) достаточно существования непрерывной функции* $W\colon \overline{D}_A \to E_1$ *такой, что для любой последовательности $\{x^k\}$, порожденной A, выполнены условия* 1), 2) *теоремы* 8.6 *и условие:*

3) *множество* $W^* = \{W(x) \mid x \in X^*\}$ *не содержит интервалов.*

В этом случае все предельные точки последовательности $\{x^k\}$ принадлежат X^* и числовая последовательность $\{W(x^k)\}$ имеет предел.

Теоремы 8.6, 8.7 представляют собой весьма общие условия сходимости итерационных (компактных) алгоритмов. Теперь сузим класс рассматриваемых алгоритмов. Сформулируем более сильные (но менее общие) достаточные условия сходимости итерационных алгоритмов. В условиях 1), 2) теорем 8.6, 8.7 в качестве меры близости точек выступала разность значений функции W в этих точках. В следующей теореме возьмем в качестве такой меры расстояние между точками.

Теорема 8.8. *Пусть для компактного алгоритма* $A = (A_k, D^k)_{k \geqslant l}$ *существуют непрерывная функция* $W\colon \overline{D}_A \to E_1$ *и множество X^* такие, что для любой последовательности $\{x^k\}$, порожденной A, выполнено условие:*

1) *если подпоследовательность* $\{x^{k_s}\} \to x \notin X^*$, *то для любого* $\delta > 0$ *найдутся такие индексы* $l_s \geqslant k_s$, *что* $\rho(x, x^k) \leqslant \delta$ *при* $k \in [k_s, l_s)$ *и*

$$W(x) = \lim_{s \to \infty} W(x^{k_s}) > \overline{\lim_{s \to \infty}}\, W(x^{l_s});$$

тогда X^* не пусто, и минимальные по значению W предельные точки последовательности $\{x^k\}$ принадлежат X^*.

Пусть, кроме того, выполнено условие:

2) *если* $\{x^{k_s}\} \to x \in X^*$, *то*

$$\lim_{s\to\infty} \rho(x^{k_s+1}, x^{k_s}) = 0;$$

тогда алгоритм сходится к X^* *по функции, т. е. минимальные по значению* W *предельные точки последовательности* $\{x^k\}$ *принадлежат* X^*, *а полуинтервал* $[\underline{\lim}_{k\to\infty} W(x^k), \overline{\lim}_{k\to\infty} W(x^k))$ *вложен в множество* $W^* = \{W(x) \mid x \in X^*\}$.

Если в дополнение к предыдущим условиям

3) *множество* $W^* \subset E_1$ *не содержит интервалов, то алгоритм сходится к* X^*, *т. е. все предельные точки* $\{x^k\}$ *принадлежат связному подмножеству* X^*, *числовая последовательность* $\{W(x^k)\}$ *имеет предел и*

$$\lim_{k\to\infty} \rho(x^{k+1}, x^k) = 0.$$

Алгоритмы, удовлетворяющие условиям теоремы 8.8, можно назвать *локальными*, а функцию $W(x)$ естественно назвать *обобщенной функцией Ляпунова* для алгоритма A.

В формулировке теоремы условие 1), грубо говоря, означает, что если алгоритм стартует из точки $x \notin X^*$, то он находит меньшие, чем $W(x)$, значения функции W в малой окрестности x. Условие 2) есть ослабленное требование

$$\lim_{k\to\infty} \rho(x^{k+1}, x^k) = 0,$$

которое выполнено как следствие всех условий теоремы 8.8.

Теорема утверждает сходимость алгоритма A к связным подмножествам множества X^*, которые на практике соответствуют локальным экстремумам (точнее, являются стационарными множествами) функции $W(x)$.

Теорема 8.8 будет основным инструментом доказательства сходимости итерационных алгоритмов невыпуклой оптимизации в этой книге. Если выполнены условия теоремы 8.8, то в силу непрерывности $W(x)$ выполнены и условия теоремы 8.7. Однако из достаточных условий теоремы 8.8 следует ряд дополнительных результатов о сходимости алгоритмов.

Доказательство теоремы 8.8. Докажем первое утверждение теоремы. Множество предельных точек последовательности $\{x^k\}$, порожденной алгоритмом, очевидно, замкнуто. Так как $W(x)$ непрерывна, то существуют минимальные по значению W предельные точки $\{x^k\}$. Теперь предположим противное: существует минимальная предельная точка $x = \lim_{s\to\infty} x^{k_s}$, не принадлежащая X^*. К подпоследовательности $\{x^{k_s}\}$ применимо минимизирующее условие 1) теоремы, которое, в частности, означает, что существует еще меньшая по значению W предельная точка $\{x^k\}$, т. е. точка $x = \lim_{s\to\infty} x^{k_s}$ не минимальна. Полученное противоречие доказывает, что X^* не пусто, и все минимальные по значению W предельные точки $\{x^k\}$ принадлежат X^*.

Докажем второе утверждение теоремы. Покажем, что интервал $[\underline{W}, \overline{W})$, где $\underline{W}$, $\overline{W}$ определены по (8.2), вложен в множество W^*. Предположим противное, тогда найдется c такое, что $c \in [\underline{W}, \overline{W})$ и $c \notin W^*$. Согласно доказанному первому утверждению теоремы, $\underline{W} \in W^*$; поэтому $c > \underline{W}$. Выберем число $d \in (c, \overline{W})$. Таким образом, интервал $[c, d]$ вложен в $(\underline{W}, \overline{W})$ и $c \notin W^*$. Числовая последовательность $\{W(x^k)\}$ бесконечное число раз пересекает отрезок $[c, d]$ от c к d, поэтому существуют такие части $\{x^{k_s}, \dots, x^{m_s}\}$ $(s = 0, 1, \dots)$ последовательности $\{x^k\}$, что либо

$$W(x^{k_s}) \leqslant c < W(x^k) < d \leqslant W(x^{m_s}), \qquad k \in (k_s, m_s), \tag{8.4}$$

либо

$$W(x^{k_s}) \leqslant c < d \leqslant W(x^{k_s+1}).$$

Подпоследовательность $\{x^{k_s}\}$ в силу компактности алгоритма имеет предельные точки. Не ограничивая общности, можем считать, что

$$\lim_{s\to\infty} x^{k_s} = x'.$$

Покажем, что $x' \notin X^*$. Предположим противное. Тогда по условию 2) теоремы

$$\lim_{s\to\infty} \rho(x^{k_s+1}, x^{k_s}) = 0,$$

и в силу непрерывности $W(x)$

$$\lim_{s\to\infty} [W(x^{k_s+1}) - W(x^{k_s})] = 0.$$

Вместе с построением (8.4) это дает

$$W(x') = \lim_{s\to\infty} W(x^{k_s}) = \lim_{s\to\infty} W(x^{k_s+1}) = c \in W^*,$$

что противоречит выбору c. Таким образом, $x' \notin X^*$.

К подпоследовательности $\{x^{k_s}\}$ применимо минимизирующее условие 1) теоремы. Выберем такое δ, что

$$\max\{W(y) \mid \rho(y, x') \leqslant \delta\} \leqslant d. \tag{8.5}$$

Согласно условию 1), найдутся такие индексы $l_s \geqslant k_s$, что для достаточно больших s будет $\rho(x^k, x') \leqslant \delta$ при $k \in [k_s, l_s)$, и

$$c \geqslant W(x') = \lim_{s\to\infty} W(x^{k_s}) > \overline{\lim_{s\to\infty}}\, W(x^{l_s}). \tag{8.6}$$

В силу (8.5) для всех $k \in [k_s, l_s)$ имеет место неравенство $W(x^k) < d$, и тогда, согласно построению (8.4), $l_s \leqslant m_s$, откуда $W(x^{l_s}) \geqslant c$. Таким образом, мы получили противоречие между соотношением (8.6) и построением (8.4). Второе утверждение теоремы доказано: $[\underline{W}, \overline{W}) \subset \subset W^*$.

Докажем третье утверждение. Согласно второму утверждению, $[\underline{W}, \overline{W}) \subset W^*$. Но так как W^* не содержит отрезков, то существует предел

$$\lim_{k \to \infty} W(x^k) = \overline{\lim_{k \to \infty}} W(x^k) = \underline{\lim_{k \to \infty}} W(x^k).$$

Теперь предположим, что существует предельная точка

$$x = \lim_{s \to \infty} x^{k_s},$$

не принадлежащая X^*. К подпоследовательности $\{x^{k_s}\}$ применимо условие 1) теоремы, которое, в частности, означает, что $\{W(x^k)\}$ предела не имеет. Полученное противоречие доказывает, что все предельные точки $\{x^k\}$ принадлежат X^*.

Заметим, что здесь мы фактически проверили условия общей теоремы 8.1.

Так как теперь все предельные точки $\{x^k\}$ принадлежат X^*, то из условия 2) и компактности алгоритма следует, что

$$\lim_{k \to \infty} \rho(x^{k+1}, x^k) = 0.$$

В конце параграфа в лемме 8.2 будет показано, что тогда множество предельных точек $\{x^k\}$ связно и, следовательно, принадлежит связному подмножеству X^*. Теорема доказана.

В предыдущих теоремах отражена с той или иной степенью подробности схема доказательства сходимости алгоритмов от противного; поэтому некоторые условия теорем носят формальный характер и вне контекста доказательства сходимости от противного психологически неудобны. Переформулируем теорему 8.8 уже без формальных предположений противного.

Пусть $\{x^k\}$ — последовательность начальных данных для алгоритма $A = (A_k, D^k)_{k \geqslant l}$. Обозначим $\{x_n^k\}_{k=0}^{\infty}$ последовательность, порожденную алгоритмом A при начальных данных $(x^0, \ldots, x^n)$, т. е.

$$\{x_n^k\}_{k=0}^{\infty} = \{x_n^k = x^k \text{ при } k \in [0, n];\ x_n^k \in A_{k-1}(x_n^0, \ldots, x_n^{k-1}) \text{ при } k > n\}.$$

Теорема 8.9. *Пусть для компактного алгоритма* $A = (A_k, D^k)_{k \geqslant l}$ *существуют непрерывная функция* W: $\overline{D}_A \to E_1$ *и множество* X^* *такие, что для любой последовательности начальных данных* $\{x^k\}$ *выполнено условие*:

1) если $\{x^{k_s}\} \to x \notin X^*$, то для любого $\delta > 0$ найдутся такие индексы $l_s \geq k_s$, что $\rho(x_{k_s}^k, x) \leq \delta$ при $k \in [k_s, l_s)$ и

$$W(x) = \lim_{s\to\infty} W(x^{k_s}) > \overline{\lim_{s\to\infty}}\, W(x_{k_s}^{l_s});$$

тогда X^* не пусто и минимальные по значению W предельные точки любой последовательности $\{y^k\}$, порожденной алгоритмом, принадлежат X^*.

Если, кроме того, выполнено условие:

2) если $\{x^{k_s}\} \to x \in X^*$, то

$$\lim_{s\to\infty} \rho(x^{k_s}, x^{k_s+1}) = 0;$$

тогда алгоритм A сходится к X^* по функции, т. е. для любой последовательности $\{y^k\}$, порожденной A, минимальные по значению W предельные точки $\{y^k\}$ принадлежат X^*, а интервал $[\underline{\lim}_{k\to\infty} W(y^k), \overline{\lim}_{k\to\infty} W(y^k))$ вложен в множество $W^* = \{W(x) \mid x \in X^*\}$.

Если в дополнение к предыдущим условиям

3) множество W^* не содержит отрезков, то алгоритм A сходится к X^*, т. е. для любой последовательности $\{y^k\}$, порожденной A, все предельные точки $\{y^k\}$ принадлежат связному подмножеству X^*, числовая последовательность $\{W(y^k)\}$ имеет предел и

$$\lim_{k\to\infty} \rho(y^k, y^{k+1}) = 0.$$

Доказательство. Пусть $\{y^k\}$ — последовательность, порожденная алгоритмом A. Очевидно, что она является и последовательностью начальных данных для A; поэтому к ней применимы условия 1), 2) теоремы, где нужно положить $x^k = y^k$ и $x_{k_s}^k = y^k$. Применительно к $\{y^k\}$ условия теоремы 8.9 совпадают с аналогичными условиями доказанной теоремы 8.8, откуда следует справедливость теоремы 8.9. Теорема доказана.

Заметим, что из условий теоремы 8.9 дополнительно следует, что локальные минимумы W вложены в X^*; точнее, если x — точка, где достигается локальный минимум W, для которого существует последовательность начальных данных $\{x^k\}$, имеющая x своей предельной точкой, то x принадлежит X^*.

Действительно, в противном случае, запуская алгоритм A из точек $x^k \to x$, согласно условию 1) теоремы 8.9, мы найдем меньшие, чем $W(x)$, значения W в сколь угодно малой окрестности x, что противоречит локальной оптимальности x.

Таким образом, в теореме 8.9 завязаны как условия сходимости алгоритмов оптимизации, так и необходимые условия экстремума функций.

В заключение параграфа докажем одну лемму о связности множеств предельных точек последовательностей, использованную при доказательстве теоремы 8.8, 8.9.

Сначала приведем несколько фактов из [2] о связности множеств в метрическом пространстве M.

Определение 8.11. Множество $X \subset M$ называется *связным*, если его нельзя представить в виде

$$X = (\Phi_1 \cap X) \cup (\Phi_2 \cap X),$$

где Φ_1 и Φ_2 — непустые замкнутые множества такие, что

$$(\Phi_1 \cap X) \cap (\Phi_2 \cap X) = \varnothing.$$

Определение 8.12. *Компонентой* множества X называется связное подмножество, не лежащее в другом, более широком связном подмножестве множества X.

Каждое множество $X \subset M$ распадается на свои связные компоненты.

Определение 8.13. Конечная последовательность точек $x^0, \ldots$ $\ldots, x^s$ метрического пространства M называется *ε-цепью* (или ε-цепью, соединяющей точку x^1 с точкой x^s), если $\rho(x^i, x^{i+1}) < \varepsilon$ для $i \in [0, s)$. Множество X называется *ε-сцепленным*, если любые две его точки могут быть соединены ε-цепью, составленной из точек множества X. Множество X называется *сцепленным*, если оно является ε-сцепленным при любом $\varepsilon > 0$.

Всякое связное множество в метрическом пространстве является сцепленным.

Всякий сцепленный компакт является связным.

Лемма 8.2. *Пусть последовательность* $\{x^k\}_{k=0}^{\infty}$ *вложена в некоторый компакт и*

$$\lim_{k\to\infty} \rho(x^k, x^{k+1}) = 0.$$

Тогда множество L *предельных точек* $\{x^k\}$ *является связным компактом.*

Доказательство. Покажем, что L является компактом, т. е. любая бесконечная последовательность точек из L имеет предельные точки. Пусть

$$\{y^n\}_{n=0}^{\infty} \subset L.$$

Так как $\{y^n\}$ являются предельными точками $\{x^k\}$, то существуют такие точки x^{k_n}, что

$$\rho(x^{k_n}, y^n) \leqslant 1/n.$$

По условию подпоследовательность $\{x^{k_n}\}$ имеет предельные точки; они же будут предельными и для $\{y^n\}$. Итак, компактность доказана.

Покажем, что L сцеплено, отсюда будет следовать связность L. Возьмем произвольное $\varepsilon > 0$. Покажем, что

$$\lim_{k\to\infty} \rho(x^k, L) = 0.$$

Заметим, что функция $\rho(x, L)$ непрерывна. Пусть

$$\overline{\lim_{k\to\infty}}\, \rho(x^k, L) = \lim_{s\to\infty} \rho(x^{k_s}, L), \quad \lim_{s\to\infty} x^{k_s} = x \in L.$$

Тогда в силу непрерывности $\rho(x, L)$

$$0 \leqslant \overline{\lim_{k\to\infty}}\, \rho(x^k, L) = \lim_{s\to\infty} \rho(x^{k_s}, L) = \rho(x, L) = 0,$$

что и требовалось доказать.

Пусть теперь для всех $k \geqslant K$ будет

$$\rho(x^k, x^{k+1}) < \varepsilon/3, \qquad \rho(x^k, L) < \varepsilon/3.$$

Пусть $a, b \in L$. Найдем ε-цепь в L, соединяющую a и b. Так как a и b — предельные точки $\{x^k\}$, то найдутся такие $k_2 > k_1 > K$, что

$$\rho(a, x^{k_1}) < \varepsilon/3, \qquad \rho(b, x^{k_2}) < \varepsilon/3.$$

Для точек $\{x^k\}_{k=k_1}^{k_2}$ по построению найдутся такие $y^k \in L$, что

$$\rho(x^k, y^k) < \varepsilon/3, \qquad k \in [k_1, k_2].$$

Очевидно, что

$$\rho(y^k, y^{k+1}) \leqslant \rho(y^k, x^k) + \rho(x^k, x^{k+1}) + \rho(x^{k+1}, y^{k+1}) \leqslant \varepsilon$$

при $k \in [k_1, k_2)$; $\rho(a, y^{k_1}) \leqslant 2\varepsilon/3$, $\rho(y^{k_2}, b) \leqslant 2\varepsilon/3$.

Таким образом, $\{a, y^{k_1}, \ldots, y^{k_2}, b\}$ — искомая ε-цепь. Так как ε произвольно, то L — сцепленное множество. Лемма доказана.

§ 9. Метод обобщенного градиента

В этом параграфе известный метод обобщенного градиента, предназначенный для минимизации выпуклых функций, распространяется на задачи безусловной минимизации невыпуклых негладких обобщенно дифференцируемых функций.

С метода обобщенного градиента начинается негладкая оптимизация. Первоначально он был применен для минимизации (с постоянным шагом) выпуклых кусочно линейных функций [134], затем были исследованы условия сходимости метода [40, 99] и построены многочисленные модификации и обобщения. Мы распространим этот важный метод на случай минимизации широкого класса (невыпуклых негладких) обобщенно дифференцируемых функций.

Напомним, что обобщенно дифференцируемые функции, вообще говоря, не имеют производных по направлениям.

Рассмотрим задачу:

$$f(x) \to \min, \qquad x \in E_n, \tag{9.1}$$

где $f(x)$ — обобщенно дифференцируемая функция такая, что $f(x) \to \infty$ при $\|x\| \to \infty$.

Определим множество (псевдо-) стационарных точек задачи (9.1):

$$X^* = \{x \in E_n \mid 0 \in G_f(x)\},$$

где $G_f(x)$ — множество псевдоградиентов $f(x)$ в точке x.

Множество X^* замкнуто ввиду замкнутости псевдоградиентного отображения G_f. Согласно теореме 3.1, все точки локального минимума $f(x)$ вложены в X^*.

Введем множество

$$F^* = \{f(x) \mid x \in X^*\}.$$

Метод обобщенного градиента имеет вид

$$x^0 \in E_n, \quad x^{k+1} = x^k - \rho_k g^k, \quad g^k \in G_f(x^k), \quad k = 0, 1, \ldots, \tag{9.2}$$

$$0 \leqslant \rho_k \leqslant R, \quad \lim_{k\to\infty} \rho_k = 0, \quad \sum_{k=0}^{\infty} \rho_k = +\infty. \tag{9.3}$$

Следующие утверждения устанавливают сходимость метода обобщенного градиента; они будут доказаны позднее.

Теорема 9.1. *Предположим, что последовательность* $\{x^k\}_{k=0}^{\infty}$, *порожденная алгоритмом* (9.2) — (9.3), *ограничена. Тогда она сходится к решению задачи* (9.1) *по функции, т. е. минимальные по значению* $f(x)$ *предельные точки* $\{x^k\}_{k=0}^{\infty}$ *принадлежат* X^*, *и все предельные точки числовой последовательности* $\{f(x^k)\}_{k=0}^{\infty}$ *составляют интервал, лежащий в множестве* F^*. *Если множество* F^* *не содержит интервалов (например, конечно или счетно), то последовательность* $\{x^k\}_{k=0}^{\infty}$ *сходится к решению задачи* (9.1), *т. е. все предельные точки* $\{x^k\}_{k=0}^{\infty}$ *принадлежат связному подмножеству* X^*, *и числовая последовательность* $\{f(x^k)\}_{k=0}^{\infty}$ *имеет предел.*

Теорема следует из доказываемой ниже леммы 9.2 и общей теоремы 8.9. Так как в нашем случае F^* замкнуто, то

$$[\varliminf_{k\to\infty} f(x^k), \quad \varlimsup_{k\to\infty} f(x^k)] \subset F^*.$$

Следующая лемма показывает, что если шаги ρ_k в методе обобщенного градиента достаточно малы, то порождаемые последовательности ограничены.

Лемма 9.1. *Пусть существует* $c > f(x^0)$, *не принадлежащее* F^*. *Тогда для любого* $d > c$ *найдется такое* $\bar{R}$, *что для любой последовательности* $\{\rho_k\}_{k=0}^{\infty}$, *удовлетворяющей* (9.3) *с* $R \leqslant \bar{R}$, *последовательности точек* $\{x^k\}$, *запущенные из* x^0 *согласно* (9.2), *не выходят из ограниченной области* $\{x \mid f(x) \leqslant d\}$.

Замечание 9.1. Требуемое в лемме 9.1 число c заведомо существует, если множество X^* (и, следовательно, F^*) ограничено или если F^* не содержит интервалов.

Замечание 9.2. Практически величина

$$R = \sup_{k\geqslant 0} \rho_k$$

подбирается следующим образом [88]. Пусть известно такое число $d > f(x^0)$, что интервал $[f(x^0), d)$ не лежит целиком в F^*. Тогда, согласно лемме 9.1, найдется такое $\bar{R}$, что при $R \leqslant \bar{R}$ последовательности $\{x^k\}$ равномерно ограничены. Запустим алгоритм (9.2) из точки x^0 с некоторыми $\{\rho_k\}_{k=0}^{\infty}$ и $R_0 = \sup_{k \geqslant 0} \rho_k$. Если на некотором шаге окажется, что $f(x^k) > d$, то прерываем процесс, выбираем другой ряд $\{\rho_k'\}_{k=0}^{\infty}$ (с меньшим $R_1 = \sup_{k \geqslant 0} \rho_k'$) и запускаем алгоритм (9.2) с новыми ρ_k' из начальной (или рекордной) точки x^0 и т. д. Если R_s устанавливается так, что $R_s \to 0$ при $s \to \infty$, то рано или поздно будет $R_s < \bar{R}$, и соответствующая траектория $\{x^k\}_{k=0}^{\infty}$ не выйдет из ограниченной области $\{x \mid f(x) \leqslant d\}$.

Доказательство теоремы 9.1 разобьем на ряд лемм. Прежде всего докажем важную геометрическую лемму.

Лемма 9.2. *Пусть P — выпуклое множество в E_n, причем $0 < \gamma \leqslant \|p\| \leqslant \Gamma < +\infty$ для всех $p \in P$. Тогда для любого набора векторов $\{p^r \in P \mid r = k, \ldots, m\}$ и для любого такого набора чисел $\{\rho_r \in E_1 \mid r = k, \ldots, m, \ \rho_r \geqslant 0\}$, что*

$$\sum_{r=k}^{m} \rho_r \geqslant \sigma > 0, \quad \sup_{k \leqslant r \leqslant m} \rho_r \leqslant R = \frac{\gamma^2 \sigma}{6\Gamma^2},$$

найдется индекс $l \in (k, m]$, для которого

$$\left(p^l, \sum_{r=k}^{l-1} \rho_r p^r \Big/ \sum_{r=k}^{l-1} \rho_r\right) \geqslant \frac{\gamma^2}{4}, \qquad \sum_{r=k}^{l-1} \rho_r \geqslant \frac{\gamma\sigma}{3\Gamma}.$$

Доказательство. Пусть

$$\sum_{r=k}^{t-1} \rho_r < \frac{\gamma\sigma}{3\Gamma} \leqslant \sum_{r=k}^{t} \rho_r, \qquad \sum_{r=k}^{m'-1} \rho_r < \sigma \leqslant \sum_{r=k}^{m'} \rho_r.$$

Предположим противное утверждению леммы: для всех $l \in (t, m']$

$$\left(p^l, \sum_{r=k}^{l-1} \rho_r p^r \Big/ \sum_{r=k}^{l-1} \rho_r\right) < \frac{\gamma^2}{4}.$$

Представим

$$\left\|\sum_{r=k}^{l} \rho_r p^r\right\|^2 = \left\|\sum_{r=k}^{l-1} \rho_r p^r\right\|^2 + 2\rho_l\left(p^l, \sum_{r=k}^{l-1} \rho_r p^r\right) + \rho_l^2 \|p^l\|^2.$$

Просуммируем эти равенства по l от $t+1$ до m':

$$\left\|\sum_{r=k}^{m'} \rho_r p^r\right\|^2 = \left\|\sum_{r=k}^{t} \rho_r p^r\right\|^2 + 2\sum_{l=t+1}^{m'} \rho_l\left(p^l, \sum_{r=k}^{l-1} \rho_r p^r\right) + \sum_{l=t+1}^{m'} \rho_l^2 \|p^l\|^2.$$

В сделанных предположениях справедливы оценки:

$$\gamma^2\sigma^2 \leqslant \Gamma^2\left(\sum_{r=k}^{t}\rho_r\right)^2 + \frac{\gamma^2}{2}\sum_{l=t+1}^{m'}\rho_l\sum_{r=k}^{l-1}\rho_r + \Gamma^2 \sup_{k\leqslant r\leqslant m'}\rho_r\sum_{l=t+1}^{m'}\rho_l \leqslant$$

$$\leqslant \Gamma^2\left(\frac{\gamma}{3\Gamma} + \frac{\gamma^2}{6\Gamma^2}\right)^2\sigma^2 + \frac{\gamma^2}{2}\sigma^2 + \Gamma^2\frac{\gamma^2\sigma}{6\Gamma^2}\sigma \leqslant \frac{11}{12}\gamma^2\sigma^2.$$

Полученное противоречие доказывает лемму.

Следующая лемма устанавливает локальное минимизирующее свойство метода обобщенного градиента: если запускать метод обобщенного градиента из нестационарной точки x с достаточно малым шагом, то метод находит меньшие, чем $f(x)$, значения функции f в малой окрестности точки старта x. Это свойство влечет выполнение условия 1) теоремы 8.9 и в основном обеспечивает сходимость метода.

Лемма 9.3. *Пусть последовательность начальных точек* $\{x^s\}_{s=0}^{\infty}$ *сходится к* $x = \lim\limits_{s\to\infty} x^s$. *Будем запускать метод обобщенного градиента* (9.2) *из начальных точек* x^s *с шагами* ρ_s^k $(k \geqslant k_s)$ *до некоторого момента* n_s. *Для каждого* s *получим последовательности* $\{x_s^k\}_{k\geqslant k_s}$:

$$x_s^{k_s} = x^s, \quad x_s^{k+1} = x_s^k - \rho_s^k g_s^k, \quad g_s^k \in G_f(x_s^k), \quad k = k_s, k_s+1, \ldots$$

Обозначим

$$\rho_s = \sup_{k\geqslant k_s}\rho_s^k, \qquad \sigma_s = \sum_{k=k_s}^{n_s-1}\rho_s^k.$$

Если $0 \notin G_f(x)$, $\sigma_s \geqslant \sigma > 0$ *и* $\lim\limits_{s\to\infty}\rho_s = 0$, *то существует такое* $\bar{\varepsilon} > 0$, *что для любого* $\varepsilon \in (0, \bar{\varepsilon}]$ *найдутся такие индексы* l_s, *что для достаточно больших* s $\|x_s^k - x\| \leqslant \varepsilon$ *при* $k \in [k_s, l_s]$ *и*

$$f(x) = \lim_{s\to\infty} f(x^s) > \overline{\lim_{s\to\infty}} f(x_s^{l_s}). \tag{9.4}$$

Доказательство. Обозначим

$$\gamma_1 = \rho(0, G_f(x)) = \inf\{\|g\| \mid g \in G_f(x)\}.$$

В силу полунепрерывности сверху G_f существует ε_1-окрестность точки x такая, что для всех $g \in G_f(y)$ $(\|y - x\| \leqslant \varepsilon_1)$ будет

$$\rho(g, G_f(x)) \leqslant \gamma_1/2.$$

Обозначим

$$\Gamma_1 = \sup\{\|g\| \mid g \in G_f(y),\ \|y - x\| \leqslant \varepsilon_1\}.$$

В силу обобщенной дифференцируемости $f(x)$ для $c_f = \gamma_1^2/(16\Gamma_1)$ существует такое $\varepsilon_2 \leqslant \varepsilon_1$, что при $\|y - x\| \leqslant \varepsilon_2$ будет

$$f(y) = f(x) + (g, y - x) + o(x, y, g), \tag{9.5}$$

где $g \in G_f(y)$ и $|o(x, y, g)| \leqslant c_f \| y - x \|$. Положим

$$\bar{\varepsilon} = \min\{\varepsilon_2, \sigma\gamma_1/2\}.$$

Зафиксируем произвольное $\varepsilon \leqslant \bar{\varepsilon}$. Пусть s таково, что $\| x^s - x \| \leqslant \varepsilon/4$ и $\rho_s \leqslant \varepsilon/(4\Gamma_1)$.

Покажем, что последовательности $\{x_s^k\}_{k \geqslant k_s}$ выходят из $\varepsilon/2$-окрестностей точек x^s. Предположим противное: $\| x_s^k - x^s \| \leqslant \varepsilon/2$ при всех $k \in [k_s, n_s]$. Тогда, с одной стороны,

$$\| x_s^{n_s} - x^s \| \leqslant \varepsilon/2 \leqslant \sigma\gamma_1/4.$$

С другой стороны, представим

$$\| x_s^{n_s} - x^s \| = \Big\| \sum_{k=k_s}^{n_s-1} \rho_s^k g_s^k \Big\| = \Big\| \sum_{k=k_s}^{n_s-1} \rho_s^k g_s^k \Big/ \sum_{k=k_s}^{n_s-1} \rho_s^k \Big\| \sum_{k=k_s}^{n_s-1} \rho_s^k.$$

При $k \in [k_s, n_s]$ имеет место

$$\| x_s^k - x \| \leqslant \| x_s^k - x^s \| + \| x^s - x \| \leqslant 3\varepsilon/4;$$

поэтому

$$\rho(g_s^k, G_f(x)) \leqslant \gamma_1/2.$$

Так как $G_f(x)$ выпукло и

$$\rho(0, G_f(x)) = \gamma_1,$$

то отсюда следует, что все g_s^k $(k \in [k_s, n_s])$ лежат в полупространстве, отстоящем от нуля на $\gamma_1/2$. То же самое имеет место и для выпуклых комбинаций векторов g_s^k $(k \in [k_s, n_s])$. В результате получаем

$$\| x_s^{n_s} - x^s \| \geqslant \sigma\gamma_1/2.$$

Полученное противоречие доказывает, что последовательности $\{x_s^k\}_{k \geqslant k_s}$ выходят из $\varepsilon/2$-окрестностей точек x^s.

Обозначим через m_s момент первого выхода последовательности $\{x_s^k\}_{k \geqslant k_s}$ из $\varepsilon/2$-окрестности x^s, т. е. для всех $k < m_s$

$$\| x_s^k - x^s \| \leqslant \varepsilon/2, \qquad \| x_s^{m_s} - x^s \| > \varepsilon/2.$$

Справедливо

$$\| x_s^{m_s} - x^s \| \leqslant \| x_s^{m_s-1} - x^s \| + \rho_s^{m_s-1} \| g_s^{m_s-1} \| \leqslant 3\varepsilon/4,$$

и для всех $k \in [k_s, m_s]$

$$\| x_s^k - x \| \leqslant \| x_s^k - x^s \| + \| x^s - x \| \leqslant \varepsilon.$$

Для $k \in [k_s, m_s]$ подставим x_s^k и g_s^k в равенство (9.5):

$$\begin{aligned} f(x_s^k) &= f(x) + (g_s^k, x_s^k - x) + o(x, x_s^k, g_s^k) \leqslant \\ &\leqslant f(x) + (g_s^k, x_s^k - x^s) + (g_s^k, x^s - x) + c_f \| x_s^k - x \| \leqslant \\ &\leqslant f(x) + (g_s^k, x_s^k - x^s) + c_f \| x_s^k - x^s \| + (\Gamma_1 + c_f) \| x^s - x \|. \quad (9.6) \end{aligned}$$

Рассмотрим скалярные произведения

$$(g_s^k, x_s^k - x^s) = -\left(g_s^k, \sum_{r=k_s}^{k-1} \rho_s^r g_s^r \Big/ \sum_{r=k_s}^{k-1} \rho_s^r\right) \sum_{r=k_s}^{k-1} \rho_s^r.$$

Для их оценки можно применить лемму 9.2, где нужно положить

$$P = \operatorname{co}\{G_f(y) \mid \|y - x\| \leqslant \varepsilon\};$$

$$p^r = g_s^r, \quad k_s \leqslant r \leqslant m_s; \quad \Gamma = \Gamma_1, \quad \gamma = \gamma_1/2.$$

Выполнены условия

$$\sum_{r=k_s}^{m_s-1} \rho_s^r \geqslant \frac{\|x_s^{m_s} - x^s\|}{\Gamma_1} \geqslant \frac{\varepsilon}{2\Gamma_1} > 0, \qquad \lim_{s\to\infty} \rho_s = 0.$$

Теперь, согласно лемме 9.2, для достаточно больших s найдутся такие индексы $l_s \leqslant m_s$, что

$$\left(g_s^{l_s}, \sum_{k=k_s}^{l_s-1} \rho_s^k g_s^k \Big/ \sum_{k=k_s}^{l_s-1} \rho_s^k\right) \geqslant \frac{\gamma_1^2}{8}, \qquad \sum_{k=k_s}^{l_s-1} \rho_s^k \geqslant \frac{\varepsilon\gamma_1}{12\Gamma_1^2}.$$

Перепишем неравенство (9.6) для $k = l_s$ с учетом полученных оценок:

$$f(x_s^{l_s}) \leqslant f(x) - \frac{\gamma_1^2}{8} \sum_{k=k_s}^{l_s-1} \rho_s^k + c_f \Gamma_1 \sum_{k=k_s}^{l_s-1} \rho_s^k + (\Gamma_1 + c_f)\|x^s - x\| \leqslant$$

$$\leqslant f(x) - \frac{\gamma_1^2}{192\Gamma_1^2} \gamma_1 \varepsilon + (\Gamma_1 + c_f)\|x^s - x\|. \qquad (9.7)$$

Итак, при достаточно малом ε для всех достаточно больших s мы указали такие индексы l_s, что $\|x_s^k - x\| \leqslant \varepsilon$ при $k \in [k_s, l_s]$ и справедливо (9.7). Для остальных s индексы l_s произвольны. Тогда для верхнего предела $\overline{\lim\limits_{s\to\infty}} f(x_s^{l_s})$ в силу (9.7) справедлива оценка (9.4). Лемма доказана.

Доказательство леммы 9.1. Нам нужно показать ограниченность в условиях леммы 9.1 последовательностей $\{x^k\}_{k=0}^{\infty}$, порожденных методом обобщенного градиента (9.2), (9.3) при достаточно малых $R = \sup\limits_{k\geqslant 0} \rho_k$.

Предположим противное: каково бы ни было $R_s \to 0$ $(s \to \infty)$, существует последовательность $\{x_s^k\}_{k=0}^{\infty}$, порожденная методом (9.2), (9.3) с $\sup\limits_{k\geqslant 0} \rho_s^k \leqslant R_s$, уходящая в бесконечность.

Выберем такое число d, что $f(x^0) < c < d < \infty$. Обозначим

$$\Gamma = \sup\{\|g\| \mid g \in G_f(y),\ f(y) \leqslant d\}.$$

Так как G_f компактнозначно и полунепрерывно сверху, то оно ограничено на компакте $\{x \mid f(x) \leqslant d\}$, т. е. $\Gamma < \infty$. Каждая числовая последовательность $\{f(x_s^k)\}_{k=0}^{\infty}$ $(s = 0, 1, \ldots)$ пересекает отрезок $[c, d]$ от c к d; поэтому найдутся части $\{x_s^{k_s}, \ldots, x_s^{n_s}\}$ $(s = 0, 1, \ldots)$ последовательностей $\{x_s^k\}_{k \geqslant 0}$ $(s = 0, 1, \ldots)$ такие, что либо

$$f(x_s^{k_s}) \leqslant c < f(x^k) < d \leqslant f(x_s^{n_s}), \qquad k \in (k_s, n_s), \tag{9.8}$$

либо

$$f(x_s^{k_s}) \leqslant c < d \leqslant f(x_s^{k_s+1}).$$

Последовательность $\{x_s^{k_s}\}_{s=0}^{\infty}$ принадлежит компактному множеству $\{x \mid f(x) \leqslant c\}$; поэтому она имеет предельные точки. Не меняя обозначений, будем считать, что

$$\lim_{s \to \infty} x_s^{k_s} = x'.$$

По построению $f(x') \leqslant c$. Покажем, что x' не принадлежит X^*. Действительно, предположим противное: $x' \in X^*$. Тогда

$$0 \leqslant \lim_{s \to \infty} \| x_s^{k_s+1} - x_s^{k_s} \| \leqslant \lim_{s \to \infty} \Gamma R_s = 0.$$

Отсюда вместе с (9.8) следует, что

$$\lim_{s \to \infty} f(x_s^{k_s}) = \lim_{s \to \infty} f(x_s^{k_s+1}) = f(x') = c \in F^*,$$

что противоречит выбору c. Итак, $x' \notin X^*$. Для достаточно малого ε в силу (9.8) и непрерывности $f(x)$ последовательности $\{x_s^k\}_{k=k_s}^{n_s} (s=0, 1, \ldots)$ выходят из ε-окрестности

$$x' = \lim_{s \to \infty} x_s^{k_s},$$

поэтому для достаточно больших s

$$\sum_{k=k_s}^{n_s-1} \rho_s^k \geqslant \frac{\| x_s^{n_s} - x_s^{k_s} \|}{\Gamma} \geqslant \frac{\varepsilon}{2\Gamma}.$$

Можно считать, что метод обобщенного градиента как бы многократно стартует из точек $x_s^{k_s} \to x' \notin X^*$. К последовательности начальных данных $\{x_s^{k_s}\}_{s=0}^{\infty}$, как теперь видно, можно применить лемму 9.3. Возьмем такое $\varepsilon < \bar{\varepsilon}(x')$, что

$$\max \{f(x) \mid \| x' - x \| \leqslant \varepsilon\} \leqslant d,$$

тогда индексы l_s из леммы 9.3 не больше n_s. Но это означает, что минимизирующее свойство (9.4) противоречит построению (9.8). Полученное противоречие доказывает лемму 9.1.

Метод обобщенного градиента в виде (9.2), (9.3) имеет главным образом теоретическое значение ввиду крайне медленной сходимости. Существует много путей его улучшения. Одна из возможностей состоит в нормировке псевдоградиентов в формулах (9.2).

Обозначим через $\bar{g}$ нормированный вектор g:

$$\bar{g} = \begin{cases} g/\|g\|, & g \neq 0, \\ 0, & g = 0. \end{cases}$$

В методе (9.2) вместо g^k можно подставить нормированные псевдоградиенты $\bar{g}^k$. Для модифицированного таким образом метода сохраняют силу теорема 9.1 и леммы 9.1, 9.3.

Лемма 9.4. *Для метода нормированного псевдоградиента остается справедливым локальное минимизирующее свойство, сформулированное в лемме* 9.3 (*в которой нужно заменить* g_s^k *на* $\bar{g}_s^k$).

Доказательство. Представим

$$x_s^{k+1} = x_s^k - \rho_s^k \bar{g}_s^k = x_s^k - \bar{\rho}_s^k g_s^k,$$

где

$$\bar{\rho}_s^k = \begin{cases} \rho_s^k/\|g_s^k\|, & g_s^k \neq 0, \\ \rho_s^k, & g_s^k = 0. \end{cases}$$

Покажем, что к последовательности $\{x_s^k\}_{k \geqslant k_s}$ в малой окрестности точки x применима лемма 9.3, откуда будет следовать лемма 9.4. Обозначим

$$\gamma_1 = \rho(0, G_f(x)) > 0.$$

В силу полунепрерывности сверху G_f существует такая ε_1-окрестность точки x, что для любого $g \in G_f(y)$ ($\|y - x\| \leqslant \varepsilon_1$) имеет место неравенство $\rho(g, G_f(x)) \leqslant \gamma_1/2$.

Обозначим по-прежнему

$$\Gamma_1 = \sup\{\|g\| \mid g \in G_f(y), \quad \|y - x\| \leqslant \varepsilon_1\}.$$

Введем индексы

$$\bar{t}_s = \sup\{t \mid \|x_s^k - x\| \leqslant \varepsilon_1 \ \forall k \in [k_s, t), \quad t \in [k_s, n_s]\}.$$

При $k \in [k_s, \bar{t}_s)$ справедливы оценки

$$\rho_s^k/\Gamma_1 \leqslant \bar{\rho}_s^k \leqslant 2\rho_s^k/\gamma_1.$$

Обозначим

$$\bar{\rho}_s = \sup_{k_s \leqslant k < \bar{t}_s} \bar{\rho}_s^k, \qquad \bar{\sigma}_s = \sum_{k=k_s}^{\bar{t}_s - 1} \bar{\rho}_s^k.$$

Очевидно, что

$$\bar{\rho}_s \leqslant 2\rho_s/\gamma_1 \to 0$$

при $s \to \infty$. Если $\bar{t}_s = n_s$, то $\bar{\sigma}_s \geqslant \sigma_s/\Gamma_1$. Если $\bar{t}_s < n_s$, то при достаточно больших s

$$\bar{\sigma}_s \geqslant \|x_s^{\bar{t}_s} - x^s\| / \Gamma_1 \geqslant \varepsilon_1/(2\Gamma_1).$$

Таким образом,

$$\bar{\sigma}_s \geqslant \min\{\varepsilon_1/2, \sigma\}/\Gamma_1 = \bar{\sigma} > 0.$$

Теперь очевидно, что к последовательностям $\{x_s^k\}_{k=s}^{\bar{t}_s}$ при достаточно больших s применима лемма 9.3, откуда и следует лемма 9.4.

Теорема 9.2. *Для алгоритма* (9.2), (9.3), *в котором векторы* g^k *заменены на нормированные* $\bar{g}^k$, *сохраняют силу результаты о сходимости в виде теоремы* 9.1 *и леммы* 9.1.

Аналогично теореме 9.1 теорема 9.2 в силу леммы 9.4 и

$$\lim_{k\to\infty} \| x^{k+1} - x^k \| \leqslant \lim_{k\to\infty} \rho_k = 0$$

следует из общей теоремы 8.9.

Замечание 9.3. Внимательный читатель может заметить, что множество псевдостационарных точек $X^* = \{x \mid 0 \in G_f(x)\}$, к которому сходится метод (9.2), может оказаться гораздо шире множества стационарных точек $\{x \mid 0 \in \partial f(x)\}$. Это обстоятельство обесценивает результаты о сходимости метода (9.2). Поэтому рассмотрим рандомизированный метод обобщенного градиента:

$$x^0 \in E_n, \qquad x^{k+1} = x^k - \rho_k g^k, \qquad g^k \in G_f(\tilde{x}^k), \tag{9.9}$$

$$\| x^k - \tilde{x}^k \| \leqslant \varepsilon_k, \qquad \lim_{k\to\infty} \varepsilon_k = 0, \tag{9.10}$$

где $\tilde{x}^k$ — случайный вектор с равномерным распределением в ε_k-окрестности точки x^k. При условиях (9.3), (9.10) последовательность $\{x^k\}$, согласно теореме 14.1, всегда сходится к X^*. Но в этом методе на каждой итерации с вероятностью единица $g^k \in \partial f(\tilde{x}^k)$, поскольку $G_f(x)$ в силу теоремы 1.12 почти всюду в E_n совпадает с $\partial f(x)$. Таким образом, в (9.9), (9.10) почти для всех траекторий $\{x^k\}$ используются только обобщенные градиенты Кларка функции $f(x)$. И так как отображение $\partial f(x)$ тоже является псевдоградиентным для $f(x)$, то в силу теоремы 14.1 метод (9.9), (9.10) почти для всех траекторий $\{x^k\}$ сходится к множеству $\{x \mid 0 \in \partial f(x)\}$. Остается заметить, что на практике вообще нет нужды прибегать к рандомизации выбора градиента, поскольку она и так происходит за счет ошибок округления.

§ 10. Метод обобщенного градиента в задаче условной оптимизации

В этом параграфе метод обобщенного градиента применяется для локальной минимизации обобщенно дифференцируемой функции при обобщенно дифференцируемых ограничениях-неравенствах и линейных ограничениях-равенствах.

Сначала рассматривается задача оптимизации только с одним ограничением-неравенством. Задача с несколькими ограничениями-неравенствами легко сводится к ней. Для ее решения применяется метод обобщенного градиента (типа [99]), использующий внутри допустимой области псевдоградиенты целевой функции, а вне ее псевдоградиенты ограничения.

Затем рассматривается задача оптимизации с несколькими ограничениями-неравенствами, упорядоченными по важности; она, вообще говоря, относится к многокритериальной оптимизации. Задача математического программирования с ограничениями-неравенствами попадает в этот класс, если ввести (произвольный) порядок среди ограничений. Для решения задачи применяется метод обобщенного градиента, последовательно минимизирующий и добивающийся выполнения убывающих по важности ограничений. При этом целевая функция считается наименее важным критерием и минимизируется в последнюю очередь.

Для решения задач оптимизации, содержащих линейные ограничения-равенства, предлагается использовать известные методы негладкой оптимизации, в том числе и рассмотренные в данном параграфе, в которых обобщенные градиенты заменены их проекциями на линейное подпространство, отвечающее линейным ограничениям.

Рассмотрим задачу математического программирования:

$$f_0(x) \to \min, \qquad x \in E_n, \tag{10.1}$$

при ограничениях

$$f_i(x) \leqslant 0, \qquad i = 1, 2, \ldots, m, \tag{10.2}$$

$f_i(x)$ — обобщенно дифференцируемые функции, $G_i(x)$ — их псевдоградиентные отображения.

Свернем все ограничения-неравенства задачи (10.1), (10.2) в одно. Это можно сделать многими способами. Например, ограничения (10.2) эквивалентны одному ограничению вида

$$h(x) \equiv \max_{1 \leqslant i \leqslant m} f_i(x) \leqslant 0, \tag{10.3}$$

или

$$h(x) \equiv \sum_{i=1}^{m} \max(0, f_i(x)) \leqslant 0. \tag{10.4}$$

Первая функция удобна тем, что при ее использовании в методах оптимизации необходимо вычислять лишь значения входящих в нее функций $f_i(x)$ и один градиент той функции, на которой максимум достигается.

Для второй функции нужно вычислять значения и градиенты всех входящих функций. Однако вторую функцию легче минимизировать, поскольку в ней минимизируется сразу несколько функций $f_i(x)$. Поэтому, если $f_i(x)$ просты и отмасштабированы, лучше использовать вторую функцию $h(x)$. Отметим, что на практике в случае ограничения (10.4) вместо нуля в правой части нужно поставить малое положительное число, чтобы правильно определять допустимость точек.

Естественно, свертывание исходных ограничений в одно негладкое ограничение вида (10.3) или (10.4) рассчитано на наихудший случай, т. е. предполагается, что $f_i(x)$ $(i = 1, 2, \ldots, m)$ являются достаточно общими обобщенно дифференцируемыми функциями. Если ограничения $f_i(x) \leqslant 0$ имеют специальный вид, то для решения задачи следует применять или разрабатывать специальные методы.

Напомним, что вычисление псевдоградиентов сложных функций (10.3), (10.4) следует проводить по правилам § 1.

1. Задачи с одним ограничением-неравенством. Так или иначе, мы будем рассматривать задачу вида

$$f(x) \to \min \tag{10.5}$$

при ограничениях

$$h(x) \leqslant 0, \qquad x \in E_n, \tag{10.6}$$

где $f(x)$ и $h(x)$ — обобщенно дифференцируемые функции.

Без существенного ограничения общности можем считать, что $h(x) \to \infty$ при $\|x\| \to +\infty$. Если это не так, то всегда вместо $h(x)$ можем взять функцию

$$H(x) = \max(h(x), \quad \|x\| - C),$$

где C — достаточно большая константа.

Определим многозначное отображение $x \to G(x)$:

$$G(x) = \begin{cases} G_f(x), & h(x) < 0, \\ \mathrm{co}\{G_f(x),\ G_h(x)\}, & h(x) = 0, \\ G_h(x), & h(x) > 0. \end{cases}$$

Согласно теореме 3.1, если x^* — точка локального минимума $f(x)$ в области $\{x \mid h(x) \leqslant 0\}$, то $0 \in G(x^*)$ и $h(x^*) \leqslant 0$.

Метод обобщенного градиента для решения задачи (10.5), (10.6) имеет следующий вид:

$$x^0 \in E_n, \quad x^{k+1} = x^k - \rho_k g^k, \quad g^k \in G(x^k), \quad k = 0, 1, \ldots, \tag{10.7}$$

$$0 \leqslant \rho_k \leqslant R, \quad \lim_{k\to\infty} \rho_k = 0, \quad \sum_{k=0}^{\infty} \rho_k = \infty. \tag{10.8}$$

Вне допустимой области метод (10.7) минимизирует функцию $h(x)$, а внутри — целевую функцию $f(x)$.

Вместо g^k в методе (10.7) можно использовать нормированные величины

$$\bar{g}^k = \begin{cases} g^k/\|g^k\|, & g^k \neq 0, \\ 0, & g^k = 0, \end{cases}$$

при этом все утверждения о сходимости метода сохраняют силу, как в задаче без ограничений.

Обозначим:

$D = \{x \mid h(x) \leqslant 0\}$ — допустимая область;

$X^* = \{x \mid x \in D,\ 0 \in G(x)\}$ — множество стационарных точек задачи;

$F^* = \{f(x) \mid x \in X^*\}$ — множество значений $f(x)$ на X^*;

$X_h^* = \{x \mid x \notin D,\ 0 \in G_h(x)\}$ — множество стационарных точек функции $h(x)$ вне области D;

$H^* = \{h(x) \mid x \in X_h^*\}$ — множество значений $h(x)$ на X_h^*.

Сформулируем результаты о сходимости метода. Следующая лемма устанавливает ограниченность последовательности $\{x^k\}_{k=0}^{\infty}$.

Лемма 10.1. *Пусть существует число $c > \max\{0, h(x^0)\}$, не принадлежащее H^*. Тогда для любого $d > c$ найдется такое $\overline{R}$, что для любой последовательности $\{\rho_k\}_{k=0}^{\infty}$, удовлетворяющей* (10.8) *с $R < \overline{R}$, последовательности $\{x^k\}_{k=0}^{\infty}$, запущенные из x^0 согласно* (10.7), *не выходят из ограниченной области $\{x \mid h(x) \leqslant d\}$.*

Поскольку вне допустимой области D алгоритм (10.7) минимизирует функцию $h(x)$, то лемма 10.1 аналогична лемме 9.1 и доказывается точно так же. Здесь также остаются в силе замечания 9.1, 9.2 относительно подбора величины R. В частности, алгоритм вида

$$x^{k+1} = \begin{cases} x^k - \rho_k g^k, & g^k \in G(x^k), \quad \text{если } h(x^k) < d, \\ x^0, & \text{если } h(x^k) \geqslant d \end{cases}$$

($d > \max\{0, h(x^0)\}$, $\{\rho_k\}_{k=0}^{\infty}$ удовлетворяют (10.8) с произвольным R и интервал $[\max\{0, h(x^0)\}, d)$ не лежит целиком в H^*) не более чем конечное число раз выходит из ограниченной области $\{x \mid h(x) \leqslant d\}$.

Теорема 10.1. *Предположим, что последовательность $\{x^k\}_{k=0}^{\infty}$ ограничена. Тогда:*

1) *минимальные по значению $h(x)$ предельные точки $\{x^k\}_{k=0}^{\infty}$ принадлежат множеству $D \cup X_h^*$, а интервал $[\varliminf_{k\to\infty} h(x^k),\ \varlimsup_{k\to\infty} h(x^k)]$ вложен в множество $H^* \cup \{h \in R^1 \mid h \leqslant 0\}$;*

2) *если существуют предельные точки $\{x^k\}_{k=0}^{\infty}$, принадлежащие D, то минимальные из них по значению $f(x)$ принадлежат X^*.*

Теорема 10.2. *Пусть $\{x^k\}_{k=0}^{\infty}$ ограничена, множества H^* и F^* не содержат интервалов. Тогда:*

1) *либо все предельные точки $\{x^k\}_{k=0}^{\infty}$ не принадлежат допустимой области D, а принадлежат X_h^*, и существует предел $\lim_{k\to\infty} h(x^k) > 0$;*

2) *либо все предельные точки $\{x^k\}_{k=0}^{\infty}$ принадлежат допустимой области D, и тогда все они принадлежат связному подмножеству X^*, а последовательность $\{f(x^k)\}_{k=0}^{\infty}$ имеет предел.*

Следствие 10.1. *Пусть последовательность $\{x^k\}_{k=0}^{\infty}$ ограничена, а функция $h(x)$ не имеет стационарных точек вне допустимой области, т. е.*

$$\{x \mid 0 \in G_h(x)\} \subset D.$$

Тогда все предельные точки $\{x^k\}_{k=0}^{\infty}$ принадлежат допустимой области, и, следовательно,

1) *минимальные из них по значению $f(x)$ принадлежат X^*, а*

$$[\varliminf_{k\to\infty} f(x^k),\ \varlimsup_{k\to\infty} f(x^k)] \subset F^*;$$

2) *если F^* не содержит интервалов, то все предельные точки $\{x^k\}_{k=0}^{\infty}$ принадлежат связному подмножеству X^* и существует предел* $\lim_{k\to\infty} f(x^k)$.

Теоремы 10.1, 10.2 и следствие 10.1 обобщают теорему 9.1 о сходимости метода обобщенного градиента на случай задачи минимизации с ограничениями. Прежде чем доказывать теоремы, докажем минимизирующее свойство метода обобщенного градиента при наличии ограничения.

Лемма 10.2. *Пусть последовательность начальных точек $\{x^s\}_{s=0}^{\infty}$ сходится к $x = \lim_{s\to\infty} x^s$. Будем запускать метод обобщенного градиента (10.7) из начальных точек x^s с шагами ρ_s^k ($k \geqslant k_s$) до некоторого момента n_s. Для каждого s получим последовательности $\{x_s^k\}_{k=k_s}^{n_s}$:*

$$x_s^{k_s} = x^s, \quad x_s^{k+1} = x_s^k - \rho_s^k g_s^k,$$

$$g_s^k \in G(x_s^k), \qquad k = k_s, \ k_s + 1, \ldots$$

Обозначим

$$\rho_s = \sup_{k \geqslant k_s} \rho_s^k, \qquad \sigma_s = \sum_{k=k_s}^{n_s - 1} \rho_s^k.$$

Если $0 \notin G(x)$, $\sigma_s \geqslant \sigma > 0$ и $\lim_{s\to\infty} \rho_s = 0$, то существует такое $\bar{\varepsilon} > 0$, что для любого $\varepsilon \in (0, \bar{\varepsilon}]$ найдутся такие индексы $l_s \geqslant k_s$, что $\| x_s^k - x \| \leqslant \varepsilon$ при $k \in [k_s, l_s]$ и

1) $h(x) = \lim_{s\to\infty} h(x^s) > \overline{\lim_{s\to\infty}} h(x_s^{l_s}), \qquad h(x) > 0;$

2) $f(x) = \lim_{s\to\infty} f(x^s) > \overline{\lim_{s\to\infty}} f(x_s^{l_s}), \quad h(x) < 0;$

3) $f(x) = \lim_{s\to\infty} f(x^s) > \overline{\lim_{s\to\infty}} f(x_s^{l_s}), \quad 0 \geqslant \overline{\lim_{s\to\infty}} h(x_s^{l_s}), \quad h(x) = 0.$

Доказательство. Утверждения 1), 2) леммы следуют при достаточно малом $\bar{\varepsilon}$ из леммы 9.3.

Рассмотрим случай $h(x) = 0$. Обозначим $\gamma_1 = \rho(0, G(x))$. Нетрудно проверить, что отображение G локально ограничено и замкнуто и, следовательно, полунепрерывно сверху. Поэтому существует такая ε_1-окрестность точки x, что для всех $g \in G(y)$ ($\|y - x\| \leqslant \varepsilon_1$) имеет место неравенство

$$\rho(g, G(x)) \leqslant \gamma_1/2.$$

Обозначим

$$\Gamma_1 = \sup \{\|g\| \mid g \in G(y), \ \|y - x\| \leqslant \varepsilon_1\} < \infty.$$

В силу обобщенной дифференцируемости $f(x)$ и $h(x)$ для любых констант $c_f, c_h > 0$ найдется такое $\varepsilon_2 \leqslant \varepsilon_1$, что справедливы соотношения

$$f(y) = f(x) + (g_f, y - x) + o_f(x, y, g_f),$$
$$h(y) = h(x) + (g_h, y - x) + o_h(x, y, g_h), \tag{10.9}$$

где $|o_f(x, y, g_f)| \leqslant c_f \|y - x\|$ и $|o_h(x, y, g_h)| \leqslant c_h \|y - x\|$ при $\|y - x\| \leqslant \varepsilon_2$ и $g_f \in G_f(y)$, $g_h \in G_h(y)$. Константы c_f и c_h будут уточнены позднее.

Положим

$$\bar{\varepsilon} = \min\{\varepsilon_2, \sigma\gamma_1/2\}.$$

Зафиксируем произвольное $\varepsilon \in (0, \bar{\varepsilon}]$. Пусть s таково, что $\|x^s - x\| \leqslant \varepsilon/4$ и $\rho_s \leqslant \varepsilon/(4\Gamma_1)$.

Покажем, что последовательности $\{x_s^k\}_{k=k_s}^{n_s}$ выходят из $\varepsilon/2$-окрестностей точек x^s. Предположим противное: $\|x_s^k - x^s\| \leqslant \varepsilon/2$ при всех $k \in [k_s, n_s]$. Тогда, с одной стороны, имеем

$$\|x_s^{n_s} - x^s\| \leqslant \varepsilon/2 \leqslant \sigma\gamma_1/4.$$

С другой стороны, представим

$$\|x_s^{n_s} - x^s\| = \left\| \sum_{k=k_s}^{n_s-1} \rho_s^k g_s^k \right\| = \left\| \sum_{k=k_s}^{n_s-1} \rho_s^k g_s^k \Big/ \sum_{k=k_s}^{n_s-1} \rho_s^k \right\| \sum_{k=k_s}^{n_s-1} \rho_s^k.$$

Здесь для всех $k \in [k_s, n_s]$

$$\|x_s^k - x\| \leqslant \|x_s^k - x^s\| + \|x^s - x\| \leqslant 3\varepsilon/4;$$

поэтому

$$\rho(g_s^k, G(x)) \leqslant \gamma_1/2.$$

Так как $G(x)$ выпукло и $\rho(0, G(x)) = \gamma_1$, то все g_s^k $(k \in [k_s, n_s])$ лежат в полупространстве, отстоящем от нуля на $\gamma_1/2$. То же самое имеет место и для выпуклых комбинаций векторов g_s^k $(k \in [k_s, n_s])$. Таким образом, с другой стороны, имеем

$$\|x_s^{n_s} - x^s\| \geqslant \frac{\gamma_1}{2} \sum_{k=k_s}^{n_s-1} \rho_s^k \geqslant \frac{\sigma\gamma_1}{2}.$$

Полученное противоречие доказывает, что последовательности $\{x_s^k\}_{k=k_s}^{n_s}$ выходят из $\varepsilon/2$-окрестностей точек x_s.

Обозначим через m_s момент первого такого выхода, т. е.

$$m_s = \max\{m \mid \|x_s^k - x^s\| \leqslant \varepsilon/2 \quad \forall k \in [k_s, m),\ m \in [k_s, n_s]\}.$$

Справедливо

$$\|x_s^{m_s} - x\| \leqslant \rho_s^{m_s-1} \|g_s^{m_s-1}\| + \|x_s^{m_s-1} - x^s\| + \|x^s - x\| \leqslant \varepsilon.$$

Представим

$$g_s^k = \lambda_{fs}^k g_{fs}^k + \lambda_{hs}^k g_{hs}^k, \qquad g_{fs}^k \in G_f(x_s^k), \qquad g_{hs}^k \in G_h(x_s^k), \qquad \lambda_{fs}^k + \lambda_{hs}^k = 1.$$

Для $k \in [k_s, m_s]$ подставим x_s^k и $g_{fs}^k \in G_f(x_s^k)$ в равенство (10.9):

$$f(x_s^k) = f(x) + (g_{fs}^k, x_s^k - x) + o_f(x, x_s^k, g_{fs}^k) \leqslant$$

$$\leqslant f(x) + (g_{fs}^k, x_s^k - x^s) + (g_{fs}^k, x^s - x) + c_f \| x_s^k - x \| \leqslant$$

$$\leqslant f(x) + (g_{fs}^k, x_s^k - x^s) + c_f \| x_s^k - x^s \| + (\Gamma_1 + c_f) \| x^s - x \|. \quad (10.10)$$

Аналогично

$$h(x_s^k) \leqslant h(x) + (g_{hs}^k, x_s^k - x^s) + c_h \| x_s^k - x^s \| + (\Gamma_1 + c_h) \| x^s - x \|. \quad (10.11)$$

Теперь рассмотрим скалярные произведения:

$$(g_s^k, x_s^k - x^s) = -\left(g_s^k, \sum_{r=k_s}^{k-1} \rho_s^r g_s^r \Big/ \sum_{r=k_s}^{k-1} \rho_s^r\right) \sum_{r=k_s}^{k-1} \rho_s^r,$$

где $g_s^r \in G(x_s^r)$ $(r \in [k_s, m_s])$. Для их оценки можно применить лемму 9.2, где нужно положить

$$P = \operatorname{co}\{G(y) \mid \| y - x \| \leqslant \varepsilon_1\};$$

$$p^r = g_s^r, \qquad r \in [k_s, m_s]; \qquad \Gamma = \Gamma_1; \qquad \gamma = \gamma_1/2.$$

Выполнены условия

$$\sum_{r=k_s}^{m_s-1} \rho_s^r \geqslant \frac{\| x_s^{m_s} - x^s \|}{\Gamma_1} \geqslant \frac{\varepsilon}{2\Gamma_1}, \qquad \lim_{s\to\infty} \rho_s = 0.$$

Теперь, согласно лемме 9.2, для достаточно больших s найдутся такие индексы $l_s \leqslant m_s$, что

$$\left(g_s^{l_s}, \sum_{k=k_s}^{l_s-1} \rho_s^k g_s^k \Big/ \sum_{k=k_s}^{l_s-1} \rho_s^k\right) \geqslant \frac{\gamma_1^2}{8}, \qquad \sum_{k=k_s}^{l_s-1} \rho_s^k \geqslant \frac{\varepsilon\gamma_1}{12\Gamma_1^2}. \quad (10.12)$$

Напомним, что мы рассматриваем случай $h(x) = 0$. Покажем, что при соответствующем выборе константы c_h при достаточно больших s будет $h(x_s^{l_s}) \leqslant 0$, что доказывает вторую часть утверждения 3) леммы. Действительно, предположим, что $h(x_s^{l_s}) > 0$. Тогда $g_s^{l_s} = g_{hs}^{l_s}$, и из (10.11) получаем

$$0 < h(x_s^{l_s}) \leqslant -\frac{\gamma_1^2}{8} \sum_{k=k_s}^{l_s-1} \rho_s^k + c_h \Gamma_1 \sum_{k=k_s}^{l_s-1} \rho_s^k + (\Gamma_1 + c_h) \| x^s - x \|.$$

Если $c_h \leqslant \gamma_1^2/(16\Gamma_1)$, то

$$0 \leqslant -\frac{\gamma_1^2}{192\Gamma_1^2}\gamma_1\varepsilon + (\Gamma_1 + c_h)\|x^s - x\|,$$

что невозможно при достаточно больших s. Это противоречие доказывает вторую часть утверждения 3) леммы.

Теперь имеем две возможности для достаточно больших s: $h(x_s^{l_s})<0$ и $h(x_s^{l_s}) = 0$. При $h(x_s^{l_s})<0$ имеет место равенство $g_s^{l_s} = g_{fs}^{l_s}$; поэтому неравенство (10.10) при $k = l_s$ можно переписать в виде

$$f(x_s^{l_s}) \leqslant f(x) - \frac{\gamma_1^2}{8}\sum_{k=k_s}^{l_s-1}\rho_s^k + c_f\Gamma_1\sum_{k=k_s}^{l_s-1}\rho_s^k + (\Gamma_1 + c_f)\|x^s - x\|.$$

Если $c_f \leqslant \gamma_1/(16\Gamma_1)$, то

$$f(x_s^{l_s}) \leqslant f(x) - \frac{\gamma_1^2}{192\Gamma_1^2}\gamma_1\varepsilon + (\Gamma_1 + c_f)\|x^s - x\|.$$

Отсюда для s таких, что $h(x_s^{l_s})<0$, следует справедливость утверждения 3) леммы.

Рассмотрим случай $h(x_s^{l_s}) = 0$. Из (10.11) при $k = l_s$ получаем

$$\left(g_{hs}^{l_s}, \sum_{k=k_s}^{l_s-1}\rho_s^k g_s^k\right) \leqslant c_h\left\|\sum_{k=s}^{l_s-1}\rho_s^k g_s^k\right\| + (\Gamma_1 + c_h)\|x^s - x\|.$$

Из оценки (10.12) имеем

$$\lambda_{fs}^{l_s}\left(g_{fs}^{l_s}, \sum_{k=k_s}^{l_s-1}\rho_s^k g_s^k\right) \geqslant \frac{\gamma_1^2}{8}\sum_{k=k_s}^{l_s-1}\rho_s^k - \lambda_{hs}^{l_s}\left(g_{hs}^{l_s}, \sum_{k=k_s}^{l_s-1}\rho_s^k g_s^k\right) \geqslant$$

$$\geqslant \left(\frac{\gamma_1^2}{8} - c_h\Gamma_1\right)\sum_{k=k_s}^{l_s-1}\rho_s^k - (\Gamma_1 + c_h)\|x^s - x\|. \quad (10.13)$$

Теперь подберем подходящее значение для константы c_h; все предыдущие рассуждения верны при $c_h = \gamma_1^2/(16\Gamma_1)$. При достаточно больших s правая и левая части (10.13) положительны, поэтому, учитывая, что $0 \leqslant \lambda_{fs}^{l_s} \leqslant 1$, получаем

$$\left(g_{fs}^{l_s}, \sum_{k=k_s}^{l_s-1}\rho_s^k g_s^k\right) \geqslant \frac{\gamma_1^2}{16}\sum_{k=k_s}^{l_s-1}\rho_s^k - (\Gamma_1 + c_h)\|x^s - x\|.$$

Подставляя эту оценку в (10.10), получаем для $k = l_s$ при достаточно больших s

$$f(x_s^{l_s}) \leqslant f(x) - \frac{\gamma_1^2}{16}\sum_{k=k_s}^{l_s-1}\rho_s^k + c_f\Gamma_1\sum_{k=k_s}^{l_s-1}\rho_s^k + (2\Gamma_1 + c_f + c_h)\|x^s - x\|.$$

Теперь подберем значение c_f (c_h было указано ранее). Все проведенные рассуждения верны при $c_f = \gamma_1^2/(32\Gamma_1)$. Тогда

$$f(x_s^{l_s}) \leqslant f(x) - \frac{\gamma_1^2}{384\Gamma_1^2}\gamma_1\varepsilon + (2\Gamma_1 + c_f + c_h)\|x^s - x\|.$$

Отсюда для s таких, что $h(x_s^{l_s}) = 0$, также следует первая часть утверждения леммы. При небольших s индексы l_s произвольны. Лемма доказана.

Имея для метода (10.7) минимизирующее свойство в виде леммы 10.2 и при ограниченных $\{x^k\}_{k=0}^{\infty}$ свойство $\lim\limits_{k\to\infty}\|x^{k+1} - x^k\| = 0$, можно вывести справедливость теорем сходимости 10.1, 10.2 из общих результатов о сходимости алгоритмов § 8.

Доказательство теоремы 10.1. Вне допустимой области D метод (10.7) минимизирует функцию $h(x)$, как в задаче без ограничений, при этом роль функции Ляпунова играет $h(x)$, роль множества решений — множество $D \cup X_h^*$. Функция $h(x)$ на $D \cup X_h^*$ пробегает значения из множества $\{h \in E_1 | h \leqslant 0\} \cup H^*$. Отметим, что H^* замкнуто. Поэтому первая часть теоремы 10.1 следует из теоремы 9.1 сходимости метода обобщенного градиента в задаче без ограничений, или, другими словами, из (10.8), утверждения 1) леммы 10.2 и общей теоремы 8.9.

Вторая часть теоремы 10.1 следует из условий 2), 3) леммы 10.2. Действительно, предположим противное. Пусть

$$x = \lim_{s\to\infty} x^{k_s}$$

есть одна из минимальных по значению $f(x)$ предельных точек $\{x^k\}_{k=0}^{\infty}$ в области D, не принадлежащая X^*. Можем считать, что алгоритм (10.7) как бы многократно стартует из точек x^{k_s} $(s = 0, 1, \ldots)$; при этом получаются последовательности $\{x^k\}_{k\geqslant k_s}$. К ним применима лемма 10.2.

Если x — внутренняя точка области D, то из утверждения 2) леммы 10.2 следует, что при достаточно малом ε существует еще меньшая по значению f внутренняя предельная точка $\{x^k\}_{k=0}^{\infty}$, что противоречит минимальности x.

Если x лежит на границе D, то в силу утверждения 3) леммы 10.2 опять существует меньшая по значению f допустимая предельная точка $\{x^k\}_{k=0}^{\infty}$, что опять противоречит минимальности x. Теорема 10.1 доказана.

Доказательство теоремы 10.2. В теореме 10.1 показано, что интервал $[\varliminf\limits_{k\to\infty} h(x^k),\ \varlimsup\limits_{k\to\infty} h(x^k)]$ вложен в множество

$$H^* \cup \{h \in E_1 | h \leqslant 0\}.$$

Отсюда следует, что либо интервал $[\varliminf_{k\to\infty} h(x^k), \varlimsup_{k\to\infty} h(x^k)]$ целиком вложен в множество $\{h \in E_1 \mid h \leqslant 0\}$, и тогда все предельные точки $\{x^k\}_{k=0}^{\infty}$ принадлежат D, либо существует предел

$$\lim_{k\to\infty} h(x^k) > 0,$$

и тогда предельные точки $\{x^k\}_{k=0}^{\infty}$ не принадлежат допустимой области D. В этом последнем случае все предельные точки $\{x^k\}$ принадлежат X_h^*. Действительно, если это не так, то существует предельная точка

$$\lim_{s\to\infty} x^{k_s} = x \notin X^*.$$

Можно считать, что алгоритм (10.7) как бы многократно стартует из точек x^{k_s} $(s = 0, 1, \ldots)$, при этом получаются последовательности $\{x^k\}_{k \geqslant k_s}$. К ним применима лемма 10.2. Минимизирующее свойство 1) этой леммы, в частности, отрицает возможность существования предела $\lim_{k\to\infty} h(x^k)$. Полученное противоречие доказывает первое утверждение теоремы 10.2.

Докажем второе утверждение теоремы. Как было показано выше, либо все предельные точки $\{x^k\}_{k=0}^{\infty}$ не принадлежат D, либо все они принадлежат D. Предположим, что все они принадлежат D. Будем рассматривать значения $f(x)$ на точках x^k независимо от того, допустимые они или недопустимые. Можем считать, что метод (10.7) минимизирует функцию $f(x)$ на D, причем имеет место минимизирующее свойство 3) леммы 10.2 и в силу ограниченности $\{x^k\}$ выполнено $\lim_{k\to\infty} \| x^{k+1} - x^k \| = 0$. Тогда утверждение 2) теоремы 10.2 следует из общей теоремы 8.9.

2. Задачи с несколькими ограничениями-неравенствами. Рассматривая метод (10.7) решения задачи (10.5), (10.6), мы видим, что он сначала как бы минимизирует функцию $h(x)$, а затем функцию $f(x)$ на множестве $\{x \mid h(x) \leqslant 0\}$. Ничто не мешает увеличить число последовательно минимизируемых функций. Таким образом, мы приходим к методу решения задачи математического программирования с несколькими ограничениями как задачи многокритериальной оптимизации.

Итак, рассмотрим задачу математического программирования:

$$f_{m+1}(x) \to \min, \qquad x \in E_n, \tag{10.14}$$

при ограничениях

$$f_j(x) \leqslant c_j, \qquad j = 1, 2, \ldots, m, \tag{10.15}$$

где будем считать, что ограничения $f_j(x) \leqslant c_j$ упорядочены (по важности).

Ограничение с меньшим индексом будем считать более важным, чем ограничение с бóльшим индексом, и, следовательно, нужно добиваться его выполнения в первую очередь, т. е. минимизировать соответствующую функцию $f_j(x)$ до уровня c_j. Лишь добившись выполнения всех ограничений, мы приступаем к минимизации целевой функции $f_{m+1}(x)$, с которой также можно связать некоторый уровень $c_{m+1} \geqslant -\infty$, т. е., по существу, не различать ограничения и целевую функцию.

При этом, если ограничений много, мы можем и не добиться выполнения менее важных ограничений, а процесс оптимизации остановится на каком-то ограничении i^* $(1 \leqslant i^* \leqslant m+1)$. Соответствующая точка x^* будет решением задачи

$$f_{i^*}(x) \to \min \tag{10.16}$$

при ограничениях

$$f_j(x) \leqslant c_j, \qquad 1 \leqslant j < i^*. \tag{10.17}$$

Таким образом, под решением, возможно, несовместной задачи (10.14), (10.15) с упорядоченными ограничениями естественно понимать такое оптимальное решение x^* совместной задачи математического программирования вида (10.16), (10.17), что

$$f_{i^*}(x^*) > c_{i^*}.$$

Если множество $\{x \mid f_1(x) \leqslant c_1\}$ пусто, то можно считать, что задача (10.14), (10.15) вообще не имеет решений.

Введение порядка на множестве ограничений есть один из способов рассмотрения несовместных задач математического программирования.

Для задач с упорядоченным множеством ограничений можно также ввести функцию полезности. В этом плане нам удобно обратиться к задачам максимизации.

Пусть необходимо максимизировать функцию $\varphi_1(x)$ на множестве $X \subset E_n$, чтобы превысить уровень c_1, затем необходимо максимизировать $\varphi_2(x)$ на множестве $\{x \in X \mid f_1(x) \geqslant c_1\}$, чтобы превысить уровень c_2, затем функцию $\varphi_3(x)$ на множестве $\{x \in X \mid \varphi_1(x) \geqslant C_1, \varphi_2(x) \geqslant C_2\}$, чтобы превысить уровень c_3, и т. д.

Заметим, что в ряду $\varphi_1(x), \ldots, \varphi_j(x), \ldots, \varphi_m(x)$ одна и та же функция может повторяться несколько раз под разными индексами j, но с бóльшими значениями уровней c_j, что будет отражать убывающую важность больших значений этой функции. Проблему упорядочивания целевых функций по важности мы здесь не обсуждаем.

Для рассматриваемой задачи построим функцию полезности $\Phi(\varphi)$, где $\varphi = (\varphi_1, \ldots, \varphi_m) \in E_m$. Эта функция должна быть тем больше, чем больше выполнено ограничений и чем больше значение функции в первом невыполненном ограничении. Мы предположим, что $\varphi_j(x) \geqslant 0$ $(j = 1, \ldots, m)$ при $x \in X$. Этими свойствами обладает следующая функция:

$$\Phi(\varphi) = \begin{cases} \varphi_1, & 0 \leqslant \varphi_1 < c_1, \\ \varphi_2 + c_1, & c_1 \leqslant \varphi_1,\ 0 \leqslant \varphi_2 < c_2, \\ \varphi_3 + c_1 + c_2, & c_1 \leqslant \varphi_1,\ c_2 \leqslant \varphi_2,\ 0 \leqslant \varphi_3 < c_3, \\ \dots\dots\dots\dots & \dots\dots\dots\dots \\ \varphi_m + c_1 + c_2 + \ldots + c_{m-1}, & c_1 \leqslant \varphi_1, \ldots, c_{m-1} \leqslant \varphi_{m-1}, \\ & 0 \leqslant \varphi_m < c_m, \\ c_1 + c_2 + \ldots + c_m, & c_1 \leqslant \varphi_1, \ldots, c_m \leqslant \varphi_m. \end{cases}$$

Если есть функция полезности, то задачу многокритериальной оптимизации можно решать «в лоб», подставив критерии $\varphi_j(x)$ в функцию полезности $\Phi(\varphi)$ и максимизируя получившуюся функцию по x. При этом все зависит от свойств функции $\Phi(\varphi)$ и от той информации, которой мы о ней располагаем.

В данном случае функция $\Phi(\varphi)$ разрывна, однако она является квазивогнутой, т. е. множества $\{\varphi \in E_m \mid \Phi(\varphi) \geqslant c\}$ выпуклы (они имеют вид $\{\varphi \in E_m \mid \varphi_1 \geqslant c_1, \ldots, \varphi_k \geqslant c_k, \varphi_{k+1} \geqslant c\}$, где $c_1 + c_2 + \ldots$ $\ldots + c_k \leqslant c < c_1 + c_2 + \ldots + c_{k+1}$).

Если мы подставим квазивогнутые функции $\varphi(x)$ в $\Phi(\varphi)$, то получим квазивогнутую функцию $\Phi(\varphi(x))$, так как выпукло множество

$$\{x \mid \Phi(\varphi(x)) \geqslant c\} = \{x \mid \varphi_1(x) \geqslant c_1, \ldots, \varphi_k(x) \geqslant c_k, \varphi_{k+1}(x) \geqslant c\},$$

где $c_1 + c_2 + \ldots + c_k \leqslant c < c_1 + c_2 + \ldots + c_{k+1}$.

Существующие методы максимизации квазивогнутых функций [75] обоснованы в предположении их непрерывности. Мы рассмотрим более общую ситуацию, когда функции $\varphi_j(x)$ обобщенно дифференцируемы, и построим метод минимизации функции $\Phi(\varphi(x))$.

Чтобы максимизировать Φ в точке $\varphi \in E_m$, нужно максимизировать первую функцию φ_j в ряду $\varphi_1, \varphi_2, \ldots, \varphi_m$, не достигшую своего уровня c_j. Применительно к задаче минимизации (10.14), (10.15) это означает, что нужно минимизировать функцию $f_i(x)$ первого из невыполненных ограничений. Поэтому рассмотрим следующий метод решения задачи (10.14), (10.15):

$$x^0 \in E_n, \quad x^{k+1} = x^k - \rho_k g^k, \quad g^k \in G_{i_k}(x^k), \quad k = 0, 1, \ldots,$$

$$i_k = \min\{j \mid 1 \leqslant j \leqslant m + 1, \ f_j(x^k) > c_j \quad (c_{m+1} = -\infty)\}.$$

Мы можем рассмотреть даже несколько более общий метод:

$$x^0 \in E_n, \quad x^{k+1} = x^k - \rho_k g^k, \quad g^k \in G(x^k), \tag{10.18}$$

$$0 \leqslant \rho_k \leqslant R, \quad \lim_{k\to\infty} \rho_k = 0, \quad \sum_{k=0}^{\infty} \rho_k = \infty, \tag{10.19}$$

где

$$G(x) = \operatorname{co}\{G_{i(x)}(x); \ G_j(x), \ j \in J(x)\}, \tag{10.20}$$

$$i(x) \equiv \min\{j \mid 1 \leqslant j \leqslant m + 1, f_j(x) \geqslant c_j \quad (c_{m+1} = -\infty)\}, \tag{10.21}$$

$$J(x) \equiv \{j \mid 1 \leqslant j < i(x), \ f_j(x) = c_j\}, \tag{10.22}$$

$G_j(x)$ — псевдоградиентное отображение функции $f_j(x)$ задачи (10.14), (10.15) ($j = 1, 2, \ldots, m + 1$).

Сформулируем аналог леммы 10.2 для этого метода. Сначала докажем полунепрерывность сверху отображения $G(x)$.

Лемма 10.3. *Функция $i(x)$ и отображения $x \to J(x)$, $x \to i(x) \cup$ $\cup\, J(x)$ полунепрерывны сверху.*

Доказательство. Согласно определению (10.21) индекса $i(x)$, имеет место $f_{i(x)}(x) > c_{i(x)}$. В силу непрерывности $f_{i(x)}$ неравенство $f_{i(x)}(y) > c_{i(x)}$ сохраняется для y из некоторой окрестности x. Здесь $i(y) \leqslant i(x)$ и функция $i(x)$ полунепрерывна сверху.

Если

$$f_j(x) \neq c_j,$$

то $f_j(y) \neq c_j$ для y из некоторой окрестности x. Отсюда следует, что $J(y) \subset J(x)$, т. е. отображение $J(x)$ полунепрерывно сверху (и замкнуто).

Докажем, что отображение $x \to i(x) \cup J(x)$ замкнуто; в данном случае это эквивалентно его полунепрерывности сверху. В силу уже доказанной замкнутости $J(x)$ достаточно показать, что если $x^k \to x$, $i(x^k) \to i$, то $i \in i(x) \cup J(x)$. Согласно сказанному выше, $i \leqslant i(x)$. Так как

$$f_{i(x^k)}(x^k) > c_{i(x^k)},$$

то для достаточно больших k

$$f_i(x^k) > c_i,$$

и в силу непрерывности $f_i(x)$ имеет место $f_i(x) \geqslant c_i$. Отсюда следует, что либо $i = i(x)$, либо $i \in J(x)$. Лемма доказана.

Лемма 10.4. *Отображение $x \to G(x)$, где $G(x)$ построено согласно* (10.20)—(10.22), *локально ограничено, замкнуто и, следовательно, полунепрерывно сверху.*

Доказательство. Напомним, что псевдоградиентные отображения $G_j(x)$ функций $f_j(x)$ локально ограничены и замкнуты ($j = 1, 2, \ldots, m+1$). Поэтому и $G(x)$ локально ограничено. Докажем замкнутость G. Пусть $x^k \to x$, $g^k \in G(x^k)$ и $g^k \to g$; нужно показать, что $g \in G(x)$.

Представим

$$g^k = \lambda^k_{i(x^k)} g^k_{i(x^k)} + \sum_{j \in J(x^k)} \lambda^k_j g^k_j, \quad \lambda^k_{i(x^k)}, \quad \lambda^k_j \geqslant 0,$$

$$\lambda^k_{i(x^k)} + \sum_{j \in J(x^k)} \lambda^k_j = 1, \quad g^k_{i(x^k)} \in G_{i(x^k)}(x^k), \quad g^k_j \in G_j(x^k).$$

Для достаточно больших k, согласно лемме 10.3, $J(x^k) \subset J(x)$; поэтому можно представить

$$g^k = \lambda^k_{i(x^k)} g^k_{i(x^k)} + \sum_{j \in J(x)} \lambda^k_j g^k_j,$$

где

$$\lambda^k_j = 0 \quad \text{и} \quad g^k_j \in G_j(x) \quad \text{при} \quad j \notin J(x^k).$$

В силу конечности набора функций $f_j(x)$ существует такая подпоследовательность индексов $\{k_s\}$, что $i(x^{k_s}) = i$, $J(x^{k_s}) \subset J(x)$ и

$$\lambda^{k_s}_{i(x^{k_s})} = \lambda^{k_s}_i \to \lambda_i \geqslant 0, \quad \lambda^{k_s}_j \to \lambda_j, \quad \lambda_i + \sum_{j \in J(x)} \lambda_j = 1,$$

$$g^{k_s}_{i(x^{k_s})} = g^{k_s}_i \to g_i, \quad g^{k_s}_j \to g_j, \quad j \in J(x).$$

В силу замкнутости отображений G_j $(j = 1, 2, \ldots, m+1)$ $g_i \in G_i(x)$, $g_j \in G_j(x)$ $(j \in J(x))$. Согласно лемме 10.3, либо $i = i(x)$, либо $i \in J(x)$; поэтому

$$g = \lim_{k \to \infty} g^k = \lambda_i g_i + \sum_{j \in J(x)} \lambda_j g_j \in G(x).$$

Лемма доказана.

Лемма 10.5. *Пусть последовательность начальных точек* $\{x^s\}_{s=0}^{\infty}$ *сходится к* $x = \lim\limits_{s \to \infty} x^s$. *Будем запускать метод* (10.18) *из начальных точек* x^s *в шагами* ρ^k_s $(k \geqslant k_s)$ *до некоторого момента* n_s. *Для каждого* s *получим последовательности* $\{x^k_s\}_{k=k_s}^{n_s}$:

$$x^{k_s}_s = x^s, \quad x^{k+1}_s = x^k_s - \rho^k_s g^k_s, \quad g^k_s \in G(x^k_s), \quad k = k_s, \ k_s + 1, \ldots$$

Обозначим

$$\rho_s = \sup_{k \geqslant s} \rho^k_s, \quad \sigma_s = \sum_{k=k_s}^{n_s - 1} \rho^k_s.$$

Если $0 \notin G(x)$, $\sigma_s \geqslant \sigma > 0$ *и* $\lim\limits_{s \to \infty} \rho_s = 0$, *то существует таков* $\bar{\varepsilon}(x) > 0$, *что для любого* $\varepsilon \in (0, \bar{\varepsilon}]$ *найдутся такие индексы* l_s, *что для достаточно больших* s $\|x^k_s - x\| \leqslant \varepsilon$ *при* $k \in [k_s, l_s)$ *и*

1) $f_{i(x)}(x) = \lim\limits_{s \to \infty} f_{i(x)}(x^s) > \overline{\lim\limits_{s \to \infty}} f_{i(x)}(x^{l_s}_s)$, *где* $i(x) = \min\{j \mid 1 \leqslant j \leqslant m+1$, $f_j(x) > c_j$ $(c_{m+1} = -\infty)\}$;

2) $c_j \geqslant \overline{\lim\limits_{s \to \infty}} f_j(x^{l_s}_s)$, $j \in J(x)$, *где* $J(x) = \{j \mid 1 \leqslant j < i(x),\ f_j(x) = c_j\}$.

Доказательство. Доказательство этой леммы в основном повторяет доказательство леммы 10.2. Обозначим

$$\gamma = \rho(0, G(x)) > 0.$$

В силу полунепрерывности сверху $G(x)$ (лемма 10.4) существует такая ε_1-окрестность точки x, что для всех $g \in G(y)$ $(\|y - x\| \leqslant \varepsilon_1)$ имеет место $\rho(g, G(x)) \leqslant \gamma/2$. Обозначим

$$\Gamma = \sup\{\|g\| \mid g \in G(y), \ \|y - x\| \leqslant \varepsilon_1\} < \infty.$$

В силу леммы 10.3 существует ε_2-окрестность точки x такая, что для всех $\{y \mid \|y - x\| \leqslant \varepsilon_2\}$ имеет место $J(y) \subset J(x)$, $i(y) \in i(x) \cup J(x)$.

В силу обобщенной дифференцируемости функций $f_j(x)$ для $a = \gamma^2/(32\Gamma)$ найдется такое $\varepsilon_3 \leqslant \varepsilon_2$, что

$$f_j(y) = f_j(x) + (g_j, y - x) + o_j(x, y, g_j), \tag{10.23}$$

где $|o_j(x, y, g_j)| \leqslant a\|y - x\|$ при $\|y - x\| \leqslant \varepsilon_3$ и $g_j \in G_j(y)$ $(j = 1, \ldots, m+1)$. Положим

$$\bar{\varepsilon} = \min\{\varepsilon_3, \sigma\gamma/2\}.$$

Зафиксируем произвольное $\varepsilon \in (0, \bar{\varepsilon}]$. Пусть s таково, что $\|x^s - x\| \leqslant \varepsilon/4$ и $\rho_s \leqslant \varepsilon/(4\Gamma)$.

Как и в лемме 10.2, последовательности $\{x_s^k\}_{k=k_s}^{n_s}$ выходят из $(\varepsilon/2)$-окрестностей точек x_s. Индексы m_s обозначают моменты первого такого выхода. Справедливо $\|x_s^k - x\| \leqslant \varepsilon$ при $k \in [k_s, m_s]$.

Для каждого s представим

$$g_s^k = \sum_{j \in i(x_s^k) \cup J(x_s^k)} \lambda_{sj}^k g_{sj}^k, \quad g_{sj}^k \in G_j(x_s^k),$$

$$\lambda_{sj}^k \geqslant 0, \qquad \sum_{j \in i(x_s^k) \cup J(x_s^k)} \lambda_{sj}^k = 1,$$

Для $k \in [k_s, m_s]$ подставим x_s^k и g_{sj}^k, где

$$j \in i(x_s^k) \cup J(x_s^k),$$

в соотношения (10.23):

$$f_j(x_s^k) \leqslant f_j(x) + (g_{sj}^k, x_s^k - x^s) + a\|x_s^k - x^s\| + (\Gamma + a)\|x^s - x\|. \tag{10.24}$$

Рассмотрим скалярные произведения

$$(g_s^k, x_s^k - x^s) = -\left(g_s^k, \sum_{r=k_s}^{k-1} \rho_s^r g_s^r \Big/ \sum_{r=k_s}^{k-1} \rho_s^r\right) \sum_{r=k_s}^{k-1} \rho_s^r.$$

Как и в доказательстве леммы 10.2, для достаточно больших s найдутся такие индексы $l_s \leqslant m_s$, что

$$\left(g_s^{l_s}, \sum_{k=k_s}^{l_s-1} \rho_s^k g_s^k \Big/ \sum_{k=k_s}^{l_s-1} \rho_s^k\right) \geqslant \frac{\gamma^2}{8}, \qquad \sum_{k=k_s}^{l_s-1} \rho_s^k \geqslant \frac{\varepsilon\gamma}{12\Gamma^2}. \tag{10.25}$$

Обозначим $i_s = i(x_s^{l_s})$, $J_s = J(x_s^{l_s})$. Имеем

$$g_s^{l_s} = \lambda_{si_s}^{l_s} g_{si_s}^{l_s} + \sum_{j \in J_s} \lambda_{sj}^{l_s} g_{sj}^{l_s}. \tag{10.26}$$

Для $j \in J_s \subset J(x)$ и $k = l_s$ из (10.24) получаем

$$c_j = f_j(x_s^{l_s}) \leqslant c_j + (g_{sj}^{l_s}, x_s^{l_s} - x^s) + a\Gamma \sum_{k=k_s}^{l_s-1} \rho_s^k + (\Gamma + a)\|x^s - x\|,$$

откуда

$$-(g_{sj}^{l_s}, x_s^{l_s} - x^s) \leqslant a\Gamma \sum_{k=k_s}^{l_s-1} \rho_s^k + (\Gamma + a) \| x^s - x \|. \qquad (10.27)$$

Из (10.25) — (10.27) получаем

$$\lambda_{si_s}^{l_s} \left(g_{si_s}^{l_s}, \sum_{k=k_s}^{l_s-1} \rho_s^k g_s^k \right) \geqslant \frac{\gamma^2}{8} \sum_{k=k_s}^{l_s-1} \rho_s^k - \sum_{j \in J_s} \lambda_{sj}^{l_s} \left(g_{sj}^{l_s}, \sum_{k=k_s}^{l_s-1} \rho_s^k g_s^k \right) \geqslant$$

$$\geqslant \frac{\gamma^2}{8} \sum_{k=k_s}^{l_s-1} \rho_s^k - \sum_{j \in J_s} \lambda_{sj}^{l_s} \left(a\Gamma \sum_{k=k_s}^{l_s-1} \rho_s^k + (\Gamma + a) \| x^s - x \| \right) \geqslant$$

$$\geqslant \frac{\gamma^2}{16} \sum_{k=k_s}^{l_s-1} \rho_s^k - (\Gamma + a) \| x^s - x \|. \qquad (10.28)$$

При достаточно больших s правая и левая части (10.28) положительны, $0 \leqslant \lambda_{si_s}^{l_s} \leqslant 1$; поэтому

$$\left(g_{si_s}^{l_s}, \sum_{k=k_s}^{l_s-1} \rho_s^k g_s^k \right) \geqslant \frac{\gamma^2}{16} \sum_{k=k_s}^{l_s-1} \rho_s^k - (\Gamma + a) \| x^s - x \|.$$

Эта оценка, очевидно, выполнена и когда $J_s = \varnothing$. Подставляя ее в (10.24) для $k = l_s$ и $j = i_s$, получим

$$f_{i_s}(x_s^{l_s}) \leqslant f_{i_s}(x) - \frac{\gamma^2}{16} \sum_{k=k_s}^{l_s-1} \rho_s^k + a\Gamma \sum_{k=k_s}^{l_s-1} \rho_s^k + 2(\Gamma + a) \| x^s - x \| \leqslant$$

$$\leqslant f_{i_s}(x) - \frac{\gamma^2}{32} \sum_{k=k_s}^{l_s-1} \rho_s^k + 2(\Gamma + a) \| x^s - x \| \leqslant$$

$$\leqslant f_{i_s}(x) - \frac{\gamma^2}{384\Gamma^2} \gamma\varepsilon + 2(\Gamma + a) \| x^s - x \|. \qquad (10.29)$$

Теперь покажем, что при достаточно больших s

$$f_j(x_s^{l_s}) \leqslant c_j, \qquad j \in J(x),$$

откуда будет следовать второе утверждение леммы. Действительно, предположим противное: при сколь угодно больших s найдутся такие $j \in J(x)$, что $f_j(x_s^{l_s}) > c_j$. Определим индексы

$$j_s = \min \{ j \mid j \in J(x),\ f_j(x_s^{l_s}) > c_j \}.$$

Очевидно, что

$$i_s = i(x_s^{l_s}) \leqslant j_s.$$

Согласно выбору $\varepsilon \leqslant \varepsilon_2$, имеет место

$$i_s \in i(x) \cup J(x).$$

Вместе с предыдущим неравенством это дает $i_s \in J(x)$. Отсюда с учетом того, что

$$f_{i_s}(x_s^{l_s}) > c_{i_s},$$

следует противоположное неравенство: $i_s \geqslant j_s$. Итак, мы показали, что $i_s = j_s$. В этом случае одновременно выполнены $c_{i_s} < f_{i_s}(x_s^{l_s})$ и $f_{i_s}(x) = c_{i_s}$. Подставляя эти оценки в (10.29), получаем противоречие для достаточно больших s:

$$c_{i_s} \leqslant c_{i_s} - \frac{\gamma^2}{384\Gamma^2}\gamma\varepsilon + (\Gamma + a)\|x^s - x\|.$$

Значит, при достаточно больших s

$$f_j(x_s^{l_s}) \leqslant c_j \quad \text{для} \quad j \in J(x).$$

Второе утверждение леммы доказано.

Докажем первое утверждение леммы. Имеем

$$i_s = i(x_s^{l_s}) \in i(x) \cup J(x).$$

Уже показано, что для достаточно больших s

$$f_j(x_s^{l_s}) \leqslant c_j \quad \text{для} \quad j \in J(x).$$

Так как

$$f_{i_s}(x_s^{l_s}) > c_{i_s},$$

то для достаточно больших s

$$i_s = i(x_s^{l_s}) = i(x).$$

Тогда первое утверждение леммы следует из неравенства (10.29). Лемма доказана.

Следствие 10.2. *Чтобы точка x была точкой локального минимума задачи* (10.14), (10.15), *необходимо выполнение следующих условий:*

$$0 \in G(x), \quad f_j(x) \leqslant c_j, \quad j = 1, 2, \ldots, m.$$

Действительно, условие $f_j(x) \leqslant c_j$ ($j = 1, 2, \ldots, m$) является условием допустимости. Предположим, что x — точка локального минимума, но $0 \notin G(x)$. Будем многократно ($s = 0, 1, \ldots$) запускать алгоритм (10.18) из начальной точки $x_s \equiv x$ с шагами ρ_s^k ($k \geqslant k_s$) до некоторого момента n_s так, что

$$\lim_{s\to\infty} \rho_s = \lim_{s\to\infty} \sup_{k_s \leqslant k \leqslant n_s} \rho_s^k = 0, \quad \sum_{k=k_s}^{n_s-1} \rho_s^k \geqslant \sigma > 0.$$

Тогда, согласно лемме 10.5, в сколь угодно малой окрестности точки x алгоритм найдет точки с меньшим, чем $f_{m+1}(x)$, значением функции f_{m+1}, что противоречит локальной оптимальности x. Таким образом, в минимизирующем свойстве леммы 10.5 содержатся необходимые условия экстремума. Этот момент уже отмечался в замечании к теореме 8.9.

Так как на первом этапе алгоритм (10.18) минимизирует функцию $h(x) = f_1(x)$, то ограниченность последовательностей, порожденных алгоритмом, можно обеспечить так же, как это сделано для метода (10.7) в лемме 10.1. Обозначим

$$X_i^* = \{x \in E_n \mid 0 \in G(x);\ \ f_j(x) \leqslant c_j,\ \ 1 \leqslant j \leqslant i;\ \ f_i(x) > c_i\},$$
$$F_i^* = \{f_i(x) \mid x \in X_i^*\}.$$

Теорема 10.3. *Предположим, что последовательность* $\{x^k\}_{k=0}^{\infty}$, *порожденная алгоритмом* (10.18) — (10.22), *ограничена, а множества* F_i^* $(i = 1, 2, \ldots, m+1)$ *не содержат отрезков. Тогда последовательность* $\{x^k\}_{k=0}^{\infty}$ *сходится к решению одной из задач* (10.16), (10.17), *где* $1 \leqslant i^* \leqslant m+1$, *т. е. все предельные точки* $\{x^k\}_{k=0}^{\infty}$ *являются допустимыми для задачи* (10.16), (10.17) *и принадлежат множеству* $X_{i^*}^*$, *а также существует предел* $\lim\limits_{k\to\infty} f_{i^*}(x^k) > c_{i^*}$.

Доказательство. На первом этапе алгоритм минимизирует функцию $f_1(x)$. При этом роль функции Ляпунова играет сама функция $f_1(x)$, роль множества решений играет множество

$$X_1^{**} \equiv X_1^* \cup \{x \in E_n \mid f_1(x) \leqslant c_1\}.$$

В силу

$$\lim_{k\to\infty} \| x^{k+1} - x^k \| = 0,$$

леммы 10.5 (с $l(x) = 1$) и общей теоремы 8.9 все предельные точки $\{x^k\}_{k=0}^{\infty}$ принадлежат X_1^{**} и полуинтервал $[\varliminf\limits_{k\to\infty} f(x^k), \varlimsup\limits_{k\to\infty} f(x^k))$ вложен в множество

$$F_1^* \cup \{f \in E_1 \mid f \leqslant c_1\}.$$

Так как F_1^* не содержит отрезков, то либо существует предел $\lim\limits_{k\to\infty} f_1(x^k) > c_1$ и все предельные точки $\{x^k\}_{k=0}^{\infty}$ принадлежат X_1^*, либо все предельные точки $\{x^k\}_{k=0}^{\infty}$ принадлежат множеству $\{x \in E_n \mid f_1(x) \leqslant c_1\}$.

Если имеет место последний случай, то можем считать, что алгоритм минимизирует функцию $f_2(x)$ при ограничении $f_1(x) \leqslant c_1$. Теперь в качестве функции Ляпунова выступает $f_2(x)$, а в качестве множества решений — множество

$$X_2^{**} = X_2^* \cup \{x \in E_n \mid f_1(x) \leqslant c_1,\ f_2(x) \leqslant c_2\}.$$

В силу

$$\lim_{k\to\infty} \| x^{k+1} - x^k \| = 0,$$

леммы 10.5 (с $l(x) = 2$) и общей теоремы 8.9 последовательность $\{x^k\}_{k=0}^{\infty}$ сходится к указанному множеству решений, и полуинтервал $[\underline{\lim}_{k\to\infty} f_2(x^k), \overline{\lim}_{k\to\infty} f_2(x^k))$ вложен в множество

$$F_2^* \cup \{f \in E_1 \mid f \leqslant c_2\}.$$

Так как F_2^* не содержит интервалов, то либо существует предел $\lim\limits_{k\to\infty} f_2(x^k) > c_2$ и все предельные точки $\{x^k\}_{k=0}^{\infty}$ принадлежат X_2^*, либо $\overline{\lim\limits_{k\to\infty}} f_2(x^k) \leqslant c_2$ и все предельные точки $\{x^k\}_{k=0}^{\infty}$ принадлежат множеству $\{x \in E_n \mid f_1(x) \leqslant c_1,\ f_2(x) \leqslant c_2\}$ и т. д. Теорема доказана.

Следствие 10.3. *Если все стационарные точки отображения G лежат в допустимой области исходной задачи* (10.14), (10.15):

$$\{x \in E_n \mid 0 \in G(x)\} \subset \{x \in E_n \mid f_j(x) \leqslant c_j,\ 1 \leqslant j \leqslant m\},$$

то в условиях теоремы 10.3 *последовательность* $\{x^k\}_{k=0}^{\infty}$ *сходится к решению*

$$X^* = \{x \in E_n \mid 0 \in G(x),\ f_j(x) \leqslant c_j,\ 1 \leqslant j \leqslant m\}$$

исходной задачи и существует предел $\lim\limits_{k\to\infty} f_{m+1}(x^k)$.

3. **Задачи с линейными ограничениями.** До сих пор мы рассматривали задачи математического программирования с ограничениями-неравенствами. Теперь рассмотрим задачи математического программирования с общими ограничениями-неравенствами и линейными ограничениями-равенствами. Пусть необходимо решить задачу

$$f(x) \to \min \tag{10.30}$$

при ограничениях

$$f_j(x) \leqslant c_j, \qquad j = 1, 2, \ldots, m, \tag{10.31}$$

$$Ax = b, \qquad e \leqslant x \leqslant d, \tag{10.32}$$

где $x = (x_1, \ldots, x_n) \in E_n$, $f(x)$ и $f_j(x)$ ($j = 1, \ldots, m$) — обобщенно-дифференцируемые функции, A — матрица размера $m_1 \times n$, $b \in E_{m_1}$, e и d — n-мерные векторы, быть может содержащие неограниченные компоненты.

Существует много возможностей для построения алгоритмов решения этой задачи; например, можно обобщать методы линеаризации или приведенного градиента. Однако в контексте проделанной работы проще всего, по-видимому, воспользоваться методом проекции обобщенного градиента на линейное подпространство, заданное ограничениями $Ax = 0$.

Пусть заданы функция $f(x)$ ($x \in E_n$) и линейное многообразие

$$L = \{x \in E_n \mid Ax = b\}.$$

Обозначим через

$$L_0 = \{x \in E_n \mid Ax = 0\}$$

подпространство, связанное с L. Можно рассматривать сужение функции $f(x)$ на L. Если $f(x)$ выпукла в E_n, то она останется выпуклой и на выпуклом подмножестве E_n, в частности на L. Можно рассматривать субградиенты функции $f(x)$, рассматриваемой как функция на L. Покажем, как вычислить эти субградиенты через обычные субградиенты $f(x)$ $(x \in E_n)$. Имеем субградиентное неравенство

$$f(y) \geqslant f(x) + (g, y - x), \qquad x, y \in E_n, \qquad g \in \partial f(x).$$

Разложим вектор g по подпространству L_0 и ортогональному дополнению к L_0, т. е. представим $g = g_0 + g'$, где g_0 — проекция g на L_0. Когда $x, y \in L$, имеет место $(g, y - x) = (g_0, y - x)$, и, таким образом,

$$f(y) \geqslant f(x) + (g_0, y - x) \quad \text{для всех} \quad x, y \in L,$$

т. е. g_0 — субградиент f как функции на L. Оператор проектирования пространства E_n на подпространство L_0, как известно [120], имеет вид

$$P = I - A(A^T A)^{-1} A^T,$$

где I — единичная матрица размера $n \times n$, A^T — транспонированная матрица A.

Для минимизации $f(x)$ на L теперь можно применять известные методы выпуклой оптимизации (метод обобщенного градиента, метод эллипсоидов, методы с растяжением пространства), в которых следует использовать субградиенты $f(x)$ как функции на L. Если $f(x)$ должна минимизироваться на L при дополнительных выпуклых ограничениях-неравенствах, то функции в этих ограничениях также должны рассматриваться как выпуклые на L с соответствующими субградиентами. То же самое относится к ограничениям вида $e \leqslant x \leqslant d$. Их можно заменить и одним ограничением вида

$$h(x) \equiv \max_{1 \leqslant i \leqslant n} \max\{e_i - x_i, x_i - d_i\} \leqslant 0,$$

или

$$h(x) \equiv \sum_{i=1}^{n} (\max\{0, e_i - x_i\} + \max\{0, x_i - d_i\}) \leqslant 0.$$

Проектирование субградиентов уменьшает размерность рабочего пространства, и тем самым уменьшается трудоемкость методов выпуклой оптимизации, поскольку оценки трудоемкости монотонно зависят от размерности [73].

Сказанное выше о выпуклых функциях целиком переносится на обобщенно дифференцируемые функции. Пусть $f(x)$ — обобщенно дифференцируемая функция, $G_f(x)$ — ее псевдоградиентное отображение. Рассмотрим многозначное отображение

$$PG_f(x) = \{g_0 \in L_0 \mid g_0 = Pg, \ g \in G_f(x)\},$$

где P — оператор проектирования пространства E_n на L_0. Множества $PG_f(x)$, как и $G_f(x)$, непусты, выпуклы, замкнуты и ограничены, а отображение $x \to PG_f(x)$ полунепрерывно сверху. Пусть $x, y \in L$, $g \in G(y)$ и $g = Pg + g'$; тогда

$$o(x, y, Pg) = f(y) - f(x) - (Pg, y - x) =$$
$$= f(y) - f(x) - (g, y - x) = o(x, y, g)$$

удовлетворяет требуемым условиям малости (1.3). Таким образом, обобщенно дифференцируемая в E_n функция $f(x)$ является и обобщенно дифференцируемой как функция в L с псевдоградиентным отображением $PG_f(x)$.

Теперь мы можем рассматривать сужение исходной задачи (10.30)—(10.32) на множество L и для ее решения применять рассмотренные ранее методы обобщенного градиента (10.7) и ведущего ограничения (10.18)—(10.22), в которых вместо псевдоградиентов функций в исходном пространстве E_n фигурируют их проекции на подпространство L_0. Начальную точку для методов нужно взять из L. Ограничения $e \leqslant x \leqslant d$ следует учитывать наравне с остальными ограничениями-неравенствами; в частности, их можно свертывать в одно или ранжировать по важности.

§ 11. Построение релаксационных методов невыпуклой негладкой оптимизации

В этом параграфе излагается схема построения релаксационных алгоритмов локальной минимизации обобщенно дифференцируемых функций.

В релаксационных методах негладкой оптимизации приходится решать две взаимосвязанные проблемы: строить направление убывания функции и подбирать шаг в этом направлении. Построим релаксационный метод, в котором ключевая проблема поиска направления спуска сведена к чисто геометрической задаче, что позволяет говорить о ее сложности и поиске оптимального алгоритма решения. Кроме того, метод предоставляет широкие возможности для оптимизации выбора шага. И наконец, предлагаемый метод останавливается через конечное число итераций по достижении заданной точности решения. Если допустимое отклонение устремлять к нулю, то полученное приближенное решение будет стремиться к точному. Приводятся также некоторые оценки трудоемкости метода на классе выпуклых задач.

Пусть требуется решить задачу

$$f(x) \to \min, \qquad x \in E_n; \tag{11.1}$$

$f(x)$ — обобщенно дифференцируемая функция; $G_f(x)$ — ее псевдоградиентное отображение.

Согласно теореме 8.9, все точки локального минимума $f(x)$ принадлежат множеству (решений)

$$X^* = \{x \in E_n \mid 0 \in G_f(x)\}.$$

Множество X^* будем считать точным решением задачи (11.1). Наряду с ним рассмотрим некоторые приближенные решения $X^*_{\varepsilon\delta}$.

Введем обозначения·

$$\Delta(C, D) = \sup_{c \in C} \inf_{d \in D} \| c - d \|$$

— уклонение множества C от множества D;

$$\gamma_\delta(x) = \rho\left(0, \mathrm{co} \bigcup_{\{y \mid \|y - x\| \leqslant \delta\}} G_f(y)\right);$$

$$X^*_{\varepsilon\delta} = \{x \in E_n \mid \gamma_\delta(x) \leqslant \varepsilon\}.$$

Лемма 11.1. *Функция $\delta \to \gamma_\delta(x)$ монотонно по δ убывает, а функция $(\delta, x) \to \gamma_\delta(x)$ полунепрерывна снизу, т. е. при $(\delta_k, x^k) \to (\delta, x)$ и $\gamma_{\delta_k}(x^k) \to \gamma$ имеет место неравенство*

$$\gamma \geqslant \gamma_\delta(x).$$

Доказательство. Если δ возрастает, то множество $\mathrm{co}\{G_f(y) \mid \|y - x\| \leqslant \delta\}$ расширяется; следовательно, $\gamma_\delta(x)$ убывает. Покажем, что функция $(\delta, x) \to \gamma_\delta(x)$ полунепрерывна снизу. Напомним, что

$$\gamma_{\delta_k}(x) = \inf\left\{\|g\| \mid g \in \mathrm{co} \bigcup_{\{y \mid \|y - x\| \leqslant \delta_k\}} G_f(y)\right\}.$$

Поэтому для числовой последовательности $\varepsilon_k \to 0$ существуют такие векторы g_i^k и числа λ_i^k ($i = 1, 2, \dots, n + 1$; n — размерность пространства), что

$$\|g^k\| \leqslant \gamma_{\delta_k}(x^k) + \varepsilon_k, \quad g^k = \sum_{i=1}^{n+1} \lambda_i^k g_i^k,$$

$$g_i^k \in G_f(y_i^k), \quad \|y_i^k - x^k\| \leqslant \delta_k, \quad \lambda_i^k \geqslant 0, \quad \sum_{i=1}^{n+1} \lambda_i^k = 1.$$

В силу локальной ограниченности G_f существует такая подпоследовательность индексов $\{k_s\}_{s=0}^{\infty}$, что

$$\lim_{s \to \infty} g_i^{k_s} = g_i, \quad \lim_{s \to \infty} y_i^{k_s} = y_i, \quad \lim_{s \to \infty} \lambda_i^{k_s} = \lambda_i.$$

Очевидно, что

$$\|y_i - y\| \leqslant \delta, \quad \lambda_i \geqslant 0, \quad \sum_{i=1}^{n+1} \lambda_i = 1,$$

и в силу замкнутости G_f $g_i \in G_f(y_i)$. Введем обозначения:

$$\delta'_s = \delta_{k_s}, \quad g = \sum_{i=1}^{n+1} \lambda_i g_i \in \mathrm{co} \bigcup_{\{y \mid \|y - x\| \leqslant \delta\}} G_f(y).$$

Переходя в неравенстве

$$\|g^{k_s}\| \leqslant \gamma_{\delta'_s}(x^{k_s}) + \varepsilon_{k_s}$$

к пределу по s, получаем

$$\|g\| \leqslant \lim_{s\to\infty} \gamma_{\delta'_s}(x^{k_s}) = \gamma.$$

Но $\gamma_\delta(x) \leqslant \|g\|$. Таким образом,

$$\gamma \geqslant \|g\| \geqslant \gamma_\delta(x),$$

что и требовалось доказать.

Лемма 11.2. *Многозначное отображение* $(\varepsilon, \delta) \to X^*_{\varepsilon\delta}$ *монотонно возрастает, т. е.* $X^*_{\varepsilon_1\delta_1} \subset X^*_{\varepsilon_2\delta_2}$ *при* $\varepsilon_1 \leqslant \varepsilon_2$, $\delta_1 \leqslant \delta_2$, *и замкнуто, т. е. при* $(\varepsilon_k, \delta_k) \to (\varepsilon, \delta)$, $x^k \in X^*_{\varepsilon_k\delta_k}$ *и* $\{x^k\} \to x$ *имеем* $x \in X^*_{\varepsilon\delta}$.

Доказательство. Пусть $\varepsilon_1 \leqslant \varepsilon_2$, $\delta_1 \leqslant \delta_2$, $\gamma_{\delta_1}(x) \leqslant \varepsilon_1$. В силу монотонного убывания $\gamma_\delta(x)$ по δ имеем

$$\gamma_{\delta_2}(x) \leqslant \gamma_{\delta_1}(x) \leqslant \varepsilon_1 \leqslant \varepsilon_2;$$

тогда $x \in X^*_{\varepsilon_2\delta_2}$, что и требовалось доказать.

Пусть

$$(\varepsilon_k, \delta_k) \to (\varepsilon, \delta), \qquad x^k \in X^*_{\varepsilon_k,\delta_k}, \qquad \{x^k\} \to x.$$

Тогда

$$\gamma_{\delta_k}(x^k) \leqslant \varepsilon_k.$$

Переходя к пределу по k, в силу полунепрерывности снизу $\gamma_\delta(x)$ по (δ, x) получаем

$$\gamma_\delta(x) \leqslant \lim_{k\to\infty} \gamma_{\delta_k}(x^k) \leqslant \varepsilon,$$

т. е. $x \in X^*_{\varepsilon\delta}$, что и требовалось доказать.

Следствие 11.1. *Множество X^* вместе со своей δ-окрестностью вложено в $X^*_{\varepsilon\delta}$. Пусть D—произвольное замкнутое ограниченное множество в E_n. Многозначное отображение $(\varepsilon, \delta) \to X^*_{\varepsilon\delta} \cap D$ ограничено и замкнуто и, следовательно, полунепрерывно сверху. Поэтому при $X^* \cap D \neq \varnothing$ имеем $X^*_{\varepsilon\delta} \cap D \neq \varnothing$, и*

$$\Delta(X^*_{\varepsilon\delta} \cap D, X^* \cap D) \to 0 \quad \textit{для} \quad (\varepsilon, \delta) \to (0, 0),$$

т. е. множество $X^*_{\varepsilon\delta} \cap D$ является приближенным решением задачи (11.1).

Приводимый ниже алгоритм A1 строит последовательность приближений $\{x^k\}$ к множеству $X^*_{\varepsilon\delta} \cap \{x \in E_n \mid f(x) \leqslant f(x^0)\}$, причем он сходится к этому множеству за конечное число итераций. Этот алгоритм допускает определенную свободу действий на некоторых шагах, т. е., по существу, содержит целый класс алгоритмов.

Алгоритм А1.

Шаг 0. Выбрать начальное приближение $x^0 \in E_n$, точности $\varepsilon > 0$, $\delta > 0$ и параметры $\Lambda \geqslant \varepsilon$, $\Delta \geqslant \delta$, α, β, θ $(0 < \alpha, \beta, \theta < 1)$, положить $k = 0$ (начало цикла по k).

Шаг 1. Выбрать $\varepsilon_0 \in [\varepsilon, \Lambda]$ и $\delta_0 \in [\delta, \Delta]$, положить $i = 0$ (начало цикла по i).

Шаг 2. Выбрать начальное пробное направление $l^0 \in E_n$, $\| l^0 \| > 0$. Вычислить пробную точку $y^1 = x^k - \delta_i l^0 / \| l^0 \|$ и произвольный обобщенный градиент $g^1 \in G_f(y^1)$. Положить $j = 1$ (начало цикла по j).

Шаг 3. Вычислить новое пробное направление l^j (правило будет описано ниже).

Шаг 4. Если $\| l^j \| \leqslant \varepsilon$ и $\delta_i \leqslant \delta$, то стоп.

Шаг 5. Если $\| l^j \| \leqslant \varepsilon_i$, то $\varepsilon_{i+1} = \alpha\varepsilon_i$, $\delta_{i+1} = \beta\delta_i$, заменить i на $i + 1$, перейти к шагу 2.

Шаг 6. Вычислить новую пробную точку $y^{j+1} = x^k - \delta_i l^j / \| l^j \|$ и произвольный обобщенный градиент $g^{j+1} \in G_f(y^{j+1})$.

Шаг 7. Если $(g^{j+1}, l^j) < \theta\varepsilon_i^2/2$, то заменить j на $j + 1$, перейти к шагу 3.

Шаг 8. Если $f(y^{j+1}) > f(x^k) - \theta\delta_i\varepsilon_i^2/(4 \| l^j \|)$, то $\varepsilon_{i+1} = \varepsilon_i$, $\delta_{i+1} = \beta\delta_i$, заменить i на $i + 1$, перейти к шагу 2.

Шаг 9. Найти следующее приближение x^{k+1} такое, что $f(x^{k+1}) \leqslant f(y^{j+1})$; в частности, можно взять $x^{k+1} = y^{j+1}$. Заменить k на $k + 1$, перейти к шагу 1.

Замечание 11.1. В цикле по k алгоритм вырабатывает последовательность приближений x^k, в цикле по i подбирает при фиксированном x^k величину пробного шага δ_i, а в цикле по j отыскивает при фиксированных x^k и δ_i направление спуска l^j. На шаге 9 может использоваться произвольный алгоритм отыскания лучшей, чем y^{j+1} (по значению f), точки x^{k+1}; например, можно проводить одномерную оптимизацию f в направлении l^j.

В алгоритме А1 на шаге 3 используется правило L построения пробных направлений l^j из накопленных обобщенных градиентов g^1, g^2, ..., g^j. Правило L должно удовлетворять следующим двум условиям:

У1) $l^j \in \operatorname{co}\{g^1, \ldots, g^j\}$, $j = 1, 2, \ldots$;

У2) если последовательность $\{g^j\}_{j=1}^{\infty}$ вложена в такое выпуклое множество P, что $0 < \gamma \leqslant \| g \| \leqslant \Gamma < \infty$ для $g \in P$, то найдется такой индекс m, что

$$(g^{m+1}, l^m) \geqslant \theta\gamma^2/2, \quad 0 < \theta < 1.$$

Этим условиям удовлетворяют, как будет показано в леммах 11.3—11.5, например, следующие правила построения пробных направлений l^j.

L1. Пусть числа $\{\rho_r\}_{r=1}^{\infty}$ таковы, что $0 \leqslant \rho_r \leqslant \rho < \infty$, $\sum\limits_{r=1}^{\infty} \rho_r = \infty$.

Пусть

$$l^j = \sum_{r=1}^{j} \rho_r g^r \Big/ \sum_{r=1}^{j} \rho_r = \Big(1 - \rho_j \Big/ \sum_{r=1}^{j} \rho_r\Big) l^{j-1} + \Big(\rho_j \Big/ \sum_{r=1}^{j} \rho_r\Big) g^j.$$

В частности, при $\rho_r \equiv \rho$ имеем

$$l^j = \frac{1}{j} \sum_{r=1}^{j} \rho_r g^r = \left(1 - \frac{1}{j}\right) l^{j-1} + \frac{g^j}{j}.$$

L2. $l^j \in \mathrm{co}\{l^{j-1}, g^j\}$ и $\| l^j \|$ минимальна, $l^1 = g^1$.

L3. $l^j \in \mathrm{co}\{g^1, \ldots, g^j\}$ и $\| l^j \|$ минимальна.

L4. $l^j \in \mathrm{co}\{l^{j-1}, g^i,\ i \in I_j\}$, где $j \in I_j \subset \{1, \ldots, j\}$, и $\| l^j \|$ минимальна.

Теорема 11.1. *Пусть $f(x)$ — обобщенно дифференцируемая функция, множество $D = \{x \in E_n \mid f(x) \leqslant f(x^0)\}$ ограничено. Тогда алгоритм* A1 *останавливается через конечное число итераций в точке x^k, принадлежащей множеству $X^*_{\varepsilon\delta} \cap D$.*

Доказательство. По построению (см. шаги 8, 9) рассматриваемый алгоритм является релаксационным, поэтому все приближения x^k принадлежат множеству D. Если алгоритм останавливается через конечное число итераций в точке x^k на шаге 5, то по условию остановки соответствующие $\| l^j \| \leqslant \varepsilon$ и $\delta_i \leqslant \delta$, поэтому $x^k \in X^*_{\varepsilon\delta} \cap D$.

Конечность алгоритма докажем от противного. Итак, предположим, что алгоритм работает бесконечно долго. Покажем, что тогда он вырабатывает бесконечную последовательность точек $\{x^k\}_{k=0}^{\infty}$, т. е. зацикливание на шагах 5, 7, 8 невозможно. На шаге 5 оно невозможно, так как в этом случае рано или поздно станет $\varepsilon_i \leqslant \varepsilon$ и $\delta_i \leqslant \delta$, и алгоритм остановится. На шаге 7 оно невозможно в силу условия У2 на правило L. На шаге 8 оно невозможно в силу свойств рассматриваемых функций. Например, для выпуклых функций $f(x)$ шаг 8 всегда пропускается, так как на шаге 8 всегда

$$f(y^{j+1}) \leqslant f(x^k) + (g^{j+1}, y^{j+1} - x^k) = f(x^k) - \delta_i (g^{j+1}, l^j / \| l^j \|) \leqslant \leqslant f(x^k) - \theta \varepsilon_i^2 \delta_i / (2 \| l^j \|). \quad (11.2)$$

Числовая последовательность $\{f(x^k)\}_{k=0}^{\infty}$ монотонно убывает и ограничена снизу, поэтому существует конечный предел $\lim\limits_{k \to \infty} f(x^k)$. Но, с другой стороны, в выпуклом случае всегда $\varepsilon_i \geqslant \alpha^p \Lambda$ и $\delta_i \geqslant \beta^p \Delta$, где показатель степени p таков, что $\alpha^p \Lambda \leqslant \varepsilon$ и $\beta^p \Delta \leqslant \delta$. Кроме того, $\| l^j \| \leqslant \Gamma < +\infty$. Таким образом, $f(x^k)$ каждый раз убывает, согласно (11.2), не менее чем на некоторую фиксированную величину, и, следовательно, $\{f(x^k)\}_{k=0}^{\infty}$ не может иметь предела. В выпуклом случае теорема доказана.

Покажем, что если $f(x)$ обобщенно дифференцируема, то зацикливание на шаге 8 также невозможно. Действительно, при зацикливании на этом шаге будет $\lim\limits_{i \to \infty} \delta_i = 0$ и $\varepsilon_i \geqslant \gamma = \alpha^p \Lambda$, где p таково, что $\alpha^p\Lambda \leqslant \varepsilon$ и $\beta^p\Delta \leqslant \delta$. Обозначим

$$\Gamma = \sup\{\|g\| \mid g \in G_f(y),\ f(y) \leqslant f(x^0)\}.$$

Имеем $\Gamma < +\infty$ ввиду ограниченности на компакте компактнозначного полунепрерывного сверху отображения G_f. При достаточно малых δ_l в силу обобщенной дифференцируемости $f(x)$ на шаге 8 имеем

$$f(y^{j+1}) \leqslant f(x^k) + (g^{j+1}, y^{j+1} - x^k) + o(x^k, y^{j+1}, g^{j+1}) \leqslant$$

$$\leqslant f(x^k) - \delta_l\left(g^{j+1}, \frac{l^j}{\|l^j\|}\right) + \frac{\theta\gamma^2}{4\Gamma}\|y^{j+1} - x^k\| \leqslant$$

$$\leqslant f(x^k) - \frac{\theta\varepsilon_l^2\delta_l}{2\|l^j\|} + \frac{\theta\gamma^2\delta_l}{4\Gamma} \leqslant f(x^k) - \frac{\theta\varepsilon_l^2\delta_l}{4\|l^j\|},$$

т. е. зацикливание на шаге 8 невозможно.

Итак, если алгоритм не останавливается, то он вырабатывает бесконечную последовательность приближений $\{x^k\}$. Так как $\{f(x^k)\}_{k=0}^{\infty}$ монотонно убывает и ограничена снизу, то существует предел $\lim\limits_{k\to\infty} f(x^k)$. Пусть $\lim\limits_{s\to\infty} x^{k_s} = x'$. По условию перехода с шага 8 на шаг 9 имеем

$$f(x^{k_s+1}) \leqslant f(x^{k_s}) - \theta\varepsilon_l^2\delta_l/(4\|l^j\|). \tag{11.3}$$

Заметим, что $\|l^j\| \leqslant \Gamma < +\infty$ и $\varepsilon_l \geqslant \gamma = \alpha^p\Lambda > 0$, где p таково, что $\alpha^p\Lambda \leqslant \varepsilon$ и $\beta^p\Delta \leqslant \delta$. Поэтому, если мы покажем, что

$$\lim_{l\to\infty}\delta_l = \lim_{s\to\infty}\|y^{j_s+1} - x^{k_s}\| > 0, \tag{11.4}$$

то, переходя в (11.3) к пределу по s, получим противоречие со сходимостью $\{f(x^k)\}_{k=0}^{\infty}$, и теорема будет доказана.

Итак, покажем, что (11.4) выполняется. Запишем разложение для $f(x)$ в точке x':

$$f(y) = f(x') + (g_y, y - x') + o(x', y, g_y).$$

Согласно определению обобщенно дифференцируемой функции, для $a = \theta\gamma^2/(12\Gamma)$ существует такое $\delta'(a)$, что для всех $\{y \mid \|y - x'\| \leqslant \delta'\}$ и всех $g_y \in G_f(y)$

$$o(x', y, g_y) \leqslant a\|y - x'\|.$$

Для x^{k_s} и $\{y \mid \|y - x'\| \leqslant \delta'\}$ можем написать

$$f(y) \leqslant f(x') + (g_y, y - x') + a\|y - x'\| \leqslant$$

$$\leqslant f(x^{k_s}) + (g_y, y - x^{k_s}) + a\|y - x^{k_s}\| + f(x') -$$

$$- f(x^{k_s}) + (g_y, x^{k_s} - x') + a\|x^{k_s} - x'\| \leqslant$$

$$\leqslant f(x^{k_s}) + (g_y, y - x^{k_s}) + a\|y - x^{k_s}\| + C\|x^{k_s} - x'\|,$$

где $C = 2\Gamma + a$. Возьмем $\delta'' = \delta'/2$. Покажем, что для достаточно больших k_s

$$\| y^{j+1} - x^{k_s} \| \geqslant \beta\delta''.$$

Найдем такое K, что для всех $k_s \geqslant K$

$$\| x^{k_s} - x' \| \leqslant \min \{\delta'', a\beta\delta''/(2C)\}.$$

Для $k = k_s$ алгоритм на шаге 8 будет строить такие y^{j+1}, l^j, ε_i, δ_i, что

$$(g^{j+1}, l^j) \geqslant \theta\varepsilon_i^2/2, \quad \| y^{j+1} - x^{k_s} \| = \delta_i.$$

Покажем, что при $k_s \geqslant K$ для всех соответствующих δ_i выполняется неравенство $\delta_i \geqslant \beta\delta''$. Действительно, если это не так, то при некоторых $k_s \geqslant K$ и i

$$\beta\delta'' \leqslant \delta_i \leqslant \delta'', \quad \| y^{j+1} - x^{k_s} \| = \delta_i, \quad (g^{j+1}, l^j) \geqslant \theta\varepsilon_i^2/2.$$

Отметим, что $\| y^{j+1} - x' \| \leqslant \delta'$ и $\varepsilon_i \geqslant \gamma > 0$. Подставим эти y^{j+1} и g^{j+1} в полученное выше неравенство:

$$f(y^{j+1}) \leqslant$$
$$\leqslant f(x^{k_s}) + (g^{j+1}, y^{j+1} - x^{k_s}) + a\| y^{j+1} - x^{k_s} \| + C\| x^{k_s} - x' \| \leqslant$$
$$\leqslant f(x^{k_s}) - \delta_i \left(g^{j+1}, \frac{l^j}{\| l^j \|}\right) + a\delta_i + \frac{a\beta\delta''}{2} \leqslant$$
$$\leqslant f(x^{k_s}) - \frac{\theta\varepsilon_i^2\delta_i}{2\| l^j \|} + \frac{\theta\gamma^2(2\delta_i + \beta\delta'')}{24\Gamma} \leqslant f(x^{k_s}) - \frac{\theta\delta_i\varepsilon_i^2}{4\| l^j \|}.$$

Но это означает, что на шаге 8 δ_i дальше не может уменьшаться. Таким образом, для достаточно больших k_s

$$\delta_i = \| y^{j+1} - x^{k_s} \| \geqslant \beta\delta'',$$

что и требовалось показать. Теорема доказана.

Замечание 11.2. На шаге 1 алгоритма при $k \geqslant 1$ можно полагать

$$\delta_0 = \max \{\| x^k - x^{k-1} \|, \delta\},$$
$$\varepsilon_0 = \max \{(f(x^{k-1}) - f(x^k))/\| x^k - x^{k-1} \|, \varepsilon\};$$

при этом теорема остается в силе.

Замечание 11.3. Если на шаге 4 вместо «стоп» ввести деление ε и δ на число, большее единицы, то в силу следствия 11.1 и теоремы 11.1 итерационный алгоритм будет сходиться предельными точками к множеству X^*.

Замечание 11.4. Алгоритм A1 может быть распространен на задачи условной оптимизации так же, как это сделано для метода обобщенного градиента в § 9.

Как видно из доказательства теоремы 11.1, правило L построения пробных направлений l^j решает следующую задачу.

Задача ПН (поиска направления). Пусть имеется множество $P \subset E_n$ такое, что $0 < \gamma \leqslant \| g \| \leqslant \Gamma < +\infty$ для всех $g \in \operatorname{co} P$. Мы получаем информацию о множестве P в виде последовательности векторов $g^j \in P$ $(j = 1, 2, \ldots)$. По имеющейся в каждый момент j информации $\{g^1, g^2, \ldots, g^j\}$ необходимо построить векторы $l^j \in \operatorname{co}\{g^1, g^2, \ldots, g^j\}$ $(j = 1, 2, \ldots)$ так, чтобы прийти к неравенству

$$(g^{j+1}, l^j) \geqslant \theta\gamma^2/2.$$

Сразу не очевидно, что существуют правила построения направлений l^j, решающие задачу ПН. Но мы покажем, что правила L1—L4 решают указанную задачу, и приведем некоторые оценки количества шагов, необходимых для ее решения. Здесь естественно возникает вопрос об оптимальном правиле, который мы, однако, оставляем открытым.

Определение 11.1. *Трудоемкостью правила* L на задаче ПН будем называть число

$$m_1(P, L, \theta) = \sup_{\{g^j\}_{j=1}^{\infty} \subset P} \min\{j \mid (g^{j+1}, l^j) \geqslant \theta\gamma^2/2,\ 0 < \theta < 1\}.$$

Лемма 11.3. *Правило* L1 *решает задачу ПН. Если в нем* $\rho_r \equiv \rho$, *т. е.*

$$l^j = \frac{1}{j} \sum_{r=1}^{j} g^r,$$

то для трудоемкости справедлива оценка

$$m_1 \leqslant \frac{2}{1-\theta} \frac{\Gamma^2}{\gamma^2}.$$

Доказательство. Пусть $\{g^j\}_{j=1}^{\infty}$ — произвольная последовательность векторов из выпуклого множества $P \subset E_n$ такого, что

$$0 < \gamma \leqslant \| g \| \leqslant \Gamma < +\infty \quad \text{для} \quad g \in P.$$

Пусть для всех $j < m$ имеет место

$$(g^{j+1}, l^j) \leqslant \theta\gamma^2/2, \qquad 0 < \theta < 1.$$

Просуммируем равенства

$$\Big\| \sum_{r=1}^{j} \rho_r g^r \Big\|^2 = \Big\| \sum_{r=1}^{j-1} \rho_r g^r \Big\|^2 + 2\rho_j \Big(g^j, \sum_{r=1}^{j-1} \rho_r g^r\Big) + \rho_j^2 \| g^j \|^2$$

по j от 2 до m:

$$\Big\| \sum_{r=1}^{m} \rho_r g^r \Big\|^2 = \rho_1^2 \| g^1 \|^2 + 2 \sum_{j=2}^{m} \rho_j \Big(g^j, \sum_{r=1}^{j-1} \rho_r g^r\Big) + \sum_{j=2}^{m} \rho_j^2 \| g^j \|^2.$$

Справедливы оценки

$$\gamma^2\Big(\sum_{r=1}^{m}\rho_r\Big)^2\leqslant\Big\|\sum_{r=1}^{m}\rho_r g^r\Big\|^2\leqslant\rho_1\Gamma^2+\theta\gamma^2\sum_{j=2}^{m}\rho_j\sum_{r=1}^{j-1}\rho_r+\rho\Gamma^2\sum_{j=2}^{m}\rho_j\leqslant$$

$$\leqslant\rho_1\Gamma^2+\theta\gamma^2\Big(\sum_{j=1}^{m}\rho_j\Big)^2+\rho\Gamma^2\sum_{j=1}^{m}\rho_j,$$

откуда

$$(1-\theta)\frac{\gamma^2}{\Gamma^2}\leqslant\rho_1/\Big(\sum_{j=1}^{m}\rho_j\Big)^2+\rho/\sum_{j=1}^{m}\rho_j.$$

Отсюда ясно, что число m не может быть сколь угодно большим, т. е. найдется m такое, что

$$(g^{m+1},l^m)\geqslant\theta\gamma^2/2.$$

Если $\rho_r\equiv\rho$, то

$$(1-\theta)\gamma^2/\Gamma^2\leqslant 1/m^2+1/m\leqslant 2/m,$$

т. е. $m\leqslant\dfrac{2}{1-\theta}\dfrac{\Gamma^2}{\gamma^2}$. Лемма доказана.

Лемма 11.4. *Правило* L2 *решает задачу ПН и, следовательно, удовлетворяет условиям* У1, У2. *Более того, для трудоемкости вида*

$$m_2(P,\mathrm{L}2)=\sup_{\{g^j\}_{j=1}^{\infty}\subset P}\min\Big\{j\,|\,(g^{j+1},l^j)\geqslant\frac{1}{2}\|l^j\|^2\Big\}$$

справедлива оценка

$$m_2\leqslant 1+8\frac{\Gamma^2}{\gamma^2}\ln\frac{\Gamma}{\gamma}.$$

Так как из $(g^{j+1},l^j)\geqslant\|l^j\|^2/2$ *следует* $(g^{j+1},l^j)\geqslant\theta\gamma^2/2$, $0<\theta<1$, *то для трудоемкости* $m_1(P,\mathrm{L}2,\theta)$ *из определения* 11.1 *справедлива та же оценка.*

Доказательство. Пусть P — выпуклое множество,

$$0<\gamma\leqslant\|g\|\leqslant\Gamma<+\infty,\quad g\in P;\qquad\{g^j\}_{j=1}^{\infty}\subset P.$$

Векторы l^j строятся рекуррентно:

$$l^1=g^1;\qquad l^j\in\mathrm{co}\{l^{j-1},g^j\},$$

$\|l^j\|$ минимальна. Пусть для всех $j\leqslant m$

$$(g^j,l^{j-1})\leqslant\|l^{j-1}\|^2/2,$$

или

$$(g^j,l^{j-1}/\|l^{j-1}\|)\leqslant\|l^{j-1}\|/2.$$

Постараемся получить оценки вида

$$\|l^j\|\leqslant q\|l^{j-1}\|,\qquad 0<q<1,$$

при условии, что проекции векторов g^j на направления l^{j-1} не превосходят $\| l^{j-1} \|/2$. Необходимо рассмотреть несколько случаев. Если

$$\| g^j \| \leqslant \| l^{j-1} \|/2,$$

то

$$\| l^j \| \leqslant \| g^j \| \leqslant \| l^{j-1} \|/2.$$

Если

$$\| l^{j-1} \|/2 \leqslant \| g^j \| \leqslant \| l^{j-1} \|/\sqrt{2},$$

то из геометрических соображений

$$\| l^j \| \leqslant \| g^j \| \leqslant \| l^{j-1} \|/\sqrt{2}.$$

Пусть теперь

$$\| g^j \| \geqslant \| l^{j-1} \|/\sqrt{2}.$$

Обозначим $x = (g^j, l^{j-1}/\| l^{j-1} \|)$. Из геометрических соображений находим

$$\| l^j \|^2 = \| l^{j-1} \|^2 \frac{\| g^j \|^2 - x^2}{\| g^j \|^2 - 2x \| l^{j-1} \| + \| l^{j-1} \|^2}.$$

Максимум правой части по x при $x \leqslant \| l^{j-1} \|/2$ достигается при $x = \| l^{j-1} \|/2$; поэтому

$$\| l^j \|^2 \leqslant \| l^{j-1} \|^2 \left(1 - \frac{\| l^{j-1} \|^2}{4 \| g^j \|^2}\right) \leqslant \| l^{j-1} \|^2 \left(1 - \frac{\gamma^2}{4\Gamma^2}\right).$$

Поскольку

$$1 - \gamma^2/(2\Gamma^2) > 1/\sqrt{2},$$

то при

$$(g^j, l^{j-1}/\| l^{j-1} \|) \leqslant \| l^{j-1} \|/2$$

всегда имеем

$$\| l^j \|^2 \leqslant \| l^{j-1} \|^2 (1 - \gamma^2/(4\Gamma^2)).$$

Для $j = m$ получаем

$$\gamma^2 \leqslant \| l^m \|^2 \leqslant \| l^1 \|^2 (1 - \gamma^2/(4\Gamma^2))^{m-1} \leqslant \Gamma^2 (1 - \gamma^2/(4\Gamma^2))^{m-1},$$

откуда

$$m \leqslant 1 + 8 \frac{\Gamma^2}{\gamma^2} \ln \frac{\Gamma}{\gamma}.$$

Лемма доказана.

Л е м м а 11.5. *Правила* L3, L4 *решают задачу* ПН *и, следовательно, удовлетворяют условиям* У1, У2. *Для трудоемкостей* $m_2(P, \mathrm{L3})$ *и* $m_2(P, \mathrm{L4})$ *заведомо справедлива оценка из леммы* 11.4:

$$\max \{m_2(P, \mathrm{L3}), m_2(P, \mathrm{L4})\} \leqslant 1 + 8 \frac{\Gamma^2}{\gamma^2} \ln \frac{\Gamma}{\gamma}.$$

Доказательство. Пусть решается задача ПН, причем векторы l^j строятся по правилам L3 или L4. Пусть для всех $j \leqslant m$ выполняется неравенство

$$(g^j, l^{j-1}) < \| l^{j-1} \|^2/2.$$

Рассмотрим вектор $\bar{l}^j$, построенный из l^{j-1} и g^j согласно правилу L2. Очевидно, $\| l^j \| \leqslant \| \bar{l}^j \|$. При доказательстве леммы 11.4 было получено неравенство

$$\| \bar{l}^j \|^2 \leqslant \| l^{j-1} \|^2 (1 - \gamma^2/(4\Gamma^2)),$$

откуда, как и в лемме 11.4, следует оценка

$$m \leqslant 1 + 8 \frac{\Gamma^2}{\gamma^2} \ln \frac{\Gamma}{\gamma}.$$

Лемма доказана.

Лемма 11.6. *Если в задаче* ПН $P = \{a^1, \ldots, a^N\}$, *то*

$$m_2 (P, \mathrm{L3}) \leqslant N + 1.$$

Доказательство. Для l^j, вычисленных по правилу L3, справедливо

$$(g^j, l^j) \geqslant \| l^j \|^2.$$

Очевидно, $g^{N+1} \in \{a^1, \ldots, a^N\}$; поэтому $l^{N+1} = l^N$ и $(g^{N+1}, l^N) \geqslant \| l^N \|^2$, т. е.

$$m_2 (P, \mathrm{L3}) \leqslant N + 1.$$

Лемма доказана.

Алгоритм А2. В алгоритме А1 для вычисления направлений на шаге 3 и при проверке условий на шаге 7 вместо векторов $g^1, \ldots, g^{j+1}$ можно использовать нормированные величины $\bar{g}^1, \ldots, \bar{g}^{j+1}$, где $\bar{g} = g/\|g\|$ ($\bar{g} = 0$, если $g = 0$). Тогда шаг 8 алгоритма следует заменить на шаг 8′.

Шаг 8′. Если

$$f(y^{j+1}) > f(x^k) - \theta\delta_i\varepsilon_i^2 \| g^{j+1} \| / (4 \| l^j \|),$$

то $\varepsilon_{i+1} = \varepsilon_i$, $\delta_{i+1} = \beta\delta_i$, заменить i на $i + 1$, перейти к шагу 2.

Доказательство сходимости модифицированного таким образом алгоритма к множеству $\bar{X}_{\varepsilon\delta} \cap \{x \in E_n \mid f(x) \leqslant f(x^0)\}$, где

$$\bar{X}_{\varepsilon\delta} = \{x \mid \bar{\gamma}_\delta (x) \leqslant \varepsilon\},$$

$$\bar{\gamma}_\delta (x) = \inf \left\{ \| \bar{g} \| \mid \bar{g} \in \mathrm{co} \bigcup_{\{y \mid \|y-x\| \leqslant \delta\}} \bar{G}_f (y) \right\},$$

$$\bar{G}_f (y) = \{\bar{g} \in E_n \mid g \in G_f (y)\},$$

незначительно отличается от доказательства сходимости алгоритма А1.

Алгоритм А3. В леммах 11.4, 11.5 доказаны более сильные свойства правил L2, L3, чем те, которые использованы в алгоритме А1. Поэтому алгоритм А1 можно модифицировать для правил L2, L3 построения l^j, а именно заменить шаги 7, 8 на следующие:

Шаг 7″. Если $(g^{j+1}, l^j) < \| l^j \|^2/2$, то заменить j на $j+1$, перейти к шагу 3.

Шаг 8″. Если $f(y^{j+1}) > f(x) - \delta_i \| l^j \|/4$, то $\varepsilon_{i+1} = \varepsilon_i$, $\delta_{i+1} = \beta\delta_i$, заменить i на $i+1$, перейти к шагу 2.

Для алгоритма А3 сохраняет силу теорема 11.1.

Оценим трудоемкость предложенных алгоритмов в случае минимизации выпуклой функции. Трудоемкость зависит от количества итераций (циклов) алгоритма, которые подсчитываются с помощью переменных i, j, k (см. замечание 11.1). Рассмотрим самый простой случай. Пусть на шаге 2 алгоритмов всегда принимается $\varepsilon_0 = \delta$, $\delta_0 = \delta$. Тогда в выпуклом случае i-итерации не производятся и всегда $\|l^j\| \geqslant \varepsilon$, $\delta_i = \delta$.

Обозначим

$$D = \{x \mid f(x) \leqslant f(x^0)\},$$

$$\Gamma = \sup \{\| g \| \mid g \in G_f(y),\ y \in D\};$$

очевидно, что $\| l^j \| \leqslant \Gamma$. Таким образом, для алгоритма А1 на k-итерации значение выпуклой функции $f(x)$ уменьшается не меньше, чем на $\theta\delta\varepsilon^2/(2\Gamma)$. Следовательно, общее число k-итераций не превосходит величины

$$\frac{2\Gamma}{\theta\delta\varepsilon^2} [f(x^0) - \min_{x \in D} f(x)].$$

Обозначим через d диаметр области D; тогда

$$[f(x^0) - \min_{x \in D} f(x)] \leqslant d\Gamma.$$

Поэтому общее число k-итераций не превосходит $2d\Gamma^2/(\theta\delta\varepsilon^2)$. Теперь все зависит от числа j-итераций на каждой k-итерации. Обозначим

$$\Gamma_\delta(x) = \sup \{\| g \| \mid g \in G_f(y),\ \| y - x \| \leqslant \delta\},$$

$$\gamma_\delta(x) = \inf \{\| g \| \mid g \in \mathrm{co}\, \{G_f(y) \mid \| y - x \| \leqslant \delta\}\}.$$

Величина $Q_\delta(x) = \Gamma_\delta(x)/\gamma_\delta(x)$ характеризует локальную овражность функции $f(x)$. Согласно лемме 11.3, при $l^j = \frac{1}{j} \sum_{r=1}^{j} g^r$ число j-итераций на каждой k-итерации не превосходит величины $2Q_\delta^2(x)/(1-\theta)$.

Итак, общая трудоемкость достижения множества $X^*_{\varepsilon\delta}$ алгоритмом А1, в котором $l^j = \frac{1}{j} \sum_{r=1}^{j} g^r$ и на шаге 9 $x^{k+1} = y^{j+1}$, оценивается следующим образом. Значения функции $f(x)$ вообще не вычисляются, а необходимое количество вычислений градиента составляет

$$N_g^1 \leqslant \frac{4d\Gamma^2}{\theta(1-\theta)\delta\varepsilon^2} \sup_{x \in X^*_{\varepsilon\delta}} Q_\delta^2(x) \leqslant \frac{4d\Gamma^4}{\theta(1-\theta)\delta\varepsilon^4}.$$

Аналогичная оценка для алгоритма А3, в котором используется процедура L2 и на шаге 9 $x^{k+1} = y^{j+1}$, имеет вид

$$N_g^2 \leqslant 16 \frac{d\Gamma}{\delta\varepsilon} \left(\sup_{x \in X^*_{\varepsilon\delta}} \{Q_\delta^2(x) \ln Q_\delta(x)\} + \frac{1}{8} \right) \leqslant 16 \frac{d\Gamma}{\delta\varepsilon} \left(\frac{\Gamma^2}{\varepsilon^2} \ln \frac{\Gamma}{\varepsilon} + \frac{1}{8} \right).$$

Если минимизируемая функция $f(x)$ кусочно линейна и состоит из M кусков, то для алгоритма А3, в котором используется L3 и на шаге 9 $x^{k+1} = y^{j+1}$, аналогичная оценка имеет вид

$$N_g^3 \leqslant 16 \frac{d\Gamma}{\delta\varepsilon} M.$$

§ 12. О многоэкстремальной оптимизации

Невыпуклые функции могут иметь много локальных экстремумов, не совпадающих с абсолютным минимумом функции на допустимом множестве. Рассмотренные до сих пор алгоритмы предназначались лишь для отыскания локальных минимумов. Проблема глобальной оптимизации функций сложна и актуальна. Обзор результатов, полученных в этой области, имеется в работах [114, 117, 165, 181].

В настоящем параграфе приводятся некоторые дополнительные результаты о трех подходах в многомерной глобальной оптимизации:

методе комбинирования локального и нелокального поиска;

методе аппроксимации исходной многоэкстремальной задачи многоэкстремальными задачами специального вида;

методе сглаживания локальных экстремумов.

1. Комбинированные алгоритмы. Естественный путь решения реальных многоэкстремальных задач состоит в разумном сочетании алгоритмов локальной и глобальной оптимизации. Например, в комбинированных алгоритмах [1, 12, 82] этапы работы локального и глобального алгоритма чередуются. При этом каждый раз алгоритмы работают не до конца, а при выполнении некоторых условий происходит переключение с одного на другой. Частным случаем этой схемы можно считать и метод перебора локальных экстремумов [28]. Встречаются и другие виды комбинирования локальных и глобальных алгоритмов [36].

Сходимость комбинированных алгоритмов необходимо специально доказывать, потому что без соблюдения определенных условий они могут не сходиться даже к локальным экстремумам, несмотря на то что построены на основе сходящихся алгоритмов.

Сначала мы приведем два комбинированных алгоритма, в которых основную роль в смысле обеспечения сходимости к локальным экстремумам играет локальный спуск, а глобальный поиск используется для нахождения подходящих точек старта для локального алгоритма. Затем, наоборот, приведем пример алгоритма, в котором основную роль играет глобальный поиск, а как промежуточное звено используется локальный спуск.

Пусть необходимо

$$f(x) \to \min, \qquad x \in E_n, \tag{12.1}$$

где $f(x)$ — многоэкстремальная обобщенно дифференцируемая функция.

В следующем комбинированном алгоритме для локального спуска используется метод обобщенного градиента, а для нелокального поиска может быть использован любой другой алгоритм.

Алгоритм К1.

Шаг 0. Выбрать начальную точку $x^0 \in E_n$, числа σ, $\Delta > 0$ и последовательности $\{\varepsilon_k\}$, $\{\delta_k\}$, $\{\rho_k\}$ такие, что

$$\varepsilon_k, \delta_k, \rho_k \geqslant 0;$$

$$\lim_{k\to\infty} \varepsilon_k = \lim_{k\to\infty} \delta_k = \lim_{k\to\infty} \rho_k = 0, \quad \sum_{k=0}^{\infty} \rho_k = \infty. \tag{12.2}$$

Положить $k = 0$, $\sigma_0 = 0$.

Шаг 1. Вычислить значение $f(x^k)$ и произвольный вектор $g^k \in G_f(x^k)$.

Шаг 2. Если $f(x^k) \geqslant \min\limits_{0\leqslant r\leqslant k} f(x^r) + \Delta$, то положить $x^{k+1} = \arg\min\{f(x) \mid x = x^r,\ r \in [0, k]\}$, $\sigma_{k+1} = 0$; заменить k на $k+1$ и перейти к шагу 1.

Шаг 3. Если $\sigma_k \geqslant \sigma$ или $\|g^k\| \leqslant \varepsilon_k$, то найти каким-либо алгоритмом $\tilde{A}$ новую точку старта $\tilde{x}$ такую, что $f(\tilde{x}) \leqslant f(x^k) + \delta_k$, и положить $x^{k+1} = \tilde{x}$, $\sigma_{k+1} = 0$; заменить k на $k+1$ и перейти к шагу 1.

Шаг 4. Сделать шаг градиентного метода

$$x^{k+1} = x^k - \rho_k g^k / \|g^k\|$$

и положить $\sigma_{k+1} = \sigma_k + \rho_k$. Заменить k на $k+1$ и перейти к шагу 1.

Замечание 12.1. В качестве вспомогательного алгоритма $\tilde{A}$ может выступать, например, какой-либо алгоритм глобального поиска в окрестности точки x^k. Алгоритм $\tilde{A}$ должен быть результативным; в частности, за неимением лучшей точки можно взять $\tilde{x} = x^k$ или $\tilde{x} = \arg\min\{f(x) \mid x = x^r,\ r \in [0, k]\}$.

Теорема 12.1. *Пусть в задаче* (12.1) *функция* $f(x)$ *обобщенно дифференцируема и множество* $D = \{x \in E_n \mid f(x) \leqslant f(x^0) + \Delta\}$ *ограничено; последовательность точек* $\{x^k\}$ *построена по алгоритму* К1. *Тогда все предельные точки* $\{x^k\}$ *принадлежат* D, *среди минимальных из них по значению* f *найдутся точки, принадлежащие* $X^* = \{x \in E_n \mid 0 \in G_f(x)\}$, *и интервал* $[\varliminf_{k\to\infty} f(x^k), \varlimsup_{k\to\infty} f(x^k)]$ *вложен в множество* $F^* = \{f(x) \mid x \in X^*\}$. *Если* F^* *не содержит интервалов, то все предельные точки* $\{x^k\}$ *принадлежат* X^*, *и числовая последовательность* $\{f(x^k)\}$ *имеет предел.*

Доказательство. Заметим, что в К1 фактически скомбинированы три алгоритма: A_1 — градиентный метод, A_2 — возврат в рекордную точку траектории и $A_3 = \tilde{A}$ — поиск новой точки старта $\tilde{x}$. Для удобства введем символьную функцию $A: k \to \{A_1, A_2, A_3\}$, которая показывает, какой из алгоритмов — A_1, A_2 или A_3 — работает на k-й итерации, т. е. каким алгоритмом порождается точка x^{k+1}. Выделим индексы $k_s (s = 0, 1, \ldots)$, при которых происходит переключение с одного алгоритма на другой, т. е. в моменты k_s имеет место либо $\sigma_{k_s} = 0$, либо $\sigma_{k_s} \geqslant \sigma$, либо $\| g^{k_s} \| \leqslant \varepsilon_{k_s}$, либо $f(x^{k_s}) \geqslant \min\limits_{0 \leqslant k \leqslant k_s} f(x^k) + \Delta$.

В силу условия (12.2) и по построению алгоритма К1 последовательность $\{k_s\}$ бесконечна.

Алгоритм К1 может выйти за пределы ограниченного множества D только на один шаг градиентного метода, и поскольку $\lim\limits_{k \to \infty} \rho_k = 0$, то все предельные точки $\{x^k\}$ принадлежат D. Таким образом, первое утверждение теоремы доказано.

Докажем второе утверждение: среди минимальных по значению f предельных точек $\{x^k\}$ имеются точки, принадлежащие X^*. Предположим противное. Пусть

$$\lim_{t \to \infty} x^{k_t} = x' \notin X^*, \quad f(x') = \lim_{t \to \infty} f(x^{k_t}) = \varliminf_{k \to \infty} f(x^k). \tag{12.3}$$

Найдем индексы $k_{s(t)} \in \{k_s\}$ такие, что $k_{s(t)-1} \leqslant k_t < k_{s(t)}$ $(t = 0, 1, \ldots)$. Без ограничения общности можем считать, что существует $x'' = \lim\limits_{t \to \infty} x^{k_{s(t)}}$.

Покажем, что

$$f(x') = \lim_{t \to \infty} f(x^{k_t}) = \varlimsup_{t \to \infty} \max_{k_t \leqslant k \leqslant k_{s(t)}} f(x^k). \tag{12.4}$$

Рассмотрим несколько случаев.

Пусть $A(k_t) = A_1$ для всех t. Предположим, что (12.4) не имеет места. Тогда в силу непрерывности $f(x)$

$$\varlimsup_{t \to \infty} \max_{k_t \leqslant k \leqslant k_{s(t)}} \| x^k - x \| > 0$$

и, следовательно,

$$\varlimsup_{t \to \infty} \sum_{k=k_t}^{k_{s(t)}-1} \rho_k > 0. \tag{12.5}$$

Можно считать, что метод обобщенного градиента как бы многократно стартует из точек x^{k_t} $(t = 0, 1, \ldots)$. При условии (12.5) из леммы 9.3 следует существование предельных точек $\{x^k\}$ со значениями f меньшими, чем в x', что противоречит (12.3).

Если $A(k_t) = A_2$ для всех t, то $k_{s(t)} = k_t + 1$ и (12.4) выполнено в силу того, что

$$f(x^{k_t+1}) \leqslant f(x^{k_t}) - \Delta.$$

Если $A(k_t) = A_3$ для всех t, то $k_{s(t)} = k_t + 1$ и (12.4) справедливо в силу выполнения условий

$$f(x^{k_t+1}) \leqslant f(x^{k_t}) + \delta_{k_t}, \quad \lim_{t \to \infty} \delta_{k_t} = 0.$$

Теперь очевидно, что и в общем случае, когда $A(k_t) \in \{A_1, A_2, A_3\}$, свойство (12.4) выполнено.

Без ограничения общности будем считать, что $\lim\limits_{t\to\infty} x^{k_{s(t)}} = x''$. По предположению противного $x'' \notin X^*$, а в силу (12.4) справедливо $f(x') \geqslant f(x'')$. Можно считать, что из точек $\{x^{k_{s(t)}}\}$ стартуют алгоритмы A_1, A_2 или A_3. Если $A(k_{s(t)}) = A_1$ для всех t, то либо $\sigma_{k_{s(t)+1}} \geqslant \sigma > 0$, либо $\| g^{k_{s(t)+1}} \| \leqslant \varepsilon_{k_{s(t)+1}}$. В любом случае $\varliminf\limits_{t\to\infty} \| x^{k_{s(t)+1}} - x'' \| > 0$, и, следовательно,

$$\sum_{k=k_{s(t)}}^{k_{s(t)+1}-1} \rho_k \geqslant c' > 0.$$

Отсюда в силу леммы 9.3 следует существование других предельных точек последовательности $\{x^k\}$ с меньшими значениями f, чем в x' и x''. Пришли к противоречию.

Если $A(k_{s(t)}) = A_2$ для всех t, то в силу неравенства $f(x^{k_{s(t)+1}}) \leqslant f(x^{k_{s(t)}}) - \Delta$ точка x' не является минимальной предельной точкой $\{x^k\}$ (по значению f). Получили противоречие.

Если $A(k_{s(t)}) = A_3$ для всех t, то $k_{s(t)+1} = k_{s(t)} + 1$. Без ограничения общности можем считать, что $\lim\limits_{t\to\infty} x^{k_{s(t)+1}} = x''' \notin X^*$. Так как $f(x^{k_{s(t)+1}}) \leqslant f(x^{k_{s(t)}}) + \delta_{k_{s(t)}}$ и $\lim\limits_{t\to\infty} \delta_{k_{s(t)}} = 0$, то

$$f(x') \geqslant f(x'') \geqslant \lim_{t\to\infty} f(x^{k_{s(t)+1}}) = f(x'''). \tag{12.6}$$

Теперь по построению комбинированного алгоритма К1 в точках $x^{k_{s(t)+1}}$ могут быть задействованы при достаточно больших t только алгоритмы A_1 или A_2, которые, как показано выше, порождают предельные точки $\{x^k\}$ с меньшими значениями f, чем в x'. Опять пришли к противоречию.

Легко видеть, что общий случай, когда $A(k_t) \in \{A_1, A_2, A_3\}$, сводится к рассмотренным частным случаям. Таким образом, полученные выше противоречия доказывают второе утверждение теоремы.

Докажем третье утверждение. Обозначим

$$\underline{f} = \varliminf_{k\to\infty} f(x^k), \quad \bar{f} = \varlimsup_{k\to\infty} f(x^k).$$

Покажем, что $[\underline{f}, \bar{f}) \subset F^*$. Предположим противное. Тогда существует число c такое, что $c \in [\underline{f}, \bar{f})$ и $c \notin F^*$. Согласно доказанному второму утверждению теоремы, $\underline{f} \in F^*$; следовательно, $c > \underline{f}$. Выберем число d так, что $\underline{f} < c < d < \bar{f}$. Числовая последовательность $\{f(x^k)\}$ пересекает интервал $[c, d]$ от c к d бесконечное число раз; поэтому можно выделить индексы $\{k_t\}$ и $\{m_t\}$ такие, что $\lim\limits_{t\to\infty} x^{k_t} = x'$ и либо

$$f(x^{k_t}) \leqslant c < f(x^k) < d \leqslant f(x^{m_t}), \quad k \in (k_t, m_t), \tag{12.7}$$

либо

$$f(x^{k_t}) \leqslant c < d \leqslant f(x^{k_t+1}).$$

Для подпоследовательности $\{x^{k_t}\}$ можно доказать справедливость соотношения (12.4) так же, как это сделано выше. Из (12.4) и (12.7) следует, что для больших t $k_{s(t)} < m_t$ и

$$f(x') = \lim_{t\to\infty} f(x^{k_t}) = \lim_{t\to\infty} f(x^{k_{s(t)}}) = c. \tag{12.8}$$

Без ограничения общности будем считать, что $\lim\limits_{t\to\infty} x^{k_{s(t)}} = x''$. Из (12.8) в силу выбора $c \notin F^*$ следует, что $x'' \notin X^*$. Теперь можем считать, что из точек $x^{k_{s(t)}}$ $(t = 0, 1, \ldots)$ стартует алгоритм A_1, A_2 или A_3. Рассмотрим ряд частных случаев.

Если $A(k_{s(t)}) = A_1$ для всех t, то по построению комбинированного алгоритма либо $\sigma_{k_{s(t)+1}} \geqslant \sigma > 0$, либо $\| g^{k_{s(t)+1}} \| \leqslant \varepsilon_{k_{s(t)+1}}$. Отсюда с учетом того, что $x'' \notin X^*$, следует

$$\lim_{t\to\infty} \| x^{k_{s(t)+1}} - x'' \| > 0, \qquad \sum_{k=k_{s(t)}}^{k_{s(t)+1}-1} \rho_k \geqslant c' > 0.$$

Можно считать, что процесс градиентного спуска как бы многократно стартует из точек последовательности $\{x^{k_{s(t)}}\}$, которая сходится к $x'' \notin X^*$. В данной ситуации к последовательности $\{x^k\}_{k \geqslant k_{s(t)}}$ $(t = 0, 1, \ldots)$ применима лемма 9.3. В этой лемме фигурирует параметр ε, который выберем так, что $\max\{f(x) \mid \| x - x'' \| \leqslant \varepsilon\} < d$. Тогда, согласно лемме, существует подпоследовательность $\{x^{l_t}\}$ такая, что $\| x^k - x'' \| \leqslant \varepsilon$ при $k \in (k_{s(t)}, l_t)$ и

$$f(x'') = \lim_{t\to\infty} f(x^{k_{s(t)}}) > \overline{\lim_{t\to\infty}} f(x^{l_t}). \tag{12.9}$$

Полученное соотношение противоречит (12.7).

Если $A(k_{s(t)}) = A_2$ для всех t, то

$$f(x^{k_{s(t)+1}}) \leqslant f(x^{k_{s(t)}}) - \Delta, \tag{12.10}$$

что также противоречит (12.7).

Пусть $A(k_{s(t)}) = A_3$ для всех t. Без ограничения общности можем считать, что $\lim\limits_{t\to\infty} x^{k_{s(t)+1}} = x'''$. Для $\{x^{k_{s(t)+1}}\}$ справедливо соотношение (12.6). Из (12.6) и (12.7) следует, что для достаточно больших t $k_{s(t)+1} < m_t$ и

$$f(x') = f(x'') = \lim_{t\to\infty} f(x^{k_{s(t)+1}}) = c,$$

откуда в силу выбора $c \notin F^*$ следует, что $x''' \notin X^*$. Теперь по построению комбинированного алгоритма при достаточно больших t в точках $x^{k_{s(t)}+1}$ могут быть задействованы только алгоритмы A_1 или A_2. Эти алгоритмы, как уже было показано выше для аналогичной ситуации (см. (12.9), (12.10)), обладают минимизирующими свойствами, вступающими в противоречие с построением (12.7).

Легко видеть, что общий случай, когда $A(k_{s(t)}) \in \{A_1, A_2, A_3\}$, сводится к изучению рассмотренных частных случаев. Таким образом, полученные выше противоречия доказывают, что $[\underline{f}, \bar{f}) \subset F^*$. Теперь в силу ограниченности множества D и замкнутости отображения G_f множество F^* замкнуто; поэтому $[\underline{f}, \bar{f}] \subset F^*$. Третье утверждение теоремы доказано.

Пусть теперь F^* не содержит интервалов. Так как $[\underline{f}, \bar{f}] \subset F^*$, то $\underline{f} = \bar{f}$, т. е. числовая последовательность $\{f(x^k)\}$ имеет предел.

Покажем, что все предельные точки $\{x^k\}$ принадлежат X^*. Предположим противное; тогда существует подпоследовательность $\{x^{k_t}\}$ такая, что $\lim\limits_{t\to\infty} x^{k_t} = x' \notin X^*$. Рассмотрим индексы $\{k_{s(t)}\}$ такие, что $k_{s(t)-1} < k_t \leqslant k_{s(t)}$ $(t = 0, 1, \ldots)$. Не ограничивая общности, можем считать, что $\lim\limits_{t\to\infty} x^{k_{s(t)-1}} = x''$.

Заметим, что $A(k_{s(t)-1}) = A_2$ выполнено лишь для конечного числа индексов t, так как в противном случае получаем противоречие со сходимостью $\{f(x^k)\}$.

Если $A(k_{s(t)-1}) = A_3$ для бесконечного множества номеров t, то для тех же достаточно больших t $A(k_{s(t)}) = A_1$. Напомним, что $x' \notin X^*$. Поскольку имеет место либо $\sigma_{k_{s(t)+1}} \geqslant \sigma$, либо $\| g^{k_{s(t)+1}} \| \leqslant \varepsilon_{k_{s(t)+1}}$, то $\rho_{k_{s(t)}} + \ldots + \rho_{k_{s(t)+1}} \geqslant c' > 0$. К последовательности $\{x^{k_{s(t)}}\}$ применима лемма 9.3, утверждение которой противоречит уже доказанной сходимости $\{f(x^k)\}$.

Пусть $A(k_{s(t)-1}) = A_1$ для бесконечного множества индексов t; без ограничения общности можем считать, что это выполнено для всех t. Имеет место $x'' \in X^*$, так как в противном случае, применяя к последовательности начальных точек $\{x^{k_{s(t)-1}}\}$ лемму 9.3, получаем противоречие с уже доказанной сходимостью $\{f(x^k)\}$.

В силу замкнутости множества X^* существует $\delta > 0$ такое, что $\rho(x', X^*) \geqslant \delta > 0$.

Последовательности $\{x^{k_{s(t)-1}}, \ldots, x^{k_t}\}$ $(t = 0, 1, \ldots)$ движутся от x'' к x', причем

$$\lim_{t\to\infty} \| x^{k_{s(t)-1}} - x' \| = \| x'' - x' \|, \quad \lim_{t\to\infty} \| x^{k_t} - x' \| = 0,$$

$$\| x'' - x' \| \geqslant \rho(x', X^*) \geqslant \delta > 0.$$

Числовые последовательности $\{\| x^{k_{s(t)-1}} - x' \|, \| x^{k_{s(t)-1}+1} - x' \|, \ldots$ $\ldots, \| x^{k_t} - x' \|\}$ $(t = 0, 1, \ldots)$ бесконечное число раз пересекают уровень δ; поэтому существует подпоследовательность $\{x^{k_{r(t)}}\}$ такая, что $k_{s(t)-1} < k_{r(t)} < k_t$ и при достаточно больших t $\| x^{k_{r(t)}} - x' \| \geqslant \delta$ и $\| x^{k_{r(t)}+1} - x' \| < \delta$.

Не меняя обозначений, будем считать, что $\lim\limits_{t \to \infty} x^{k_{r(t)}} = x'''$. Так как $\lim\limits_{k \to \infty} \rho_k = 0$, то

$$\lim_{t \to \infty} \| x^{k_{r(t)}} - x' \| = \| x''' - x' \| = \delta < \rho(x', X^*).$$

Отсюда следует, что $x''' \notin X^*$.

Обозначим $\Gamma = \sup\{\|g\| \mid g \in G_f(y),\ y \in D\}$. При достаточно больших t имеем

$$\frac{\delta}{2} \leqslant \| x^{k_t} - x^{k_{r(t)}} \| \leqslant \Gamma \sum_{k=k_{r(t)}}^{k_t - 1} \rho_k;$$

следовательно,

$$\sum_{k=k_{r(t)}}^{k_t - 1} \rho_k \geqslant \frac{\delta}{2\Gamma}.$$

Можно считать, что метод обобщенного градиента как бы многократно стартует из точек $x^{k_{r(t)}} \to x''' \notin X^*$. В этой ситуации опять применима лемма 9.3, из которой, в частности, следует расходимость последовательности $\{f(x^k)\}_{k=0}^{\infty}$, что противоречит ранее установленной сходимости этой последовательности. Теорема полностью доказана.

Алгоритм К2.

Вспомним релаксационные алгоритмы А1—А3 из § 11. Они вырабатывают конечную последовательность приближений к множеству приближенных решений $X^*_{\varepsilon\delta}$ задач (11.1), (11.2). В текущей точке в них отыскиваются направление l убывания минимизируемой функции и точка $y = x^k - \delta l / \| l \|$. На шаге 9 в этих алгоритмах можно полагать $x^{k+1} = y$, а можно, как отмечалось, находить каким-либо вспомогательным алгоритмом точку x^{k+1} такую, что $f(x^{k+1}) \leqslant f(y)$. Этот вспомогательный алгоритм может быть не только алгоритмом линейного поиска в направлении l, но и каким-либо (возможно, эвристическим) алгоритмом глобального поиска в окрестности точек x^k или y.

Таким образом, по существу, релаксационные алгоритмы А1—А3 из § 11 являются комбинированными; в них после каждого релаксационного шага допускается использовать глобальный поиск для нахождения не худшей, чем уже достигнутая точка y, точки x^{k+1}. В § 11 доказана локальная сходимость такого комбинированного алгоритма.

Рассмотрим теперь комбинированный алгоритм, в котором основную роль играет ограниченный глобальный поиск в окрестности текущего приближения.

Пусть необходимо

$$f(x) \to \min$$

при ограничении

$$x \in D \subset E_n,$$

где $f(x)$ — непрерывная функция такая, что множества $\{x \in E_n \mid x \in D, f(x) \leqslant c\}$ ограничены для любого $c \in E_1$.

Пусть $K(x)$ — многозначное отображение такое, что отображение $x \to K(x) \cap D$ полунепрерывно снизу, т. е. из $x^k \to x$ и $y \in K(x) \cap D$ следует существование $y^k \in K(x^k) \cap D$ таких, что $y^k \to y$.

Например, если $D = E_n$ и $K(x)$ полунепрерывно снизу, то $K(x) \cap D = K(x)$ полунепрерывно снизу. В частности, отображение

$$x \to K(x) = \{y \in E_n \mid y - x \in K\},$$

где K — произвольное множество в E_n, полунепрерывно и сверху, и снизу.

Лемма 12.1. *Пусть D — выпуклое множество, $K_r(x)$ — шар (в некоторой норме) радиуса r с центром в точке x. Тогда многозначное отображение $x \to K_r(x) \cap D$ полунепрерывно снизу в D.*

Доказательство. Пусть $x^k \to x$, $x^k \in D$ и $y \in K_r(x) \cap D$. Покажем, что существуют $y^k \in K_r(x^k) \cap D$ такие, что $\{y^k\} \to y$.

Обозначим

$$t_k = \sup\{t \mid 0 \leqslant t \leqslant 1, \quad x^k + t(y - x^k) \in K_r(x^k)\},$$

$$y^k = x^k + t_k(y - x^k) = (1 - t_k)x^k + t_k y.$$

Так как x^k, $y \in D$, то $y^k \in D$. Таким образом, $y^k \in K_r(x^k) \cap D$.

Покажем, что $\lim\limits_{k \to \infty} t_k = 1$; тогда $\| y^k - y \| = (1 - t_k) \| y - x^k \| \to 0$ при $k \to \infty$, и лемма доказана. Предположим противное, тогда существует подпоследовательность $t_{k_s} \leqslant 1 - \varepsilon$, $0{,}5 \leqslant \varepsilon < 1$. Возьмем $t' = 1 - \varepsilon/2$, $t_{k_s} < t'$. Рассмотрим $y' = x^{k_s} + t'(y - x^{k_s})$. Справедлива оценка

$$\| x^{k_s} - y' \| = t' \| y - x^{k_s} \| \leqslant t' (\| y - x \| + \| x^{k_s} - x \|).$$

При $\| x^{k_s} - x \| \leqslant \frac{1 - t'}{t'} \| y - x \|$ имеет место $\| x^{k_s} - y \| \leqslant \| y - x \| \leqslant r$. Следовательно, для достаточно больших s существует $t' > t_{k_s}$ такое, что $y' = x^{k_s} + t'(y - x^{k_s}) \in K_r(x^{k_s})$, что противоречит выбору $t_{k_s} = \sup\{t \mid 0 \leqslant t \leqslant 1, \quad x^{k_s} + t(y - x^{k_s}) \in K_r(x^{k_s})\}$. Лемма доказана.

Рассмотрим следующий идеализированный алгоритм.

Алгоритм К3.

Пусть заданы произвольная точка $x^0 \in D$ и число $r > 0$. Найдем точку x^1 глобального минимума $f(x)$ на множестве $K_r(x^0) \cap D$. Из точек x^1 осуществим спуск любым промежуточным (в частности, тем же самым) алгоритмом так, что получим точку x^2, $f(x^2) \leqslant f(x^1)$. Затем опять находим точку x^3 глобального минимума $f(x)$ на множестве $K_r(x^2) \cap D$. Из нее опять осуществим спуск каким-либо промежуточным алгоритмом, получим точку x^4, $f(x^4) \leqslant f(x^3)$, и т. д.

Теорема 12.2. *Пусть последовательность* $\{x^k\}$ *порождена алгоритмом* КЗ, *где отображение* $x \to K_r(x) \cap D$ *полунепрерывно снизу. Обозначим через* $\{x^{k_m}\}$ *подпоследовательность тех точек* x^k, *в* $K_r(x^k)$*-окрестности которых производился глобальный поиск. Тогда последовательность* $\{x^k\}$ *ограничена, и если* x' *— какая-либо предельная точка* $\{x^{k_m}\}$, *то* x' *является точкой глобального минимума* $f(x)$ *на множестве* $K_r(x') \cap D$.

Доказательство. Алгоритм КЗ релаксационный, т. е. последовательность $\{f(x^k)\}$ монотонно убывает. Последовательность $\{x^k\}$ лежит в ограниченном множестве $\{x \mid x \in D,\ f(x) \leqslant f(x^0)\}$; поэтому $\{f(x^k)\}$ ограничена снизу. Следовательно, существует конечный предел $\lim\limits_{k\to\infty} f(x^k)$.

Теперь предположим противное: существует предельная точка $x' = \lim\limits_{s\to\infty} x^{k_{m_s}}$ последовательности $\{x^{k_m}\}$, не являющаяся точкой глобального минимума $f(x)$ на $K_r(x') \cap D$. Тогда существует точка $x'' \in K_r(x') \cap D$ такая, что $f(x'') \leqslant f(x')$; очевидно, что справедлива оценка $\| x'' - x' \| \leqslant r$. В силу полунепрерывности снизу отображения $x \to K_r(x) \cap D$ существует последовательность $y^s \to x''$, $y^s \in K(x^{k_{m_s}}) \cap D$. Выберем $\varepsilon < f(x') - f(x'')$. Так как $f(x^{k_{m_s}}) \to f(x')$, $f(y^s) \to f(x'')$, то для достаточно больших s

$$f(x^{k_{m_s}}) - f(y^s) \geqslant \varepsilon,$$

откуда $f(x^{k_{m_s}}) \geqslant f(y^s) + \varepsilon \geqslant f(x^{k_{m_s}+1}) + \varepsilon$, что противоречит сходимости последовательности $\{f(x^k)\}$. Теорема доказана.

Комбинированный алгоритм КЗ является идеализированным, поскольку в нем не указаны условия остановки процедур глобального поиска. Восполним этот пробел.

Пусть алгоритм глобальной минимизации $f(y)$ на множестве $K_r(x^k) \cap D$ порождает последовательность приближений $\{y^s\}$ и известна оценка точности приближений:

$$f(y^s) - f_k^* \leqslant \varepsilon_s \to 0, \quad s \to \infty,$$

где $f_k^* = \min\{f(y) \mid y \in K_r(x^k) \cap D\}$; для алгоритмов, рассматриваемых в следующем разделе параграфа, такая оценка имеется. Тогда можно ограничиться нахождением $y^s \in K_r(x^k) \cap D$ такого, что

$$\varepsilon_s \leqslant \alpha(f(x^k) - f(y^s)), \qquad \alpha \geqslant 0.$$

Из неравенства

$$f(y^s) - f_k^* \leqslant \varepsilon_s \leqslant \alpha(f(x^k) - f(y^s))$$

с учетом $f(x^{k+1}) \leqslant f(y^s)$ следует

$$f(x^{k+1}) \leqslant f(y^s) \leqslant \frac{\alpha}{\alpha+1} f(x^k) + \frac{1}{\alpha+1} f_k^*.$$

Этого достаточно, чтобы в рассматриваемом случае сохранила силу теорема 12.2.

2. **Метод аппроксимаций.** В этом пункте мы распространим метод невыпуклых аппроксимаций многоэкстремальных задач [29, 95, 96] на многомерные многоэкстремальные задачи с невыпуклыми ограничениями и наметим некоторые пути решения возникающих при этом вспомогательных многоэкстремальных задач специального вида.

Пусть $f(x)$ $(x \in D \subset E_n)$ — непрерывная функция. Предположим, что имеется семейство $\varphi(y, x)$ $(y \in D)$ *нижних аппроксимаций функции* $f(x)$ такое, что

1) $f(x) \geqslant \varphi(y, x)$ для всех $y, x \in D$;

2) $f(y) = \varphi(y, y)$ для всех $y \in D$;

3) $\varphi(y, x)$ локально липшицевы по x равномерно по y, т. е. для любого компакта $K \subset D$ существует константа L_K такая, что для любых $x, y, z \in K$ имеет место

$$|\varphi(y, x) - \varphi(y, z)| \leqslant L_K \|x - z\|.$$

Если $f(x)$ $(x \in D)$ дифференцируема, причем ее градиент $\nabla f(x)$ удовлетворяет условию Липшица в D с константой L, то в качестве нижней аппроксимации можно взять касательный к графику $f(x)$ в точке y параболоид:

$$\varphi(y, x) = f(y) + \frac{1}{2L}\|\nabla f(y)\|^2 - \frac{L}{2}\left\|x - y - \frac{1}{L}\nabla f(y)\right\|^2. \qquad (12.4')$$

Если $f(x)$ липшицева в D с константой Липшица L, то в качестве нижних аппроксимаций $f(x)$ можно взять функции

$$\varphi(y, x) = f(y) - L\|y - x\|. \qquad (12.5')$$

Опишем метод невыпуклых аппроксимаций [96] для решения следующей многоэкстремальной задачи:

$$f(x) \to \min \qquad (12.6')$$

при ограничении

$$x \in D \subset E_n, \qquad (12.7')$$

где $f(x)$ — непрерывная на D функция, имеющая нижние аппроксимации $\varphi(y, x)$, удовлетворяющие условиям 1)—3), D — замкнутое ограниченное множество.

Возьмем произвольную точку $x^0 \in D$. Найдем точку $x^1 \in D$ с возможно меньшим значением функции $\varphi_0(x) = \varphi(x^0, x)$. Затем найдем точку $x^2 \in D$ с возможно меньшим значением функции $\varphi_1(x) = \max(\varphi(x^0, x), \varphi(x^1, x))$ и т. д.

Обозначим

$$\varphi_k(x) = \max_{0 \leqslant i \leqslant k} \varphi(x^i, x).$$

В описанном методе необходимо решать последовательность специальных многоэкстремальных задач вида

$$\varphi_k(x) \to \min \qquad (12.8')$$

при ограничении

$$x \in D \subset E_n. \qquad (12.9')$$

Лемма 12.2. *Имеет место*

$$\lim_{k \to \infty} [f(x^k) - \varphi_{k-1}(x^k)] = 0.$$

Теорема 12.3. *Пусть точки x^k являются точками глобального минимума функций $\varphi_{k-1}(x)$ $(k = 1, 2, \ldots)$. Тогда все предельные точки $\{x^k\}_{k=0}^{\infty}$ являются точками глобального минимума $f(x)$ на D.*

Лемма 12.2 и теорема 12.3 в более общей ситуации будут доказаны ниже.

Замечание 12.2. Величина $\varepsilon_k = f(x^k) - \varphi_{k-1}(x^k)$ дает оценку точности приближения x^k:

$$f(x^k) - \min_{x \in D} f(x) \leqslant f(x^k) - \varphi_{k-1}(x^k).$$

Распространим описанный метод для решения задач следующего вида:

$$f_0(x) \to \min$$

при ограничениях

$$f_i(x) \leqslant 0, \quad i = 1, 2, \ldots, m, \quad x \in D \subset E_n,$$

где $f(x)$ и $f_i(x)$ $(i = 1, 2, \ldots, m)$ — непрерывные функции, имеющие нижние аппроксимации $\varphi_i(y, x)$ $(i = 0, 1, \ldots, m)$; множество D замкнуто и ограничено.

Мы сразу свернем все ограничения-неравенства в одно (хотя это и не обязательно), т. е. введем функцию

$$h(x) = \max_{1 \leqslant i \leqslant m} f_i(x)$$

и рассмотрим эквивалентную задачу:

$$f(x) \to \min \qquad (12.10')$$

при ограничениях

$$h(x) \leqslant 0, \quad x \in D. \qquad (12.11)$$

Заметим, что если

$$h(y) = \max_{1 \leqslant i \leqslant m} f_i(y) = f_{i^*(y)}(y) \qquad (12.12)$$

и $\varphi_{i^*(y)}(y, x)$ — нижняя аппроксимация функции $f_{i^*(y)}(x)$, то

$$\psi(y, x) = \varphi_{i^*(y)}(y, x) \qquad (12.13)$$

также нижняя аппроксимация функции $h(x)$. В дальнейшем будем считать известными нижние аппроксимации $\psi(y, x)$.

Метод решения задачи (12.10), (12.11) заключается в построении последовательности точек $\{x^k\} \subset D$ по следующему правилу.

Точка $x^0 \in D$ произвольна. Пусть уже построены точки $x^0, x^1, \dots$ $\dots, x^k \in D$. Обозначим

$$\varphi_k(x) = \max_{0 \leqslant i \leqslant k} \varphi(x^i, x),$$

$$\psi_k(x) = \max_{0 \leqslant i \leqslant k} \{\psi(x^i, x) \mid h(x^i) > 0\}.$$

Тогда точку x^{k+1} найдем как решение следующей многоэкстремальной задачи специального вида:

$$\varphi_k(x) \to \min \tag{12.14}$$

при ограничениях

$$\psi_k(x) \leqslant 0, \quad x \in D. \tag{12.15}$$

Лемма 12.3. *Если исходная задача* (12.10), (12.11) *не имеет допустимых решений, то при некотором k и аппроксимирующая задача* (12.14), (12.15) *не имеет решений. Если исходная задача имеет допустимые решения, то последовательность $\{x^k\}$ бесконечна, и все предельные точки $\{x^k\}$ принадлежат допустимой области* (12.11).

Доказательство. Покажем, что если последовательность $\{x^k\}$ допустимых решений задач (12.14), (12.15) бесконечна, то все предельные точки $\{x^k\}$ принадлежат допустимой области задачи (12.10), (12.11). В этом случае область (12.11) не пуста. Таким образом, если областб (12.11) пуста, то последовательность $\{x^k\}$ не может быть бесконечной, т. е. при некотором k аппроксимирующая задача (12.14), (12.15) не имеет допустимых решений.

Итак, пусть последовательность точек $\{x^k\}$ из ограниченного множества D бесконечна. Предположим, что существует подпоследовательность $\{x^{k_m}\}$ такая, что

$$\lim_{m \to \infty} x^{k_m} = x', \qquad h(x') > 0.$$

Последовательности $\{\psi_k(x)\}$ $(x \in D)$ монотонно возрастают и ограничены сверху, поэтому существует $\psi(x) = \lim_{k \to \infty} \psi_k(x)$. Функции $\psi(x^k, x)$, а следом и $\psi_k(x)$ липшицевы в D равномерно по k; поэтому функция $\psi(x)$ также липшицева в D.

Теперь по построению $\psi_{k_m - 1}(x^{k_m}) \leqslant 0$; переходя здесь к пределу по m, получим $\psi(x') \leqslant 0$. Но с другой стороны, $\psi(x^{k_m}) \geqslant \psi_{k_m}(x^{k_m}) = h(x^{k_m})$ при достаточно больших m. Переходя здесь к пределу по m, имеем $\psi(x') \geqslant h(x') > 0$. Полученное противоречие доказывает лемму.

Лемма 12.4. *Пусть последовательность $\{x^k\}$ допустимых решений задачи* (12.14), (12.15) *бесконечна. Тогда*

$$\lim_{k \to \infty} [f(x^k) - \varphi_{k-1}(x^k)] = 0.$$

Доказательство. Предположим противное. Тогда существует $\{x^{k_m}\}$ такая, что $\lim_{m\to\infty} x^{k_m} = x'$ и

$$\lim_{m\to\infty} [f(x^{k_m}) - \varphi_{k_m-1}(x^{k_m})] = \varepsilon > 0.$$

Последовательности $\{\varphi_k(x)\}$ $(x \in D)$ монотонно возрастают и ограничены сверху функцией $f(x)$; поэтому существует $\varphi(x) = \lim_{k\to\infty} \varphi_k(x)$. Функции $\{\varphi_k(x)\}$ очевидно, липшицевы в D равномерно по k; поэтому $\varphi(x)$ также липшицева в D. Тогда

$$\lim_{m\to\infty} [f(x^{k_m}) - \varphi_{k_m-1}(x^{k_m})] = f(x') - \varphi(x') = \varepsilon > 0.$$

Но с другой стороны, $\varphi(x^{k_m}) \geqslant \varphi_{k_m}(x^{k_m}) = f(x^{k_m})$; переходя здесь к пределу по m, имеем $\varphi(x') \geqslant f(x')$. Полученное противоречие доказывает лемму.

Доказательство леммы 12.2 не отличается от доказательства леммы 12.4.

Теорема 12.4. *Пусть точка x^{k+1} является точкой глобального минимума задачи* (12.14), (12.15); *последовательность $\{x^k\}$ бесконечна. Тогда все предельные точки $\{x^k\}$ являются точками глобального решения задачи* (12.10), (12.11).

Доказательство. Предположим противное. Тогда существует подпоследовательность $\{x^{k_m}\}$ такая, что $\lim_{m\to\infty} x^{k_m} = x'$, и для некоторой допустимой точки x'' задачи (12.10), (12.11) имеет место $f(x') > f(x'')$. По лемме 12.4

$$\lim_{m\to\infty} \varphi_{k_m-1}(x^{k_m}) = f(x') > f(x'').$$

Но с другой стороны, $\varphi_{k_m-1}(x^{k_m}) \leqslant f(x'')$ и $\overline{\lim_{m\to\infty}} \varphi_{k_m-1}(x^{k_m}) \leqslant f(x'')$. Полученное противоречие доказывает теорему.

Теорема 12.3 доказывается аналогично.

В рассматриваемом методе необходимо решать последовательность специальных многоэкстремальных задач (12.8), (12.9) или (12.14), (12.15). Наметим некоторые пути их решения. Для этого нам придется преобразовывать экстремальные задачи.

Будем называть две задачи *эквивалентными*, если известен алгоритм, позволяющий по решению одной задачи восстановить решение другой. Эквивалентные задачи одновременно либо имеют решения, либо не имеют решения.

Две задачи на минимум, имеющие ограниченные решения, заведомо эквивалентны, если известен алгоритм, позволяющий для любой допустимой точки одной задачи указать допустимую точку другой с не большим значением целевой функции.

Действительно, тогда оптимальные значения целевых функций задач совпадают, и существующие по условию соответствия между точками допустимых областей позволяют восстанавливать решения одной задачи по решению другой.

Рассмотрим вспомогательные задачи (12.8), (12.9).

Пусть $f(x)$ в (12.6) допускает нижние аппроксимации вида (12.4) с некоторой константой L. В частности, если $f(x)$ является функцией максимума, то нижние аппроксимации могут строиться по формулам типа (12.12), (12.13).

Теперь задача (12.8), (12.9) имеет вид:

$$\max_{1\leqslant i\leqslant k}\left(a_i - \frac{L}{2}\|x-b^i\|^2\right)\to\min_x$$

при ограничении

$$x\in D\subset E_n,$$

где

$$a_i = f(x^i) + \frac{1}{2L}\|\nabla f(x^i)\|^2, \quad b^i = x^i + \frac{1}{L}\nabla f(x^i).$$

Эта задача эквивалентна следующей:

$$\Phi(x, z_1)\equiv z_1\to\min_{x,z_1}$$

при ограничениях

$$L(b^i, x) + a_i - \frac{L}{2}\|b^i\|^2 - \frac{L}{2}\|x\|^2\leqslant z_1,$$

$$i = 1, 2, \ldots, k; \qquad x\in D.$$

Делая замену переменных $z = z_1 + \frac{L}{2}\|x\|^2$, получаем следующую эквивалентную задачу:

$$z - \frac{L}{2}\|x\|^2\to\min_{x,z} \tag{12.16}$$

при ограничениях

$$L(b^i, x) - z\leqslant\frac{L}{2}\|b^i\|^2 - a_i, \qquad i = 1, 2, \ldots, k; \tag{12.17}$$

$$x\in D. \tag{12.18}$$

Если множество D задается линейными ограничениями, то имеем многоэкстремальную задачу минимизации вогнутой квадратичной функции на множестве, задаваемом линейными ограничениями. В этом случае все локальные минимумы задачи (12.16)—(12.18) находятся в вершинах многогранника (12.17), (12.18). Некоторые алгоритмы минимизации вогнутой функции на линейном многограннике, основанные на направленном переборе вершин и отсечении бесперспективных областей многогранника, рассмотрены в работах [8, 126, 188].

Заметим, что находить локальные минимумы задачи (12.16)—(12.18) сравнительно легко.

Построим последовательность $\{y^s\}$, оканчивающуюся в каком-либо локальном минимуме задачи. Точка $y^0\in D$ произвольна. Пусть уже построена точка y^s. Линеаризуем целевую функцию (12.16) в точке y^s

и найдем y^{s+1} как решение следующей задачи линейного программирования:

$$z - L(y^s, y) \to \min_{y,z} \tag{12.19}$$

при ограничениях

$$L(b^i, y) - z \leqslant \frac{L}{2} \| b^i \|^2 - a_i, \qquad i = 1, 2, \ldots, k; \tag{12.20}$$

$$y \in D. \tag{12.21}$$

Поскольку в последовательности задач (12.19)—(12.21) меняется только целевая функция, то оптимальное решение одной задачи является допустимым для следующей. Здесь важное значение имеет выбор точки $y^0 \in D$.

Спустившись в локальный минимум задачи (12.16)—(12.18), нужно каким-либо способом, например перебором окрестных вершин многогранника или методами отсечений [8, 126, 188], найти точку с меньшим, чем уже достигнутое, значением целевой функции (12.6) и затем опять спуститься в новый локальный экстремум и т. д.

По существу, мы описали идею комбинированного алгоритма для решения задачи (12.16)—(12.18).

К сожалению, в других случаях не удается простыми преобразованиями привести вспомогательные задачи (12.8), (12.9) или (12.14), (12.15) к известному виду. Однако мы сделаем некоторые преобразования этих задач.

Пусть $f(x)$ и $h(x)$ в (12.10), (12.11) допускают нижние аппроксимации вида (12.4), (12.13) с константами L_f и L_h. Тогда вспомогательная задача (12.14), (12.15) имеет вид

$$\max_{1 \leqslant i \leqslant k} \left[a_i - \frac{1}{2} L_f \| x - b^i \|^2 \right] \to \min_x$$

при ограничениях

$$\max_{\{i | 1 \leqslant i \leqslant k, h(x^i) > 0\}} \left[c_i - \frac{1}{2} L_h \| x - d^i \|^2 \right] \leqslant 0, \qquad x \in D,$$

где

$$a_i = f(x^i) + \frac{1}{2L_f} \| \nabla f(x^i) \|^2, \qquad b^i = x^i + \frac{1}{L_f} \nabla f(x^i),$$

$$c_i = h(x^i) + \frac{1}{2L_h} \| \nabla f_{j^*(x^i)}(x^i) \|^2, \quad d^i = x^i + \frac{1}{L_h} \nabla f_{j^*(x^i)}(x^i),$$

$$h(x^i) = \max_{1 \leqslant j \leqslant m} f_j(x^i) = f_{j^*(x^i)}(x^i).$$

Эта задача эквивалентна следующей:

$$W(x, t, z) \equiv z \to \min_{x,t,z} \tag{12.22}$$

при ограничениях

$$L_f(b^i, x) - \frac{1}{2} L_f t - z \leqslant \frac{1}{2} L_f \| b^i \|^2 - a_i, \qquad i = 1, 2, \ldots, k; \quad (12.23)$$

$$L_h(d^j, x) - \frac{1}{2} L_h t \leqslant \frac{1}{2} L_h \| d^j \|^2 - c_j,$$

$$j \in \{i \mid 1 \leqslant i \leqslant k, \ h(x^i) > 0\}; \quad (12.24)$$

$$x \in D, \qquad \| x \|^2 \geqslant t > 0; \quad (12.25)$$

здесь переменными являются x, t, z и имеется только одно нелинейное вогнутое ограничение $\| x \|^2 \geqslant t > 0$.

Пусть теперь минимизируемая многоэкстремальная функция $f(x)$ в задаче (12.6), (12.7) удовлетворяет в D условию Липшица с константой L и нижние аппроксимации $\varphi(x^i, x)$ имеют вид (12.5). В этом случае вспомогательные задачи (12.8), (12.9) имеют вид

$$\max_{1 \leqslant i \leqslant k} (f_i - L \| x - x^i \|) \to \min_x$$

при ограничении

$$x \in D,$$

где $f_i = f(x^i)$.

Эта задача эквивалентна следующей:

$$V(x, z) \equiv z \to \min_{x, z}$$

при ограничениях

$$f_i - L \| x - x^i \| \leqslant z, \qquad i = 1, 2, \ldots, k;$$

$$z \leqslant \min_{1 \leqslant i \leqslant k} f_i, \qquad x \in D.$$

В свою очередь эта задача эквивалентна следующей:

$$V(x, z) \equiv z \to \min_{x, z}$$

при ограничениях

$$2L^2(x^i, x) - 2f_i z + z^2 - L^2 \| x \|^2 \leqslant c_i^2, \qquad i = 1, 2, \ldots, k;$$

$$z \leqslant \min_{1 \leqslant i \leqslant k} f_i; \qquad x \in D;$$

где $c_i = L^2 \| x^i \|^2 - f_i^2$.

Вводя дополнительную переменную t, приходим к новой эквивалентной задаче:

$$U(x, z, t) \equiv z \to \min_{x, z, t} \quad (12.26)$$

при ограничениях

$$2L^2(x^i, x) - 2f_i z + z^2 - Lt \leqslant c_i, \qquad i = 1, \ldots, k; \quad (12.27)$$

$$z \leqslant \min_{1 \leqslant i \leqslant k} f_i, \qquad x \in D; \quad (12.28)$$

$$\| x \|^2 \geqslant t \geqslant 0; \quad (12.29)$$

здесь переменными являются x, z, t и имеется только одно нелинейное вогнутое ограничение $\|x\|^2 \geqslant t \geqslant 0$.

Рассмотрим преобразованные вспомогательные задачи (12.22)—(12.25), (12.26)—(12.29). Обозначим оптимальное значение целевой функции в них через z^*. Линеаризуя в y невыпуклое ограничение $t - \|x\|^2 \leqslant 0$, получим $t - 2(y, x) + \|y\|^2 \leqslant 0$. Подставив последнее вместо ограничения $t - \|x\|^2 \leqslant 0$ в задачи (12.22)—(12.25), (12.26)—(12.29), получим задачи линейного и выпуклого программирования. Обозначим оптимальное решение таких задач через $z^*(y)$. Так как $t - \|x\|^2 \leqslant t - 2(y, x) + \|y\|^2$, то $z^* \leqslant z^*(y)$. Имеет место

$$\{(t, x) \in E_{n+1} \mid x \in D,\ t - \|x\|^2 \leqslant 0\} =$$

$$= \bigcup_{y \in D} \{(t, x) \mid x \in D,\ t - 2(y, x) + \|y\|^2 \leqslant 0\},$$

поэтому $z^* = \inf_{y \in D} z^*(y)$. Таким образом, взяв сетку точек $\{y^s\} \subset D$ и решив набор преобразованных вспомогательных задач с линеаризованным в точках y^s невыпуклым ограничением $t - \|x\|^2 \leqslant 0$, получим оценку оптимального значения z^*. Для разных y^s задачи (12.22)—(12.25) с линеаризованным ограничением $t - \|x\|^2 \leqslant 0$ различаются только одним линейным ограничением, поэтому их целесообразно решать двойственным симплекс-методом.

3. **Метод сглаживания.** Сглаживание или усреднение функций давно используется в математике. В частности, в данной книге оно широко используется для построения методов локальной минимизации липшицевых функций. Для целей глобальной оптимизации оно стало применяться начиная с работы [138]. Дело в том, что сглаживание (усреднение) функции устраняет мелкие локальные экстремумы, сохраняя общую картину рельефа функции. Минимизируя сглаженную функцию, мы надеемся проскочить локальные экстремумы исходной функции и попасть в район глобального минимума.

Пусть необходимо

$$f(x) \to \min_{x \in E_n},$$

где $f(x)$ — непрерывная функция такая, что $f(x) \to +\infty$ при $\|x\| \to + \infty$.

Обозначим куб с гранью h и центром x как

$$K_h(x) = \{y \in E_n \mid \max_{1 \leqslant i \leqslant n} |y_i - x_i| \leqslant h/2\},$$

$S_h(x)$ — его поверхность, $N(y)$ — внешняя нормаль к $S_h(x)$ в точке $y \in S_h(x)$.

Рассмотрим сглаженные (усредненные) функции:

$$f_h(x) = \frac{1}{h^n} \int_{K_h(x)} f(y)\, dy_1 dy_2 \ldots dy_n.$$

Градиент $f_h(x)$, когда $f(x)$ непрерывна, вычисляется поверхностным интегралом

$$\nabla f_h(x) = \frac{1}{h^n} \int\limits_{S_h(x)} f(y) N(y)\, dS.$$

Метод глобальной оптимизации, основанный на сглаживании, заключается в следующем.

Начиная с произвольной точки $x^0 \in E_n$, минимизируем сглаженную функцию $f_{h_1}(x)$ с достаточно большим параметром сглаживания h_1 до точности $\|\nabla f_{h_1}(x^1)\| \leqslant \varepsilon_1$. Затем, начиная с x^1, минимизируем функцию $f_{h_2}(x)$ с меньшим параметром сглаживания $h_2 < h_1$ до точности $\|\nabla f_{h_2}(x^2)\| \leqslant \varepsilon_2$ и т. д.

В этом методе

$$\lim_{k\to\infty} h_k = \lim_{k\to\infty} \varepsilon_k = 0.$$

Когда функция $f(x)$ непрерывно дифференцируема, отображение $(x, h) \to \nabla f_h(x)$ непрерывно. Поэтому последовательность $\{x^k\}$ не может сходиться своими предельными точками к нестационарным точкам $f(x)$. Используя лемму 2.5, аналогичное свойство можно доказать и для локально липшицевых функций $f(x)$.

Однако каковы же глобальные свойства метода? В этой связи интересно следующее наблюдение [81].

Пусть $f(x)$ зависит от скалярной переменной x. Рассмотрим сглаженные функции ($h \geqslant 0$)

$$f_h(x) = \frac{1}{h} \int\limits_{x-h/2}^{x+h/2} f(y)\, dy, \qquad f_0(x) = f(x).$$

Их градиенты в этом случае имеют вид

$$\nabla f_h(x) = \left[f\left(x + \frac{h}{2}\right) - f\left(x - \frac{h}{2}\right)\right] h^{-1}, \qquad h > 0.$$

Построим картину стационарных точек сглаженных функций $f_h(x)$ в плоскости $E_x \times E_h = \{(x, h) | x \in E_x, h \in E_h\}$, т. е. построим множество $T = \{(x, h) \in E_x \times E_h | \nabla f_h(x) = 0\}$.

Заметим, что если $f(x^1) = f(x^2)$ $(x^1 \neq x^2)$, то точка $((x^1 + x^2)/2, |x^2 - x^1|) \in T$. Таким образом, для построения T достаточно проследить за всеми параллельными оси E_x отрезками, концы которых лежат на графике $f(x)$. Непрерывно перемещая эти отрезки вверх или вниз и следя за их длиной, легко построить T.

Картина множества T наиболее прозрачна, когда все экстремумы $f(x)$ различны. Тогда T состоит из непрерывных непересекающихся линий или полос, связанных с экстремумами на оси E_x. Причем любая линия, начинающаяся в локальном экстремуме $(x^*, 0)$ (на оси E_x), ограничена и замыкается на другой локальный экстремум $(x^{**}, 0)$ (на оси E_x). К стационарным неэкстремальным точкам линии не подходят. И только линия, начинающаяся в глобальном минимуме $(x_{\min}, 0)$, уходит на бесконечность в плоскости $E_x \times E_h$.

Если не все экстремумы $f(x)$ различны, то картина T несколько сложнее. Однако по-прежнему T распадается на связные ограниченные компоненты, соединяющие локальные экстремумы, и неограниченную связную компоненту, соединяющую все глобальные минимумы $f(x)$ и бесконечность, а также главные максимумы между соседними глобальными минимумами.

Минимизируя $f_h(x)$ при достаточно большом параметре сглаживания h, мы, очевидно, попадем на стационарную линию из T, ведущую в глобальный минимум. Однако непросто пройти по этой линии в глобальный экстремум, поскольку она может вести себя весьма причудливо, в частности может иметь изломы. Точки на этой линии могут быть локальными минимумами, локальными максимумами или просто стационарными по x для соответствующих сглаженных функций $f_h(x)$.

ГЛАВА 4

НЕМОНОТОННЫЕ МЕТОДЫ С УСРЕДНЕНИЕМ НАПРАВЛЕНИЙ СПУСКА

В последнее время большое внимание привлекают релаксационные методы негладкой оптимизации, которые, с одной стороны, вытекают из методов сопряженных градиентов гладкой оптимизации, а с другой — обобщают известные методы минимизации функции максимума. Основная черта их заключается в том, что направление спуска определяется усредненными субградиентами, вычисленными в некоторой окрестности текущей точки. Такие алгоритмы еще называют *ε-субградиентными методами*. В этих методах основным и весьма сложным вопросом является построение направления значительного убывания функции.

В данной главе исследуются немонотонные методы минимизации недифференцируемых функций с усреднением конечных разностей или обобщенных градиентов. Немонотонные методы, как правило, не сложнее итерационных процедур, рассмотренных в предыдущих главах. Для их построения не требуется дополнительной памяти и сложных регулировок параметров. Следует отметить, что оптимальный быстросходящийся метод овражного шага [74] был построен в классе немонотонных методов. В § 14 будет показано, что по своей структуре он близок к методам усредненных градиентов. В дальнейшем процедуры усреднения вида

$$z^{k+1} = (1 - a_k) z^k + a_k H(x^k, \alpha_k)$$

будут широко использоваться в гл. 5 при построении численных методов минимизации липшицевых функций с ограничениями, а также в гл. 7 для решения стохастических экстремальных задач.

§ 13. Методы с усреднением направлений спуска

В гл. 2 изучались конечно-разностные методы минимизации липшицевых функций. Им присущи недостатки обычных одношаговых процедур. Чтобы сделать процесс поиска экстремума более регулярным, целесообразно усреднять векторы направлений спуска, которые вычисляются на предыдущих итерациях. Усреднение направлений выгодно проводить в задачах минимизации функций овражного типа, поскольку среднее направление указывает движение вдоль оврага.

Пусть требуется минимизировать локально липшицеву функцию $f(x)$. Определим последовательность точек

$$x^{k+1} = x^k - \rho_k \sum_{t=r_k}^{k} H(x^t, \alpha_t). \tag{13.1}$$

Здесь $\{r_k\}$ — последовательность моментов (в которые происходит возобновление суммирования), определяющая число усредняемых векторов. Векторы $H(x^k, \alpha_k)$ определяются по одной из следующих формул:

$$H(x^k, \alpha_k) = \frac{1}{2\alpha_k} \sum_{i=1}^{n} [f(\tilde{x}_1^k, \ldots, x_i^k + \alpha_k, \ldots, \tilde{x}_n^k) -$$

$$- f(\tilde{x}_1^k, \ldots, x_i^k - \alpha_k, \ldots, \tilde{x}_n^k)] e_i, \qquad (13.2)$$

$$H(x^k, \alpha_k) = \sum_{i=1}^{n} \frac{f(\tilde{x}^k + \Delta_k e_i) - f(\tilde{x}^k)}{\Delta_k} e_i, \qquad (13.3)$$

$$H(x^k, \alpha_k) = \sum_{i=1}^{p} \frac{f(\tilde{x}^k + \Delta_k \mu_i^k) - f(\tilde{x}^k)}{\Delta_k} \mu_i^k, \qquad (13.4)$$

аналогичных соответственно формулам (4.6), (4.7), (6.1).

В методе обобщенного градиентного спуска

$$H(x^k, \alpha_k) = g(\tilde{x}^k). \qquad (13.5)$$

В практических расчетах усреднение векторов $H(x^k, \alpha_k)$ проводится по тем точкам, которые содержатся в ε_k-окрестности точки x^{r_k}, т. е. следующее восстановление происходит, когда

$$\| x^k - x^{r_k} \| > \varepsilon_k.$$

Для того чтобы не учитывать старую информацию, устремляют ε_k к нулю. В дальнейшем будем считать, что число усредненных векторов ограничено некоторой величиной M, т. е. $k - r_k \leqslant M$.

Теорема 13.1. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < \infty, \quad \frac{\rho_k}{\alpha_k} \to 0, \quad \frac{\Delta_k}{\alpha_k} \to 0,$$

$$\frac{|\alpha_k - \alpha_{k+1}|}{\rho_k} \to 0, \quad \alpha_k \to 0, \quad \lim_{k\to\infty} \frac{\alpha_{k-1}}{\alpha_k} = 1,$$

$$\overline{\lim_{k\to\infty}} \frac{\rho_k}{\rho_{k+1}} \leqslant C < \infty.$$

Тогда предельные точки последовательности $\{x^k\}$ с вероятностью 1 принадлежат множеству

$$X^* = \{x^* \mid 0 \in \partial f(x^*)\},$$

последовательность $\{f(x^k)\}$ с вероятностью 1 сходится.

Доказательство во многом совпадает с доказательством теоремы 4.1. Из теоремы о среднем имеем

$$f(x^{k+1}, \alpha_k) = f(x^k, \alpha_k) + (\nabla f(x^k + \tau(x^{k+1} - x^k), \alpha_k), \ x^{k+1} - x^k) - $$
$$- \frac{1}{k - r_k + 1}\left(\sum_{t=r_k}^{k} \nabla f(x^t, \alpha_t), \ x^{k+1} - x^k\right) +$$
$$+ \frac{1}{k - r_k + 1}\left(\sum_{t=r_k}^{k} \nabla f(x^t, \alpha_t), \ x^{k+1} - x^k\right), \qquad 0 \leqslant \tau \leqslant 1.$$

Оценим величины

$$\| \nabla f(x^k + \tau(x^{k+1} - x^k), \alpha_k) - \nabla f(x^i, \alpha_i)\|, \qquad i = r_k, \ldots, k.$$

Заметим, что

$$\| \nabla f(y, \alpha_k) - \nabla f(y, \alpha_{r_k})\| \leqslant C\left[\frac{|\alpha_k - \alpha_{r_k}|}{\alpha_{r_k}} + \frac{|\alpha_k - \alpha_{r_k}|}{\alpha_k}\right],$$
$$\| \nabla f(y, \alpha_k) - \nabla f(y, \alpha_{k-1})\| \leqslant C\left[\frac{|\alpha_k - \alpha_{k-1}|}{\alpha_{k-1}} + \frac{|\alpha_k - \alpha_{k-1}|}{\alpha_k}\right],$$
$$|f(y, \alpha_{k+1}) - f(y, \alpha_k)| \leqslant C\,|\alpha_{k+1} - \alpha_k|.$$

Поэтому

$$\| \nabla f(x^k + \tau(x^{k+1} - x^k), \alpha_k) - \nabla f(x^i, \alpha_i)\| =$$
$$= \| \nabla f(x^k + \tau(x^{k+1} - x^k), \alpha_k) - \nabla f(x^i, \alpha_k) + \nabla f(x^i, \alpha_k) -$$
$$- \nabla f(x^i, \alpha_i)\| \leqslant \frac{C}{\alpha_k}\sum_{t=i}^{k} \rho_t + C\left[\frac{|\alpha_k - \alpha_i|}{\alpha_i} + \frac{|\alpha_k - \alpha_i|}{\alpha_k}\right],$$
$$\| \nabla f(x^k + \tau(x^{k+1} - x^k), \alpha_k) - \nabla f(x^k, \alpha_k)\| \leqslant \frac{C\rho_k}{\alpha_k}.$$

Отсюда вытекает, что

$$f(x^{k+1}, \alpha_{k+1}) \leqslant f(x^k, \alpha_k) + C\,|\alpha_k - \alpha_{k+1}| +$$
$$+ C\frac{\rho_k}{\alpha_k}\sum_{t=r_k}^{k} \rho_t + C\rho_k \sum_{t=r_k}^{k-1}\left[\frac{|\alpha_t - \alpha_k|}{\alpha_t} + \frac{|\alpha_t - \alpha_k|}{\alpha_k}\right] -$$
$$- \frac{\rho_k}{k - r_k + 1}\left(\sum_{t=r_k}^{k} \nabla f(x^t, \alpha_t), \ \sum_{t=r_k}^{k} (H(x^t, \alpha_t) - \nabla f(x^t, \alpha_t) +\right.$$
$$\left. + \nabla f(x^t, \alpha_t))\right) \leqslant f(x^k, \alpha_k) + C\,|\alpha_k - \alpha_{k+1}| + C\frac{\rho_k}{\alpha_k}\sum_{t=r_k}^{k} \rho_t +$$

$$+ C\rho_k \sum_{t=r_k}^{k-1} \left[\frac{|\alpha_t - \alpha_k|}{\alpha_t} + \frac{|\alpha_t - \alpha_k|}{\alpha_k} \right] -$$

$$- \frac{\rho_k}{k - r_k + 1} \left\| \sum_{t=r_k}^{k} \nabla f(x^t, \alpha_t) \right\|^2 + \frac{\rho_k}{k - r_k + 1} \left(\sum_{t=r_k}^{k} \nabla f(x^t, \alpha_t), \right.$$

$$\left. \sum_{t=r_k}^{k} (\nabla f(x^t, \alpha_t) - H(x^t, \alpha_t)) \right). \tag{13.6}$$

Для доказательства сходимости достаточно проверить выполнение условий (4.12), (4.13).

Пусть $x^s \to x' \notin X^*$ $(s \in S)$ и $2\bar{\delta}$-окрестности точек x^s $(s \geqslant \bar{s})$ не пересекаются с X^*. Предположим, что условие (4.12) не выполняется, т. е. все точки x^k $(k \geqslant s \geqslant \bar{s})$ содержатся в δ-окрестности точки x^s, $\delta \leqslant \bar{\delta}/2$. Поскольку $\rho_k \to 0$, то для достаточно больших s точки x^{s-M}, x^{s-M+1}, ..., x^{s-1} также принадлежат δ-окрестности x^s. Поэтому из леммы 2.5 следует, что при достаточно больших s и довольно малых δ выполняется неравенство

$$\left\| \sum_{t=r_k}^{k} \nabla f(x^t, \alpha_t) \right\| \geqslant \sigma > 0.$$

Из условий теоремы следует, что

$$\frac{\rho_{r_k}}{\alpha_k} = \frac{\rho_{r_k}}{\rho_{r_k+1}} \times \ldots \times \frac{\rho_k}{\alpha_k} \to 0;$$

аналогично

$$\frac{\rho_{r_k+1}}{\alpha_k} \to 0, \ \ldots, \ \frac{\rho_{k-1}}{\alpha_k} \to 0.$$

Точно так же

$$\frac{|\alpha_i - \alpha_k|}{\alpha_i} \to 0, \quad \frac{|\alpha_i - \alpha_k|}{\alpha_k} \to 0, \quad i = r_k, \ldots, k.$$

Следовательно, при достаточно больших k

$$C|\alpha_k - \alpha_{k+1}| + \frac{C}{\alpha_k} \sum_{t=r_k}^{k} \rho_t + C \sum_{t=r_k}^{k-1} \left[\frac{|\alpha_t - \alpha_k|}{\alpha_t} + \frac{|\alpha_t - \alpha_k|}{\alpha_k} \right] \leqslant \frac{\sigma}{2M}.$$

Тогда в силу сходимости с вероятностью 1 ряда

$$\sum_{k=0}^{\infty} \rho_k \left(\sum_{t=r_k}^{k} \nabla f(x^t, \alpha_t), \sum_{t=r_k}^{k} (\nabla f(x^t, \alpha_t) - H(x^t, \alpha_t)) \right)$$

из неравенства (13.6) при больших s вытекает

$$f(x^k, \alpha_k) \leqslant f(x^s, \alpha_s) - \frac{\sigma}{2M} \sum_{t=s}^{k-1} \rho_t, \tag{13.7}$$

что при $k \to \infty$ противоречит ограниченности $\{f(x^k, \alpha_k)\}$. Следовательно,

$$k(s) = \min\{r \mid \|x^r - x^s\| > \delta, \ r > s\} < \infty.$$

По определению

$$x^k(s) \notin U_\delta(x^s) \equiv \{x \mid \|x - x^s\| \leqslant \delta\},$$

но так как $\rho_k \to 0$, то для достаточно больших s

$$x^{k(s)} \in U_{2\delta}(x^s),$$

т. е. неравенство (13.7) остается в силе при $k = k(s)$. Из соотношения

$$x^{k(s)} = x^s - \sum_{t=s}^{k(s)-1} \rho_t \left(\sum_{p=r_t}^{t} H(x^p, \alpha_p) \right)$$

вытекает, что

$$\sum_{t=s}^{k(s)-1} \rho_t > \frac{\delta}{CM}.$$

Поэтому из неравенства (13.7) получаем

$$f(x^{k(s)}, \alpha_{k(s)}) \leqslant f(x^s, \alpha_s) - \frac{\sigma\delta}{2CM^2},$$

откуда непосредственно следует (4.13). Теорема доказана.

В качестве направления движения можно брать вектор d^k, который принадлежит выпуклой оболочке векторов $H(x^{r_k}, \alpha_{r_k}), \ldots, H(x^k, \alpha_k)$, т. е.

$$x^{k+1} = x^k - \rho_k d^k, \tag{13.8}$$

$$d^k = \sum_{t=r_k}^{k} \lambda_t^k H(x^t, \alpha_t), \qquad \sum_{t=r_k}^{k} \lambda_t^k = 1, \qquad \lambda_t^k \geqslant 0.$$

Т е о р е м а 13.2. *Пусть выполнены условия теоремы* 13.1. *Тогда предельные точки последовательности* $\{x^k\}$ (13.8) *с вероятностью* 1 *принадлежат множеству*

$$X^* = \{x^* \mid 0 \in \partial f(x^*)\},$$

последовательность $\{f(x^k)\}$ *сходится с вероятностью* 1.

Доказательство практически аналогично предыдущему, поскольку вектор, принадлежащий выпуклой оболочке векторов $H(x^{r_k}, \alpha_{r_k}), \ldots$ $\ldots, H(x^k, \alpha_k)$, представим в виде (13.3). Поэтому

$$f(x^{k+1}, \alpha_{k+1}) \leqslant f(x^k, \alpha_k) + \left(\sum_{t=r_k}^{k} \lambda_t^k (\nabla f(x^k + \tau(x^{k+1} - x^k), \alpha_k) - \right.$$

$$\left. - \nabla f(x^t, \alpha_t)), x^{k+1} - x^k \right) + \left(\sum_{t=r_k}^{k} \lambda_t^k \nabla f(x^t, \alpha_t), x^{k+1} - x^k \right).$$

Отсюда следует, что

$$f(x^{k+1}, \alpha_{k+1}) \leqslant f(x^k, \alpha_k) + C|\alpha_k - \alpha_{k+1}| +$$
$$+ C\frac{\rho_k}{\alpha_k}\sum_{t=r_k}^{k} \rho_t + C\rho_k \sum_{t=r_k}^{k}\left[\frac{|\alpha_t-\alpha_k|}{\alpha_t} + \frac{|\alpha_t-\alpha_k|}{\alpha_k}\right] -$$
$$- \rho_k \left\|\sum_{t=r_k}^{k} \lambda_t^k \nabla f(x^t, \alpha_t)\right\|^2 + \rho_k \left(\sum_{t=r_k}^{k} \lambda_t^k \nabla f(x^t, \alpha_t),\right.$$
$$\left.\sum_{t=r_k}^{k} \lambda_t^k (\nabla f(x^t, \alpha_t) - H(x^t, \alpha_t))\right). \qquad (13.9)$$

Таким образом, получили неравенство, аналогичное (13.6).

В релаксационных методах негладкой оптимизации типа (13.8) в качестве d^k обычно берут ближайший к нулю вектор из выпуклой оболочки накопленных векторов, т. е. решают дополнительно специальную задачу квадратичного программирования. Если не требовать монотонности процесса, то, как показывают практические расчеты, методы типа (13.8) не имеют преимущества в скорости сходимости по сравнению с методами (13.1) с простейшим усреднением и без запоминания информации.

Заметим, что условиям теоремы 13.1 удовлетворяют также шаговые множители

$$\rho_k' = \rho_k n_k, \quad \rho_k'' = \rho_k / n_k, \quad 1 \leqslant n_k \leqslant N < \infty.$$

Поэтому в выбранном направлении движения можно проводить грубую одномерную минимизацию. Обычно в численных расчетах векторы $H(x^k, \alpha_k)$, как и сам вектор спуска в (13.1), выбирают нормированными.

Усреднение направлений спуска выгодно использовать также в методах условной минимизации. Пусть

$$X = X_1 \cap X_2, \quad X_1 = \{x \mid f_i(x) \leqslant 0, \quad i = 1, \ldots, m\},$$

X_2 — выпуклое ограниченное и замкнутое множество, причем X_2 имеет непустое пересечение с множеством внутренних точек X_1. Метод минимизации $f_0(x)$ при $x \in X$ с усреднением субградиентов определяется последовательностью точек

$$x^{k+1} = \pi_{X_2}(x^k - \rho_k \bar{x}^k), \qquad (13.10)$$

где

$$\bar{x}^k = \sum_{t=r_k}^{k} g^{\nu_t}(x^t);$$

$$\nu_t = \begin{cases} 0, & \max\limits_{1\leqslant i\leqslant m} f_i(x^t) \leqslant 0; \\ j, & \max\limits_{1\leqslant i\leqslant m} f_i(x^t) = f_j(x^t) > 0. \end{cases}$$

Здесь $f_\nu(x)$ — выпуклые функции, $g^\nu(x) \in \partial f_\nu(x)$ $(\nu = 0, 1, \ldots, m)$, $\pi_{X_2}(x)$ — оператор проектирования точки x на множество X_2, в точках x^{r_k} происходит возобновление суммирования.

Теорема 13.3. *Пусть*

$$\rho_k \to 0, \qquad \sum_{k=0}^{\infty} \rho_k = \infty.$$

Тогда предельные точки последовательности $\{x^k\}$ принадлежат множеству

$$X^* = \{x^* = \arg\min_{x \in X} f_0(x)\}.$$

Доказательство. Обозначим

$$Q_\tau = \{x \mid x \in X_1, \ f_0(x) \leqslant \tau\}.$$

Пусть $X_\delta^* \subset Q_\tau$ — множество внутренних точек множества Q_τ, которые отстоят от границы Q_τ не менее чем на δ. При достаточно малом δ множество X_δ^* непусто. Вначале покажем, что предельные точки последовательности $\{x^k\}$ принадлежат множеству Q_τ; отсюда, устремив τ к $f_0(x^*)$, получим утверждение теоремы.

Пусть существует последовательность

$$x^s \to x' \notin Q_\tau, \qquad s \in S.$$

Найдутся такие $\bar{s}$ и $\bar{\varepsilon}$, что, начиная с некоторого номера $s \geqslant \bar{s}$, $2\bar{\varepsilon}$-окрестности точек x^s не пересекаются с Q_τ. Покажем, что для любых $\varepsilon \leqslant \bar{\varepsilon}/2$ и достаточно больших s

$$k(s) = \min\{r \mid \| x^r - x^s \| > \varepsilon, \ r > s\} < \infty.$$

Предположим противное:

$$x^k \in U_\varepsilon(x^s), \qquad s \geqslant \bar{s}.$$

Так как $\rho_k \to 0$, то для достаточно больших s точки $x^{s-M}, \ldots, x^{s-1}$ также принадлежат $U_\varepsilon(x^s)$. Пусть $x^k \in X_1$; тогда

$$0 > -\sigma \geqslant \max_{y \in X_\delta^*} (f_0(y) - f_0(x^k)).$$

Если $x^k \notin X_1$, то

$$0 > -\sigma \geqslant \max_{y \in X_\delta^*} (f_j(y) - f_j(x^k)), \qquad j = 1, \ldots, m.$$

Поэтому из этих неравенств для достаточно малых ε вытекает, что

$$\max_{y \in X_\delta^*} \sum_{t=r_k}^{k} (g^{\nu_t}(x^t), \ y - x^k) \leqslant -\frac{\sigma}{2}.$$

Положим

$$W(x) = \min_{y \in X_\delta^*} \| x^k - y \|^2 = \| x^k - y(x^k) \|^2, \qquad y(x^k) \in X_\delta^*.$$

Тогда

$$W(x^{k+1}) \leqslant \| x^{k+1} - y(x^k)\|^2 = \| \pi_{X_2}(x^k - \rho_k \bar{x}^k) - y(x^k)\|^2 \leqslant$$
$$\leqslant \| x^k - \rho_k \bar{x}^k - y(x^k)\|^2 = W(x^k) + 2\rho_k (\bar{x}^k, y(x^k) - x^k) +$$
$$+ \rho_k^2 \| \bar{x}^k \|^2 \leqslant W(x^k) + C\rho_k^2 - \rho_k \sigma.$$

Для больших s имеем $C\rho_k \leqslant \sigma/2$, поэтому

$$0 \leqslant W(x^k) \leqslant W(x^s) - \frac{\sigma}{2} \sum_{t=s}^{k-1} \rho_t.$$

Перейдя к пределу по $k \to \infty$, получим противоречие с условием ограниченности снизу $W(x)$. Нетрудно показать, что

$$\overline{\lim_{s\to\infty}}\, W(x^{k(s)}) < \lim_{s\to\infty} W(x^s).$$

Таким образом, выполняется условие (4.13). Заметим, что $W(x)$ на Q_τ принимает континуум значений, однако это не является препятствием для доказательства сходимости, поскольку $W(x)$ на Q_τ принимает значения из $[0, \delta]$. Следовательно, предельные точки последовательности $\{x^k\}$ принадлежат Q_τ, откуда, устремив τ к $f_0(x^*)$, получаем требуемое утверждение теоремы.

Конечно-разностный метод минимизации $f_0(x)$ при $x \in X$ имеет следующий вид:

$$x^{k+1} = \pi_{X_2}(x^k - \rho_k \bar{x}^k), \tag{13.11}$$

где

$$\bar{x}^k = \sum_{t=r_k}^{k} H^{\nu_t}(x^t, \alpha_t);$$

$$\nu_t = \begin{cases} 0, & \max\limits_{1\leqslant i\leqslant m} f_i(x^t) \leqslant 0, \\ j, & \max\limits_{1\leqslant i\leqslant m} f_i(x^t) = f_j(x^t) > 0; \end{cases}$$

векторы $H^{\nu_t}(x^t, \alpha_t)$ определяются по формулам (13.2) — (13.4).

Теорема 13.4. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < \infty, \quad \frac{\Delta_k}{\alpha_k} \to 0, \quad \alpha_k \to 0.$$

Тогда с вероятностью 1 *предельные точки последовательности* $\{x^k\}$ *принадлежат множеству* X^*.

Доказательство теоремы аналогично предыдущему.

§ 14. Методы усредненных градиентов

В этом параграфе исследуется класс нерелаксационных методов минимизации обобщенно дифференцируемых функций, в которых направление движения является выпуклой комбинацией некоторого числа обобщенных градиентов, полученных на предыдущих итерациях, а шаг регулируется адаптивно. Усреднение предыдущих градиентов можно рассматривать как простой прием, улучшающий работу градиентного метода. Показано, что специальными случаями метода усредненных градиентов являются известные методы тяжелого шарика [100] и овражного шага [14, 74], обобщенные для минимизации (невыпуклых негладких) обобщенно дифференцируемых функций при ограничениях. Как известно, в гладком выпуклом случае эти методы намного превосходят по эффективности градиентный метод. Кроме того, эти методы можно использовать для глобальной минимизации функций при ограничениях.

Построим метод усредненных градиентов для решения следующей задачи:

$$f(x) \to \min \tag{14.1}$$

при ограничении

$$h(x) \leqslant 0, \tag{14.2}$$

где $f(x)$ и $h(x)$ — обобщенно дифференцируемые функции, $G_f(x)$ и $G_h(x)$—их псевдоградиентные отображения, $h(x) \to +\infty$ при $\|x\| \to +\infty$. Распространение метода на задачи с более общими ограничениями в принципе можно провести так же, как это сделано для метода обобщенного градиента в § 10. Обозначим

$$G(x) = \begin{cases} G_f(x), & h(x) < 0, \\ \text{co}\{G_f(x),\ G_h(x)\}, & h(x) = 0, \\ G_h(x) & h(x) > 0, \end{cases} \tag{14.3}$$

$$G_\delta(x) = \text{co}\{G(y) \mid \|y - x\| \leqslant \delta\}.$$

Сходимость метода усредненных градиентов будем трактовать как следствие определенной устойчивости метода обобщенного градиента, к изучению которой сейчас перейдем.

1. Устойчивость метода обобщенного градиента. Рассмотрим следующий метод:

$$x^0 \in E_n, \quad x^{k+1} = x^k - \rho_k P^k, \quad P^k = g^k + \Delta^k, \qquad k = 0, 1, \ldots, \tag{14.4}$$

$$g^k \in G_{\delta_k}(x^k), \tag{14.5}$$

$$0 \leqslant \rho_k \leqslant \rho, \qquad 0 \leqslant \delta_k \leqslant \delta, \qquad \|\Delta^k\| \leqslant \Delta, \tag{14.6}$$

$$\lim_{k\to\infty} \rho_k = \lim_{k\to\infty} \delta_k = \lim_{k\to\infty} \|\Delta^k\| = 0, \qquad \sum_{k=0}^{\infty} \rho_k = +\infty. \tag{14.7}$$

Этот метод отличается от метода обобщенного градиента (9.2), (9.3) или (10.7), (10.8) тем, что в нем, не теряя сходимости, можно брать текущий градиент g^k не в точке x^k, а в некоторой близкой точке y: $\|y - x^k\| \leqslant \delta_k$. Более того, в качестве направления движения P^k можно брать произвольную выпуклую комбинацию градиентов, вычисленных с ошибкой в соседних с x^k точках. Данное свойство вместе с общими условиями (14.7) выражает определенную устойчивость метода обобщенного градиента (9.2), (9.3) и (10.7), (10.8).

Отметим, что указанная устойчивость отличается от реальной вычислительной устойчивости метода обобщенного градиента вида

$$x^{k+1} = x^k - \rho_k g^k, \qquad g^k \in G(x^k), \qquad k = 0, 1, \ldots \tag{14.8}$$

При практических расчетах по любому методу неизбежны ошибки величин, входящих в запись метода. Метод устойчив, если при малых ошибках он приходит в малую окрестность решения. Например, вычислительная машина всегда производит действия с рациональными числами ограниченной разрядности. Если вычисления происходят в пределах разрядной сетки, то они абсолютно точные. Однако если в результате какой-либо операции над числами, находящимися в пределах разрядной сетки, результат операции выходит за пределы разрядной сетки, то он округляется или обрезается. Здесь в вычисления закрадываются ошибки округления. Это приводит к тому, что, например, в методе обобщенного градиента следующая точка x^{k+1} вычисляется не по (14.8), а по формуле

$$x^{k+1} = \overline{\bar{x}^k - \overline{\bar{\rho}_k \cdot \bar{g}_k}}, \qquad g^k \in G(\bar{x}^k), \tag{14.9}$$

где черта означает округление (т. е. в этой формуле величины без черты обозначают числа, машинные представления которых могут выходить за пределы разрядной сетки, а величины с чертой — рациональные числа в пределах разрядной сетки). Если в (14.9) ошибки округления на каждой итерации сделать малыми по сравнению с величиной шага ρ_k, то изучение сходимости алгоритма (14.9) можно свести к исследованию сходимости процедур вида (14.4), (14.7) [85].

Следующая лемма устанавливает локальную устойчивость метода обобщенного градиента.

Лемма 14.1. *Пусть последовательность начальных точек* $\{x^s\}$ *сходится к* $x = \lim\limits_{s\to\infty} x^s$. *Для каждого* s *рассмотрим последовательности* $\{x_s^k\}_{k=k_s}^{n_s}$ *такие, что*

$$x_s^{k_s} = x^s, \qquad x_s^{k+1} = x_s^k - \rho_s^k P_s^k, \tag{14.10}$$

$$P_s^k = Q_s^k + \Delta_s^k, \quad Q_s^k \in G_{\delta_s^k}(x_s^k), \qquad k_s \leqslant k \leqslant n_s. \tag{14.11}$$

Обозначим

$$\rho_s = \sup_{k_s \leqslant k \leqslant n_s} \rho_s^k, \qquad \delta_s = \sup_{k_s \leqslant k \leqslant n_s} \delta_s^k,$$

$$\Delta_s = \sup_{k_s \leqslant k \leqslant n_s} \|\Delta_s^k\|, \quad \sigma_s = \sum_{k=k_s}^{n_s - 1} \rho_k.$$

Если $0 \notin G(x)$ *и* $\sigma_s \geqslant \sigma > 0$, *то существует* $\bar\varepsilon = \bar\varepsilon(x, \sigma)$ *такое, что для любого* $\varepsilon \in (0, \bar\varepsilon]$ *найдутся* $\bar\rho = \bar\rho\ (x, \varepsilon), \bar\delta = \bar\delta(x, \varepsilon)$ *и* $\bar\Delta = \bar\Delta(x, \varepsilon)$ *такие, что для* $\{x_s^k\}_{k=k_s}^{n_s}$, *определенных формулами* (14.10), (14.11), *где* $\delta_s^k \leqslant \bar\delta$, $\rho_s^k \leqslant \bar\rho$ *и* $\Delta_s^k \leqslant \bar\Delta$, *найдутся индексы* l_s *такие, что* $\| x^k - x \| \leqslant \varepsilon$ *при* $k \in [k_s, l_s]$ *и*

1) $f(x) = \lim\limits_{s\to\infty} f(x^s) > \overline{\lim\limits_{s\to\infty}} f(x_s^{l_s}), \quad h(x) < 0;$

2) $h(x) = \lim\limits_{s\to\infty} h(x^s) > \overline{\lim\limits_{s\to\infty}} h(x_s^{l_s}), \quad h(x) > 0;$

3) $f(x) = \lim\limits_{s\to\infty} f(x^s) > \overline{\lim\limits_{s\to\infty}} f(x_s^{l_s}),$

$$C(x)\bar\delta \geqslant \overline{\lim_{s\to\infty}} h(x_s^{l_s}), \quad C(x) > 0, \quad h(x) = 0.$$

Доказательство. Напомним, что $G_\delta(x) = \mathrm{co}\{G(y) \mid \|y - x\| \leqslant \delta\}$. В силу полунепрерывности сверху $G(x)$ существует ε_1-окрестность точки x такая, что $\gamma_1 = \rho(0, G_{\varepsilon_1}(x)) > 0$. Обозначим

$$\Gamma_f = \sup\{\|g\| \mid g \in G_f(y),\ \|y - x\| \leqslant \varepsilon_1\},$$

$$\Gamma_h = \sup\{\|g\| \mid g \in G_h(y),\ \|y - x\| \leqslant \varepsilon_1\},$$

$$\Gamma = \max(\Gamma_f, \Gamma_h) + \gamma_1 < +\infty.$$

Возьмем $\varepsilon_2 = \min(\varepsilon_1, \sigma\gamma_1)$. В силу обобщенной дифференцируемости $f(x)$ и $h(x)$ для любых констант C_f и C_h найдется $\varepsilon_3 \leqslant \varepsilon_2$ такое, что при $\|y - x\| \leqslant \varepsilon_3$

$$f(y) = f(x) + (g_f, y - x) + O_f(x, y, g_f),$$

$$h(y) = h(x) + (g_h, y - x) + O_h(x, y, g_h),$$

где $|O_f(x, y, g_f)| \leqslant C_f \|y - x\|$, $|O_h(x, y, g_h)| \leqslant C_h \|y - x\|$, $g_f \in G_f(y)$, $g_h \in G_h(y)$. Константы C_f и C_h будут уточнены позднее. Если $h(x) \neq 0$, то найдется $\varepsilon_4 \leqslant \varepsilon_3$ такое, что для всех y, $\|y - x\| \leqslant \varepsilon_4$, будет $|h(y) - h(x)| \leqslant |h(x)|/2$. Если $h(x) = 0$, то полагаем $\varepsilon_4 = \varepsilon_3$.

Положим $\bar\varepsilon = \varepsilon_4$. Зафиксируем произвольное $\varepsilon \in (0, \bar\varepsilon]$. Возьмем $\bar\delta_1 = \varepsilon/4$, $\bar\rho_1 = \varepsilon/(4\Gamma)$, $\bar\Delta_1 = \gamma_1/2$. Пусть для $s \geqslant S$ имеет место $\|x^s - x\| \leqslant \varepsilon/4$, $\delta_s \leqslant \bar\delta_1$, $\Delta_s \leqslant \bar\Delta_1$, $\rho_s \leqslant \bar\rho_1$.

Обозначим через $m_s = m_s(\varepsilon)$ максимальный индекс $k \in [k_s, n_s]$ такой, что для всех $r \in [k_s, k)$ имеет место $\|x_s^r - x\| \leqslant \varepsilon/2$.

Покажем, что $\varepsilon/2 \leqslant \|x_s^{m_s} - x\| \leqslant 3\varepsilon/4$. Докажем сначала левое неравенство. Действительно, если $\|x_s^{m_s} - x\| \leqslant \varepsilon/2$, то $m_s = n_s$, и получаем противоречие: $\varepsilon_2 > 3\varepsilon/4 \geqslant \|x_s^{n_s} - x^s\| \geqslant \sigma\gamma_1$. Далее

$$\|x_s^{m_s} - x\| \leqslant \|x_s^{m_s - 1} - x\| + \rho_s^{m_s - 1} \|P_s^{m_s - 1}\| \leqslant 3\varepsilon/4.$$

Отсюда

$$\frac{3\varepsilon}{4} \geqslant \left\| \sum_{k=k_s}^{m_s-1} \rho_k P^k \right\| \geqslant \frac{\gamma_1}{2} \sum_{k=k_s}^{m_s-1} \rho_k,$$

$$\sum_{k=k_s}^{m_s-1} \rho_k \leqslant a = \frac{3\varepsilon}{2\gamma_1}.$$

Рассмотрим $g_s^k \in G_{\delta_s^k}(x_s^k)$. Справедливо представление:

$$g_s^k = \sum_{i=1}^{n+1} \lambda_s^{ki} g_s^{ki} = \sum_{i=1}^{n+1} \lambda_s^{ki} \alpha_s^{ki} g_{fs}^{ki} + \sum_{i=1}^{n+1} \lambda_s^{ki} \beta_s^{ki} g_{hs}^{ki} =$$

$$= \left(\sum_{i=1}^{n+1} \mu_s^{ki} g_{fs}^{ki}\right) \sum_{i=1}^{n+1} \lambda_s^{ki} \alpha_s^{ki} + \left(\sum_{i=1}^{n+1} \nu_s^{ki} g_{hs}^{ki}\right) \sum_{i=1}^{n+1} \lambda_s^{ki} \beta_s^{ki} =$$

$$= \bar{\alpha}_s^k \bar{g}_{fs}^k + \bar{\beta}_s^k \bar{g}_{hs}^k, \tag{14.12}$$

где λ_s^{ki}, α_s^{ki}, β_s^{ki}, μ_s^{ki}, ν_s^{ki}, $\bar{\alpha}_s^k$, $\bar{\beta}_s^k \geqslant 0$,

$$\sum_{i=1}^{n+1} \lambda_s^{ki} = \sum_{i=1}^{n+1} \mu_s^{ki} = \sum_{i=1}^{n+1} \nu_s^{ki} = \alpha_s^{ki} + \beta_s^{ki} = \bar{\alpha}_s^k + \bar{\beta}_s^k = 1,$$

$$g_s^{ki} \in G(x_s^{ki}), \qquad g_{fs}^{ki} \in G_f(x_s^{ki}), \qquad g_{hs}^{ki} \in G_h(x_s^{ki}),$$

$$\bar{g}_{fs}^k \in \operatorname{co}\{G_f(y) \mid \|y - x_s^k\| \leqslant \delta_s^k\},$$

$$\bar{g}_{hs}^k \in \operatorname{co}\{G_h(y) \mid \|y - x_s^k\| \leqslant \delta_s^k\}.$$

При $\|x^s - x\| \leqslant \varepsilon/4$, $\delta_s \leqslant \varepsilon/4$, $k_s \leqslant k \leqslant m_s$, $1 \leqslant i \leqslant n+1$ имеет место $\|x_s^{ki} - x\| \leqslant \|x_s^{ki} - x_s^k\| + \|x_s^k - x\| \leqslant \varepsilon \leqslant \varepsilon_3$. Поэтому справедливо разложение

$$f(x_s^{ki}) = f(x) + (g_{fs}^{ki}, x_s^{ki} - x) + O_f(x, x_s^{ki}, g_{fs}^{ki}) \leqslant$$

$$\leqslant f(x) + (g_{fs}^{ki}, x_s^{ki} - x^s) + C_f \|x_s^{ki} - x^s\| + (\Gamma_f + C_f)\|x^s - x\|.$$

В этом разложении x_s^{ki} $(1 \leqslant i \leqslant n+1)$ можно приближенно заменить на x_s^k; имеет место неравенство

$$f(x_s^k) \leqslant f(x) + (g_{fs}^{ki}, x_s^k - x^s) + C_f \|x_s^k - x^s\| +$$

$$+ (2\Gamma_f + C_f)\delta_s + (\Gamma_f + C_f)\|x^s - x\|.$$

Умножим эти неравенства на μ_s^{ki} и просуммируем по i:

$$f(x_s^k) \leqslant f(x) + (\bar{g}_{fs}^k, x_s^k - x^s) + C_f \|x_s^k - x^s\| +$$

$$+ (2\Gamma_f + C_f)\delta_s + (\Gamma_f + C_f)\|x^s - x\|. \tag{14.13}$$

Представим

$$x_s^k = \bar{x}_s^k + z_s^k,$$

$$\bar{x}_s^k = x^s - \sum_{r=k_s}^{k-1} \rho_s^r g_s^r, \quad z_s^k = -\sum_{r=s}^{k-1} \rho_s^r \Delta_s^r.$$

Справедлива оценка при $k \in [k_s, m_s]$:

$$\| z_s^k \| \leqslant (\sup_{s \leqslant r \leqslant n_s} \Delta_s^r) \sum_{r=k_s}^{m_s-1} \rho_s^r \leqslant a\Delta_s.$$

Заменим в правой части (14.13) x_s^k на $\bar{x}_s^k$:

$$f(x_s^k) \leqslant f(x) + (\bar{g}_{fs}^k, \bar{x}_s^k - x^s) + C_f \| \bar{x}_s^k - x^s \| +$$
$$+ (2\Gamma_f + C_f)\delta_s + (\Gamma_f a + C_f)\Delta_s + (\Gamma_f + C_f)\| x^s - x \|. \quad (14.14)$$

Аналогично

$$h(x_s^k) \leqslant h(x) + (\bar{g}_{hs}^k, \bar{x}_s^k - x^s) + C_h \| \bar{x}_s^k - x^s \| +$$
$$+ (2\Gamma_h + C_h)\delta_s + (\Gamma_h a + C_h)\Delta_s + (\Gamma_h + C_h)\| x^s - x \|. \quad (14.15)$$

Теперь рассмотрим скалярное произведение

$$(g_s^k, \bar{x}_s^k - x_s) = -\left(g_s^k, \sum_{r=k_s}^{k-1} \rho_s^r g_s^r \Big/ \sum_{r=k_s}^{k-1} \rho_s^r\right) \sum_{r=k_s}^{k-1} \rho_s^r.$$

Для $s \geqslant S$ и $k \in [k_s, m_s]$ имеет место $g_s^k \in G_{\varepsilon_1}(x)$, поэтому $\gamma_1/2 \leqslant \|P_s^k\| \leqslant \leqslant \Gamma$, и

$$\sum_{k=k_s}^{m_s-1} \rho_s^k \geqslant \| x_s^{m_s} - x^s \|/\Gamma \geqslant \varepsilon/(4\Gamma).$$

Очевидно, что для оценки скалярных произведений $(g_s^k, \bar{x}_s^k - x^s)$ при $s \geqslant S$ применима лемма 9.2, в которой нужно положить

$$P = G_{\varepsilon_1}(x), \qquad p^k = g_s^k, \qquad \sigma = \varepsilon/(4\Gamma), \qquad \rho = \varepsilon\gamma_1^2/(24\Gamma^3).$$

Положим $\bar{\rho} = \min(\bar{\rho}_1, \rho)$. Если $s \geqslant S$ и $\rho_s \leqslant \bar{\rho}$, то, согласно лемме 9.2, найдутся индексы $l_s \leqslant m_s$, для которых $\| x_s^k - x \| \leqslant \varepsilon$ при $k \in \in [k_s, l_s]$, и

$$\left(g_s^{l_s}, \sum_{k=k_s}^{l_s-1} \rho_s^k g_s^k \Big/ \sum_{k=k_s}^{l_s-1} \rho_s^k\right) \geqslant \frac{\gamma_1^2}{4},$$

$$\sum_{k=k_s}^{l_s-1} \rho_s^k \geqslant \frac{\varepsilon\gamma_1}{12\Gamma^2}. \quad (14.16)$$

Необходимо рассмотреть три случая: $h(x)<0$, $h(x)>0$ и $h(x)=0$.

Пусть $h(x)<0$. При достаточно больших s и $k\in[k_s, m_s]$ имеет место $g_s^k=\bar g_{fs}^k$, поэтому при $k=l_s$ неравенство (14.14) можно переписать в виде

$$f(x_s^{l_s})\leqslant f(x)-\left(\frac{\gamma_1^2}{4}-\Gamma C_f\right)\sum_{k=k_s}^{l_s-1}\rho_s^k+$$
$$+(2\Gamma_f+C_f)\,\delta_s+(\Gamma_f a+C_f)\,\Delta_f+(\Gamma_f+C_f)\,\|x^s-x\|. \qquad (14.17)$$

Теперь конкретизируем C_f (а значит, и $\bar\varepsilon$): возьмем $C_f=\gamma_1^2/(8\Gamma)$. Имеем $\varepsilon\in(0,\bar\varepsilon]$ и $\bar\rho=\min(\bar\rho_1,\rho)$. Положим

$$\bar\delta_f=\frac{\varepsilon\gamma_1^3}{300\Gamma^2(2\Gamma_f+C_f)},\quad \bar\Delta_f=\frac{\varepsilon\gamma_1^3}{300\Gamma^2(\Gamma_f a+C_f)}.$$

Все проведенные ранее рассуждения превратятся из условно верных в безусловно верные. В результате из (14.16), (14.17) при $\rho_s\leqslant\bar\rho$, $\delta_s\leqslant\bar\delta_f$ и $\Delta_s\leqslant\bar\Delta_s$ получим

$$f(x_s^{l_s})\leqslant f(x)-\frac{\varepsilon\gamma_1^3}{300\Gamma^2}+(\Gamma_f+C_f)\,\|x^s-x\|, \qquad (14.18)$$

откуда следует утверждение 1) леммы.

Утверждение 2) доказывается аналогично.

Пусть теперь $h(x)=0$. Докажем утверждение 3) леммы.

Покажем, что при соответствующем выборе констант C_h, $\bar\delta$ и $\bar\Delta$ при достаточно больших s будет $\min\{h(y)\mid\|y-x_s^{l_s}\|\leqslant\delta_s\}\leqslant 0$, что влечет $h(x_s^{l_s})\leqslant 2\Gamma_h\delta_s=C(x)\,\delta_s$ и доказывает вторую часть утверждения 3) леммы. Действительно, предположим, что $\min\{h(y)\mid\|y-x_s^{l_s}\|\leqslant\leqslant\delta_s\}>0$. Тогда $g_s^{l_s}=\bar g_{hs}^{l_s}$, и из (14.15) получаем

$$0\leqslant h(x_s^{l_s})\leqslant-\left(\frac{\gamma_1^2}{4}-C_h\Gamma\right)\sum_{k=k_s}^{l_s-1}\rho_s^k+(2\Gamma_h+C_h)\,\delta_s+$$
$$+(\Gamma_h a+C_h)\,\Delta_s+(\Gamma_h+C_h)\,\|x^s-x\|.$$

При $C_h\leqslant\gamma_1^2/(8\Gamma)$ и

$$\delta_s\leqslant\frac{1}{300}\frac{\varepsilon\gamma_1^3}{\Gamma^2(2\Gamma_h+C_h)},\quad \Delta_s\leqslant\frac{1}{300}\frac{\varepsilon\gamma_1^3}{\Gamma^2(\Gamma_h a+C_h)}$$

для достаточно больших s получаем противоречие:

$$0\leqslant-\frac{\varepsilon\gamma_1^3}{300\Gamma^2}+(\Gamma_h+C_h)\,\|x^s-x\|.$$

Вторая часть утверждения 3) леммы доказана.

Теперь в зависимости от расположения $x_s^{l_s}$ рассмотрим два случая:

а) $\max\{h(y)\,|\,\|y-x_s^{l_s}\|\leqslant\delta_s\}<0$; тогда $g_s^{l_s}=\bar g_{fs}^{l_s}$, справедливы неравенства (14.14), (14.17) и (14.18), откуда при $C_f\leqslant\gamma_1^2/(8\Gamma)$ (и соответствующих $\varepsilon\leqslant\bar\varepsilon$), а также при $\delta_s\leqslant\bar\delta_f$ и $\Delta_s\leqslant\bar\Delta_s$ следует первая часть утверждения 3) леммы;

б) $\max\{h(y)\,|\,\|y-x_s^{l_s}\|\leqslant\delta_s\}\geqslant 0$; тогда $h(x_s^{l_s})\geqslant-2\Gamma_h\delta_s$ и из (14.15) при $k=l_s$ получаем

$$\left(\bar g_{hs}^{l_s},\ \sum_{k=k_s}^{l_s-1}\rho_s^k g_s^k\right)\leqslant C_h\left\|\sum_{k=k_s}^{l_s-1}\rho_s^k g_s^k\right\|+(4\Gamma_h+C_h)\,\delta_s+$$
$$+(\Gamma_h a+C_h)\,\Delta_s+(\Gamma_h+C_h)\,\|x^s-x\|. \qquad (14.19)$$

Из (14.12), (14.16), (14.19) следует

$$\bar\alpha_s^{l_s}\left(\bar g_{fs}^{l_s},\ \sum_{k=k_s}^{l_s-1}\rho_s^k g_s^k\right)\geqslant\frac{\gamma_1^2}{4}\sum_{k=k_s}^{l_s-1}\rho_s^k-\bar\beta_s^{l_s}\left(\bar g_{hs}^{l_s},\ \sum_{k=k_s}^{l_s-1}\rho_s^k g_s^k\right)\geqslant$$
$$\geqslant\left(\frac{\gamma_1^2}{4}-C_h\Gamma\right)\sum_{k=k_s}^{l_s-1}\rho_s^k-(4\Gamma_h+C_h)\,\delta_s-$$
$$-(\Gamma_h a+C_h)\,\Delta_s-(\Gamma_h+C_h)\,\|x^s-x\|. \qquad (14.20)$$

Теперь выберем величину C_h; все проведенные рассуждения верны при $C_h=\gamma_1^2/(8\Gamma)$. Если

$$\delta_s\leqslant\bar\delta\leqslant\frac{1}{200}\frac{\varepsilon\gamma_1^3}{\Gamma^2(4\Gamma_h+C_h)},\qquad \Delta_s\leqslant\bar\Delta\leqslant\frac{1}{200}\frac{\varepsilon\gamma_1^3}{\Gamma^2(\Gamma_h a+C_h)},$$

то при достаточно больших s из (14.20) следует

$$\left(\bar g_{fs}^{s},\ \sum_{k=k_s}^{l_s-1}\rho_s^k g_s^k\right)\geqslant\frac{\gamma_1^2}{8}\sum_{k=k_s}^{l_s-1}\rho_s^k-(4\Gamma_h+C_h)\,\delta_s-$$
$$-(\Gamma_h a+C_h)\,\Delta_s-(\Gamma_h+C_h)\,\|x^s-x\|. \qquad (14.21)$$

Из (14.14), (14.21) при $k=l_s$ и достаточно больших s получаем

$$f(x_s^{l_s})\leqslant f(x)-\left(\frac{\gamma_1^2}{8}-\Gamma C_f\right)\sum_{k=k_s}^{l_s-1}\rho_s^k+(\Gamma_f a+\Gamma_h a+C_f+C_h)\,\Delta_s+$$
$$+(2\Gamma_f+4\Gamma_h+C_f+C_h)\,\delta_s+(\Gamma_f+\Gamma_h+C_f+C_h)\,\|x^s-x\|.$$

Конкретизируем теперь C_f (так как C_h было уже указано ранее, то тем самым мы конкретизируем и $\bar\varepsilon$): положим $C_f=\gamma_1^2/(16\Gamma)$. Имеем $0\leqslant\varepsilon\leqslant\bar\varepsilon$ и $\bar\rho=\min(\bar\rho_1,\rho)$. Теперь определим $\bar\delta$ и $\bar\Delta$. Положим

$$\bar\delta=\frac{\varepsilon\gamma_1^3}{200\Gamma^2(2\Gamma_f+4\Gamma_h+C_f+C_h)},\qquad \bar\Delta=\frac{\varepsilon\gamma_1^3}{200\Gamma^2(\Gamma_f a+\Gamma_h a+C_f+C_h)}.$$

Все проведенные ранее рассуждения превратятся из условно верных в безусловно верные. В результате при $\rho_s \leqslant \bar{\rho}$, $\delta_s \leqslant \bar{\delta}$ и $\Delta_s \leqslant \bar{\Delta}$ получим

$$f(x_s^{l_s}) \leqslant f(x) - \frac{\varepsilon\gamma^3}{600\Gamma^2} + (\Gamma_f + \Gamma_h + C_f + C_h)\, \| x^s - x \|,$$

откуда также следует первая часть утверждения 3) леммы. Итак, мы указали $\bar{\varepsilon}(x, \sigma)$ и $\bar{\rho}(x, \varepsilon)$, $\bar{\delta}(x, \varepsilon)$, $\bar{\Delta}(x, \varepsilon)$ ($\varepsilon \in (0, \bar{\varepsilon}]$), при которых выполнены утверждения 1) — 3) леммы.

Теорема 14.1. *Пусть в задаче* (14.1), (14.2) $f(x)$ *и* $h(x)$ — *обобщенно дифференцируемые функции, причем* $h(x) \to +\infty$ *при* $\|x\| \to +\infty$. *Пусть множество* $H^* = \{h(x) \mid 0 \in G_h(x),\ h(x) > 0\}$ *не содержит интервалов. Тогда при достаточно малых* ρ, δ *и* Δ *из* (14.6) *последовательность* (14.4) — (14.7) *ограничена, и*

либо все предельные точки $\{x^k\}$ *не принадлежат допустимой области* $D = \{x \in E_n \mid h(x) \leqslant 0\}$, *и тогда все они принадлежат* $X_h^* = \{x \mid 0 \in G_h(x),\ h(x) > 0\}$ *и существует предел* $\lim\limits_{k\to\infty} h(x^k) > 0$;

либо все предельные точки $\{x^k\}$ *принадлежат* D, *и тогда минимальные из них по значению* f *принадлежат* $X^* = \{x \in E_n \mid 0 \in G(x),\ h(x) \leqslant 0\}$, *а отрезок* $[\varliminf\limits_{k\to\infty} f(x^k),\ \varlimsup\limits_{k\to\infty} f(x^k)]$ *вложен в множество* $F^* = \{f(x) \mid x \in X^*\}$. *Если при этом* F^* *не содержит интервалов, то все предельные точки* $\{x^k\}$ *принадлежат связному подмножеству* X^* *и существует предел* $\lim\limits_{k\to\infty} f(x^k)$.

Данная теорема о сходимости метода (14.4) — (14.7) доказывается точно так же, как леммы 9.1, 10.1 и теоремы 10.1, 10.2 о сходимости метода обобщенного градиента (10.7), (10.8) ввиду справедливости леммы 14.1, аналогичной лемме 10.2.

2. **Метод усредненных градиентов**. Метод предназначен для решения задачи (14.1), (14.2). Он вырабатывает бесконечную последовательность точек $\{x^k\}$ согласно следующим правилам:

$$x^0 \in E_n \text{ — произвольно,} \tag{14.22}$$

$$x^{k+1} = x^k - \rho_k \nu_k P^k, \quad P^k = Q^k + \Delta^k, \quad \rho_k \geqslant 0, \quad \nu_k > 0, \tag{14.23}$$

$$Q^k = \sum_{r=r_k}^{k} \lambda_{kr} g^r, \quad g^r \in G(x^r), \quad \lambda_{kr} \geqslant 0, \tag{14.24}$$

где x^k — текущее приближение, P^k — направление движения, Q^k — усредненный градиент, Δ^k — некоторая малая добавка, $\rho_k = \rho_k(x^0, g^0, \ldots, x^k, g^k)$ — регулируемые шаговые множители, $\nu_k = \nu_k(x^0, g^0, \ldots, x^k, g^k)$ — регулируемые нормирующие множители, $\lambda_{kr} = \lambda_{kr}(x^0, g^0, \ldots, x^k, g^k)$ — регулируемые усредняющие множители.

Выполнены следующие условия.

Для любого компакта $K \subset E_n$ найдутся числа $\nu_K > 0$, $\lambda_K > 0$, $N_K < +\infty$, $\Lambda_K < +\infty$ такие, что если $\{x^k\} \subset K$ и $g^k \in G(x^k)$, то для всех k имеет место

$$0 < \nu_K \leqslant \nu_k(x^0, g^0, \ldots, x^k, g^k) \leqslant N_K < +\infty, \tag{14.25}$$

$$0 < \lambda_K \leqslant \sum_{r=r_k}^{k} \lambda_{kr}(x^0, g^0, \ldots, x^k, g^k) \leqslant \Lambda_K < +\infty. \tag{14.26}$$

Для любой ограниченной последовательности $\{x^k\}$ и последовательности $\{g^k\}$ ($g^k \in G(x^k)$) имеет место

$$\lim_{k\to\infty} \Delta^k(x^0, g^0, \ldots, x^k, g^k) = 0, \tag{14.27}$$

$$\max_{r_k \leqslant r \leqslant k} \| x^k - x^r \| \leqslant \varepsilon_k(x^0, g^0, \ldots, x^k, g^k), \tag{14.28}$$

$$\lim_{k\to\infty} \varepsilon_k(x^0, g^0, \ldots, x^k, g^k) = 0 \tag{14.29}$$

(функция ε_k мажорирует $\max\limits_{r_k \leqslant r \leqslant k} \| x^k - x^r \|$),

$$\lim_{k\to\infty} \rho_k(x^0, g^0, \ldots, x^k, g^k) = 0. \tag{14.30}$$

Функции $\rho_k(x^0, g^0, \ldots, x^k, g^k)$ обладают следующим свойством: если $\{x^k\} \to x$, $g^k \in G(x^k)$ и $0 \notin G(x)$, то

$$\sum_{k=0}^{\infty} \rho_k(x^0, g^0, \ldots, x^k, g^k) = +\infty. \tag{14.31}$$

З а м е ч а н и е 14.1. Вектор Q^k в (14.23) является некоторым средним из предыдущих обобщенных градиентов, что объясняет название метода. Нетрудно видеть, что в случае безусловной минимизации функции $f(x)$ вектор Q^k есть некоторый обобщенный градиент усредненной функции

$$\varphi_k(x) = \sum_{r=r_k}^{k} \lambda_{kr} f(x - (x^k - x^r)).$$

Поэтому метод усредненных градиентов может также рассматриваться как метод решения специфической нестационарной (предельной) экстремальной задачи, в которой минимизируется последовательность функций $\varphi_k(x)$.

З а м е ч а н и е 14.2. В методе усредненных градиентов (14.22) — (14.24) можно использовать нормированные обобщенные градиенты

$$\bar{g}^r = g^r/(\| g^r \| + \varepsilon), \qquad \varepsilon > 0.$$

В этом случае усредненные градиенты Q^k имеют вид

$$Q^k = \sum_{r=r_k}^{k} \lambda_{kr} \bar{g}^r = \sum_{r=r_k}^{k} \frac{\mu_{kr}}{\| g^r \| + \varepsilon} g^r,$$

где

$$\mu_{kr} \geqslant 0, \qquad \sum_{r=r_k}^{k} \mu_{kr} = 1.$$

Можно также нормировать направления движения P^k; для этого следует взять нормирующие множители вида

$$\nu_k(x^0, g^0, \ldots, x^k, g^k) = \frac{1}{\|P^k\| + \varepsilon}, \qquad \varepsilon > 0.$$

Ввиду ограниченности отображения G на компакте метод усредненных градиентов с нормировкой обобщенных градиентов и векторов P^k удовлетворяет условиям (14.25) — (14.31).

Следующая лемма устанавливает локальное минимизирующее свойство метода усредненных градиентов, которое в основном и обеспечивает его сходимость.

Лемма 14.2. *Предположим, что последовательность* $\{x^k\}$, *порожденная алгоритмом* (14.22) — (14.31), *ограничена и существует предельная точка* $x' = \lim_{s\to\infty} x^{k_s}$ *такая, что* $0 \notin G(x)$. *Тогда найдется* $\bar\varepsilon > 0$ *такое, что для любого* $\varepsilon \in (0, \bar\varepsilon]$ *существуют индексы* $l_s \geqslant k_s$, *для которых* (*при достаточно больших* s) $\|x^k - x'\| \leqslant \varepsilon$ *при* $k \in [k_s, l_s]$, *и справедливо*:

1) $f(x') = \lim_{s\to\infty} f(x^{k_s}) > \overline{\lim_{s\to\infty}} f(x^{l_s})$, $h(x) < 0$;

2) $h(x') = \lim_{s\to\infty} h(x^{k_s}) > \overline{\lim_{s\to\infty}} h(x^{l_s})$, $h(x) > 0$;

3) $f(x') = \lim_{s\to\infty} f(x^{k_s}) > \overline{\lim_{s\to\infty}} f(x^{l_s})$, $0 \geqslant \overline{\lim_{s\to\infty}} h(x^{l_s})$, $h(x) = 0$.

Доказательство. Будем считать, что $\{x^k\} \subset K$, где K — компакт. Обозначим $\gamma = \rho(0, G(x')) > 0$,

$$\Gamma = \sup_{k\geqslant 0} \|g^k\| < +\infty, \quad G_\varepsilon(x) = \mathrm{co}\,\{G(y) \mid \|y - x\| \leqslant \varepsilon\}.$$

В силу полунепрерывности сверху выпуклозначного отображения G найдется $\varepsilon' > 0$ такое, что $\rho(0, G_{\varepsilon'}(x')) \geqslant \gamma/2$. Представим

$$x^{k+1} = x^k - \rho_k \nu_k P^k = x^k - \bar\rho_k \bar P^k,$$

где

$$\bar P^k = \bar Q^k + \bar\Delta^k = P^k / \sum_{r=r_k}^{k} \lambda_{kr}, \quad \bar Q^k = Q^k / \sum_{r=r_k}^{k} \lambda_{kr},$$

$$\bar\Delta^k = \Delta^k / \sum_{r=r_k}^{k} \lambda_{kr}, \quad \bar\rho_k = \rho_k \nu_k \sum_{r=r_k}^{k} \lambda_{kr}.$$

Покажем, что вся последовательность $\{x^k\}$ не может сходиться к точке x. Предположим противное. Тогда для достаточно больших k в силу (14.29) $\bar Q^k \in G_{\varepsilon_k}(x^k) \subset G_{\varepsilon'}(x')$, а в силу (14.26), (14.27) $\|\bar\Delta^k\| \leqslant$

$\leqslant \gamma/4$. Отсюда для достаточно больших t и s получаем соотношение

$$2\varepsilon' \geqslant \| x^t - x^s \| = \Big\| \sum_{k=s}^{t-1} \bar{\rho}_k \bar{P}^k \Big/ \sum_{k=s}^{t-1} \bar{\rho}_k \Big\| \sum_{k=s}^{t-1} \bar{\rho}_k \geqslant \frac{1}{4} \gamma \nu_K \lambda_K \sum_{k=s}^{t-1} \rho_k,$$

где правая часть в силу сделанного предположения и условия (14.31) стремится к бесконечности при $t \to +\infty$. Полученное противоречие доказывает, что вся последовательность не может сходиться к x'. Поэтому существует $\varepsilon'' < \varepsilon'$ такое, что индексы

$$m_s(\varepsilon'') = \sup \{ m \geqslant k_s \mid \| x^k - x' \| \leqslant \varepsilon'' \text{ при } k \in [k_s, m)\} < +\infty.$$

Рассмотрим последовательность $\{x_s^k = x^k,\ k \in [k_s, m_s(\varepsilon'')]\}$. Покажем, что при достаточно больших s к ним применима лемма 14.1. Действительно, $x_s^{k_s} = x^{k_s} \to x'$, $0 \notin G(x')$;

$$x_s^{k+1} = x_s^k - \bar{\rho}_k \bar{P}^k, \qquad k \in [k_s, m_s(\varepsilon''));$$

$$\bar{P}^k = \bar{Q}^k + \bar{\Delta}^k; \qquad \| \bar{\Delta}^k \| \leqslant \| \Delta^k \| / \lambda_K \to 0, \qquad k \to \infty;$$

$$\bar{Q}^k \in G_{\varepsilon_k}(x^k) \subset G_{\delta_s}(x^k); \qquad \delta_s = \sup_{k \geqslant k_s} \varepsilon_k \to 0, \qquad s \to +\infty;$$

$$\bar{\rho}_k = \rho_k \nu_k \sum_{r=r_k}^{k} \lambda_{kr} \leqslant N_K \Lambda_K \rho_k \to 0, \qquad k \to +\infty.$$

Для s и $k \geqslant k_s$ таких, что $\| x^{k_s} - x' \| \leqslant \varepsilon''/2$ и $\| \bar{\Delta}^k \| \leqslant \| \Delta^k \| / \lambda_K \leqslant \Gamma$, имеет место

$$\sigma_s = \sum_{k=k_s}^{m_s(\varepsilon'')-1} \bar{\rho}_k \geqslant \| x^{m_s(\varepsilon'')} - x^{k_s} \| \Big/ \max_{k_s \leqslant k \leqslant m_s(\varepsilon'')} \| \bar{P}^k \| = \frac{\varepsilon''}{8\Gamma} = \sigma' > 0.$$

Итак, к последовательностям $\{x_s^k = x^k,\ k \in [k_s, m_s(\varepsilon'')]\}$ $(s = 0, 1, \ldots)$ применима лемма 14.1, откуда с учетом того, что $\lim\limits_{s \to \infty} \delta_s = 0$ и $\lim\limits_{s \to \infty} \sup\limits_{k \geqslant k_s} \rho_k = 0$, следуют утверждения данной леммы. Лемма доказана.

Следующая лемма устанавливает условия ограниченности последовательностей $\{x^k\}$, построенных методом усредненных градиентов.

Лемма 14.3. *Пусть* $h(x) \to +\infty$ *при* $\| x \| \to +\infty$, *и пусть существует* $c > \max(0, h(x^0))$, *не принадлежащее* $H^* = \{h(x) \mid 0 \in G_h(x),\ h(x) > 0\}$. *Последовательность* $\{x^k\}$ *построена согласно* (14.22) — (14.24) *при условиях* (14.27) — (14.31). *Тогда для любого* $d > c$ *найдутся числа* R, V *и* Δ *такие, что если* $\sup\limits_{k \geqslant 0} \rho_k \leqslant R$, $\sup\limits_{k \geqslant 0} \varepsilon_k \leqslant V$ *и* $\sup\limits_{k \geqslant 0} \| \Delta^k \| \leqslant \Delta$, *то последовательность* $\{x^k\}$ *не выходит из ограниченной области* $\{x \mid h(x) \leqslant d\}$.

Доказательство. Предположим противное. Тогда для некоторого $d > c$ найдутся последовательности $\{x_s^k\}_{k \geqslant 0}$, запущенные для каждого s из точки x^0 согласно (14.22) — (14.31), где $\sup\limits_{k\geqslant 0} \rho_s^k \leqslant R_s \to 0$, $\sup\limits_{k\geqslant 0} \varepsilon_s^k \leqslant V_s \to 0$ и $\sup\limits_{k\geqslant 0} \| \Delta_s^k \| \leqslant \Delta_s \to 0$ $(s \to \infty)$, и уходящие на бесконечность; индекс s обозначает номер запущенной последовательности. Обозначим

$$K = \{x \mid h(x) \leqslant d\}, \qquad \Gamma = \sup \{\| g \| \mid g \in G(y),\ y \in K\} < +\infty.$$

Выделим индексы k_s и t_s такие, что

$$h(x_s^k) < d \text{ при } k \in [0, k_s], \tag{14.32}$$

$$h(x_s^{k_s}) \leqslant c < h(x^k) < d \leqslant h(x_s^{t_s}), \qquad k \in (k_s, t_s). \tag{14.33}$$

Последовательность $\{x_s^{k_s}\}$ ограничена. Не меняя обозначений, будем считать, что $\lim\limits_{s\to\infty} x_s^{k_s} = x'$. Представим

$$x_s^{k+1} = x_s^k - \rho_s^k \nu_s^k P_s^k = x_s^k - \bar\rho_s^k \bar P_s^k, \qquad k \geqslant k_s,$$

где

$$\bar\rho_s^k = \rho_s^k \nu_s^k \sum_{r=r_k}^{k} \lambda_{kr},$$

$$\bar P_s^k = \bar Q_s^k + \bar\Delta_s^k = P^k \Big/ \sum_{r=r_k}^{k} \lambda_{kr} (x_s^0, g_s^0, \ldots, x_s^k, g_s^k),$$

$$\bar Q_s^k = Q_s^k \Big/ \sum_{r=r_k}^{k} \lambda_{kr}, \qquad \bar\Delta_s^k = \Delta_s^k \Big/ \sum_{r=r_k}^{k} \lambda_{kr}.$$

Покажем, что к последовательности $\{x_s^k,\ k_s \leqslant k \leqslant t_s\}$ $(s = 0, 1, \ldots)$ применима лемма 14.1.

В силу (14.32), (14.33) и (14.27), (14.28) при $k < t_s$ имеет место

$$\bar\rho_s^k \leqslant \nu_K \Lambda_K \rho_s^k \leqslant N_K \Lambda_K R_s \to 0, \qquad s \to +\infty; \tag{14.34}$$

$$\| \bar\Delta_s^k \| \leqslant \| \Delta_s^k \| / \lambda_K \leqslant \Delta_s / \lambda_K \to 0, \qquad s \to +\infty; \tag{14.35}$$

$$\| \bar P_s^k \| \leqslant \| \bar Q_s^k \| + \| \bar\Delta_s^k \| \leqslant \Gamma + \frac{1}{\lambda_K} \sup_{s\geqslant 0} \Delta_s < +\infty; \tag{14.36}$$

$$\max_{r_k \leqslant r \leqslant k} \| x_s^k - x_s^r \| \leqslant \varepsilon_s^k \leqslant V_s \to 0, \qquad s \to +\infty; \tag{14.37}$$

$$\bar Q_s^k \in \operatorname{co} \{G(y) \mid \| y - x_s^k \| \leqslant V_s\}.$$

В силу (14.34), (14.36) $\lim\limits_{s\to\infty} \| x^{k_s+1} - x^{k_s} \| = 0$, откуда с учетом (14.33) следует

$$h(x') = \lim_{s\to\infty} h(x_s^{k_s}) = \lim_{s\to\infty} h(x_s^{k_s+1}) = c \notin H^*.$$

Поэтому $0 \notin G_h(x')$, $h(x') > 0$.

Найдем $\delta > 0$ такое, что $\max\{h(y) \mid \|y - x\| \leqslant \delta\} < d$. В силу (14.33) и непрерывности $h(y)$ последовательности $\{x_s^k,\ k_s \leqslant k \leqslant t_s\}$ $(s = 0, 1, \ldots)$ выходят из δ-окрестности x'. Обозначим через m_s моменты первого такого выхода: $m_s \leqslant t_s$. Для всех $k \in [k_s, m_s]$ выполнено (14.36); поэтому при достаточно больших s имеет место $\sum\limits_{k=k_s}^{m_s-1} \bar{\rho}_s^k \geqslant \sigma > 0$.

Таким образом, к последовательностям $\{x_s^k,\ k_s \leqslant k \leqslant m_s\}$ применима лемма 14.1. Однако минимизирующее свойство 2) этой леммы противоречит построению (14.33). Лемма доказана.

Для фактического подбора параметров метода применяется следующий прием.

Пусть известны оценки:

$$R_0 \geqslant \sup_{k \geqslant 0} \rho_k^0, \qquad V_0 \geqslant \sup_{k \geqslant 0} \varepsilon_k^0, \qquad \Delta_0 \geqslant \sup_{k \geqslant 0} \|\Delta_0^k\|,$$

справедливые для всех последовательностей, порождаемых методом Запустим метод из точки x^0 до тех пор, пока $h(x^k) < d$. Если в некоторый момент τ_1 окажется $h(x^{\tau_1}) > d$, то вернемся в начальную точку, выберем $\{\rho_k^1\}$, $\{\varepsilon_k^1\}$, $\{\Delta_1^k\}$ с меньшими оценками:

$$R_1 \geqslant \sup_{k \geqslant 0} \rho_k^1, \qquad V_1 \geqslant \sup_{k \geqslant 0} \varepsilon_k^1, \qquad \Delta_1 \geqslant \sup_{k \geqslant 0} \|\Delta_1^k\|$$

и опять запустим метод из точки x^0 с новыми параметрами. Если в некоторый момент τ_2 $h(x^{\tau_2}) > d$, то снова вернемся в начальную точку, выберем $\{\rho_k^2\}$, $\{\varepsilon_k^2\}$, $\{\Delta_2^k\}$ с еще меньшими оценками:

$$R_2 \geqslant \sup_{k \geqslant 0} \rho_k^2, \qquad V_2 \geqslant \sup_{k \geqslant 0} \varepsilon_k^2, \qquad \Delta_2 \geqslant \sup_{k \geqslant 0} \|\Delta_2^k\|$$

и опять запустим метод из точки x^0 с новыми параметрами и т. д.

Если $\lim\limits_{s \to \infty} R_s = \lim\limits_{s \to \infty} V_s = \lim\limits_{s \to \infty} \Delta_s = 0$, то в силу леммы 14.3 метод после конечного числа рестартов не будет выходить из ограниченной области $\{x \mid h(x) \leqslant d\}$.

Теорема 14.2. *Пусть в задаче* (14.1), (14.2) *функции* $f(x)$ *и* $h(x)$ *обобщенно дифференцируемы, причем* $h(x) \to +\infty$ *при* $\|x\| \to \infty$, *и множество* $H^* = \{h(x) \mid 0 \in G_h(x),\ h(x) > 0\}$ *не содержит интервалов. Пусть последовательность* $\{x^k\}$, *построенная согласно* (14.22) — (14.31), *ограничена. В этом случае*

либо все предельные точки $\{x^k\}$ *не принадлежат допустимой области* $D = \{x \mid h(x) \leqslant 0\}$; *они принадлежат связному подмножеству множества*

$$X_h^* = \{x \mid 0 \in G_h(x),\ h(x) > 0\}$$

и существует предел $\lim\limits_{k \to \infty} h(x^k) > 0$;

либо все предельные точки $\{x^k\}$ *принадлежат допустимой области* D, *тогда минимальные из них принадлежат множеству* $X^* = \{x \mid 0 \in G(x),\ h(x) \leqslant 0\}$ *и отрезок* $[\varliminf\limits_{k \to \infty} f(x^k),\ \varlimsup\limits_{k \to \infty} f(x^k)]$ *вложен в* *множество $F^* = \{f(x) \mid x \in X^*\}$. Если при этом F^* не содержит интервалов, то все предельные точки $\{x^k\}$ принадлежат связному подмножеству множества X^* и существует предел $\lim\limits_{k\to\infty} f(x^k)$.*

Доказательство. По предположению теоремы последовательность $\{x^k\}$ принадлежит некоторому компакту K. Минимальные по значению h предельные точки $\{x^k\}$ принадлежат $X_h^* \cup D$, ибо в противном случае ввиду свойства 2) леммы 14.2 они не могут быть минимальными.

Обозначим $\underline{h} = \varliminf\limits_{k\to\infty} h(x^k)$, $\overline{h} = \varlimsup\limits_{k\to\infty} h(x^k)$, $H_-^* = H^* \cup \{h \in E_1 \mid h \leqslant 0\}$.

Покажем, что $[\underline{h}, \overline{h}) \subset H_-^*$. Предположим противное. Ранее уже доказано, что $\underline{h} \in H_-^*$. Поэтому существует c такое, что $\underline{h} < c < \overline{h}$ и $c \notin H_-^*$. Выберем d так, что $\underline{h} < c < d < \overline{h}$. Последовательность $\{h(x^k)\}$ бесконечное число раз пересекает отрезок $[c, d]$ от c к d; поэтому существуют индексы k_s и t_s такие, что $\lim\limits_{s\to\infty} x^{k_s} = x'$ и

$$h(x^{k_s}) \leqslant c < h(x^k) < d \leqslant h(x^{t_s}), \qquad k \in (k_s, t_s). \tag{14.38}$$

Ввиду (14.30) имеет место $h(x') = \lim\limits_{s\to\infty} h(x^{k_s}) = \lim\limits_{s\to\infty} h(x^{k_s+1}) = c$; поэтому $x' \notin X_h^* \cup D$. Возьмем ε' такое, что

$$\max\{h(y) \mid \|y - x'\| \leqslant \varepsilon'\} < d.$$

Тогда для $\varepsilon \leqslant \varepsilon'$ минимизирующее свойство 2) леммы 14.2 противоречит построению (14.38). Полученное противоречие доказывает, что $[\underline{h}, \overline{h}) \subset H_-^*$. Так как H_-^* замкнуто, то $[\underline{h}, \overline{h}] \subset H_-^*$.

Если H^* не содержит интервалов, то из доказанного следует, что либо существует предел $\lim\limits_{k\to\infty} h(x^k) > 0$ и тогда все предельные точки $\{x^k\}$ не принадлежат D, либо $\overline{h} = \varlimsup\limits_{k\to\infty} h(x^k) \leqslant 0$ и тогда все предельные точки $\{x^k\}$ принадлежат D. В первом случае все предельные точки $\{x^k\}$ принадлежат X_h^*, ибо в противном случае из свойства 2) леммы 14.2 следует расходимость $\{h(x^k)\}$.

Пусть все предельные точки $\{x^k\}$ принадлежат D. Теперь будем считать, что минимизируется функция $f(x)$ на D. Минимальные по значению $f(x)$ предельные точки $\{x^k\}$ принадлежат X^*, ибо в противном случае из свойства 3) леммы 14.2 следует невозможное — существование еще меньших предельных точек $\{x^k\}$. Так же, как это сделано для $\{h(x^k)\}$, можно доказать, что полуинтервал $[\varliminf\limits_{k\to\infty} f(x^k), \varlimsup\limits_{k\to\infty} f(x^k))$ вложен в F^*. В силу ограниченности D и замкнутости G множество F^* замкнуто; поэтому $[\varliminf\limits_{k\to\infty} f(x^k), \varlimsup\limits_{k\to\infty} f(x^k)] \subset F^*$.

Если F^* не содержит интервалов, то из доказанного следует, что существует $\lim\limits_{k\to\infty} f(x^k)$. Тогда все предельные точки $\{x^k\}$ принадлежат X^*, ибо в противном случае получаем противоречие — из свойства 3) леммы 14.2 следует расходимость $\{f(x^k)\}$.

В силу ограниченности последовательности $\{x^k\}$ и условий (14.25)—(14.30) справедливо $\lim_{k\to\infty}\|x^{k+1}-x^k\|=0$. Тогда, согласно лемме 8.3, множество предельных точек последовательности $\{x^k\}$ связно, т. е. $\{x^k\}$ сходится к связному подмножеству множества X^*. Теорема доказана.

3. Регулировка параметров в методе усредненных градиентов. Соотношения (14.25) — (14.31) представляют собой общие условия сходимости метода усредненных градиентов. Конкретизации подлежат величины ρ_k, ε_k, r_k, λ_{kr}. Величины λ_{kr} определяют способ усреднения предыдущих градиентов, r_k и ε_k определяют число учитываемых (запоминаемых) предыдущих градиентов, ρ_k определяют величину шага вдоль выбранного среднего направления. Все эти параметры должны удовлетворять общим условиям (14.25) — (14.31).

Рассмотрим некоторые регулировки шага ρ_k.

Заметим, что если ρ_k удовлетворяют условиям (14.30), (14.31), то $\rho'_k=M_k\rho_k\,(1/N\leqslant M_k\leqslant N<\infty)$ также удовлетворяет (14.30), (14.31), т. е. в направлении P^k можно делать одномерную минимизацию, если, например, минимизируемая функция убывает.

Регулировка шага R1. Простейшей является программная регулировка шага, при которой заранее задается последовательность $\{\rho_k\}$ такая, что

$$\rho_k\geqslant 0,\quad \lim_{k\to\infty}\rho_k=0,\quad \sum_{k=0}^{\infty}\rho_k=\infty.$$

Регулировка шага R2. Рассмотрим адаптивную (ступенчатую) регулировку шага, в которой признаком уменьшения величины ρ_k является малость расстояния $\rho\,(0,\{g^k,g^{k+1},\dots,g^{s_k}\})$ от нуля до выпуклой оболочки накопленных обобщенных градиентов. Пусть числовые последовательности $\{M_k\}$, $\{u_k\}$, $\{\theta_k\}$ таковы, что $u_k\geqslant 0$, $M_k>0$, $\theta_k>0$;

$$\lim_{k\to\infty}M_k=\lim_{k\to\infty}\theta_k=0,\quad \sup_{k\geqslant 0}u_k<\infty,\quad \sum_{k=0}^{\infty}u_k=\infty.$$

Пусть в дополнение к условиям (14.25) — (14.31) $\lim_{k\to\infty}r_k=\infty$. В этой регулировке шаг — $\rho_k=M_{t_k}u_k$, а индексы t_k и s_k изменяются так:

$$s_0=t_0=0;$$

если $\rho\,(0,\mathrm{co}\,\{g^k,g^{k-1},\dots,g^{s_k}\})\geqslant\theta_{t_k}$, то $s_{k+1}=s_k$, $t_{k+1}=t_k$;

если $\rho\,(0,\mathrm{co}\,\{g^k,g^{k-1},\dots,g^{s_k}\})<\theta_{t_k}$, то $s_{k+1}=k$, $t_{k+1}=t_k+1$.

Докажем сходимость метода усредненных градиентов с регулировкой шага R2. Для этого достаточно проверить условия (14.30), (14.31). Условия (14.25) — (14.29) предполагаются выполненными.

Предположим, что $x^k\to x$, $0\notin G(x)$. Последовательности $\{t_k\}$ и, следовательно, $\{s_k\}$ не могут стремиться к бесконечности, ибо в таком случае при достаточно больших k получается противоречие:

$$\frac{1}{2}\rho\,(0,G(x))\leqslant\rho\,(0,\mathrm{co}\,\{g^{s_{k+1}},g^{s_{k+1}-1},\dots,g^{s_k}\})<\theta_{t_k}\to 0.$$

Итак, для достаточно больших $k \geqslant s$ $t_k = t$ и $s_k = s$; поэтому

$$\sum_{k=0}^{\infty} \rho_k = \sum_{k=0}^{s-1} \rho_k + M_t \sum_{k=s}^{\infty} u_k = \infty.$$

Условие (14.31) проверено. Фактически оно означает невозможность $x^k \to x$, $0 \notin G(x)$.

Проверим теперь условие (14.30). Для простоты будем считать $\Delta^k \equiv 0$. Предположим, что $\{x^k\}$ принадлежит некоторому компакту K. Покажем, что $t_k \to \infty$, откуда будет следовать, что $\lim\limits_{k\to\infty} \rho_k = 0$. Предположим противное: $\lim\limits_{k\to\infty} t_k = t < \infty$ и, следовательно, $\lim\limits_{k\to\infty} s_k = s < \infty$. Тогда для $k \geqslant s$ $t_k = t$ и $s_k = s$; поэтому

$$\rho(0, \operatorname{co}\{g^k, g^{k-1}, \ldots, g^s\}) \geqslant \theta_t \text{ и } \{x^k\} \subset K.$$

Так как $r_k \to \infty$, то для достаточно больших $k \geqslant k'$ $r_k \geqslant s$. Тогда получаем

$$\| x^k - x^{k'} \| = \left\| \sum_{r=k'}^{k-1} \rho_r P^r \right\| \geqslant \lambda_K \theta_t M_s \sum_{r=k'}^{k-1} u_r \to \infty,$$

что противоречит ограниченности $\{x^k\}$. Условие (14.30) проверено. Сходимость метода усредненных градиентов с регулировкой шага R2 доказана.

Если $P^k = \sum\limits_{r=r_k}^{k} \mu_{kr} \bar{g}^r$, где

$$\sum_{r=r_k}^{k} \mu_{kr} = 1, \qquad \mu_{kr} \geqslant 0, \qquad \bar{g}^r = g^r/(\| g^r \| + \varepsilon), \qquad \varepsilon > 0,$$

то в регулировке R2 вместо g^r можно использовать $\bar{g}^r$.

Для управления шагом ρ_k можно применять близкую к R2 регулировку из [90], особенность которой в том, что расстояние $\rho(0, \operatorname{co}\{g^k, g^{k-1}, \ldots, g^{s_k}\})$, во-первых, не вычисляется точно, а, во-вторых, его приближенное значение от итерации к итерации лишь подправляется.

Регулировка шага R3. В основе этой регулировки лежит следующее соображение. Пусть вначале $\rho_k = \mathrm{const}$, тогда $\sum\limits_{k=0}^{\infty} \rho_k = +\infty$. Если траектория метода попадает в окрестность решения, то расстояние $\| x^k - x^s \|$ от точки отсчета x^s до текущей точки x^k начнет колебаться вокруг некоторого значения; поэтому $\| x^k - x^s \| \Big/ \sum\limits_{r=s}^{k-1} \rho_r \to 0$. Это обстоятельство служит признаком достижения окрестности решения и сигнализирует о необходимости уменьшения шага.

Пусть $\{M_k\}$, $\{\theta_k\}$, $\{u_k\}$ — числовые последовательности такие, что $M_k > 0$, $\theta_k > 0$, $u_k \geqslant 0$,

$$\lim_{k\to\infty} M_k = \lim_{k\to\infty} \theta_k = 0, \qquad \sup_{k\geqslant 0} u_k < \infty, \qquad \sum_{k=0}^{\infty} u_k = \infty.$$

В этой регулировке шаг — $\rho_k = M_{t_k} u_k$, а индексы t_k и s_k с ростом k меняются следующим образом:

$t_0 = s_0 = 0$;

если $\| x^k - x^{s_k} \| \Big/ \sum\limits_{r=s_k}^{k-1} \rho_k \geqslant \theta_{t_k}$, то $s_{k+1} = s_k$, $t_{k+1} = t_k$;

если $\| x^k - x^{s_k} \| \Big/ \sum\limits_{r=s_k}^{k-1} \rho_r < \theta_{t_k}$, то $s_{k+1} = k$, $t_{k+1} = t_k + 1$.

Регулировка шага R4. Численные эксперименты показывают, что регулировку шага R3 можно заметно улучшить, если в ней вместо отношения $\| x^k - x^{s_k} \| \Big/ \sum\limits_{r=s_k}^{k-1} \rho_r$ использовать величину

$$R_k = \min_{s_k \leqslant t < k} \left(\| x^k - x^t \| \Big/ \sum_{r=t}^{k-1} \rho_r \right).$$

При этом необходимо помнить точки траектории $x^{s_k}, x^{s_k+1}, \ldots, x^k$.

Регулировка шага R5. Чтобы не запоминать траекторию в регулировке R4, можно, не теряя эффективности, вместо R_k использовать величину

$$T_k = \min \left(\| x^k - x^{s_k} \| \Big/ \sum_{r=s_k}^{k-1} \rho_r, \ \| x^k - x^{k_{\min}} \| \Big/ \sum_{r=k_{\min}}^{k-1} \rho_r \right),$$

где индексы $k_{\min}$ таковы, что

$$h(x^{k_{\min}}) = \min_{s_k \leqslant r < k} h(x^r),$$

если $\min\limits_{s_k \leqslant r < k} h(x^r) > 0$, и

$$f(x^{k_{\min}}) = \min_{\{r | s_k \leqslant r < k, h(x^r) \leqslant 0\}} f(x^r),$$

если $\min\limits_{s_k \leqslant r < k} h(x^r) \leqslant 0$.

Для доказательства сходимости метода усредненных градиентов с регулировками шага R3 — R5 достаточно проверить условия (14.30), (14.31). Проверим их для регулировки R4.

Предположим, что $\lim\limits_{k\to\infty} x^k = x$, $0 \notin G(x)$. Очевидно, в этом случае

$\{x^k\}$ принадлежит некоторому компакту K. Для простоты считаем $\Delta^k \equiv 0$. Обозначим последовательность

$$\{k_l\}_{l\geqslant 0} = \{k \mid R_k < \theta_{t_k}\} = \{k \mid s_{k+1} = k\}.$$

Индексы t_k не могут стремиться к $+\infty$, ибо в таком случае $s_k \to \infty$ и получается противоречие: с одной стороны,

$$R_{k_l} = \min_{k_{l-1}\leqslant k<k_l} \left(\| x^{k_l} - x^k \| \Big/ \sum_{r=k}^{k_l-1} \rho_k\right) < \theta_{t_{k_l}} \to 0, \qquad l \to \infty,$$

а с другой стороны, при достаточно больших k_l

$$R_{k_l} \geqslant \frac{1}{2} \lambda_K \rho (0, G(x)) > 0.$$

Итак, для $k \geqslant k'$ все $t_k = t < \infty$, поэтому

$$\sum_{k=0}^{\infty} \rho_k = \sum_{k=0}^{k'-1} \rho_k + M_t \sum_{k=k'}^{\infty} u_k = +\infty.$$

Условие (14.31) проверено. По существу, оно означает невозможность $x^k \to x$, $0 \notin G(x)$.

Проверим условие (14.30). Пусть $\{x^k\}$ ограничена. Покажем, что $\lim\limits_{k\to\infty} t_k = \infty$; отсюда будет следовать, что $\lim\limits_{k\to\infty} \rho_k = 0$. Предположим противное: $\lim\limits_{k\to\infty} t_k = t < \infty$. Отсюда следует, что и $\lim\limits_{k\to\infty} s^k = s < \infty$. Тогда для $k \geqslant s$ имеет место

$$\| x^k - x^s \| \geqslant \theta_t \sum_{r=s}^{k-1} \rho_r = \theta_t M_t \sum_{r=s}^{k-1} u_r \to \infty,$$

что противоречит ограниченности $\{x^k\}$. Условие (14.30) проверено.

Рассмотрим некоторые способы усреднения градиентов, т. е. рассмотрим регулировки усредняющих множителей λ_{kr} в методе (14.22) — (14.24).

Усреднение предыдущих градиентов применяется для того, чтобы придать алгоритму противоовражные свойства. Если алгоритм начинает болтаться поперек оврага минимизируемой функции, то соответствующее среднее из предыдущих градиентов может указывать приемлемое направление вдоль оврага.

Процедура усреднения P1. В качестве направления P^k в методе (14.22) — (14.24) можно брать ближайший к началу координат вектор из выпуклой оболочки предыдущих градиентов $\{g^{r_k}, \dots, g^k\}$. Такое P^k является направлением наискорейшего спуска в точке x^k для аппроксимирующей функции

$$\bar{f}(x) = f(x^k) + \max_{r_k\leqslant r\leqslant k} (g^r, x - x^k).$$

В этом случае построение направлений P^k сводится к решению последовательности связанных задач квадратичного программирования специального вида. Для решения этих задач можно применять алгоритмы из [33, 170, 186].

Процедура усреднения Р2. Направление P^k в методе (14.22) — (14.24) можно вычислять рекуррентно:

$$P^0 = g^0, \qquad P^k = (1-\alpha_k)\,P^{k-1} + \alpha_k g^k, \qquad 0 \leqslant \alpha_k \leqslant 1,$$

причем периодически необходимо производить так называемое восстановление, т. е. полагать $\alpha_k = 1$. Обозначим через r_k ближайший момент восстановления, т. е. $r_k = \max\{r \mid r \leqslant k, \alpha_r = 1\}$. Тогда $P^k = \sum_{r=r_k}^{k-1} \lambda_{kr} g^r$, где

$$\lambda_{kr_k} = \prod_{s=r_k}^{k} (1-\alpha_s), \qquad \lambda_{kk} = \alpha_k,$$

$$\lambda_{kr} = \alpha_r \prod_{s=r+1}^{k} (1-\alpha_s); \qquad \sum_{r=r_k}^{k} \lambda_{kr} = 1.$$

Здесь α_k можно выбирать, например, из условия

$$\alpha_k = \arg \min_{0\leqslant\alpha\leqslant 1} \|(1-\alpha)\; P^{k-1} + \alpha g^k\|,$$

при этом α_k легко вычисляется аналитически.

Процедура усреднения Р3. Можно применять среднеарифметическое усреднение предыдущих градиентов:

$$P^k = \frac{1}{k - r_k + 1} \sum_{r=r_k}^{k} g^r.$$

Если функция r_k монотонно возрастает, то

$$P^k = \left(1 - \frac{1}{k-r_k+1}\right) P^{k-1} + \frac{1}{k-r_k+1}\, g^k.$$

В процедурах Р1 — Р3 можно использовать нормированные обобщенные градиенты $\bar{g}^r$, особенно это полезно в процедуре Р3.

Ключевой проблемой в методе усредненных градиентов является регулировка величин r_k и ε_k, определяющих число учитываемых (усредняемых, запоминаемых) предыдущих градиентов. Из численных экспериментов следует, что число $k - r_k + 1$ усредняемых градиентов необходимо увеличивать с ростом k; в то же время они должны браться из все меньших и меньших окрестностей точек x^k.

Например, для регулировок шага R2—R4 и усреднения Р3 число r_k можно регулировать следующим образом:

$$r_{k+1} = r_k, \quad \text{если} \quad \sum_{r=r_k}^{k} \rho_r \leqslant M_{t_k};$$

$$r_{k+1} = k, \quad \text{если} \quad \sum_{r=r_k}^{k} \rho_r > M_{t_k}.$$

При этом можно полагать, например, $u_k = (k - r_k + 1)^{1/2}$.

4. **Метод тяжелого шарика**. Специальным случаем метода усредненных градиентов является известный метод тяжелого шарика [100]. Если минимизируется гладкая функция $f(x)$, то последний имеет вид

$$x^0 \in E_n, \qquad x^1 = x^0 - \alpha_0 \nabla f(x^0),$$

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k) + \beta_k (x^k - x^{k-1}), \qquad k \geqslant 1, \tag{14.39}$$

где $\nabla f(x)$ — градиент функции f в точке x, $\alpha_k \geqslant 0$, $\beta_k \geqslant 0$ — шаговые множители. Введение инерционного члена $\beta_k (x^k - x^{k-1})$ в итерационный процесс приводит к ускорению сходимости, траектория движения становится плавной и проходит по дну оврага минимизируемой функции. На гладких овражных функциях процесс (14.39) обладает заметным преимуществом перед градиентным методом [100].

Метод тяжелого шарика можно использовать для глобальной минимизации функций, поскольку за счет инерционного члена он может проскакивать локальные минимумы.

Распространим этот метод на задачи минимизации (невыпуклых негладких) обобщенно дифференцируемых функций при ограничениях.

Пусть решается задача (14.1), (14.2). На задачи с более общими ограничениями метод распространяется точно так же, как это сделано для метода обобщенного градиента в § 10.

Пусть

$$x^0 \in E_n, \qquad x^{k+1} = x^k - \rho_k P^k, \qquad \rho_k \geqslant 0, \tag{14.40}$$

$$P^0 = g^0, \qquad P^k = (1 - \gamma_k) P^{k-1} + \gamma_k g^k, \tag{14.41}$$

$$g^k \in G(x^k), \qquad 0 \leqslant \gamma_k \leqslant 1, \qquad k = 0, 1, \dots, \tag{14.42}$$

где x^k — текущее приближение, P^k — направление движения, ρ_k — шаговый множитель, γ_k — параметр усреднения; отображение $G(x)$ определено в (14.3). Выполнены условия

$$\lim_{k \to \infty} \rho_k = 0, \qquad \sum_{k=0}^{\infty} \rho_k = +\infty, \qquad \lim_{k \to \infty} \frac{\rho_k}{\gamma_k} = 0. \tag{14.43}$$

Метод (14.40) — (14.42) можно привести к виду (14.39):

$$x^{k+1} = x^k - \rho_k \gamma_k g^k - \rho_k (1 - \gamma_k) P^{k-1} =$$

$$= x^k - \rho_k \gamma_k g^k + \frac{\rho_k}{\rho_{k-1}} (1 - \gamma_k)(x^k - x^{k-1}).$$

Из соотношения (14.41) следует

$$P^k = \sum_{r=0}^{k} \mu_r^k g^r; \quad \mu_r^k \geqslant 0, \quad \sum_{r=0}^{k} \mu_r^k = 1;$$

где

$$\mu_0^k = \prod_{i=1}^{k} (1 - \gamma_i), \quad \mu_k^k = \gamma_k; \quad \mu_r^k = \gamma_r \prod_{i=r+1}^{k} (1 - \gamma_i),$$

$$0 < r < k.$$

Представим $(0 \leqslant r_k \leqslant k)$

$$P^k = \sum_{r=0}^{k} \mu_r^k g^r = \sum_{r=r_k}^{k} \mu_r^k g^r + \sum_{r=0}^{r_k - 1} \mu_r^k g^r.$$

Обозначим

$$\bar{Q}^k = \sum_{r=r_k}^{k} \mu_r^k g^r \Big/ \sum_{r=r_k}^{k} \mu_r^k, \quad \bar{\rho}_k = \rho_k \sum_{r=r_k}^{k} \mu_r^k, \tag{14.44}$$

$$\bar{\Delta}^k = \sum_{r=0}^{r_k-1} \mu_r^k g^r \Big/ \sum_{r=r_k}^{k} \mu_r^k. \tag{14.45}$$

Перепишем метод (14.40) в следующем виде:

$$x^{k+1} = x^k - \rho_k P^k = x^k - \bar{\rho}_k (\bar{Q}^k + \bar{\Delta}^k). \tag{14.46}$$

Процесс (14.46) по форме является одним из методов усредненных градиентов. Ограниченность последовательностей (14.40), (14.46) можно обеспечить так же, как это делалось для метода обобщенного градиента в § 9, 10 и для метода усредненных градиентов, т. е. либо за счет выбора достаточно малых шагов ρ_k, либо используя механизм возврата в начальную точку. Подберем индексы r_k в (14.44), (14.45) так, чтобы выполнялись условия (14.25) — (14.31) сходимости метода усредненных градиентов.

Лемма 14.4. *При условиях* (14.43) *существуют индексы* $r_k \leqslant k$ *такие, что* $\lim\limits_{k\to\infty} r_k = +\infty$ *и одновременно*

$$\lim_{k\to\infty} \sum_{r=r_k}^{k} \rho_r = 0, \quad \lim_{k\to\infty} \sum_{r=r_k}^{k} \gamma_k = +\infty. \tag{14.47}$$

Доказательство. Сначала построим последовательность индексов $\{r'_k\}$ такую, что

$$r'_k \leqslant k, \quad \lim_{k\to\infty} r'_k = +\infty, \quad \lim_{k\to\infty} \sum_{r=r'_k}^{k} \rho_k = +\infty.$$

Это можно сделать, например, следующим образом.

Положим $\rho = \sup_{r\geqslant 0} \rho_r$, $c_0 = 0$, $r'_0 = 0$, $k = 0$.

Пусть уже вычислены c_k и r'_k; вычислим c_{k+1} и r'_{k+1}.

Если $\sum\limits_{r=r'_k}^{k} \rho_r < c_k$, то положим $c_{k+1} = c_k$, $r'_{k+1} = r'_k$, иначе $c_{k+1} = c_k + 2\rho$, $r'_{k+1} = r'_k + 1$.

Ясно, что $\lim\limits_{k\to\infty} r_k = +\infty$, $\lim\limits_{k\to\infty} c_k = +\infty$ и

$$\sum_{r=r'_k}^{k} \rho_r \geqslant c_k - \rho \to +\infty \quad \text{при} \quad k \to +\infty.$$

Обозначим $\delta_k = \sup\limits_{r\geqslant r'_k} \rho_r/\gamma_r \to 0$, $k \to +\infty$. Теперь найдем индексы r_k из условия

$$\sum_{r=r_k}^{k} \gamma_r \leqslant \min\left(\frac{1}{\sqrt{\delta_k}}, \sum_{r=r'_k}^{k} \gamma_r\right) < \sum_{r=r_k-1}^{k} \gamma_r.$$

Здесь $\sum\limits_{r=r_k}^{k} \gamma_r \leqslant \sum\limits_{r=r'_k}^{k} \gamma_r$; поэтому $r_k \geqslant r'_k$.

При достаточно больших k

$$\sum_{r=r'_k}^{k} \gamma_r \geqslant \sum_{r=r'_k}^{k} \rho_r \to +\infty, \qquad k \to +\infty;$$

поэтому

$$\sum_{r=r_k}^{k} \gamma_r \geqslant \min\left(\frac{1}{\sqrt{\delta_k}}, \sum_{r=r'_k}^{k} \gamma_r\right) - \gamma_{r_k-1} \to +\infty, \qquad k \to \infty.$$

В то же время имеем

$$\sum_{r=r_k}^{k} \rho_r \leqslant \sup_{r'_k\leqslant r\leqslant k} \frac{\rho_r}{\gamma_r} \sum_{r=r_k}^{k} \gamma_r \leqslant \sqrt{\delta_k} \to 0, \qquad k \to +\infty.$$

Лемма доказана.

Теперь для процесса (14.46) проверим условия (14.25) — (14.31) сходимости метода усредненных градиентов.

Заметим, что

$$\sum_{r=r_k}^{k} \mu_r^k = \sum_{r=r_k}^{k-1} \alpha_r \prod_{i=r+1}^{k} (1-\gamma_i) + \gamma_k = 1 - \prod_{i=r_k}^{k} (1-\gamma_i),$$

$$\sum_{r=0}^{r_k-1} \mu_r^k = \prod_{i=r_k}^{k} (1-\gamma_i),$$

$$\prod_{i=r_k}^{k} (1-\gamma_i) = \exp\left\{\sum_{i=r_k}^{k} \ln(1-\gamma_i)\right\} \leqslant \exp\left\{-\sum_{i=r_k}^{k} \gamma_i\right\}.$$

При условиях (14.43), (14.47) и ограниченности $\{x^k\}$ имеет место

$$\lim_{k\to\infty} \bar{\rho}_k \leqslant \lim_{k\to\infty} \rho_k = 0,$$

$$\sum_{k=0}^{\infty} \bar{\rho}_k \geqslant \sum_{k=0}^{k'} \bar{\rho}_k + \frac{1}{2} \sum_{k=k'+1}^{\infty} \rho_k = +\infty;$$

$$\max_{r_k\leqslant s\leqslant k} \| x^k - x^s \| = \sup_{r_k\leqslant s\leqslant k} \left\| \sum_{r=s}^{k-1} \rho_r \sum_{t=0}^{r} \mu_t^r g^t \right\| \leqslant$$

$$\leqslant \left(\sup_{0\leqslant t\leqslant k} \| g^t \|\right) \sum_{r=r_k}^{k} \rho_r \to 0, \qquad k\to\infty;$$

$$\| \bar{\Delta}^k \| \leqslant \left(\sup_{0\leqslant r<r_k} \| g^r \|\right) \sum_{r=0}^{r_k-1} \mu_r^k \left(1 - \sum_{r=0}^{r_k-1} \mu_r^k\right) =$$

$$= \left(\sup_{0\leqslant r\leqslant r_k} \| g^r \|\right) \prod_{i=r_k}^{k} (1-\gamma_i) \Big/ \left(1 - \prod_{i=r_k}^{k} (1-\gamma_i)\right) \to 0, \quad k\to\infty.$$

Таким образом, в предположении ограниченности последовательности $\{x^k\}$ при условиях (14.43) выполнены все условия (14.27) — (14.31) сходимости метода (14.46) и, следовательно, метода (14.40) — (14.42). Поэтому для метода тяжелого шарика справедливы утверждения теоремы 14.2. Сходимость процесса (14.46) следует также из сходимости метода (14.4) — (14.7).

Так же, как и для метода усредненных градиентов, в процедуре (14.40) — (14.42) можно использовать нормированные обобщенные градиенты (см. замечание 14.2).

Опишем механизм возврата в начальную точку, гарантирующий ограниченность метода тяжелого шарика.

Пусть последовательности $\{\rho_k\}_{k\geqslant 0}$, $\{\gamma_k\}_{k>0}$ удовлетворяют условиям (14.42), (14.43). Запустим метод тяжелого шарика из точки x_0 с параметрами $\{\rho_k\}_{k\geq 0}$, $\{\gamma_k\}_{k\geq 0}$. Если в некоторый момент τ_1 случится $h(x^{\tau_1}) > d$, то вернемся в начальную точку и переобозначим $x^{\tau_1} = x^0$. Опять запустим метод из точки $x^{\tau_1} = x^0$ от момента τ_1 с параметрами $\{\rho_k\}_{k\geq\tau_1}$, $\{\gamma_k\}_{k>\tau_1}$. Если в некоторый момент τ_2 окажется, что $h(x^{\tau_2}) > d$, то вернемся в точку x^0 и переобозначим $x^{\tau_2} = x^0$. Снова запустим метод из точки $x^{\tau_2} = x^0$ от момента τ_2 с параметрами $\{\rho_k\}_{k\geq\tau_2}$, $\{\gamma_k\}_{k>\tau_2}$. И так далее.

Обозначим $\{\gamma'_k\} = \{1, \gamma_1, \ldots, \gamma_{\tau_1-1}, 1, \gamma_{\tau_1+1}, \ldots, \gamma_{\tau_2-1}, 1, \gamma_{\tau_2+1}, \ldots\}$.

Очевидно, последовательности $\{\rho_k\}$, $\{\gamma'_k\}$ удовлетворяют условию (14.43); поэтому существуют индексы $r_k \to +\infty$, удовлетворяющие (14.47) (где вместо γ_k стоят γ'_k).

Формально можно считать, что точки $\{x^{\tau_s}, x^{\tau_s+1}, \ldots, x^{\tau_{s+1}-1}\}$ генерируются методом усредненных градиентов вида (14.44) — (14.46), в котором фактически учитываются только градиенты g^r с $r \geq \tau_s$, так как $\gamma'_{\tau_s} = 1$. Очевидно, что $R_s = \sup\limits_{k\geq\tau_s} \bar{\rho}_k \to 0$, $\Delta_s = \sup\limits_{k\geq\tau_s} \|\bar{\Delta}^k\| \to 0$, если $\tau_s \to \infty$.

Обозначим $\Gamma = \sup\{\|g\| \mid g \in G(y),\ h(y) \leq d\}$. Определим индексы $\{r'_k\}$ такие, что $r'_k = r_k$, если $\max\{\tau_s \mid \tau_s \leq k\} \leq r_k$, и $r'_k = \max\{\tau_s \mid r_k \leq \tau_s \leq k\}$ — в противном случае. Пусть $\varepsilon'_k = \max\limits_{r'_k\leq r\leq k} \|x^k - x^r\|$.

Для усредненных градиентов $\bar{Q}^k$ в данном случае имеет место $\bar{Q}^k \in \mathrm{co}\{G(y) \mid \|y - x^k\| \leq \varepsilon'_k\}$. Справедлива оценка

$$\varepsilon'_k = \max_{r'_k\leq r\leq k} \|x^k - x^r\| \leq \Gamma \sum_{r=r'_k}^{k} \rho_r \leq \Gamma \sum_{r=r_k}^{k} \rho_r \to 0, \quad k\to\infty.$$

Следовательно, $V_s = \sup\limits_{k\geq\tau_s} \varepsilon'_k \to 0$, если $\tau_s \to \infty$.

Поэтому в силу леммы 14.3 метод тяжелого шарика лишь конечное число раз выходит из ограниченной области $\{x \mid h(x) \leq d\}$.

5. Метод овражного шага. Оказывается, специальным случаем рассмотренного выше метода усредненных градиентов является также известный метод овражного шага. Он был предложен в [14], затем значительно обобщен и исследован в [74, 76]. В гладком выпуклом случае метод обладает высокой скоростью сходимости (порядка $O(k^{-2})$, где k — число итераций) и является оптимальным для задач этого класса.

Метод овражного шага может быть использован для глобальной минимизации функций. Он проходит по оврагам минимизируемой функции и обладает определенной инерцией (см. (14.51)), что позволяет проскакивать локальные минимумы. Поэтому с его помощью, регулируя шаговые множители, можно грубо просмотреть овраги функции на предмет обнаружения областей локальных минимумов. Затем из найденных перспективных точек можно сделать локальные спуски.

Мы распространим этот метод на задачи минимизации (невыпуклых негладких) обобщенно дифференцируемых функций при ограничениях. Запишем метод овражного шага для решения задачи (14.1), (14.2). На задачи с более общими ограничениями он распространяется точно так же, как это сделано для метода обобщенного градиента в § 10.

Метод вырабатывает две последовательности $\{x^k\}$ и $\{y^k\}$ согласно следующим правилам:

$$x^0 = y^0 = \bar{x} \in E_n, \tag{14.48}$$

$$y^{k+1} = x^k - \rho_k g^k, \quad g^k \in G(x^k), \tag{14.49}$$

$$x^{k+1} = y^{k+1} + \lambda_k (y^{k+1} - y^k), \quad k = 0, 1, \ldots \tag{14.50}$$

Здесь из точки x^k делается шаг градиентного спуска с шаговым множителем ρ_k, а вдоль направления $y^{k+1} - y^k$ делается овражный шаг с шаговым множителем $\lambda_k \geqslant 0$. Множества $G(x)$ определены в (14.3). Можно использовать нормированные обобщенные градиенты.

Заметим, что в методе (14.48) — (14.50) в точках y^k градиенты функций задачи не вычисляются; поэтому представим метод в форме, не содержащей вспомогательных точек y^k. Исключив y^k и y^{k+1} из (14.50) с помощью (14.49), получим

$$\begin{gathered} x^0 \in E_n, \quad x^1 = x^0 - (1 + \lambda_0) \rho_0 g^0, \\ x^{k+1} = x^k - (1 + \lambda_k) \rho_k g^k + \lambda_k \rho_{k-1} g^{k-1} + \lambda_k (x^k - x^{k-1}), \\ g^k \in G(x^k), \quad g^{k-1} \in G(x^{k-1}), \quad k = 0, 1, \ldots \end{gathered} \tag{14.51}$$

В правой части формулы (14.51) имеется инерционный член $\lambda_k (x^k - x^{k-1})$, что позволяет последовательности $\{x^k\}$, порожденной методом, проскакивать локальные экстремумы.

Из этих соотношений следует

$$x^{k+1} = x^k - \rho_k g^k - \sum_{r=0}^{k} \lambda_r \lambda_{r+1} \ldots \lambda_k \rho_r g^r,$$

$$\begin{aligned} x^{k+1} - x^s = &-(1 + \lambda_s + \lambda_s \lambda_{s+1} + \ldots + \lambda_s \lambda_{s+1} \ldots \lambda_k) \rho_s g^s - \\ &-(1 + \lambda_{s+1} + \lambda_{s+1} \lambda_{s+2} + \ldots + \lambda_{s+1} \ldots \lambda_k) \rho_{s+1} g^{s+1} - \ldots \\ &\ldots - (1 + \lambda_k) \rho_k g^k + (\lambda_s + \lambda_s \lambda_{s+1} + \ldots + \lambda_s \lambda_{s+1} \ldots \lambda_k)(x^s - x^{s-1}) + \\ &+ (\lambda_s + \lambda_s \lambda_{s+1} + \ldots \lambda_s \lambda_{s+1} \ldots \lambda_k) \rho_{s-1} g^{s-1}. \end{aligned}$$

Введем обозначения:

$$\lambda_{ji} = \lambda_i \lambda_{i+1} \ldots \lambda_j,$$

$$\sigma_{qp} = \lambda_p + \lambda_p \lambda_{p+1} + \ldots + \lambda_p \lambda_{p+1} \ldots \lambda_q.$$

В этих обозначениях имеем

$$x^{k+1} = x^k - \rho_k g^k - \sum_{r=0}^{k} \lambda_{kr}\rho_r g^r,$$

$$x^{k+1} - x^s = -\sum_{r=s}^{k} (1+\sigma_{kr})\rho_r g^r + \sigma_{ks}(x^s - x^{s-1}) + \sigma_{ks}\rho_{s-1} g^{s-1}.$$

Сначала рассмотрим метод овражного шага с так называемым восстановлением. Будем полагать в некоторые моменты (восстановления) k_s овражный шаг λ_{k_s} равным нулю. Для удобства положим $k_0 = 0$. Тогда

$$x^{k+1} = x^k - \rho_k g^k - \sum_{r=k_s}^{k} \lambda_{kr}\rho_r g^r = x^k - \bar{\rho}_k \bar{P}_k, \quad k_s \leqslant k < k_{s+1}, \tag{14.52}$$

$$\bar{\rho}_k = \rho_k + \sum_{r=k_s}^{k} \lambda_{kr}\rho_r, \quad \bar{P}^k = \left(\rho_k g^k + \sum_{r=k_s}^{k} \lambda_{kr}\rho_r g^r\right)/\bar{\rho}_k.$$

При $k = k_s$ $x^{k_s+1} = x^{k_s} - \rho_{k_s} g^{k_s}$, $\bar{\rho}_{k_s} = \rho_{k_s}$, $\bar{P}^{k_s} = g^{k_s}$. Теперь ясно, что этот метод принадлежит классу методов усредненных градиентов (при соответствующем выборе параметров). Ограниченность метода можно обеспечить так же, как это сделано для метода усредненных градиентов (14.23): возвращать процесс в начальную точку, если он забирается слишком высоко по функции $h(x)$, и в этот момент производить восстановление.

Пусть шаги ρ_k удовлетворяют условиям (14.30), (14.31). Обозначим $r_k = \max\limits_{s \geqslant 0}\{k_s \mid k_s \leqslant k\}$. Предположим, что выполнены условия

$$\max_{r\in[r_k, k]} \sigma_{kr} \leqslant C < +\infty, \quad \lim_{k\to\infty} \sum_{r=r_k}^{k} \rho_r = 0, \tag{14.53}$$

что имеет место, например, когда $\lambda_k \leqslant \lambda < +\infty$ и $k - r_k \leqslant \text{const}$.

Так как $\lambda_{kr} \leqslant \sigma_{kr} \leqslant C$ при $r \in [r_k, k]$, то

$$\sum_{r=r_k}^{k} \lambda_{kr}\rho_r \leqslant \sum_{r=r_k}^{k} \sigma_{kr}\rho_r \leqslant C \sum_{r=r_k}^{k} \rho_r \to 0, \quad k \to \infty.$$

Отсюда с учетом (14.53) получаем

$$\lim_{k\to\infty}\left(\rho_k + \sum_{r=r_k}^{k} \lambda_{kr}\rho_r\right) = \lim_{k\to\infty} \sum_{r=r_k}^{k} (1+\sigma_{kr})\rho_r = 0,$$

и таким образом выполнены все условия сходимости (14.25) — (14.31) метода усредненных градиентов (14.52).

Рассмотрим метод овражного шага (14.48) — (14.50) без восстановления. В преобразованной форме он имеет вид

$$x^{k+1} = x^k - (1+\lambda_k)\rho_k g^k + \lambda_k \rho_{k-1} g^{k-1} + \lambda_k (x^k - x^{k-1}) =$$
$$= x^k - \rho_k g^k - \sum_{r=0}^{k} \lambda_{kr} \rho_r g^r.$$

Для простоты рассмотрим программные регулировки параметров. Пусть

$$0 \leqslant \lambda_k \leqslant \lambda < 1, \quad \rho_k \geqslant 0. \quad \lim_{k\to\infty} \rho_k = 0, \quad \sum_{k=0}^{\infty} \rho_k = +\infty. \qquad (14.54)$$

Отметим, что для выпуклого гладкого случая в [74] найдены оптимальные множители λ_k; они обладают свойствами $\lambda_k \leqslant 1$ и $\lim\limits_{k\to\infty} \lambda_k = 1$.

Представим

$$\sum_{r=0}^{k} \lambda_{kr} \rho_r g^r = \sum_{r=r_k}^{k} \lambda_{kr} \rho_r g^r + \lambda_{kr_k} \sum_{r=0}^{r_k-1} \lambda_{(r_k-1)r} \rho_r g^r.$$

Предположим, что существуют индексы $r_k \leqslant k$ такие, что $\lim\limits_{k\to\infty} r_k = +\infty$,

$$\lim_{k\to\infty} \sum_{r=r_k}^{k} \rho_r = 0, \quad \lim_{k\to\infty} \sum_{r=r_k}^{k} \lambda^{r_k - r} \rho_r = +\infty. \qquad (14.55)$$

Покажем, что это предположение выполнено, если, например, $\rho_k = c/k^\alpha$, $0 < \alpha < 1$. Имеет место

$$\sum_{r=r_k}^{k} \rho_r \leqslant \frac{c}{1-\alpha} \left[(k+1)^{1-\alpha} - r_k^{1-\alpha}\right] \leqslant \frac{c}{1-\alpha} \frac{k+1-r_k}{k^\alpha}.$$

Для монотонно убывающих $\{\rho_k\}$ с использованием суммирования по частям получаем $(q = 1/\lambda > 1)$

$$\sum_{r=r_k}^{k} \lambda^{r_k - r} \rho_r = \sum_{r=r_k}^{k} q^{r-r_k} \rho_r \geqslant \frac{1}{q-1} \left(q^{k+1-r_k} \rho_k - \rho_{r_k}\right).$$

Выберем r_k так, чтобы $r_k \to +\infty$ и $\sum\limits_{r=r_k}^{k} \rho_r$ убывала достаточно медленно при $k \to \infty$, например возьмем $r_k \sim k + 1 - k^\beta$, $0 < \beta < \alpha < 1$. Тогда условия (14.55) заведомо выполнены.

Обозначим

$$\bar{\rho}_k = \rho_k + \sum_{r=r_k}^{k} \lambda_{kr} \rho_r, \quad \bar{Q}^k = \left(\rho_k g^k + \sum_{r=r_k}^{k} \lambda_{kr} \rho_r g^r\right) / \bar{\rho}_k,$$

$$\bar{\Delta}^k = \frac{\lambda_{kr_k}}{\bar{\rho}_k} \sum_{r=0}^{r_k-1} \lambda_{(r_k-1)r} \rho_r g^r.$$

Представим

$$x^{k+1} = x^k - \rho_k(\bar{Q}^k + \bar{\Delta}^k). \tag{14.56}$$

Ясно, что этот метод попадает в схему метода усредненных градиентов и также метода (14.4), (14.5). Проверим условия сходимости (14.28) — (14.33). В силу (14.54), (14.55) для ограниченных последовательностей $\{x^k\}$ имеют место соотношения:

$$\lim_{k\to\infty} \bar{\rho}_k \leqslant \lim_{k\to\infty}\left(\rho_k + \sum_{r=r_k}^{k} \rho_r\right) = 0, \tag{14.57}$$

$$\sum_{k=0}^{\infty} \bar{\rho}_k \geqslant \sum_{k=0}^{\infty} \rho_k = +\infty, \tag{14.58}$$

$$\left\| \sum_{r=0}^{r_k-1} \lambda_{(r_k-1)r}\rho_r g^r \right\| \leqslant \frac{1}{1-\lambda} \sup_{r\geqslant 0} \rho_r \sup_{r\geqslant 0} \| g^r \|, \tag{14.59}$$

$$\lim_{k\to\infty} \bar{\Delta}^k = 0, \ \lim_{r\to\infty} \|x^r - x^{r-1}\| = 0, \tag{14.60}$$

$$\max_{r_k\leqslant r\leqslant k} \| x^k - x^r \| \leqslant$$

$$\leqslant \sum_{r=r_k}^{k} (1+\sigma_{kr})\rho_r \| g^r \| + \sigma_{kr_k} \| x^{r_k} - x^{r_k-1} \| + \sigma_{kr_k}\rho_{r_k-1} \| g^{r_k-1} \| \leqslant$$

$$\leqslant \frac{1}{1-\lambda} \sup_{r\geqslant 0} \| g^r \| \sum_{r=r_k}^{k} \rho_r + \frac{1}{1-\lambda} \| x^{r_k} - x^{r_k-1} \| +$$

$$+ \frac{\lambda}{1-\lambda}\rho_{r_k-1} \| g^{r_k-1} \| \to 0, \ k\to\infty. \tag{14.61}$$

Таким образом, для метода (14.56) в предположении ограниченности $\{x^k\}$ выполнены все условия (14.7) и (14.25) — (14.31); поэтому для метода овражного шага справедливы утверждения теорем 14.1 и 14.2. То же имеет место, если в методе овражного шага (14.48) — (14.50) использовать нормированные обобщенные градиенты.

Опишем детальнее механизм возврата в начальную точку, гарантирующий ограниченность метода овражного шага.

Пусть последовательности $\{\rho_k\}_{k\geqslant 0}$, $\{\lambda_k\}_{k>0}$ удовлетворяют условиям (14.54). Запустим метод овражного шага из точки x^0 до тех пор, пока $h(x^k) \leqslant d$. Если в некоторый момент будет $h(x^{\tau_1}) > d$, то вернемся в начальную точку и переобозначим $x^{\tau_1} = x^0$. Опять запустим метод из точки $x^{\tau_1} = x^0$ от момента τ_1 с параметрами $\{\rho_k\}_{k\geqslant\tau_1}$, $\{\lambda_k\}_{k>\tau_1}$. Если в некоторый момент τ_2 окажется $h(x^{\tau_2}) > d$, то вернемся в начальную точку и переобозначим $x^{\tau_2} = x^0$. Снова запустим метод из точки $x^{\tau_2} = x_0$ от момента τ_2 с параметрами $\{\rho_k\}_{k\geqslant\tau_2}$, $\{\lambda_k\}_{k>\tau_2}$. И так далее.

Обозначим $\{\lambda_k'\} = \{0, \lambda_1, \ldots, \lambda_{\tau_1-1}, 0, \lambda_{\tau_1+1}, \ldots, \lambda_{\tau_2-1}, 0, \ldots\}$.

Очевидно, последовательности $\{\rho_k\}_{k\geqslant 0}$, $\{\lambda_k'\}_{k\geqslant 0}$ удовлетворяют условиям (14.54), (14.55), где вместо λ_k стоят λ_k'; индексы $r_k \to +\infty$.

Формально можно считать, что точки $\{x^{\tau_s}, x^{\tau_s+1}, \ldots, x^{\tau_{s+1}-1}\}$ генерируются методом усредненных градиентов вида (14.56), в котором фактически учитываются только градиенты g^r с $r \geqslant \tau_s$, так как $\lambda_{\tau_s}' = 0$. Очевидно, $R_s = \sup\limits_{k\geqslant\tau_s} \bar{\rho}_k \to 0$, если $\tau_s \to \infty$. В силу (14.60) $\Delta_s = \sup\limits_{k\geqslant\tau_s} \| \Delta^k \| \to 0$, если $\tau_s \to \infty$.

Обозначим $\Gamma = \sup\{\|g\| \mid g \in G(y),\ h(y) \leqslant d\}$. Определим индексы $\{r_k'\}$ такие, что $r_k' = r_k$, если $\max\{\tau_s \mid \tau_s \leqslant k\} \leqslant r_k$, и $r_k' = \max\{\tau_s \mid r_k \leqslant \tau_s \leqslant k\}$ — в противном случае. Пусть $\varepsilon_k' = \max\limits_{r_k' \leqslant r \leqslant k} \| x^k - x^r \|$. Для усредненных градиентов $\bar{Q}^k$ в данном случае имеет место $\bar{Q}^k \in \mathrm{co}\,\{G(y) \mid \| y - x^k \| \leqslant \varepsilon_k'\}$. Очевидно, выполнены соотношения (14.57)—(14.60), а также следующие оценки, аналогичные (14.61):

если $r_k' = \tau_s$, то

$$\varepsilon_k' \leqslant \sum_{r=\tau_s}^{k} (1 + \sigma_{kr}) \rho_r \| g^r \| \leqslant \frac{\Gamma}{1-\lambda} \sum_{r=\tau_s}^{k} \rho_r \leqslant \frac{\Gamma}{1-\lambda} \sum_{r=r_k}^{k} \rho_r \to 0;$$

если $r_k' > \tau_s$, то

$$\varepsilon_k' \leqslant \frac{\Gamma}{1-\lambda} \sum_{r=r_k}^{k} \rho_r + \sigma_{kr_k} \| x^{r_k} - x^{r_k-1} \| + \sigma_{kr_k} \rho_{r_k-1} \| g^{r_k-1} \| \to 0.$$

Кроме того, для $k > \tau_s$ имеет место

$$\| x^k - x^{k-1} \| \leqslant \bar{\rho}_k (\| \bar{Q}^k \| + \| \bar{\Delta}^k \|) \leqslant (\Gamma + \sup_{k\geqslant 0} \| \bar{\Delta}^k \|) \sup_{k\geqslant\tau_s} \bar{\rho}_k.$$

Отсюда следует, что $V_s = \sup\limits_{k\geqslant\tau_s} \varepsilon_k' \to 0$, если $\tau_s \to +\infty$. Поэтому в силу леммы 14.3 метод овражного шага лишь конечное число раз выходит из ограниченной области $\{x \mid h(x) \leqslant d\}$.

В заключение параграфа заметим, что при построении методов усредненных градиентов мы опирались на устойчивость (лемма 14.1 и теорема 14.1) исходного метода обобщенного градиента. Аналогично, имея какой-либо устойчивый в том же смысле базовый метод вида $x^{k+1} \in x^k - \rho_k G(x^k)$, где $G(x)$ уже не градиентное отображение, можно строить методы усредненных направлений и аналоги методов тяжелого шарика и овражного шага.

ГЛАВА 5

РЕШЕНИЕ ЭКСТРЕМАЛЬНЫХ ЗАДАЧ С ЛИПШИЦЕВЫМИ ФУНКЦИЯМИ ПРИ ОГРАНИЧЕНИЯХ

В этой главе изучаются более сложные, чем в гл. 2, задачи минимизации липшицевых функций с наличием ограничений. Для задач с линейными ограничениями исследуются аналоги известных в нелинейном программировании методов условного и приведенного градиентов. Для задач с нелинейными ограничениями рассматривается обобщенный вариант метода возможных направлений и методы штрафных функций. Показано, что точные штрафы для липшицевых функций, как и для выпуклых функций, обладают глобальным свойством.

§ 15. Метод условного градиента

Метод условного градиента был вначале предложен для нахождения экстремума гладких функций при наличии ограничений. Наиболее прямой подход к решению задачи

$$f(x) \to \min \tag{15.1}$$

при

$$x \in D, \tag{15.2}$$

где D — выпуклое ограниченное и замкнутое множество, заключается в линеаризации функции $f(x)$: для данной допустимой точки $x^k \in D$ находят решение $\bar{x}^k$ задачи, которая имеет те же ограничения (15.2), а в качестве функции цели—линейную аппроксимацию функции $f(x)$ в точке x^k. Затем полагают, что направление $d^k = \bar{x}^k - x^k$ является подходящим направлением для поиска минимума. Поэтому для определения направления движения d^k решают задачу

$$(\nabla f(x), y) \to \min \tag{15.3}$$

при условии

$$y \in D. \tag{15.4}$$

В том случае, когда множество D образовано линейными ограничениями, задача (15.3), (15.4) решается известными методами линейного программирования.

Оказалось, что общую схему метода условного градиента можно сохранить и при решении негладких экстремальных задач. Основное отличие заключается в построении вспомогательной последовательности таких векторов z^k, что

$$z^{k+1} = z^k + a_k (H(x^k, \alpha_k) - z^k), \tag{15.5}$$

где a_k — некоторые неотрицательные множители, $H(x^k, \alpha_k)$ определяется по одной из формул:

$$\frac{1}{2\alpha_k} \sum_{i=1}^{n} [f(\tilde{x}_1^k, \ldots, x_i^k + \alpha_k, \ldots, \tilde{x}_n^k) -$$

$$- f(\tilde{x}_1^k, \ldots, x_i^k - \alpha_k, \ldots, \tilde{x}_n^k)] e_i, \qquad (15.6)$$

$$\sum_{i=1}^{n} \frac{f(\tilde{x}^k + \Delta_k e_i) - f(\tilde{x}^k)}{\Delta_k} e_i, \qquad (15.7)$$

где $\tilde{x}^k$ — независимые случайные величины, равномерно распределенные в n-мерном кубе с центром x^k и ребром $2\alpha_k$.

Направление z^{k+1} представляет собой выпуклую комбинацию z^k и вектора $H(x^k, \alpha_k)$, что делает его похожим на выбор спуска в методах сопряженных градиентов. Подробнее со свойствами операции усреднения (15.5) можно ознакомиться в гл. 7.

Конечно-разностные методы решения задачи (15.1), (15.2) при липшицевой функции $f(x)$ определяются соотношениями

$$x^{k+1} = x^k + \rho_k(\bar{x}^k - x^k), \qquad 0 \leqslant \rho_k \leqslant 1, \qquad (15.8)$$

где $\bar{x}^k$ — решение задачи

$$\min_x (z^k, x), \qquad (15.9)$$

$$x \in D. \qquad (15.10)$$

Если существует способ вычисления обобщенного градиента функции $f(x)$, то в формуле (15.5)

$$H(x^k, \alpha_k) = g(\tilde{x}^k), \qquad (15.11)$$

где $g(\tilde{x}^k) \in \partial f(\tilde{x}^k)$.

Пусть область D определяется неравенствами

$$f_i(x) \leqslant 0, \qquad i = 1, \ldots, m,$$

и $f_i(x)$ — выпуклые функции. Полагаем, что выполняется условие регулярности. Тогда множество X^* стационарных точек задачи (15.1), (15.2) выразим условием: $x^* \in X^*$, если найдутся такие векторы $g(x^k) \in \partial f(x^k)$, $g^i(x^k) \in \partial f_i(x^k)$, что

$$g(x^*) = \sum_{i=1}^{m} \lambda_i g^i(x^*), \qquad \lambda_i \leqslant 0,$$

$$\lambda_i f_i(x^*) = 0, \qquad i = 1, \ldots, m.$$

Теорема 15.1. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty}\rho_k=\infty,\qquad \sum_{k=0}^{\infty}a_k^2<\infty,\qquad \frac{\rho_k}{\alpha_k a_k}\to 0,$$

$$\sum_{k=0}^{\infty}\frac{\rho_k^2}{\alpha_k^2}<\infty,\qquad \frac{|\alpha_k-\alpha_{k+1}|}{\rho_k}\to 0,\qquad \alpha_k\to 0.$$

Тогда последовательность $f(x^k)$ сходится с вероятностью 1 и предельные точки x^k с вероятностью 1 принадлежат множеству X^*.

Доказательство. Следуя описанной в §4 технике доказательства, достаточно показать, что выполняются условия (4.12), (4.13). Как это будет видно из дальнейшего, полезно рассмотреть вспомогательную задачу

$$(g(y),\ x)\to \min \tag{15.12}$$

при условиях

$$f_i(x)\leqslant 0,\qquad i=1,\ \dots,\ m, \tag{15.13}$$

где $g(y)\in\partial f(y)$, а точка y удовлетворяет ограничениям (15.13). Пусть $\bar y$ — решение этой задачи. Для любых x со свойством (15.13)

$$(g(y),\ \bar y-x)\leqslant 0.$$

Если $y\notin X^*$, то

$$(g(y),\bar y-y)\leqslant\sigma(y)<0. \tag{15.14}$$

Действительно, предположим, что $(g(y),\ \bar y-y)=0$; иначе говоря, y — решение задачи (15.12), (15.13). Согласно необходимым условиям экстремума задачи (15.12), (15.13),

$$g(y)=\sum_{i=1}^{m}\lambda_i g^i(y),\qquad \lambda_i\leqslant 0,$$

$$\lambda_i f_i(y)=0,\qquad i=1,\ \dots,\ m.$$

Но эти соотношения означают, что $y\in X^*$. Получили противоречие. В действительности имеет место более общий факт:

$$\max_{g(y)\in\partial f(x)}(g(y),\bar y-y)\leqslant\hat\sigma(y)<0. \tag{15.15}$$

Докажем это. Предположим, что существуют такие последовательности $\{\bar y^s\}$ и $\{g^s(y)\}\in\partial f(y)$, что

$$(g^s(y),\ \bar y^s-y)\to 0,$$

где $\bar y^s$ — решение задачи $\min(g^s(y),\ x)$, $x\in D$. Пусть для простоты обозначений $g^s(y)\to\hat g(y)\in\partial f(y)$. Из неравенства (15.14) вытекает, что

$$(\hat g(y),\ \hat y-y)\leqslant\sigma(y)<0,$$

где

$$(\hat{g}(y), \hat{y}) = \min_{x \in D} (\hat{g}(y), x).$$

Выбираем номер s столь большим, чтобы выполнялись неравенства

$$(g^s(y), \bar{y}^s - y) \geqslant \sigma(y)/2,$$

$$\| g^s(y) - \hat{g}(y) \| \leqslant |\sigma(y)| / (8C),$$

где $\| x \| \leqslant C$ для всех $x \in D$. Тогда из соотношения

$$(\hat{g}(y) - g^s(y), \hat{y} - y) + (g^s(y), \hat{y} - y) \leqslant \sigma(y)$$

следует, что

$$(g^s(y), \hat{y} - y) \leqslant \frac{3}{4} \sigma(y).$$

Таким образом,

$$(g^s(y), \hat{y} - y) \leqslant \frac{3}{4} \sigma(y), \qquad (g^s(y), \overline{y^s} - y) \geqslant \frac{\sigma(y)}{2}.$$

Получилось противоречие, так как

$$(g^s(y), y^s) \leqslant (g^s(y), \hat{y}).$$

Итак, выполняется неравенство (15.15). Отсюда следует, что для точек y, принадлежащих достаточно малой окрестности точки $y' \notin X^*$, выполняется соотношение

$$\max \{(g(y), \bar{y} - y) \mid g(y) \in \partial f(y)\} \leqslant \hat{\sigma} < 0. \qquad (15.16)$$

Для сглаженных функций имеем

$$f(x^{k+1}, \alpha_k) = f(x^k, \alpha_k) + (\nabla f(x^k + \tau(x^{k+1} - x^k), \alpha_k), \quad x^{k+1} - x^k) =$$
$$= f(x^k, \alpha_k) + \rho_k (\nabla f(x^k + \tau(x^{k+1} - x^k), \alpha_k) - \nabla f(x^k, \alpha_k), \bar{x}^k - x^k) +$$
$$+ \rho_k (\nabla f(x^k, \alpha_k), \bar{x}^k - x^k) \leqslant f(x^k, \alpha_k) + C\rho_k^2/\alpha_k +$$
$$+ \rho_k (\nabla f(x^k, \alpha_k), \bar{x}^k - x^k),$$

откуда

$$f(x^{k+1}, \alpha_{k+1}) \leqslant f(x^k, \alpha_k) + C\rho_k^2/\alpha_k + C\,|\alpha_k - \alpha_{k+1}| +$$
$$+ \rho_k (\nabla f(x^k, \alpha_k), \bar{x}^k - x^k). \qquad (15.17)$$

Предположим, что существует подпоследовательность $x^s \to x' \notin X^*$ $(s \in S)$. Пусть условие (4.12) не выполняется, т. е. все точки x^k $(k \geqslant s)$ содержатся в достаточно малой δ-окрестности точки x^s, не пересекающейся с X^*. Из неравенства (15.16) следует, что

$$(g(x^s), \hat{x}^s - x^s) \leqslant \hat{\sigma} < 0,$$

где $\hat{x}^s$ — решение задачи (15.12), (15.13). Для достаточно малых δ

$$(g(x^s),\ \hat{x}^s - x^k) \leqslant \hat{\sigma}/2.$$

Отметим теперь интересное свойство операции z^k : если выполняются условия теоремы, то с вероятностью 1

$$(z^k - \nabla f(x^k, \alpha_k)) \to 0, \qquad k \to \infty.$$

Доказательство этого факта приведено в гл. 7.

Согласно лемме 2.5, начиная с некоторого номера k, расстояние от вектора $\nabla f(x^k, \alpha_k)$ до множества $\partial f(x^s)$ сколь угодно мало, если точки x^k принадлежат достаточно малой окрестности x^s. Тогда в силу того, что $g(x^s)$ — произвольный вектор из $\partial f(x^s)$, для достаточно больших k

$$(z^k,\ \hat{x}^s - x^k) \leqslant \hat{\sigma}/4.$$

Следовательно,

$$(z^k,\ \bar{x}^k - x^k) \leqslant \hat{\sigma}/4,$$

где $\bar{x}^k$ — решение задачи (15.9), (15.10). Поэтому, начиная с некоторого номера k, имеем

$$(\nabla f(x^k, \alpha_k),\ \bar{x}^k - x^k) \leqslant \hat{\sigma}/8, \qquad k \geqslant \bar{s}.$$

Таким образом, в неравенстве (15.17) мы оценили величину $(\nabla f(x^k, \alpha_k),$ $\bar{x}^k - x^k)$. Отсюда следует, что

$$f(x^{k+1}, \alpha_{k+1}) \leqslant f(x^s, \alpha_s) + \frac{\hat{\sigma}}{16} \sum_{r=s}^{k} \rho_r, \tag{15.18}$$

поскольку для достаточно больших k

$$C\left[\frac{\rho_k}{\alpha_k} + \frac{|\alpha_k - \alpha_{k+1}|}{\rho_k}\right] \leqslant \frac{|\hat{\sigma}|}{16}.$$

Переходя в неравенстве (15.18) к пределу по $k \to \infty$, получим противоречие с условием ограниченности $f(x)$. Следовательно, выполняется условие (4.12):

$$k(s) = \min\{r \mid \|x^r - x^s\| > \delta,\ r > s\} < \infty.$$

Из соотношения

$$x^{k(s)} = x^s + \sum_{r=s}^{k(s)-1} \rho_r (\bar{x}^r - x^r)$$

вытекает, что

$$\sum_{r=s}^{k(s)-1} \rho_r \geqslant \frac{\delta}{C}.$$

Поэтому из неравенства (15.18) в силу равномерной сходимости $f(x, \alpha) \to f(x)$ вытекает условие (4.13):

$$\overline{\lim_{s\to\infty}} f(x^{k(s)}) < \lim_{s\to\infty} f(x^s).$$

Теорема доказана.

Следует отметить, что если вместо z^k в соотношения (15.9) подставить непосредственно $H(x^k, \alpha_k)$ или $g(\tilde{x}^k)$, то последовательность $\{x^k\}$ может не сходиться к X^*.

Метод (15.8) — (15.10) обладает более широкими возможностями по сравнению с методом проектирования на допустимое множество. Операция проектирования равносильна решению экстремальной задачи с квадратичной функцией цели. В методе условного градиента операция проектирования заменяется решением задачи с линейной функцией цели; поэтому этот метод целесообразно применять в том случае, когда трудно выполнить операцию проектирования. Его выгодно использовать также в задачах, ограничения которых имеют блочную структуру, когда вспомогательная задача линейного программирования решается достаточно просто.

§ 16. Метод приведенного градиента

Метод приведенного градиента был предложен Вулфом для решения задачи нелинейного программирования с линейными ограничениями

$$\min f(x) \tag{16.1}$$

при условиях

$$Ax = b, \quad x \geqslant 0, \tag{16.2}$$

где A — матрица размера $m \times n$ и ранга m, b — вектор из E_m.

Метод основан на сокращении части базисных переменных с помощью внебазисных независимых переменных, в результате чего получается новая, приведенная задача, содержащая только внебазисные переменные. К приведенной задаче, имеющей простейшие ограничения, может применяться один из методов безусловной минимизации, причем очередное направление спуска выбирается так, чтобы новое приближение удовлетворяло условиям (16.2). Таким методом чаще всего является метод сопряженных градиентов, или квазиньютоновский метод. Схема приведенного градиента обобщает симплекс-процедуру линейного программирования; в ней широко используются приемы работы с разреженными матрицами [71].

Покажем, как метод приведенного градиента можно распространить на задачу с липшицевой функцией $f(x)$. Предположим, что любые m столбцов матрицы A линейно независимы, и если в равенстве $Ax = b$ ненулевые компоненты вектора x отвечают линейно независимым столбцам матрицы A, то таких компонент имеется ровно m. Тогда каждая точка множества $S = \{x \mid Ax = b, x \geqslant 0\}$ имеет по крайней мере m положительных компонент.

Пусть точка x удовлетворяет (16.2). Матрица A может быть представлена в виде $[B, N]$, а вектор x — в виде $[x_B^T, x_N^T]^T$, где B — неособенная матрица размера $m \times m$, $x_B > 0$. Вектор x_B называется *базисным*. Компоненты небазисного вектора x_N могут быть либо положительными, либо нулевыми. Пусть

$$g(x) = [g_B^T(x), \quad g_N^T(x)]^T,$$

где $g_B(x)$ — обобщенный градиент функции $f(x)$ по переменным базисного вектора x_B, а $g_N(x)$ — обобщенный градиент по внебазисным переменным x_N.

Представим вектор направления движения в виде $[d_B^T, d_N^T]^T$. Заметим, что равенство $0 = Ad = Bd_B + Nd_N$ выполняется, если для любого d_N положить $d_B = -B^{-1}Nd_N$. Назовем

$$r^T = [r_B^T, r_N^T] = g^T(x) - g_B^T(x) B^{-1}A = g_N^T(x) - g_B^T(x) B^{-1}N$$

приведенным градиентом. Исследуем $(g(x), d)$:

$$(g(x), d) = (g_B(x), d_B) + (g_N(x), d_N) = [g_N^T(x) - g_B^T(x) B^{-1}N] d_N = \\ = (r_N, d_N).$$

Выберем d_N так, чтобы $(r_N, d_N) < 0$ и $d_j \geqslant 0$, если $x_j = 0$. Для каждой внебазисной компоненты j положим $d_j = -r_j$, если $r_j \leqslant 0$, и $d_j = -x_j r_j$, если $r_j > 0$. Это обеспечивает выполнение неравенства $d_j \geqslant 0$, если $x_j = 0$. Кроме того, $(g(x), d) \leqslant 0$, и строгое неравенство имеет место, если $d_N \neq 0$.

Лемма 16.1. *Пусть d выбирается описанным выше способом, $g(x)$ — произвольный вектор из множества $\partial f(x)$. Тогда, если $d = 0$, то в точке x выполняются условия Куна — Таккера.*

Доказательство. Прежде всего заметим, что вектор d является возможным направлением, так как

$$Ad = Bd_B + Nd_N = B(-B^{-1}Nd_N) + Nd_N = 0.$$

Если переменная x_j базисная, то по предположению $x_j > 0$. Если x_j внебазисная, то компонента d_j может быть отрицательной только в том случае, если $x_j > 0$. Таким образом, $d_j \geqslant 0$, если $x_j = 0$, и, следовательно, d — возможное направление. Кроме того,

$$(g(x), d) = \sum_{j \in I} r_j d_j,$$

где I — множество индексов небазисных переменных. По определению d_j очевидно, что либо $d = 0$, либо $(g(x), d) < 0$.

Пусть в точке x выполняются условия Куна — Таккера. Тогда существуют такие векторы $g(x) \in \partial f(x)$, $u = (u_B^T, u_N^T)^T \geqslant 0$ и v, что

$$\begin{gathered}[g_B^T(x), g_N^T(x)] + v^T[B, N] - [u_B^T, u_N^T] = 0, \\ (u_B, x_B) = 0, \quad (u_N, x_N) = 0.\end{gathered} \qquad (16.3)$$

Так как $x_B > 0$ и $u_B \geqslant 0$, то из равенства $(u_B, x_B) = 0$ вытекает, что $u_B = 0$. Из первого равенства в (16.3) следует, что $v^T = -g_B^T(x) B^{-1}$. Подставляя это выражение во второе равенство в (16.3), получаем

$$u_N^T = g_N^T(x) - g_B^T(x) B^{-1} N.$$

Другими словами, $u_N = r_N$. Таким образом, условия Куна — Таккера сведены к соотношениям $r_N \geqslant 0$ и $(r_N, x_N) = 0$. Однако в силу определения равенство $d = 0$ справедливо тогда и только тогда, когда $r_N \geqslant 0$ и $(r_N, x_N) = 0$. Поэтому, если $d = 0$, то в точке x выполняются условия Куна — Таккера.

Метод приведенного градиента решения задачи (16.1), (16.2), когда $f(x)$ — липшицева функция, определяется следующим образом.

Начальный этап. Выбирают допустимую точку x^1, удовлетворяющую условиям $Ax = b$, $x \geqslant 0$. Положить $k = 1$ и перейти к шагу 1.

Шаг 1. Выбирают I_k — множество индексов m штук наибольших компонент вектора x^k:

$$x_B = \{x_i, i \in I_k\}, \quad B = \{a^j, j \in I_k\}, \quad N = \{a^j, j \notin I_k\}.$$

Вычисляют приведенный градиент

$$r_N^T = (z_N^k)^T - (z_B^k)^T B^{-1} N.$$

Определяют направление спуска d^k:

$$d_j^k = \begin{cases} -r_j, & j \notin I_k, \quad r_j \leqslant 0, \\ -x_j r_j, & j \in I_k, \quad r_j > 0, \end{cases}$$
$$d_B^k = -B^{-1} N d_N. \tag{16.4}$$

Шаг 2. Последовательности точек x^k, z^k пересчитывают по формулам

$$x^{k+1} = x^k + \sigma_k d^k,$$
$$z^{k+1} = z^k + a_k (H(x^k, \alpha_k) - z^k), \tag{16.5}$$

$$\sigma_k = \min\{\rho_k, \lambda\},$$

$$\lambda = \begin{cases} \min\limits_{1 \leqslant j \leqslant n} \left\{ -\dfrac{x_j^k}{d_j^k} \,\middle|\, d_j^k < 0 \right\}, & \exists d_j^k < 0, \\ \infty, \quad d^k \geqslant 0. \end{cases} \tag{16.6}$$

Полагают $k = k + 1$ и переходят к шагу 1.

Теорема 16.1. *Пусть выполняются условия теоремы* 15.1. *Тогда последовательность* $f(x^k)$ *сходится с вероятностью* 1 *и предельные точки* x^k *с вероятностью* 1 *удовлетворяют условиям Куна — Таккера.*

Доказательство. Для доказательства сходимости метода приведенного градиента (16.5) нужно учесть следующие два результата. Первый заключается в том, что

$$(z^k, d^k) < \gamma < 0 \tag{16.7}$$

для всех точек x^k с достаточно большими номерами k, которые содержатся в окрестности точки y, не пересекающейся с X^* — множеством точек Куна — Таккера. Неравенство (16.7) вытекает из лемм 2.5, 16.1, а также свойств z^k и $\partial f(x)$.

Следующий важный момент доказательства заключается в том, что в указанных выше точках шаговый множитель σ_k равен ρ_k. Обозначим через I_s множество индексов m наибольших компонент вектора x^s, использованного для вычисления d^s. Предположим, что существует подпоследовательность $x^s \to \hat{x}$ $(s \in S \subset \{1, \dots, k, \dots\})$, для которой

$$\inf\{\delta_s \mid s \in S\} = 0,$$

где

$$\delta_s = \min\left[\min_i \left\{-\frac{x_i^s}{d_i^s} \,\middle|\, d_i^s < 0\right\}, \infty\right].$$

Тогда существует такое множество индексов $S' \subset S$, что $\delta_s = -x_p^s/d_p^s$ сходятся к нулю при $s \in S'$, а также $I_s = \hat{I}$, где $x_p^s > 0$, $d_p^s < 0$, p — элемент множества $\{1, 2, \dots, n\}$, $\hat{I}$ — множество индексов m наибольших компонент вектора $\hat{x}$. Заметим, что в соответствии с (16.4) последовательность $\{d_p^s\}$ $(s \in S')$ ограничена, и так как $\delta_s \to 0$ $(s \in S')$, то $x_p^s \to 0$. Таким образом, $\hat{x}_p = 0$, т. е. $p \notin \hat{I}$. Но $I_s = \hat{I}$ для $s \in S'$; следовательно, $p \notin I_s$. Так как $d_p^s < 0$, то из (16.4) следует $d_p^s = -x_p^s r_p^s$. Отсюда вытекает

$$\delta_s = -x_p^s/d_p^s = 1/r_p^s.$$

Это показывает, что $r_p^s \to \infty$, а это невозможно в силу ограниченности r^s. Таким образом, существует такое $\delta > 0$, что точка $x^k + \lambda d^k$ допустима для всех $\lambda \in [0, \delta]$ и для всех номеров k, начиная с некоторого.

В дальнейшем доказательство аналогично приведенному для теоремы 15.1.

§ 17. Метод возможных направлений

Перейдем теперь к исследованию численного метода решения общей задачи нелинейного программирования: найти

$$\min f_0(x) \tag{17.1}$$

при условиях

$$f_i(x) \leqslant 0, \qquad i = 1, \dots, m, \tag{17.2}$$

где $f_0(x)$ — липшицева функция, $f_i(x)$ — непрерывно дифференцируемые функции. Здесь используются идеи метода возможных направлений Топкиса — Бейнотта (см. [6]) и метода условного градиента, рассмотренного в § 15.

Напомним, что ненулевой вектор d называется возможным направлением в точке $x \in D$ (D — непустое множество из E_n), если существует такое $\delta > 0$, что $x + \lambda d \in D$ для всех $\lambda \in [0, \delta]$. Метод решения исходной задачи определяется соотношением

$$x^{k+1} = x^k + \gamma_k d^k. \tag{17.3}$$

Вектор x^k является допустимой точкой. Здесь при определении направления движения учитываются как активные, так и неактивные ограничения. Этим он отличается от метода возможных направлений Зойтендейка, в котором поиск направления спуска происходит на основе почти активных ограничений. Возможное направление d^k находится из решения задачи линейного программирования

$$\min y \tag{17.4}$$

при условиях

$$\begin{gathered}(z^k, d) - y \leqslant 0,\\ f_i(x^k) + (\nabla f_i(x^k), d) - y \leqslant 0, \qquad i = 1, \ldots, m,\\ -1 \leqslant d_j \leqslant 1, \qquad j = 1, \ldots, n.\end{gathered} \tag{17.5}$$

Последовательность $\{z^k\}$ строится с помощью операции усреднения по формуле

$$z^{k+1} = z^k + a_k (H(x^k, \alpha_k) - z^k),$$

а вектор $H(x^k, \alpha_k)$ определяется по одной из формул: (15.6), (15.7), (15.11).

Остается определить величину шага γ_k. Пусть область D, определяемая ограничениями (17.2), является ограниченным множеством, для которого выполняется условие регулярности. Положим

$$\rho'_k = \max \{\lambda \mid x^k + \lambda d^k \in D\}.$$

Тогда в формуле (17.3) $\gamma_k = \min \{\rho'_k, \rho_k\}$, где ρ_k выбирается из соотношений

$$\sum_{k=0}^{\infty} \rho_k = \infty, \qquad \rho_k \to 0. \tag{17.6}$$

Определим множество X^* стационарных точек задачи (17.1), (17.2): $x^* \in X^*$, если существуют такие векторы $g^\nu(x^*) \in \partial f_\nu(x^*)$ ($\nu = 0, 1, \ldots, m$) и числа $\lambda_i \leqslant 0$ ($i = 1, \ldots, m$), что

$$g^0(x^*) = \sum_{i=1}^{m} \lambda_i g^i(x^*),$$

$$\lambda_i f_i(x^*) = 0, \qquad i = 1, \ldots, m.$$

Теорема 17.1. *Пусть выполняются условия теоремы* 15.1, *а также соотношение* (17.6). *Тогда предельные точки последовательности x^k с вероятностью* 1 *принадлежат множеству X^* и последовательность $f_0(x^k)$ сходится почти наверное.*

Доказательство. Как и при исследовании сходимости метода условного градиента, рассмотрим вспомогательную задачу

$$\min y \tag{17.7}$$

при условиях

$$\begin{gathered}(g^0(x), d) - y \leqslant 0, \\ f_i(x) + (\nabla f_i(x), d) - y \leqslant 0, \qquad i = 1, \ldots, m, \\ -1 \leqslant d_j \leqslant 1, \qquad j = 1, \ldots, n,\end{gathered} \tag{17.8}$$

где $g^0(x) \in \partial f_0(x)$.

Пусть $x \notin X^*$. Тогда легко показать, что $\bar{y}(x) < 0$, где $\bar{y}(x)$ — решение задачи (17.7), (17.8). Отсюда следует, что для точек u, принадлежащих достаточно малой окрестности точки $x \notin X^*$, выполняется неравенство

$$\max_{g^0(u) \in \partial f_0(u)} \bar{y}(u) < \varepsilon < 0. \tag{17.9}$$

Следующий важный момент доказательства заключается в том, что для всех точек

$$x^k \in U_{2\bar{\delta}}(x^s) \equiv \{x \mid \| x - x^s \| \leqslant 2\bar{\delta}\}$$

при достаточно больших s и малых $\bar{\delta}$ таких, что $x^s \to x' \notin X^*$ $(s \in S)$, выбор шагового множителя в методе (17.3) происходит с $\gamma_k = \rho_k$, т. е. удовлетворяет соотношениям (17.6).

В силу соотношения (17.9) найдется ненулевой вектор $\bar{d}$, который удовлетворяет неравенствам

$$\begin{gathered}(g^0(x^k), \bar{d}) \leqslant \varepsilon, \\ f_i(x^k) + (\nabla f_i(x^k), \bar{d}) \leqslant \varepsilon, \qquad i = 1, \ldots, m, \\ -1 \leqslant \bar{d}_j \leqslant 1, \qquad j = 1, \ldots, n.\end{gathered}$$

Используя результат леммы 2.5, можно подобрать вектор $g^0(x^s) \in \in \partial f_0(x^s)$ и числа s и $\bar{\delta}$ такими, чтобы выполнялись неравенства

$$\begin{gathered}(\nabla f_0(x^k, \alpha_k), \bar{d}) \leqslant \varepsilon/2, \\ f_i(x^k) + (\nabla f_i(x^k), \bar{d}) \leqslant \varepsilon/2, \qquad i = 1, \ldots, m, \\ -1 \leqslant \bar{d}_j \leqslant 1, \qquad j = 1, \ldots, n.\end{gathered}$$

С вероятностью 1

$$(z^k - \nabla f_0(x^k, \alpha_k)) \to 0;$$

поэтому для оптимального решения d^k задачи (17.4), (17.5) справедливы неравенства

$$(z^k,\ d^k) \leqslant \varepsilon/4,$$
$$f_i(x^k) + (\nabla f_i(x^k),\ d^k) \leqslant \varepsilon/4, \qquad i = 1, \dots, m, \tag{17.10}$$
$$-1 \leqslant d_j^k \leqslant 1, \qquad j = 1, \dots, n.$$

Так как функции $f_i(x^k)$ $(i = 1, \dots, m)$ непрерывно дифференцируемы, то из (17.10) вытекает существование такого τ, что для всех $\lambda \in [0, \tau]$

$$f_i(x^k) + (\nabla f_i(x^k + \lambda d^k), d^k) \leqslant \varepsilon/8. \tag{17.11}$$

Поскольку $f_i(x^k) \leqslant 0$ для всех i, то по теореме о среднем получаем

$$f_i(x^k + \lambda d^k) = f_i(x^k) + \lambda(\nabla f_i(x^k + \mu_{ik}\lambda d^k), d^k) =$$
$$= (1 - \lambda) f_i(x^k) + \lambda[f_i(x^k) + (\nabla f_i(x^k + \mu_{ik}\lambda d^k), d^k)] \leqslant$$
$$\leqslant \lambda[f_i(x^k) + (\nabla f_i(x^k + \mu_{ik}\lambda d^k), d^k)], \tag{17.12}$$

где $0 \leqslant \mu_{ik} \leqslant 1$. В силу того, что $\lambda\mu_{ik} \in [0, \tau]$, из (17.10) и (17.11) следует

$$f_i(x^k + \lambda d^k) \leqslant \lambda\varepsilon/8.$$

Это означает, что точка $x^k + \lambda d^k$ допустима для любого $\lambda \in [0, \tau]$. Таким образом, для достаточно больших k величина шага $\gamma_k = \rho_k$. В дальнейшем доказательство опирается на вывод условий (4.12), (4.13) и проводится в какой-то мере уже стандартно.

§ 18. Методы штрафных функций

1. Конечно-разностные методы негладких штрафов. Одним из эффективных способов решения экстремальных задач с ограничениями является сведение этих задач с помощью штрафных функций к задачам безусловной минимизации. Наиболее интересны точные штрафные функции, которые сводят исходную задачу

$$f_0(x) \to \min \tag{18.1}$$

при условиях

$$f_i(x) \leqslant 0, \qquad i = 1, \dots, m, \tag{18.2}$$
$$x \in X \tag{18.3}$$

к однократной минимизации негладкой функции.

В выпуклом программировании показано [37], что таковой является функция вида

$$\Phi(x) = f_0(x) + \sum_{i=1}^{m} r_i \max\{0, f_i(x)\}, \quad r_i > 0. \tag{18.4}$$

Пусть $(x^*,\ u^*)$ — седловая точка функции Лагранжа

$$L(x, u) \equiv f_0(x) + \sum_{i=1}^{m} u_i f_i(x),$$

$$x \in X, \quad u \geqslant 0,$$

т. е. точка, для которой выполняются соотношения

$$f_0(x^*) + \sum_{i=1}^{m} u_i f_i(x^*) \leqslant f_0(x^*) + \sum_{i=1}^{m} u_i^* f_i(x^*) \leqslant f_0(x) + \sum_{i=1}^{m} u_i^* f_i(x),$$
$$x \in X, \quad u \geqslant 0. \tag{18.5}$$

Условия (18.5) можно еще записать так:

$$L(x^*, u^*) = \min_{x \in X} \max_{u \geqslant 0} L(x, u) = \max_{u \geqslant 0} \min_{x \in X} L(x, u).$$

Известно, что если $(x^*,\ u^*)$ — седловая точка функции $L(x,\ u)$, $x \in X$, $u \geqslant 0$, то x^* является оптимальной точкой задачи (18.1) — (18.3). Обратное утверждение справедливо для задачи выпуклого программирования в предположении, что ограничения (18.2), (18.3) удовлетворяют условию регулярности Слейтера, т. е. существует вектор $\bar{x} \in X$, для которого $f_i(\bar{x}) < 0$ $(i = 1,\ \ldots,\ m)$. Важность условия Слейтера связана с тем, что из него следует ограниченность множества компонент вектора u. Если $f_i(x)$ $(i = 0,\ 1, \ldots,\ m)$ — выпуклые функции, X — выпуклое множество, то для функции $\Phi(x)$ (18.4) при $r_i > u_i^*$ задача минимизации эквивалентна задаче (18.1) — (18.3) в смысле совпадения их оптимальных значений и оптимальных множеств.

Нужно отметить, что даже при гладких $f_i(x)$ функция $\Phi(x)$ не будет гладкой; поэтому для ее минимизации нужно применять методы недифференцируемой оптимизации. Субградиентные методы минимизации $\Phi(x)$ хорошо известны; поэтому мы рассмотрим конечно-разностные методы.

В дальнейшем X — ограниченное замкнутое множество, на которое можно легко проектировать; $\Phi(x)$ — функция максимума специального вида, поэтому для ее минимизации применим следующий конечно-разностный метод:

$$x^{k+1} = \pi_X\left(x^k - \rho_k\left[H^0(x^k, \alpha_k) + \sum_{i=1}^{m} r_i^+ H^i(x^k, \alpha_k)\right]\right), \tag{18.6}$$

$$r_i^+ = \begin{cases} 0, & f_i(x^k) \leqslant 0, \\ r_i, & f_i(x^k) > 0, \end{cases}$$

где векторы $H^i(x^k, \alpha_k)$ определяются по одной из формул (15.6), (15.7).

Теорема 18.1. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < \infty, \quad \frac{\Delta_k}{\rho_k} \to 0, \quad \alpha_k \to 0.$$

Тогда с вероятностью 1 *предельные точки последовательности* $\{x^k\}$ *принадлежат множеству* X^* *минимумов задачи* (18.1) — (18.3).

Доказательство. Требуется показать, что выполняются условия (4.12), (4.13). Пусть $x^s \to x' \notin X^*$ $(s \in S)$. Можно указать такие положительные числа $\bar{s}$ и $\bar{\delta}$, что для всех точек

$$x \in U_{2\bar{\delta}}(x^s) \equiv \{x \mid \| x - x^s \| \leqslant 2\bar{\delta}\}, \quad s \leqslant \bar{s},$$

выполняются неравенства

$$\Phi(x) - \Phi(x^*) \geqslant \sigma > 0, \quad x^* \in X^*.$$

Тогда в силу равномерной сходимости функций $f_i(x, \alpha_k)$ к $f_i(x)$ $(i = 0, 1, \ldots, m)$ в указанных точках при достаточно больших k имеем

$$f_0(x^k, \alpha_k) + \sum_{i=1}^{m} r_i^+ f_i(x^k, \alpha_k) - f_0(x^*, \alpha_k) - \sum_{i=1}^{m} r_i^+ f_i(x^*, \alpha_k) \geqslant \frac{\sigma}{2} > 0.$$

Поэтому

$$\left(\nabla f_0(x^k, \alpha_k) + \sum_{i=1}^{m} r_i^+ \nabla f_i(x^k, \alpha_k), x^* - x^k\right) \leqslant -\frac{\sigma}{2}.$$

Покажем, что $k(s) < \infty$, где

$$k(s) = \min\{r \mid \| x^r - x^s \| > \delta, r > s\}, \qquad \delta \leqslant \bar{\delta}/2.$$

Допустим противное: $x^k \in U_\delta(x^s)$ для всех $k > s$. Для простоты выкладок считаем, что векторы $H^i(x^k, \alpha_k)$ определяются по формулам (15.6). Тогда имеет место цепочка неравенств

$$\| x^{k+1} - x^* \|^2 \leqslant \left\| x^k - \rho_k \left[H^0(x^k, \alpha_k) + \sum_{i=1}^{m} r_i^+ H^i(x^k, \alpha_k)\right] - x^* \right\|^2 =$$

$$= \| x^k - x^* \|^2 - 2\rho_k \left(H^0(x^k, \alpha_k) + \sum_{i=1}^{m} r_i^+ H^i(x^k, \alpha_k), x^k - x^*\right) +$$

$$+ 2\rho_k^2 \left\| H^0(x^k, \alpha_k) + \sum_{i=1}^{m} r_i^+ H^i(x^k, \alpha_k) \right\|^2 \leqslant \| x^k - x^* \|^2 -$$

$$- 2\rho_k \left(\nabla f_0(x^k, \alpha_k) + \sum_{i=1}^{m} r_i^+ \nabla f_i(x^k, \alpha_k), x^k - x^*\right) +$$

$$+ 2\rho_k \left(\nabla f_0(x^k, \alpha_k) - H^0(x^k, \alpha_k) + \sum_{i=1}^{m} r_i^+ [\nabla f_i(x^k, \alpha_k) -\right.$$

$$\left. - H^i(x^k, \alpha_k)], x^k - x^*\right) + C\rho_k^2 \leqslant \| x^k - x^* \|^2 - \sigma\rho_k +$$

$$+ 2\rho_k \left(\nabla f_0(x^k, \alpha_k) - H^0(x^k, \alpha_k) + \sum_{i=1}^{m} r_i^+ [\nabla f_i(x^k, \alpha_k) -\right.$$

$$\left. - H^i(x^k, \alpha_k)], x^k - x^*\right) + C\rho_k^2. \qquad (18.7)$$

Положим

$$W(x^k) = \min_{x^* \in X^*} \| x^k - x^* \|^2 = \| x^k - x^*(k) \|^2.$$

Тогда из (18.7) следует, что

$$W(x^{k+1}) \leqslant W(x^k) - \sigma\rho_k + 2\rho_k \Big(\nabla f_0(x^k, \alpha_k) - H^0(x^k, \alpha_k) +$$

$$+ \sum_{i=1}^{m} r_i^+ [\nabla f_i(x^k, \alpha_k) - H^i(x^k, \alpha_k)], x^k - x^*\Big) + C\rho_k^2. \quad (18.8)$$

Просуммировав (18.8) по k и устремив k к бесконечности, получим противоречие, поскольку с вероятностью 1

$$\sum_{k=0}^{\infty} \rho_k \Big(\nabla f_0(x^k, \alpha_k) - H^0(x^k, \alpha_k) + \sum_{i=1}^{m} r_i^+ [\nabla f_i(x^k, \alpha_k) -$$

$$- H^i(x^k, \alpha_k)], x^k - x^*\Big) + \sum_{k=0}^{\infty} C\rho_k^2 < \infty, \quad (18.9)$$

т. е. условие (4.12) доказано.

Из свойств операции проектирования имеем

$$\delta < \| x^{k(s)} - x^s \| \leqslant \sum_{r=s}^{k(s)-1} \| x^r - x^s \| \leqslant C \sum_{r=s}^{k(s)-1} \rho_r.$$

Из соотношения (18.9) вытекает, что для достаточно больших s

$$\sum_{r=s}^{k(s)} \Big\{2\rho_r \Big(\nabla f_0(x^r, \alpha_r) - H^0(x^r, \alpha_r) + \sum_{i=1}^{m} r_i^+ [\nabla f_i(x^r, \alpha_r) -$$

$$- H^i(x^r, \alpha_r)], x^r - x^*\Big) + C\rho_r^2\Big\} \leqslant \frac{\sigma}{2} \sum_{r=s}^{k(s)-1} \rho_r. \quad (18.10)$$

Для достаточно больших s $x^{k(s)} \in U_{2\delta}(x^s)$; поэтому из неравенства (18.10) имеем

$$W(x^{k(s)}) \leqslant W(x^s) - \sigma\delta/(2C),$$

откуда вытекает условие (4.13). Теорема доказана.

Замечание. Негладкие функции $\Phi(x)$ (18.4) имеют преимущество перед штрафными функциями

$$F(x) = f_0(x) + \sum_{i=1}^{m} r_i (\max\{0, f_i(x)\})^2,$$

поскольку минимизация $F(x)$ при условиях $x \in X$ дает приближенное решение задачи (18.1) — (18.3), которое стремится к точному только при $r_i \to \infty$.

Негладкая штрафная функция $\Phi(x)$ (18.4) является, как правило, функцией овражного типа, поэтому для ее минимизации целесообразно использовать процедуры усреднения векторов $H(x^k, \alpha_k)$ $(k=0,1, \ldots, m)$, которые рассматривались в гл. 4.

2. Глобальный характер негладких штрафов. Рассмотрим теперь задачу

$$f_0(x) \to \min \tag{18.11}$$

при условиях

$$f_i(x) \leqslant 0, \qquad i = 1, \ldots, m, \tag{18.12}$$

где $f_i(x)$ — липшицевы функции. Возникает вопрос: будет ли в этом случае функция (18.4) точной штрафной функцией? Исследуем следующую штрафную функцию:

$$P(x, \lambda) = f_0(x) + \lambda \sum_{i=1}^{m} \max(0, f_i(x)), \quad \lambda > 0. \tag{18.13}$$

Пусть $f_0(x) \to \infty$ при $x \to \infty$, $-\infty < \inf f_0(x) < \infty$, и пусть в точках минимума x^* задачи (18.11), (18.12) выполняются условия регулярности: существует такое d, что

$$f_i^0(x^*; d) < 0,$$

$$I(x^*) \equiv \{i \mid f_i(x) = 0\} \quad \forall i \in I(x^*).$$

Обозначим через Y множество точек минимума задачи (18.11), (18.12).

Теорема 18.2. *Существует такое $\bar{\lambda} > 0$, что для $\lambda \geqslant \bar{\lambda}$*

$$\arg\min P(x, \lambda) \in Y.$$

Доказательство проведем от противного. Выберем сколь угодно большое $\lambda > 0$. Пусть x_λ — точка минимума $P(x, \lambda)$, причем $x_\lambda \notin Y$. Естественно считать, что x_λ не удовлетворяет ограничениям (18.12), так как иначе $x_\lambda \in Y$. Поэтому существует индекс j такой, что $f_j(x_\lambda) > 0$. По предположению существует последовательность

$$\{x_{\lambda_s}\}, \quad s \to \infty, \quad \max_{1 \leqslant i \leqslant m} f_i(x_{\lambda_s}) > 0.$$

Без ограничения общности считаем, что $x_{\lambda_s} \to x'$ при $s \to \infty$. Если в свою очередь точка x' также не удовлетворяет ограничениям (18.12), то

$$P(x_{\lambda_s}, \lambda_s) \to \infty, \qquad s \to \infty.$$

Поэтому

$$f_i(x') \leqslant 0, \qquad i = 1, \ldots, m.$$

Поскольку

$$P(x_{\lambda_s}, \lambda_s) < f_0(x^*), \qquad x^* \in Y,$$

то $x_{\lambda_s} \to x' \in Y$. Так как x_{λ_s} — точка минимума $P(x, \lambda_s)$, то из необходимого условия экстремума следует, что

$$0 \in \partial P(x_{\lambda_s}, \lambda_s) =$$

$$= \partial \Big(f_0(x_{\lambda_s}) + \lambda_s \sum_{i \in I^+(x_{\lambda_s})} f_i(x_{\lambda_s}) + \lambda_s \sum_{i \in I(x_{\lambda_s})} \max(0, f_i(x_{\lambda_s})\Big), \qquad (18.14)$$

$$I^+(x_{\lambda_s}) \equiv \{i \mid f_i(x_{\lambda_s}) > 0\},$$

$$I(x_{\lambda_s}) \equiv \{i \mid f_i(x_{\lambda_s}) = 0\}.$$

В силу свойств обобщенных градиентов суммы и максимума липшицевых функций из формулы (18.14) вытекает, что

$$0 \in \partial f_0(x_{\lambda_s}) + \lambda_s \sum_{i \in I^+(x_{\lambda_s})} \partial f_i(x_{\lambda_s}) + \sum_{i \in I(x_{\lambda_s})} \bar{\lambda}_i \partial f_i(x_{\lambda_s}), \qquad (18.15)$$

$$\bar{\lambda}_i = \lambda_s \mu_i, \quad 0 \leqslant \mu_i \leqslant 1.$$

Заметим, что для достаточно больших s выполняется включение

$$I(x_{\lambda_s}) \cup I^+(x_{\lambda_s}) \subset I(x') = \{i \mid f_i(x') = 0\}.$$

Соотношение (18.15) означает, что найдутся векторы $g^i(x_{\lambda_s}) \in \partial f_i(x_{\lambda_s})$ $(i = 0, 1, \ldots, m)$, для которых

$$g^0(x_{\lambda_s}) + \lambda_s \sum_{i \in I^+(x_{\lambda_s})} g^i(x_{\lambda_s}) + \sum_{i \in I(x_{\lambda_s})} \bar{\lambda}_i g^i(x_{\lambda_s}) = 0. \qquad (18.16)$$

Поскольку множество $\{1, \ldots, m\}$ конечно, то можно выбрать подпоследовательность λ_s, для которой $f_i(x_{\lambda_s}) > 0$. Из свойства замкнутости отображений $x \to \partial f_i(x)$ $(i = 0, 1, \ldots, m)$ вытекает, что

$$g^0(x_{\lambda_s}) \to \bar{g}^0(x') \in \partial f_0(x'),$$

$$g^j(x_{\lambda_s}) \to \bar{g}^j(x') \in \partial f_j(x').$$

Поэтому, перейдя к пределу по $s \to \infty$ в соотношении (18.16), получим противоречие с ограниченностью множителей Лагранжа в точках множества Y. Теорема доказана.

§ 19. Конечно-разностный метод минимизации липшицевых функций с ограничениями

Заметим, что выбор штрафных коэффициентов r_i точной штрафной функции $\Phi(x)$ (18.4) представляет собой непростую задачу, поскольку компоненты u_i^* $(i = 1, \ldots, m)$ седловой точки функции Лагранжа известны лишь в исключительных случаях. Предлагаемый метод, с одной стороны, близок к методу штрафных функций, а с другой —

использует идею известного метода линеаризации [105]. Его особенность заключается в том, что штрафной коэффициент вычисляется аналитически.

Рассмотрим задачу

$$f(x) \to \min \tag{19.1}$$

при условии

$$h(x) \leqslant 0, \tag{19.2}$$

где $f(x)$ и $h(x)$ — локально липшицевы функции. Условие (19.2) не сужает задачу, так как система $f_i(x) \leqslant 0$ $(i = 1, \ldots, m)$ с помощью функции $h(x) = \max\limits_{1 \leqslant i \leqslant m} f_i(x)$ сводится к (19.2).

Метод решения задачи (19.1), (19.2) определяется соотношением

$$x^{k+1} = x^k + \rho_k d^k. \tag{19.3}$$

где вектор d^k — решение задачи квадратичного программирования

$$(z_f^k, d) + \frac{1}{2} \| d \|^2 \to \min \tag{19.4}$$

при дополнительном ограничении

$$(z_h^k, d) + h(x^k) \leqslant 0, \tag{19.5}$$

которое присутствует в том случае, когда выполняется условие $h(x^k) \geqslant$ $\geqslant 0$. Здесь x^0, z_f^0, z_h^0 — произвольные начальные приближения,

$$\begin{aligned} z_f^{k+1} &= z_f^k + a_k (H_f(x^k, \alpha_k) - z_f^k), \\ z_h^{k+1} &= z_h^k + a_k (H_h(x^k, \alpha_k) - z_h^k), \end{aligned} \tag{19.6}$$

векторы $H_f(x^k, \alpha_k)$, $H_h(x^k, \alpha_k)$ определяются по формулам (15.6), (15.7).

Далее будет показано, что вектор направления движения d^k вычисляется в виде

$$d^k = -z_f^k - u_k z_h^k,$$

где коэффициент u_k определяется аналитически.

Для доказательства сходимости метода (19.3) — (19.6) удобнее считать, что d^k — решение задачи (19.4), (19.5). Определим множество X^* решений задачи (19.1), (19.2): $x^* \in X^*$, если найдутся такие $g_f(x^*) \in$ $\in \partial f(x^*)$, $g_h(x^*) \in \partial h(x^*)$, $\lambda \geqslant 0$, что

$$\begin{aligned} &g_f(x^*) + \lambda g_h(x^*) = 0, \\ &\lambda h(x^*) = 0, \quad h(x^*) \leqslant 0. \end{aligned} \tag{19.7}$$

Соотношения (19.7) удовлетворяют необходимым условиям экстремума, если выполняется условие регулярности: существует такое d, что

$$h^0(x^*; d) < 0, \quad h(x^*) = 0. \tag{19.8}$$

Из (19.8), в частности, следует, что $0 \notin \partial h(x^*)$ при $h(x^*) = 0$. Считаем, что множество X^* ограничено и все точки x^k принадлежат ограниченной области (см. замечание к теореме 4.1).

Рассмотрим вспомогательную задачу

$$(g_f(y), y) + \frac{1}{2}\|d\|^2 \to \min$$

при дополнительном ограничении

$$(g_h(y), d) + h(y) \leqslant 0, \tag{19.9}$$

если

$$h(y) \geqslant 0, \quad g_f(y) \in \partial f(y), \quad g_h(y) \in \partial h(y).$$

Полагаем, что задача (19.9) разрешима при любом y, принадлежащем произвольному ограниченному множеству из E_n.

Пусть $d(y)$ — решение задачи (19.9). Тогда, если $y \notin X^*$, то $d(y) \neq 0$. Предположим, что, напротив, $d(y) = 0$. Легко заметить, что в этом случае $h(y) \leqslant 0$. Вектор $d(y)$ является решением задачи (19.9) тогда и только тогда, когда существует такое $\lambda(y) \geqslant 0$, что

$$g_f(y) + d(y) + \lambda(y) g_h(y) = 0,$$

$$\lambda(y)((g_h(y), d(y)) + h(y)) = 0, \quad h(y) \geqslant 0.$$

Поэтому эти соотношения при $d(y) = 0$ означают, что в точке y выполняются необходимые условия экстремума (19.7). Следовательно, если $y \notin X^*$, то $d(y) \neq 0$. В силу замкнутости множеств $\partial f(x)$, $\partial h(x)$ и замкнутости отображений $x \to \partial f(x)$, $x \to \partial h(x)$ отсюда вытекает, что для точек x, находящихся в достаточно малой окрестности точки $x' \notin X^*$, справедливы неравенства

$$\inf \| d(x) \| \geqslant c(x') > 0,$$

где нижняя грань берется по всевозможным $g_f(x) \in \partial f(x)$, $g_h(x) \in \partial h(x)$.

Теорема 19.1. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} \left(\frac{\rho_k}{\alpha_k}\right)^2 < \infty, \quad \frac{\rho_k}{\alpha_k a_k} \to 0,$$

$$\sum_{k=0}^{\infty} a_k^2 < \infty, \quad \frac{|\alpha_k - \alpha_{k+1}|}{\rho_k} \to 0, \quad \alpha_k \to 0, \quad \frac{\Delta_k}{\alpha_k} \to 0.$$

Тогда с вероятностью 1 *предельные точки последовательности* $\{x^k\}$ *из* (19.3) — (19.6) *принадлежат множеству* X^*, $f(x^k)$ *сходится почти наверное.*

Доказательство. Для сглаженной функции $h(x, \alpha_k)$ имеем

$$h(x^{k+1}, \alpha_k) = h(x^k, \alpha_k) + (\nabla h(x^k + \beta(x^{k+1} - x^k), \alpha_k), \ x^{k+1} - x^k) =$$
$$= h(x^k, \alpha_k) + (\nabla h(x^k, \alpha_k), x^{k+1} - x^k) +$$
$$+ (\nabla h(x^k + \beta(x^{k+1} - x^k), \alpha_k) - \nabla h(x^k, \alpha_k), x^{k+1} - x^k), \qquad 0 \leqslant \beta \leqslant 1.$$

Отсюда следует, что если $h(x^k) \geqslant 0$, то

$$h(x^{k+1}, \alpha_{k+1}) \leqslant$$

$$\leqslant h(x^k, \alpha_k) + C\frac{\rho_k^2}{\alpha_k} + C|\alpha_k - \alpha_{k+1}| + \rho_k(\nabla h(x^k, \alpha_k) - z_h^k, d^k) +$$

$$+ \rho_k(z_h^k, d^k) \leqslant h(x^k, \alpha_k) - \rho_k h(x^k) + C|\alpha_k - \alpha_{k+1}| + C\frac{\rho_k^2}{\alpha_k} + C\rho_k\gamma_k,$$

где

$$\gamma_k = |(\nabla h(x^k, \alpha_k) - z_h^k, d^k)|.$$

В гл. 7 показывается, что при условиях данной теоремы с вероятностью 1

$$\gamma_k \to 0, \qquad (z_f^k - \nabla f(x^k, \alpha_k)) \to 0.$$

Из условий (4.12), (4.13) легко заметить, что все предельные точки последовательности $\{x^k\}$ удовлетворяют ограничению $h(x) \leqslant 0$.

Для функции $f(x, \alpha_k)$ имеем

$$f(x^{k+1}, \alpha_k) = f(x^k, \alpha_k) + (\nabla f(x^k + \beta(x^{k+1} - x^k), \alpha_k), \quad x^{k+1} - x^k) =$$

$$= f(x^k, \alpha_k) + (\nabla f(x^k + \beta(x^{k+1} - x^k), \alpha_k) - \nabla f(x^k, \alpha_k), x^{k+1} - x^k) +$$

$$+ (\nabla f(x^k, \alpha_k), x^{k+1} - x^k) \leqslant f(x^k, \alpha_k) + C\frac{\rho_k^2}{\alpha_k} +$$

$$+ \rho_k(\nabla f(x^k, \alpha_k) - z_f^k, d^k) + \rho_k(z_f^k, d^k), \qquad 0 \leqslant \beta \leqslant 1.$$

Отсюда следует, что

$$f(x^{k+1}, \alpha_{k+1}) \leqslant f(x^k, \alpha_k) + C\frac{\rho_k^2}{\alpha_k} + C|\alpha_k - \alpha_{k+1}| +$$

$$+ \rho_k(z_f^k, d^k) + \rho_k(\nabla f(x^k, \alpha_k) - z_f^k, d^k). \qquad (19.10)$$

Доказательство сходимости последовательности $\{x^k\}$ к X^* проводится из условий (4.12), (4.13). Пусть условие (4.12) не выполняется, т. е. все точки x^k $(k > \bar{s}, \bar{s} \in S)$ содержатся в достаточно малой δ-окрестности точки x^s, не пересекающейся с X^*, $x^s \to x' \notin X^*$ $(s \in S)$. Если $h(x^k) < 0$, то $z_f^k = -d^k$. Поэтому при достаточно больших k $d^k \neq 0$, что следует из леммы 2.5 и условий, что с вероятностью 1

$$(z_f^k - \nabla f(x^k, \alpha_k)) \to 0, \qquad 0 \notin \partial f(x').$$

Пусть $h(x^k) \geqslant 0$. Так как d^k есть решение задачи (19.4), (19.5), то найдется такое $\lambda_k \geqslant 0$, что

$$z_f^k + d^k + \lambda_k z_h^k = 0, \qquad (19.11)$$

$$\lambda_k((z_h^k, d^k) + h(x^k)) = 0.$$

Отсюда

$$(z_f^k, d^k) = \lambda_k h(x^k) - \| d^k \|^2. \tag{19.12}$$

Рассмотрим задачу

$$\begin{aligned} &(g_f(x'), d) + \frac{1}{2} \| d \|^2 \to \min, \\ &(g_h(x'), d) + h(x') \leqslant 0, \quad h(x') \geqslant 0, \end{aligned} \tag{19.13}$$

где

$$g_f(x') \in \partial f(x'), \qquad g_h(x') \in \partial h(x').$$

Достаточно изучать случай $h(x') = 0$. Как уже отмечалось выше, $d(x') \neq 0$, где $d(x')$ — решение задачи (19.13). Кроме того,

$$(g_f(x'), d(x')) + \frac{1}{2} \| d(x') \|^2 \leqslant \varphi(x') < 0,$$

поскольку $d(x') = 0$ не является решением задачи (19.13).

Покажем, что существует допустимый вектор задачи (19.4), (19.5), незначительно отличающийся от $d(x')$. Пусть

$$\tilde{g}_f(x') = \min_{y \in \partial f(x')} \| z_f^k - y \|,$$

$$\tilde{g}_h(x') = \min_{y \in \partial h(x')} \| z_h^k - y \|.$$

Будем искать решение d^k задачи (19.4), (19.5) в виде

$$d = d(x') + \Delta d, \qquad \Delta d \equiv -\mu \frac{\tilde{g}_h(x')}{\| \tilde{g}_h(x') \|},$$

поскольку при выполнении условия регулярности $0 \notin \partial h(x')$. Тогда $(z_h^k, d) + h(x^k) =$

$$= (z_h^k - \tilde{g}_h(x'), d(x')) + (\tilde{g}_h(x'), d(x')) - \mu \frac{(z_h^k, g_h(x'))}{\| \tilde{g}_h(x') \|} + h(x^k).$$

В силу того что с вероятностью 1

$$(z_h^k - \nabla h(x^k, \alpha_h)) \to 0,$$

при достаточно больших k и малом δ величина $(z_h^k - \tilde{g}_h(x'), d(x'))$ также мала. Так как $(\tilde{g}_h(x'), d(x')) \leqslant 0$, то можно подобрать столь малое $\mu > 0$, чтобы выполнялось неравенство

$$(z_h^k, d) + h(x^k) \leqslant 0.$$

Оценим теперь значение целевой функции (19.4) в точке $d = d(x') + \Delta d$:

$$(z_f^k, d) + \frac{1}{2}\|d\|^2 =$$

$$= (z_f^k - \tilde{g}_f(x'), d) + (\tilde{g}_f(x'), d(x') + \Delta d) + \frac{1}{2}\|d(x')\|^2 =$$

$$= (\tilde{g}_f(x'), d(x')) + (\tilde{g}_f(x'), \Delta d) + \frac{1}{2}\|d(x')\|^2 + (d(x'), \Delta d) +$$

$$+ \frac{1}{2}\|\Delta d\|^2 + (z_f^k - \tilde{g}_f(x'), d).$$

По тем же соображениям для достаточно больших номеров k

$$(z_f^k, d) + \frac{1}{2}\|d\|^2 < \frac{\varphi(x')}{2} < 0.$$

Отсюда следует, что, начиная с некоторого номера k, имеем

$$\|d^k\| > \sigma > 0.$$

Из соотношений (19.10), (19.12) следует, что

$$f(x^{k+1}, \alpha_{k+1}) \leqslant f(x^k, \alpha_k) + \rho_k \lambda_k h(x^k) -$$

$$- \rho_k \|d^k\|^2 + C\frac{\rho_k^2}{\alpha_k} + C|\alpha_k - \alpha_{k+1}| + \rho_k(\nabla f(x^k, \alpha_k) - z_f^k, d^k).$$

Из условий теоремы вытекает, что, начиная с некоторого номера k, при достаточно малом δ

$$\lambda_k h(x^k) + C\left[\frac{\rho_k}{\alpha_k} + \frac{|\alpha_k - \alpha_{k+1}|}{\rho_k}\right] + (\nabla f(x^k, \alpha_k) - z_f^k, d^k) \leqslant \frac{\sigma^2}{2}.$$

Поэтому

$$f(x^{k+1}, \alpha_{k+1}) \leqslant f(x^s, \alpha_s) - \frac{\sigma^2}{2}\sum_{r=s}^{k}\rho_r.$$

Переходя к пределу по $k \to \infty$, получим противоречие. Следовательно, выполняется условие (4.12):

$$k(s) \equiv \min\{r \mid \|x^r - x^s\| > \delta, r > s\} < \infty.$$

Из соотношения

$$x^{k(s)} = x^s + \sum_{r=s}^{k(s)-1}\rho_r d^r$$

следует, что

$$\sum_{r=s}^{k(s)-1}\rho_r > \frac{\delta}{C},$$

поэтому

$$\overline{\lim_{s\to\infty}} f(x^{k(s)}) \leqslant \lim_{s\to\infty} f(x^s) - \sigma^2\delta/(2C),$$

т. е. выполняется условие (4.13).

Легко заметить, что задача (19.4), (19.5) решается аналитически. Если $h(x^k) < 0$, то $d^k = -z_f^k$. Если же $h(x^k) \geqslant 0$, то из соотношений (19.11) следует, что

$$d^k = -z_f^k - u_k z_h^k,$$

где

$$u_k = \max\left\{0, \frac{h(x^k) - (z_f^k, z_h^k)}{\| z_h^k \|^2}\right\}.$$

§ 20. Конечно-разностный метод Эрроу — Гурвица с усреднением

Градиентные методы минимизации выпуклых негладких функций на основе известной процедуры Эрроу — Гурвица [137] к настоящему времени исследованы весьма подробно. В этом параграфе изучается конечно-разностный метод решения задачи

$$f_0(x) \to \min \tag{20.1}$$

при ограничениях

$$f_i(x) \leqslant 0, \qquad i = 1, \ldots, m, \tag{20.2}$$

$$x \in X, \tag{20.3}$$

где $f_i(x)$ $(i = 1, \ldots, m)$ — выпуклые функции, X — выпуклое ограниченное замкнутое множество из E_n.

Введем функцию Лагранжа

$$L(x, u) = f_0(x) + \sum_{i=1}^{m} u_i f_i(x).$$

Рассмотрим следующий метод решения задачи (20.1) — (20.3):

$$x^{k+1} = \pi_X\left(x^k - \rho_k\left[H^0(x^k, \alpha_k) + \sum_{i=1}^{m} u_i^k H^i(x^k, \alpha_k)\right]\right), \tag{20.4}$$

$$u^{k+1} = \pi_U(u^k + \rho_k L_u(x^k, u^k)). \tag{20.5}$$

Здесь $L_u(x, u)$ — градиент функции $L(x, u)$ по переменным u. Ограниченность оптимальных двойственных переменных u следует из условий регулярности, в частности условия Слейтера, поэтому множество U, на которое осуществляется проектирование в формуле (20.5), считаем также выпуклым, замкнутым и ограниченным.

Векторы конечных разностей $H^i(x^k, \alpha_k)$ $(i = 0, 1, \ldots, m)$ определяются по одной из формул, аналогичных (15.6), (15.7). Сходимость траекторий $\{x^k\}$, $\{u^k\}$ к множеству седловых точек лагранжиана $L(x, u)$

будет исследована в чезаровском смысле. Именно, пусть

$$\hat{x}^k = \sum_{s=0}^{k} \rho_s x^s \Big/ \sum_{s=0}^{k} \rho_s, \tag{20.6}$$

$$\hat{u}^k = \sum_{s=0}^{k} \rho_s u^s \Big/ \sum_{s=0}^{k} \rho_s. \tag{20.7}$$

Оказывается, что при естественных предположениях предельные точки последовательностей $\{\hat{x}^k\}$, $\{\hat{u}^k\}$ принадлежат множеству седловых точек $L(x, u)$.

Важно отметить, что усреднение (20.6), (20.7) позволяет избежать наложения дополнительного условия о строгой выпуклости целевой функции $f_0(x)$ или неодинакового выбора шаговых множителей по координатам x и u.

Теорема 20.1. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < \infty, \quad \frac{\Delta_k}{\alpha_k} \to 0, \quad \alpha_k \to 0, \quad \sum_{k=0}^{\infty} \rho_k \alpha_k < \infty. \tag{20.8}$$

Тогда с вероятностью 1 *предельные точки последовательностей* $\{\hat{x}^k\}$, $\{\hat{u}^k\}$ *принадлежат множеству седловых точек* $L(x, u)$.

Доказательство проведем для случая, когда векторы $H^i(x^k, \alpha_k)$ $(i = 0, 1, \ldots, m)$ определяются по формуле (15.6).

Введем обозначения.

$$v^k = H^0(x^k, \alpha_k) + \sum_{i=1}^{m} u_i^k H^i(x^k, \alpha_k),$$

$$\gamma_k = H^0(x^k, \alpha_k) - \nabla f_0(x^k, \alpha_k) + \sum_{i=1}^{m} u_i^k (H^i(x^k, \alpha_k) - \nabla f_i(x^k, \alpha_k)),$$

(X^*, u^*) — множество седловых точек функции $L(x, u)$. Справедливы неравенства

$$\begin{gathered}\|x^{k+1} - x^*\|^2 \leqslant \\ \leqslant \|x^k - \rho_k v^k - x^*\|^2 = \|x^k - x^*\|^2 + 2\rho_k (v^k, x^* - x^k) + \rho_k^2 \|v^k\|^2 \leqslant \\ \leqslant \|x^k - x^*\|^2 + 2\rho_k \Big(\nabla f_0(x^k, \alpha_k) + \sum_{i=1}^{m} u_i^k \nabla f_i(x^k, \alpha_k), x^* - x^k\Big) + \\ + 2\rho_k(\gamma_k, x^* - x^k) + C\rho_k^2 \leqslant \|x^k - x^*\|^2 + 2\rho_k (L(x^*, u^k) - L(x^k, u^k)) + \\ + C\rho_k\alpha_k + 2\rho_k(\gamma_k, x^* - x^k) + C\rho_k^2 \leqslant \|x^k - x^*\|^2 + \\ + 2\rho_k(\bar{L}(x^*) - L(x^k, u^k)) + C\rho_k\alpha_k + C\rho_k^2 + 2\rho_k(\gamma_k, x^* - x^k),\end{gathered} \tag{20.9}$$

$$\bar{L}(x) = \max_{u \in U} L(x, u).$$

Для произвольного вектора двойственных переменных u выполняются неравенства

$$\| u^{k+1} - u \|^2 \leqslant \| u^k + \rho_k L_u (x^k, u^k) - u \|^2 \leqslant$$
$$\leqslant \| u^k - u \|^2 + 2\rho_k (L (x^k, u^k) - L (x^k, u)) + C\rho_k^2. \qquad (20.10)$$

Сложив неравенства (20.9), (20.10), получим

$$\| x^{k+1} - x^* \|^2 + \| u^{k+1} - u \|^2 \leqslant \| x^k - x^* \|^2 + \| u^k - u \|^2 +$$
$$+ 2\rho_k (\overline{L} (x^*) - L (x^k, u)) + C\rho_k\alpha_k + C\rho_k^2 + 2\rho_k (\gamma_k, x^* - x^k).$$

После суммирования по $k = 0, 1, \ldots, s$ имеем

$$\Big(\sum_{k=0}^{s} \rho_k\Big)^{-1} \sum_{k=0}^{s} \rho_k (L (x^k, u) - \overline{L} (x^*)) \leqslant$$
$$\leqslant \Big[\| x^1 - x^* \|^2 + \| u^1 - u \|^2 - \| x^{s+1} - x^* \|^2 - \| u^{s+1} - u \|^2 +$$
$$+ \sum_{k=0}^{s} C (\rho_k^2 + \rho_k\alpha_k) + 2 \sum_{k=0}^{s} \rho_k (\gamma_k, x^* - x^k)\Big] \Big(2 \sum_{k=0}^{s} \rho_k\Big)^{-1}.$$

Из определения $\hat{x}^k$, $\hat{u}^k$ и выпуклости L по x получим

$$L (\hat{x}^k, u) - \overline{L} (x^*) \leqslant$$
$$\leqslant \Big[C + \sum_{k=0}^{s} C (\rho_k^2 + \rho_k\alpha_k) + 2 \sum_{k=0}^{s} \rho_k (\gamma_k, x^* - x^k)\Big] \Big(2 \sum_{k=0}^{s} \rho_k\Big)^{-1}.$$

Таким образом, $\overline{L} (\hat{x}^k) \to \overline{L} (x^*)$ с вероятностью 1, поскольку с вероятностью 1 сходится ряд $\sum_{k=0}^{\infty} \rho_k (\gamma_k, x^* - x^k)$. Другими словами, предельные точки последовательности $\{\hat{x}^k\}$ с вероятностью 1 принадлежат множеству X^*. Аналогично показывается, что с вероятностью 1

$$\underline{L} (\hat{u}_k) \to \max_u \underline{L} (u), \qquad \underline{L} (u) \equiv \min_{x \in X} L (x, u).$$

Теорема доказана.

СЛУЧАЙНЫЕ ЛИПШИЦЕВЫ И СЛУЧАЙНЫЕ ОБОБЩЕННО ДИФФЕРЕНЦИРУЕМЫЕ ФУНКЦИИ

В этой главе изучаются свойства случайных липшицевых и случайных обобщенно дифференцируемых функций, необходимые для обоснования методов стохастического программирования. Доказывается измеримость некоторых обобщенно градиентных отображений по совокупности детерминированных и случайных переменных и приводятся некоторые результаты о перестановке знаков математического ожидания и обобщенного дифференцирования.

§ 21. Измеримые многозначные отображения

Нам понадобятся некоторые сведения из теории (измеримых) многозначных отображений.

Пусть E_n — n-мерное евклидово пространство; 2^{E_n} — множество подмножеств E_n; $\mathcal{B}$ (или $\mathcal{B}_{E_n}$) — σ-алгебра борелевских подмножеств E_n; θ — множество; Σ — σ-алгебра измеримых подмножеств Θ; $\mathcal{B} \times \Sigma$ — минимальная алгебра в $E_n \times \Theta$, содержащая все множества вида $B \times T$, где $B \in \mathcal{B}$, $T \in \Theta$.

1. **Многозначные отображения.** Приведем некоторые общие сведения о многозначных отображениях (подробнее см. [60, 77]).

Определение 21.1. *Многозначным отображением* из пространства X в пространство Y называется всякое отображение

$$G : X \to 2^Y, \quad \operatorname{dom} G = \{x \in X \mid G(x) \neq \varnothing\}.$$

Отображение G называется замкнутозначным (выпуклозначным, компактнозначным), если его значения $G(x)$ являются замкнутыми (выпуклыми, компактными) множествами в Y.

Определение 21.2. Многозначное отображение $G : E_n \to 2^{E_m}$ называется *замкнутым*, если его график $\Gamma = \{(x, y) \in E_{n+m} \mid x \in \operatorname{dom} G, y \in G(x)\}$ является замкнутым множеством в E_{n+m}, т. е. из $x^k \to x$ и $g^k \to g$ $(g^k \in G(x^k))$ следует $g \in G(x)$.

Определение 21.3. Многозначное отображение $G : E_n \to 2^{E_m}$ называется *полунепрерывным сверху*, если для любой точки $x \in \operatorname{dom} G$ и любой окрестности нуля U в E_m найдется окрестность нуля V в E_n такая, что $G(x + V) \subset G(x) + U$, т. е. если для любой точки $x \in \operatorname{dom} G$ и для любого $\varepsilon > 0$ найдется $\delta > 0$ такое, что $\rho(g, G(x)) < \varepsilon$ для всех $g \in G(y)$, $\| y - x \| \leqslant \delta$.

Лемма 21.1. *Многозначное отображение* $G: E_n \to 2^{E_m}$ *полунепрерывно сверху тогда и только тогда, когда для любого* $x \in \operatorname{dom} G \subset E_n$ *и для любых* $x^k \to x$, $x^k \in \operatorname{dom} G$ *уклонение*

$$\Delta(G(x^k), G(x)) = \sup_{g^k \in G(x^k)} \inf_{g \in G(x)} \| g - g^k \| \to 0, \quad k \to \infty.$$

Теорема 21.1. *Пусть* $G: E_n \to 2^{E_m}$ — *многозначное отображение. Тогда:*

замкнутое отображение G *замкнутозначно;*

замкнутозначное полунепрерывное сверху отображение G *замкнуто;*

замкнутое локально ограниченное (ограниченное в каждом компакте) отображение G *полунепрерывно сверху;*

конечное (ограниченное в каждой точке) полунепрерывное сверху отображение G *локально ограничено (ограничено в каждом компакте); в частности, компактнозначное полунепрерывное сверху отображение локально ограничено.*

2. Измеримые многозначные отображения. Рассмотрим многозначные отображения из измеримого пространства (Θ, Σ) в E_n. Пусть

$$G: \Theta \to 2^{E_n}, \quad \operatorname{dom} G = \{\theta \in \Theta \mid G(\theta) \neq \varnothing\}.$$

Определение 21.4. Однозначное отображение $g: \Theta \to E_n$ называется *сечением* (селектором) многозначного отображения $G: \Theta \to 2^{E_n}$, если $g(\theta) \in G(\theta)$ для всех $\theta \in \Theta$.

Определение 21.5. Многозначное отображение $G: \Theta \to 2^{E_n}$ называется *измеримым*, если для любого замкнутого множества $C \subset E_n$ имеет место

$$G^{-1}(C) = \{\theta \in \Theta \mid G(\theta) \cap C \neq \varnothing\} \in \Sigma.$$

Определение 21.6. Многозначное отображение $G: \Theta \to 2^{E_n}$ называется *нормальным*, если оно измеримо и множества $G(\theta)$ замкнуты для всех $\theta \in \operatorname{dom} G$.

Теорема 21.2 [151, 144, 175]. *Следующие утверждения относительно замкнутозначного отображения* $G: \Theta \to 2^{E_n}$ *эквивалентны:*

а) G *измеримо;*

б) *для любого компакта* $K \subset E_n$ *множество*

$$G^{-1}(K) = \{\theta \in \Theta \mid G(\theta) \cap K \neq \varnothing\} \in \Sigma;$$

в) *для любого открытого множества* $B \subset E_n$

$$G^{-1}(B) = \{\theta \in \Theta \mid G(\theta) \cap B \neq \varnothing\} \in \Sigma;$$

г) *существует представление Кастена* [144] *для* G, *т. е.* $\operatorname{dom} G$ *измеримо, и существует счетное семейство измеримых сечений* g_i:

$\Theta \to E_n$ $(l = 1, 2, \ldots)$ *отображения* G *такое, что множество* $\{g_l(\theta), l = 1, 2, \ldots\}$ *плотно в* $G(\theta)$ *для* $\theta \in \Theta$.

Лемма 21.2. *Замкнутое отображение* $G: E_n \to 2^{E_m}$ *является* $\mathcal{B}_{E_n}$*-измеримым (борелевским) и, следовательно, имеет сечения, являющиеся борелевскими функциями.*

Действительно, для любого компакта $K \subset E_m$ множество $\{x \in E_n \mid G(x) \cap K \neq \varnothing\}$ замкнуто и, следовательно, борелево; тогда по теореме 21.2 отображение G $\mathcal{B}_{E_n}$-измеримо и имеет борелевские сечения.

Лемма 21.3 [49]. *Пусть многозначное отображение* $G: x \to G(x) \subset E_m$ *измеримо; тогда отображение* co G: $x \to \mathrm{co}\, G(x) \subset E_m$ *также измеримо.*

Определение 21.7. Всякую функцию $\varphi: E_n \times \Theta \to E_1 \cup \{+\infty\}$ будем называть *интегрантом.*

Обозначим через φ_θ функцию из E_n в E_1 такую, что $\varphi_\theta(x) = \varphi(x, \theta)$; ее надграфик по определению есть множество

$$\mathrm{epi}\, \varphi_\theta = \{(x, \alpha) \in E_{n+1} \mid \alpha \geqslant \varphi_\theta(x)\}.$$

Интегрант φ называется нормальным, если многозначное отображение $\theta \to \mathrm{epi}\varphi_\theta$ нормально.

Интегрант φ называется выпуклым, если функции $\varphi_\vartheta(x)$ выпуклы для всех $\vartheta \in \Theta$.

Определение 21.8. Функция $f: E_n \times \Theta \to E_1$ называется *каратеодориевой*, если она измерима по $\vartheta \in \Theta$ при каждом $x \in E_n$ и непрерывна по $x \in E_n$ для всех $\vartheta \in \Theta$.

Лемма 21.4 [110]. *Каратеодориева функция* f *является нормальным интегрантом.*

Лемма 21.5 [175]. *Если* φ *— нормальный выпуклый интегрант, то многозначное отображение*

$$\theta \to K(\theta) = \{x \in E_n \mid \varphi(x, \theta) \leqslant 0\}$$

нормально.

Определение 21.9. Отображение $g: E_n \times \Theta \to E_m$ называется $\mathcal{B} \times \Sigma$*-измеримым* (или $\mathcal{B}_{E_n} \times \Sigma$-измеримым), если $\{(x,\theta) \in E_n \times \Theta \mid g(x, \theta) \leqslant c\} \in \mathcal{B}_{E_n} \times \Sigma$ для любого $c \in E_m$.

Определение 21.10. Отображение g называется *каратеодориевым*, если оно измеримо по $\vartheta \in \Theta$ при каждом $x \in E_n$ и непрерывно по $x \in E_n$ для всех $\vartheta \in \Theta$.

Лемма 21.6 [151]. *Каратеодориева функция* f *и, как следствие, каратеодориево отображение* g $\mathcal{B} \times \Sigma$*-измеримы.*

Замечание 21.1. В дальнейшем нам придется также иметь дело с отображениями вида $G: E_n \times \Theta \to 2^{E_m}$ и интегрантами вида $\varphi: E_m \times (E_n \times \Theta) \to E_1 \cup \{+\infty\}$. На них очевидным образом переносятся приведенные выше определения и утверждения.

Например, многозначное отображение $G: E_n \times \Theta \to 2^{E_m}$ называется $\mathcal{B} \times \Sigma$-измеримым, если для любого замкнутого множества $C \subset E_m$

$$G^{-1}(C) = \{(x, \theta) \mid G(x, \theta) \cap C \neq \varnothing\} \in \mathcal{B} \times \Sigma.$$

Интегрант $\varphi(g, x, \theta)$: $E_m \times (E_n \times \Theta) \to E_1 \cup \{+\infty\}$ называется нормальным, если многозначное отображение

$$(x, \theta) \to \operatorname{epi} \varphi_{x,\vartheta} = \{(g, \alpha) \in E_{m+1} \mid \alpha \geqslant \varphi(g, x, \theta)\}$$

$\mathcal{B} \times \Sigma$-измеримо и множества $\operatorname{epi} \varphi_{x,\theta}$ непусты и замкнуты для всех (x, θ).

Замечание 21.2. $\mathcal{B} \times \Sigma$-измеримое отображение $G(x, \theta)$: $E_n \times \times \Theta \to 2^{E_m}$ является измеримым по переменной θ при фиксированном x ввиду аналогичного свойства измеримых функций [10, с. 126] и представления Кастена измеримых многозначных отображений (теорема 21.2).

Определение 21.11. *Интегралом (Лебега) многозначного отображения* $G : \Theta \to 2^{E_n}$ называется множество интегралов от всевозможных интегрируемых сечений отображения G:

$$\int_{\Theta} G(\theta) P(d\theta) = \left\{ \int_{\Theta} g(\theta) P(d\theta) \mid g(\theta) \in G(\theta) \right\}.$$

Теорема 21.3 [144]. *Пусть* (Θ, Σ, P) — *измеримое пространство с конечной положительной мерой, многозначное отображение* G: $\Theta \to 2^{E_n}$ *нормально,* $\sup\{\|g\| \mid g \in G(\vartheta)\} \leqslant L(\vartheta)$, *где* $L(\vartheta)$ — *интегрируемая функция. Тогда множество* $\int G(\vartheta) P(\vartheta)$ *компактно в* E_n, *а множество* $\int_{\Theta} \operatorname{co} G(\vartheta) P(d\vartheta)$ *выпукло и компактно в* E_n, *и если мера* P *непрерывна, то*

$$\int_{\Theta} G(\theta) P(d\theta) = \int_{\Theta} \operatorname{co} G(\theta) P(d\theta).$$

Лемма 21.7. *Пусть* (Θ, Σ, P) — *измеримое пространство с конечной положительной мерой, многозначное отображение* $G(x, \theta)$: $E_n \times \Theta \to 2^{E_m}$ *полунепрерывно сверху по* x *в* $K \subset E_n$ *для всех* θ *и измеримо по* θ *при каждом* $x \in K$, *множества* $G(x, \theta)$ — *непустые компакты при всех* (x, θ), $\sup\{\|g\| \mid g \in G(x, \theta), x \in K\} \leqslant L_K(\theta)$, *где* $L_K(\vartheta)$ — *интегрируемая функция. Тогда множества*

$$G(x) = \int_{\Theta} G(x, \theta) P(d\theta)$$

суть непустые компакты (т. е. ограниченные замкнутые множества) и многозначное отображение $x \to G(x)$ *полунепрерывно сверху в* K.

Доказательство. Компактность $G(x)$ следует из теоремы 21.3. Докажем полунепрерывность сверху отображения G. Обозначим

$$\Delta(B, A) = \sup_{b \in B} \inf_{a \in A} \|b - a\|, \qquad A, B \subset E_n.$$

Необходимо показать, что $\lim_{y \to x} \Delta(G(y), G(x)) = 0$. По определению

21.11 для любого $g_z \in G(z)$ существует измеримое сечение $g_z(\theta)$ отображения $G(z, \cdot)\colon \Theta \to 2^{E_m}$ такое, что $g_z = \int\limits_{\Theta} g_z(\theta) P(d\theta)$. Поэтому

$$\Delta(G(y), G(x)) = \sup_{g_y \in G(y)} \inf_{g_x \in G(x)} \| g_y - g_x \| \leqslant$$
$$\leqslant \sup_{g_y(\theta) \in G(y,\theta)} \inf_{g_x(\theta) \in G(x,\theta)} \int\limits_{\Theta} \| g_y(\theta) - g_x(\theta) \| P(d\theta), \tag{21.1}$$

где $g_x(\theta)$, $g_y(\theta)$ — измеримые (и интегрируемые) сечения отображений $G(x, \cdot)$, $G(y, \cdot)$ соответственно. Покажем, что в (21.1) inf можно внести под знак интеграла, т. е.

$$\inf_{g_x(\theta) \in G(x,\theta)} \int\limits_{\Theta} \| g_y(\theta) - g_x(\theta) \| P(d\theta) = \int\limits_{\Theta} \inf_{g_x \in G(x,\theta)} \| g_y(\theta) - g_x \| P(d\theta). \tag{21.2}$$

Покажем, что в (21.2) левая часть не меньше правой. Для интегрируемой функции $g_x(\theta) \in G(x, \theta)$

$$\| g_y(\theta) - g_x(\theta) \| \geqslant \inf_{g_x \in G(x,\theta)} \| g_y(\theta) - g_x \|, \tag{21.3}$$

причем правая часть (21.3) измерима и интегрируема. Действительно, пусть $\{g_x^m(\theta),\ m = 1, 2, \ldots\}$ — счетное плотное семейство сечений $G(x, \cdot)$. В силу непрерывности $\| g_y(\theta) - g_x \|$ по g_x

$$\inf_{g_x \in G(x,\theta)} \| g_y(\theta) - g_x \| = \inf_m \| g_y(\theta) - g_x^m(\theta) \| =$$
$$= \lim_{M \to \infty} \min_{0 \leqslant m \leqslant M} \| g_y(\theta) - g_x^m(\theta) \|. \tag{21.4}$$

Функция $\min\limits_{0 \leqslant m \leqslant M} \| g_y(\theta) - g_x^m(\theta) \|$ измерима, так как измеримы лебеговские множества ($c \in E_1$)

$$\{\theta \mid \min_{0 \leqslant m \leqslant M} \| g_y(\theta) - g_x^m(\theta) \| \leqslant c\} = \bigcup_{0 \leqslant m \leqslant M} \{\theta \mid \| g_y(\theta) - g_x^m(\theta) \| \leqslant c\}.$$

Тогда левая часть (21.4) измерима как предел измеримых функций. Кроме того,

$$\inf_{g_x \in G(x,\theta)} \| g_y(\theta) - g_x \| \leqslant 2L_K(\theta).$$

Таким образом, правая часть в (21.3) интегрируема. Проинтегрируем (21.3) и в левой части возьмем inf по $g_x(\theta) \in G(x, \theta)$; тогда получим

$$\inf_{g_x(\theta) \in G(x,\theta)} \int\limits_{\Theta} \| g_y(\theta) - g_x(\theta) \| P(d\theta) \geqslant \int\limits_{\Theta} \inf_{g_x \in G(x,\theta)} \| g_y(\theta) - g_x \| P(d\theta). \tag{21.5}$$

Теперь докажем обратное неравенство. Покажем, что inf в правой части (21.5) на самом деле достигается не поточечно, а на измеримых сечениях $G(x, \cdot)$. Пусть $\{g_x^m(\theta),\ m = 1, 2, \ldots\}$ — счетное плотное семейство измеримых сечений $G(x, \cdot)$. Возьмем произвольное $\varepsilon > 0$. Рассмотрим множество

$$\Theta_\varepsilon^m = \{\theta \mid \| g_y(\theta) - g_x^m(\theta) \| \leqslant \inf_{g_x \in G(x,\theta)} \| g_y(\theta) - g_x \| + \varepsilon\}.$$

Определим функцию $h_{x,\varepsilon}^m\colon \Theta \to E_n$ следующим образом:

$$h_{x,\varepsilon}^m(\theta) = \begin{cases} g_x^1(\theta), & \theta \in \Theta_\varepsilon^1, \\ g_x^2(\theta), & \theta \in \Theta_\varepsilon^2 \setminus \Theta_\varepsilon^1, \\ \cdots\cdots\cdots\cdots & \\ g_x^m(\theta), & \theta \in \Theta_\varepsilon^m \setminus \bigcup\limits_{i=1}^{m-1} \Theta_\varepsilon^i, \\ g_x^1(\theta), & \theta \in \Theta \setminus \bigcup\limits_{i=1}^{m} \Theta_\varepsilon^i. \end{cases}$$

Очевидно, функция $h_{x,\varepsilon}^m$ является измеримым сечением $G(x, \cdot)$, причем на множестве $T_\varepsilon^m = \bigcup\limits_{i=1}^{m} \Theta_\varepsilon^i$

$$\| g_y(\theta) - h_{x,\varepsilon}^m(\theta) \| \leqslant \inf_{g_x \in G(x,\theta)} \| g_y(\theta) - g_x \| + \varepsilon.$$

Покажем, что $P_\varepsilon^m = P(\theta \setminus T_\varepsilon^m) \to 0$ $(m \to \infty)$. Действительно, $\{T_\varepsilon^m\}_{m \geqslant 1}$ — монотонно возрастающая последовательность измеримых множеств, причем $P(T_\varepsilon^m) \leqslant P(\Theta) < +\infty$. Поэтому предельное множество $T_\varepsilon^\infty = \bigcup\limits_{i=1}^{\infty} \Theta_\varepsilon^i$ измеримо и $P_\varepsilon^\infty = P(T_\varepsilon^\infty) \leqslant P(\Theta)$. Если $P_\varepsilon^\infty < P(\Theta)$, то на множестве $\Theta \setminus T_\varepsilon^\infty$: $P(\Theta \setminus T_\varepsilon^\infty) > 0$ для всех m выполнено

$$\| g_y(\theta) - g_x^m(\theta) \| > \inf_{g_x \in G(x,\theta)} \| g_y(\theta) - g_x \| + \varepsilon.$$

Это означает, что семейство $\{g_x^m(\theta)\}_{m=1}^\infty$ на множестве $\Theta \setminus T_\varepsilon^\infty$ не плотно в $G(x, \theta)$. Полученное противоречие доказывает, что $P_\varepsilon^m = P(\Theta \setminus T_\varepsilon^m) \to 0$ $(m \to \infty)$. Тогда

$$\inf_{g_x(\theta) \in G(x,\theta)} \int_\Theta \| g_y(\theta) - g_x(\theta) \| P(d\theta) \leqslant \int_\Theta \| g_y(\theta) - h_{x,\varepsilon}^m(\theta) \| P(d\theta) =$$

$$= \Big(\int_{T_\varepsilon^m} + \int_{\Theta \setminus T_\varepsilon^m} \Big) \| g_y(\theta) - h_{x,\varepsilon}^m(\theta) \| P(d\theta) \leqslant$$

$$\leqslant \Big(\int_{\Theta} - \int_{\Theta \setminus T_\varepsilon^m}\Big) \inf_{g_x \in G(x,\theta)} \| g_y(\vartheta) - g_x \| P(d\vartheta) +$$

$$+ \varepsilon + \int_{\Theta \setminus T_\varepsilon^m} \| g_y(\vartheta) - h_{x,\varepsilon}^m(\vartheta) \| P(d\vartheta), \qquad (21.6)$$

причем остаточные интегралы в правой части (21.6) по множествам $\Theta \setminus T_\varepsilon^m$ в силу абсолютной непрерывности интеграла Лебега стремятся к нулю при $m \to +\infty$. Переходя в (21.6) к пределу сначала по $m \to +\infty$, а затем по $\varepsilon \to 0$, получаем

$$\inf_{g_x(\theta) \in G(x,\theta)} \int_{\Theta} \| g_y(\theta) - g_x(\theta) \| P(d\theta) \leqslant \int_{\Theta} \inf_{g_x \in G(x,\theta)} \| g_y(\theta) - g_x \| P(d\theta). \qquad (21.7)$$

Из (21.5) и (21.7) следует требуемое равенство (21.2). Из (21.1) и (21.2) получаем

$$\Delta(G(y), G(x)) \leqslant \sup_{g_y(\theta) \in G(x,\theta)} \int_{\Theta} \inf_{g_x \in G(x,\theta)} \| g_y(\theta) - g_x \| P(d\theta). \qquad (21.8)$$

Совершенно аналогично доказывается, что под знак интеграла в (21.8) можно внести sup. Необходимо только показать, что функция $\inf \{ \| g_y - g_x \| \mid g_x \in G(x,\theta) \}$ непрерывна по g_y. В самом деле,

$$\inf_{g_x \in G(x,\theta)} \| g_y^2 - g_x \| - \| g_y^1 - g_y^2 \| \leqslant$$

$$\leqslant \inf_{g_x \in G(x,\theta)} \| g_y^1 - g_x \| \leqslant \inf_{g_x \in G(x,\theta)} \| g_y^2 - g_x \| + \| g_y^1 - g_y^2 \|.$$

Итак,

$$\Delta(G(y), G(x)) \leqslant \int_{\Theta} \Delta(G(y,\theta), G(x,\theta))\, P(d\theta).$$

Так как $\Delta(G(y,\theta), G(x,\theta)) \leqslant 2L_K(\theta)$, $L_K(\theta)$ интегрируема и $\Delta(G(y,\theta), G(x,\theta)) \to 0$ при $y \to x$ $(\theta \in \Theta)$, то в силу теоремы Лебега о предельном переходе под знаком интеграла $\Delta(G(y), G(x)) \to 0$ при $y \to x$. Лемма доказана.

§ 22. Случайные липшицевы функции

В настоящем параграфе изучается субдифференциал Кларка случайных липшицевых функций. Показано, что он является измеримым по совокупности детерминированных и случайных переменных многозначным отображением. Этот результат используется для доказательства измеримости траекторий некоторых методов стохастической оптимизации. Получена формула, относящаяся к субдифференциальному исчислению, а именно что субдифференциал математического ожидания случайной липшицевой функции вложен в математическое ожидание субдифференциала этой функции (которое понимается как интеграл от многозначного отображения).

Заметим, что для выпуклых функций субдифференциал Кларка совпадает с субдифференциалом выпуклой функции (множеством субградиентов). Таким образом, имеем включение субдифференциала математического ожидания случайной выпуклой функции в математическое ожидание ее субдифференциала. Обратное включение (при доказанной измеримости субдифференциала) тривиально следует из субградиентного неравенства для выпуклых функций. Таким образом в выпуклом случае имеем известный результат — равенство субдифференциала математического ожидания и математического ожидания субдифференциала. Доказательство этого факта для функций в абстрактных пространствах, сведения исторического и библиографического характера имеются в [48].

Лемма 22.1. *Пусть* (Θ, Σ) — *измеримое пространство, функция* $f: E_n \times \Theta \to E_1$ *измерима по* $\theta \in \Theta$ *при каждом* $x \in E_n$ *и липшицева по* x *в каждом компакте* $K \subset E_n$ *для всех* $\theta \in \Theta$, $\partial_x f(x, \theta)$ — *субдифференциал Кларка функции* $f(\cdot, \theta)$ *в точке* x. *Тогда многозначное отображение* $\partial_x f$: $E_n \times \Theta \to 2^{E_n}$ $\mathscr{B} \times \Sigma$-*измеримо и, следовательно, имеет* $\mathscr{B} \times \Sigma$-*измеримые сечения. Отображение* $\partial_x f(x, \cdot)$: $\Theta \to 2^{E_n}$ *измеримо по* θ *при фиксированном* x.

Доказательство. Напомним, что

$$\partial f(x, \theta) = \{g \in E_n \mid (g, d) \leqslant f^0(x, \theta; d) \ \ \forall d \in E_n\},$$

$$f^0(x, \theta; d) = \lim_{\delta \to 0} \sup_{\substack{\|y\| \leqslant \delta \\ 0 < \lambda \leqslant \delta}} [f(x + y + \lambda d, \theta) - f(x + y, \theta)]/\lambda,$$

где $f^0(x, \theta; d)$ — производная Кларка функции $f(\cdot, \theta)$ по направлению $d \in E_n$. Покажем, что $f^0(\cdot, \cdot; d)$ $\mathscr{B} \times \Sigma$-измерима. Для этого достаточно показать, что $\mathscr{B} \times \Sigma$-измерима функция

$$f^0_\delta(x, \theta; d) = \sup_{\substack{\|y\| \leqslant \delta \\ 0 < \lambda \leqslant \delta}} [f(x + y + \lambda d, \theta) - f(x + y, \theta)]/\lambda. \tag{22.1}$$

Пусть $\{y_l\}$ и $\{\lambda_m\}$ — счетные всюду плотные множества в

$$\{y \in E_n \mid \|y\| \leqslant \delta\} \text{ и } \{\lambda \in E_1 \mid 0 < \lambda \leqslant \delta\}.$$

Тогда

$$f^0_\delta(x, \theta; d) = \sup_{\substack{\|y\| \leqslant \delta \\ 0 < \lambda \leqslant \delta}} [f(x + y + \lambda d, \theta) - f(x + y, \theta)]/\lambda =$$

$$= \lim_{\varepsilon \to +0} \sup_{\substack{\|y\| \leqslant \delta \\ \varepsilon \leqslant \lambda \leqslant \delta}} [f(x + y + \lambda d, \theta) - f(x + y, \theta)]/\lambda =$$

$$= \lim_{\varepsilon \to +0} \sup_{\substack{\|y_l\| \leqslant \delta \\ \varepsilon \leqslant \lambda_m \leqslant \delta}} [f(x + y_l + \lambda_m d, \theta) - f(x + y_l, \theta)]/\lambda_m =$$

$$= \sup_{l, m} [f(x + y_l + \lambda_m d, \theta) - f(x + y_l, \theta)]/\lambda_m =$$

$$= \lim_{N \to \infty} \max_{l, m \leqslant N} [f(x + y_l + \lambda_m d, \theta) - f(x + y_l, \theta)]/\lambda_m.$$

Заметим, что функции $f(x+y_l+\lambda_m d, \theta)$ и $f(x+y_l, \theta)$ каратеодориевы на $E_n \times \Theta$ и, следовательно, $\mathscr{B} \times \Sigma$-измеримы (лемма 21.6). Функция

$$\max_{l,m\leqslant N} [f(x+y_l+\lambda_m d, \theta) - f(x+y_l, \theta)]/\lambda_m,$$

очевидно, $\mathscr{B} \times \Sigma$-измерима. Тогда $f_\delta^0(\cdot, \cdot; d)$ измерима как предел измеримых функций. Следовательно, и $f^0(\cdot, \cdot; d)$ $\mathscr{B} \times \Sigma$-измерима. Заметим, что $f^0(x, \theta; \cdot)$ непрерывна. Представим

$$\partial_x f(x, \theta) = \{g \in E_n \mid \psi(g, x, \theta) \equiv \max_{\|d\|\leqslant 1} [(g, d) - f^0(x, \theta; d)] \leqslant 0\}.$$

Функция $\psi(g, \cdot, \cdot)$ $\mathscr{B}\times\Sigma$-измерима при каждом g (аналогично $f_\delta^0(\cdot, \cdot; d)$), выпукла и непрерывна по g; по лемме 21.4 $\psi(g, x, \theta)$ — нормальный выпуклый интегрант на $E_n \times (E_n \times \Theta)$. Теперь в силу леммы 21.5 многозначное отображение

$$(x, \theta) \to \partial_x f(x, \theta) = \{g \in E_n \mid \psi(g, x, \theta) \leqslant 0\}$$

$\mathscr{B} \times \Sigma$-измеримо и по теореме 21.2 имеет $\mathscr{B} \times \Sigma$-измеримые сечения. В силу замечания 21.2 из измеримости $\partial_x f(x, \theta)$ по двум переменным (x, θ) следует измеримость по одной переменной θ. Лемма доказана.

Следствие 22.1. *Если функция $f(x, \theta)$ выпукла по x при каждом θ и измерима по θ при каждом x, то ее субградиентное отображение (оно совпадает с субдифференциалом Кларка) $\partial_x f(x, \theta)$ $\mathscr{B} \times \Sigma$-измеримо, а отображение $\partial_x f(x, \cdot)$-измеримо.*

Теорема 22.1. *Пусть (Θ, Σ, P) — пространство с конечной положительной мерой (например, вероятностное пространство), функция $f: E_n \times \Theta \to E_1$ измерима по $\theta \in \Theta$ при каждом $x \in E_n$ и липшицева по x в каждом компакте $K \subset E_n$ с конечной константой Липшица $L_K(\theta)$ для всех $\theta \in \Theta$, $\partial_x f(x, \theta)$—субдифференциал Кларка функции $f(\cdot, \theta)$ в точке x. Тогда $\partial_x f(x, \cdot): \Theta \to 2^{E_n}$ — выпуклозначное нормальное отображение. Если, кроме того, $f(x, \theta)$ и $L_K(\theta)$ интегрируемы, то функция $F(x) = \int_\Theta f(x, \theta) P(d\theta)$ липшицева в K с константой $L_K = \int_\Theta L_K(\theta) P(d\theta)$ и справедливо включение*

$$\partial F(x) \subset \int_\Theta \partial_x f(x, \theta) P(d\theta). \tag{22.2}$$

Доказательство. Напомним, что множества $\partial_x f(x, \theta)$ выпуклы и компактны (§ 2). Измеримость отображения $\partial_x f(x, \cdot): \Theta \to 2^{E_n}$ доказана в лемме 22.1. Следовательно, отображение $\partial_x f(x, \cdot)$ нормально.

Липшицевость $F(x)$ очевидна: для $x, y \in K$ справедливо

$$|F(y) - F(x)| \leqslant \int_\Theta |f(y, \theta) - f(x, \theta)| P(d\theta) \leqslant \|y - x\| \int_\Theta L_K(\theta) P(d\theta).$$

Напомним, что

$$\partial F(x) = \{g \in E_n \mid (g, d) \leqslant F^0(x; d) \quad \forall d \in E_n\}, \tag{22.3}$$

$$F^0(x; d) = \lim_{\delta \to +0} \sup_{\substack{\|y\| \leqslant \delta \\ 0 < \lambda \leqslant \delta}} [f(x + y + \lambda d) - f(x + y)]/\lambda.$$

Докажем неравенство

$$F^0(x; d) \leqslant \int_{\Theta} f^0(x; \theta) P(d\theta). \tag{22.4}$$

Пусть $f^0_\delta(x, \theta; d)$ определена формулой (22.1). Для y, λ таких, что $\|y\| \leqslant \delta$ и $0 < \lambda \leqslant \delta$, имеет место

$$f^0_\delta(x, \theta; d) \geqslant [f(x + y + \lambda d, \theta) - f(x + y, \theta)]/\lambda.$$

Функции f^0_δ и f^0 $\mathscr{B} \times \Sigma$-измеримы и, следовательно, измеримы (см. замечание 21.2). Кроме того, $|f^0_\delta(x, \theta; d)| \leqslant L_K(\theta)$ и $|f^0(x, \theta; d)| \leqslant L_K(\theta)$; поэтому они интегрируемы. Тогда

$$\int_{\Theta} f^0_\delta(x, \theta; d) P(d\theta) \geqslant [f(x + y + \lambda d) - f(x + y)]/\lambda,$$

$$\int_{\Theta} f^0_\delta(x, \theta; d) P(d\theta) \geqslant \sup_{\substack{\|y\| \leqslant \delta \\ 0 < \lambda \leqslant \delta}} [f(x + y + \lambda d) - f(x + y)]/\lambda.$$

В последнем неравенстве перейдем к пределу по $\delta \to 0$. В силу теоремы Лебега предел можно внести под знак интеграла; в результате получаем неравенство (22.4). Обозначим $\mathsf{M}\partial_x f(x, \theta) = \int_{\Theta} \partial_x f(x, \theta) P(d\theta)$ и $\mathsf{M} f^0(x, \theta; d) = \int_{\Theta} f^0(x, \theta; d) P(d\theta)$.

Докажем равенство

$$\mathsf{M}\partial_x f(x, \theta) = \{g \in E_n \mid (g, d) \leqslant \mathsf{M} f^0(x, \theta; d) \quad \forall d \in E_n\}. \tag{22.5}$$

Сначала докажем, что левая часть (22.5) вложена в правую. Пусть $g \in \mathsf{M}\partial_x f(x, \theta)$. Согласно определению 21.10, существует интегрируемая функция $g(\theta)$ такая, что $g(\theta) \in \partial_x f(x, \theta)$ и $g = \int_{\Theta} g(\theta) P(d\theta)$. Тогда по определению $\partial_x f(x, \theta)$ справедливо $(g(\theta), d) \leqslant f^0(x, \theta; d)$ для любого $d \in E_n$. Интегрируя это неравенство, получаем

$$(g, d) \leqslant \mathsf{M} f^0(x, \theta; d) \quad \forall d \in E_n.$$

Следовательно, g принадлежит правой части (22.5), что и требовалось доказать.

Теперь предположим, что равенство (22.5) не имеет места, т. е. существует $g' \notin \mathsf{M}\partial_x f(x, \theta)$ и

$$g' \in \{g \in E_n \mid (g, d) \leqslant \mathsf{M} f^0(x, \theta; d) \quad \forall d \in E_n\}.$$

По теореме 21.3 $\mathsf{M}\partial_x f(x, \theta)$ — выпуклый компакт; поэтому существует $g'' \in \mathsf{M}\partial_x f(x, \theta)$ такое, что

$$\| g' - g'' \| = \rho(g', \mathsf{M}\partial_x f(x, \theta)).$$

Пусть $d = g' - g''$. В силу сделанного предположения для любого $g \in \mathsf{M}\partial_x f(x, \theta)$

$$(g, d) < (g', d) \leqslant \mathsf{M}f^0(x, \theta; d). \tag{22.6}$$

Покажем, что этого не может быть. Определим

$$G_d(x, \theta) = \{g \in E_n \mid g \in \partial_x f(x, \theta),\ (g, d) = f^0(x, \theta; d)\}.$$

Покажем, что $\theta \to G_d(x, \theta)$ — измеримое отображение. В самом деле,

$$G_d(x, \theta) = \{g \in E_n \mid \psi(g, x, \theta) = \sup_{\|l\| \leqslant 1} [(g, l) - f^0(x, \theta; l)] \leqslant 0,$$

$$\psi_d(g, x, \theta) \equiv (g, d) - f^0(x, \theta; d) = 0\} =$$

$$= \{g \in E_n \mid \varphi(g, x, \theta) \equiv \max(\psi(g, x, \theta),\ -\psi_d(g, x, \theta)) \leqslant 0\}.$$

Функции $\psi(g, x, \theta), -\psi_d(g, x, \theta)$, $\varphi(g, x, \theta)$ выпуклы и непрерывны по g и измеримы по θ, так как этим свойством обладают функции $(g, l) - f^0(x, \theta; l)$, а операции sup и max не нарушают этих свойств. Следовательно, $\varphi(\cdot, x, \cdot)$ — нормальный выпуклый интегрант на $E_n \times \Theta$; в силу леммы 21.5 отображение $G_d(x, \cdot)$ измеримо. По теореме 21.2 существует измеримое (и интегрируемое) сечение $g(\theta) \in G_d(x, \theta)$, т. е. $g(\theta) \in \partial_x f(x, \theta)$ и $(g(\theta), d) = f^0(x, \theta; d)$. Интегрируя эти выражения, получаем

$$g = \int_\Theta g(\theta) P(d\theta) \in \mathsf{M}\partial_x f(x, \theta),$$

$$(g, d) = \left(\int_\Theta g(\theta) P(d\theta),\ d\right) = \mathsf{M}f^0(x, \theta; d),$$

что противоречит (22.6). Итак, равенство (22.5) доказано. Теперь из (22.3) — (22.5) следует требуемое включение (22.2).

З а м е ч а н и е 22.1. В работе [180] доказано неравенство (22.4), а включение (22.2) установлено при следующих предположениях:

1) существуют производные по направлениям $d \in E_n$ функций $f(\cdot, \theta)$:

$$f'(x, \theta; d) = \lim_{\lambda \to +0} [f(x + \lambda d, \theta) - f(x, \theta)]/\lambda;$$

2) имеет место равенство $f'(x, \theta; d) = f^0(x, \theta; d)$.

Поскольку производная Кларка $f^0(x, \theta; d)$ выпукла по d, то фактически в [180] предполагается выпуклость по d производной по направлениям $f'(x, \theta; d)$, т. е. локальная выпуклость $f(\cdot, \theta)$ [49].

§ 23. Случайные обобщенно дифференцируемые функции и исчисление стохастических обобщенных градиентов

В этом параграфе изучаются случайные обобщенно дифференцируемые функции и их случайные псевдоградиентные отображения. Показано, что класс обобщенно дифференцируемых функций замкнут относительно операции взятия математического ожидания. При этом математическое ожидание псевдоградиентного отображения (как интеграл от многозначного отображения) случайной функции само является некоторым псевдоградиентным отображением математического ожидания этой функции. В этом отношении обобщенно дифференцируемые функции аналогичны выпуклым функциям.

Благодаря указанным свойствам, а также в силу устойчивости метода обобщенного градиента оказывается возможным распространить метод стохастических обобщенных градиентов (стохастических квазиградиентов [41]) на задачи стохастического программирования с обобщенно дифференцируемыми функциями. Это будет сделано в следующем параграфе. Существенную роль при этом играют стохастические обобщенные градиенты функций задачи, которые являются измеримыми по совокупности детерминированных и случайных переменных сечениями случайных псевдоградиентных отображений. В этом параграфе исследованы вопросы существования и построения таких сечений.

1. Случайные обобщенно дифференцируемые функции.

Теорема 23.1. *Пусть (Θ, Σ, P) — измеримое пространство с конечной положительной мерой (например, вероятностное пространство). Функция $f\colon E_n \times \Theta \to E_1$ обобщенно дифференцируема по $x \in E_n$ для $\theta \in \Theta$ и интегрируема по θ при каждом x, $G_f\colon E_n \times \Theta \to 2^{E_n}$ — ее измеримое по θ при каждом x псевдоградиентное (по x) при каждом θ отображение (например, $G_f(x, \theta) = \partial_x f(x, \theta)$). Пусть для каждого компакта $K \subset E_n$ существует интегрируемая функция $L_K(\theta)$ такая, что*

$$\sup\{\|g\| \mid g \in G_f(x, \theta),\ x \in K\} \leqslant L_K(\theta).$$

Тогда $F(x) = \int\limits_{\Theta} f(x, \theta)\, P(d\theta)$ обобщенно дифференцируема и $x \to G_F(x) = \int\limits_{\Theta} G_f(x, \theta)\, P(d\theta)$ — некоторое псевдоградиентное отображение для $F(x)$.

Доказательство. Необходимо показать,

во-первых, что множества $G_F(x)$ не пусты, выпуклы и компактны, — это следует из теорем 21.2, 21.3;

во-вторых, что отображение $G_F\colon E_n \to 2^{E_n}$ полунепрерывно сверху, — это следует из леммы 21.7;

в-третьих, что справедливо разложение

$$F(y) = F(x) + (g, y - x) + o(x, y, g),$$

где $o(x, y^k, g^k)/\| y^k - x \| \to 0$ для любых последовательностей $y^k \to x$ и $g^k \in G_F(y^k)$.

По определению 21.11 для каждого $g \in G_F(y)$ существует интегрируемое сечение $g_y(\theta)$ отображения $G_f(y, \cdot)$ такое, что $g = \mathsf{M} g_y(\theta) = \int_\Theta g_y(\theta) P(d\theta)$. В силу обобщенной дифференцируемости $f(\cdot, \theta)$ справедливо разложение

$$f(y, \theta) = f(x, \theta) + (g, y - x) + o(x, y, g; \theta), \tag{23.1}$$

где $o(x, y, g; \theta)/\| y - x \| \to 0$ при произвольных $y \to x$ и $g \in G_f(y, \theta)$. Подставим сечение $g_y(\theta)$ в (23.1):

$$f(y, \theta) = f(x, \theta) + (g_y(\theta), y - x) + o(x, y, g_y(\theta); \theta). \tag{23.2}$$

Очевидно, $o(x, y, g_y(\theta); \theta)$ интегрируема при интегрируемых $g_y(\theta)$. Проинтегрировав (23.2), получим

$$o(x, y, g) = F(y) - F(x) - (g, y - x) = \int_\Theta o(x, y, g_y(\theta); \theta) P(d\theta),$$

где $g = \mathsf{M} g_y(\theta)$.

Пусть теперь заданы произвольные последовательности $y^k \to x$ и $g^k \in G_F(y^k)$. Покажем, что $o(x, y^k, g^k)/\| y^k - x \| \to 0$. Действительно, по определению интеграла для каждого $g^k \in G_F(y^k)$ существует интегрируемое сечение $g^k(\theta)$ отображения $G_f(y^k, \cdot)$ такое, что $g^k = \int_\Theta g^k(\theta) \times P(d\theta)$. Пусть $\| y^k - x \| \leqslant \varepsilon$. Так как $| o(x, y^k, g^k(\theta); \theta) | / \| y^k - x \| \leqslant L_\varepsilon(\theta)$, где $L_\varepsilon(\theta)$ — некоторая интегрируемая функция, и $o(x, y^k, g^k(\theta); \theta)/\| y^k - x \| \to 0$ при $k \to \infty$ для всех θ, то в силу теоремы Лебега о предельном переходе под знаком интеграла

$$\lim_{k \to \infty} \frac{o(x, y^k, g^k)}{\| y^k - x \|} = \int_\Theta \lim_{k \to \infty} \frac{o(x, y^k, g^k(\theta); \theta)}{\| y^k - x \|} P(d\theta) = 0.$$

Теорема доказана.

2. Исчисление стохастических обобщенных градиентов. Рассмотрим задачу стохастического программирования

$$F(x) \equiv \int_\Theta f(x, \theta) P(d\theta) \to \min, \tag{23.3}$$

где $f(x, \theta)$ удовлетворяет условиям теоремы 23.1. Пусть $G_f(x, \theta)$ — требуемое в условиях теоремы 23.1 псевдоградиентное (по x) отображение для функции $f(\cdot, \theta)$. Например, им может быть, согласно теореме 1.10 и лемме 22.1, субдифференциал Кларка $\partial_x f(x, \theta)$; далее будут даны другие примеры подходящих $G_f(x, \theta)$. Пусть $g(x, \theta)$ есть $\mathscr{B} \times \Sigma$-измеримое сечение $G_f(x, \theta)$. В наших предположениях оно существует, так как в силу леммы 22.1, теоремы 21.2 и теоремы 1.10 существует $\mathscr{B} \times \Sigma$-измеримое сечение $g(x, \theta) \in \partial_x f(x, \theta) \subset G_f(x, \theta)$.

Простейший стохастический алгоритм для решения задачи (23.3) имеет вид

$$x^0 \in E_n, \quad x^{k+1} = x^k - \rho_k g(x^k, \theta^k), \quad k = 0, 1, \ldots, \tag{23.4}$$

где x^0 — начальная точка, x^k и x^{k+1} — последовательные приближения к решению задачи, ρ_k — шаговый множитель, θ^k — независимая реализация случайного параметра $\theta \in \Theta$, $g(x^k, \theta^k)$ — градиент подынтегральной функции $f(x, \theta^k)$ в точке x^k, являющийся стохастической оценкой градиента минимизируемой функции $F(x)$.

Пусть (Ω, Σ_ω) обозначает измеримое пространство, являющееся прямым счетным произведением измеримых пространств (Θ, Σ). Обозначим через ω последовательность $\{\theta^k\}$; очевидно, $\omega \in \Omega$. Определим отображение $\pi_k \colon \Omega \to \Theta$, являющееся проектором пространства $\Omega = \Theta \times \Theta \times \ldots$ на k-й сомножитель Θ, т. е. $\pi_k(\omega) = \theta^k$. Будем рассматривать точки траектории $\{x^k\}$ как функции от ω. Чтобы для анализа траекторий $\{x^k(\omega)\}$ $(\omega \in \Omega)$ можно было применять методы теории случайных процессов, необходимо прежде всего быть уверенным в измеримости $\{x^k(\omega)\}$ как функций на (Ω, Σ_ω).

Измеримость траектории $\{x^k(\omega)\}$ алгоритма (23.4) устанавливается рекуррентно. Функция $x^0(\omega) = x^0$ измерима как константа. Если $x^k(\omega)$ измерима, то $g^k(\omega) = g(x^k(\omega), \pi_k(\omega))$ измерима как суперпозиция измеримого отображения $z^k = (x^k, \pi_k)\colon \Omega \to E_n \times \Theta$ и $\mathscr{B} \times \Sigma$-измеримого отображения $g\colon E_n \times \Theta \to E_n$. Следовательно, вектор-функция $x^{k+1}(\omega)$ также измерима.

Таким образом, при построении методов стохастической оптимизации существенную роль играют измеримые по совокупности детерминированных и случайных переменных сечения обобщенно градиентных отображений функций, стоящих под знаком математического ожидания. Рассмотрим вопросы существования и построения таких сечений.

Пусть (Θ, Σ, P) — вероятностное пространство.

Определение 23.1. Функция $f\colon E_n \times \Theta \to E_1$ называется *случайной обобщенно дифференцируемой* (непрерывно дифференцируемой, выпуклой, липшицевой), если она измерима по $\theta \in \Theta$ при каждом $x \in E_n$ и обобщенно дифференцируема (непрерывно дифференцируема, выпукла, липшицева) по $x \in E_n$ для всех $\theta \in \Theta$.

Определение 23.2. *Стохастическим псевдоградиентом* (обобщенным градиентом) случайной обобщенно дифференцируемой функции $f(x, \theta)$ называется всякое $\mathscr{B} \times \Sigma$-измеримое сечение $g(x, \theta)$ ее любого псевдоградиентного отображения $G_f(x, \theta)$.

Лемма 23.1. *Всякая случайная обобщенно дифференцируемая функция имеет стохастические псевдоградиенты (обобщенные градиенты).*

Доказательство. По теореме 1.10 $\partial_x f(x, \theta) \subset G_f(x, \theta)$, а, согласно лемме 22.1, отображение $\partial_x f(x, \theta)$ $\mathscr{B} \times \Sigma$-измеримо; поэтому (теорема 21.2) оно имеет $\mathscr{B} \times \Sigma$-измеримые сечения.

Лемма 23.2. *Пусть отображение $G_f(x, \theta)$ $\mathscr{B} \times \Sigma$-измеримо; тогда его сечение*

$$g(x, \theta) = \arg\min \{\|g\| \mid g \in G_f(x, \theta)\}$$

также $\mathscr{B} \times \Sigma$-измеримо.

Утверждение леммы вытекает из результатов работы [175, следствие 4.3].

Лемма 23.3. *Градиент $g(x, \theta)$ случайной непрерывно дифференцируемой функции $f(x, \theta)$ $\mathscr{B} \times \Sigma$-измерим. Если $f(x, \theta)$ интегрируема по $\theta \in \Theta$ при каждом x и для любого компакта $K \subset E_n$ $\sup\{\|g(x, \theta)\| \mid y \in K\} \leqslant L_K(\theta)$, где $L_K(\theta)$ — интегрируемая функция, то функция $F(x) \equiv \int\limits_\Theta f(x, \theta) P(d\theta)$ непрерывно дифференцируема, $g(x) \equiv \equiv \int\limits_\Theta g(x, \theta) P(d\theta)$ — ее градиент.*

Доказательство. Градиент $g(x, \theta)$ непрерывен по x и измерим по θ; по лемме 21.6 он $\mathscr{B} \times \Sigma$-измерим. Пусть $K = \{y \mid \|y - x\| \leqslant \varepsilon\}$; по условию леммы $\sup\{\|g(x, \theta)\| \mid \|y - x\| \leqslant \varepsilon\} \leqslant L_K(\theta)$, причем функция $L_K(\theta)$ интегрируема. В силу теоремы о среднем для непрерывно дифференцируемой функции $f(\cdot, \theta)$ справедливы неравенства

$$\frac{|f(y, \theta) - f(x, \theta) - (g(x, \theta), y - x)|}{\|y - x\|} \leqslant 2L_K(\theta),$$

$$\|g(y, \theta) - g(x, \theta)\| \leqslant 2L_K(\theta).$$

Теперь в силу теоремы Лебега о предельном переходе под знаком интеграла имеем

$$\lim_{y \to x} \frac{F(y) - F(x) - (g(x), y - x)}{\|y - x\|} =$$

$$= \int\limits_\Theta \lim_{y \to x} \frac{f(y, \theta) - f(x, \theta) - (g(x, \theta), y - x)}{\|y - x\|} P(d\theta) = 0;$$

$$\lim_{y \to x} \|g(y) - g(x)\| \leqslant \int\limits_\Theta \lim_{y \to x} \|g(y, \theta) - g(x, \theta)\| P(d\theta) = 0,$$

что означает непрерывную дифференцируемость $F(x)$. Лемма доказана.

Лемма 23.4. *Пусть функция $f_\alpha(x, \theta)$ $(\alpha \in K \subset E_m,\ x \in E_n,\ \theta \in \Theta)$ измерима по θ и непрерывна по (α, x) вместе со своим градиентом по x $g_\alpha(x, \theta)$; K — компакт в E_m. Пусть*

$$f(x, \theta) = \max\{f_\alpha(x, \theta) \mid \alpha \in K\},$$

$$A(x, \theta) = \{\alpha \in K \mid f_\alpha(x, \theta) = f(x, \theta)\}.$$

Тогда функция $f(x, \theta)$ обобщенно дифференцируема по x, множество е· псевдоградиентов в точке x дается формулой

$$G_f(x, \theta) = \mathrm{co}\{g \in E_n \mid g = g_\alpha(x, \theta),\ \alpha \in A(x, \theta)\},$$

многозначные отображения $(x, \theta) \to G_f(x, \theta)$ *и* $(x, \theta) \to A(x, \theta)$ $\mathscr{B} \times \Sigma$-*измеримы, и если* $\alpha(x, \theta)$ *есть* $\mathscr{B} \times \Sigma$-*измеримое сечение* $A(x, \theta)$, *то* $g_{\alpha(x,\theta)}(x, \theta)$ *есть* $\mathscr{B} \times \Sigma$-*измеримое сечение* $G_f(x, \theta)$. *Если, кроме того,* $f(x, \theta)$ *интегрируема по* θ *и для любого компакта* $B \subset E_n$ *имеет место*

$$\sup \{\| g_\alpha(x, \theta)\| \mid \alpha \in K,\ x \in B\} \leqslant L_{K,B}(\theta),$$

где $L_{K,B}(\theta)$ *интегрируема, то функция* $F(x) = \int\limits_{\Theta} f(x, \theta) P(d\theta)$ *обобщенно дифференцируема,* $G_F(x) = \int\limits_{\Theta} G_f(x, \theta) P(d\theta)$ — *некоторое псевдоградиентное отображение для* $F(x)$.

Доказательство. Функция $f(x, \theta)$ слабовыпукла по x [91]; поэтому она обобщенно дифференцируема (теорема 1.3) и $G_f(x, \theta)$ — ее некоторое псевдоградиентное отображение.

Нетрудно показать, что для любого компакта $D \subset E_m$ функция $\max \{F_\alpha(x, \theta) \mid \alpha \in K \cap D\}$ непрерывна по x и измерима по θ, т. е. она каратеодориева и, следовательно (лемма 21.6), $\mathscr{B} \times \Sigma$-измерима. Тогда отображение $A(x, \theta)$ $\mathscr{B} \times \Sigma$-измеримо в силу $\mathscr{B} \times \Sigma$-измеримости множеств

$$\{(x, \theta) \mid A(x, \theta) \cap D \neq \varnothing\} = \{(x, \theta) \mid \max_{\alpha \in K \cap D} f_\alpha(x, \theta) = f(x, \theta)\}.$$

Отображение $g_\alpha(x, \theta)$ непрерывно по α, непрерывно по x и измеримо по θ (как предел измеримых конечных разностей); согласно лемме 21.6, оно $\mathscr{B}_{E_m} \times (\mathscr{B}_{E_n^+} \times \Sigma)$-измеримо. Если $\alpha(x, \theta)$ есть $\mathscr{B} \times \Sigma$-измеримое сечение $A(x, \theta)$, то сечение $g_{\alpha(x,\theta)}(x, \theta)$ отображения $G_f(x, \theta)$ $\mathscr{B} \times \Sigma$-измеримо как суперпозиция измеримых отображений.

Пусть $\{\alpha_l(x, \theta)\}_{l=1}^{\infty}$ — счетное плотное семейство измеримых сечений $A(x, \theta)$; тогда $\{g_{\alpha_l(x,\theta)}(x, \theta)\}_{l=1}^{\infty}$ — счетное плотное семейство измеримых сечений отображения

$$G_f'(x, \theta) = \{g_\alpha(x, \theta) \mid \alpha \in A(x, \theta)\}.$$

По теореме 21.2 $G_f'(x, \theta)$ $\mathscr{B} \times \Sigma$-измеримо, а по лемме 21.2 $G_f(x, \theta) = \operatorname{co} G_f'(x, \theta)$ также $\mathscr{B} \times \Sigma$-измеримо.

Обобщенная дифференцируемость $F(x)$ и псевдоградиентность $G_F(x)$ следуют из теоремы 23.1. Лемма доказана.

Замечание 23.1. Отображение $A(x, \theta)$ и сечение $\alpha(x, \theta) \in A(x, \theta)$ представляют собой точные решения задачи $\max \{f_\alpha(x, \theta) \mid \alpha \in K\}$. На практике эта задача решается некоторым алгоритмом (точно или приближенно), т. е. по (x, θ) строится некоторое решение $\tilde{\alpha}(x, \theta)$ и подставляется в $g_\alpha(x, \theta)$. Чтобы гарантировать $\mathscr{B} \times \Sigma$-измеримость функции $g_{\tilde{\alpha}(x,\theta)}(x, \theta)$, необходимо специально проверить, является ли результат $\tilde{\alpha}(x, \theta)$ работы алгоритма $\mathscr{B} \times \Sigma$-измеримым.

Пусть $f_0(y, \theta)$ $(y \in E_m)$ и $f_i(x, \theta)$ $(x \in E_n,\ i = 1, \ldots, m)$ — случайные обобщенно дифференцируемые функции; $G_0(y, \theta)$ и $G_i(x, \theta)$ $(i = 1, 2, \ldots, m)$ — их псевдоградиентные отображения; $L_0(\theta, K_0)$ и $L_i(\theta, K)$ — соответствующие константы Липшица на множествах $K_0 \subset E_m$, $K \subset E_n$, причем

$$\sup \{\|g\| \mid g \in G_0(y, \theta),\ y \in K_0\} \leqslant L_0(\theta, K_0),$$

$$\sup \{\|g\| \mid g \in G_i(x, \theta),\ x \in K\} \leqslant L_i(x, K).$$

Введем вектор-функцию $z(x, \theta) = (f_1(x, \theta), \ldots, f_m(x, \theta))$ и $n \times m$-матрицу $[g^1 \ldots g^m]$. Рассмотрим функцию

$$f(x, \theta) = f_0(z(x, \theta), \theta) = f_0(f_1(x, \theta), \ldots, f_m(x, \theta), \theta). \tag{23.5}$$

Ее свойства изложены далее в леммах 23.5 — 23.8.

Лемма 23.5. *Функция $f(x, \theta)$, определенная формулой* (23.5), *является случайной обобщенно дифференцируемой, а отображение $x \to G_f(x, \theta)$, где*

$$G_f(x, \theta) = \mathrm{co}\{g \in E_n \mid g = [g^1 \ldots g^m]\, g^0,$$

$$g^0 \in G_0(z(x, \theta), \theta),\ g^i \in G_i(x, \theta),\ i = 1, \ldots, m\}, \tag{23.6}$$

является псевдоградиентным по x для $f(x, \theta)$.

Действительно, функция $f_0(y, \theta)$ каратеодориева и, следовательно, $\mathscr{B}_{E_m} \times \Sigma$-измерима, а вектор-функция $z(x, \cdot)$ по условию измерима (при фиксированном x). Тогда $f(x, \cdot)$ измерима по θ при фиксированном x как суперпозиция указанных измеримых функций. Обобщенная дифференцируемость $f(\cdot, \theta)$ по x и псевдоградиентность $G_f(x, \theta)$ при каждом θ доказаны в теореме 1.6.

Лемма 23.6. *Пусть для каждого компакта $K_0 \subset E_m$ и каждого компакта $K \subset E_n$ функции $f_0(y, \theta)$ и $f_i(x, \theta)$ $(i = 1, 2, \ldots, m)$ ограничены на $K_0 \times \Theta$ и $K \times \Theta$ соответственно, а соответствующие константы Липшица $L_0(\theta, K_0)$ и $L_i(\theta, K)$ ограничены на Θ. Тогда сложная функция $f(x, \theta)$ удовлетворяет условию Липшица на K и интегрируема вместе со своей константой Липшица, а функция $F(x) = \int\limits_{\Theta} f(x, \theta) P(d\theta)$ обобщенно дифференцируема и $G_F(x) = \int\limits_{\Theta} G_f(x, \theta) P(d\theta)$ — некоторое ее псевдоградиентное отображение.*

Действительно, липшицевость сложной функции $f(x, \theta)$ легко проверяется, интегрируемость $f(x, \theta)$ и $G_f(x, \theta)$ следует из их измеримости (леммы 23.5, 23.7) и ограниченности, а обобщенная дифференцируемость $F(x)$ и псевдоградиентность $G_F(x)$ следуют из теоремы 23.1.

Лемма 23.7. *Пусть $g^0(y, \theta)$ является $\mathscr{B}_{E_m} \times \Sigma$-измеримым сечением $G_0(y, \theta)$, а $g^i(x, \theta)$ есть $\mathscr{B}_{E_n} \times \Sigma$-измеримое сечение $G_i(x, \theta)$ $(i = 1, 2, \ldots, m)$. Тогда функция*

$$g(x, \theta) = [g^1(x, \theta) \ldots g^m(x, \theta)]\, g^0(z(x, \theta), \theta)$$

есть $\mathscr{B} \times \Sigma$*-измеримое сечение* $G_f(x, \theta)$ *(см.* (23.6)). *Если* $G_0(y, \theta)$ *и* $G_i(x, \theta)$ *измеримы по совокупности переменных, то и* $G_f(x, \theta)$ $\mathscr{B} \times \Sigma$*-измеримо.*

Доказательство. Заметим, что измеримость вектор-функции эквивалентна ее покоординатной измеримости; поэтому для доказательства измеримости $g(x, \theta)$ достаточно показать измеримость $g^0(z(x, \theta), \theta)$. Вектор-функция $z(x, \theta)$ каратеодориева, и в силу леммы 21.6 она $\mathscr{B} \times \Sigma$-измерима; тогда $g^0(z(x, \theta), \theta)$ $\mathscr{B} \times \Sigma$-измерима как суперпозиция измеримых функций.

Докажем $\mathscr{B} \times \Sigma$-измеримость отображения $(x, \theta) \to G_f(x, \theta)$. Обозначим

$$G'_f(x, \theta) = \{g \in E_n \mid g = [g^1 g^2 \dots g^m]\, g^0,$$

$$g^0 \in G_0(z(x, \theta), \theta),\ g^i \in G_i(x, \theta),\ i = 1, \dots, m\},$$

т. е. $\mathrm{co}\, G'_f(x, \theta) = G_f(x, \theta)$. В силу леммы 21.3 достаточно показать измеримость $(x, \theta) \to G'_f(x, \theta)$. Покажем существование счетного плотного семейства измеримых по совокупности переменных сечений отображения G'; по теореме 21.2 отсюда будет следовать измеримость G'. По теореме 21.2 существуют счетные плотные семейства $\mathscr{B} \times \Sigma$-измеримых сечений $\{g^0_{\mu_0},\ \mu_0 = 1, 2 \dots\}$ и $\{g^i_{\mu_i},\ \mu_i = 1, 2, \dots\}$ отображений G_0 и G_i соответственно. Очевидно, что функция

$$g_\mu(x, \theta) = [g^1_{\mu_1}(x, \theta) \dots g^m_{\mu_m}(x, \theta)]\, g^0_{\mu_0}(z(x, \theta), \theta),$$

где $\mu = (\mu_0, \mu_1, \dots, \mu_m)$, является $\mathscr{B} \times \Sigma$-измеримым сечением G'. Семейство $\{g_\mu\}$ счетно. Покажем, что множество значений $\{g_\mu(x, \theta)\}$ плотно в $G'_f(x, \theta)$ для всех (x, θ). Возьмем произвольное $g' \in G'_f(x, \theta)$. По построению существуют $g'_0 \in G_0(z(x, \theta), \theta)$ и $g'_i \in G_i(x, \theta)$ такие, что $g' = [g'_1 \dots g'_m]\, g'_0$. В силу плотности семейств сечений $g^i_{\mu_i}$ существуют $g^0_{\mu_0^k}(z(x, \theta), \theta) \to g'_0$ и $g^i_{\mu_i^k}(x, \theta) \to g'_i$. Тогда

$$g_{\mu^k}(x, \theta) = [g^1_{\mu_1^k}(x, \theta) \dots g^m_{\mu_m^k}(x, \theta)]\, g^0_{\mu_0^k}(z(x, \theta), \theta) \to g'.$$

Лемма 23.7 доказана.

Лемма 23.8. *Рассмотрим обобщенно дифференцируемую (см. теорему* 1.5) *функцию максимума*

$$f(x, \theta) = \max_{1 \leqslant i \leqslant m} f_i(x, \theta).$$

Как показано в § 1, *ее псевдоградиентное по* x *отображение задается формулами*

$$G_f(x, \theta) = \mathrm{co}\{g \in E_n \mid g_i \in G_i(x, \theta),\ i \in I(x, \theta)\}, \tag{23.7}$$

$$I(x, \theta) = \{i \mid 1 \leqslant i \leqslant m,\ f_i(x, \theta) = f(x, \theta)\}. \tag{23.8}$$

Если $g_i(x, \theta)$ $(i = 1, \dots, m)$ *являются* $\mathscr{B} \times \Sigma$*-измеримыми сечениями* $G_i(x, \theta)$, *а* $i(x, \theta)$ *есть* $\mathscr{B} \times \Sigma$*-измеримое сечение отображения*

$(x, \theta) \to I(x, \theta)$, *то*

$$g(x, \theta) = g_{i(x,\theta)}(x, \theta)$$

есть $\mathcal{B} \times \Sigma$*-измеримое сечение отображения* $G_f(x, \theta)$.

Если отображения G_i $(i = 1, \ldots m)$ $\mathcal{B} \times \Sigma$*-измеримы, то и* G_f $\mathcal{B} \times \Sigma$*-измеримо.*

Замечание 23.2. Для построения измеримого сечения отображения G_f в лемме 23.8 необходимы $B \times \Sigma$-измеримые сечения отображения $(x, \theta) \to I(x, \theta)$.

Покажем, что отображение $(x, \theta) \to I(x, \theta)$ $\mathcal{B} \times \Sigma$-измеримо. Конечное множество $I(x, \theta)$ замкнуто. Для любого замкнутого $B \subset E_1$ множество

$$\{(x, \theta) \mid I(x, \theta) \cap B \neq \varnothing\} = \{(x, \theta) | \max_{\substack{i \in B \\ 1 \leqslant i \leqslant m}} f_i(x, \theta) = f(x, \theta)\}$$

$\mathcal{B} \times \Sigma$-измеримо, так как $f(x, \theta)$ и $f_i(x, \theta)$ каратеодориевы. Следовательно, отображение $(x, \theta) \to I(x, \theta)$ $\mathcal{B} \times \Sigma$-измеримо, и по теореме 21.2 существуют $\mathcal{B} \times \Sigma$-измеримые сечения отображения $(x, \theta) \to I(x, \theta)$. Рассмотрим, например,

$$i(x, \theta) = \min\{i \mid i \in I(x, \theta)\}.$$

Функция $i(x, \theta)$ $\mathcal{B} \times \Sigma$-измерима, так как $\mathcal{B} \times \Sigma$-измеримы лебеговские множества $(c \in E_1)$

$$\{(x, \theta) | i(x, \theta) \leqslant c\} = \{(x, \theta) | I(x, \theta) \cap \{r \mid r \leqslant c\} \neq \varnothing\}.$$

Замечание 23.3. Если в условиях леммы 23.8 функции $f_i(x, \theta)$ $(i = 1, 2, \ldots, m)$ удовлетворяют условию Липшица по x в компакте $K \subset E_n$ с константами $L_i(\theta)$, то функция максимума $f(x, \theta)$ удовлетворяет в K условию Липшица с константой $L(\theta) = \max\limits_{1 \leqslant i \leqslant m} L_i(\theta)$. Если, кроме того, $f_i(x, \theta)$ интегрируемы вместе со своими константами Липшица, то и функция максимума интегрируема вместе со своей константой Липшица, и в силу теоремы 23.1 функция $F(x) = \int\limits_{\Theta} f(x, \theta) P(d\theta)$ обобщенно дифференцируема.

Доказательство леммы 23.8. Сечение $g(x, \theta)$ $\mathcal{B} \times \Sigma$-измеримо, так как $\mathcal{B} \times \Sigma$-измеримы лебеговские множества $(c \in E_n)$

$$\{(x, \theta) | g(x, \theta) \leqslant c\} = \bigcup_{k=1}^{m} (\{(x, \theta) | i(x, \theta) = k\} \cap \{(x, \theta) | g_k(x, \theta) \leqslant c\}).$$

Докажем $\mathcal{B} \times \Sigma$-измеримость G_f, если G_i $(i = 1, \ldots, m)$ $\mathcal{B} \times \Sigma$-измеримы. В силу леммы 21.3 достаточно показать измеримость отображения

$$(x, \theta) \to G_f'(x, \theta) = \{g \in E_n \mid g \in G_i(x, \theta),\ i \in I(x, \theta)\}.$$

Пусть $\{g_i^{\nu_i}(x, \theta)\}_{\nu_i=1}^{\infty}$ $(i = 1, \ldots, m)$ — счетные плотные семейства $\mathcal{B} \times \Sigma$-измеримых сечений G_i. Достаточно рассмотреть случай $m = 2$.

Определим функции:

$$h_1^{\nu_1\nu_2}(x, \theta) = \begin{cases} g_1^{\nu_1}(x, \theta), & f_1(x, \theta) \geqslant f_2(x, \theta), \\ g_2^{\nu_2}(x, \theta), & f_1(x, \theta) < f_2(x, \theta); \end{cases}$$

$$h_2^{\nu_1\nu_2}(x, \theta) = \begin{cases} g_1^{\nu_1}(x, \theta), & f_1(x, \theta) > f_2(x, \theta), \\ g_2^{\nu_2}(x, \theta), & f_1(x, \theta) \leqslant f_2(x, \theta). \end{cases}$$

Напомним, что $f_1(x, \theta)$ и $f_2(x, \theta)$ каратеодориевы; поэтому $h_1^{\nu_1\nu_2}(x, \theta)$ и $h_2^{\nu_1\nu_2}(x, \theta)$ $\mathcal{B} \times \Sigma$-измеримы. Очевидно, набор $\{h_1^{\nu_1\nu_2}(x, \theta), h_2^{\nu_1\nu_2}(x, \theta)\}_{\nu_1,\nu_2=1}^{\infty}$ является счетным плотным семейством измеримых сечений отображения G'. Лемма доказана.

ГЛАВА 7

РЕШЕНИЕ СТОХАСТИЧЕСКИХ ЭКСТРЕМАЛЬНЫХ ЗАДАЧ

В настоящей главе основное внимание уделяется численным методам решения стохастических задач, в которых не существует непрерывных производных функций цели и ограничений, а значения функций не вычисляются точно и имеют вероятностную природу.

Рассматривается следующая экстремальная задача:

$$F_0(x) \equiv \mathsf{M} f_0(x, \theta) \to \min_x$$

при ограничениях

$$F_i(x) = \mathsf{M} f_i(x, \theta) \leqslant 0, \qquad i = 1, 2, \ldots, m, \tag{C}$$

$$x \in X \subset E_n,$$

где θ — элементарное событие некоторого вероятностного пространства (Θ, Σ, P), M — знак математического ожидания.

Основная трудность решения задачи (C) состоит в том, что при фиксированном x вычисление $F_i(x)$ $(i = 0, 1, \ldots, m)$ требует значительных усилий, так как это связано с вычислением многомерных интегралов. Вместо таких значений $F_i(x)$ $(i = 0, 1, \ldots, m)$ обычно имеется возможность наблюдать при данном x значения случайных величин $f_i(x, \theta)$. Поэтому в задачах (C) нельзя широко применять методы нелинейного программирования, и для их решения развиваются свои стохастические методы.

Изучаемые в этой главе численные методы позволяют решать задачу (C) в условиях, когда законы распределения θ неизвестны, но существует способ вычисления случайных величин $f_i(x, \theta)$ $(i = 0, 1, \ldots$ $\ldots, m)$. Эти же методы применимы и тогда, когда вероятностные свойства θ заданы, но вычисления функций $F_i(x)$ $(i = 0, 1, \ldots, m)$ либо невозможны (неизвестна зависимость функций $f_i(x, \theta)$ от θ), либо слишком сложны.

Различают прямые и непрямые методы стохастического программирования. Если известны вероятностные свойства θ и задана аналитическая зависимость $f_i(x, \theta)$ от θ, то математические ожидания в (C), по крайней мере в принципе, могут быть вычислены. Тогда задача (C) формально перестает быть стохастической. В непрямых методах стремятся получить зависимости $F_i(x)$ $(i = 0, 1, \ldots, m)$, а затем применять методы нелинейного программирования. Иными словами, вместо стохастической задачи рассматривается ее детерминированный вариант. Поэтому успех применения непрямых методов в значительной степени зависит от вероятностных свойств θ и свойств функций $f_i(x, \theta)$.

Методы, основанные на информации о значениях $f_i(x, \theta)$ или аналогов их градиентов, называются прямыми методами стохастического программирования. При решении практических задач функциональные зависимости $f_i(x, \theta)$, как правило, нелинейны, а векторы x и θ имеют большую размерность. Поэтому найти аналитический вид функций $F_i(x, \theta)$ чаще всего не представляется возможным, поскольку это связано с определением математических ожиданий, являющихся многомерными интегралами Лебега.

Если вероятностное распределение θ задано, то вычисление $f_i(x, \theta)$ сводится к моделированию случайных величин с заданным законом распределения. Этот процесс обычно осуществляется на ЭВМ с помощью преобразований одного или нескольких независимых значений случайной величины, равномерно распределенной в интервале $[0, 1]$.

Существенная особенность прямых методов состоит в том, что они применимы для решения задач, в которых законы распределения θ неизвестны. В процессе оптимизации сложных объектов могут возникнуть ситуации, когда формулировка вероятностных свойств θ представляет значительную трудность и необходимо решить задачу без аналитического исследования вероятностной модели. Для применения прямых методов требуются только наблюдения $\theta_1, \theta_2, \ldots, \theta_k, \ldots$ параметра θ. Поэтому эти методы могут быть широко использованы для оптимизации систем на основе имитационного моделирования. Прямые методы составляют основной предмет исследования данной главы.

Сделаем еще некоторые общие замечания о технологии решения задач стохастического программирования. Во-первых, стохастичность задачи может ярко проявляться не во всей допустимой области, а только, скажем, в некоторой окрестности решения, т. е. вдали от решения дисперсия градиентов целевой функции может быть мала по сравнению с самими градиентами и задача почти детерминирована. Поэтому вначале следует применять методы нелинейного программирования, устойчивые по отношению к ошибкам используемых градиентов. Если известен аналитический вид случайных функций задачи, то, пронаблюдав одну или несколько реализаций случайных параметров задачи, можно вначале интеграл — математическое ожидание — заменить интегральной суммой и для минимизации этой суммы применить методы нелинейного программирования (негладкой оптимизации). По ходу действий нужно периодически следить за дисперсией градиентов целевой функции. Если она становится сравнимой с самими градиентами, то следует использовать стохастические аналоги устойчивых методов нелинейного программирования. Если дисперсия станет больше самих градиентов, то следует либо путем дополнительных наблюдений уменьшить эту дисперсию, либо переходить на простой метод стохастических градиентов, так как сложные методы в такой ситуации становятся не лучше простого.

§ 24. Методы усредненных стохастических градиентов

В [41] для решения задачи выпуклого стохастического программирования предложен метод типа стохастической аппроксимации — метод стохастических квазиградиентов, использующий градиенты случайной функции, стоящей под знаком математического ожидания, и операцию проектирования на выпуклое множество. В [73] этот метод усовершенствован и распространен на бесконечномерный случай. В [91] метод стохастических квазиградиентов распространен на задачи стохастического программирования со слабо выпуклыми функциями (см. § 1). В настоящем параграфе рассматривается метод стохастических обобщенных градиентов для решения задачи стохастического программирования с обобщенно дифференцируемыми функциями при ограничениях. Рассматриваются также методы с усреднением стохастических градиентов, в частности стохастические аналоги методов тяжелого шарика и овражного шага (см. § 14).

Рассмотрим задачу невыпуклого стохастического программирования:

$$F(x) \equiv \mathsf{M}\, f(x, \theta) \to \min_x \tag{24.1}$$

при ограничениях

$$h(x) \leqslant 0, \qquad x \in E_n, \tag{24.2}$$

где $\theta \in \Theta$; (Θ, Σ, P) — вероятностное пространство; $f(x, \theta)$: $E_n \times \Theta \to E_1$ — случайная (измеримая по θ при каждом x) обобщенно дифференцируемая (по x для всех θ) функция, интегрируемая вместе со своей константой Липшица $L_K(\theta)$ для каждого компакта $K \subset E_n$; функция h: $E_n \to E_1$ обобщенно дифференцируема, причем $h(x) \to +\infty$ при $\|x\| \to +\infty$.

По теореме 23.1 функция F обобщенно дифференцируема. Предположим, что известны некоторое измеримое (по θ при каждом x) псевдоградиентное (по x для всех θ) отображение $G_f(x, \theta)$ функции $f(x, \theta)$, а также $\mathcal{B} \times \Sigma$-измеримое сечение $g_f(x, \theta)$ отображения $(x, \theta) \to G_f(x, \theta)$. Как показано в § 23, требуемые G_f и g_f существуют. По теореме 23.1 $G_F(x) = \int_\Theta G_f(x, \theta)\, P(d\theta)$ есть некоторое псевдоградиентное отображение для $F(x)$ и $g_f(x) = \mathsf{M}\, g_f(x, \theta) \in G_F(x)$. Предположим, что также известны псевдоградиентное отображение $G_h(x)$ и его борелевское сечение $g_h(x) \in G_h(x)$. Требуемое сечение в силу замкнутости $G_h: E_n \to 2^{E_n}$ существует (лемма 21.2).

Обозначим

$$G(x) = \begin{cases} G_F(x), & h(x) < 0, \\ \mathrm{co}\,\{G_F(x),\ G_h(x)\}, & h(x) = 0, \\ G_h(x), & h(x) > 0; \end{cases} \tag{24.3}$$

$$g(x, \theta) = \begin{cases} g_f(x, \theta), & h(x) < 0, \\ g_h(x, \theta), & h(x) \geqslant 0. \end{cases}$$

Функция $g(x, \theta)$ $\mathscr{B} \times \Sigma$-измерима, так как $\mathscr{B} \times \Sigma$-измеримы множества $(c \in E_n)$

$$\{(x, \theta) | g(x, \theta) \leqslant c\} = (\{(x, \theta) | g_f(x, \theta) \leqslant c\} \cap$$
$$\cap \{(x, \theta) | h(x) < 0\}) \cup (\{(x, \theta) | g_h(x, \theta) \leqslant c\} \cap \{(x, \theta) | h(x) \geqslant 0\}).$$

Обозначим множество псевдостационарных точек задачи (24.1), (24.2) $X^* = \{x \in E_n \mid 0 \in G(x)\}$.

Введем обозначения:

$\Theta^i = \Theta \times \ldots \times \Theta$ — i-я декартова степень Θ;

$\theta^i = (\theta_0, \ldots, \theta_{i-1})$ — элемент пространства Θ^i;

$\Sigma^i = \Sigma \times \ldots \times \Sigma = \sigma\{A_0 \times \ldots \times A_{i-1} \mid A_k \in \Sigma,\ 0 \leqslant k < i\}$ — σ-алгебра, являющаяся произведением i σ-алгебр Σ;

$P^i = P \times \ldots \times P$ — вероятностная мера на (Θ^i, Σ^i), являющаяся произведением i мер P; $(\Theta^i, \Sigma^i, P^i)$ — соответствующее вероятностное пространство;

$\Omega = \Theta^\infty = \Theta \times \Theta \times \ldots$ — счетное прямое произведение пространств Θ;

$\Sigma_\omega = \Sigma^\infty = \Sigma \times \Sigma \times \ldots = \sigma\{A_0 \times A_1 \times \ldots \mid A_k \in \Sigma,\ k \geqslant 0\}$ — счетное произведение σ-алгебр Σ.

Меры P^i на (Θ^i, Σ^i), согласно теореме Ионеску Тулча (см. [133]), можно единственным образом продолжить до некоторой вероятностной меры P_ω на (Ω, Σ_ω) так, что выполняется соотношение

$$P^i\{A \mid A \in \Sigma^i\} = P_\omega\{A \times \Theta \times \Theta \times \ldots \mid A \in \Sigma^i\}.$$

Тогда $(\Omega, \Sigma_\omega, P_\omega)$ — вероятностное пространство.

1. **Метод усредненных стохастических градиентов**. Метод предназначен для решения задачи (24.1), (24.2); он порождает случайную последовательность $\{x^k(\omega)\}$ приближений к решению задачи согласно следующим соотношениям:

$$x^0(\omega) = x^0 \in E_n, \tag{24.4}$$

$$x^{k+1}(\omega) = x^k(\omega) - \rho_k P^k(\omega), \tag{24.5}$$

$$P^k(\omega) = \sum^{k} \lambda_{kr} g^r(\omega) | H_r(\omega), \tag{24.6}$$

$$g^r(\omega) = g(x^r(\omega), \pi_r(\omega)),\ \pi_r(\omega) = \theta_r, \tag{24.7}$$

$$H_r(\omega) = H_r(x^0(\omega), \ldots, x^r(\omega), g^0(\omega), \ldots, g^{r-1}(\omega)) \geqslant \nu > 0, \tag{24.8}$$

где $\omega = (\theta_0, \theta_1, \ldots) \in \Omega$; $\theta_r\ (r = 0, 1, \ldots)$ — независимые реализации случайного параметра $\theta \in \Theta$; отображение $g(x, \theta)$ определено в (24.3) и ограничено по x в каждом компакте $K \subset E_n$ равномерно по θ; $H_r(x^0, \ldots, x^r, g^0, \ldots, g^{r-1})$ — борелевская функция от своих аргументов, ограниченная в каждом компакте из $(E_n)^{2r+1}$; ν — некоторая положительная константа; ρ_k, r_k, λ_{kr} — детерминированные функции от переменных k и r, удовлетворяющие следующим условиям:

$$\lambda_{kr} \geqslant 0, \quad \sum_{r=r_k}^{k} \lambda_{kr} = 1 \tag{24.9}$$

(для удобства считаем, что $\lambda_{kr} = 0$ при $r \in [0, r_k)$);

$$0 \leqslant \rho_k \leqslant \rho, \quad \lim_{k\to\infty} \sum_{r=r_k}^{k} \rho_k = 0, \quad \sum_{k=0}^{\infty} \rho_k = +\infty; \tag{24.10}$$

$$\sum_{r=0}^{\infty}\left(\sum_{k=r}^{\infty} \lambda_{kr}\rho_r\right)^2 = \sum_{k,s=0}^{\infty} \rho_k\rho_s \sum_{r=\max(r_k, r_s)}^{\min(k,s)} \lambda_{hr}\lambda_{hs} < +\infty. \tag{24.11}$$

Замечание 24.1. Предполагается, что соотношения (24.4) — (24.8) выполнены для всех $\omega \in \Omega$. Случайный процесс $\{x^k(\omega)\}$ рассматривается в вероятностном пространстве $(\Omega, \Sigma_\omega, P_\omega)$. Векторы $g^r(\omega)$ называются стохастическими (в Ω) обобщенными градиентами функций задачи в точках $x^r(\omega)$, величины $H_r(\omega)$ являются нормирующими коэффициентами, а направления $P^k(\omega)$ можно назвать усредненными стохастическими градиентами.

Замечание 24.2. Измеримость траектории $\{x^k(\omega)\}_{k=0}^{\infty}$ устанавливается рекуррентно. Функция $x^0(\omega) = x^0$, очевидно, измерима. Если $x^0, \ldots, x^k(\omega), g^0(\omega), \ldots, g^{k-1}(\omega)$ измеримы, то $g^k(\omega)$ измерима как суперпозиция измеримого отображения $z^k(\omega) = (x^k(\omega), \pi_k(\omega))$: $\Omega \to E_n \times \Theta$ и $\mathcal{B} \times \Sigma$-измеримого отображения g: $E_n \times \Theta \to E_n$ (в этом месте используются $\mathcal{B} \times \Sigma$-измеримость сечений $g_f(x, \theta)$ и борелевость $g_h(x)$). Величина $H_r(\omega)$ измерима как суперпозиция измеримого отображения $y^r(\omega) = (x^0(\omega), \ldots, x^r(\omega), g^0(\omega), \ldots, g^{r-1}(\omega)) : \Omega \to (E_n)^{2r+1}$ и борелевской функции H_r: $(E_n)^{2r+1} \to E_1$. Отсюда следует, что и $x^{k+1}(\omega)$ измерима.

Замечание 24.3. Вне допустимой области $\{x \mid h(x) \leqslant 0\}$ задачи (24.1), (24.2) метод усредненных стохастических градиентов (24.4) — (24.11) фактически работает как детерминированный метод усредненных градиентов (14.22) — (14.33). Поэтому его ограниченность (для любого ω) можно обеспечить так же, как в чисто детерминированном случае, т. е. либо за счет выбора достаточно малых шагов $\rho_k \leqslant \rho$, либо используя механизм возврата в начальную точку (см. § 14).

Замечание 24.4. Если существует оценка $\|g_f(x, \theta)\| \leqslant L_f(x)$ для всех θ, где $L_f(x)$ — непрерывная функция, то условие ограниченности $g_f(x, \theta)$ по x в каждом компакте из E_n равномерно по θ заведомо выполнено.

Замечание 24.5. Пусть известна оценка $\|\mathsf{M} g_f(x, \theta)\| \leqslant H(x)$, где $H(x)$ — непрерывная функция; тогда нормирующие коэффициенты можно взять в следующем виде ($\nu > 0$):

$$H_r(\omega) = \begin{cases} H(x^r(\omega)) + \nu, & h(x^r(\omega)) < 0, \\ \|g_h(x^r(\omega))\| + \nu, & h(x^r(\omega)) \geqslant 0. \end{cases}$$

В другом варианте можно полагать

$$H_0(\omega) = \text{const}, \qquad H_r(\omega) = \|P^{r-1}(\omega)\| + \nu, \qquad \nu > 0, \qquad r = 1, 2, \ldots$$

Замечание 24.6. Условие (24.11) выполнено, если

$$\sum_{k=0}^{\infty}\left(\rho_k \sum_{r=r_k}^{k} \rho_r\right) < +\infty,$$

что имеет место, например, при $k - r_k \leqslant \text{const}\,(k)$ или при $\rho_k = c/k^{\alpha}$, $k - (c_1 + c_2 k^{\beta}) \leqslant r_k \leqslant k$, $0{,}5 < \alpha \leqslant 1$, $0 < \beta < 2\alpha - 1$; c, c_1, $c_2 > 0$. При указанных α, β выполнено также (24.10).

Замечание 24.7. Можно, следуя [73, 143], для ускорения сходимости метода усреднять минимизирующую последовательность $\{x^k(\omega)\}_{k=0}^{\infty}$, например, следующим образом:

$$\bar{x}^0(\omega) = x^0(\omega),$$

$$\bar{x}^k(\omega) = (1 - \tau_k)\,\bar{x}^{k-1}(\omega) + \tau_k x^k(\omega), \qquad k \geqslant 1, \tag{24.12}$$

$$\tau_0 = 1, \qquad 0 \leqslant \tau_k \leqslant 1, \qquad \sum_{k=0}^{\infty} \tau_k = +\infty. \tag{24.13}$$

Если $\{\rho_k\}$ удовлетворяют (24.10), то в силу леммы Абеля—Дини *) условиям (24.13) удовлетворяет последовательность

$$\tau_k = \rho_k \Big/ \sum_{r=0}^{k} \rho_r, \qquad k \geqslant 0. \tag{24.14}$$

Лемма 24.1. Пусть $X^* = \bigcup_{j \in J} X_j^*$, где X_j^* $(j \in J)$ — связные множества. Тогда $\{\bar{x}^k(\omega)\}$ сходится при условиях (24.13) к множеству $\bar{X}^* = \bigcup_{j \in J} \operatorname{co} X_j^*$.

Доказательство. Обозначим $\Phi_r^k = \prod_{t=r}^{k} (1 - \tau_t)$ и для удобства положим $\Phi_{k+1}^k = 1$. Заметим, что $\Phi_r^k \leqslant \exp\left(-\sum_{t=r}^{k} \tau_t\right)$ и при условии (24.13) $\lim_{k\to\infty} \Phi_r^k = 0$ для любого фиксированного r. Для любых $s \in (0, k]$ справедливо представление

$$\bar{x}^k(\omega) = \sum_{r=0}^{k} \tau_r \Phi_{r+1}^k x^r(\omega) = \sum_{r=s}^{k} \mu_r^k x^r(\omega) +$$

$$+ \Phi_s^k \left(\sum_{r=0}^{s-1} \nu_r^k x^r(\omega) - \sum_{r=s}^{k} \mu_r^k x^r(\omega)\right); \tag{24.15}$$

---

*) Фихтенгольц Г. М. Курс дифференциального и интегрального исчисления. Т. 2.— М.: Наука, 1966.— 290 с.

$$\mu_r^k = \tau_r \Phi_{r+1}^k/(1 - \Phi_s^k), \qquad r \in [s, k];$$

$$\nu_r^k = \tau_r \Phi_{r+1}^k/\Phi_s^k, \qquad r \in [0, s);$$

$$\sum_{r=0}^{s-1} \nu_r^k = \sum_{r=s}^{k} \mu_r^k = 1.$$

Как будет показано (теорема 24.1), при условиях (24.9) — (24.11) последовательность $\{x^k(\omega)\}_{k=0}^{\infty}$ сходится к некоторому связному подмножеству $X^*_{j(\omega)}$ множества решений X^*. В (24.15) первое слагаемое $\sum_{r=s}^{k} \mu_r^k x^r(\omega)$ при достаточно больших s и всех $k \geqslant s$ будет находиться в сколь угодно малой окрестности выпуклой оболочки $\operatorname{co} X^*_{j(\omega)}$, а второе слагаемое стремится к нулю при $k \to +\infty$. Поэтому $\{\bar{x}^k\}_{k=0}^{\infty}$ сходится в сколь угодно малую окрестность $\operatorname{co} X^*_{j(\omega)}$. Лемма доказана.

Замечание 24.8. Если в задаче (24.1), (24.2) имеются еще линейные ограничения $Nx = b$, где N — матрица размера $m \times n$, а b есть m-мерный вектор, то можно воспользоваться методом проекции (стохастического) градиента на линейное подпространство $L_0 = \{x \mid Nx = 0\}$, как это было сделано в § 10 для метода обобщенного градиента. В § 10 было показано, что проекция псевдоградиента функции $F(x)$ на подпространство L_0 является псевдоградиентом этой функции, рассматриваемой только уже на линейном многообразии $L_b = \{x \mid Nx = b\}$. Спроектированные псевдоградиенты можно использовать в любых методах оптимизации $F(x)$, работающих в многообразии L_b. Ввиду перестановочности операции проектирования на L_0 и операции взятия математического ожидания спроектированные стохастические градиенты функции $F(x) = \mathsf{M} f(x, \theta)$ будут стохастическими градиентами функции $F(x)$, рассматриваемой уже как функция на многообразии L_b. Поэтому, используя спроектированные градиенты, можно запускать рассматриваемые методы стохастического программирования в линейном многообразии L_b и решать задачи с линейными ограничениями.

Теорема 24.1. *Пусть задача невыпуклого стохастического программирования* (24.1), (24.2) *решается методом усредненных стохастических градиентов* (24.4) — (24.11). *Пусть множество* $H^* = \{h(x) \mid 0 \in G_h(x),\ h(x) > 0\}$ *не содержит интервалов. Предположим, что последовательность* $\{x^k(\omega)\}_{k=0}^{\infty}$ *ограничена для любого* ω; *это имеет место, например, при достаточно малых величинах* $\rho = \sup_{k \geqslant 0} \rho_k$ *и* $R = \sup_{k \geqslant 0} \sum_{r=r_k}^{k} \rho_r$. *Тогда:*

либо все предельные точки $\{x^k(\omega)\}$ *не принадлежат допустимой области* $D = \{x \mid h(x) \leqslant 0\}$, *а принадлежат множеству* $X_h^* = \{x \mid 0 \in G_h(x),\ h(x) > 0\}$ *и существует предел* $\lim_{k \to \infty} h(x^k(\omega)) > 0$;

либо все предельные точки $\{x^k(\omega)\}$ *принадлежат допустимой области* D; *в этом случае минимальные из них по значению* F P_ω*-почти наверное* (P_ω-п.н.) *принадлежат множеству* $X^* = \{x \in E_n \mid 0 \in G(x),\ h(x) \leqslant 0\}$ *и интервал* $[\underline{\lim}_{k\to\infty} F(x^k(\omega)), \overline{\lim}_{k\to\infty} F(x^k(\omega))]$ P_ω-п. н. *вложен в множество* $F^* = \{f(x) \mid x \in X^*\}$. *Если при этом* F^* *не содержит интервалов, то все предельные точки* $\{x^k(\omega)\}$ P_ω-п.н. *принадлежат связной компоненте* X^* *и* P_ω-п.н. *существует предел* $\lim_{k\to\infty} F(x^k(\omega))$.

Лемма 24.2. *Пусть существует* $c > \max(0, h(x^0))$, *не принадлежащее* H^*. *Тогда для любого* $d > c$ *найдутся достаточно малые* ρ' *и* R' *такие, что для любой последовательности* $\{\rho_k\}$, *удовлетворяющей* (24.10) *с* $\rho \leqslant \rho'$ *и* $R \leqslant R'$, *для всех* ω *последовательности* $\{x^k(\omega)\}$, *запущенные из* x^0 *согласно* (24.4) — (24.9), *не выходят из ограниченной области* $\{x \mid h(x) \leqslant d\}$.

Доказательство. Лемма справедлива в силу того, что вне области $D_\varepsilon = \{x \mid h(x) \geqslant \varepsilon > 0\}$ для достаточно больших k метод усредненных стохастических градиентов минимизирует детерминированную функцию $h(x)$ и поэтому работает как детерминированный метод усредненных градиентов. Действительно, предположим противное; тогда для некоторого $d > c$ найдутся $\{\rho_s^k\}$ $(s = 0, 1, \ldots)$ с $\sup_{k\geqslant 0} \rho_s^k = \rho_s' \to 0$ и $\sup_{k\geqslant 0} \sum_{r=r_k}^{k} \rho_s^k = R_s' \to 0$ такие, что соответствующие последовательности $\{x_s^k(\omega^s)\}$ $(\omega^s \in \Omega)$, запущенные из точки x^0 для $s = 0, 1, \ldots$ $(\omega^s \in \Omega)$, уходят на бесконечность. Выделим индексы $k_s(\omega^s)$ и $t_s(\omega^s)$ такие, что при $k \in (k_s, t_s)$

$$h(x_s^{k_s}(\omega^s)) \leqslant c < h(x_s^k(\omega^s)) < d \leqslant h(x_s^s(\omega^s)). \tag{24.16}$$

Последовательность $\{x_s^{k_s}(\omega^s)\}_{s=0}^{\infty}$ ограничена; не теряя общности, будем считать, что $\lim_{s\to\infty} x_s^{k_s}(\omega^s) = x'$. В предположениях (24.10) и в силу (24.16) имеет место

$$\lim_{s\to\infty} h(x_s^{k_s}(\omega^s)) = \lim_{s\to\infty} h(x_s^{k_s+1}(\omega^s)) = c \notin H^*;$$

поэтому $x' \notin X_h^*$. Представим

$$x_s^{k+1}(\omega^s) = x_s^k(\omega^s) - \rho_s^k P_s^k(\omega^s) = x_s^k(\omega^s) - \bar{\rho}_s^k \bar{P}_s^k(\omega^s),$$

где

$$\bar{\rho}_k = \rho_k \sum_{r=r_k}^{k} \lambda_{kr}/H_r(\omega^s),$$

$$\bar{P}^k(\omega^s) = \sum_{r=r_k}^{k} \frac{\lambda_{kr}}{H_r(\omega^s)} g^r(\omega^s) \Big/ \sum_{r=r_k}^{k} \frac{\lambda_{kr}}{H_r(\omega^s)}.$$

Для достаточно больших s при $k \in [k_s, t_s]$ и $r \geqslant r_k$ имеет место $g^r(\omega^s) = g_h(x_s^r(\omega^s)) \in G_h(x_s^r(\omega^s))$, и метод усредненных стохастических градиентов работает как детерминированный метод усредненных градиентов. В этом случае

$$\| g^r(\omega) \| \leqslant \Gamma = \sup \{ \| g(x, \theta) \| \mid h(x) \leqslant d \} < +\infty,$$

$$\overline{P}_s^k(\omega^s) \in \mathrm{co} \{ G_h(y) \mid \| y - x^k(\omega^s) \| \leqslant \delta_s^k \},$$

$$\delta_s^k = \frac{\Gamma}{\ } \sum_{r=r_k}^{k} \rho_s^r \leqslant \varepsilon_s = \frac{\Gamma}{\nu} \sup_{k \geqslant k_s} \sum_{r=r_k}^{k} \rho_s^r \to 0, \qquad s \to \infty.$$

Возьмем $\delta > 0$ такое, что $\max \{ h(y) \mid \| y - x \| \leqslant \delta \} < d$. В силу (24.16) и непрерывности $h(y)$ последовательности $\{x_s^k(\omega^s)\}_{k=k_s}^{t_s}$ $(s = 0, 1, \ldots)$ выходят из δ-окрестности точки x'. Обозначим через m_s моменты первого такого выхода. При достаточно больших s и $k \in [k_s, m_s)$ имеет место $\| \overline{P}_s^k \| \leqslant \Gamma < +\infty$ и $\sum_{k=k_s}^{m_s - 1} \bar{\rho}_s^k \geqslant \sigma > 0$. К последовательностям $\{x_s^k(\omega^s),\ k \in [k_s, m_s]\}$ применима лемма 14.1, минимизирующее свойство 2) которой противоречит построению (24.16). Лемма доказана.

Лемма 24.3. *Предположим, что последовальности $\{x^k(\omega)\}$, построенные согласно* (24.4) — (24.11), *ограничены равномерно по* ω. *Пусть $\bar{g}^r(\omega) = M_\omega g^r(\omega) \mid x^0(\omega), \ldots, x^r(\omega)$ — условное математическое ожидание по ω функции $g^r(\omega)$ при фиксированных $x^0(\omega), \ldots, x^r(\omega)$; опрeделим величины*

$$\xi_0^k(\omega) = \sum_{t=0}^{k-1} \rho_t \sum_{r=r_t}^{t} \frac{\lambda_{tr}}{H_r(\omega)} (g^r(\omega) - \bar{g}^r(\omega)). \tag{24.17}$$

Тогда случайная последовательность $\{\xi_0^k(\omega)\}_{k=0}^{\infty}$ P_ω-п. н. имеет предел.

Доказательство. Обычно такого рода утверждения легко следуют из теории мартингалов [133]. Здесь некоторая трудность в том, что последовательность $\{\xi_0^k(\omega)\}_{k=0}^{\infty}$ не является мартингалом относительно потока σ-алгебр, порожденных случайными величинами $\{x^k(\omega)\}_{k=0}^{\infty}$. Мы искусственно построим последовательность $\{\bar{\xi}_0^k(\omega)\}_{k=0}^{\infty}$, близкую к $\{\xi_0^k(\omega)\}_{k=0}^{\infty}$ и являющуюся мартингалом относительно потока σ-алгебр, порожденных $\{x^k(\omega)\}_{k=0}^{\infty}$. Можно преобразовать:

$$\xi_0^k = \sum_{t=0}^{k-1} \rho_t \sum_{r=r_t}^{t} \lambda_{tr} (g^r - \bar{g}^r)/H_r = \sum_{r=0}^{k-1} \left(\sum_{t=r}^{k-1} \lambda_{tr} \rho_t \right) (g^r - \bar{g}^r)/H_r =$$

$$= \sum_{r=0}^{k-1} \left(\sum_{t=r}^{\infty} \lambda_{tr} \rho_t \right) (g^r - \bar{g}^r)/H_r - \sum_{r=0}^{k-1} \left(\sum_{t=k}^{\infty} \lambda_{tr} \rho_t \right) (g^r - \bar{g}^r)/H_r.$$

Случайная последовательность

$$\xi_0^k(\omega)=\sum_{r=0}^{k-1}\left(\sum_{t=r}^{\infty}\lambda_{tr}\rho_t\right)(g^r-\bar{g}^r)/H_r$$

образует мартингал относительно $\{x^k(\omega)\}_{k=0}^{\infty}$. Пусть

$$\Gamma=\sup_{\omega\in\Omega}\sup_{k\geqslant 0}\|g^k(\omega)\|<\infty.$$

Для случайных величин

$$\delta^k(\omega)=\sum_{r=0}^{k-1}\left(\sum_{t=k}^{\infty}\lambda_{tr}\rho_t\right)(g^r-\bar{g}^r)/H_r$$

справедливы оценки

$$\|\delta^k\|\leqslant\frac{2\Gamma}{\nu}\sum_{r=0}^{k}\left(\sum_{t=k}^{\infty}\lambda_{tr}\rho_t\right)=\frac{2\Gamma}{\nu}\sum_{t=k}^{\infty}\rho_t\sum_{r=0}^{k}\lambda_{tr}=$$

$$=\frac{2\Gamma}{\nu}\sum_{t=k}^{\infty}\rho_t\sum_{r=r_t}^{k}\lambda_{tr}\leqslant\frac{2\Gamma}{\nu}\sum_{t=k}^{q(k)}\rho_t\leqslant\frac{2\Gamma}{\nu}\sum_{t=r_{q(k)}}^{q(k)}\rho_t,$$

где величины r_t и $r_{q(k)}$ получаются из r_k заменой k на t и $q(k)$ соответственно; $q(k)=\sup\{t\mid t\geqslant k, r_t\leqslant k\}$. Так как $q(k)\to+\infty$ при $k\to\infty$, то в силу условия (24.10) $\|\delta^k\|\to 0$ при $k\to\infty$.

Справедливо

$$M\|\bar{\xi}_0^k(\omega)\|^2\leqslant\frac{4\Gamma^2}{\nu^2}\sum_{r=0}^{\infty}\left(\sum_{t=r}^{\infty}\lambda_{tr}\rho_t\right)^2<+\infty;$$

отсюда $M\|\bar{\xi}_0^k(\omega)\|\leqslant 1+M\|\bar{\xi}_0^k\|^2<+\infty$, и так как $\{\bar{\xi}_0^k(\omega)\}_{k=0}^{\infty}$ есть мартингал, то последовательность $\{\bar{\xi}_0^k(\omega)\}$, а следовательно, и последовательность $\{\xi_0^k(\omega)\}$ P_ω-п. н. имеют предел (см. [133, с. 496]). Лемма доказана.

Лемма 24.4. *Пусть* ω *таково, что* $\{\xi_0^k(\omega)\}_{k=0}^{\infty}$ *имеет предел. Предположим, что* $\lim\limits_{s\to\infty}x^{k_s}(\omega)=x(\omega)\notin X^*$. *Обозначим*

$$m_s(\varepsilon,\omega)=\sup\{m\mid\|x^k(\omega)-x\|\leqslant\varepsilon\ \text{ при }\ k\in[k_s,m)\}.$$

Тогда найдется $\bar{\varepsilon}(\omega)$ *такое, что для любого* $\varepsilon\in(0,\bar{\varepsilon}]$ *существуют индексы* $l_s(\omega)\in[k_s(\omega),m_s(\varepsilon,\omega)]$ *такие, что*

1) $F(x(\omega))=\lim\limits_{s\to\infty}F(x^{k_s}(\omega))>\overline{\lim\limits_{s\to\infty}}F(x^{l_s}(\omega))$, $\quad h(x(\omega))<0$;

2) $h(x(\omega))=\lim\limits_{s\to\infty}h(x^{k_s}(\omega))>\overline{\lim\limits_{s\to\infty}}h(x^{l_s}(\omega))$, $\quad h(x(\omega))>0$;

3) $F(x(\omega))=\lim\limits_{s\to\infty}F(x^{k_s}(\omega))>\overline{\lim\limits_{s\to\infty}}F(x^{l_s}(\omega))$, $\quad 0\geqslant\overline{\lim\limits_{s\to\infty}}h(x^{l_s}(\omega))$, $h(x(\omega))=0$.

Доказательство. Справедливо представление

$$x^{k+1}(\omega) = x^{k_s}(\omega) - \sum_{t=k_s}^{k} \rho_t \sum_{r=r_t}^{t} \lambda_{tr} g^r / H_r =$$

$$= x^{k_s}(\omega) - \sum_{t=k_s}^{k} \rho_t \sum_{r=r_t}^{t} \lambda_{tr} \bar{g}^r / H_r - \xi_{k_s}^{k+1}(\omega) =$$

$$= x^{k_s}(\omega) - \sum_{t=k_s}^{k} \rho_t \bar{P}^t(\omega) - \xi_{k_s}^{k+1}(\omega) = \bar{x}_{k_s}^{k+1}(\omega) - \xi_{k_s}^{k+1}(\omega),$$

где

$$\bar{x}_{k_s}^{k+1}(\omega) = \sum_{t=k_s}^{k} \rho_t \bar{P}^t(\omega) = \bar{x}_{k_s}^{k}(\omega) - \rho_k \bar{P}^k(\omega), \tag{24.18}$$

$$\bar{P}^t(\omega) = \sum_{r=r_t}^{t} \lambda_{tr} \bar{g}^r(\omega)/H_r(\omega), \tag{24.19}$$

$$\bar{g}^r(\omega) = \mathsf{M}_\omega g^r(\omega) \mid x^0(\omega), \ldots, x^r(\omega), \tag{24.20}$$

$$\xi_n^m(\omega) = \sum_{t=n}^{m-1} \rho_t \sum_{r=r_t}^{t} \lambda_{tr} (g^r(\omega) - \bar{g}^r(\omega))/H_r(\omega).$$

Вместо $\{x^k(\omega)\}_{k=0}^{\infty}$ будем изучать поведение близких последовательностей $\{\bar{x}_{k_s}^k(\omega)\}_{k \geqslant k_s}$ $(s = 0, 1, \ldots)$, генерируемых при фиксированном ω детерминированным методом усредненных градиентов (24.18) — (24.20), в котором используются псевдоградиенты функций F и h, взятые не в точках $\bar{x}^k(\omega)$, а в близких точках $x^k(\omega)$, причем

$$\| \bar{x}_{k_s}^k(\omega) - x^k(\omega) \| \leqslant \| \xi_{k_s}^k(\omega) \| \leqslant \sup_{k \geqslant k_s} \| \xi_{k_s}^k(\omega) \| = \delta_{k_s}(\omega)$$

и $\lim\limits_{s \to \infty} \delta_{k_s}(\omega) = 0$, так как по условию леммы сходится ряд $\xi_0^\infty(\omega)$. Заметим, что

$$| F(x^k(\omega)) - F(\bar{x}_{k_s}^k(\omega)) | \leqslant L_F \| \xi_{k_s}^k(\omega) \|,$$

$$| h(x^k(\omega)) - h(\bar{x}_{k_s}^k(\omega)) | \leqslant L_h \| \xi_{k_s}^k(\omega) \|,$$

где L_F и L_h — константы Липшица функций F и h для компакта, содержащего $\{x^k(\omega)\}_{k=0}^{\infty}$. Отсюда следует, что разности $| F(x^k(\omega)) - F(\bar{x}_{k_s}^k(\omega)) |$ и $| h(x^k(\omega)) - h(\bar{x}_{k_s}^k(\omega)) |$ могут быть сделаны равномерно по k сколь угодно малыми при достаточно больших s.

Пусть $\lim\limits_{s\to\infty} \bar{x}_{k_s}^{k_s}(\omega) = \lim\limits_{s\to\infty} x^{k_s}(\omega) = x(\omega)$, $0 \notin G(x)$. Покажем, что последовательности $\{\bar{x}_{k_s}^{k}(\omega)\}_{k\geqslant k_s}$ обладают следующим локальным минимизирующим свойством.

Существует $\bar{\varepsilon}(\omega)$ такое, что для любого $\varepsilon \in (0, \bar{\varepsilon}(\omega)]$ найдутся индексы $l_s(\omega) \geqslant k_s$ такие, что для достаточно больших s выполнено $\| x_{k_s}^{k}(\omega) - x(\omega) \| \leqslant \varepsilon$ при $k \in [k_s, l_s)$ и имеет место:

1) $F(x(\omega)) = \lim\limits_{s\to\infty} F(\bar{x}_{k_s}^{k_s}(\omega)) > \overline{\lim\limits_{s\to\infty}} F(\bar{x}_{k_s}^{l_s}(\omega))$, $h(x(\omega)) < 0$;

2) $h(x(\omega)) = \lim\limits_{s\to\infty} h(\bar{x}_{k_s}^{k_s}(\omega)) > \overline{\lim\limits_{s\to\infty}} h(\bar{x}_{k_s}^{l_s}(\omega))$, $h(x(\omega)) > 0$;

3) $F(x(\omega)) = \lim\limits_{s\to\infty} F(\bar{x}_{k_s}^{k_s}(\omega)) > \overline{\lim\limits_{s\to\infty}} F(\bar{x}_{k_s}^{l_s}(\omega))$, $0 \geqslant \overline{\lim\limits_{s\to\infty}} h(\bar{x}_{k_s}^{l_s}(\omega))$, $h(x(\omega)) = 0$.

Указанное минимизирующее свойство последовательностей $\{\bar{x}_{k_s}^{k}(\omega)\}_{k\geqslant k_s}$ $(s = 0, 1, \ldots)$ аналогично свойству леммы 14.2 и означает просто устойчивость метода усредненных градиентов (24.18) — (24.20) по отношению к ошибкам величин, входящих в запись метода, а именно по отношению к ошибкам в точках, где берутся псевдоградиенты функций. Из этого свойства устойчивости детерминированного метода усредненных градиентов, которое в свою очередь следует из устойчивости метода обобщенного градиента (лемма 14.1), ввиду близости последовательностей $\{x^{k}(\omega)\}_{k\geqslant k_s}$ и $\{\bar{x}_{k_s}^{k}(\omega)\}_{k\geqslant k_s}$ следует утверждение леммы.

Итак, покажем, что к последовательностям $\{\bar{x}_{k_s}^{k}(\omega)\}_{k\geqslant k_s}$ $(s = 0, 1, \ldots)$ применима лемма 14.1, откуда и будет следовать требуемое минимизирующее свойство этих последовательностей. Обозначим

$$\Gamma = \sup_{k\geqslant 0} \| g^{k}(\omega) \|, \quad \rho_{k_s}^{k} = \rho_k, \quad k \geqslant k_s; \quad \bar{\rho}_s = \sup_{k\geqslant k_s} \rho_k;$$

$$\delta_{k_s}^{k} = \| \xi_{k_s}^{k} \| + \frac{\Gamma}{\nu} \sum_{r=r_k}^{k} \rho_r, \quad \delta_s = \sup_{k\geqslant k_s} \| \bar{\xi}_{k_s}^{k} \| + \frac{\Gamma}{\nu} \sup_{k\geqslant k_s} \sum_{r=r_k}^{k} \rho_r.$$

Отметим, что

$$\lim_{s\to\infty} \bar{\rho}_s = \lim_{s\to\infty} \delta_s = 0 \quad \text{и} \quad \sum_{k\geqslant k_s} \rho_{k_s}^{k} = +\infty,$$

$$\bar{P}^{k}(\omega) \in \operatorname{co}\{G(y) \mid \| y - \bar{x}_{k_s}^{k}(\omega) \| \leqslant \delta_{k_s}^{k}\}, \qquad k \geqslant k_s.$$

Значит, все условия леммы 14.1 выполнены, откуда следует минимизирующее свойство последовательностей $\{\bar{x}_{k_s}^{k}(\omega)\}_{k\geqslant k_s}$ и $\{x^{k}(\omega)\}_{k\geqslant k_s}$. Лемма доказана.

Д о к а з а т е л ь с т в о т е о р е м ы 24.1. Ограниченность всех последовательностей $\{x^{k}(\omega)\}_{k=0}^{\infty}$ $(\omega \in \Omega)$ при достаточно малых $\rho = \sup\limits_{k\geqslant 0} \rho_k$ и

$R = \sup\limits_{k \geqslant 0} \sum\limits_{r=r_k}^{k} \rho_r$ следует из леммы 24.2. Из (24.5) — (24.10) следует, что $\lim\limits_{k\to\infty} \| x^{k+1}(\omega) - x^k(\omega) \| = 0$ для всех $\omega \in \Omega$. Обозначим через Ω' множество тех ω, для которых последовательность $\{\xi_0^k(\omega)\}_{k=0}^{\infty}$ (см. (24.17)) имеет предел. В силу леммы 24.3 $P_\omega(\Omega') = 1$. Теперь для каждого $\omega \in \Omega'$, используя лемму 24.4, можем доказать утверждение теоремы 24.1 точно так же, как это сделано для детерминированного метода в теореме 14.2. Теорема доказана.

Рассмотрим кратко стохастические аналоги других методов невыпуклой негладкой оптимизации.

2. Метод стохастических обобщенных градиентов задается следующими соотношениями:

$$x^0(\omega) = x^0 \in E_n,$$

$$x^{k+1}(\omega) = x^k(\omega) - \rho_k g^k(\omega)/H_k(\omega),$$

$$g^k(\omega) = g(x^k(\omega), \pi_k(\omega)), \quad \pi_k(\omega) = \theta_k, \quad k = 0, 1, \ldots,$$

$$0 \leqslant \rho_k \leqslant \rho, \quad \sum_{k=0}^{\infty} \rho_k = +\infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < +\infty.$$

Здесь ω, $g^k(\omega)$, $H_k(\omega)$ те же, что и в (24.4) — (24.8). Очевидно, данный метод является частным случаем метода усредненных стохастических градиентов; поэтому для него также справедлива теорема 24.1. Для ускорения сходимости метода последовательность $\{x^k(\omega)\}$ можно усреднять согласно (24.12), (24.14).

3. Стохастический метод тяжелого шарика имеет вид:

$$x^0(\omega) = x^0 \in E_n, \quad P^0(\omega) = g^0(\omega), \tag{24.21}$$

$$x^{k+1}(\omega) = x^k(\omega) - \rho_k P^k(\omega), \tag{24.22}$$

$$P^k(\omega) = (1 - \gamma_k) P^{k-1}(\omega) + \gamma_k g^k(\omega)/H_k(\omega), \tag{24.23}$$

$$g^k(\omega) = g(x^k(\omega), \pi_k(\omega)), \quad \pi_k(\omega) = \theta_k, \tag{24.24}$$

$$0 < \gamma \leqslant \gamma_k \leqslant 1, \quad 0 \leqslant \rho_{k+1} \leqslant \rho_k, \quad \sum_{k=0}^{\infty} \rho_k = +\infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < +\infty. \tag{24.25}$$

Здесь ω, $g^k(\omega)$, $H_k(\omega)$, $\pi_k(\omega)$ имеют тот же смысл и удовлетворяют тем же условиям, что и в методе (24.4) — (24.8) усредненных стохастических градиентов. Отметим, что $\lim\limits_{k\to\infty} \rho_k/\gamma_k = 0$.

Напомним, что уравнение движения (24.22) можно преобразовать к следующему виду:

$$x^{k+1}(\omega) = x^k(\omega) - \rho_k\gamma_k g^k(\omega)/H_k(\omega) - \rho_k(1 - \gamma_k) P^{k-1}(\omega) =$$

$$= x^k(\omega) - \rho_k\gamma_k g^k(\omega)/H_k(\omega) + \frac{\rho_k}{\rho_{k-1}}(1 - \gamma_k)(x^k(\omega) - x^{k-1}(\omega)).$$

Для данного метода выполнены утверждения о сходимости в виде теоремы 24.1. Поясним, почему это так.

Во-первых, стохастический метод тяжелого шарика (24.21) — (24.25) можно представить в виде (14.44) — (14.46). Тогда вне области $D_\varepsilon = \{x \mid h(x) \geqslant \varepsilon > 0\}$ для достаточно больших k он будет работать как детерминированный метод усредненных градиентов, и его ограниченность равномерно по ω можно обеспечить с помощью механизма возврата в начальную точку, описанного в § 14.

Во-вторых, метод можно трактовать как некоторый модифицированный детерминированный метод усредненных градиентов.

Справедливо представление:

$$P^k(\omega) = \sum_{r=0}^{k} \lambda_{kr} g^r / H_r, \quad \lambda_{kr} \geqslant 0, \quad \sum_{r=0} \lambda_{kr} = 1,$$

$$\lambda_{k0} = \prod_{i=1}^{k} (1-\gamma_i), \quad \lambda_{kk} = \gamma_k; \quad \lambda_{kr} = \gamma_r \prod_{i=r+1} (1-\gamma_i),$$

$$0 < r < k.$$

Теперь представим траекторию:

$$x^{k+1}(\omega) = x^{k_s}(\omega) - \sum_{t=k_s}^{k} \sum_{r=0}^{t} \lambda_{tr} g^r / H_r =$$

$$= x^{k_s}(\omega) - \sum_{t=k_s}^{k} \rho_t \sum_{r=r_t} \lambda_{tr} \bar{g}^r / H_r - \xi_k^{k+1}(\omega) =$$

$$= x^{k_s}(\omega) - \sum_{t=k_s}^{k} \rho_t \bar{P}^t(\omega) - \xi_{k_s}^{k+1}(\omega) = \bar{x}_{k_s}^{k+1}(\omega) - \xi_{k_s}^{k+1}(\omega),$$

где

$$\bar{x}_k^{k+1}(\omega) = \sum_{t=k_s}^{k} \rho_t \bar{P}^t(\omega) = \bar{x}_{k_s}^k(\omega) - \rho_k \bar{P}^k(\omega),$$

$$\bar{P}^t(\omega) = \sum_{r=0}^{t} \lambda_{tr} \bar{g}^r(\omega) / H_r(\omega),$$

$$\bar{g}^r(\omega) = M_\omega g^r(\omega) \mid x^0(\omega), \ldots, x^r(\omega),$$

$$\xi_k^{k+1}(\omega) = \sum_{t=k_s}^{k} \rho_t \sum_{r=0} \lambda_{tr} (g^r(\omega) - \bar{g}^r(\omega)) / H_r(\omega).$$

Представим

$$\bar{P}^k = \sum_{r=0}^{k} \lambda_{kr} \bar{g}^r / H_r = \sum_{r=r_k} \lambda_{kr} \bar{g}^r / H_r + \sum_{r=0}^{r_k - 1} \lambda_{kr} \bar{g}^r / H_r.$$

Обозначим

$$Q^k = \sum_{r=r_k}^{k} \lambda_{kr} \frac{\bar{g}^r}{H_r} \Big/ \sum_{r=r_k}^{k} \lambda_{kr}, \quad \rho'_k = \rho_k \sum_{r=r_k}^{k} \lambda_{kr},$$

$$\Delta^k = \sum_{r=0}^{r_k-1} \lambda_{kr} \frac{\bar{g}^r}{H_r} \Big/ \sum_{r=r_k}^{k} \lambda_{kr}.$$

Так как $\lim\limits_{k\to\infty} \rho_k/\gamma_k \leqslant \lim\limits_{k\to\infty} \rho_k/\gamma = 0$, то существуют r_k, удовлетворяющие условиям леммы 14.4.

Предположим, что последовательности $\{x^k(\omega)\}$ ограничены равномерно по ω. Обозначим $\Gamma = \sup\limits_{\omega\in\Omega} \sup\limits_{k\geqslant 0} \| g^k(\omega) \| < \infty$.

Выберем r_k согласно лемме 14.4, тогда $\lim\limits_{k\to\infty} \Delta^k = 0$. С учетом введенных обозначений перепишем

$$\bar{x}_{k_s}^{k+1}(\omega) = \bar{x}_{k_s}^{k}(\omega) - \rho_k \bar{P}^k(\omega) = \bar{x}_{k_s}^{k}(\omega) - \rho'_k (Q^k(\omega) + \Delta^k(\omega)), \tag{24.26}$$

$$x^{k+1}(\omega) = \bar{x}_{k_s}^{k+1}(\omega) - \xi_{k_s}^{k+1}(\omega).$$

Для каждого ω соотношение (24.26) можно рассматривать как детерминированный метод усредненных градиентов, в котором используются псевдоградиенты функций задачи, взятые не в точках $\bar{x}_{k_s}^{k}(\omega)$, а в близких точках $x^k(\omega)$, и усреднение градиентов осуществляется с некоторой ошибкой $\Delta^k(\omega)$.

Если мы покажем, что $\sup\limits_{k\geqslant k_s} \| \xi_{k_s}^{k}(\omega) \| \to 0$ P_ω-п. н. при $s\to\infty$, то, используя устойчивость метода обобщенного градиента (теорема 14.1), можно доказать минимизирующее свойство и сходимость P_ω-п. н. стохастического метода тяжелого шарика точно так же, как это сделано для метода усредненных стохастических градиентов в лемме 24.4 и теореме 24.1.

Таким образом, необходимо показать, что P_ω-п. н. сходится последовательность $\{\xi_0^k(\omega)\}$, где

$$\xi_0^k(\omega) = \sum_{t=0}^{k-1} \rho_t \sum_{r=0}^{t} \lambda_{tr} (g^r(\omega) - \bar{g}^r(\omega))/H_r(\omega), \quad k \geqslant 0.$$

Так же, как в лемме 24.3, представим

$$\xi_0^k = \sum_{t=0}^{k-1} \rho_t \sum_{r=0}^{t} \lambda_{tr} (g^r - \bar{g}^r)/H_r = \sum_{r=0}^{k-1} \left(\sum_{t=r}^{k-1} \rho_t \lambda_{tr} \right) (g^r - \bar{g}^r)/H_r =$$

$$= \sum_{r=0}^{k-1} \left(\sum_{t=r}^{\infty} \lambda_{tr}\rho_t \right) (g^r - \bar{g}^r)/H_r - \sum_{r=0}^{k-1} \left(\sum_{t=k}^{\infty} \lambda_{tr}\rho_t \right) (g^r - \bar{g}^r)/H_r,$$

где

$$\sum_{t=r}^{\infty} \lambda_{tr}\rho_t \leqslant \rho_r \sum_{t=r}^{\infty} (1-\gamma)^{t-r} = \rho_r/\gamma < +\infty;$$

$$\sum_{t=k}^{\infty} \lambda_{tr}\rho_t \leqslant \rho_k \sum_{t=k}^{\infty} (1-\gamma)^{t-r} = \rho_k/\gamma < +\infty \quad \text{при } r \leqslant k.$$

Случайная последовательность

$$\bar{\xi}_0^k(\omega) = \sum_{r=0}^{k-1} \left(\sum_{t=r}^{\infty} \lambda_{tr}\rho_t \right) (g^r - \bar{g}^r)/H_r$$

образует мартингал относительно потока σ-алгебр, порожденных случайными величинами $\{x^k(\omega)\}_{k=0}^{\infty}$. Имеет место

$$M_\omega \| \bar{\xi}_0^k(\omega) \|^2 \leqslant \frac{4\Gamma^2}{\nu^2} \sum_{r=0}^{\infty} \left(\sum_{t=r}^{\infty} \lambda_{tr}\rho_t \right)^2 \leqslant \frac{4\Gamma^2}{\nu^2} \sum_{r=0}^{\infty} \rho_r^2 \left(\sum_{t=r}^{\infty} \lambda_{tr} \right)^2 \leqslant$$

$$\leqslant \frac{4\Gamma^2}{\nu^2} \sum_{r=0}^{\infty} \rho_r^2 \left(\sum_{t=r}^{\infty} (1-\gamma)^{t-r} \right)^2 \leqslant \frac{4\Gamma^2}{\gamma^2\nu^2} \sum_{r=0}^{\infty} \rho_r^2 < +\infty.$$

Поэтому мартингал $\{\bar{\xi}_0^k(\omega)\}_{k=0}^{\infty}$ имеет предел P_ω-п. н. Для случайного вектора

$$\delta^k(\omega) = \sum_{r=0}^{k-1} \left(\sum_{t=k}^{\infty} \lambda_{tr}\rho_t \right) (g^r - \bar{g}^r)/H_r$$

справедливы оценки

$$\| \delta^k(\omega) \| \leqslant \frac{2\Gamma}{\nu} \rho_k \sum_{r=0}^{k-1} \sum_{t=k}^{\infty} \lambda_{tr} \leqslant$$

$$\leqslant \frac{2\Gamma}{\nu} \rho_k \sum_{r=0}^{k-1} \sum_{t=k}^{\infty} (1-\gamma)^{t-r} \leqslant \frac{2\Gamma}{\gamma\nu} \rho_k \to 0, \quad k \to \infty.$$

Таким образом, последовательность $\{\xi_0^k(\omega)\}_{k=0}^{\infty}$ имеет предел P_ω-п. н. Сходимость стохастического метода тяжелого шарика обоснована.

4. **Стохастический метод овражного шага.** Детерминированный метод овражного шага подробно рассмотрен в § 14. Его стохастический аналог для решения задачи (24.1), (24.2) имеет вид

$$x^0(\omega) = y^0(\omega) = x^0 \in E_n, \tag{24.27}$$

$$y^{k+1}(\omega) = x^k(\omega) - \rho_k g^k(\omega), \tag{24.28}$$

$$g^k(\omega) = g(x^k(\omega), \pi_k(\omega)), \quad \pi_k(\omega) = \theta_k, \tag{24.29}$$

$$x^{k+1}(\omega) = y^{k+1}(\omega) + \lambda_k (y^{k+1}(\omega) - y^k(\omega)), \quad k = 0, 1, \dots, \tag{24.30}$$

где множители градиентного шага ρ_k и овражного шага λ_k удовлетворяют условиям:

$$0 \leqslant \rho_{k+1} \leqslant \rho_k \leqslant \rho, \quad \sum_{k=0}^{\infty} \rho_k = +\infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < +\infty, \tag{24.31}$$

$$0 \leqslant \lambda_k \leqslant \lambda < 1. \tag{24.32}$$

Здесь θ_k, ω, $g^k(\omega)$, $\pi_k(\omega)$ имеют тот же смысл, что и для метода усредненных стохастических градиентов (24.4) — (24.8), а $x^k(\omega)$, $y^k(\omega)$, ρ_k, λ_k аналогичны соответствующим величинам в методе овражного шага (14.48) — (14.50).

Предположим, что существуют индексы $r_k \leqslant k$ такие, что $\lim\limits_{k\to\infty} r_k = +\infty$ и

$$\lim_{k\to\infty} \sum_{r=r_k}^{k} \rho_r = 0, \quad \lim_{k\to\infty} \sum_{r=r_k}^{k} \lambda^{r_k - r} \rho_r = +\infty. \tag{24.33}$$

В § 14 было показано, что такие r_k существуют для $\rho_k = \text{const}/k^{\alpha}$, $0 < \alpha < 1$.

Заметим, что если для некоторых r_k выполнены условия (24.33), то имеет место

$$\lim_{k\to\infty} \sum_{r=0}^{k} \lambda^{k-r} \rho_r = 0. \tag{24.34}$$

Действительно, представим

$$\sum_{r=0}^{k} \lambda^{k-r} \rho_r = \sum_{r=r_k}^{k} \lambda^{k-r} \rho_k + \lambda^{k-r_k} \sum_{r=0}^{r_k-1} \lambda^{r_k-r} \rho_r.$$

Справедливы оценки

$$\sum_{r=r_k}^{k} \lambda^{k-r} \rho_r \leqslant \sum_{r=r_k}^{k} \rho_k, \quad \sum_{r=0}^{r_k-1} \lambda^{r_k-r} \rho_r \leqslant \frac{\lambda\rho}{1-\lambda} < +\infty,$$

$$\sum_{r=0}^{k} \lambda^{k-r} \rho_r \leqslant \sum_{r=r_k}^{k} \rho_r \left(1 + \frac{\lambda^{k-r_k}}{\sum\limits_{r=r_k}^{k} \lambda^{k-r} \rho_r} \frac{\lambda\rho}{1-\lambda} \right) \to 0,$$

$$k \to \infty.$$

Как и в детерминированном случае, метод (24.27) — (24.32) можно привести к виду

$$x^0(\omega) = x^0 \in E_n, \quad x^1(\omega) = x^0 - (1 + \lambda_0) \rho_0 g^0(\omega),$$

$$x^{k+1}(\omega) = x^k(\omega) - (1 + \lambda_k) \rho_k g^k(\omega) + {}$$
$$+ \lambda_k \rho_{k-1} g^{k-1}(\omega) + \lambda_k (x^k(\omega) - x^{k-1}(\omega)).$$

Из этих соотношений следует, что

$$x^{k+1}(\omega) = x^k(\omega) - \rho_k g^k(\omega) - \sum_{r=0}^{k} \lambda_r \lambda_{r+1} \ldots \lambda_k \rho_r g^r(\omega). \qquad (24.35)$$

Таким образом, метод (24.27) — (24.30), в сущности, работает, как метод усредненных стохастических градиентов. Поскольку $\lambda_r \leqslant \lambda < 1$, то основную роль в правой части (24.35) играют стохастические градиенты $g^r(\omega)$ с r, близкими к k, т. е. стохастические градиенты, взятые в точках, близких к $x^k(\omega)$.

Для метода (24.27) — (24.30) также справедлива теорема 24.1. Доказательство этого факта проводится так же, как доказательство сходимости стохастического метода тяжелого шарика. В частности, ограниченность последовательности $\{x^k(\omega)\}$ равномерно по ω можно обеспечить с помощью механизма возврата в начальную точку, описанного в § 14.

Представим соотношение (24.35) в виде

$$x^{k+1}(\omega) = x^k(\omega) - \rho_k P^k(\omega), \qquad (24.36)$$

$$P^k(\omega) = \sum_{r=0}^{k} \lambda_{kr} g^r(\omega),$$

$$\lambda_{kk} = 1 + \lambda_k, \quad \lambda_{kr} = \frac{\rho_r}{\rho_k} \lambda_r \lambda_{r+1} \ldots \lambda_k \leqslant \frac{\rho_r}{\rho_k} \lambda^{k-r+1}, \qquad 0 \leqslant r < k.$$

Начиная с некоторого момента k_s, наряду с (24.36) будем рассматривать близкий процесс

$$\bar{x}_{k_s}^{k_s}(\omega) = x^{k_s}(\omega),$$

$$\bar{x}_{k_s}^{k+1}(\omega) = \bar{x}_{k_s}^{k}(\omega) - \rho_k \bar{P}^k(\omega) = x^{k_s}(\omega) - \sum_{t=k_s}^{k} \rho_t \bar{P}^t(\omega) =$$

$$= x^{k_s}(\omega) - \rho_k \bar{g}^k(\omega) - \sum_{r=0}^{k} \lambda_r \lambda_{r+1} \ldots \lambda_k \rho_r \bar{g}^r(\omega), \qquad (24.37)$$

$$\bar{P}^k(\omega) = \sum_{r=0}^{k} \lambda_{kr} \bar{g}^r(\omega),$$

$$\bar{g}^r(\omega) = M_\omega g^r(\omega) \mid x^0(\omega), \ldots, x^k(\omega).$$

Между $x^{k+1}(\omega)$ и $\bar{x}_{k_s}^{k+1}(\omega)$ есть следующая связь:

$$x^{k+1}(\omega) = \bar{x}_{k_s}^{k+1}(\omega) - \xi_{k_s}^{k+1}(\omega),$$
$$\xi_{k_s}^{k+1}(\omega) = \sum_{t=k_s}^{k} \rho_t \sum_{r=0}^{t} \lambda_{tr}(g^r - \bar{g}^r). \qquad (24.38)$$

Таким образом, можно считать, что процесс (24.37) для каждого ω как бы порождается детерминированным методом овражного шага с тем отличием, что градиенты для этого метода берутся не в точках $\bar{x}^k_{k_s}(\omega)$, а в близких точках $x^k(\omega)$, причем $\|\bar{x}^k_{k_s}(\omega) - x^k(\omega)\| = \|\xi^k_{k_s}(\omega)\|$. Если ряд $\xi^\infty_0(\omega)$ сходится, то $\sup\limits_{k \geq k_s} \|\xi^k_{k_s}(\omega)\| \to 0$ при $s \to \infty$.

Соотношение (24.37) можно привести к виду (14.56), и тогда, используя устойчивость метода обобщенного градиента (теорема 14.1), можно доказать минимизирующее свойство и сходимость P_ω-п. н. стохастического метода овражного шага точно так же, как это сделано для метода усредненных стохастических градиентов в лемме 24.4 и теореме 24.1.

Итак, осталось показать, что ряд $\xi^\infty_0(\omega)$ имеет предел P_ω-п. н. Как и в лемме 24.3, представим

$$\xi^k_0(\omega) = \sum_{r=0}^{k-1}\left(\sum_{t=r}^{k-1}\lambda_{tr}\rho_t\right)(g^r - \bar{g}^r) = \bar{\xi}^k_0(\omega) - \delta^k(\omega),$$

где

$$\bar{\xi}^k_0(\omega) = \sum_{r=0}^{k-1}\left(\sum_{t=r}^{\infty}\lambda_{tr}\rho_t\right)(g^r - \bar{g}^r),$$

$$\delta^k(\omega) = \sum_{r=0}^{k-1}\left(\sum_{t=k}^{\infty}\lambda_{tr}\rho_t\right)(g^r - \bar{g}^r).$$

Предположим, что последовательность $\{x^k(\omega)\}_{k=0}^\infty$ ограничена равномерно по ω. Обозначим $\Gamma = \sup\limits_{\omega\in\Omega}\sup\limits_{r\geq 0}\|g^r(\omega)\| < \infty$. Случайная последовательность $\{\bar{\xi}^k_0(\omega)\}$ образует мартингал относительно потока σ-алгебр, порожденных случайными векторами $\{x^k(\omega)\}$. Справедливы оценки

$$M_\omega\|\bar{\xi}^k_0(\omega)\|^2 \leq 4\Gamma^2\sum_{r=0}^{\infty}\left(\sum_{t=r}^{\infty}\lambda_{tr}\rho_t\right)^2 \leq$$

$$\leq 4\Gamma^2\sum_{r=0}^{\infty}\rho_r^2\left(1 + \sum_{t=r}^{\infty}\lambda^{t-r}\right)^2 \leq 4\Gamma^2\left(1 + \frac{1}{1-\lambda}\right)^2\sum_{r=0}^{\infty}\rho_r^2 < \infty. \qquad (24.39)$$

Поэтому мартингал $\{\bar{\xi}^k_0(\omega)\}$ имеет предел P_ω-п. н.

Для случайных величин $\delta^k(\omega)$ справедливо

$$\|\delta^k(\omega)\| \leq 2\Gamma\sum_{r=0}^{k-1}\sum_{t=k}^{\infty}\lambda_{tr}\rho_t \leq 2\Gamma\sum_{r=0}^{k-1}\rho_r\sum_{t=k}^{\infty}\lambda^{t-r+1} \leq$$

$$\leq 2\Gamma\sum_{r=0}^{k}\lambda^{k-r}\rho_r \to 0, \quad k \to \infty \quad (\text{см. } (24.34)).$$

Таким образом, последовательность $\{\xi^k_0(\omega)\}$ имеет предел P_ω-п. н. Сходимость стохастического метода овражного шага обоснована.

§ 25. Методы конечных разностей в стохастическом программировании

1. **Метод стохастической аппроксимации.** Это один из первых прямых методов решения стохастической задачи на безусловный экстремум. Впервые он был предложен в работе [174] для поиска в пространстве E_n корня функции $F(x) = \mathsf{M} f(x, \theta)$. Затем в [155] подобные процедуры использовались для минимизации $F(x)$.

Метод стохастической аппроксимации Кифера — Вольфовица, предложенный для минимизации $F(x)$, определяется соотношением

$$x^{k+1} = x^k - \rho_k \sum_{i=1}^{n} \frac{f(x^k + \Delta_k e_i, \theta^k) - f(x^k - \Delta_k e_i, \theta^k)}{\Delta_k} e_i, \qquad (25.1)$$

где θ^k — независимые наблюдения θ. Сходимость последовательности (25.1) к минимуму $F(x)$ обычно исследуется в предположении, что $F(x)$ — непрерывно дифференцируемая функция.

Небольшое видоизменение процедур стохастической аппроксимации приводит к стохастическим методам минимизации негладких функций вероятностной природы. Построение этих методов связано со сглаживанием функций приемом, который изучался в гл. 1.

Пусть при фиксированном θ функция $f(x, \theta)$ удовлетворяет локальному условию Липшица, т. е. для некоторого ограниченного множества X в E_n существует такая константа $L(\theta)$ в $[0, \infty)$, что

$$|f(x, \theta) - f(y, \theta)| \leqslant L(\theta) \, \|x - y\|$$

для любых точек $x, y \in X$. Далее считаем, что $\mathsf{M} L(\theta) < \infty$.

Возьмем

$$F(x, \alpha) = \frac{1}{(2\alpha)^n} \int_{x_1-\alpha}^{x_1+\alpha} \cdots \int_{x_n-\alpha}^{x_n+\alpha} F(y_1, \ldots, y_n) \, dy_1 \ldots dy_n,$$

где $F(x) = \mathsf{M} f(x, \theta)$.

Из результатов гл. 2 легко заметить, что случайный вектор $\xi(x, \alpha)$, определяемый формулой

$$\xi(x, \alpha) = \frac{1}{2\alpha} \sum_{i=1}^{n} [f(\tilde{x}_1, \ldots, x_i + \alpha, \ldots, \tilde{x}_n, \theta^k) -$$

$$- f(\tilde{x}_1, \ldots . x_i - \alpha, \ldots, \tilde{x}_n, \theta^k)] \, e_i, \qquad (25.2)$$

удовлетворяет условию

$$\mathsf{M}(\xi(x, \alpha) \mid x) = \nabla F(x, \alpha),$$

где $\tilde{x}_i$ ($i = 1, \ldots, n$) — независимые случайные величины, равномерно распределенные на отрезках $[x_i - \alpha, x_i + \alpha]$, θ^k — независимые наблюдения θ.

По аналогии с формулами (4.7), (6.1) введем конечные разности следующего вида:

$$\sum_{i=1}^{n} \frac{f(\tilde{x} + \Delta e_i, \theta^k) - f(\tilde{x}, \theta^k)}{\Delta} e_i, \tag{25.3}$$

$$\sum_{i=1}^{p} \frac{f(\tilde{x} + \Delta \mu_i^k, \theta^k) - f(\tilde{x}, \theta^k)}{\Delta} \mu_i^k, \tag{25.4}$$

для которых

$$M(\xi(x, \alpha) \mid x) = \nabla F(x, \alpha) + b, \quad \| b \| \leqslant C\Delta/\alpha$$

Следовательно, стохастические конечно-разностные методы минимизации $F(x)$, обобщающие процедуру стохастической аппроксимации Кифера — Вольфовица (25.1), определяются соотношениями

$$x^{k+1} = x^k - \rho_k \xi(x^k, \alpha_k), \tag{25.5}$$

где вектор $\xi(x^k, \alpha_k)$ вычисляется по одной из формул (25.2) — (25.4).

Теорема 25.1. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < \infty, \quad \frac{\rho_k}{\alpha_k} \to 0, \quad \frac{\Delta_k}{\alpha_k} \to 0,$$

$$\alpha_k \to 0, \quad \frac{|\alpha_k - \alpha_{k+1}|}{\rho_k} \to 0, \quad ML^2(\theta) < \infty.$$

Тогда все предельные точки последовательности $\{x^k\}$ с вероятностью 1 принадлежат множеству $X^* \equiv \{x^* \mid 0 \in \partial F(x^*)\}$; последовательность $\{F(x^k)\}$ сходится с вероятностью 1.

Доказательство проводится аналогично данному для теоремы 4.1. Соотношение (4.10) не совсем очевидно. Из условия $ML^2(\theta) < \infty$ вытекает, что с вероятностью 1

$$\sum_{k=0}^{\infty} \rho_k (\xi(x^k, \alpha_k) - \nabla F(x^k, \alpha_k)) < \infty,$$

откуда с вероятностью 1

$$\rho_k (\xi(x^k, \alpha_k) - \nabla F(x^k, \alpha_k)) \to 0.$$

Аналогичный факт имеет место и для формул (25.3) — (25.4). Поэтому из соотношения

$$x^{k+1} = x^k - \rho_k \nabla F(x^k, \alpha_k) + \rho_k (\nabla F(x^k, \alpha_k) - \xi(x^k, \alpha_k))$$

следует (4.10). Вывод условий (4.12), (4.13) проводится так же, как и в теореме 4.1.

2. Игровая стохастическая задача. Метод стохастической аппроксимации (25.5) может быть использован для решения более сложных задач, когда классический метод (25.1) неприменим. Стохастические задачи игрового типа возникают при выборе оптимальных решений в условиях, когда на результат решения оказывают влияние несколько факторов, которые можно разбить на три группы: одна группа контролируется первым игроком, другая — вторым игроком, интересы которого противоположны интересам первого игрока, а третья — случайные переменные θ.

Пусть первый игрок принимает решение x из допустимого множества X, второй — решение y из допустимого множества Y, а состояние природы θ является элементарным событием некоторого вероятностного пространства. Предположим, что второй игрок делает свой выбор после первого игрока и ему известны выбор x первого игрока и состояние природы θ. Если при этом $f(x, y, \theta)$ — сумма, которую второй игрок получает от первого, когда природа находится в состоянии θ, то ожидаемый проигрыш первого игрока составит

$$F(x) = M \max_{y \in Y} f(x, y, \theta), \qquad x \in X. \tag{25.6}$$

Тогда первому игроку следует выбрать x, минимизирующее функцию $F(x)$ при условии $x \in X$.

Трудность решения этой задачи заключается в том, что только в редких случаях можно найти явный вид $F(x)$. Кроме того, функция $F(x)$ не будет, вообще говоря, непрерывно дифференцируемой даже при достаточно гладкой $f(x, y, \theta)$.

Предположим, что X, Y — выпуклые ограниченные замкнутые множества, и пусть при каждом x и θ существует такое $y(x, \theta)$, что

$$f(x, y(x, \theta), \theta) = \max_{y \in Y} f(x, y, \theta),$$

где $f(x, y, \theta)$ — выпуклая по x функция.

Тогда для минимизации $F(x)$ (25.6) применим метод

$$x^{k+1} = \pi_X (x^k - \rho_k \eta (x^k, \alpha_k)), \tag{25.7}$$

в котором вектор $\eta(x^k, \alpha_k)$ определяется по одной из следующих формул:

$$\frac{1}{2\alpha_k} \sum_{i=1}^{n} [f(\tilde{x}_1^k, \ldots, x_i^k + \alpha_k, \ldots, \tilde{x}_n^k, y(x^k, \theta^k), \theta^k) -$$

$$- f(\tilde{x}_1^k, \ldots, x_i^k - \alpha_k, \ldots, \tilde{x}_n^k, y(x^k, \theta^k), \theta^k)]\, e_i, \tag{25.8}$$

$$\sum_{i=1}^{n} \frac{f(\tilde{x}^k + \Delta_k e_i, y(x^k, \theta^k), \theta^k) - f(\tilde{x}^k, y(x^k, \theta^k), \theta^k)}{\Delta_k}\, e_i, \tag{25.9}$$

$$\sum_{i=1}^{p} \frac{f(\tilde{x}^k + \Delta_k \mu_i^k, y(x^k, \theta^k), \theta^k) - f(\tilde{x}^k, y(x^k, \theta^k), \theta^k)}{\Delta_k}\, e_i. \tag{25.10}$$

Теорема 25.2. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < \infty, \quad \frac{\Delta_k}{\alpha_k} \to 0, \quad \alpha_k \to 0, \quad \mathsf{M} L^2(\theta) < \infty.$$

Тогда с вероятностью 1 *предельные точки последовательности* $\{x^k\}$ *принадлежат множеству минимумов функции* (25.6).

Доказательство почти не отличается от доказательства теоремы 25.1. Следует учесть, что

$$\mathsf{M}(\eta(x^k, \alpha_k) \mid x^k, y(x^k, \theta^k), \theta^k) = \nabla f(x^k, y(x^k, \theta^k), \theta^k, \alpha_k) + b_k,$$

где $f(x, y(x, \theta), \theta, \alpha)$ — сглаженная по x функция вида (2.1), $\|b_k\| \leqslant \leqslant CL(\theta^k)\Delta_k/\alpha_k$. При исследовании сходимости метода основную роль играет неравенство

$$F(x) - F(x^k) \geqslant \mathsf{M}(g(x^k, y(x^k, \theta^k), \theta^k \mid x^k), x - x^k), \tag{25.11}$$

где $g(x^k, y(x^k, \theta^k), \theta^k) \in \partial f(x^k, y(x^k, \theta^k), \theta^k)$.

Очевидно, что

$$\begin{aligned} f(x, y(x, \theta^k), \theta^k) - f(x^k, y(x^k, \theta^k), \theta^k) \geqslant \\ \geqslant f(x, y(x^k, \theta^k), \theta^k) - f(x^k, y(x^k, \theta^k), \theta^k) \geqslant \\ \geqslant (g(x^k, y(x^k, \theta^k), \theta^k), \ x - x^k). \end{aligned}$$

Взяв условное математическое ожидание от обеих частей этого неравенства, получим (25.11).

3. **Скорость сходимости.** Пусть выполняются предположения, сформулированные в § 7, $F(x)$ — сильно выпуклая функция,

$$\mathsf{M}(F(x) - f(x, \theta))^2 \leqslant \sigma^2, \qquad x \in X,$$

X — выпуклый компакт, L — константа Липшица функции $F(x)$ на множестве $D \supset X$. Исследуем поведение следующего метода:

$$x^{k+1} = \pi_X(x^k - \rho_k \xi(x^k, \alpha_k)), \tag{25.12}$$

где $\xi(x^k, \alpha_k)$ определяется по формуле (25.2).

Теорема 25.3. *Пусть выполнены условия* (7.13) *и, кроме того,* $\beta - 4\mu > 0$. *Тогда для последовательности* (25.12) *имеет место оценка*

$$\varphi(k)\, \mathsf{M}\|x^k - x^*\|^2 \leqslant C,$$

где

$$\varphi(k) = \begin{cases} k^{\psi}, & \beta < 1 \quad \text{или} \quad \beta = 1, \quad 2\lambda b > \psi, \\ k^{\psi}/\ln k, & \beta = 1, \qquad 2b\lambda = \psi, \\ k^{2b\lambda}, & \beta = 1, \qquad 2b\lambda < \psi, \end{cases}$$

$$\psi = \min\{\mu, \beta - 4\mu\},$$

причем при $2b\lambda > \psi$, $\beta = 1$

$$\lim_{p\to\infty} \sup_{k\geqslant p} k^{\psi} \mathsf{M} \| x^k - x^* \|^2 \leqslant d\,(2b\lambda - \psi)^{-1},$$

где

$$d = \begin{cases} d_1 = 2Lb\sqrt{L\lambda^{-1} n \sqrt{n}}, & \text{если} \quad \mu < 1/5, \\ d_2 = n\,(\sigma b/m)^2, & \text{если} \quad \mu > 1/5, \\ d_3 = d_1 + d_2, & \text{если} \quad \mu = 1/5. \end{cases}$$

Скорость сходимости почти наверное метода характеризует

Теорема 25.4. *Пусть выполнены условия теоремы* 7.4, *а также* $\psi + \beta + 1 > 0$. *Тогда*

$$\mathsf{P}\,\{\sup_{t\geqslant k} t^{\nu} \| x^t - x^* \| \geqslant a\} \leqslant C\,(k^{\nu}/a)^2\,[\varphi_1(k)]^{-1}, \quad a > 0,$$

где

$$0 \leqslant \nu < \min\left\{\frac{\psi + \beta - 1}{2},\, b\lambda\right\},$$

$$\varphi_1(k) = \begin{cases} \varphi(k), & \text{если} \quad \beta = 1, \\ k^{\psi+\beta-1}, & \text{если} \quad \beta < 1. \end{cases}$$

Из теорем 25.3, 25.4 следует, что если $\beta = 1$, $10\, b\lambda L > 3$, $\mu = 1/5$, то как среднеквадратическая скорость сходимости, так и скорость сходимости почти наверное являются наибольшими. Тогда

$$\lim_{t\to\infty} \sup_{k\geqslant t} k^{1/5} \mathsf{M} \| x^k - x^* \|^2 \leqslant d_3 \left(\frac{2b\lambda L}{3} - \frac{1}{5}\right)^{-1}.$$

В том случае, когда направление спуска выбирается по формуле (25.4), среднеквадратичное отклонение k-го приближения от точки минимума при оптимально выбранных параметрах имеет порядок $k^{-1/6}$.

Теоремы 7.1 — 7.4, 25.3, 25.4 показывают, что случайные помехи в целевой функции в значительной степени влияют на скорость сходимости конечно-разностных методов.

§ 26. Операция усреднения

Операция усреднения случайных функций $f(x, \theta)$, градиентов, а также конечно-разностных аппроксимаций оказывается очень полезной при решении самых разнообразных задач стохастического программирования. Как правило, структура методов, изучаемых в этой книге, следующая: строится некоторая последовательность точек $x^{k+1} = x^k + \rho_k d(x^k)$, удовлетворяющих условию $x^k \in X$, где X — ограниченное множество в E_n, а вектор $d(x^k)$ (в общем случае случайный) задает направление движения из точки x^k.

Рассмотрим последовательность

$$z^{k+1} = z^k + a_k\,(f(x^k, \theta^k) - z^k), \tag{26.1}$$

где z^0 — начальное приближение, a_k — положительные множители, θ^k — независимые наблюдения параметра θ. Заметим, что если $a_k = 1/(k+1)$, $z^0 = 0$, то

$$z^{k+1} = \frac{f(x^0, \theta^0) + \ldots + f(x^k, \theta^k)}{k+1};$$

поэтому можно говорить, что z^k получается с помощью операции усреднения. Различие между операцией усреднения (26.1) и обычным усреднением, проводимым в пределах закона больших чисел, заключается в том, что распределение вероятностей случайной величины $f(x, \theta)$ зависит от многомерного параметра x, который изменяется в процессе поиска минимума $F(x)$.

Вначале покажем, что если случайная величина $f(x, \theta)$ ограничена для всех x, θ, т. е. $|f(x, \theta)| \leqslant C$ (C — некоторая константа), то усредненные оценки z^k ($k = 0, 1, \ldots$) также равномерно ограничены. С практической точки зрения это требование не является существенным, поскольку при реализации численных методов на ЭВМ оно выполняется автоматически. Несколько позже условие ограниченности $f(x, \theta)$ будет опущено.

Лемма 26.1. *Предположим, что*

$$|f(x, \theta)| \leqslant C, \qquad \sum_{k=0}^{\infty} a_k = \infty, \quad a_k \to 0.$$

Тогда в итерационной процедуре (26.1) *величины z^k равномерно ограничены.*

Доказательство. Не ограничивая общности, считаем $a_k < 1$; поэтому

$$|z^{k+1}| \leqslant |z^k| (1 - a_k) + Ca_k.$$

Отсюда, полагая по определению

$$\prod_{l=k+1}^{k} (1 - a_l) = 1,$$

получаем неравенство

$$|z^{k+1}| \leqslant |z^0| \prod_{s=0}^{k} (1 - a_s) + C \sum_{s=0}^{k} a_s \prod_{l=s+1}^{k} (1 - a_l). \tag{26.2}$$

Покажем, что первый член стремится к нулю при $k \to \infty$. Из разложения

$$e^{-a_k} = 1 - \frac{a_k}{1!} + \frac{a_k^2}{2!} - \frac{a_k^3}{3!} + \ldots$$

следует, что

$$e^{-a_k} \geqslant 1 - a_k.$$

Поэтому

$$\varphi \equiv \exp\left\{-\sum_{s=0}^{k} a_s\right\} \geqslant \prod_{s=0}^{k} (1-a_s).$$

Величина φ стремится к нулю, так как $\sum\limits_{k=0}^{\infty} a_k = \infty$; значит, $\prod\limits_{s=0}^{k} (1-a_s) \to 0.$

Представим второй член суммы (26.2) в виде

$$\sum_{s=0}^{k} a_s \prod_{l=s+1}^{k} (1-a_l) = \sum_{s=0}^{k} \left(\prod_{l=s+1}^{k} (1-a_l) - \prod_{l=s}^{k} (1-a_l) \right).$$

Отсюда

$$\sum_{s=0}^{k} a_s \prod_{l=s+1}^{k} (1-a_l) = 1 - \prod_{l=0}^{k} (1-a_l) < 1.$$

Лемма доказана.

Если не предполагать ограниченности $f(x, \theta)$, то усредненные оценки z^k могут принимать сколь угодно большие значения. В этом случае z^k определяются по формуле

$$z^{k+1} = \pi_z (z^k + a_k (f(x^k, \theta^k) - z^k)),$$

где z — ограниченное множество. В дальнейшем эта ситуация особо не оговаривается, и для простоты обозначений считаем, что оценки z^{k+1} определяются по формуле (26.1). Сформулируем основное утверждение этого параграфа. Обозначим через B_k σ-алгебру, индуцированную случайными векторами $x^0, z^0, \ldots, x^k, z^k$.

Теорема 26.1. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty} a_k = \infty, \quad \sum_{k=0}^{\infty} a_k^2 < \infty, \quad \frac{\rho_k}{a_k} \to 0,$$

$$\mathsf{M} \| s(x^k) \|^2 < \infty, \quad \mathsf{M} f^2(x, \theta) < \infty,$$

и пусть функция $F(x) = \mathsf{M} f(x, \theta)$ удовлетворяет условию Липшица в области X. Тогда предел последовательности $\{z^k - F(x^k)\}$ с вероятностью 1 равен нулю.

Доказательство проводим от противного, используя схему, описанную в §4. Сначала покажем, что с вероятностью 1

$$|z^{k+1} - z^k| \to 0.$$

Заметим, что

$$\mathsf{M}(f(x^k, \theta^k) \mid B_k) = F(x^k);$$

поэтому из условий теоремы с вероятностью 1

$$\sum_{k=0}^{\infty} a_k (F(x^k) - f(x^k, \theta^k)) < \infty.$$

Следовательно, из необходимого условия сходимости рядов вытекает, что

$$a_k (f(x^k, \theta^k) - F(x^k)) \to 0.$$

Поэтому из соотношений

$$z^{k+1} = z^k + a_k (f(x^k, \theta^k) - z^k) =$$
$$= z^k + a_k (F(x^k) - z^k) + a_k (f(x^k, \theta^k) - F(x^k))$$

вытекает, что с вероятностью 1

$$|z^{k+1} - z^k| \to 0.$$

Предположим, что существует сходящаяся подпоследовательность

$$(z^s(\omega) - F(x^s(\omega))) \to d \neq 0.$$

Можно указать такие положительные числа $\bar{s}$ и $\bar{\delta}$, что $2\bar{\delta}$-окрестности точек $(z^s - F(x^s))$, $s \leqslant \bar{s}$, не содержат нуля. Покажем, что $k(s) < \infty$, где

$$k(s) = \min \{r \,|\, |(z^r - F(x^r)) - (z^s - F(x^s))| > \delta, \ r > s\},$$
$$\delta \leqslant \bar{\delta}/2.$$

Выведем неравенство, из которого будет следовать этот результат. Все ограниченные величины обозначим буквой C. Опуская некоторые выкладки, получим

$$(z^{k+1} - F(x^{k+1}))^2 \leqslant$$
$$\leqslant (z^k - F(x^k) + F(x^k) - F(x^{k+1}) + a_k (f(x^k, \theta^k) - z^k))^2 \leqslant$$
$$\leqslant (z^k - F(x^k))^2 + 2(F(x^k) - F(x^{k+1}))^2 + 2a_k^2 (f(x^k, \theta^k) - z^k)^2 +$$
$$+ 2(z^k - F(x^k))(F(x^k) - F(x^{k+1})) +$$
$$+ 2a_k (z^k - F(x^k))(f(x^k, \theta^k) - z^k) \leqslant (z^k - F(x^k))^2 +$$
$$+ C\rho_k^2 + Ca_k^2 + C\rho_k + T_k - 2a_k (z^k - F(x^k))^2, \tag{26.3}$$

где

$$T_k = C\rho_k^2 (\| d(x^k) \|^2 - \mathrm{M} \| d(x^k) \|^2 | B_k) + 4a_k^2 (f^2(x^k, \theta^k) -$$
$$- \mathrm{M} f^2(x^k, \theta^k) | B_k) + C\rho_k (\| d(x^k) \| - \mathrm{M} \| d(x^k) \| | B_k) +$$
$$+ 2a_k (z^k - F(x^k))(f(x^k, \theta^k) - F(x^k)).$$

Заметим, что из неравенства

$$\mathrm{M} \| d(x^k) \|^2 | B_k < \infty$$

вытекает

$$\mathrm{M} \| d(x^k) \| | B_k < \infty,$$

поскольку

$$\| d(x^k) \| \leqslant 1 + \frac{1}{4} \| d(x^k) \|^2.$$

Из условий теоремы $\sum_{k=0}^{\infty} T_k < \infty$ с вероятностью 1. Согласно предположению, для точек $z^k - F(x^k)$, принадлежащих δ-окрестности точки $z^s - F(x^s)$, $k \geqslant \bar{s}$, выполняется соотношение

$$(z^k - F(x^k))^2 > \bar{\delta}^2.$$

По условию теоремы сумма $C(\rho_k^2 + a_k^2 + \rho_k)$, начиная с некоторого номера $\bar{s}$ $(k \geqslant \bar{s})$, не превосходит $\bar{\delta}^2 a_k/2$. Поэтому, просуммировав неравенство (26.3), получим

$$(z^k - F(x^k))^2 \leqslant (z^s - F(x^s))^2 + \sum_{r=s}^{k-1} T_r - \sum_{r=s}^{k-1} a_r. \qquad (26.4)$$

Устремив k к бесконечности, получим противоречие с условиями ограниченности $F(x)$ в области X и ограниченности z^k. Следовательно, найдется конечный номер

$$k(s) \equiv \min \{ r \mid |(z^r - F(x^r)) - (z^s - F(x^s))| > \delta, \ r > s \}.$$

Из соотношений

$$|z^{k(s)} - z^s| \leqslant \left| \sum_{k=s}^{k(s)-1} [a_k (F(x^k) - z^k) + a_k (f x^k, \theta^k) - F(x^k))] \right|,$$

$$|F(x^{k(s)}) - F(x^s)| \leqslant C \sum_{k=s}^{k(s)} \rho_k \| d(x^k) \|,$$

а также из сходимости рядов

$$\sum_{k=0}^{\infty} \rho_k (\| d(x^k) \| - M \| d(x^k) \| | B_k),$$

$$\sum_{k=0}^{\infty} a_k (f(x^k, \theta^k) - F(x^k))$$

вытекает, что для достаточно больших номеров s

$$C \sum_{k=s}^{k(s)-1} a_k > \frac{\delta}{2}.$$

По построению

$$|z^{k(s)} - F(x^{k(s)}) - z^s + F(x^s)| > \delta,$$

но для достаточно больших s

$$|z^{k(s)}-F(x^{k(s)})-z^s+F(x^s)|<2\delta,$$

а также

$$\sum_{k=s}^{k(s)-1} T_k<\frac{\bar{\delta}\delta}{4C};$$

поэтому все рассуждения, проведенные при выводе неравенства (26.4), остаются справедливыми и при $k = k(s)$. Таким образом,

$$(z^{k(s)}-F(x^{k(s)}))^2\leqslant(z^s-F(x^s))^2-\bar{\delta}^2\delta/(4C). \qquad (26.5)$$

Перейдя к пределу в (26.5), получим

$$\overline{\lim_{s\to\infty}}\,(z^{k(s)}-F(x^{k(s)}))^2\leqslant d^2-\bar{\delta}^2\delta/(4C).$$

Для положительных чисел a, b таких, что

$$\overline{\lim_{s\to\infty}}\,(z^{k(s)}-F(x^{k(s)}))^2<a<b<d^2,$$

последовательность $(z^k-F(x^k))^2$ пересекает интервал (a, b) слева направо бесконечно много раз. Поэтому существует последовательность пар $\{x^r\}$ $(r\in R)$, $\{x^p\}$ $(p\in P)$, для которой

$$(z^r-F(x^r))^2\leqslant a,\qquad (z^p-F(x^p))^2\geqslant b,$$

$$a<(z^k-F(x^k))^2<b,\qquad r<k<p.$$

С подпоследовательностью $\{(z^r-F(x^r))^2\}$ поступим так же, как и с $\{(z^s-F(x^s))^2\}$ $(s\in S)$. Пусть

$$k(r)=\min\{t\,|\,|(z^t-F(x^t))-(z^r-F(x^r))|>\delta,\ t>r\}.$$

Для достаточно больших r и достаточно малых δ справедливо соотношение $r<k(r)<p$; поэтому

$$(z^{k(r)}-F(x^{k(r)}))^2>(z^r-F(x^r))^2,$$

что противоречит неравенству (26.5), записанному для подпоследовательности $\{(z^r-F(x^r)\}$. Полученное противоречие доказывает теорему.

Результат, аналогичный теореме 26.1, справедлив и для операции усреднения $\nabla f(x, \theta)$. Пусть

$$y^{k+1}=y^k+a_k(\xi(k)-y^k),$$

$$M\xi(k)\,|\,B_k=\nabla F(x^k)+b_k,$$

где b_k — некоторая случайная помеха, появляющаяся, например, при рассмотрении конечно-разностной аппроксимации

$$\xi(k)=\sum_{i=1}^{k}\frac{f(x^k+\alpha_k e_i,\theta^k)-f(x^k)}{\alpha_k}\,e_i.$$

Если функции $f(x, \theta)$ непрерывно дифференцируемы, то вектор $\xi(k)$ можно положить равным $\nabla f(x^k, \theta^k)$, θ^k — независимые наблюдения параметра θ.

Теорема 26.2. *Пусть выполнены условия теоремы* 26.1,

$$\mathsf{M}\|b_k\| \leqslant r_k, \qquad r_k \to 0,$$

$$\mathsf{M}\|b_k\|^2 < \infty, \qquad \mathsf{M}\|\xi(k)\|^2 < \infty,$$

градиент функции $F(x)$ удовлетворяет в X условию Липшица. Тогда с вероятностью 1

$$(y^k - \nabla F(x^k)) \to 0, \qquad k \to \infty.$$

Доказательство этой теоремы с незначительными изменениями проводится так же, как и в теореме 26.1.

В задачах минимизации функции $F(x) = \mathsf{M}f(x, \theta)$, где $f(x, \theta)$ удовлетворяют локальному условию Липшица с константой $L(\theta)$, операция усреднения

$$v^{k+1} = v^k + a_k(\xi(x^k, \alpha_k) - v^k)$$

характеризуется тем, что с вероятностью 1

$$(v^k - \nabla F(x^k, \alpha_k)) \to 0, \qquad k \to \infty.$$

Здесь $F(x, \alpha)$ — сглаженная функция, вектор $\xi(x^k, \alpha_k)$ определяется по формулам (25.2) — (25.4). Если $f(x, \theta)$ — выпуклая по x функция, то

$$\xi(x^k, \alpha_k) = g(\tilde{x}^k, \theta^k), \qquad g(\tilde{x}^k, \theta^k) \in \partial f(\tilde{x}^k, \theta^k). \tag{26.6}$$

Теорема 26.3. *Пусть выполнены условия теоремы* 26.1,

$$\sum_{k=0}^{\infty}\left(\frac{\rho_k}{\alpha_k}\right)^2 < \infty, \quad \frac{\rho_k}{\alpha_k a_k} \to 0, \quad \alpha_k \to 0,$$

$$\frac{\Delta_k}{\alpha_k} \to 0, \quad \frac{|\alpha_k - \alpha_{k+1}|}{\alpha_k a_k} \to 0, \quad \mathsf{M}L^2(\theta) < \infty.$$

Тогда с вероятностью 1

$$(v^k - \nabla F(x^k, \alpha_k)) \to 0, \qquad k \to \infty.$$

Доказательство. Для простоты записи считаем, что

$$\mathsf{M}\xi(x^k, \alpha_k) \mid B_k = \nabla F(x^k, \alpha_k).$$

Выполняется следующая цепочка неравенств:

$$\|v^{k+1} - \nabla F(x^{k+1}, \alpha_{k+1})\|^2 = \|v^k - \nabla F(x^k, \alpha_k) + a_k(\xi(x^k, \alpha_k) - v^k) +$$
$$+ (\nabla F(x^k, \alpha_k) - \nabla F(x^{k+1}, \alpha_k)) + (\nabla F(x^{k+1}, \alpha_k) - \nabla F(x^{k+1}, \alpha_{k+1}))\|^2 \leqslant$$

$$\leqslant \| v^k - \nabla F(x^k, \alpha_k)\|^2 + 2a_k (v^k - \nabla F(x^k, \alpha_k), \xi(x^k, \alpha_k) \pm$$

$$\pm \nabla F(x^k, \alpha_k) - v^k) + 2(v^k - \nabla F(x^k, \alpha_k), \nabla F(x^k, \alpha_k) - \nabla F(x^{k+1}, \alpha_k)) +$$

$$+ 2(v^k - \nabla F(x^k, \alpha_k), \nabla F(x^{k+1}, \alpha_k) - \nabla F(x^{k+1}, \alpha_{k+1})) +$$

$$+ 3a_k^2 \| \xi(x^k, \alpha_k) - v^k\|^2 + 3\|\nabla F(x^k, \alpha_k) - \nabla F(x^{k+1}, \alpha_k)\|^2 +$$

$$+ 3\|\nabla F(x^{k+1}, \alpha_k) - \nabla F(x^{k+1}, \alpha_{k+1})\|^2 \leqslant \| v^k - \nabla F(x^k, \alpha_k)\|^2 -$$

$$- 2a_k \| v^k - \nabla F(x^k, \alpha_k)\|^2 + 2a_k (v^k - \nabla F(x^k, \alpha_k),$$

$$\xi(x^k, \alpha_k) - \nabla F(x^k, \alpha_k)) + C\frac{\rho_k}{\alpha_k}\|d(x^k)\| + \frac{C|\alpha_k - \alpha_{k+1}|}{\alpha_k} +$$

$$+ Ca_k^2\|\xi(x^k, \alpha_k)\|^2 + Ca_k^2 + C\left(\frac{\rho_k}{\alpha_k}\right)^2\|d(x^k)\|^2 + C\frac{|\alpha_k - \alpha_{k+1}|^2}{\alpha_k^2} \leqslant$$

$$\leqslant \| v^k - \nabla F(x^k, \alpha_k)\|^2 - 2a_k\| v^k - \nabla F(x^k, \alpha_k)\|^2 +$$

$$+ C\frac{\rho_k}{\alpha_k} + C\frac{|\alpha_k - \alpha_{k+1}|}{\alpha_k} + Ca_k^2 + C\frac{\rho_k^2}{\alpha_k^2} + C\frac{|\alpha_k - \alpha_{k+1}|^2}{\alpha_k^2} + T_k, \quad (26.7)$$

где

$$T_k = 2a_k (v^k - \nabla F(x^k, \alpha_k), \ \xi(x^k, \alpha_k) - \nabla F(x^k, \alpha_k)) +$$

$$+ C\frac{\rho_k}{\alpha_k}(\|d(x^k)\| - \mathsf{M}\|d(x^k)\|\,|B_k) + Ca_k^2(\|\xi(x^k, \alpha_k)\|^2 -$$

$$- \mathsf{M}\|\xi(x^k, \alpha_k)\|^2|B_k) + C\frac{\rho_k^2}{\alpha_k^2}(\|d(x^k)\|^2 - \mathsf{M}\|d(x^k)\|^2|B_k).$$

Из условий теоремы вытекает, что $\sum\limits_{k=0}^{\infty} T_k < \infty$. При выводе (26.7) использовалось неравенство

$$(c_1 + c_2 + \ldots + c_n)^2 \leqslant n(c_1^2 + \ldots + c_n^2).$$

В дальнейшем доказательство проводится аналогично теореме 26.1

§ 27. Стохастическая оптимизация на основе операции усреднения

В этом параграфе исследуются численные методы минимизации функции

$$F(x) = \mathsf{M}f(x, \theta) \quad (27.1)$$

при условии

$$x \in X, \quad (27.2)$$

где X — ограниченное множество из E_n.

1. Методы с усреднением направлений спуска. Рассмотрим вначале задачу минимизации гладкой выпуклой функции $F(x)$ (27.1) при условии (27.2), где X — выпуклое множество.

Метод определяется рекуррентными соотношениями

$$x^{k+1} = \pi_X (x^k - \rho_k y^k), \tag{27.3}$$

$$y^{k+1} = y^k + a_k (\xi(k) - y^k), \tag{27.4}$$

где

$$M\xi(k)/B_k = \nabla F(x^k) + b_k,$$

$\pi_X(x)$ — оператор проектирования точки x на множество X. Если функция $f(x, \theta)$ при каждом θ непрерывно дифференцируема по x, то в качестве $\xi(k)$ можно взять $\xi(k) = \nabla f(x^k, \theta^k)$, где θ^k — независимые наблюдения параметра θ.

Целесообразность применения операции усреднения обусловлена тем, что необходимо уменьшить влияние помех на процесс поиска, сделать его более регулярным. Особенность метода (27.3), (27.4) состоит в том, что при достаточно большом номере k вектор, указывающий направление движения, близок к антиградиенту функции цели. Однако это преимущество сказывается на начальной стадии минимизации для достаточно далекого начального приближения x^0. Операция усреднения y^k включает всю старую информацию о случайных градиентах, вычисленных на предыдущих итерациях; иными словами, процедура y^k обладает инерционностью, которая ухудшает свойства алгоритма в окрестности решения. В работе [100] показано, что вблизи решения методы с усреднением направлений спуска уступают в скорости сходимости одношаговым процедурам, например стохастическому квазиградиентному методу.

Теорема 27.1. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} a_k^2 < \infty, \quad \frac{\rho_k}{a_k} \to 0,$$

$$M\|\xi(k)\|^2 < \infty, \quad M\|b_k\|^2 < \infty, \quad M\|b_k\| \leqslant r_k,$$

$$r_k \to 0,$$

$\nabla F(x)$ *удовлетворяет в* X *условию Липшица. Тогда с вероятностью* 1 *предельные точки последовательности* x^k *принадлежат множеству* X^* *решений задачи* (27.1), (27.2) *и с вероятностью* 1

$$F(x^k) \to F(x^*), \qquad x^* \in X^*.$$

Доказательство. Достаточно показать, что выполняются условия (4.12), (4.13). Предположим, что существует подпоследовательность $x^s \to x' \notin X^*$ $(s \in S)$. Можно указать такие положительные числа $\bar{s}$ и $\bar{\delta}$, что $2\bar{\delta}$-окрестности точек x^s $(s \geqslant \bar{s})$ не пересекаются с X^*. Построим в каждой точке x^s $(s \geqslant \bar{s})$ окрестности радиуса $\delta \leqslant \bar{\delta}/2$. Далее для удобства все ограниченные величины обозначаем буквой C.

Справедливо неравенство

$$\| x^{k+1} - x^* \|^2 \leqslant \| x^k - \rho_k y^k - x^* \|^2 = \| x^k - x^* \|^2 -$$
$$- 2\rho_k (y^k, x^k - x^*) + \rho_k^2 \| y^k \|^2 \leqslant \| x^k - x^* \|^2 -$$
$$- 2\rho_k (y^k - \nabla F(x^k),\ x^k - x^*) + 2\rho_k (\nabla F(x^k), x^* - x^k), \qquad (27.5)$$

где $x^* \in X^*$, точка x^k принадлежит δ-окрестности x^s $(s \geqslant \bar{s})$. Тогда найдется такое $\varepsilon > 0$, что

$$F(x^k) - F(x^*) \geqslant \varepsilon,$$
$$(\nabla F(x^k), x^* - x^k) \leqslant F(x^*) - F(x^k) \leqslant -\varepsilon.$$

Из теоремы 26.2 вытекает, что величина

$$-2(y^k - \nabla F(x^k), x^* - x^k) + C\rho_k^2,$$

начиная с некоторого номера k, не превосходит ε. Поэтому из (27.5) следует, что

$$\| x^{k+1} - x^* \|^2 \leqslant \| x^k - x^* \|^2 - \varepsilon\rho_k.$$

Положим

$$W(x^k) = \min_{x^* \in X^*} \| x^k - x^* \|^2 = \| x^k - x^*(k) \|^2.$$

Тогда из предыдущего неравенства имеем

$$W(x^{k+1}) \leqslant \| x^{k+1} - x^*(k) \|^2 \leqslant W(x^k) - \varepsilon\rho_k. \qquad (27.6)$$

Просуммировав неравенство (27.6), получим

$$W(x^k) \leqslant W(x^s) - \varepsilon \sum_{r=s}^{k-1} \rho_r, \qquad (27.7)$$

что при $k \to \infty$ противоречит ограниченности $W(x)$ на замкнутом ограниченном множестве $U_\delta(x^s)$. Следовательно, выполняется условие (4.12), т. е. найдется конечный номер $k(s)$ такой, что

$$k(s) = \min_{r > s} \{ r \mid \| x^r - x^s \| > \delta \}.$$

Неравенство (27.7) справедливо и для $k = k(s)$. Из соотношения

$$\delta \leqslant \| x^{k(s)} - x^s \| \leqslant \sum_{r=s}^{k(s)-1} \rho_r \| y^r \|$$

вытекает

$$\sum_{r=s}^{k(s)-1} \rho_r > \frac{\delta}{C},$$

что при подстановке в (27.7) дает

$$W(x^{k(s)}) \leqslant W(x^s) - \varepsilon\delta/C,$$

а это не что иное, как (4.13). Теорема доказана.

Если $f(x, \theta)$ — выпуклая при каждом θ функция, то метод решения задачи (27.1), (27.2) определяется соотношениями

$$x^{k+1} = \pi_X (x^k - \rho_k v^k), \qquad (27.8)$$

$$v^{k+1} = v^k + a_k (\xi (x^k, \alpha_k) - v^k), \qquad (27.9)$$

вектор $\xi (x^k, \alpha_k)$ вычисляется по формулам (25.2) — (25.4), (26.6).

Теорема 27.2. *Пусть выполнены условия теоремы* 26.3, *и пусть*

$$\sum_{k=0}^{\infty} \rho_k = \infty.$$

Тогда с вероятностью 1 *предельные точки последовательности* x^k *принадлежат множеству* X^* *решений задачи* (27.1), (27.2) *и с вероятностью* 1

$$F(x^k) \to F(x^*).$$

Доказательство. Справедливы соотношения

$$\| x^{k+1} - x^* \|^2 \leqslant \| x^k - \rho_k v^k - x^* \|^2 =$$
$$= \| x^k - x^* \|^2 - 2\rho_k (v^k, x^k - x^*) + \rho_k^2 \| v^k \|^2 \leqslant$$
$$\leqslant \| x^k - x^* \|^2 - 2\rho_k (v^k - \nabla F (x^k, \alpha_k), x^k - x^*) -$$
$$- 2 (\nabla F (x^k, \alpha_k), x^k - x^*) + C\rho_k^2,$$

где $F(x, \alpha)$ —сглаженная функция. Далее доказательство проводится аналогично доказательству теоремы 27.1 в силу того, что с вероятностью 1

$$(v^k - \nabla F (x^k, \alpha_k)) \to 0.$$

Если применить операцию усреднения для безусловной минимизации функции $F(x) = \mathsf{M} f(x, \theta)$, где $f(x, \theta)$ — липшицева функция, то получим метод

$$x^{k+1} = x^k - \rho_k v^k, \qquad (27.10)$$

$$v^{k+1} = v^k + a_k (\xi (x^k, \alpha_k) - v^k). \qquad (27.11)$$

В соотношениях (27.11) векторы $\xi (x^k, \alpha_k)$ определяются по формулам (26.2) — (26.4).

Теорема 27.3. *Пусть выполнены условия теоремы* 26.3, *и пусть* $\sum_{k=0}^{\infty} \rho_k = \infty$. *Тогда с вероятностью* 1 *предельные точки последовательности* x^k *принадлежат множеству*

$$X^* = \{x^* \mid 0 \in \partial F (x^*)\},$$

$F(x^k)$ *с вероятностью* 1 *сходится.*

Доказательство сходимости метода (27.10), (27.11) основано на том, что с вероятностью 1

$$(v^k - \nabla F(x^k, \alpha_k)) \to 0.$$

Оно вытекает из неравенств

$$F(x^{k+1}, \alpha_k) = F(x^k, \alpha_k) + (\nabla F(x^k + \tau(x^{k+1} - x^k), \alpha_k), x^{k+1} - x^k) =$$
$$= F(x^k, \alpha_k) + (\nabla F(x^k + \tau(x^{k+1} - x^k), \alpha_k) \pm$$
$$\pm \nabla F(x^k, \alpha_k), x^{k+1} - x^k) \leqslant F(x^k, \alpha_k) +$$
$$+ C\rho_k^2/\alpha_k - \rho_k(\nabla F(x^k, \alpha_k), v^k), \quad 0 \leqslant \tau \leqslant 1,$$

а также из доказательства теоремы 4.1.

2. Стохастические методы условного градиента. Операция проектирования точки y на множество X равносильна решению экстремальной задачи (с квадратичной функцией цели) о минимизации $\| y - x \|^2$ при $x \in X$. Метод условного градиента отличается от предыдущих алгоритмов тем, что вместо операции проектирования рассматривается операция линеаризации, равносильная решению задачи с линейной функцией цели. Поэтому данный метод эффективно применим в том случае, когда трудно осуществить операцию проектирования.

Рассмотрим метод минимизации непрерывно дифференцируемой функции $F(x) = \mathsf{M} f(x, \theta)$ при $x \in X$ в предположении, что имеется возможность наблюдать случайный вектор $\xi(k)$, для которого

$$\mathsf{M}\xi(k) \mid B_k = \nabla F(x^k) + b_k.$$

Определим последовательности точек

$$x^{k+1} = x^k + \rho_k(\bar{x}^k - x^k), \qquad 0 \leqslant \rho_k \leqslant 1, \tag{27.12}$$

$$y^{k+1} = y^k + a_k(\xi(k) - y^k), \tag{27.13}$$

где $\bar{x}^k$ — решение задачи

$$\min_{x \in X}(y^k, x). \tag{27.14}$$

Простые примеры показывают, что метод, у которого вектор $\bar{x}^k$ есть решение задачи

$$\min_{x \in X}(\xi(k), x),$$

может не сходиться к множеству решений задачи (27.1), (27.2).

Теорема 27.4. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty} \rho_k = 0, \quad \sum_{k=0}^{\infty} a_k^2 < \infty, \quad \frac{\rho_k}{a_k} \to 0,$$

$$\mathsf{M}\| \xi(k) \|^2 < \infty, \quad \mathsf{M}\| b_k \|^2 < \infty, \quad \mathsf{M}\| b_k \| \leqslant r_k, \quad r_k \to 0,$$

градиент функции $F(x)$ удовлетворяет в X условию Липшица. Тогда с вероятностью 1 предельные точки последовательности x^k (27.12) — (27.14) принадлежат множеству X^*, удовлетворяющему необходимым условиям экстремума задачи (27.1), (27.2), *последовательность $F(x^k)$ сходится с вероятностью* 1.

Доказательство во многом совпадает с доказательством метода условного градиента, рассмотренного в § 15. Из теоремы о среднем имеем

$$F(x^{k+1}) = F(x^k) + (\nabla F(x^k + \tau(x^{k+1} - x^k), x^{k+1} - x^k)) =$$

$$= F(x^k) + \rho_k (\nabla F(x^k + \tau(x^{k+1} - x^k)) - \nabla F(x^k), \bar{x}^k - x^k) +$$

$$+ \rho_k (\nabla F(x^k), \bar{x}^k - x^k) \leqslant F(x^k) + C\rho_k^2 + \rho_k (\nabla F(x^k), \bar{x}^k - x^k), \quad (27.15)$$

$$0 \leqslant \tau \leqslant 1.$$

В неравенстве (27.15) нужно оценить величину $\rho_k (\nabla F(x^k), \bar{x}^k - x^k)$. Для этого рассмотрим вспомогательную задачу минимизации по x функции $(\nabla F(z), x)$ при условиях $x \in X$, $z \in X$. Пусть $\bar{z}$ — решение данной задачи. Для точек z, принадлежащих достаточно малой окрестности точки $z' \notin X^*$, справедливо неравенство

$$(\nabla F(z), \bar{z} - z) \leqslant \sigma < 0. \quad (27.16)$$

Предположим, что $x^s \to x' \notin X^*$ $(s \in S)$. Пусть условие (4.12) не выполняется, т. е. все точки x^k $(k \geqslant s)$ содержатся в достаточно малой δ-окрестности точки x^s, не пересекающейся с X^*. Из неравенства (27.16) имеем

$$(\nabla F(x^k), \tilde{x}^k - x^k) \leqslant \sigma < 0,$$

где $\tilde{x}^k$—решение предыдущей вспомогательной задачи. Так как с вероятностью 1

$$(y^k - \nabla F(x^k)) \to 0,$$

то для достаточно больших $k \geqslant s$

$$(y^k, \tilde{x}^k - x^k) \leqslant \sigma/2.$$

Следовательно,

$$(y^k, \bar{x}^k - x^k) \leqslant \sigma/2,$$

где $\bar{x}^k$ — решение задачи (27.14). Отсюда в силу предыдущего замечания получаем

$$(\nabla F(x^k), \bar{x}^k - x^k) \leqslant \sigma/4.$$

Суммируя неравенство (27.16) и учитывая, что $\rho_k \to 0$, для достаточно больших s имеем

$$F(x^k) \leqslant F(x^s) + \frac{\sigma}{8} \sum_{r=s}^{k-1} \rho_r. \quad (27.17)$$

Поэтому, переходя к пределу по $k \to \infty$, получим противоречие с ограниченностью $F(x)$. Таким образом, выполняется условие (4.12), т. е. существует

$$k(s) = \min_{r > s} \{r \mid \| x^r - x^s \| > \delta\}.$$

Из неравенства (27.17) непосредственно следует, что

$$\overline{\lim_{s \to \infty}} F(x^{k(s)}) < \lim_{s \to \infty} F(x^s),$$

а это означает, что справедливо условие (4.13). Теорема доказана.

Обсудим теперь задачу минимизации функции

$$F(x) = \mathsf{M} f(x, \theta), \qquad x \in X,$$

где $f(x, \theta)$ удовлетворяет в $D \supset X$ условию Липшица с константой $L(\theta)$. Рассматриваемый метод является стохастическим вариантом метода условного градиента, изложенного в § 15. Определяются последовательности точек

$$x^{k+1} = x^k + \rho_k (\bar{x}^k - x^k), \tag{27.18}$$

$$v^{k+1} = v^k + a_k (\xi(x^k, \alpha_k) - v^k), \tag{27.19}$$

$$(v^k, \bar{x}^k) = \min_{x \in X} (v^k, x). \tag{27.20}$$

Здесь $\xi(x^k, \alpha_k)$ — случайные векторы, вычисляемые по формулам (25.2) — (25.4). Если $f(x, \theta)$ — выпуклая по x функция, то

$$\xi(x^k, \alpha_k) = g(\tilde{x}^k, \theta^k), \qquad g(\tilde{x}^k, \theta^k) \in \partial f(\tilde{x}^k, \theta^k).$$

Теорема 27.5. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} a_k^2 < \infty, \quad \sum_{k=0}^{\infty} \left(\frac{\rho_k}{\alpha_k}\right)^2 < \infty, \quad \alpha_k \to 0,$$

$$\frac{\rho_k}{\alpha_k a_k} \to 0, \quad \frac{|\alpha_k - \alpha_{k+1}|}{\alpha_k a_k} \to 0, \quad \mathsf{M} L^2(\theta) < \infty.$$

Тогда $F(x^k)$ *сходится с вероятностью* 1 *и предельные точки последовательности* x^k (27.18) — (27.20) *с вероятностью* 1 *принадлежат множеству* X^*, *которое удовлетворяет необходимым условиям экстремума задачи* (27.1), (27.2).

Доказательство полностью совпадает с доказательством теоремы 15.1.

3. **Скорость сходимости стохастического метода условного градиента.** Приведем без доказательства оценку скорости сходимости метода (27.12) — (27.14). Пусть C_L, C_ξ, C_b, C_ρ, C_a, r, t, β — положительные константы,

$$\alpha = \min\left(r - t, \frac{t}{2}, \beta\right), \qquad 0 < t < r \leqslant 1,$$

и градиент функции $F(x)$ удовлетворяет неравенству

$$\|\nabla F(x)-\nabla F(y)\|\leqslant C_L\|x-y\|,\quad x, y\in X.$$

Пусть при $k\geqslant N$ выполняются следующие условия:

$$\rho_k=C_\rho k^{-r},\quad a_k=C_a k^{-t},$$

$$\alpha N^{-(1-r)}<C_\rho\leqslant N^r,\quad 2\alpha N^{t-1}<2C_a\leqslant N^t,$$

$$\|\xi(k)\|\leqslant C_\xi,\quad \|b_k\|\leqslant C_b k^{-\beta}.$$

Число N вводится для того, чтобы выбрать константы C_ρ и C_a по возможности большими, поскольку при доказательстве теоремы используются условия $\rho_h\leqslant 1$, $a_k\leqslant 1$.

Теорема 27.6. *При $k\geqslant N$ имеет место оценка*

$$\mathsf{M}(F(x^k)-F(x^*))\leqslant C_1 k^{-\alpha},\quad x^*\in X^*, \tag{27.21}$$

где

$$C_1=\max\{(\mathsf{M}F(x^N)-F(x^*))N^\alpha,$$

$$\left(2C_xC_aC_\rho+\frac{1}{2}C_LC_x^2C_\rho^2\right)(C_\rho-\alpha N^{-(1-r)})^{-1}\},$$

$$C_x=\max_{x,y\in X}\|x-y\|.$$

Выберем константы r и t так, чтобы показатель степени α в формуле (27.21) был максимальным, т. е. решим задачу

$$\alpha\equiv\min\left(r-t,\ \frac{t}{2},\ \beta\right)\to\max,\quad 0<t<r\leqslant 1.$$

Легко заметить, что оптимальное решение есть

$$r=1,\quad t=\alpha/3,\quad \alpha=\min(1/3,\beta).$$

4. Примеры. Проиллюстрируем применение стохастического метода условного градиента к решению некоторых стохастических задач.

Многопродуктовая задача. Предположим, что требуется сделать заказ $x=(x_1,\ldots,x_n)$ на поставку n неоднородных продуктов при условии, что спрос характеризуется случайным вектором $\theta=(\theta_1,\ldots,\theta_r)$ и задан коэффициент взаимозаменяемости λ_{js} s-го продукта спроса j-м продуктом заказа (вектор заказа может состоять из продуктов, не входящих в вектор спроса). Значение x_j представим как $x_j=\sum_{s=1}^{r}x_{js}$ и рассмотрим функцию потерь $f(x,\theta)$. Если функция потерь учитывает лишь затраты, связанные с недоиспользованием одних продуктов $\left(\text{в случае } \sum_{j=1}^{n}\lambda_{js}x_{js}>\theta_s\right)$, и потери от дефицита других (в случае

$\sum_{j=1}^{n} \lambda_{js} x_{js} \leqslant \theta_s\Big)$, то естественно считать, что

$$f(x, \theta) = \sum_{s=1}^{r} \max \left\{\alpha_s \Big(\sum_{j=1}^{n} \lambda_{js} x_{js} - \theta_s\Big), \quad \beta_s \Big(\theta_s - \sum_{j=1}^{n} \lambda_{js} x_{js}\Big)\right\},$$

где α_s — затраты на хранение единицы s-го продукта, β_s — затраты от его дефицита. Тогда математическое ожидание потерь равно

$$F(x) = \sum_{s=1}^{r} \mathsf{M} \max \left\{\alpha_s \Big(\sum_{j=1}^{n} \lambda_{js} x_{js} - \theta_s\Big), \quad \beta_s \Big(\theta_s - \sum_{j=1}^{n} \lambda_{js} x_{js}\Big)\right\}. \quad (27.22)$$

Требуется найти вектор x, минимизирующий (27.22), при условии

$$\sum_{j=1}^{n} a_j x_j \leqslant a, \quad x_j = \sum_{s=1}^{r} x_{js}, \quad x_{js} \geqslant 0, \quad (27.23)$$

где a_j — объем единицы j-го продукта, a — вместимость склада.

Заметим, что задача линейного программирования (27.20) имеет тривиальное решение: все x_{js} равны нулю, кроме $x_{js} = a/a_j$, для которого достигается

$$\min_{j,s} \left\{v_{js} \frac{a}{a_j}\right\} < 0, \quad j = 1, \ldots, n, \quad s = 1, \ldots, r$$

(v_{js} определяется по формуле (27.19)).

Однопродуктовая задача с учетом затрат на перевозки. Пусть имеется однородный продукт, размещенный на m складах. Допустимое количество продукта на складе i есть a_i ($i = 1, \ldots, m$). В рассматриваемый период продукт будет требоваться на n рынках, причем потребность рынка j не может быть заранее определена, и ее следует считать случайной величиной θ_j. Обозначим через

$$y_j = \sum_{i=1}^{m} x_{ij}, \qquad j = 1, \ldots, n,$$

количество продукта, перевезенного на j-й рынок. Тогда убыток при условии, что спрос есть θ_j, равен

$$f_j(y_j, \theta_j) = \begin{cases} \alpha_j \Big(\sum_{i=1}^{m} x_{ij} - \theta_j\Big), & \sum_{i=1}^{m} x_{ij} \geqslant \theta_j, \\ \beta_j \Big(\theta_j - \sum_{i=1}^{m} x_{ij}\Big), & \sum_{i=1}^{m} x_{ij} < \theta_j. \end{cases}$$

Требуется определить совокупность $x = (x_{ij}, i = 1, \ldots m, j = 1, \ldots$ $\ldots, n)$, минимизирующую сумму транспортных издержек и ожидаемых убытков

$$F(x) =$$

$$= \sum_{i,j} a_{ij}x_{ij} + \sum_{j=1}^{n} \mathsf{M} \max \left\{\alpha_j\left(\sum_{i=1}^{m} x_{ij} - \theta_j\right), \quad \beta_j\left(\theta_j - \sum_{i=1}^{m} x_{ij}\right)\right\} \quad (27.24)$$

при условии

$$\sum_{j=1}^{n} x_{ij} \leqslant a_i, \quad i = 1, \ldots, m, \quad x_{ij} \geqslant 0. \quad (27.25)$$

В силу специфики ограничений (27.25) задача линейного программирования в методе условного градиента разбивается на m подзадач, каждая из которых имеет тривиальное решение: для каждого i только одно значение $x_{ij} = a_i$ больше нуля, остальные x_{ij} равны нулю.

Стохастическая модель распределения площадей под сельскохозяйственные культуры. Рассмотрим одну модель оптимального распределения посевных площадей с учетом того, что урожайность на разных участках носит случайный характер.

Примем следующие обозначения: n — количество культур, m — количество участков земли различного вида, S_j — площадь j-го участка, $a_{ij}(\theta)$ — урожайность i-й культуры на j-м участке, γ_i — удельный вес i-й культуры в общем производстве рассматриваемой сельскохозяйственной продукции, x_{ij} — площадь, занимаемая под i-ю культуру на j-м участке.

Требуется выбрать такой план распределения посевных площадей, чтобы выполнялись ограничения по площадям для земельных участков каждого вида, и максимизировать математическое ожидание количества комплектов сельскохозяйственной продукции в заданных пропорциях, т. е.

$$F(x) \equiv \mathsf{M}\left(\min_i \frac{1}{\gamma_i} \sum_{j=1}^{m} a_{ij}(\theta)\, x_{ij}\right) \to \max,$$

$$\sum_{i=1}^{n} x_{ij} \leqslant S_j, \qquad j = 1, \ldots, m,$$

$$x_{ij} \geqslant 0, \quad i = 1, \ldots, n.$$

В силу свойств ограничений описанной модели задача линейного программирования в стохастическом методе условного градиента разбивается на m подзадач, каждая из которых имеет очевидное решение.

5. **Стохастический метод приведенного градиента.** Исследуем теперь стохастический вариант метода, изложенного в § 16, для решения задачи минимизации

$$F(x) = \mathsf{M} f(x, \theta)$$

при условиях

$$x \in X = \{x\colon\ Ax = b, \quad x \geqslant 0\},$$

где A — матрица размера $m \times n$, b — вектор из E_m. Для множества X выполняются предпосылки, сформулированные там же.

Стохастический вариант метода приведенного градиента решения задачи (27.1), (27.2), когда $f(x, \theta)$ — липшицева функция, определяется следующим образом.

Начальный этап. Выбрать допустимую точку x^1, удовлетворяющую условиям $Ax = b$, $x \geqslant 0$. Положить $k = 1$ и перейти к шагу 1.

Шаг 1. Выбирают I_k — множество индексов m штук наибольших компонент вектора x^k,

$$x_B = \{x_i,\ i \in I_k\},\quad B = \{a^j,\ j \in I_k\},\quad N = \{a^j,\ j \notin I_k\}.$$

Вычисляют приведенный градиент

$$r_N^T = (v_N^k)^T - (v_B^k)^T B^{-1} N. \tag{27.26}$$

Определяют направление спуска d^k:

$$d_j^k = \begin{cases} -r_j, & j \notin I_k,\quad r_j \leqslant 0, \\ -x_j r_j, & j \notin I_k,\quad r_j > 0, \end{cases} \tag{27.27}$$

$$d_B^k = -B^{-1} N d_N.$$

Шаг 2 Последовательности точек x^k, v^k пересчитывают по формулам

$$x^{k+1} = x^k + \sigma_k d^k,$$

$$v^{k+1} = v^k + a_k (\xi(x^k, \alpha_k) - v^k),$$

$$\sigma_k = \min(\rho_k, \lambda), \tag{27.28}$$

$$\lambda = \begin{cases} \min\left\{-\dfrac{x_j^k}{d_j^k}\ \middle|\ d_j^k < 0\right\}, & \exists d_j^k < 0, \\ \infty, & d^k \geqslant 0. \end{cases}$$

Полагают $k = k + 1$ и переходят к шагу 1. Векторы $\xi(x^k, \alpha_k)$ вычисляют по формулам (25.2) — (25.4). Если $f(x, \theta)$ — выпуклая по x функция, то

$$\xi(x^k, \alpha_k) = g(\tilde{x}^k, \theta^k).$$

Теорема 27.7. *Пусть выполнены условия теоремы 27.5. Тогда последовательность $F(x^k)$ сходится с вероятностью 1 и предельные точки последовательности x^k с вероятностью 1 удовлетворяют условиям Куна — Таккера.*

Доказательство проводится аналогично доказательству теоремы 16.1.

Рассмотрим теперь ситуацию, когда функция $F(x)$ непрерывно дифференцируема и существует случайный вектор $\xi(k)$ такой, что

$$M\xi(k) \mid B_k = \nabla F(x^k) + b_k.$$

В частности, если $f(x, \theta)$ при каждом θ непрерывно дифференцируема по x, то можно взять $\xi(k) = \nabla f(x^k, \theta^k)$. В этом случае метод (27.26) — (27.28) упрощается. Усреднение v^k проводить не надо, а приведенный градиент r_N в (27.26) определяется по формуле

$$r_N^T = \xi_N^T(k) - \xi_B^T(k) B^{-1} N. \qquad (27.26)$$

Теорема 27.8. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < \infty, \quad \mathsf{M}\, \| b_k \| \leqslant r_k, \quad r_k \to 0,$$

$$\mathsf{M}\, \| b_k \|^2 < \infty, \quad \mathsf{M}\, \| \xi(k) \|^2 < \infty,$$

$\nabla F(x)$ *удовлетворяет в X условию Липшица. Тогда $F(x^k)$ сходится с вероятностью 1 и предельные точки последовательности x^k, определяемой по формулам* (27.26) — (27.28), *с вероятностью* 1 *удовлетворяют условиям Куна — Таккера.*

6. Стохастические методы возможных направлений. В предыдущих пунктах этого параграфа рассматривались задачи минимизации $F(x) \equiv \mathsf{M} f(x, \theta)$ при условиях $x \in X$. В том случае, когда множество X простой структуры, использовалась операция проектирования на X. В методе условного градиента операция проектирования заменялась минимизацией в X некоторой линейной формы. В методе приведенного градиента значения базисных переменных пересчитывались путем изменения независимых внебазисных переменных в соответствии с требованием сохранения допустимости.

Обсудим теперь более сложную задачу, когда область X определяется нелинейными ограничениями $f_i(x) \leqslant 0$ $(i = 1, \ldots, m)$. Полагаем, что X ограничено и выполняется условие регулярности.

Пусть $F(x)$, $f_i(x)$ $(i = 1, \ldots, m)$ — непрерывно дифференцируемые функции. Для решения задачи (27.1), (27.2) рассмотрим следующий стохастический вариант метода возможных направлений, изложенного в § 17:

$$x^{k+1} = x^k - \gamma_k d^k, \qquad (27.29)$$

где возможное направление d^k при заданной допустимой точке x^k находится из решения следующей задачи линейного программирования:

$$\min \sigma \qquad (27.30)$$

при условиях

$$\begin{gathered} (y^k, d) - \sigma \leqslant 0, \\ f_i(x^k) + (\nabla f_i(x^k), d) - \sigma \leqslant 0, \qquad i = 1, \ldots, m, \\ -1 \leqslant d_j \leqslant 1, \qquad j = 1, \ldots, n. \end{gathered} \qquad (27.31)$$

Последовательность y^k находится с помощью операции усреднения

$$y^{k+1} = y^k + a_k(\xi(k) - y^k), \qquad (27.32)$$

где $\xi(k)$ — случайный вектор такой, что

$$M\xi(k)/B_k = \nabla F(x^k) + b_k.$$

Положим

$$\rho_k' = \max\{\lambda \mid x^k + \lambda d^k \in X\},$$

$$\gamma_k = \min\{\rho_k', \rho_k\}. \qquad (27\ 33)$$

Теорема 27.9. *Пусть выполнены условия теоремы* 27.4. *Тогда с вероятностью* 1 *предельные точки последовательности* x^k, *определяемой по формулам* (27.29) — (27.33), *удовлетворяют условиям Куна — Таккера, последовательность* $F(x^k)$ *сходится с вероятностью* 1.

Доказательство. Из теоремы о среднем имеем

$$F(x^{k+1}) \leqslant F(x^k) + C\gamma_k^2 + \gamma_k(\nabla F(x^k), d^k). \qquad (27.34)$$

В неравенстве (27.34) нужно оценить $(\nabla F(x^k), d^k)$. Это можно сделать, рассмотрев вспомогательную задачу

$$\min \sigma$$

при условиях

$$(\nabla F(z), d) \leqslant \sigma,$$

$$f_i(z) + (\nabla f_i(z), d) \leqslant \sigma, \qquad i = 1, \ldots, m, \qquad (27.35)$$

$$-1 \leqslant d_j \leqslant 1.$$

Пусть $(\bar\sigma(z), \bar d(z))$ — решение этой задачи. Тогда имеет место следующее свойство: для точек z, принадлежащих достаточно малой окрестности точки $z' \notin X^*$ (множество точек Куна — Таккера), выполняется неравенство $\bar\sigma(z) < \sigma < 0$. Отсюда следует, что

$$(\nabla F(z), \bar d(z)) \leqslant \sigma < 0.$$

Очевидно, что для доказательства теоремы достаточно проверить условия (4.12), (4.13). Пусть существует подпоследовательность

$$x^s \to x' \notin X^*, \qquad s \in S \subset \{1, 2, \ldots, k, \ldots\}.$$

Существуют такое целое $\bar s$ и положительное число $\bar\delta$, что $U_{2\bar\delta}(x^s) \cap \cap X^* = \varnothing$ $(s \geqslant \bar s)$. Поэтому

$$(\nabla F(x^s), \bar d(x^s)) \leqslant \sigma < 0,$$

$$f_i(x^s) + (\nabla f_i(x^s), \quad \bar d(x^s)) \leqslant \sigma,$$

$$-1 \leqslant \bar d_j(x^s) \leqslant 1, \qquad j = 1, \ldots, n,$$

где $\bar d(x^s)$ — решение задачи (27.35).

Пусть условие (4.12) не выполняется, т. е. все точки x^k $(k \geqslant s)$ принадлежат $U_\delta(x^s)$ $(\delta \leqslant \bar\delta/2)$. Тогда для достаточно малых δ из предыдущих неравенств вытекает, что

$$(\nabla F(x^k), \bar{d}(x^s)) \leqslant \sigma/2,$$

$$f_i(x^k) + (\nabla f_i(x^k), \quad \bar{d}(x^s)) \leqslant \sigma/2,$$

$$-1 \leqslant \bar{d}_j(x^s) \leqslant 1, \qquad j = 1, \ldots, n.$$

Поскольку с вероятностью 1 $(y^k - \nabla F(x^k)) \to 0$, то, начиная с некоторого номера $k \geqslant s$,

$$(y^k, \bar{d}(x^s)) \leqslant \sigma/4.$$

Поэтому для оптимального решения d^k задачи (27.30) — (27.31) выполняются соотношения

$$(y^k, d^k) \leqslant \sigma/4,$$

$$f_i(x^k) + (\nabla f_i(x^k), d^k) \leqslant \sigma/4, \tag{27.36}$$

$$-1 \leqslant d^k \leqslant 1, \qquad j = 1, \ldots, n.$$

Из неравенств (27.36) легко показать, что величина шага γ_k равна ρ_k. Следовательно, из неравенства (27.34) для достаточно больших k имеем

$$F(x^k) \leqslant F(x^s) + \frac{\sigma}{8} \sum_{r=s}^{k-1} \rho_r. \tag{27.37}$$

Переходя к пределу по $k \to \infty$, получим противоречие. Поэтому

$$k(s) = \min_{r>s} \{r \mid \| x^r - x^s \| > \delta\} < \infty.$$

В свою очередь из соотношения

$$x^{k(s)} = x^s + \sum_{r=s}^{k(s)-1} \rho_r d^r$$

вытекает, что

$$\sum_{r=s}^{k(s)-1} \rho_r > \frac{\delta}{C}.$$

Следовательно, из неравенства (27.37) заключаем, что выполняется условие (4.13)

$$\overline{\lim_{s\to\infty}} F(x^{k(s)}) < \lim_{s\to\infty} F(x^s).$$

Теорема доказана.

Рассмотрим теперь задачу минимизации функции

$$F(x) \equiv \mathsf{M} f(x, \theta)$$

в предположении, что $f(x, \theta)$ удовлетворяет по x в $D \supset X$ условию Липшица с константой $L(\theta)$; $f_i(x)$ $(i = 1, \ldots, m)$ — непрерывно дифференцируемые функции.

Определим последовательности

$$x^{k+1} = x^k + \gamma_k d^k,$$

$$v^{k+1} = v^k + a_k (\xi (x^k, \alpha_k) - v^k),$$

где d^k — решение задачи

$$\min \sigma$$

при условиях

$$(v^k, d) - \sigma \leqslant 0,$$

$$f_i (x^k) + (\nabla f_i (x^k), d) - \sigma \leqslant 0, \qquad i = 1, \ldots, m,$$

$$-1 \leqslant d_j \leqslant 1, \quad j = 1, \ldots, n.$$

Величина шага γ_k выбирается по формуле (27.33). В конечно-разностных методах минимизации $F(x)$ векторы $\xi(x^k, \alpha_k)$ вычисляются по формулам (25.2) — (25.4). Если $f(x, \theta)$ — выпуклая функция, то в стохастическом квазиградиентном методе

$$\xi (x^k, \alpha_k) = g (\tilde{x}^k, \theta^k) \in \partial f (\tilde{x}^k, \theta^k).$$

Теорема 27.10. *Пусть выполнены условия теоремы* 27.5. *Тогда* $\{F(x^k)\}$ *сходится с вероятностью* 1 *и предельные точки последовательности* $\{x^k\}$ *с вероятностью* 1 *удовлетворяют условию Куна — Таккера.*

Доказательство проводится аналогично доказательству теорем 17.1 и 27.9.

7. Стохастические методы минимизации липшицевых функций. Пусть требуется решить задачу

$$F(x) \equiv \mathsf{M} f(x, \theta) \to \min \qquad (27.38)$$

при условии

$$h(x) \leqslant 0, \qquad (27.39)$$

где $f(x, \theta)$, $h(x)$ — липшицевы функции. Метод решения задачи (27.38), (27.39) определяется соотношениями

$$x^{k+1} = x^k + \rho_k d^k,$$

где

$$d^k = -z_F^k, \qquad h(x^k) < 0,$$

$$d^k = -z_F^k - u_k z_h^k, \qquad h(x^k) \geqslant 0,$$

$$u_k = \max \left\{ 0, \frac{h(x^k) - (z_F^k, z_h^k)}{\| z_h^k \|^2} \right\}.$$

Векторы z_F^k, z_h^k вычисляются по формулам

$$z_F^{k+1} = z_F^k + a_k (\xi_F (x^k, \alpha_k) - z_F^k),$$

$$z_h^{k+1} = z_h^k + a_k (\xi_h (x^k, \alpha_k) - z_h^k).$$

В конечно-разностных методах ξ_F, ξ_h определяются соотношениями (25.2) — (25.4). Если f, h — выпуклые функции, то в стохастическом квазиградиентном методе

$$\xi_F(x^k, \alpha_k) = g(\tilde{x}^k, \theta_k) \in \partial f(\tilde{x}^k, \theta^k),$$

$$\xi_h(x^k, \alpha_k) = g(\tilde{x}^k) \in \partial h(\tilde{x}^k).$$

Обозначим через X^* множество точек, удовлетворяющих необходимому условию экстремума задачи (27.38), (27.39): $x^* \in X^*$, если найдутся такие векторы $g_F(x^*) \in \partial F(x^*)$, $g_h(x^*) \in \partial h(x^*)$, что

$$g_F(x^*) + \lambda g_h(x^*) = 0, \quad \lambda \geqslant 0,$$

$$\lambda h(x^*) = 0, \qquad h(x^*) \leqslant 0.$$

Имеет место следующий результат.

Теорема 27.11. *Пусть выполнены условия теоремы* 27.5. *Тогда* $F(x^k)$ *сходится с вероятностью* 1 *и предельные точки последовательности* x^k *с вероятностью* 1 *принадлежат* X^*.

§ 28. Стохастический конечно-разностный метод Эрроу — Гурвица

Рассмотрим теперь решение общей задачи стохастического программирования:

$$F_0(x) \equiv \mathsf{M} f_0(x, \theta) \to \min \tag{28.1}$$

при ограничениях

$$F_i(x) \equiv \mathsf{M} f_i(x, \theta) \leqslant 0, \qquad i = 1, \ldots, m, \tag{28.2}$$

$$x \in X, \tag{28.3}$$

где $f_i(x, \theta)$ $(i = 0, 1, \ldots, m)$ — выпуклые липшицевы по x (при каждом θ) на $D \supset X$ функции, X — выпуклое ограниченное замкнутое множество в E_n. Изучаемый метод определяется соотношениями

$$x^{k+1} = \pi_X\left(x^k - \rho_k\left[\xi^0(x^k, d^k) + \sum_{i=1}^{m} u_i^k \xi^i(x^k, \alpha_k)\right]\right), \tag{28.4}$$

$$u^{k+1} = \pi_U(u^k + \rho_k L_u(x^k, u^k, \theta^k)), \tag{28.5}$$

где $L_u(x, u, \theta)$ — градиент функции

$$L(x, u, \theta) = f_0(x, \theta) + \sum_{i=1}^{m} u_i f_i(x, \theta)$$

по переменным u.

Векторы $\xi^{\nu}(x^k, \alpha_k)$ $(\nu = 0, 1, \ldots, m)$ вычисляются по одной из формул:

$$\frac{1}{2\alpha_k} \sum_{i=n}^{n} [f_{\nu}(\tilde{x}_1^k, \ldots, x_i + \alpha_k, \ldots, \tilde{x}_n^k, \theta^k) -$$

$$- f_{\nu}(\tilde{x}_1^k, \ldots, x_i - \alpha_k, \ldots, \tilde{x}_n^k, \theta^k)] e_i, \quad (28.6)$$

$$\sum_{i=1}^{n} \frac{f_{\nu}(\tilde{x}^k + \Delta_k e_i, \theta^k) - f_{\nu}(\tilde{x}^k, \theta^k)}{\Delta_k} e_i. \quad (28.7)$$

Сходимость траекторий $\{x^k\}$, $\{u^k\}$ к множеству седловых точек функции Лагранжа

$$L(x, u) = F_0(x) + \sum_{i=1}^{m} u_i F_i(x)$$

исследуется в чезаровском смысле.

Обозначим

$$\hat{x}^k = \sum_{s=0}^{k} \rho_s x^s \Big/ \sum_{s=0}^{k} \rho_s,$$

$$\hat{u}^k = \sum_{s=0}^{k} \rho_s u^s \Big/ \sum_{s=0}^{k} \rho_s,$$

$K_{\nu}(\theta)$ — константа Липшица функции $f_{\nu}(x, \theta)$ $(\nu = 0, 1, \ldots, m)$ на D.

Теорема 28.1. *Пусть выполнены условия*

$$\sum_{k=0}^{\infty} \rho_k = \infty, \quad \sum_{k=0}^{\infty} \rho_k^2 < \infty, \quad \frac{\Delta_k}{\alpha_k} \to 0, \quad \sum_{k=0}^{\infty} \rho_k \alpha_k < \infty, \quad \alpha_k \to 0,$$

$$\mathsf{M} K_{\nu}^2(\theta) < \infty, \quad \mathsf{M} f_{\nu}^2(x, \theta) < \infty, \quad \nu = 0, 1, \ldots, m.$$

Тогда с вероятностью 1 *предельные точки последовательностей* $\{\hat{x}^k\}$, $\{\hat{u}^k\}$ *принадлежат множеству седловых точек* $L(x, u)$.

Доказательство проводим для случая, когда векторы $\xi^{\nu}(x^k, \alpha_k)$ $(\nu = 0, 1, \ldots, m)$ определяются по формулам (28.6). Ограниченные величины обозначаем буквой C.

Пусть

$$v^k = \xi_0(x^k, \alpha_k) + \sum_{i=1}^{m} u_i^k \xi_i(x^k, \alpha_k),$$

$$\gamma_k = \xi_0(x^k, \alpha_k) - \nabla F_0(x^k, \alpha_k) + \sum_{i=1}^{m} u_i^k (\xi_i(x^k, \alpha_k) - \nabla F_i(x^k, \alpha_k)),$$

(X^*, U^*) — множество седловых точек функции $L(x, u)$, $x^* \in X^*$.

Имеют место неравенства

$$\| x^{k+1} - x^* \|^2 \leqslant \| x^k - \rho_k v^k - x^* \|^2 \leqslant$$
$$\leqslant \| x^k - x^* \|^2 + 2\rho_k (v^k, x^k - x^*) + \rho_k^2 \| v^k \|^2 \leqslant$$
$$\leqslant \| x^k - x^* \|^2 + 2\rho_k (\nabla F_0(x^k, \alpha_k) + \sum_{i=1}^m u_i^k (\nabla F_i(x^k, \alpha_k), \ x^* - x^k)) +$$
$$+ 2\rho_k (\gamma_k, x^* - x^k) + C\rho_k^2 + \rho_k^2 (\| v^k \|^2 - M \| v^k \|^2) \leqslant$$
$$\leqslant \| x^k - x^* \|^2 + 2\rho_k (\overline{L}(x^*) - L(x^k, u^k)) + C\rho_k \alpha_k +$$
$$+ C\rho_k^2 + 2\rho_k (\gamma_k, x^* - x^k) + \rho_k^2 (\| v^k \|^2 - M \| v^k \|^2), \qquad (28.8)$$

где $\overline{L}(x) = \max_{u \in U} L(x, u)$.

Для двойственных переменных u выполняются неравенства

$$\| u^{k+1} - u \|^2 \leqslant \| u^k + \rho_k L_u (x^k, u^k, \theta^k) - u \|^2 \leqslant \| u^k - u \|^2 +$$
$$+ 2\rho_k (L(x^k, u^k, \theta^k) - L(x^k, u, \theta^k)) + \rho_k^2 \| L_u (x^k, \ u^k, \theta^k) \|^2 \leqslant$$
$$\leqslant \| u^k - u \|^2 + 2\rho_k (L(x^k, u^k, \theta^k) - L(x^k, u, \ \theta^k)) +$$
$$+ C\rho_k^2 + \rho_k^2 (\| L_u (x^k, u^k, \theta^k) \|^2 - M \| L_u (x^k, u^k, \theta^k) \|^2). \qquad (28.9)$$

Сложив неравенства (28.8) и (28.9), получим

$$\| x^k - x^* \|^2 + \| u^{k+1} - u \|^2 \leqslant \| x^k - x^* \|^2 + \| u^k - u \|^2 +$$
$$+ 2\rho_k (\overline{L}(x^*) - L(x^k, u)) + 2\rho_k (L(x^k, u^k, \theta^k) - L(x^k, u^k)) +$$
$$+ 2\rho_k (L(x^k, u) - L(x^k, u, \ \theta^k)) + 2\rho_k (\gamma_k, x^* - x^k) +$$
$$+ \rho_k^2 (\| v^k \|^2 - M \| v^k \|^2) + \rho_k^2 (\| L_u (x^k, u^k, \theta^k) \|^2 -$$
$$- M \| L_u (x^k, u^k, \theta^k) \|^2) + C\rho_k^2 + C\rho_k \alpha_k.$$

Отсюда следует, что

$$\sum_{s=0}^k \rho_s (L(\hat{x}^k, u) - \overline{L}(x^*)) \leqslant C + \frac{1}{2} \sum_{s=0}^k C (\rho_s^2 + \rho_s \alpha_s) +$$
$$+ \frac{1}{2} \sum_{s=0}^k \{2\rho_s (L(x^s, u^s, \theta^s) - L(x^s, u^s)) + 2\rho_s (L(x^s, u) - L(x^s, u, \theta^s)) +$$
$$+ 2\rho_s (\gamma_s, x^* - x^s) + \rho_s^2 (\| v^s \|^2 - M \| v^s \|^2) +$$
$$+ \rho_s^2 \{ \| L_u (x^s, u^s, \theta^s) \|^2 - M \| L_u (x^s, u^s, \theta^s) \|^2 \}. \qquad (28.10)$$

Следовательно, $\overline{L}(\hat{x}^k) \to \overline{L}(x^*)$ с вероятностью 1, поскольку п. н. сходятся ряды в соотношении (28.10). Аналогично получаем, что с вероятностью 1

$$\underline{L}(\hat{u}^k) \to \max_u \underline{L}(u), \quad \underline{L}(u) = \min_{x \in X} L(x, u).$$

Теорема доказана.

БИБЛИОГРАФИЧЕСКИЕ УКАЗАНИЯ

Приводимые ниже литературные указания ни в коей мере не претендуют на полноту: выделены в основном те работы, которые имеют непосредственное отношение к излагаемому материалу, а также основные монографии; при ссылках монографиям отдается предпочтение перед статьями.

Предисловие

Общие проблемы оптимизации рассматриваются в работах В. М. Глушкова [16], Н. Н. Красовского [54], В. С. Михалевича и В. Л. Волковича [62], Н. Н. Моисеева [67], Л. С. Понтрягина, В. Г. Болтянского, Р. В. Гамкрелидзе и Е. Ф. Мищенко [101], А. Н. Тихонова и В. Я. Арсенина [118].

С теорией линейного программирования и линейной алгебры можно познакомиться по монографиям Данцига [32], Д. К. Фаддеева и В. Н. Фаддеевой [120], Муртафа [71]. Теория и методы нелинейного программирования изложены в книгах Базара и Шетти [6], Ф. П. Васильева [11], Ю. Г. Евтушенко [36], И. И. Еремина и Н. Н. Астафьева [38], Зангвилла [46], В. Г. Карманова [51], Н. Н. Моисеева, Ю. П. Иванилова, Е. М. Столяровой [68], Ортеги и Рейнболдта [92], Полака [98], Б. Т. Поляка [100], Б. Н. Пшеничного и Ю. М. Данилина [105], Химмельблау [125] и сборнике [131].

Глава 1

Элементы теории выпуклого анализа содержатся почти в каждом учебнике по оптимизации, исчерпывающий материал имеется в монографии Рокафеллара [109], большое место отведено в книгах В. Ф. Демьянова и Л. В. Васильева [34], Б. Н. Пшеничного [104].

§ 1. Различные классы негладких функций изучались в работах Данскина [31], В. Ф. Демьянова и В. Н. Малоземова [33], В. Ф. Демьянова и Л. В. Васильева [34], А. Д. Иоффе и В. М. Тихомирова [49], В. И. Норкина [80, 83], Е. А. Нурминского [88], Б. Н. Пшеничного [103, 104], Н. З. Шора [136], Кларка [145], Льюенбергера [162], Мангасарьяна [163], Мирицы [169], Миффлина [167], Рокафеллара [176—178], Бихайна [143].

Обобщенные градиенты вводились для невыпуклых локально липшицевых [145], квазидифференцируемых [34, 103], допускающих верхнюю выпуклую аппроксимацию [104], обобщенно дифференцируемых [80, 83], слабо выпуклых [88] функций. Различные свойства обобщенных градиентов изучены для локально липшицевых [146], квазидифференцируемых [34] и обобщенно дифференцируемых [83] функций. Численные методы минимизации невыпуклых негладких функций, использующие обобщенные градиенты, предложены для липшицевых [22], обобщенно дифференцируемых [80], слабо выпуклых [88] и полугладких [166] функций.

§ 2. Дифференциальные свойства локально липшицевых функций систематически начал изучать Кларк [145]; он ввел понятия «обобщенная производная по направлению» и «субдифференциал». Изложение § 2 основано на работах [146, 22]. Операция сглаживания или усреднения функций давно используется в математике. Для целей негладкой оптимизации она начала применяться А. М. Гупалом [23], В. Я. Катковником [52]. Свойства локально липшицевых функций изучались также в работах Рокафеллара [176—178], Хириарт-Урути [153]. Теорема о среднем установлена в [159]. Нетрадиционное определение локально липшицевых функций дал Мирица [169].

§ 3. С проблематикой теории необходимых условий экстремума для негладких функций можно ознакомиться по монографиям Б. Н. Пшеничного [103, 104]. Необходимые условия экстремума для задач с локально липшицевыми функциями изучали Кларк [146], Варга [10], А. М. Гупал [22], Хириарт-Урути [152, 153], Миффлин [167]. Своеобразные необходимые условия экстремума установлены для квазидифференцируемых функций [34]. Вопрос об ограниченности множителей Куна — Таккера при более общих условиях изучался в [150].

Глава 2

§ 4. Конечно-разностные методы минимизации локально липшицевых функций были предложены А. М. Гупалом в статье [23] и развивались в книге [22]. Численные методы минимизации на основе операции сглаживания исследовались в книге В. Я. Катковника [52]. Численные методы оптимизации без вычисления градиентов изучались для гладких задач в книге Б. Н. Пшеничного и Ю. М. Данилина [105], для негладких выпуклых задач в книге А. С. Немировского и Д. Б. Юдина [73].

§ 5. Методы решения минимаксных задач рассматривались в книгах Данскина [31], В. Ф. Демьянова и В. Н. Малоземова [33], В. В. Федорова [122], Н. З. Шора [136], Е. А. Нурминского [91]. Построение конечных разностей для функций максимума выполнено в статье [17].

§ 6. С методами случайного поиска можно ознакомиться по книге Л. А. Растригина [106]; сходимость и скорость сходимости этих методов исследовались В. Г. Кармановым [51]. Результат § 6 содержится в книге [22].

§ 7. Сложность задач и эффективность методов выпуклой оптимизации исследованы А. С. Немировским и Д. Б. Юдиным [73]. Результаты § 7 содержатся в статье [27]. Скорость сходимости конечно-разностных методов исследовал Г. Г. Мураускас [70].

§ 8. Метод функций Ляпунова для исследования сходимости релаксационных методов оптимизации изложен в книгах Базара и Шетти [6], Ю. Г. Евтушенко [36], Б. Т. Поляка [100]. Для нерелаксационных алгоритмов этот метод обобщен в работах [5, 46, 87, 91].

Первые необходимые и достаточные условия сходимости алгоритмов нелинейного программирования сформулировал Зангвилл [46]. На основе этих результатов Е. А. Нурминский [87, 91] предложил весьма общие достаточные условия сходимости алгоритмов нелинейного программирования, нашедшие широкое применение при исследовании нерелаксационных алгоритмов нелинейного и стохастического программирования. Затем эти результаты обобщались многими авторами. В данном параграфе необходимые и достаточные условия сходимости сформулированы в теоремах 8.1—8.3; теорема 8.1 является самой общей. Л. Г. Баженов [5] преобразовал необходимые и достаточные условия Зангвилла [46] к виду, удобному для применения к нерелаксационным алгоритмам нелинейного программирования, и избавился в них от формальных предположений противного. Близкие к теореме 8.3 необходимые и достаточные условия сходимости получил С. П. Урясьев.

Достаточные условия сходимости в данном параграфе сформулированы в теоремах 8.4—8.9. Условие 1) теоремы 8.8 является ослаблением аналогичного требования Е. А. Нурминского [87]; оно использовалось В. И. Норкиным [80]. Дальнейшее его ослабление (условие 1) теоремы 8.6) найдено С. П. Урясьевым. Условие 2) теоремы 8.8 введено Е. А. Нурминским [89] и ослаблено (условие 2) теоремы 8.6) С. И. Ляшко [59]. Условие 3) теоремы 8.8 эквивалентно тому, что W^* нигде в E_1 не плотно [80], и является обобщением требования Е. А. Нурминского [87], что W^* не более чем счетно.

Первоначально Е. А. Нурминский [87] примерно в условиях теоремы 8.8 доказал, что все предельные точки последовательности $\{x^k\}$ принадлежат X^*. Он же показал [91], что без условия 3) теоремы 8.8 существуют предельные точки $\{x^k\}$, принадлежащие X^*, а В. И. Норкин [84] отметил, что это минимальные по значению $W(x)$ предельные точки $\{x^k\}$. Ю. М. Ермольев [41] и Ю. М. Ермольев и П. И. Верченко [42] заметили, что при условиях Е. А. Нурминского [87] $\{W(x^k)\}$ имеет предел. М. В. Михалевич [65] показал, что без условия 3) в теореме 8.8 имеет место $\varliminf_{k\to\infty} W(x^k) \leqslant \sup\{w \in E_1 \mid w \in W^*\}$. Следующий шаг сделал П. А. Дорофеев [35]; он показал, что в условиях Е. А. Нурминского [87] без предположений о множестве W^* имеет место сходимость $\{x^k\}$ одной предельной точкой и по функционалу. Связность множества предельных точек $\{x^k\}$ в условиях леммы 8.3 отмечена в [92, с. 459].

В работе С. К. Завриева [44] предложены другие общие достаточные условия сходимости алгоритмов невыпуклой оптимизации. В одной части эти условия являются более общими, чем достаточные условия сходимости данного параграфа, а в другой — более конкретными: в них нет формальных предположений противного. Однако характер полученных в [44] утверждений о сходимости алгоритмов тот же самый; в частности, получены условия сходимости алгоритмов невыпуклой оптимизации по функционалу, доказана теорема, аналогичная теореме 8.8, а также доказана лемма 8.3.

§ 9, 10. Метод обобщенного градиента предложен Н. З. Шором [134]; условия его сходимости для выпуклых функций исследованы в [40, 99, 136]. Метод обобщался и усовершенствовался в [18, 34, 38, 73, 136]; он также был распространен для решения некоторых невыпуклых негладких задач оптимизации [4, 80, 88, 136, 166]. Методы обобщенного градиента с растяжением пространства подробно изучаются в книге [136]. На основе метода эллипсоидов [141], являющегося частным случаем метода [135], Л. Г. Хачиян [124] построил полиномиальный алгоритм в линейном программировании.

Задачи с упорядоченными по важности критериями (задачи лексикографической оптимизации) рассмотрены в монографиях В. В. Подиновского и В. М. Гаврилова [97], В. В. Федорова [122], а также в статье [121].

Несобственные задачи изучены в монографии И. И. Еремина, В. Д. Мазурова и Н. Н. Астафьева [39].

В книгах [13, 136] описаны многочисленные приложения метода обобщенного градиента и его обобщений.

§ 11. Релаксационные методы минимизации негладкой выпуклой функции максимума изложены в монографии В. Ф. Демьянова и В. Н. Малоземова [33]. Релаксационные методы минимизации общих выпуклых функций предложены Лемарешалем [160] и Вулфом [185]; релаксационным методам выпуклой оптимизации посвящены также статьи [107, 108, 130, 157, 161, 166, 172, 179]. Релаксационные методы минимизации полугладких функций при ограничениях предложены Миффлином [166, 168] и развиты Кивилом [156, 157]. Метод данного параграфа укладывается в общую схему, изложенную в [172].

§ 12. Обзоры методов глобальной оптимизации сделаны Мак-Кормиком [165], Р. Г. Стронгиным [114], А. Г. Сухаревым [117] (см. также сборник [181]). Заметное место отведено проблемам глобальной оптимизации в книгах Д. И. Батищева [7], В. П. Вулатова [8], Ф. П. Васильева [11], И. Б. Моцкуса [69], Л. А. Растригина [106], А. Г. Сухарева [116].

Комбинированные алгоритмы глобальной оптимизации изучались в [1, 11, 12, 69, 82, 86].

Метод невыпуклых аппроксимаций предложен и обоснован в работах С. А. Пиявского [95, 96], Ю. М. Данилина и С. А. Пиявского [29], Ю. М. Данилина [30] и получил широкое распространение. Обобщение на задачи с невыпуклыми ограничениями является новым.

Метод сглаживания для целей глобальной оптимизации предложен Д. Б. Юдиным [138] и изучался в работах [52, 81, 94, 106]. Метод одномерной глобальной оптимизации при невыпуклых ограничениях предложен в [115].

Методы глобальной минимизации вогнутой функции при линейных ограничениях предложены в [3, 8, 126, 154, 188]. Применение методов тяжелого шарика и овражного шага для целей глобальной оптимизации обсуждается в [100], а случайных инерционных методов — в [106].

Глава 4

Усреднение стохастических градиентов для построения сходящихся методов стохастической оптимизации применялось в [20, 22, 41, 52, 91, 128].

Методы минимизации негладких функций с усреднением обобщенных градиентов предлагались в [22, 84, 85, 129, 171].

§ 13. Немонотонные методы с усреднением конечно-разностных аналогов градиента исследовались в [22].

§ 14. Вопросы устойчивости численных методов оптимизации обсуждаются в книгах В. Г. Карманова [51], Б. Т. Поляка [100], а разностных схем — в монографии А. А. Самарского [111].

Устойчивость методов выпуклой оптимизации по отношению к случайным ошибкам вычисления градиента исследована в монографиях Ю. М. Ермольева [41], Кушнера и Кларка [158], А. С. Немировского и Д. Б. Юдина [73], в статье М. В. Михалевича [66].

Устойчивость метода обобщенного градиента по отношению к детерминированным погрешностям определения градиента исследовалась в монографиях Н. З. Шора [136], А. С. Немировского и Д. Б. Юдина [73].

Устойчивость субградиентного метода по отношению к ошибкам в положении точек, где берутся градиенты, установлена в [50, 132].

Метод усредненных градиентов для минимизации обобщенно дифференцируемых функций предложен в [84]. Адаптивные регулировки шага в нерелаксационных методах негладкой оптимизации рассматриваются в [37, 90, 91, 100, 129, 136].

Метод тяжелого шарика описан в [100]. Метод овражного шага предложен И. М. Гельфандом и М. Л. Цетлиным [14], а исследован и обобщен Ю. Е. Нестеровым [74, 76]. Обоснование этих методов для минимизации невыпуклых негладких функций является новым.

Глава 5

§ 15. Метод условного градиента для квадратичных задач был развит Франком и Вулфом [148]. Аналог метода условного градиента для минимизации выпуклых функций исследовался в [78], а для минимизации липшицевых функций — в [23].

§ 16. Метод приведенного градиента исследован Вулфом [184]; он является обобщением симплекс-метода на задачи нелинейного программирования и поэтому получил широкое распространение [6, 71, 131, 170].

§ 17. Методы возможных направлений были предложены Зойтендейком [47]; они изучаются также в книге [6].

§ 18. Метод штрафных функций был введен Курантом, ему посвящена монография Фиакко и Мак-Кормика [123], центральное место он занимает в монографии В. В. Федорова [122].

Существование точных штрафных функций установлено И. И. Ереминым [37] и Зангвиллом [187]; они изучались также в [164]. Глобальные свойства точных штрафных функций изучались в [64, 142, 146].

§ 19. Конечно-разностный метод минимизации липшицевых функций при ограничениях построен как аналог метода линеаризации Б. Н. Пшеничного [105] и исследован в [25].

§ 20. Метод Эрроу — Гурвица предложен в книге [137], чезаровское усреднение исследовалось в [73]. Конечно-разностный метод из § 20 изучался в [26].

Глава 6

§ 21. Сведения по многозначным отображениям имеются в книгах [34, 50, 60, 77, 104].

Измеримые многозначные отображения являются обобщением измеримых функций. С измеримыми функциями достаточно ознакомиться по книге А. Н. Колмогорова и С. В. Фомина [53]; много полезных сведений по измеримости имеется в монографии Варги [10]. Сведения по измеримым многозначным отображениям можно найти в обзоре [182] и в работах [10, 48, 49, 57, 110, 144, 175].

Мы следуем в основном работам Кастена [144], Рокафеллара [175], Химельберга [151]. Результаты Кастена изложены также в [10]. Лемма 21.7 установлена в [79, 86].

§§ 22, 23. Случайные липшицевы функции изучались в [79] (см. замечание 22.1), а обобщенно дифференцируемые — в [86].

Глава 7

С задачами и методами стохастического программирования можно познакомиться по монографиям Ю. М. Ермольева [41], Д. Б. Юдина [139, 140], Ю. М. Ермольева и А. И. Ястремского [43], по обзору Ветса [183]. Поведение методов оптимизации в присутствии случайных помех обсуждается в книге Б. Т. Поляка [100].

Методы стохастического программирования берут начало со статьи Роббинса и Монро [174]. Конечно-разностный метод стохастической аппроксимации был предложен в работе Кифера и Вольфовица [155]. Методу стохастической аппроксимации посвящены монографии Вазана [9], М. Б. Невельсона и Р. З. Хасьминского [72]. Приемы ускорения сходимости метода стохастической аппроксимации указаны в [112], адаптивная регулировка шага рассматривалась в [119].

§ 24. Задачи выпуклого стохастического программирования изучались в [41, 73, 50], обобщение на слабо выпуклый случай сделано в [91], а на обобщенно дифференцируемый случай — в [86].

Трактовка сходимости стохастических методов как следствия устойчивости их детерминированных вариантов принята в книге [50]. Условия сходимости стохастических аналогов детерминированных методов изучались в [89].

Сведения по теории вероятностей содержатся в книгах Лоэва [58], Б. Н. Прохорова и Ю. А. Розанова [102], А. Н. Ширяева [133].

Методы с усреднением стохастических градиентов изучались в [20—22, 41, 52, 85, 86, 100, 128].

Поведение стохастического метода тяжелого шарика в квадратичном случае обсуждается в [100]. Обоснование стохастических методов тяжелого шарика и овражного шага для решения задач невыпуклой стохастической оптимизации является новым.

§ 25. Конечно-разностные методы липшицева стохастического программирования, обобщающие классический метод Кифера и Вольфовица, построены в книге [22]. Скорость сходимости таких методов исследовалась в [70].

§ 26, 27. Операция усреднения впервые использовалась Я. З. Цыпкиным [128] и Ю. М. Ермольевым [41], она связана с отслеживанием минимума нестационарной квадратичной функции. Изложение материала § 26, 27 основано на работе [22]. Скорость сходимости стохастического метода условного градиента исследовал С. В. Пашко [93]. Теоремы 26.1—26.3 можно также получить как следствие результатов работы [45].

§ 28. Рассматривается стохастический вариант конечно-разностного метода Эрроу — Гурвица, изложенного в § 20.

СПИСОК ЛИТЕРАТУРЫ

1. Алгоритмы оптимизации проектных решений /Под ред. А. И. Половинкина.— М. : Сов. радио, 1976.
2. Александров П. С. Введение в теорию множеств и общую топологию.— М. : Наука, 1977.
3. Ащепков А. Г., Белов Б. И., Булатов В. П. и др. Методы решения задач математического программирования и оптимального управления.— Новосибирск: Наука, — 1984.
4. Баженов Л. Г. Об условиях сходимости метода минимизации почти-дифференцируемых функций // Кибернетика.— 1972.— № 4.— С. 71—72.
5. Баженов Л. Г. Необходимые и достаточные условия сходимости итеративных процедур // Докл. АН УССР. Сер. А.— 1985.— № 2.— С. 64—65.
6. Базара М., Шетти К. Нелинейное программирование. Теория и алгоритмы.— М. : Мир, 1982.
7 Батищев Д. И. Методы оптимального проектирования.— М. : Радио и связь, 1984.
8. Булатов В. П. Методы погружения в задачах оптимизации.— Новосибирск: Наука, 1977.
9. Вазан М. Стохастическая аппроксимация.— М. : Мир, 1972.
10. Варга Дж. Оптимальное управление дифференциальными и функциональными уравнениями.— М. : Наука, 1977.
11. Васильев Ф. П. Численные методы решения экстремальных задач.— М.: Наука, 1980.
12. Вилков А. В., Жидков Н. П., Щедрин Б. М. Методы отыскания глобального минимума функции одного переменного // ЖВМ и МФ.— 1975.— Т. 15, № 4.— С. 1040—1042.
13. Вычислительные методы выбора оптимальных проектных решений / Под ред. В. С. Михалевича.— Киев : Наукова думка, 1977.
14. Гельфанд И. М., Цетлин М. Л. Принципы нелокального поиска в системах автоматической оптимизации // Докл. АН СССР.— 1961.— Т. 137, № 2.— С. 295—298.
15. Гихман И. И., Скороход А. В. Введение в теорию случайных процессов.— М. : Наука, 1977.
16. Глушков В. М. Основы безбумажной информатики.— М. : Наука, 1982.
17. Глушкова О. В., Гупал А. М. Численные методы минимизации функций максимума без вычисления градиентов // Кибернетика.— 1980.— № 5.— С. 141—143.
18. Гольштейн Е. Г. Обобщенный градиентный метод отыскания седловых точек // Эконом. и мат. методы.— 1972.— Т. 8, № 4.— С. 569—579.
19. Горбачук В. И. Общие условия сходимости алгоритмов к произвольно указанным множествам // Кибернетика.— 1983.— № 6.— С. 112—113.
20. Гупал А. М., Баженов Л. Г. Стохастический аналог метода сопряженных градиентов // Кибернетика.— 1972.— № 1.— С. 125—126.
21. Гупал А. М., Баженов Л. Г. Стохастический метод линеаризации // Кибернетика.— 1972.— № 3.— С. 116—117.
22. Гупал А. М. Стохастические методы решения негладких экстремальных задач.— Киев : Наукова думка, 1979.
23. Гупал А. М. Алгоритмы поиска экстремума недифференцируемых функций с ограничениями / ИК АН УССР.— Киев, 1976, 18 с.— Препринт 76—8.

24. Гупал А. М. Об одном методе минимизации почти-дифференцируемых функций // Кибернетика.— 1977.— № 2.— С. 73—75.
25. Гупал А. М. Метод минимизации функций, удовлетворяющих условию Липшица // Кибернетика.— 1980.— № 5.— С. 91—92.
26. Гупал А. М., Дубровский В. Б. Конечно-разностный метод Эрроу — Гурвица с усреднением // Кибернетика.— 1985.— № 4.
27. Гупал А. М., Лоскутов В. Г. Эффективность конечно-разностных методов минимизации недифференцируемых функций // Кибернетика.— 1982.— № 2.— С. 115—117.
28. Гурин Л. С., Лобач В. П. Комбинация метода Монте-Карло с методом скорейшего спуска при решении некоторых экстремальных задач // ЖВМ и МФ.— 1962.— Т. 2, № 3.— С. 499—502.
29. Данилин Ю. М., Пиявский С. А. Об одном алгоритме отыскания абсолютного минимума // Теория оптимальных решений. Вып. 2.— Киев : ИК АН УССР, 1967.— С. 25—37.
30. Данилин Ю. М. Оценка эффективности одного алгоритма отыскания абсолютного минимума // ЖВМ и МФ.— 1971.— Т. 11, № 4.— С. 1026—1031.
31. Данскин Дж. М. Теория максимина.— М. : Сов. радио, 1970.
32. Данциг Дж. Линейное программирование, его обобщения и применения.— М. : Прогресс, 1966.
33. Демьянов В. Ф., Малоземов В. Н. Введение в минимакс.— М. : Наука, 1972.
34. Демьянов В. Ф., Васильев Л. В. Недифференцируемая оптимизация.— М. : Наука, 1981.
35. Дорофеев П. А. О некоторых свойствах метода обобщенного градиента // ЖВМ и МФ.— 1985.— Т. 25, № 2.— С. 181—189.
36. Евтушенко Ю. Г. Методы решения экстремальных задач и их применение в системах оптимизации.— М. : Наука, 1982.
37. Еремин И. И. Метод штрафов в выпуклом программировании // Кибернетика.— 1967.— № 4.— С. 63—67.
38. Еремин И. И., Астафьев Н. Н. Введение в теорию линейного и выпуклого программирования.— М. : Наука, 1976.
39. Еремин И. И., Мазуров В. Д., Астафьев Н. Н. Несобственные задачи линейного и выпуклого программирования.— М. : Наука, 1983.
40. Ермольев Ю. М. Методы решения нелинейных экстремальных задач // Кибернетика.— 1966.— № 4.— С. 1—17.
41. Ермольев Ю. М. Методы стохастического программирования.— М. : Наука, 1976.
42. Ермольев Ю. М., Верченко П. И. О методе линеаризации в предельных экстремальных задачах // Кибернетика.— 1976.— № 2.— С. 65—69.
43. Ермольев Ю. С., Ястремский А. И. Стохастические модели и методы в экономическом планировании.— М. : Наука, 1979.
44. Завриев С. К. Стохастические градиентные методы решения минимаксных задач.— М. : Изд-во МГУ, 1984.
45. Завриев С. К. Об одном аналоге усиленного закона больших чисел для сумм зависимых случайных величин // Программное обеспечение и модели исследования операций / Под ред. Л. Н. Королева и П. С. Краснощекова.— М. : Изд-во МГУ, 1985.
46. Зангвилл У. Нелинейное программирование: единый подход.— М. : Сов. радио, 1973.
47 Зойтендейк Г. Методы возможных направлений.— М. : ИЛ, 1963.
48 Иоффе А. Д., Левин В. Л. Субдифференциалы выпуклых функций // Труды ММО, т. 26.— М. : Наука, 1972.— С. 3—73.
49. Иоффе А. Д., Тихомиров В. М. Теория экстремальных задач.— М. : Наука, 1974.
50. Итеративные методы в теории игр и программировании / Под ред. В. З. Беленького и В. А. Волконского.— М. : Наука, 1979.
51. Карманов В. Г. Математическое программирование.— М. : Наука, 1979.
52. Катковник В. Я. Линейные оценки и стохастические задачи оптимизации.— М. : Наука, 1976.

53. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа.— М. : Наука, 1968.
54. Красовский Н. Н. Теория управления движением.— М. : Наука, 1968.
55. Кружков С. Н. К вопросу о дифференцируемости почти всюду функций многих переменных // Вестник МГУ. — 1976.— № 6.— С. 67—70.
56. Левин А. Ю. Об одном алгоритме минимизации выпуклых функций // Докл. АН СССР.— 1965.— Т. 160, № 6.— С. 1244—1247.
57. Левин В. Л. Измеримые сечения многозначных отображений в топологические пространства и верхние огибающие интегрантов Каратеодори // Докл. АН СССР.— Т. 252, № 3.— С. 535—539.
58. Лоэв М. Теория вероятностей.— М. : ИЛ, 1962.
59. Ляшко С. И. Метод линеаризации при решении экстремальных задач для функций с неизвестным параметром // Математические методы исследования операций и теории надежности.— Киев : ИК АН УССР, 1978.— С. 13—20.
60. Макаров В. Л., Рубинов А. М. Математическая теория экономической динамики и равновесия.— М. : Наука, 1973.
61. Михалевич В. С. Методы последовательной оптимизации // Кибернетика.— 1965.— № 1.— С. 45—55.
62. Михалевич В. С., Волкович В. Л. Вычислительные методы исследования и проектирования сложных систем.— М. : Наука, 1982.
63. Михалевич В. С., Гупал А. М., Болгарин И. В. Численные методы решения нестационарных и предельных экстремальных задач без вычисления градиентов // Кибернетика.— 1984.— № 5.— С. 56—57.
64. Михалевич В. С., Гупал А. М., Дубровский В. Б. Глобальный характер точных негладких штрафов // Кибернетика.— 1985.— № 4.
65. Михалевич М. В. Обобщенный стохастический метод центров // Кибернетика.— 1980.— № 2.— С. 122—125.
66. Михалевич М. В. Об устойчивости методов стохастического программирования к ошибкам вычисления стохастических квазиградиентов // Кибернетика.— 1980.— № 3.— С. 117—119.
67. Моисеев Н. Н. Элементы теории оптимальных систем.— М. : Наука, 1975.
68. Моисеев Н. Н., Иванилов Ю. П., Столярова Е. М. Методы оптимизации.— М. : Наука, 1978.
69. Моцкус И. Б. Многоэкстремальные задачи в проектировании.— М. : Наука, 1967.
70. Мураускас Г. Г. Исследование скорости сходимости итеративных стохастических алгоритмов оптимизации и оценивания: Автореф. дис. ... канд. физ.-мат. наук.— Киев : ИК АН УССР, 1983.
71. Муртаф Б. Современное линейное программирование.— М. : Мир, 1984.
72. Невельсон М. Б., Хасьминский Р. З. Стохастическая аппроксимация и рекуррентное оценивание.— М. : Наука, 1972.
73. Немировский А. С., Юдин Д. Б. Сложность задач и эффективность методов оптимизации.— М. : Наука, 1979.
74. Нестеров Ю. Е. Метод решения задачи выпуклого программирования со скоростью сходимости $O(1/k^2)$ // Докл. АН СССР.— 1983.— Т. 269, № 3.— С. 543—547.
75. Нестеров Ю. Е. Методы минимизации негладких выпуклых и квазивыпуклых функций // Эконом. и мат. методы.— 1984.— Т. 20, № 3.— С. 519—531.
76. Нестеров Ю. Е. Об одном классе методов безусловной минимизации выпуклой функции, обладающих высокой скоростью сходимости // ЖВМ и МФ.— 1984.— Т. 24, № 7.— С. 1090—1093.
77. Никайдо Х. Выпуклые структуры и математическая экономика.— М. : Мир, 1972.
78. Николаева Н. Д. Об одном алгоритме для решения задач выпуклого программирования // Эконом. и мат. методы.— 1974.— Т. 10, № 5.— С. 941—946.
79. Норкин В. И. О случайных липшицевых функциях // Кибернетика.— 1986.— № 2.
80. Норкин В. И. Нелокальные алгоритмы минимизации недифференцируемых функций // Кибернетика.— 1978.— № 5.— С. 44—48.
81. Норкин В. И. Два замечания: о методе сглаживания в многоэкстремальной оптимизации и о конечно-разностном методе в недифференцируемой оптимиза-

ции // Численные методы нелинейного программирования. — Харьков: ХГУ.— 1979.
82. Норкин В. И. Об условиях сходимости комбинированных алгоритмов нелинейного программирования // Кибернетика.— 1979.— № 6.— С. 73—77.
83. Норкин В. И. Обобщенно дифференцируемые функции // Кибернетика.— 1980.— № 1.— С. 9—11.
84. Норкин В. И. Метод минимизации недифференцируемых функций с усреднением обобщенных градиентов // Кибернетика.— 1980.— № 6.— С. 88—89.
85. Норкин В. И. Об устойчивости метода обобщенного градиентного спуска // Кибернетика.— 1985.— № 4.
86. Норкин В. И. Обобщенный градиентный метод в задаче невыпуклой негладкой оптимизации: Автореф. дис. ...канд. физ.-мат. наук.— Киев : ИК АН УССР.— 1983.
87. Нурминский Е. А. Условия сходимости алгоритмов нелинейного программирования // Кибернетика.— 1972.— № 6.— С. 79—81.
88. Нурминский Е. А. Квазиградиентный метод решения задачи нелинейного программирования // Кибернетика.— 1973.— № 1.— С. 122—125.
89. Нурминский Е. А. Сходимость стохастических аналогов детерминированных процедур нелинейного программирования // Теория оптимальных решений — Киев : ИК АН УССР, 1973.— С. 60—73.
90. Нурминский Е. А., Желиховский А. А. Исследование одной регулировки шага в квазиградиентом методе минимизации слабо выпуклых функций // Кибернетика.— 1974.— № 6.— С. 101—105.
91. Нурминский Е. А. Численные методы решения детерминированных и стохастических минимаксных задач.— Киев : Наукова думка, 1979.
92. Ортега Дж., Рейнболдт В. Итерационные методы решения нелинейных систем уравнений со многими неизвестными.— М. : Мир, 1975.
93. Пашко С. В. О скорости сходимости стохастического метода линеаризации // Кибернетика.— 1984.— № 4.
94. Петерсен И. Статистическая оптимизация посредством сглаживания // Изв. АН СССР. Техн. кибернетика.— 1969.— № 2.
95. Пиявский С. А. Алгоритм отыскания абсолютного минимума функций // Теория оптимальных решений. Вып. 2.— Киев : ИК АН УССР, 1967.— С. 13—24.
96. Пиявский С. А. Один алгоритм отыскания абсолютного экстремума функции // ЖВМ и МФ.—1972.—Т. 12, № 4.— С. 888—896.
97. Подиновский В. В., Гаврилов В. М. Оптимизация по последовательно применяемым критериям.— М. : Сов. радио, 1975.
98 Полак Э. Численные методы оптимизации. Единый подход.— М. : Мир, 1974.
99. Поляк Б. Т. Один общий метод решения экстремальных задач // Докл. АН СССР.— 1967.— Т. 174, № 1.— С. 33—36.
100. Поляк Б. Т. Введение в оптимизацию.— М. : Наука, 1983.
101. Понтрягин Л. С., Болтянский В. Г., Гамкрелидзе Р. В., Мищенко Е. Ф. Математическая теория оптимальных процессов.— М. : Физматгиз, 1961.
102. Прохоров Б. Н., Розанов Ю. А. Теория вероятностей.— М. : Наука, 1967.
103. Пшеничный Б. Н. Необходимые условия экстремума.— М. : Наука, 1969.
104. Пшеничный Б. Н. Выпуклый анализ и экстремальные задачи.— М. : Наука, 1980.
105. Пшеничный Б. Н., Данилин Ю. М. Численные методы в экстремальных задачах.— М. : Наука, 1975.
106. Растригин Л. А. Системы экстремального управления.— М. : Наука, 1974.
107. Ржевский С. В. ε-субградиентный метод решения задачи выпуклого программирования // ЖВМ и МФ.— 1981.— Т. 21, № 5.— С. 1126—1132.
108. Ржевский С. В. Монотонный алгоритм поиска седловой точки негладкой функции // Кибернетика.— 1982.— № 1.— С. 95—98.

109. Рокафеллар Р. Выпуклый анализ.— М. : Мир, 1973.
110. Рокафеллар Р. Выпуклые интегральные функционалы и двойственность // Математическая экономика.— М. : Мир, 1974.— С. 222—237.
111. Самарский А. А. Теория разностных схем.— М. : Наука, 1977.
112. Саридис Дж. Самоорганизующиеся стохастические системы управления.— М. : Мир, 1980.
113. Стернберг С. Лекции по дифференциальной геометрии.— М. : Мир, 1970.
114. Стронгин Р. Г. Численные методы в многоэкстремальных задачах.— М. : Наука, 1978.
115. Стронгин Р. Г., Маркин Д. Л. Алгоритм решения многоэкстремальных задач с нелинейными невыпуклыми ограничениями // Методы математического программирования и программное обеспечение.— Свердловск: ИММ УНЦ АН СССР, 1984, с. 103—104.
116. Сухарев А. Г. Оптимальный поиск экстремума.— М. : Изд-во МГУ, 1975.
117. Сухарев А. Г. Глобальный экстремум и методы его отыскания // Математические методы исследования операций / Под ред. Н. Н. Моисеева и П. С. Краснощекова.— М. : Изд-во МГУ, 1981.— С. 4—37.
118. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач.— М. : Наука, 1979.
119. Урясьев С. П., Мирзоахмедов Ф. Регулировки шага в методах стохастического программирования // ЖВМ и МФ.— 1983.— № 6.— С. 1314—1325.
120. Фаддеев Д. К., Фаддеева В. Н. Вычислительные методы линейной алгебры.— М. : Физматгиз, 1960.
121. Федоров В. В. О задачах оптимизации с упорядоченной совокупностью ограничений // ЖВМ и МФ.— 1975.— Т. 15, № 5.— С. 1126—1137.
122. Федоров В. В. Численные методы максимина.— М. : Наука, 1979.
123. Фиакко А., Мак-Кормик Д. П. Нелинейное программирование. Методы последовательной безусловной минимизации.— М. : Мир, 1972.
124. Хачиян Л. Г. Полиномиальный алгоритм в линейном программировании // Докл. АН СССР.— 1979.— Т. 244, № 5.— С. 1093—1096.
125. Химмельблау Д. Прикладное нелинейное программирование. — М. : Мир, 1975.
126. Хоанг Туй. Вогнутое программирование при линейных ограничениях // Докл. АН СССР.— 1964.— Т. 159, № 1.— С. 32—35.
127. Цыпкин Я. З. Адаптация и обучение в автоматизированных системах.— М. : Наука, 1968.
128. Цыпкин Я. З. Обобщенные алгоритмы обучения // Автоматика и телемеханика.— 1976.— № 1.— С. 97—103.
129. Чепурной Н. Д. Метод усредненных квазиградиентов со ступенчатой регулировкой шага для минимизации слабо выпуклых функций // Кибернетика.— 1981.— № 6.— С. 131.
130. Чепурной Н. Д. Об одном релаксационном алгоритме минимизации негладких функций // Кибернетика.— 1984.— № 2.— С. 110—111.
131. Численные методы условной оптимизации / Под ред. Ф. Гилла и У. Мюррея — М. : Мир, 1977.
132. Шепилов М. А. О методе обобщенного градиента отыскания абсолютного минимума выпуклой функции // Кибернетика.— 1976.— № 4.— С. 52—57.
133. Ширяев А. Н. Вероятность.— М. : Наука, 1980.
134. Шор Н. З. Применение метода градиентного спуска для решения сетевой транспортной задачи // Материалы науч. семинара по теорет. и прикл. вопр. кибернетики и исслед. операций.— Киев: ИК АН УССР, 1962, вып. 1.— С. 9—17.
135. Шор Н. З. Метод отсечения с растяжением пространства для решения задач выпуклого программирования // Кибернетика.— 1977.— № 1.— С. 94—95.
136. Шор Н. З. Методы минимизации недифференцируемых функций и их приложения.— Киев : Наукова думка, 1979.
137. Эрроу К. Дж., Гурвиц Л., Удзава Х. Исследования по линейному и нелинейному программированию.— М. : ИЛ, 1962.
138. Юдин Д. Б. Методы количественного анализа сложных систем // Изв. АН СССР. Техн. кибернетика.— 1965.— № 1.— С. 3—13.

139. Юдин Д. Б. Математические методы управления в условиях неполной информации.— М. : Сов. радио, 1974.
140. Юдин Д. Б. Задачи и методы стохастического программирования.— М. : Сов. радио, 1979.
141. Юдин Д. Б., Немировский А. С. Информационная сложность и эффективные методы решения выпуклых экстремальных задач // Эконом. и мат. методы.— 1976.— Т. 12, № 2.— С. 357—369.
142. Bazaraa M. S., Good J. J. Sufficient conditions for a globally exact penalty function without convexity // Math. Progr. Study 19.— 1982.— V. 19.— P. 1—15.
143. Bihain A. Optimization of upper semidifferentiable functions // J. Optimiz. Theory Appl.— 1984.— V. 44.— P. 545—568.
144. Castaing C. Sur le multi-applications measurables // Revue Française d'Informatique et de Recherche Operationnelle.— 1967.— № 1.— P. 91—126.
145. Clarke F. H. Generalized gradients and applications // Trans. Amer. Math. Soc.— 1975.— V. 205.— P. 247—262.
146. Clarke F. H. A new approach to Lagrange multipliers // Math. Oper. Res.— 1976.— V. 1, № 2.— P. 165—174.
147. Debreu G. Integration of correspondences // Proc. 5-th Berckeley Symp. on Math. Stat.— Prob. 11, part 1.— P. 351—372.
148. Frank M., Wolfe P. An algorithm for guadratic programming // Naval Res. Log. Quart.— 1956.— V. 3, № 1—2.— P. 113—120.
149. Fubian V. Stochastic approximations of constrained minima // Trans. 4-th Prague Conf. Inf. Theory. Stat. Dec. Funct. Random Process, 1965.— Prague.— 1967.— P. 277—290.
150. Hien Nguyen V., Strodiot J. J., Mifflin R. On conditions to have bounded multipliers in locally lipschitz programming // Math. Progr.— 1980. — V. 18, № 1.— P. 100—106.
151. Himelberg C. J. Measurable relations // Fundamenta Mathematicae.— 1975.— V. 87, № 1.— P. 53—72.
152. Hiriart-Urruty J. B. On optimality conditions in nondifferentiable programming // Math. Progr.— 1978.— V. 14.— P. 73—86.
153. Hiriart-Urruty J. B. Refinements of necessary optimality conditions in nondifferentiable programming. I // Appl. Math. Optim.— 1979.— № 5.— P. 63—82; II // Math. Progr. Study 19.— 1982.— P. 120—139.
154. Hoang Tuy. Concave Minimization under Linear Constraints with Special Structure // Optimization.— 1985.— V. 16, № 3.— P. 335—352.
155. Kiefer J., Wolfowitz J. Stochastic estimation of the maximum of a regression function // Ann. Math. Stat.— 1952.— V. 23, № 3.— P. 462—466.
156. Kiviel K. C. An aggregate subgradient method for nonsmooth convex minimization // Math. Progr.— 1983.— V. 27.— P. 320—341.
157. Kiviel K. C. Methods of Descent for Nondifferentiable Optimization // Lecture Notes in Mathematics.— 1985.— V. 1133.— P. 362.
158. Kushner H. J., Clark D. S. Stochastic approximation methods for constrained and unconstrained systems.— N. Y., Heidelberg, Berlin: Springer, 1978.
159. Lebourg M. G. Valeur moyenne pour gradient generalize // C. R. Acad. Sc Paris.— 1975.— V. 281, ser. A, № 19.— P. 775—779.
160. Lemarechal C. An extension of Davidon methods to nondifferentiable problems // Math. Progr. Study 3.— 1975.— P. 95—109.
161. Lemarechal C., Strodiot J. J., Bihain A. On a bundle algorithm for nonsmooth minimization: Nonlinear Progr. 4/Eds O. L. Mangasarian, R. R. Meyer, S. M. Robinson.— N. Y.: Academic Press, 1981.— P. 245—281.
162. Leuenberger D. G. Quasi-convex programming // SIAM J. Appl. Math.— 1968.— V. 16.— P. 1090—1095.
163. Mangasarian O. L. Pseudo-convex functions // SIAM J. Control.— 1965.— V. 3, № 3.— P. 281—290.
164. Mangasarian O. L. Exact penalty functions in nonlinear programming // Math. Progr.— 1979.— № 17.— P. 251—269.

165. McCormick G. P. Attempts to calculate global solutions of problems that may have local minima: Numerical Methods for Nonlinear Optimization/ Ed. F. Lootsma.— Academic Press, London, N. Y. 1972.— P. 209—221.
166. Mifflin R. An algorithm for constrained optimization with semismooth functions // Math. Oper. Res.— 1977.— № 2.— P. 191—207.
167. Mifflin R. Semismooth and semiconvex functions in constrained optimization // SIAM J. Contr. and Optim.— 1977.— V. 15, № 6.— P. 959—972.
168. Mifflin R. A modification and an extentions of Lemareshal's algorithm for nonsmooth minimization // Math. Progr. Study 17.— 1982.— P. 77—90.
169. Mirica S. A note on the generalized differentiability of mappings // Nonl. Anal. Theory, Methods, Applications.— 1980.— V. 4, № 3.— P. 567—575.
170. Murtagh B. A., Saunders M. Large scale linearly constrained optimization // Math. Progr.— 1978.— V. 14.— P. 41—72.
171. Nurminski E. A. On ε-subgradient methods of non-differentiable optimization // Lect. Notes in Control and Information Sciences.— 1979.— № 14.— P. 187—195.
172. Polak E., Mayne D. Q. Algorithm models for nondifferentiable optimization // SIAM J. Control and Optimization.— 1985.— V. 23, № 3.— P. 477—491.
173. Rademacher H. Uber partielle und total differenzierbarkeit von funktionen mehrerer variabeln und über die transformation der doppelintegrale // Math. Annalen.— 1919.— V. 79.— P. 340—359.
174. Robbins H., Monro S. A stochastic approximation method // Ann. Math. Stat.— 1951.— V. 22, N 3.— P. 400—407.
175. Rockafellar R. T. Measurable dependence of convex sets and functions on parametries // J. Math. Anal. Appl.— 1969.— V. 28.— P. 4—25.
176. Rockafellar R. T. Directionally lipschitzian functions and subdifferential calculus // Proc. London Math. Soc.— 1979.— V. 39.— P. 331—355.
177. Rockafellar R. T. Generalized directional derivatives and subgradients of nonconvex functions // Canad. J. Math.— 1980.— V. 32.— P. 257—280.
178. Rockafellar R. T. The theory of subgradients and its applications to problems of optimization. Convex and nonconvex functions. Research Notes in Mathematics 1 / Eds K. H. Hoffman, R. Wille.— Berlin: Heldermann.— 1981.
179. Strodiot J. J., Hien Nguyen V., Heukemez N. ε-optimal solution in nondifferentiable convex programming and related questions // Math. Prog.— 1982.— V. 25.— P. 307—323.
180. Thibault Z. Quelques proprietés des sous-differentiels de fonctions reels localment lypchitziennes definies sur un espace de Banach separable // C. R. Acad. Sc. Paris.— 1976.— V. 282, Ser. A, № 10.— P. 507—610.
181. Towards global optimization / Eds L. C. W. Dixon, G. P. Szego.— Amsterdam : North-Holland, 1975.
182. Wagner D. H. Survey of measurable selection theorems // SIAM J. Control and Opt.— 1977.— V. 15, № 5.— P. 859—903.
183. Wets R. Stochastic programming. Solution techniques and approximation schemes // Math. Program. State art. 11-th Int. Symp. Bonn., 23—27 Aug. 23—27 1982. Berlin e. a. 1983.— P. 566—603.
184. Wolfe P. Methods for nonlinear constraints.— Heidelberg: North-Holland, 1967.— P. 120—131.
185. Wolfe P. A method of conjugate subgradients for minimizing nondifferentiable functions // Math. Progr. Study 3.— 1975.— P. 145—173.
186. Wolfe P. Finding a nearest point in a polytope // Math. Progr.— 1976.— V. 11.— P. 128—149.
187. Zangwill W. I. Nonlinear programming via penalty functions // Management Science.— 1967.— V. 13.— P. 344—388.
188. Zwart P. B. Global maximization of a convex function with linear inequality constraints // Operat. Res.— 1974.— V. 22, № 3.— P. 602—609.

Владимир Сергеевич Михалевич
Анатолий Михайлович Гупал
Владимир Иванович Норкин

МЕТОДЫ НЕВЫПУКЛОЙ ОПТИМИЗАЦИИ

Редактор *Г. П. Грошев*
Художественный редактор *Г. М. Коровина*
Технический редактор *С. Я. Шкляр*
Корректоры *И. Я. Кришталь, Н. Д. Храпко*

ИБ 32283

Сдано в набор 25.08.86. Подписано к печати 10.09.87. Т-18868. Формат 60×90/16. Бумага тип. № 1. Гарнитура литературная. Печать высокая. Усл. печ. л. 17,5. Усл. кр.-отт. 17,5. Уч.-изд. л. 19,24. Тираж 4900 экз. Заказ 7-105. Цена 3 р. 10 к.

Ордена Трудового Красного Знамени
издательство «Наука»
Главная редакция физико-математической литературы
117071 Москва В-71, Ленинский проспект, 15

Киевская книжная типография научной книги.
252004 Киев 4, ул. Репина, 4