

ISBN 978-5-91601-082-4

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ  
БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ  
ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР ИМЕНИ А.А. ДОРОДНИЦЫНА  
РОССИЙСКОЙ АКАДЕМИИ НАУК

Ю.Г. ЕВТУШЕНКО

ОПТИМИЗАЦИЯ  
И БЫСТРОЕ АВТОМАТИЧЕСКОЕ  
ДИФФЕРЕНЦИРОВАНИЕ



ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР ИМЕНИ А.А. ДОРОДНИЦЫНА РАН

МОСКВА 2013

## УДК 519.626.2

Ответственный редактор  
доктор физ.-матем. наук В.И. Зубов

Излагаются основные результаты теории нелинейного программирования (НЛП), приведены теоремы Лагранжа и Куна-Таккера. Описан основной подход к численному решению динамических оптимизационных задач, базирующийся на редукции задач оптимального управления (ОУ) к задачам нелинейного программирования. Изложен общий подход к дифференцированию сложных функций, возникающих в многошаговых процессах, и показано, что из найденных результатов следуют формулы быстрого автоматического дифференцирования (БАД). Во многих случаях БАД значительно превосходит символическое дифференцирование или аппроксимацию производных с помощью разделенных разностей. Методы решения задач ОУ, основанные на идеях НЛП, оказались чрезвычайно эффективными по многим причинам: с их помощью стали очевидными многие ранее предложенные эвристические алгоритмы, возникла возможность их обобщения, они позволили использовать богатый, всесторонне развитый арсенал методов НЛП, создали конструктивную основу для построения методов оптимизации динамических систем, проинтегрированных по схемам высокого порядка точности, позволили решать сложные задачи ОУ со смешанными ограничениями. Этот подход оказался применим к оптимизации распределенных динамических систем. Идущие от НЛП численные методы приводят к близким схемам оптимизации систем, описываемых как обыкновенными дифференциальными уравнениями, так и уравнениями с частными производными. В обоих случаях после дискретизации получаются однотипные многошаговые процессы, к которым в одинаковой степени применима техника НЛП. Полученные в этом случае задачи обладают (в отличие от обычных задач НЛП) двумя замечательными свойствами: в них сравнительно просто и точно вычисляются производные целевых функций. Этим объясняется высокая эффективность градиентной техники оптимизации с использованием БАД. В качестве простейшего примера приведено решение задачи Бюргерса.

Работа выполнена при финансовой поддержке РФФИ (№ 11-026-12136-офи-м), программ фундаментальных исследований Президиума РАН П-15 и П-18 и по программе поддержки ведущих научных школ (НШ 5264.2012.1).

Рецензенты: А.И. Голиков,  
М.А. Посыпкин

Научное издание

© Федеральное государственное бюджетное учреждение науки  
Вычислительный центр им. А.А. Дородницына  
Российской академии наук, 2013

## Основные обозначения и аббревиатуры

$\mathbb{R}$	– множество вещественных чисел;
$\mathbb{R}^n$	– множество $n$ -мерных вещественных векторов;
$\mathbb{R}^{m \times n}$	– множество прямоугольных матриц размера $m \times n$ ;
$\mathbb{R}_+^n$	– множество $n$ -мерных вещественных векторов, все компоненты которых неотрицательные;
$\emptyset$	– пустое множество;
$0_n$	– нулевой $n$ -мерный вектор-столбец;
$0_n^\top$	– нулевая $n$ -мерная вектор-строка;
$0_{mn}$	– нулевая матрица размера $m \times n$ ;
$e_s$	– $s$ -мерный вектор, все компоненты которого равны единице;
$I_s$	– единичная $s \times s$ матрица;
$\langle a, b \rangle = a^\top b$	– скалярное произведение векторов $a$ и $b$ одинаковой размерности;
$\ x\ $	– евклидова норма вектора $x$ ; если $x \in \mathbb{R}^n$ , то $\ x\  = \left[ \sum_{i=1}^n (x^i)^2 \right]^{1/2}$ ;
$\ x\ _\gamma$	– гельдеровская норма вектора $x$ ; если $\gamma > 0$ , $x \in \mathbb{R}^n$ , то
	– $\ x\ _\gamma = \left[ \sum_{i=1}^n (x^i)^\gamma \right]^{1/\gamma}$ ;
$\ x\ _\infty$	– чебышевская норма, $\ x\ _\infty = \max_{1 \leq i \leq n}  x^i $ ;
$\ x\ _1$	– октаэдрическая норма, $\ x\ _1 = \sum_{i=1}^n  x^i $ ;
$D(x)$	– диагональная матрица, у которой $i$ -й диагональный элемент равен $x^i$ ;
$\text{epi } F$	– эпиграф (надграфик) функции $F$ ;
$A^\top$	– матрица, транспонированная к матрице $A$ ;
$\text{im } A$	– пространство столбцов матрицы $A$ , т.е. для матрицы $A$ размера $m \times n$ имеем $\text{im } A = \{Ax : x \in \mathbb{R}^n\}$ ;
$\text{ker } A$	– нуль-пространство матрицы $A$ , т.е. для прямоугольной матрицы $A$ размера $m \times n$ имеем $\text{ker } A = \{x \in \mathbb{R}^n : Ax = 0_m\}$ ;
$\text{pr}(x, X)$	– проекция вектора $x \in \mathbb{R}^i$ на множество $X \subset \mathbb{R}^i$ ;
$\text{dist}(x, X)$	– расстояние от точки $x$ до множества $X$ ;
$A^\parallel$	– оператор проектирования на подпространство столбцов матрицы $A$ (см. Приложение 1);
$A^\perp$	– оператор проектирования на ортогональное дополнение к подпространству столбцов матрицы $A$ (см. Приложение 1);
$A > 0_{nn}$	– если $A$ — квадратная матрица размера $n$ , то $A$ положительно определена, т.е. $z^\top A z > 0$ для любых $z \in \mathbb{R}^n$ , имеющих ненулевую норму;
$A \geq 0_{nn}$	– если $A$ — квадратная матрица размера $n$ , то $A$ положительно полуопределена, т.е. $z^\top A z \geq 0$ для любых $z \in \mathbb{R}^n$ ;
$C^i$	– класс $i$ раз непрерывно дифференцируемых функций;
$x \in S$	– $x$ принадлежит множеству $S$ ;
$T \subset S$	– множество $T$ содержится в множестве $S$ ;
$\text{int } U$	– внутренняя часть множества $U$ ;

- $V \cup W$  – объединение множеств  $V$  и  $W$ , т.е.  $V \cup W = \{x : \text{либо } x \in V, \text{ либо } x \in W\}$ ;
- $V \cap W$  – пересечение множеств  $V$  и  $W$ , т.е.  $V \cap W = \{x : x \in V \text{ и } x \in W\}$ ;
- $V \times W$  – декартово произведение множеств  $V$  и  $W$ , т.е.  $V \times W = \{[v, w] : v \in V, w \in W\}$ ;
- $x_+$  – вектор,  $i$ -я компонента которого равна  $x^i$ , если  $x^i \geq 0$ , и равна нулю, если  $x^i \leq 0$ ;
- $x \geq y$  – для каждой  $i$ -й компоненты векторов  $x$  и  $y$  имеет место неравенство  $x^i \geq y^i$ ;
- $\nabla\Omega(u)$  – градиент функции  $\Omega(u)$  (вектор-столбец),  $i$ -я компонента которого равна  $d\Omega(u)/du^i$ ;
- $\Delta\Omega(u)$  – матрица вторых производных функции  $\Omega(u)$ ;
- $g_{x^\top}$  – прямоугольная матрица Якоби,  $(ij)$ -элемент которой вычисляется по формуле  $\partial g^i / \partial x^j$ ,  $g_{x^\top} = (g_x^\top)^\top$ ;
- $g_x^\top$  – прямоугольная матрица,  $(ij)$ -элемент которой вычисляется по формуле  $\partial g^j / \partial x^i$ ,  $(g_x^\top)^\top = g_{x^\top}$ ;
- $L$  – функция Лагранжа;
- БАД – быстрое автоматическое дифференцирование;
- БМ – безусловная минимизация;
- ЛП – линейное программирование;
- МП – математическое программирование;
- НЛП – нелинейное программирование;
- ОУ – оптимальное управление;
- у.р.о. – условие регулярности ограничений;
- – окончание доказательства;
- – окончание примера.

## Введение

Вопросы оптимизации возникали на протяжении всей истории человечества. Леонард Эйлер писал: “В мире не происходит ничего, в чем бы не был виден смысл какого-нибудь максимума или минимума”. В древнем мире были математически сформулированы первые оптимизационные задачи (задачи Евклида, Архимеда, Аполлония, Дидоны и многие другие). С развитием цивилизации оптимизационные задачи усложнялись, охватывая все новые и новые области человеческой деятельности.

Теория оптимизации создавалась на протяжении последних четырехсот лет. Первые важные теоретические результаты в этой области были получены в XVII веке П. Ферма, Г. Лейбницем, И. Ньютоном, братьями Бернулли. В дальнейшем, в XVIII и XIX веках, значительный вклад в теорию оптимизации внесли Л. Эйлер, Ж. Лагранж, А. Лежандр, Ж. Фурье, К. Вейерштрасс, В. Гамильтон, С. Якоби, А. Пуанкаре, М.В. Остроградский, П.Л. Чебышев и многие другие.

Особенно интенсивные работы по созданию теории и разработке численных методов решения задач оптимизации начались в середине прошлого столетия. Эти исследования были обусловлены двумя причинами: во-первых, в это время были созданы первые электронно-вычислительные машины, дававшие возможность реализовывать многочисленные методы; во-вторых, проектирование сложных технических систем, конструирование ракет, самолетов, объектов атомной энергетики и многих других технических систем привели к необходимости решения разнообразных практических задач оптимизации. Не только технические задачи, но и многие научные проблемы требовали для своего решения использования эффективных численных методов оптимизации. Укажем, например, разнообразные обратные задачи, возникающие при обработке геофизических наблюдений, при моделировании процессов атмосферной циркуляции, физической метрологии, океанографии и т.д.

За последние сто лет была создана теоретическая база оптимизации, определились ее главные направления: линейное и нелинейное программирование (ЛП и НЛП), оптимальное управление (ОУ), дискретная оптимизация, глобальная оптимизация, многокритериальная оптимизация. Процесс разработки численных методов продолжается, он обусловлен большим спросом со стороны многочисленных предприятий, создающих и совершенствующих различные системы, их стремлением максимально повысить эффективность проектируемых изделий. Многие методы оптимизации были реализованы и собраны в разнообразных библиотеках программ. Методы решения и реализующий их оптимизационный софтвер постоянно совершенствуются, осваивая новые быстрорастущие возможности вычислительной техники. Значительное влияние на работы в области оптимизации оказало появление высокопроизводительных параллельных компьютеров, существенно расширивших область применения вычислительных методов, повышая их эффективность.

Основой оптимизации можно считать математическое программирование (МП), объединившее ЛП, НЛП и безусловную минимизацию функций многих переменных. Математическому программированию посвящено наибольшее количество публикаций. Укажем лишь некоторые книги, написанные Н.Н. Моисеевым [70], Д. Луенбергером [121], О. Мангасарьяном [123], И.И. Ереминым [60], Ф.П. Васильевым [23–25], Б.Т. Поляком [74], В.Г. Кармановым [64], А.И. Егоровым [59], Д. Бертсекасом [19, 95] и автором [50].

На протяжении многих лет в ВЦ РАН разрабатывались численные методы оптимизации, проводилась их апробация для решения различных тестовых и практических задач. Основное внимание было уделено созданию библиотек программ для задач ЛП, НЛП, многокритериальной оптимизации и оптимального управления. Среди участников этих работ можно указать Н.Н. Моисеева, В.Г. Жадана, А.И. Голикова, Л.Г. Хачияна, Н.И. Грачева, А.А. Станевичюса, М.А. Потапова и многих других. В период с 1980 г. по 1990 г. была создана диа-

логовая система оптимизации ДИСО, предназначенная для решения задач безусловной минимизации (БМ), нелинейного программирования (НЛП) и оптимального управления (ОУ). Обширная библиотека программ, вошедших в систему ДИСО, тиражировалась и широко использовалась в многочисленных проектных и исследовательских организациях [35–39]. Архитектура комплекса ДИСО, интеграция разнородных программ оптимизации, реализация техники быстрого автоматического дифференцирования — эти работы проводились автором, В.П. Мазуриком и Е.Н. Веселовым [54, 55] и многими другими.

Теория оптимального управления (ОУ), создававшаяся как развитие классического вариационного исчисления, имеет много важных приложений в самых разнообразных областях, таких как проектирование и управление роботами и специальными подвижными объектами, выбор наилучших стратегий функционирования индустриальных динамических систем. Создание численных методов существенным образом опирается на основные результаты, полученные в общей теории ОУ. Большим продвижением в этом направлении был высказанный Л.С. Понтрягиным в 1953 г. “принцип максимума”, позволивший дать каноническую формулировку необходимых условий оптимальности, заложить основу для формализации и развития нового раздела в вариационном исчислении, подготовить базу для дальнейших многочисленных и плодотворных исследований. Впоследствии “принцип максимума” был строго доказан Л.С. Понтрягиным совместно с его учениками [75]. Одно из первых доказательств принципа максимума было дано Л.И. Розоноэром [78]. Другие варианты доказательств приведены во многих книгах; укажем, например, [9, 12, 16, 20, 70, 81].

Принцип максимума оказался весьма популярным методом решения задач оптимального управления. Он особенно эффективен для аналитического решения несложных задач. Вместе с тем принцип максимума обладал одним недостатком: его нельзя было применять к важнейшему классу задач ОУ, содержащему фазовые и смешанные ограничения.

Возникли два пути преодоления этой трудности: первый был предложен А.А. Милютиным и А.Я. Дубовицким в работах [46–49]; он состоял в обобщении принципа максимума и его переносе на задачи с фазовыми ограничениями. В этом направлении работали многие их ученики и коллеги. Укажем некоторых из них и приведем часть их публикаций: В.В. Дикусар и А.А. Милютин [44], А.П. Афанасьев, В.В. Дикусар, А.А. Милютин, С.А. Чуканов [17], А.В. Дмитрук [45], А.А. Милютин [69].

Второй путь состоял в переходе от исходной задачи ОУ к ее дискретной аппроксимации, благодаря чему получалась задача нелинейного программирования, для решения которой использовались самые разнообразные методы НЛП. Этот подход развивался в частности в ВЦ РАН, некоторые его элементы даны ниже в главах 1, 3 и 4.

Для управляемых систем, описываемых обыкновенными дифференциальными уравнениями, интегральными уравнениями, уравнениями с частными производными, использовались дискретные аппроксимации (ниже это система (1.3.2)). Полученные таким образом задачи НЛП обладали важной спецификой, существенно упрощающей их решение.

Первые публикации по численным методам решения задач ОУ выполнили Д.Е. Охочимский и Т.М. Энеев, Л.И. Шатровский, А. Брайсон, А. Миеле, Г. Келли, А. Балакришнан, Дж. Лейтман, Р. Беллман. Было опубликовано значительное количество статей и книг на эту тему. Можно выделить несколько различных направлений в разработке численных методов, существенно отличающихся друг от друга. Прежде всего следует упомянуть прямые методы, основанные на спуске в пространстве управлений. Ряд исследований связан с непрямыми методами, в которых с помощью принципа максимума Л.С. Понтрягина исходная задача редуцировалась к краевой. Одна из первых публикаций на эту тему — статья [66]. В ней с помощью принципа максимума строился численный метод решения задач ОУ, аналогичный методу условного градиента, используемому в нелинейном программировании.

Цикл работ по численным методам ОУ был выполнен Н.Н. Моисеевым и его ученика-

ми. В этих работах был развит подход, основанный на вариациях в пространстве состояний. Другое направление, развиваемое Р.П. Федоренко, базировалось на использовании идей метода линеаризации. Методы условного градиента и проекции градиента были перенесены В.Ф. Демьяновым и А.М. Рубиновым на задачи ОУ. Немало публикаций было посвящено задачам оптимального управления системами, которые описываются уравнениями с частными производными. Упомянем книги Ф.Л. Черноусько и Н.В. Баничука [91], Ф.П. Васильева [24].

В третьей четверти XX века делались многочисленные попытки реализовать численные методы решения задач ОУ на основе идей динамического программирования, наиболее подробно описанного в книгах Р. Беллмана. Этот подход оказался эффективен лишь для задач небольшой размерности. Решение реальных практических задач требовало чрезвычайно большой машинной памяти. “Проклятие размерности” делало этот подход малопривлекательным для решения практических задач. В настоящее время ситуация заметно изменяется. Стремительный рост производительности и объема памяти вычислительных систем создали хорошие предпосылки для реализации методов, базирующихся на применении динамического программирования.

В одной из первых монографий по численным методам ОУ, написанной Н.Н. Моисеевым [70] в 1971 г., приведено подробное изложение численных схем, идущих от теории ОУ, и лишь вскользь указано на возможность иного подхода, основанного на использовании методов НЛП. Этот подход нашел самое широкое распространение в практике инженерных расчетов. Ему было посвящено наибольшее количество публикаций. Сошлемся только на первые книги автора [50, 114], а также Э. Полака [73], А.И. Пропося [76, 77]. В этих работах управление аппроксимировалось кусочно-постоянной функцией и проводилась конечномерная параметризация искомого управления, функционалы исходной задачи становились функциями конечного числа параметров. Такой подход редуцировал задачу ОУ к оптимизации функций многих переменных, причем задачи ОУ с краевыми условиями и фазовыми ограничениями сравнительно просто сводились к задачам НЛП. Предложенный Ф.Л. Черноусько метод локальных вариаций также сводил задачу ОУ к конечномерной, которая далее решалась с помощью специальных вариантов метода покоординатного спуска в пространстве управлений без использования градиентов [91].

На протяжении последних пятидесяти лет в научной литературе эпизодически появляются статьи о быстром автоматическом дифференцировании (БАД) функций многих переменных. Во многих случаях БАД значительно превосходит символическое дифференцирование или аппроксимацию производных с помощью разделенных разностей. Технике БАД было посвящено множество публикаций. Мы отсылаем читателя к трудам первого Симпозиума SIAM по автоматическому дифференцированию алгоритмов, состоявшегося в Брекенридже, штат Колорадо, в 1991 г. (см. [104]), и книге А. Гриванка [103]. Краткий обзор истории создания и развития автоматического дифференцирования и связанных с ним методов сделан М. Ири и приведен в [116]. Сошлемся также на труды конференции в Ницце [102], посвященной технике БАД, а также статьи [1, 11, 65], [97]– [103].

Начиная с 1970 г. в ВЦ АН СССР проводились работы по созданию численных методов решения задач ОУ, основанных на редукации исходной задачи к задаче НЛП; полученные при этом формулы дифференцирования оказались весьма близкими к формулам, используемым в работах, посвященных БАД. В статьях [1, 55, 111, 112, 114] был предложен общий подход к дифференцированию сложных функций и показано, что из найденных результатов можно получить формулы БАД и формулы дифференцирования для задач оптимального управления процессами, описываемыми обыкновенными дифференциальными уравнениями и уравнениями с частными производными. Некоторые исследования в этой области опубликованы в [1, 35–39, 50, 52, 53, 68, 106].

Среди многочисленных публикаций, посвященных задачам ОУ, есть работы, в которых

авторы, используя технику вариационного исчисления, определяют градиенты целевой функции (функционала) с помощью введения уравнений для импульсов (сопряженных систем). После этого исходная и сопряженная системы дискретизируются. Такой подход, однако, может привести к неточному определению градиента. Для нахождения точного градиента дискретизация обеих систем должна быть выполнена согласованно, что автоматически обеспечивает техника БАД. В качестве иллюстрации этих двух подходов в главе 4 приведено решение оптимизационной задачи для уравнения Бюргерса.

Излагаемый ниже материал является развитием и обобщением подходов, разработанных на протяжении многих лет в ВЦ РАН. Методы решения задач ОУ, базирующиеся на идеях НЛП, оказались чрезвычайно эффективными по многим причинам: с их помощью стали очевидными многие ранее предложенные эвристические алгоритмы, возникла возможность их обобщения, они позволили использовать богатый, всесторонне развитый арсенал методов НЛП и безусловной минимизации (БМ), создали конструктивную основу для построения методов оптимизации систем, проинтегрированных по схемам высокого порядка точности. Методы НЛП позволили решать сложные задачи ОУ со смешанными ограничениями. Здесь, говоря о задачах ОУ, мы не делаем различия между управляемыми системами, описываемыми обыкновенными дифференциальными уравнениями, и системами с частными производными. В обоих случаях после дискретизации получаются однотипные многошаговые процессы, к которым в одинаковой степени применима техника НЛП. Полученные в этом случае задачи обладают (в отличие от обычных задач НЛП) одним замечательным свойством: в них сравнительно просто вычисляются производные целевых функций по управлениям и управляющим параметрам. Этим объясняется высокая эффективность градиентной техники оптимизации при решении таких задач.

Теория и методы оптимизации создавались многими учеными. Интересно попытаться провести единую линию изложения важнейших результатов в теории оптимизации. Эта идея высказывалась Н.Н. Моисеевым, который на протяжении многих лет читал курс “Оптимизация” в МФТИ. В своей книге “Как далеко до завтрашнего дня” [71, с. 186] он писал: “...мне хотелось дать студентам такую интерпретацию, которая позволила бы увидеть, сколь по существу прост этот принцип (принцип максимума Л.С. Понтрягина — прим. автора), как он естественным образом связан с классическим математическим анализом. И что принцип максимума выводится практически традиционным образом, опираясь лишь на идеи Лагранжа и Лежандра”. Следуя этой идее, в данной публикации важнейшие результаты теории оптимизации излагаются исходя из хорошо известной теоремы о неявной функции. Показано, что такой подход позволяет единообразно доказать теоремы Лагранжа, Куна–Таккера, получить формулы БАД, построить градиентные методы решения задач ОУ и несложных игровых задач, дать новое доказательство принципа максимума Л.С. Понтрягина, найти эффективные расчетные схемы точного дифференцирования сложных элементарных функций многих переменных. Ниже в § 3.3 дано простейшее доказательство принципа максимума на базе общеизвестной теоремы о производной решения дифференциального уравнения по начальным условиям.

Техника БАД успешно использовалась в ВЦ РАН для решения сложных оптимизационных задач. В семидесятые годы прошлого столетия этот подход был применен к решению задач проектирования летательных аппаратов. По заказу КБ им. П.О. Сухого были созданы программы для конструирования оптимальных шпангоутов, для определения оптимальных режимов полета самолета. Часть этих работ была опубликована автором в книге “Методы решения экстремальных задач и их применение в системах оптимизации” [50]. В этой книге было указано, что технику БАД можно применять к решению простейших дифференциальных игр, в качестве примера приведены результаты численного решения задачи Р. Айзекса о долихобрахистотроне [2].



Коллективом сотрудников ВЦ РАН, в который вошли В.И. Зубов, А.Ф. Албу, А.В. Албу, выполнен цикл работ, демонстрирующих эффективность применения методологии БАД к решению практически важных реальных задач оптимального управления тепловыми процессами. Укажем важный класс задач, в которых исследуемое вещество переходит из одной фазы в другую с выделением или поглощением тепла; они называются задачами типа Стефана. Существенной чертой таких задач является наличие движущейся поверхности раздела между двумя фазами (жидкой и твердой), причем закон движения этой поверхности заранее не известен и его следует определять. Именно на этой поверхности происходит поглощение или выделение тепла, связанное с фазовым переходом. Термические свойства фаз по обеим сторонам движущейся поверхности могут оказаться различными. Задачи этого класса заметно сложнее тех, в которых отсутствует переход вещества из одной фазы в другую. Основные результаты по этой теме опубликованы в статьях [3–8].

Большое количество практических задач ОУ с помощью БАД-технологии решалось в Тверском государственном университете под руководством Е.А. Андреевой [10–12].

В данной работе мы ограничиваемся применением техники БАД только для вычисления первых производных. Вместе с тем в монографии автора [50], опубликованной в 1982 г., был приведен простейший пример задачи ОУ с фазовыми ограничениями, решенной методом Ньютона с привлечением формул БАД для нахождения вторых производных. В дальнейшем этот подход был развит; формулы, необходимые для реализации БАД-технологии, приведены в [53, 106, 111, 114].

Ниже будет показано, что методология БАД дает возможность с единой позиции рассмотреть задачи из самых различных направлений вычислительной математики. Естественно, что при таком охвате разных разделов не удалось изложить материал в обозначениях, традиционно принятых в этих областях. Этим пришлось пожертвовать ради единообразности изложения материала.

В главе 1 получены формулы вычисления градиентов и матриц производных сложных функций. С помощью теоремы о неявных функциях из условий связи определяются производные зависимых переменных по независимым переменным. Эти формулы положены в основу всего дальнейшего материала. Как частный случай из них следуют формулы так называемого “обратного” дифференцирования для многошаговых процессов. Для дискретных систем введена новая вспомогательная функция, позволяющая записать все расчетные формулы в канонической форме, что оказалось весьма удобным для практических вычислений. Из этих формул несложно получить формулы БАД для определения градиента функций многих переменных. При этом вычисление значения функции сводится к многошаговому процессу; в нем введены новые фазовые переменные, являющиеся функциями независимых переменных, по которым ищутся производные заданной функции. Найдена верхняя оценка отношения времени вычисления производных функции ко времени вычисления значения функции. Показано, что если используется техника БАД, то время вычисления частных производных элементарных функций превышает время вычисления значения функций не более чем в три раза; при этом градиент вычисляется точно.

В главе 2 с помощью найденных в первой главе формул дифференцирования получены необходимые и достаточные условия локального минимакса и максимина. Сформулированы и доказаны теоремы Лагранжа и Куна–Таккера. Для задач НЛП определена функция чувствительности, рассмотрены некоторые ее свойства, приведены условия, при которых она дифференцируема, найдены ее производные. Необходимые условия минимума Куна–Таккера можно использовать для решения задач НЛП. Особенно просто осуществить такой подход в случае, когда задача содержит только ограничения типа равенства. Соответствующие системы нелинейных уравнений решаются различными методами, разработанными для нелинейных систем. В качестве примера используется метод П.Л. Чебышева. При наличии огра-

ничений типа неравенства задача сводится к системе нелинейных равенств с помощью специальных преобразований (см., например, [95, 110, 113]).

В главе 3 методология быстрого автоматического дифференцирования переносится на задачи оптимального управления процессами, описываемыми системами обыкновенных дифференциальных уравнений. Изложение начинается с простейшей задачи Майера с терминальным функционалом. С помощью результатов главы 1 получены формулы дифференцирования терминального функционала по управлению и управляющим параметрам. Найденные формулы позволяют дать простое доказательство принципа максимума Л.С. Понтрягина, дискретного принципа максимума и принципа квазимаксима Р. Габасова и Ф.М. Кирилловой. Доказательство принципа максимума базируется на хорошо известном свойстве дифференцируемости решений систем дифференциальных уравнений по начальным условиям. Предлагаемый подход применяется для оптимизации систем, описываемых обыкновенными дифференциальными уравнениями с разрывными правыми частями.

В главе 4 рассмотрены три задачи оптимизации процессов, описываемых уравнениями с частными производными. В каждой из них задаются дискретные аппроксимации исходной системы, краевых и начальных условий. По формулам БАД автоматически однозначно определяется дискретная аппроксимация сопряженной системы и даются формулы вычисления точного градиента целевой функции. Отметим, что такой подход, несмотря на его очевидность и простоту, не является общепринятым. В качестве примера в этой главе техника БАД противопоставлена подходу, иногда применяемому в публикациях, в котором исходная и сопряженная системы дискретизируются независимо (несогласованно). Приведены результаты численных расчетов, наглядно иллюстрирующие преимущество техники БАД. Рассмотрены дискретные варианты управляемых систем, описываемых уравнением Бюргерса, а также параболическими и гиперболическими уравнениями.

Данный материал составлен на базе лекций по курсу “Дополнительные главы методов оптимизации”, читавшемуся автором в Московском физико-техническом институте и Московском государственном университете в 1990–2000 гг.

Автор выражает глубокую благодарность А.И. Голикову, сделавшему много полезных замечаний по данной рукописи, а также соавторам статей, использованных в данном тексте: В.Г. Жадану, А.А. Третьякову, М.А. Посыпкину и многим другим. Неоценимую помощь оказали Н.В. Мишина, В.У. Малкова и Н.П. Петрова в подготовке и редактировании текста.

# Глава 1. Дифференцирование сложных функций

## 1.1. Вычисление первых производных

Имеется несколько вариантов вывода формул быстрого автоматического дифференцирования (БАД). Среди них самый простой и наиболее общий способ основан на использовании хорошо известной теоремы о неявной функции. Именно этот подход будет применен ниже.

Пусть векторы  $x \in \mathbb{R}^m$  и  $u \in \mathbb{R}^r$ , дифференцируемая вектор-функция  $G(x, u)$  задает отображение  $G : \mathbb{R}^m \times \mathbb{R}^r \rightarrow \mathbb{R}^m$  и определяет систему, состоящую из  $m$  нелинейных скалярных уравнений:

$$G(x, u) = 0_m, \quad (1.1.1)$$

где  $0_s$  есть  $s$ -мерный нулевой вектор.

Введем квадратную матрицу Якоби порядка  $m$ :

$$G_{x^\top}(x, u) = \begin{bmatrix} \frac{\partial G^1(x, u)}{\partial x^1} & \cdots & \frac{\partial G^1(x, u)}{\partial x^m} \\ \frac{\partial G^2(x, u)}{\partial x^1} & \cdots & \frac{\partial G^2(x, u)}{\partial x^m} \\ \cdots & \cdots & \cdots \\ \frac{\partial G^m(x, u)}{\partial x^1} & \cdots & \frac{\partial G^m(x, u)}{\partial x^m} \end{bmatrix}.$$

Транспонированная матрица есть  $[G_{x^\top}(x, u)]^\top = G_x^\top(x, u)$ .

В системе (1.1.1) вектор  $u$  можно рассматривать как некоторый векторный параметр. При определенных предположениях из системы (1.1.1) можно определить функцию  $x(u)$  (в общем случае многозначную) такую, что

$$G(x(u), u) = 0_m. \quad (1.1.2)$$

Обычно  $x$  и  $u$  называются *векторами зависимых и независимых переменных* соответственно. Уравнения (1.1.1) называются *условиями связи*. Из условия (1.1.2) следует, что полная производная функции  $G$  по  $u$  должна быть равна нулю, т.е.

$$\frac{dG(x(u), u)}{du^\top} = G_{x^\top}(x(u), u)x_{u^\top}(u) + G_{u^\top}(x(u), u) \equiv 0_{mr}, \quad (1.1.3)$$

где  $0_{\gamma\beta}$  — прямоугольная нулевая матрица  $\gamma \times \beta$ .

Матрица Якоби  $x_{u^\top}$  преобразования  $x(u)$  представляет собой прямоугольную матрицу размера  $m \times r$ :

$$x_{u^\top}(u) = \begin{bmatrix} \frac{\partial x^1}{\partial u^1} & \frac{\partial x^1}{\partial u^2} & \cdots & \frac{\partial x^1}{\partial u^r} \\ \frac{\partial x^2}{\partial u^1} & \frac{\partial x^2}{\partial u^2} & \cdots & \frac{\partial x^2}{\partial u^r} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial x^m}{\partial u^1} & \frac{\partial x^m}{\partial u^2} & \cdots & \frac{\partial x^m}{\partial u^r} \end{bmatrix}.$$

Транспонированная к ней матрица размера  $r \times m$  обозначается  $x_u^\top(u)$ . В некоторых случаях удобно использовать формулу (1.1.3) в транспонированной форме; тогда

$$\frac{dG^\top(x(u), u)}{du} = x_u^\top(u)G_x^\top(x(u), u) + G_u^\top(x(u), u) = 0_{rm}. \quad (1.1.4)$$

Пусть  $\Gamma(x_*) \subset \mathbb{R}^m$ ,  $\Gamma(u_*) \subset \mathbb{R}^r$  — некоторые окрестности точек  $x_*$  и  $u_*$  соответственно. Эти точки таковы, что  $G(x_*, u_*) = 0_m$ . Будем говорить, что система уравнений (1.1.1) определяет на декартовом произведении  $\Gamma(x_*) \times \Gamma(u_*)$  неявную функцию  $x(u)$ , если для любого  $u \in \Gamma(u_*)$  найдется вектор-функция  $x(u) \in \Gamma(x_*)$ , такая, что  $G(x(u), u) = 0_m$ .

**Теорема 1.1.1 (о неявной функции).** Пусть выполнены следующие условия:

- а) существует пара  $[x_*, u_*] \in \mathbb{R}^{m \times r}$  такая, что  $G(x_*, u_*) = 0_m$ ;
- б) вектор-функция  $G(x, u)$  непрерывно дифференцируема в окрестности точки  $[x_*, u_*]$ ;
- в) квадратная матрица  $G_x^\top(x_*, u_*)$  не вырождена.

Тогда найдутся окрестности  $\Gamma(x_*) \subset \mathbb{R}^m$  и  $\Gamma(u_*) \subset \mathbb{R}^r$  соответственно точек  $x_*$  и  $u_*$  такие, что на  $\Gamma(x_*) \times \Gamma(u_*)$  система (1.1.1) однозначно определяет неявную дифференцируемую функцию  $x(u)$ , которая является решением системы (1.1.1), и матрица первых производных  $x_{u^\top}(u)$  удовлетворяет линейной системе (1.1.3) при всех  $u \in \Gamma(u_*)$ .

Доказательство теоремы о неявной функции можно найти в многочисленных учебниках по математическому анализу (см., например, [83]).

Из невырожденности  $G_x^\top(x_*, u_*)$  следует, что матрица  $G_x^\top(x(u), u)$  также не вырождена в некоторой окрестности точки  $u_*$ , где система (1.1.3) однозначно разрешима:

$$x_{u^\top}(u) = -[G_{x^\top}(x(u), u)]^{-1} G_{u^\top}(x(u), u).$$

Пусть задана непрерывно дифференцируемая скалярная функция  $W(x, u)$ . Определенную в теореме 1.1.1 функцию  $x(u)$  подставим в функцию  $W(x, u)$ , в результате получим сложную функцию  $\Omega(u) = W(x(u), u)$ . Полную производную функции  $\Omega(u)$  по  $u$  обозначим  $\nabla\Omega(u)$ . Матрицу  $r \times r$  вторых производных функции  $\Omega(u)$  по  $u$  обозначим  $\Delta\Omega(u)$ . В математическом программировании  $r$ -мерный вектор-столбец  $\nabla\Omega$  называется *приведенным градиентом* (reduced gradient), матрица вторых производных  $\Delta\Omega(u)$  называется *приведенным гессианом* (reduced Hessian).

Если квадратная матрица  $G_x^\top(x(u), u)$  не вырождена, то можно определить прямоугольную  $r \times m$  матрицу  $N(u)$  и  $m$ -мерный вектор  $p(u)$  по формулам

$$\underbrace{N(u)}_{r \times m} = \underbrace{x_u^\top(u)}_{r \times m} = \underbrace{[x_{u^\top}(u)]^\top}_{r \times m} = \underbrace{-G_u^\top(x(u), u)}_{r \times m} \underbrace{[G_x^\top(x(u), u)]^{-1}}_{m \times m}, \quad (1.1.5)$$

$$\underbrace{p(u)}_{m \times 1} = - \underbrace{[G_x^\top(x(u), u)]^{-1}}_{m \times m} \underbrace{W_x(x(u), u)}_{m \times 1}. \quad (1.1.6)$$

Умножим (1.1.5) справа на матрицу  $G_x^\top(x(u), u)$ , транспонируем найденное выражение, умножим (1.1.6) слева на матрицу  $G_x^\top(x(u), u)$ . Получим, что введенная  $r \times m$ -матрица  $N(u)$  и  $m$ -мерный вектор  $p(u)$  удовлетворяют следующим системам линейных уравнений:

$$G_{u^\top}(x(u), u) + G_{x^\top}(x(u), u)N^\top(u) = 0_{mr}, \quad (1.1.7)$$

$$W_x(x(u), u) + G_x^\top(x(u), u)p(u) = 0_m. \quad (1.1.8)$$

Система (1.1.7) совпадает с (1.1.3). Система (1.1.8) следует из определения (1.1.6) вектора  $p(u)$ .

Полная производная функции  $\Omega(u)$  по  $u$  находится по правилу дифференцирования сложной функции. С учетом введенных обозначений имеем

$$\nabla\Omega(u) = \frac{d\Omega(u)}{du} = W_u(x(u), u) - \overbrace{G_u^\top(x(u), u) [G_x^\top(x(u), u)]^{-1} W_x(x(u), u)}^{-N(u)} \underbrace{-p(u)}_{-p(u)}. \quad (1.1.9)$$

Используя эти представления, формулу (1.1.9) перепишем двумя различными способами:

$$\nabla\Omega(u) = W_u(x(u), u) + N(u)W_x(x(u), u), \quad (1.1.10)$$

$$\nabla\Omega(u) = W_u(x(u), u) + G_u^\top(x(u), u)p(u). \quad (1.1.11)$$

Оба выражения (1.1.10) и (1.1.11), определяющие градиент  $\nabla\Omega$ , математически эквивалентны, но с вычислительной точки зрения имеется существенное отличие. Небольшое изменение в способе вычисления градиента приводит в ряде случаев к существенному изменению трудоемкости вычислений. Основное расчетное время при определении производной тратится на нахождение матрицы  $N$  и вектора  $p$ . В первом случае из (1.1.7) находится вспомогательная прямоугольная матрица  $N$ ; во втором случае из (1.1.8) определяется  $m$ -мерный вектор  $p$ . Для нахождения матрицы  $N$  требуется решить систему (1.1.7), состоящую из  $mr$  скалярных уравнений с таким же количеством неизвестных. Для определения  $p$  необходимо только один раз решить систему (1.1.8) из  $m$  алгебраических уравнений с  $m$  неизвестными. Таким образом, использование формулы (1.1.10) примерно в  $r$  раз более трудоемко, чем расчеты по формуле (1.1.11).

Ниже мы покажем, что в задачах оптимального управления формула (1.1.10) соответствует так называемому “прямому” (или “контрвариантному”) дифференцированию, а формула (1.1.11) — “обратному” (или “ковариантному”) дифференцированию.

Полученные результаты резюмируем в виде следующей теоремы, являющейся очевидным следствием теоремы о неявной функции.

**Теорема 1.1.2.** Пусть скалярная функция  $W(x, u)$  и векторная функция  $G(x, u)$  определены и непрерывно дифференцируемы в некоторой окрестности пары  $[x_*, u_*]$ , где  $G(x_*, u_*) = 0_m$  и квадратная матрица  $G_x^\top(x_*, u_*)$  не вырождена. Тогда существует окрестность  $\Gamma(u_*)$  точки  $u_*$ , где определены  $m$ -мерные вектор-функции  $p(u)$  и  $x(u)$  такие, что

1) функция  $x(u)$ , удовлетворяющая условиям  $G(x(u), u) = 0_m$  и  $x(u_*) = x_*$ , непрерывно дифференцируема на  $\Gamma(u_*)$ , ее первая производная определяется из системы (1.1.4), матрица  $G_x^\top(x(u), u)$  не вырождена для всех  $u \in \Gamma(u_*)$ ;

2) функция  $p(u)$  вычисляется по формуле (1.1.6), причем

$$p_* = p(u_*) = - [G_x^\top(x_*, u_*)]^{-1} W_x(x_*, u_*); \quad (1.1.12)$$

3) сложная функция  $\Omega(u)$  непрерывно дифференцируема, формула Тейлора с остаточным членом в форме Пеано записывается в виде

$$\Omega(u) = \Omega(u_*) + (u - u_*)^\top \nabla\Omega(u_*) + \alpha(u_*, \|u - u_*\|)\|u - u_*\|, \quad (1.1.13)$$

где градиент вычисляется по формуле (1.1.10) или (1.1.11), скалярная функция  $\alpha$  удовлетворяет условию

$$\lim \alpha(u_*, \|u - u_*\|) = 0 \quad \text{при} \quad \|u - u_*\| \rightarrow 0. \quad (1.1.14)$$

Данная теорема будет основой для всего дальнейшего изложения. С ее помощью будут выведены формулы быстрого автоматического дифференцирования, указан путь определения производных второго порядка, в следующей главе будут получены необходимые и достаточные условия экстремума в задачах ЛП и НЛП, доказаны теоремы Лагранжа, Куна–Таккера, определена функция чувствительности и т.д.

Здесь и ниже в ряде мест удобно объединять два вектора  $x$  и  $u$  в один вектор  $z \in \mathbb{R}^n$ , где  $n = m + r$ . Будем использовать следующие обозначения:

$$z = \begin{bmatrix} x \\ u \end{bmatrix}, \quad z(u) = \begin{bmatrix} x(u) \\ u \end{bmatrix}, \quad G(z) = G(x, u), \quad G_{z^\top} = [G_{x^\top} \mid G_{u^\top}], \\ W(z) = W(x, u), \quad W_{z^\top} = [W_{x^\top} \mid W_{u^\top}]. \quad (1.1.15)$$

В теореме 1.1.1 условие  $\mathbf{e}$  о невырожденности матрицы  $G_x^\top(x_*, u_*)$  можно заменить более общим условием: матрица  $G_z^\top(z_*)$  имеет максимальный ранг, равный  $m$ , или, что эквивалентно, все  $m$  столбцов матрицы  $G_z^\top(z_*)$  линейно независимы. Выполнение этих условий обеспечивает возможность выделения из компонент вектора  $z$  подвекторов  $x \in \mathbb{R}^m$  таких, что соответствующая  $m \times m$  подматрица  $G_x^\top(x_*, u_*)$  удовлетворяет условию  $\mathbf{e}$  в теореме 1.1.1.

Приведенные выше формулы были связаны только с теоремой о неявной функции; речь не шла об оптимизации. С этого момента введенную скалярную функцию  $W(x, u)$  будем называть *целевой функцией*.

Рассмотрим задачу нелинейного программирования об отыскании минимума или максимума функции  $W(x, u)$  по  $x$  и  $u$  и при условии, что эти векторы удовлетворяют условию связи (1.1.1). Такие векторы мы называем *допустимыми*.

Совокупность точек, удовлетворяющих условию  $G(z) = G(x, u) = 0_m$ , будем обозначать

$$Z = \{z \in \mathbb{R}^n : G(z) = 0_m\}.$$

Задачу нелинейного программирования (НЛП) запишем в виде

$$W_* = \min_{z \in Z} W(z). \quad (1.1.16)$$

Множество решений этой задачи  $Z_*$  предполагается непустым. Оптимальное значение целевой функции есть  $W_* = W(z_*)$ , где  $z_* \in Z_*$ .

Для задачи (1.1.16) введем вектор множителей Лагранжа  $p \in \mathbb{R}^m$  и функцию Лагранжа

$$L(z, p) = W(z) + p^\top G(z) = L(x, u, p). \quad (1.1.17)$$

Дифференцируя  $L(z, p)$  по  $z$ , получим градиент функции Лагранжа

$$L_z(z, p) = W_z(z) + G_z^\top(z)p.$$

Транспонируя это выражение, имеем  $L_{z^\top} = [L_{x^\top} \mid L_{u^\top}]$ .

Согласно принятым обозначениям можно записать

$$\begin{aligned} L_x(x(u), u, p) &= W_x(x(u), u) + G_x^\top(x(u), u)p(u), \\ L_u(x, u, p) &= W_u(x, u) + G_u^\top(x, u)p. \end{aligned}$$

Решение задачи (1.1.16) проведем в два этапа: сначала, используя условие допустимости  $z \in Z$ , выразим зависимые переменные  $x$  в виде функций от независимых переменных  $u$ . Другими словами, разрешим систему (1.1.1) относительно  $x$ , получим ранее введенную неявную функцию  $x(u)$ . Далее функцию  $x(u)$  подставим в  $W(x, u)$ , получим сложную функцию  $\Omega(u) = W(x(u), u)$ , зависящую только от независимых переменных; задача условной минимизации (1.1.16) свелась к задаче оптимизации  $\Omega(u)$  без ограничений, т.е. к безусловной минимизации. Согласно теореме о неявной функции  $\Omega(u)$  дифференцируема в окрестности точки  $u_*$ , где выполнено условие регулярности ограничений. Воспользуемся теоремой Ферма 2.1.1, получим: для того чтобы точка  $u_*$  была точкой минимума или максимума функции  $\Omega(u)$ , необходимо, чтобы в точке  $u_*$  градиент функции  $\Omega(u)$  был равен нулю. Используя формулу (1.1.11), получим

$$\nabla \Omega(u_*) = W_u(x_*, u_*) + G_u^\top(x_*, u_*)p_* = L_u(x_*, u_*, p_*) = 0_r, \quad (1.1.18)$$

где  $x_* = x(u_*)$ ,  $p_*$  определяется из (1.1.12). Из (1.1.8) также следует

$$L_x(x_*, u_*, p_*) = W_x(x_*, u_*) + G_x^\top(x_*, u_*)p_* = 0_m. \quad (1.1.19)$$

Объединяя  $r$  равенств (1.1.18) и  $m$  равенств (1.1.19) в одно, получим систему из  $n = r + m$  скалярных уравнений

$$L_z(z_*, p_*) = W_z(z_*) + G_z^\top(z_*)p_* = 0_n, \quad (1.1.20)$$

эти равенства вместе с (1.1.2) зададут систему из  $n + m$  скалярных уравнений с таким же количеством неизвестных  $z_*$  и  $p_*$ .

Умножим (1.1.20) слева на матрицу  $G_{z^\top}(z_*)$ , получим систему, состоящую из  $m$  линейных уравнений с искомым  $m$ -мерным вектором  $p_*$ :

$$G_{z^\top}(z_*)W_z(z_*) + M(z_*)p_* = 0_m, \quad (1.1.21)$$

где введена матрица Грамма  $M(z_*) = G_{z^\top}(z_*)G_z^\top(z_*)$ . Если ранг матрицы  $G_z^\top(z_*)$  максимальный (равен  $m$ ), то система (1.1.21) однозначно разрешима:

$$p_* = -[M(z_*)]^{-1}G_{z^\top}(z_*)W_z(z_*). \quad (1.1.22)$$

Если, в частном случае, матрица  $G_z(z_*)$  квадратная, то матрица, обратная к матрице Грамма, представима в виде  $M^{-1}(z_*) = [G_z^\top(z_*)]^{-1}[G_{z^\top}(z_*)]^{-1}$  и из (1.1.22) следует формула (1.1.12), использование которой проще, однако для ее построения необходимо расщепить вектор  $z$ , выделив в нем подвектор зависимых переменных  $x$ .

Утверждение о существовании в задаче условной оптимизации (1.1.16) множителя Лагранжа  $p_*$  здесь получено из теоремы Ферма и из теоремы о неявной функции. Ниже, в § 2.4, к этому результату мы придем несколько другим путем.

## 1.2. Вычисление вторых производных

В этом разделе предполагаем, что функции  $W(x, u)$ ,  $G(x, u)$  дважды непрерывно дифференцируемы по своим аргументам. По формуле (1.1.17) определим функцию Лагранжа и к условию связи (1.1.1) добавим условие (1.1.22); в результате вместо (1.1.1) получим следующую нелинейную систему, состоящую из  $2m$  уравнений с  $2m$  неизвестными  $x$  и  $p$ :

$$G(x, u) = 0_m, \quad L_x(x, u, p) = 0_m. \quad (1.2.1)$$

Для этой системы введем матрицу Якоби

$$J(x, u, p) = \left[ \begin{array}{c|c} G_{x^\top}(x, u) & 0_{mm} \\ \hline L_{xx^\top}(x, u, p) & G_x^\top(x, u) \end{array} \right].$$

Ее определитель (якобиан) равен

$$|J(x, u, p)| = |G_{x^\top}(x, u)|^2. \quad (1.2.2)$$

Независимой переменной остается по-прежнему вектор  $u$ , зависимыми переменными будут  $x$  и  $p$ . Согласно (1.2.2) матрица  $J(x_*, u_*, p_*)$  не вырождена, если не вырождена матрица  $G_x^\top(x_*, u_*)$ . Применив теорему 1.1.1 к новой системе (1.2.1), получим, что существует окрестность  $\Gamma(u_*)$ , где однозначно определены функции  $x(u)$ ,  $p(u)$ , удовлетворяющие условиям

$$G(x(u), u) = 0_m, \quad L_x(x(u), u, p(u)) = 0_m,$$

а также условиям (1.1.3), (1.1.22). Градиент сложной функции  $W(x(u), u)$  имеет вид

$$\nabla\Omega(u) = \nabla W(x(u), u) = x_u^\top W_x(x(u), u) + W_u(x(u), u).$$

Вычисление полной производной этого выражения приводит к следующей формуле:

$$\Delta\Omega(u) = W_{uu^\top} + W_{ux^\top}x_{u^\top} + x_u^\top [W_{xu^\top} + W_{xx^\top}x_{u^\top}] + \sum_{j=1}^m x_{uu^\top}^j W_{x^j} \Big|_{[x(u),u]}. \quad (1.2.3)$$

Здесь вслед за вертикальной чертой указано, какие аргументы подставляются в производные функции  $W$  после вычисления частных производных по  $x$  и  $u$ .

Из системы (1.1.3), записанной в матричной форме, выделим  $i$ -ю строку, получим

$$G_{u^\top}^i(x(u), u) + G_{x^\top}^i(x(u), u)x_{u^\top}(u) = 0_r^\top, \quad 1 \leq i \leq m, \quad 0_r^\top = 0_{1r}.$$

Эта формула имеет место тождественно для всех  $u$ , достаточно близких к  $u_*$ , поэтому ее можно продифференцировать по  $u$  и приравнять полученное выражение нулю. В результате приходим к формуле, аналогичной (1.2.3):

$$G_{uu^\top}^i + G_{ux^\top}^i x_{u^\top} + x_u^\top [G_{xu^\top}^i + G_{xx^\top}^i x_{u^\top}] + \sum_{j=1}^m x_{uu^\top}^j G_{x^j}^i \Big|_{[x(u),u]} = 0_{rr}. \quad (1.2.4)$$

Введем прямоугольную  $r \times n$  матрицу  $K$  и транспонированную к ней  $K^\top$ :

$$K(u) = [x_u^\top(u) \mid I_r] = [N(u), I_r], \quad K^\top(u) = \begin{bmatrix} x_{u^\top}(u) \\ I_r \end{bmatrix} = \begin{bmatrix} N^\top(u) \\ I_r \end{bmatrix}, \quad (1.2.5)$$

где  $I_r$  — единичная матрица  $r \times r$ , матрица  $N$  определена в (1.1.5). Используя это обозначение, формулу (1.1.3) запишем в виде

$$\frac{dG(z(u))}{du^\top} = G_{z^\top}(z(u))K^\top(u) = 0_{mr}. \quad (1.2.6)$$

Таким образом, столбцы матрицы  $K^\top(u)$  принадлежат нуль-пространству матрицы  $G_{z^\top}(z(u))$ . Возьмем в (1.1.17) в качестве  $z$  функцию  $z(u)$  и умножим  $L_z(z(u), u)$  на матрицу  $K(u)$  слева, учтем (1.1.8), получим

$$K(u)L_z(z(u), p) = K(u)[W_z(u) + G_z^\top(z(u))p] = K(u)W_z(z(u)).$$

Отсюда следует

$$K(u)W_z(z(u)) = N(u)W_x(x(u), u) + W_u(x(u), u) = \nabla\Omega(u).$$

Введем матрицы вторых частных производных:

$$W_{zz^\top} = \left[ \begin{array}{c|c} W_{xx^\top} & W_{xu^\top} \\ \hline W_{ux^\top} & W_{uu^\top} \end{array} \right], \quad G_{zz^\top}^i = \left[ \begin{array}{c|c} G_{xx^\top}^i & G_{xu^\top}^i \\ \hline G_{ux^\top}^i & G_{uu^\top}^i \end{array} \right], \quad L_{zz^\top} = W_{zz^\top} + \sum_{j=1}^m p^j G_{x^j}^i,$$

где  $z \in \mathbb{R}^n$ ,  $n = m + r$ . Непосредственные вычисления показывают, что формулы (1.2.3) и (1.2.4) можно записать следующим компактным образом:

$$\Delta\Omega(u) = K(u)W_{zz^\top}(x(u), u)K^\top(u) + \sum_{j=1}^m x_{uu^\top}^j(u)W_{x^j}(x(u), u), \quad (1.2.7)$$

$$K(u)G_{zz^\top}^i(x(u), u)K^\top(u) + \sum_{j=1}^m x_{uu^\top}^j(u)G_{x^j}^i(x(u), u) = 0_{rr}, \quad (1.2.8)$$



Умножим (1.2.8) на  $p^i(u)$ , просуммируем по  $i$  от 1 до  $m$  и прибавим это выражение к (1.2.7). С учетом последней формулы в (1.2.5) получим

$$\begin{aligned} \Delta\Omega(u) &= K(u)L_{zz}(x(u), u, p(u))K^\top(u) + \\ &+ \sum_{j=1}^m x_{uu^\top}^j(u) \left[ W_{x^j}(x(u), u) + \sum_{i=1}^m p^i(u)G_{x^j}^i(x(u), u) \right]. \end{aligned} \quad (1.2.9)$$

Из (1.1.8) следует, что

$$W_x(x(u), u) + \sum_{i=1}^m p^i(u)G_x^i(x(u), u) = 0_m,$$

поэтому выражение, стоящее в (1.2.9) в квадратных скобках, равно нулю, и мы приходим к следующему результату:

$$\Delta\Omega(u) = K(u)L_{zz^\top}(x(u), u, p(u))K^\top(u). \quad (1.2.10)$$

Подставляя в эту формулу выражения для  $K$  и  $K^\top$  из (1.2.5), получим

$$\Delta\Omega(u) = L_{uu^\top} + L_{ux^\top}x_{u^\top} + x_u^\top(L_{xu^\top} + L_{xx^\top}x_{u^\top})\Big|_{[x(u), u]}.$$

Эта формула весьма близка к (1.2.3), но здесь в отличие от (1.2.3) удалось исключить матрицу вторых производных функции  $x(u)$  по  $u$ ; при этом вместо производных функции  $W$  здесь вычисляются соответствующие производные функции Лагранжа  $L$ .

**Теорема 1.2.1.** Пусть функции  $W(x, u)$  и  $G(x, u)$  дважды непрерывно дифференцируемы в некоторой окрестности пары  $[x_*, u_*]$ , где  $G(x_*, u_*) = 0_m$  и матрица  $G_x^\top(x_*, u_*)$  не вырождена. Тогда существует окрестность  $\Gamma(u_*)$  точки  $u_*$ , где определены функции  $x(u)$ ,  $p(u)$ , описанные в теореме 1.1.2, причем сложная функция  $\Omega(u)$  дважды непрерывно дифференцируема, формула Тейлора записывается в виде

$$\Omega(u) = \Omega(u_*) + (u - u_*)^\top \nabla\Omega(u_*) + \frac{(u - u_*)^\top \Delta\Omega(u_*)(u - u_*)}{2} + \alpha(u_*, \|u - u_*\|)\|u - u_*\|^2; \quad (1.2.11)$$

здесь функция  $\alpha(\cdot)$  удовлетворяет условию (1.1.14), приведенный градиент и приведенный гессиан вычисляются по формулам

$$\begin{aligned} \nabla\Omega(u_*) &= W_u(x_*, u_*) + G_u^\top(x_*, u_*)p_*, \\ \Delta\Omega(u_*) &= K(u_*)L_{xx^\top}(x_*, u_*, p_*)K^\top(u_*), \end{aligned}$$

где

$$\begin{aligned} p_* &= -[G_x(x_*, u_*)]^{-1}W_x(x_*, u_*), \\ K(u_*) &= [N(u_*)|J_r], \quad N(u_*) = -G_u^\top(x_*, u_*)[G_x^\top(x_*, u_*)]^{-1}. \end{aligned}$$

Формулы (1.1.13) и (1.2.11) являются ключевыми для всего последующего материала.

### 1.3. Вычисление первых производных в случае многошаговых процессов

Формулы (1.1.3) – (1.1.13) будут основой для получения различных формул дифференцирования. Ниже  $i$ -ю компоненту вектора  $x$  обозначаем  $x^i$ ; запись  $x_i$  обозначает некоторый  $i$ -й вектор. Таким образом,  $x_j^i$  есть  $i$ -я компонента вектора  $x_j$ .

В многошаговых задачах векторы  $x$  и  $u$  обычно состоят из наборов  $k$  векторов меньших размерностей, чем векторы  $x$  и  $u$ :

$$\begin{aligned} x^\top &= [x_1^\top, x_2^\top, \dots, x_k^\top], & u^\top &= [u_1^\top, u_2^\top, \dots, u_k^\top], \\ x_i &\in \mathbb{R}^s, & u_i &\in \mathbb{R}^\rho, & 1 \leq i \leq k, & x \in \mathbb{R}^m, & u \in \mathbb{R}^r, & m = ks, & r = k\rho. \end{aligned} \quad (1.3.1)$$

Допустимо, чтобы векторы  $x_i, u_i$  при разных  $i$  имели разные размерности. Для простоты изложения считаем, что их размерности одинаковые ( $s = 1$ ) и не зависят от  $i$ . Набор целых чисел от 1 до  $k$  будем обозначать  $D$ .

Пусть условие связи (1.1.1) расщеплено на  $k$  соотношений следующим образом:

$$x_j = F(j, X_j, U_j), \quad j \in D, \quad (1.3.2)$$

где  $X_j, U_j$  — некоторые заданные наборы векторов  $x_i$  и  $u_i$  соответственно;  $j$  принимает предписанные значения из  $D$ .

Обычно функция  $F(\cdot)$  определяется из дискретной аппроксимации исходной управляемой системы. Будем говорить, что соотношения (1.3.2) задают  $k$ -шаговый процесс. С каждым  $j$ -м шагом связаны векторы  $x_j$  и  $u_j$ . В многошаговых управляемых процессах обычно  $x_i$  — фазовый вектор,  $u_i$  — вектор управлений,  $x$  — полный фазовый вектор,  $x \in \mathbb{R}^m$ ,  $u$  — полный вектор управлений,  $u \in \mathbb{R}^r$ .

Для каждого  $j \in D$  введем два индексных набора  $Q_j$  и  $S_j$ , содержащие индексы всех векторов  $x_i$  и  $u_i$ , принадлежащих соответственно наборам  $X_j$  и  $U_j$ :

$$Q_j = \{i \in D : x_i \in X_j\}, \quad S_j = \{i \in D : u_i \in U_j\}.$$

На  $j$ -м шаге множество  $Q_j$  определяет все индексы “входных” фазовых векторов  $x_i$ , содержащихся в множестве  $X_j$ , необходимых для вычисления правой части (1.3.2). Аналогично на  $j$ -м шаге  $S_j$  задает множество всех индексов “входных” векторов управления  $u_i$ . Для каждого  $j \in D$  введем сопряженные индексные наборы

$$\overline{Q}_j = \{i \in D : x_j \in X_i\}, \quad \overline{S}_j = \{i \in D : u_j \in U_i\}.$$

На  $j$ -м шаге множество  $\overline{Q}_j$  определяет номера всех тех шагов  $i \in D$ , для которых  $x_j \in X_i$ . С помощью введенных индексных множеств можно записать

$$\begin{aligned} X_j &= \{x_i : i \in Q_j\}, & \overline{X}_j &= \{x_i : i \in \overline{Q}_j\}, \\ U_j &= \{u_i : i \in S_j\}, & \overline{U}_j &= \{u_i : i \in \overline{S}_j\}. \end{aligned}$$

Из приведенных определений следует, что если, например,  $i, i+2 \in Q_j$ , что эквивалентно  $x_i, x_{i+2} \in X_j$ , то после подстановки всех элементов множества  $X_j$  в правую часть (1.3.2) получается функциональная зависимость вида  $x_j = F(j, \dots, x_i, x_{i+2}, \dots, U_j)$ , т.е.  $j \in \overline{Q}_i$ ,  $j \in \overline{Q}_{i+2}$ .

Если  $s, q, \ell \in \overline{Q}_j$ , что эквивалентно включениям  $x_s, x_q, x_\ell \in \overline{X}_j$ , то из (1.3.2) следуют функциональные зависимости вида

$$x_s = F(s, \dots, x_j, \dots, U_s), \quad x_q = F(q, \dots, x_j, \dots, U_q), \quad x_\ell = F(\ell, \dots, x_j, \dots, U_\ell),$$

т.е.  $j \in Q_s$ ,  $j \in Q_q$ ,  $j \in Q_\ell$ .

Для всех  $j$  и  $i$  из множества  $D$  включение  $i \in Q_j$  эквивалентно включению  $j \in \bar{Q}_i$ ; аналогично условие  $i \in S_j$  эквивалентно условию  $j \in \bar{S}_i$ . Это следует из функциональных зависимостей

$$x_j = F(j, \dots, x_i, \dots, U_j), \quad x_j = F(j, X_j, \dots, u_i, \dots). \quad (1.3.3)$$

Вектор-функция  $G(x, u)$  в (1.1.1) может быть представлена как объединение всех вектор-функций  $F(j, X_j, U_j) - x_j$ , где  $j \in D$ . Определим прямоугольную  $\rho \times s$  матрицу производных  $N_{ij} = \partial x_j^\top / \partial u_i$ . Продифференцируем равенство (1.3.2) по  $u_i$ ; получим

$$N_{ij} = F_{u_i}^\top(j, X_j, U_j) + \sum_{q \in Q_j} N_{iq} F_{x_q}^\top(j, X_j, U_j). \quad (1.3.4)$$

Далее рассмотрим задачу минимизации введенной выше функции  $W(x, u)$ , однако вместо условий связи (1.1.1) будут использованы условия (1.3.2). Для них формула (1.1.10) запишется следующим образом:

$$\frac{d\Omega(u)}{du_i} = W_{u_i}(x, u) + \sum_{j \in D} N_{ij} W_{x_j}(x, u), \quad i \in D. \quad (1.3.5)$$

Каждому условию в (1.3.2) поставим в соответствие вектор  $p_j \in \mathbb{R}^s$ , где  $j \in D$ . Объединением этих векторов будет вектор  $p^\top = [p_1^\top, p_2^\top, \dots, p_k^\top]$ ,  $p \in \mathbb{R}^m$ .

Если исходить из формул (1.1.6) и (1.1.17), где вводился вектор  $p$ , то его можно назвать вектором множителей Лагранжа. Вместе с тем в дальнейшем этот вектор можно называть вектором импульсов, следуя при этом определению классической механики.

Функция Лагранжа для многошагового процесса (1.3.2) имеет вид

$$L(x, u, p) = W(x, u) + \sum_{j \in D} [F^\top(j, X_j, U_j) - x_j^\top] p_j. \quad (1.3.6)$$

Учитывая определение множеств  $\bar{Q}_i$  и  $\bar{S}_i$ , можно упростить формулы дифференцирования по  $x_i$  и  $u_i$ :

$$\begin{aligned} \sum_{j \in D} F_{x_i}^\top(j, X_j, U_j) p_j &= \sum_{j \in \bar{Q}_i} F_{x_i}^\top(j, X_j, U_j) p_j \\ \sum_{j \in D} F_{u_i}^\top(j, X_j, U_j) p_j &= \sum_{j \in \bar{S}_i} F_{u_i}^\top(j, X_j, U_j) p_j. \end{aligned}$$

Поэтому формулы (1.3.2), (1.1.8) и (1.1.11) запишутся в виде

$$\begin{aligned} L_{p_i}(x, u, p) &= F(i, X_i, U_i) - x_i = 0_s, \\ L_{x_i}(x, u, p) &= W_{x_i}(x, u) - p_i + \sum_{j \in \bar{Q}_i} F_{x_i}^\top(j, X_j, U_j) p_j = 0_s, \end{aligned} \quad (1.3.7)$$

$$\frac{d\Omega(u)}{du_i} = L_{u_i}(x, u, p) = W_{u_i}(x, u) + \sum_{j \in \bar{S}_i} F_{u_i}^\top(j, X_j, U_j) p_j.$$

Вместо функции Лагранжа удобно ввести новую вспомогательную функцию

$$E(x, u, p) = W(x, u) + \sum_{j \in D} F^\top(j, X_j, U_j) p_j. \quad (1.3.8)$$

Тогда соотношения (1.3.7) представимы в следующей *канонической форме*, весьма удобной для запоминания и расчетов:

$$x_i = E_{p_i}(x, u, p) = F(i, X_i, U_i), \quad (1.3.9)$$

$$p_i = E_{x_i}(x, u, p) = W_{x_i}(x, u) + \sum_{j \in \bar{Q}_i} F_{x_i}^\top(j, X_j, U_j) p_j, \quad (1.3.10)$$

$$\frac{d\Omega(u)}{du_i} = E_{u_i}(x, u, p) = W_{u_i}(x, u) + \sum_{j \in \bar{S}_i} F_{u_i}^\top(j, X_j, U_j) p_j. \quad (1.3.11)$$

Формулы (1.3.5) и (1.3.11) определяют одну и ту же производную сложной функции  $\Omega(u)$ . Матрицы  $N_{ij}$ , входящие в выражение (1.3.5), находятся из линейной матричной системы (1.3.4). Векторы  $p_i \in \mathbb{R}^m$  ( $i \in D$ ) определяются из линейной векторной системы (1.3.10).

В многошаговом процессе (1.3.2) вектор  $x_i$  представляет собой *выходной вектор*, если индексное множество  $\bar{Q}_i$  пусто. В этом случае согласно (1.3.10) вектор  $p_i$  явно не зависит от других компонент вектора  $p$  и выражается только через  $x$  и  $u$  по формуле

$$p_i = W_{x_i}(x, u). \quad (1.3.12)$$

Будем говорить, что *многошаговый процесс* (1.3.2) *явный*, если для каждого  $j \in D$  множество индексов  $\bar{Q}_j$  таково, что для любого элемента  $i \in \bar{Q}_j$  имеет место неравенство  $i < j$ . Поэтому если на  $j$ -м шаге из (1.3.2) был определен вектор  $x_j$ , то он может войти в качестве аргумента в правые части (1.3.2) только на последующих шагах  $i$ , у которых номер  $i > j$ , т.е. для любого  $i \in \bar{Q}_j$  имеет место неравенство  $i > j$ .

Для явных процессов из (1.3.4) следует, что всякая матрица  $N_{ij}$  может выражаться только через те матрицы  $N_{iq}$ , у которых  $1 \leq q < j$ . Аналогично согласно (1.3.2) всякий вектор  $x_i$  может выражаться только через те  $x_j$ , у которых  $1 \leq j < i$ . Поэтому будем говорить, что  $x_i$  и  $N_{si}$  вычислены в прямом направлении (слева направо), так как в процессе их расчетов индекс  $i$  увеличивается от 1 до  $k$ . Явные формулы наиболее популярны и особенно часто используются в дискретных аппроксимациях задач оптимального управления (ОУ), где обыкновенные дифференциальные уравнения, описывающие управляемый процесс, интегрируются по явным схемам. Однако иногда приходится иметь дело с неявными формулами.

С другой стороны, если  $j \in \bar{Q}_i$ , то имеет место первая функциональная зависимость (1.3.3), а так как процесс (1.3.2) явный, то для него  $j > i$  для всех  $i \in \bar{Q}_j$ . Поэтому для явных процессов в формуле (1.3.10) каждый вектор  $p_i$  выражается только через те векторы  $p_j$ , у которых  $j > i$ .

На последнем шаге явного процесса (1.3.2) мы получаем вектор  $x_k$ , который не принадлежит ни одному из множеств  $X_j$ ,  $1 \leq j \leq k$ ; это означает, что  $x_k$  — выходной вектор процесса (1.3.2). Из формулы (1.3.12) имеем

$$p_k = W_{x_k}(x, u). \quad (1.3.13)$$

Следовательно, начиная с  $i = k$ , можно из (1.3.13) и (1.3.10) последовательно находить все векторы  $p_i$  и, используя (1.3.11), вычислять все искомые производные.

Так как при вычислении  $p_i$  индекс  $i$  изменяется от  $k$  до 1, то говорят, что векторы  $p_i$  и производные целевой функции определяются *в обратном направлении или сверху вниз, или справа налево, или ковариантным способом*. Это делает явные процессы особенно простыми для вычислений.

Подчеркнем, что есть некоторое различие между выражениями (1.3.4)–(1.3.11) и известными “прямыми” и “обратными” формулами дифференцирования, приведенными в публикациях по БАД [94–98, 111, 112]. Наши выражения приводят к последним, если процесс (1.3.2)

явный. Отличие имеет место, когда процесс (1.3.2) неявный. Например, если  $x_i \in X_i$ , то вектор  $p_i$  входит в правую и левую часть (1.3.10); аналогично матрицы  $N_{ij}$  содержатся в обеих частях (1.3.4). Здесь не удастся последовательно определять векторы  $p_i$  и матрицы  $N_{ij}$ ; вместо этого приходится решать линейные системы (1.3.4) или (1.3.10) в целом, привлекая помимо  $p_i$  другие составляющие вектора  $p$ . Чтобы гарантировать разрешимость линейной системы (1.3.10) приходится вводить условия регулярности ограничений (1.3.2). Поэтому в общем случае приведенные здесь формулы не могут быть названы ни “прямыми”, ни “обратными”.

Имеется наглядное представление процесса вычисления значений функций и их производных в виде ориентированного графа, состоящего из  $D$  узлов. С каждым  $j$ -м узлом графа свяжем векторы  $x_j$  и  $u_j$ . В  $j$ -м узле по формуле (1.3.2) вычисляется  $x_j$ . Из всех узлов с номерами, содержащимися во множестве  $Q_j$ , проводятся ребра, входящие в узел  $j$ . В свою очередь, из узла  $j$  выходят ребра во все узлы с номерами из  $\overline{Q}_j$ . Таким образом,  $\overline{Q}_j$  есть совокупность ребер, инцидентных узлу  $j$ . Можно сказать, что  $Q_j$  — совокупность индексов всех векторов  $x_i$ ,  $i \in D$ , входящих в  $j$ -й узел;  $\overline{Q}_j$  — совокупность номеров всех тех узлов, в которые входит вектор  $x_j$ . В случае использования неявных схем возможно, что  $j \in Q_j$  и в графе появляются петли. Описанный граф можно назвать “информационным”. Он отражает функциональные зависимости между компонентами вектора  $x$ . В простейших случаях явных схем расчеты ведутся с использованием возрастающего индекса  $j$  (слева направо). В общем случае, когда управляемый процесс описывается уравнениями с частными производными, в задаче есть краевые условия; тогда необходимо дополнительно указать, в какой последовательности ведутся вычисления компонент вектора  $x$ , т.е. задать дополнительно так называемый “расчетный граф”.

Если с каждым  $i$ -м узлом связать процесс вычисления  $p_i$  по формулам (1.3.10), то получим точно такой же граф, что и для вычисления  $x_i$ , но направления всех ребер будут противоположными тем, которые имеются в информационном графе.

Ниже мы рассмотрим простейшие примеры, показывающие особенности подхода, применяемого к дифференцированию сложных функций. Во многих случаях применение формул, использующих вспомогательный вектор  $p$ , существенно менее трудоемко, чем аналогичные расчеты с использованием матрицы  $x_u^T$ .

В оставшейся части этого раздела мы считаем, что процесс (1.3.2) явный. Обозначим

$$W^k(x_1, x_2, \dots, x_k, u) = W(x, u), \quad D = \{1, 2, \dots, k\}.$$

Используя (1.3.9), выразим последний фазовый вектор  $x_k$  через предыдущие векторы  $x_{k-1}, x_{k-2}, \dots, x_1$  и подставим это выражение в  $W$ . Определим новую функцию

$$W^{k-1}(x_1, x_2, \dots, x_{k-1}, u) = W^k(x_1, x_2, \dots, x_{k-1}, F(k, X_k, U_k), u).$$

Аналогично с помощью обратной рекурсии определим последовательность скалярных функций  $W^i$ :

$$\begin{aligned} W^i(x_1, x_2, \dots, x_i, u) &= W^{i+1}(x_1, x_2, \dots, x_i, F(i+1, x_{i+1}, U_{i+1}), u), \\ W^1(u) &= W^1(F(1, X_1, U_1), u), \quad 1 \leq i \leq k-1. \end{aligned} \tag{1.3.14}$$

Здесь реализуется процесс последовательного исключения фазовых векторов  $x_i$ .

**Теорема 1.3.1.** *Предположим, что:*

1) все отображения  $F(j, X_j, U_j)$ ,  $1 \leq j \leq k$ , и скалярная функция  $W$  дифференцируемы по  $x$  и  $u$ ;

2) многошаговый процесс (1.3.2) явный.

Тогда все векторы  $p_i$  однозначно определяются из рекуррентных соотношений (1.3.10) и (1.3.13) и

$$p_i = \frac{\partial W^i(x_1, x_2, \dots, x_i, u)}{\partial x_i}, \quad 1 \leq i \leq k. \quad (1.3.15)$$

**Доказательство.** Рассмотрим матрицу  $G_x^\top(x, u)$  для явного процесса (1.3.2). Определим подматрицу размера  $s \times s$

$$Y_{ij} = \frac{\partial(F^\top(j, X_j, U_j) - x_j^\top)}{\partial x_i} = \frac{\partial G^j}{\partial x_i}.$$

Очевидно, что  $Y_{ij} = 0_{ss}$  при  $j < i$ . Если  $i = j$ , то  $Y_{ij} = -I_s$ . Следовательно,  $Y_{ij}$  — верхняя треугольная матрица, и уравнение (1.3.10) может быть переписано следующим образом:

$$p_i = W_{x_i}(x, u) + \sum_{j=i+1}^k F_{x_i}^\top(j, X_j, U_j)p_j = W_{x_i}(x, u) + \sum_{j \in \bar{Q}_j} F_{x_i}^\top(j, X_j, U_j)p_j, \quad 1 \leq i \leq k. \quad (1.3.16)$$

Линейная алгебраическая система (1.3.16) имеет единственное решение. Принимая во внимание определение сопряженного индексного множества, получим, что система (1.3.16) может быть решена с помощью обратных формул (1.3.10). Эти вычисления подобны “обратной подстановке”, которая используется в методе исключения Гаусса для решения системы линейных уравнений.

Дифференцируя (1.3.14) по  $x_i$  и применяя цепное правило дифференцирования, получим

$$\frac{\partial W^i(x_1, x_2, \dots, x_i, u)}{\partial x_i} = W_{x_i}(x, u) + \sum_{j \in \bar{Q}_i} F_{x_i}^\top(j, X_j, U_j) \frac{\partial W^j(x, u)}{\partial x_j}, \quad 1 \leq i \leq k. \quad (1.3.17)$$

Сравнивая выражения (1.3.10) и (1.3.17), можно сделать вывод о том, что утверждение (1.3.15) выполнено. Теорема 1.3.1 доказана.  $\square$

Ниже мы рассмотрим примеры, иллюстрирующие характерные свойства различных подходов, представленных для оценки градиентов. Для явных процессов во многих случаях обратный способ вычисления имеет преимущество перед прямым способом.

В ряде прикладных задач условия связи (1.3.2) задаются в неявном виде как последовательность условий

$$\Phi(j, X_j, U_j) = 0, \quad j \in D, \quad (1.3.18)$$

где смысл множеств  $X$  и  $U$  остался тем же, что и в соотношениях (1.3.2), целевой функцией является по-прежнему  $W(x, u)$ .

Этот случай сводится к рассмотренному выше с помощью представления условий (1.3.18) в виде

$$x_j = x_j + \Phi(j, X_j, U_j) = F(j, X_j, U_j), \quad j \in D.$$

Обозначим

$$\bar{E}(x, u, p) = W(x, u) + \sum_{j \in D} \Phi^\top(j, X_j, U_j)p_j.$$

Тогда соотношения (1.3.18), уравнения для определения импульсов и приведенный градиент функции  $W$  запишутся в виде

$$\begin{aligned}\bar{E}_{p_i}(x, u, p) &= \Phi(i, X_i, U_i) = 0, \quad i \in D, \\ \bar{E}_{x_i}(x, u, p) &= W_{x_i}(x, u) + \sum_{j \in \bar{Q}_i} \Phi_{x_i}^\top(j, X_j, U_j) p_j = 0,\end{aligned}\tag{1.3.19}$$

$$\frac{d\Omega(u)}{du_i} = \bar{E}_{u_i}(x, u, p) = W_{u_i}(x, u) + \sum_{j \in \bar{S}_i} \Phi_{u_i}^\top(j, X_j, U_j) p_j.\tag{1.3.20}$$

Такая форма записи особенно удобна при рассмотрении дискретных вариантов систем, описываемых уравнениями с частными производными (см. главу 4, § 4.1–§ 4.3).

## 1.4. Дифференцирование элементарных функций

В вычислительной практике для нахождения производных обычно используется приближенное численное дифференцирование. Если  $f(u)$  — функция  $r$ -мерного вектора  $u$ , то использование самой простой формулы численного дифференцирования требует по меньшей мере  $r + 1$  раз вычислить значение функции  $f(u)$  и  $r$  раз произвести вычитание и деление. Поэтому время, требуемое для численного дифференцирования функции, отнесенное ко времени вычисления значения функции, можно считать равным  $r + 1$  (здесь мы пренебрегли временем, необходимым для выполнения операций вычитания, деления, специального подбора шага дифференцирования). Результаты таких расчетов будут точными только для линейных функций; во всех остальных случаях градиент определяется с некоторой погрешностью.

Второй подход к дифференцированию состоит в использовании символьного дифференцирования. Однако при применении этой техники возникают большие трудности, если  $u$  — вектор большой размерности.

В настоящем параграфе будет описан третий подход, основанный на применении техники БАД. Ниже покажем, что при определенных условиях использование БАД приводит к замечательному результату: время, необходимое для вычисления градиента, отнесенное ко времени вычисления значения функции, не превосходит 3. Это свойство БАД открывает широкие возможности для применения градиентных методов безусловной минимизации (БМ) функций многих переменных, для решения задач НЛП и задач оптимального управления системами, описываемыми обыкновенными дифференциальными уравнениями и уравнениями с частными производными.

В этом разделе мы применим формулы, полученные выше, к дифференцированию элементарных функций векторного аргумента. Напомним, что функции  $a^x$  ( $a > 0$ ),  $x^\alpha$ ,  $\log_a x$  ( $a > 0, a \neq 1$ ),  $\sin x$ ,  $\cos x$ ,  $\operatorname{tg} x$ ,  $\operatorname{ctg} x$ ,  $\arcsin x$ ,  $\arccos x$ ,  $\operatorname{arctg} x$ ,  $\operatorname{arcctg} x$ ,  $\operatorname{sh} x$ ,  $\operatorname{ch} x$ ,  $\operatorname{th} x$  называются *основными элементарными функциями*. Предполагаем, что в компьютере содержатся программы, позволяющие точно (или с машинной точностью) вычислять значения основных элементарных функций и их производных. В случае необходимости к указанному списку основных элементарных функций могут быть добавлены программы вычисления значений и производных новых дифференцируемых функций.

Считаем, что дифференцируемая скалярная функция  $f(u)$  векторного аргумента  $u \in \mathbb{R}^r$  задана явно. Функция  $f(u)$  называется *элементарной функцией*, если она может быть представлена как конечная суперпозиция основных элементарных функций и арифметических действий. Наша задача состоит в вычислении частных производных функции  $f(u)$  по  $u$ .

Введем вектор  $x \in \mathbb{R}^k$  промежуточных (вспомогательных) фазовых переменных. Вычисление значения функции  $f(u)$  будем представлять как  $k$ -шаговый вычислительный процесс

$$x_1 = F(1, X_1, U_1), \quad x_2 = F(2, X_2, U_2), \dots, \quad x_k = F(k, X_k, U_k),\tag{1.4.1}$$

где все компоненты  $x_j \in \mathbb{R}^1$ , а последняя компонента  $x_k$  совпадает со значением функции  $f$ , т.е.  $x_k = f(u)$ ;  $X_i$  и  $U_i$  суть наборы некоторых компонент векторов  $x$  и  $u$  соответственно.

В многошаговом процессе (1.4.1) все компоненты  $x_i$  ( $1 \leq i \leq k$ ) вектора  $x$  являются скалярами, их можно записывать как  $x^i$ , т.е.  $i$ -е компоненты вектора  $x$ . Но для простоты изложения будем использовать обозначения предыдущего параграфа. Поэтому в соответствии с (1.3.1) применим нижние индексы. Символы  $x_i$  и  $u_i$  будут обозначать  $i$ -е компоненты векторов  $x$  и  $u$  соответственно. Искомые частные производные  $\partial f/\partial u^i$  будем записывать в виде  $\partial f/\partial u_i$ .

Последовательность (1.4.1) строится таким образом, чтобы процесс был явным. Поэтому всякий индекс  $i \in Q_j$  таков, что  $i < j$ . Следовательно, множества  $Q_1$  и  $\bar{Q}_k$  пустые. Предполагаем, что все функции  $F(j, X_j, U_j)$  дифференцируемы по всем компонентам векторов  $x$  и  $u$ . Введем скаляры  $p_j \in \mathbb{R}^1$ ,  $j \in D$ , и определим функцию

$$E(x, p, u) = x_k + \sum_{j=1}^k F(j, X_j, U_j)p_j.$$

В формулах (1.3.9) – (1.3.11) положим  $W = x_k$ ; тогда из (1.3.13) следует, что  $p_k = 1$ . Используя (1.3.10), последовательно (в обратном порядке) определим все  $p_i$ , начиная с  $i = k$  и заканчивая  $i = 1$ ; одновременно по формулам (1.3.11) найдем искомый градиент функции  $f(u)$ . Таким образом, получим следующие выражения:

$$p_i = \sum_{j \in \bar{Q}_i} F_{x_i}^\top(j, X_j, U_j)p_j, \quad 1 \leq i \leq k-1, \quad p_k = 1, \quad (1.4.2)$$

$$\frac{\partial f(u)}{\partial u_i} = \sum_{j \in \bar{S}_i} F_{u_i}^\top(j, X_j, U_j)p_j, \quad 1 \leq i \leq k. \quad (1.4.3)$$

Эти формулы называются *формулами быстрого автоматического дифференцирования* (БАД) элементарной функции  $f(u)$ .

Многие публикации анализируют алгоритмы автоматического дифференцирования с точки зрения алгебраической сложности вычислений, т.е. определяется общее число арифметических действий, требуемых для вычисления значения функции и ее производных. Результаты таких исследований можно найти в работах [94–104] и [112].

Ниже приводится теорема, дающая оценку сложности вычисления градиента по формулам БАД.

Пусть  $T_0$  — полное время, требуемое для вычисления значения функции  $f(u)$ ,  $T_g$  обозначает дополнительное время, требуемое для вычисления всех частных производных  $\partial f(u)/\partial u_i$ ,  $1 \leq i \leq r$ . При определении  $T_0$  и  $T_g$  будем учитывать только время выполнения арифметических операций; время, необходимое для работы с памятью компьютера и исполнения оператора присваивания, во внимание не принимается.

**Теорема 1.4.1.** *Предположим, что:*

- 1)  $f(u)$  — скалярная элементарная дифференцируемая функция векторного аргумента  $u \in \mathbb{R}^r$ ;
- 2) арифметические действия и вычисление основных элементарных функций и их производных производятся точно;
- 3) время вычисления производной каждой основной элементарной функции не больше, чем удвоенное время вычисления значения этой функции.

Тогда градиент функции  $f$ , найденный по формулам (1.4.2) и (1.4.3), определяется точно и отношение  $R = T_g/T_0$  ограничено сверху числом 3.



**Доказательство.** Предположим, что  $f$  — композиция из  $K$  основных элементарных функций, операторы  $F(i, X_i, U_i)$ , используемые при вычислении значения функции, являются унарными или бинарными. В первом случае  $F$  является какой-либо основной элементарной функцией одного аргумента. Во втором случае  $F$  — арифметическая операция между двумя переменными или одной переменной и константой. Мы называем эти операции *базисными*. Предположим, что  $T_i$  — время, требуемое для вычисления  $i$ -й основной элементарной функции, и  $N_i$  — количество вычислений  $i$ -й основной элементарной функции в процессе (1.4.1). Время, необходимое для выполнения операций сложения, вычитания, умножения и деления, обозначим соответственно  $T_+$ ,  $T_-$ ,  $T_\times$ ,  $T_/$ . Пусть  $N_+$ ,  $N_-$ ,  $N_\times$ ,  $N_/$  — соответственно количество сложений, вычитаний, умножений и делений среди бинарных операций. Предполагаем, что  $T_+ = T_- \leq T_\times \leq T_/ \leq T_i$  для любой  $i$ -й основной элементарной функции. Полное количество шагов в многошаговом процессе (1.4.2) и полное время  $T_0$  вычисления значения функции  $f(u)$  равны

$$k = \sum_{i=1}^K N_i + N_+ + N_- + N_\times + N_/,$$

$$T_0 = \sum_{i=1}^K N_i T_i + N_+ T_+ + N_- T_- + N_\times T_\times + N_/ T_/.$$

Подсчитаем время, необходимое для расчета градиента по формулам (1.4.2) и (1.4.3). Если при реализации какой-либо формулы необходимо продифференцировать  $i$ -ю основную элементарную функцию, то после этого производная умножается на  $p_q$  и результат добавляется к какой-либо сумме в правой части (1.4.2) либо (1.4.3). Время выполнения таких действий не превосходит  $2T_i + T_\times + T_+$ , а полное время дифференцирования унарных операций и их “обработки” не превысит  $\sum_{i=1}^K N_i(2T_i + T_\times + T_+)$ .

Рассмотрим дифференцирование бинарных операторов. Считаем, что операции дифференцирования предшествовало вычисление всех значений функций  $F(j, U_j, X_j)$  и все они хранятся в памяти ЭВМ. Каждый бинарный оператор умножается на некоторое число  $p_j$  и дифференцируется по обоим аргументам, результат прибавляется к какой-либо правой части в (1.4.2) или в (1.4.3). Время, необходимое для учета одного оператора сложения и вычитания, не превышает  $2(T_+ + T_-)$ . Время, необходимое для учета произведения, равно  $2(T_\times + T_+)$ .

Определим время дифференцирования отношения двух переменных. Пусть, например,  $F(s, U_s, X_s) = x_q/u_c$ . В расчетах будут фигурировать два выражения:

$$A = p_s \frac{\partial F}{\partial x_q} = \frac{p_s}{u_c}, \quad C = -p_s \frac{\partial F}{\partial u_c} = \frac{p_s x_q}{(u_c)^2} = A \cdot B, \quad \text{где } B = \frac{x_q}{u_c}.$$

Таким образом, вычисление  $A$  и  $C$  свелось к одному делению и одному умножению, т.е. требуется время, не превышающее  $2T_/$ . Эти выражения следует присвоить или, в худшем случае, прибавить или вычесть. Поэтому полное время выполнения операций деления не превосходит  $2(T_/ + T_+)$ . Здесь всюду считалось, что при расчете значений  $p_i$ ,  $1 \leq i \leq k-1$ , по формулам (1.4.2) все выражения в правых частях прибавлялись к некоторым текущим выражениям. В то же время при вычислении первого слагаемого оно не прибавляется, а присваивается. Поэтому полное время, необходимое для вычисления градиента, можно умень-

шать на  $kT_+$ . Следовательно, получается оценка

$$R = \frac{T_g}{T_0} \leq \left[ \sum_{i=1}^k N_i(2T_i + T_x + T_+) + 2N_+T_+ + 2N_-T_- + 2N_x(T_x + T_+) + 2N_/(T_+ + T_+) - kT_+ \right] / T_0 \leq M,$$

где

$$M = \left[ \sum_{i=1}^k N_i(2T_i + T_x) + N_+T_+ + N_-T_- + N_x(2T_x + T_+) + N_/(2T_+ + T_+) \right] / T_0.$$

Введенное выражение  $M$  представляет собой нелинейную функцию от  $N_i, N_+, N_-, N_x, N_+$ . Вместе с тем эти величины входят в числитель и знаменатель линейно. Чтобы найти верхнюю границу  $M$ , мы максимизируем  $M$  по этим переменным, т.е. рассматриваем задачу дробно-линейного программирования [67], решение которой очевидно: следует взять максимум из отношений коэффициентов, стоящих при  $N_i, N_+, N_-, N_x, N_+$  в числителе, к соответствующим коэффициентам в знаменателе. Принимая во внимание, что  $T_x \leq T_i$ , получаем следующую оценку:

$$R \leq M \leq \max \left[ 3, \max_{1 \leq i \leq k} \frac{2T_i + T_x}{T_i} \right] = 3,$$

что и следовало доказать.  $\square$

Теорема 1.4.1 показывает высокую эффективность БАД по сравнению с другими способами дифференцирования и дает верхнюю оценку отношения  $R$ . В ряде случаев эта величина меньше 3. В приведенном ниже примере 1.4.1 отношение  $R = 2$ . Вместе с тем в теореме 1.4.1 не учитывалось время, необходимое для работы с памятью компьютера, и время, требуемое для выполнения оператора присваивания. Поэтому при практической реализации формул БАД на компьютере величина  $R$  превосходит число 3. Близкое утверждение высказывалось и доказывалось многими авторами при различных допущениях (см., например, [94–104]). Изложенная здесь формулировка теоремы 1.4.1 взята из статьи автора [112].

В ряде случаев целесообразно расширить список основных элементарных функций, представляя некоторые арифметические выражения как дополнительные основные элементарные функции. Согласно теореме 1.4.1 для таких новых функций отношение  $T_g/T_0 \leq 3$ . Поэтому условие 3 оказывается невыполненным, вместо него следует взять следующее условие: *время вычисления производной каждой основной элементарной функции не превышает утроенное время, необходимое для определения значения этой функции*. Несложно показать, что в этом случае отношение времени вычисления градиента ко времени вычисления значения сложной функции не превышает 4.

**Пример 1.4.1.** В качестве простейшего примера рассмотрим дифференцирование произведения. Пусть  $f(u) = \prod_{i=1}^r u_i$ . Эта функция порождает следующий вычислительный процесс:

$$x_1 = u_1, \quad x_i = x_{i-1}u_i, \quad 2 \leq i \leq r, \quad W = f(u) = x_r.$$

Используя формулы (1.3.9), (1.3.10), получим

$$E = x_r + u_1p_1 + \sum_{i=2}^r x_{i-1}u_i p_i,$$

$$p_r = 1, \quad p_i = u_{i+1}p_{i+1}, \quad 1 \leq i \leq r-1,$$

$$\partial f / \partial u_1 = p_1, \quad \partial f / \partial u_j = x_{j-1}p_j, \quad 2 \leq j \leq r.$$

Очевидно, что здесь  $R = 2$ . В данном примере при численной реализации расчетов приходится вводить дополнительный  $r$ -мерный вектор  $p$ , но это обстоятельство не принималось во внимание при оценке относительной трудоемкости вычислений. ■

**Пример 1.4.2.** Рассмотрим дифференцирование оператора цикла. Пусть в описании функции  $f(u)$  содержится оператор цикла вида

$$\alpha(u) = \sum_{i=1}^{\ell} \varphi(u_i). \quad (1.4.4)$$

Если применить технику БАД непосредственно к выражению (1.4.4), то будет реализован некоторый  $l$ -шаговый процесс вычисления значения функции  $\alpha(u)$  и нахождения ее градиента. Другой вариант расчета состоит в написании процедуры  $procedure(i, \varphi(u_i), \partial\varphi(u_i)/\partial u_i)$ , зависящей от параметра  $i$ . Эта процедура при заданном  $i$  реализует  $k$ -шаговый процесс вычисления значения функции  $\varphi(u_i)$  и ее производной  $\partial\varphi(u_i)/\partial u_i$ . Производную функции  $\alpha(u)$  по  $u$  можно получить в результате работы следующего цикла:

**for**  $i := 1$  **step** 1 **until**  $\ell$  **do**  $procedure(i, \varphi(u_i), \partial\varphi(u_i)/\partial u_i)$ .

В процессе выполнения этого цикла  $\ell$  раз будет реализован  $k$ -шаговый процесс вычисления значения  $\varphi(u_i)$  и производной  $\partial\varphi(u_i)/\partial u_i$ . В результате будет получен полный набор частных производных  $\partial\varphi(u_i)/\partial u_i$ , т.е. градиент функции  $\alpha(u)$ . Такой прием не изменяет объем вычислений, но уменьшает требуемую память компьютера. Экономия происходит за счет того, что в (1.4.4) под знаком суммы стоят однотипные выражения, отличающиеся только параметром  $i$ . Благодаря этому “обработка” в отдельности каждой функции  $\varphi(u_i)$  ведется одной и той же процедурой, использующей одни и те же ячейки памяти.

Дифференцирование операторов суммы, как это имеет место в (1.4.4), требует большего объема памяти, чем последовательное применение процедуры дифференцирования к каждому отдельному слагаемому. Можно сказать, что с точки зрения рационального использования памяти второй подход при дифференцировании оператора цикла экономнее, чем последовательное независимое дифференцирование всех слагаемых. ■

**Пример 1.4.3.** Рассмотрим простейшую элементарную функцию

$$f(u) = e^{u_1+u_2} + \sin(u_1 + u_2)^2.$$

Вычисление значения этой функции представим как многошаговый процесс вида (1.3.2)

$$\begin{aligned} x_1 &= u_1 + u_2, \\ x_2 &= e^{x_1}, \\ x_3 &= (x_1)^2, \\ x_4 &= \sin x_3, \\ x_5 &= x_2 + x_4 = f(u). \end{aligned} \quad (1.4.5)$$

Приведем “входящие” индексные наборы  $Q_i, S_i$  и “выходящие” наборы  $\overline{Q}_i, \overline{S}_i$  для этого процесса:

$$\begin{array}{llll} Q_1 &= \emptyset, & S_1 &= [1, 2], & \overline{Q}_1 &= [2, 3], & \overline{S}_1 &= [1], \\ Q_2 &= [1], & S_2 &= \emptyset, & \overline{Q}_2 &= [5], & \overline{S}_2 &= [1], \\ Q_3 &= [1], & S_3 &= \emptyset, & \overline{Q}_3 &= [4], & \overline{S}_3 &= \emptyset, \\ Q_4 &= [3], & S_4 &= \emptyset, & \overline{Q}_4 &= [5], & \overline{S}_4 &= \emptyset, \\ Q_5 &= [2, 4], & S_5 &= \emptyset, & \overline{Q}_5 &= \emptyset, & \overline{S}_5 &= \emptyset, \end{array}$$

где  $\emptyset$  обозначает пустое множество.

Введем множители Лагранжа  $p_1, p_2, \dots, p_5$  и определим функцию  $E$ :

$$E(x, u, p) = x_5 + p_1(u_1 + u_2) + p_2 e^{x_1} + p_3(x_1)^2 + p_4 \sin x_3 + p_5(x_2 + x_4).$$

Согласно (1.4.2) множители Лагранжа находятся из следующих условий:

$$\begin{aligned} p_1 &= p_2 e^{x_1} + 2p_3 x_1, \\ p_2 &= p_5, \\ p_3 &= p_4 \cos x_3, \\ p_4 &= p_5, \\ p_5 &= 1. \end{aligned} \tag{1.4.6}$$

Разрешим эту линейную систему в обратном направлении, начиная с  $p_5$ , и заканчивая  $p_1$ , получим

$$p_5 = p_4 = p_2 = 1, \quad p_3 = \cos x_3, \quad p_1 = e^{x_1} + 2x_1 \cos x_3. \tag{1.4.7}$$

Поэтому формула (1.4.3) приводит к следующему результату:

$$E_{u_1} = p_1 = e^{u_1+u_2} + 2(u_1 + u_2) \cos(u_1 + u_2)^2 = \partial f(u)/\partial u_1 = \partial f(u)/\partial u_2. \tag{1.4.8}$$

В соответствии с введенной терминологией здесь полный фазовый вектор  $x$  состоит из пяти компонент, полный вектор управлений  $u$  состоит лишь из двух компонент. Привязка  $u_i$  к определенным шагам процесса произвольна.

В обратном направлении определяем последовательность функций  $W^i$  и их производных, учитывая (1.3.15), получаем те же результаты, что даны в (1.4.7).

$$\begin{aligned} W^5 &= x_5, & \partial W^5/\partial x_5 &= 1, & &= p_5 \\ W^4 &= x_2 + x_4, & \partial W^4/\partial x_4 &= 1, & &= p_4 \\ W^3 &= x_2 + \sin x_3, & \partial W^3/\partial x_3 &= \cos x_3, & &= p_3 \\ W^2 &= x_2 + \sin(x_1)^2, & \partial W^2/\partial x_2 &= 1, & &= p_2 \\ W^1 &= e^{x_1} + \sin(x_1)^2, & \partial W^1/\partial x_1 &= e^{x_1} + 2x_1 \cos x_3 &= & p_1. \end{aligned}$$

Вычислим искомые производные с помощью формул “прямого” дифференцирования (1.3.4). Несложные расчеты дают следующие соотношения:

$$\begin{aligned} N_{11} &= \partial x_1/\partial u_1 = 1, & N_{21} &= \partial x_1/\partial u_2 = 1, \\ N_{12} &= \partial x_2/\partial u_1 = N_{11} e^{x_1}, & N_{22} &= \partial x_2/\partial u_1 = N_{21} e^{x_1}, \\ N_{13} &= \partial x_3/\partial u_1 = 2N_{11} x_1, & N_{23} &= \partial x_3/\partial u_1 = 2N_{21} x_1, \\ N_{14} &= \partial x_4/\partial u_1 = N_{13} \cos x_3, & N_{24} &= \partial x_4/\partial u_1 = N_{23} \cos x_3, \\ N_{15} &= N_{12} + N_{14}, & N_{25} &= N_{22} + N_{24}. \end{aligned} \tag{1.4.9}$$

Решая эту систему, получим

$$\begin{aligned} N_{11} &= 1 = N_{21}, & N_{12} &= e^{x_1} = N_{22}, & N_{13} &= 2x_1 = N_{23}, \\ N_{14} &= 2x_1 \cos x_3 = N_{24}, & N_{15} &= e^{x_1} + 2x_1 \cos x_3 = N_{25} = \partial f/\partial u_1 = \partial f/\partial u_2. \end{aligned}$$

Таким образом, мы снова пришли к формуле (1.4.8). Обычно говорят, что расчеты в (1.4.5) и (1.4.9) ведутся *слева направо* (“сверху вниз”, “top-down”), а по формулам (1.4.6) — *справа*

налево (“снизу вверх”, “bottom-up”).

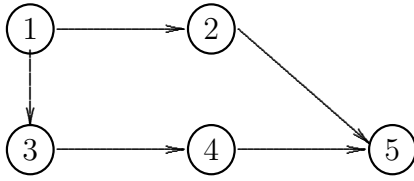


Рис. 1.4.1.  $x$ -граф

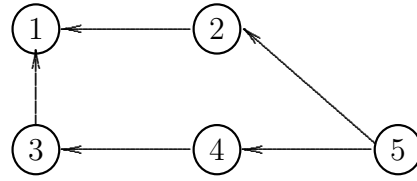


Рис. 1.4.2.  $p$ -граф

Информационный граф ( $x$ -граф) для процесса (1.4.5) изображен на рис. 1.4.1. На рис. 1.4.2 приведен  $p$ -граф, показывающий функциональную зависимость составляющих вектора  $p$ . Графы отличаются только направлением стрелок. ■

**Пример 1.4.4.** Пусть в формуле для вычисления  $f(u)$  встречаются скалярные произведения компонент вектора  $u$  на некоторый вектор констант. Расчеты производных в этом случае можно несколько упростить. Пусть в описании  $f(u)$  есть выражение

$$c + \sum_{s=\alpha}^{s=\alpha+\beta} a_s u_s.$$

В соответствии с представлением (1.4.1) необходимо ввести следующие составляющие вектора  $x$ :

$$\begin{array}{ll} x_\alpha & = a_\alpha u_\alpha, & x_{\alpha+1} & = c + x_\alpha, \\ x_{\alpha+2} & = a_{\alpha+1} u_{\alpha+1}, & x_{\alpha+3} & = x_{\alpha+1} + x_{\alpha+2}, \\ \dots & \dots & \dots & \dots \\ x_{i+2\beta} & = a_{\alpha+\beta} u_{\alpha+\beta}, & x_{\alpha+2\beta+1} & = x_{\alpha+2\beta-1} + x_{\alpha+2\beta}. \end{array}$$

Введенным  $x_\alpha, \dots, x_{\alpha+2\beta+1}$  соответствуют  $p_\alpha, \dots, p_{\alpha+2\beta+1}$ . Из (1.4.2) видно, что  $p_\alpha = p_{\alpha+1} = \dots = p_{i+2\beta+1}$ . Поэтому в рассматриваемом случае целесообразно отказаться от условия представления многошагового процесса как совокупности унарных и бинарных операций и ввести только одну вспомогательную составляющую вектора  $x$ :

$$x_\alpha = c + \sum_{s=\alpha}^{s=\alpha+\beta} a_s u_s.$$

При этом к текущим производным  $\partial\Omega/\partial u_s$  прибавим выражения  $p_\alpha a_s$ , где  $\alpha \leq s \leq \alpha + \beta$ . Этот прием сокращает число  $k$ , тем самым уменьшая размерность вектора  $p$ . При таком расширении понятия базисных операций может оказаться, что  $R = T_g/T_0 < 1$ . Можно сказать, что скалярное произведение легче продифференцировать, чем вычислить. Поэтому скалярные произведения и произведения матриц на векторы можно оформлять в виде отдельных процедур, считая их аналогом основных элементарных функций. ■

## 1.5. Оценка погрешностей

Предположим, что на каждом  $i$ -м шаге процесса (1.3.2) вектор состояния  $x_j$  определяется с погрешностью  $\varepsilon_j$ . Таким образом, вместо (1.3.2) используется следующая формула:

$$x_j = F(j, X_j, U_j) + \varepsilon_j, \quad 1 \leq j \leq k. \quad (1.5.1)$$

Вспомогательная функция  $E$  примет вид

$$E(u, x, \varepsilon) = W(x, u) + \sum_{j \in D} [F^\top(j, X_j, U_j) + \varepsilon_j^\top] p_j.$$

Здесь оба вектора  $x_j$  и  $p_j$  имеют одну и ту же размерность. Если процесс (1.5.1) явный, то можно считать, что вектор  $\varepsilon_j$  — машинная точность вычисления вектора  $F(j, X_j, U_j)$ . Если процесс (1.5.1) неявный, то  $\varepsilon_j$  представляет собой погрешность приближенного решения неявного уравнения (1.5.1) и может оказаться не малой величиной. Воспользуемся формулами (1.3.9)–(1.3.11), считая, что  $x$  и  $p$  являются сложными функциями вектора управления  $u$  и вектора  $\varepsilon^\top = [\varepsilon_1^\top, \varepsilon_2^\top, \dots, \varepsilon_k^\top]$ . Следовательно, можно записать  $\Omega(u, \varepsilon) = W(x(u, \varepsilon), u)$  и использовать канонические уравнения (1.3.10), (1.3.11). Рассматривая  $\varepsilon_i$  в этих выражениях так же, как рассматривались компоненты вектора управления  $u$ , получим точно те же выражения, что и (1.3.10), (1.3.11). Несмотря на то что ни один из векторов  $\varepsilon_j$ ,  $j \in D$ , не входит в правые части (1.3.10), векторы  $p_i$ , зависящие от  $x_j$ , будут также зависеть от  $\varepsilon$ , поэтому можно писать  $p_i = p_i(u, \varepsilon)$ . Используя (1.3.11), находим, что производная  $\Omega$  по  $\varepsilon_i$  есть

$$\frac{d\Omega(u, \varepsilon)}{d\varepsilon_i} = p_i(u, \varepsilon).$$

Предположим, что вектор управления  $u$  реализуется неточно и вместо  $u$  используется вектор  $\bar{u} = u + \delta$ . Тогда

$$\Omega(\bar{u}, \varepsilon) - \Omega(u, 0) = \sum_{i=1}^k \left[ \langle p_i(u, 0), \varepsilon_i \rangle + \left\langle \frac{d\Omega(u, 0)}{du_i}, \delta_i \right\rangle \right] + O(\|\varepsilon\|^2 + \|\delta\|^2),$$

где производные  $\Omega$  по  $u_i$  находятся из (1.3.11). Эту оценку можно использовать вместо формул интервального анализа. Теоретические и практические аспекты оценки погрешности с помощью БАД были исследованы М. Ири [117].

## 1.6. Приложение к линейной алгебре

Пусть  $A$  — квадратная матрица, состоящая из элементов  $a_{ij}$ ,  $1 \leq i \leq m$ ,  $1 \leq j \leq m$ . Предполагаем, что  $d$  — определитель матрицы  $A$  — отличен от нуля; тогда существует обратная матрица  $B = A^{-1}$ .

Определитель  $d$  матрицы  $A$  разлагается по элементам  $i$  строки матрицы  $A$  в виде

$$d = \sum_{j=1}^m a_{ij} A_{ij}; \quad (1.6.1)$$

здесь  $A_{ij} = (-1)^{(i+j)} M_{ij}$  — алгебраическое дополнение элемента  $a_{ij}$ ,  $M_{ij}$  — дополнительный минор элемента  $a_{ij}$ , т.е. минор  $(m-1)$ -го порядка, получившийся после вычеркивания из определителя матрицы  $A$   $i$ -й строки и  $j$ -го столбца.

Если  $d \neq 0$ , то существует матрица  $B = A^{-1}$ , причем каждый элемент  $b_{ji}$  этой матрицы вычисляется по формуле

$$b_{ji} = A_{ij}/d. \quad (1.6.2)$$

Определители  $A_{ij}$  и  $M_{ij}$  не зависят от элемента  $a_{ij}$ . Поэтому согласно (1.6.1) можно записать  $A_{ij} = \partial d / \partial a_{ij}$ . Учитывая (1.6.2), получаем следующую формулу для определения элемента матрицы, обратной к  $A$ :

$$b_{ji} = \frac{\partial \ln d}{\partial a_{ij}}.$$

Используя это выражение, по формулам БАД определим элементы обратной матрицы.

## Комментарии к главе 1

§ 1.1. Доказательство теоремы о неявных функциях приведены, например, в книгах А.М. Тер-Крикорова и М.И. Шабунина [83], Г.М. Фихтенгольца [89].

§ 1.2 – § 1.4. Формулы БАД можно найти во многих статьях и книгах. Укажем лишь некоторые из них, а именно [1, 65, 68, 94–104, 106, 111]. Данный здесь материал приведен в определенную систему, ориентированную на изложение наиболее общих формул БАД. Идейной основой изложения послужила статья автора “Вычисление точных градиентов в распределенных динамических системах”, опубликованная в 1998 г. в журнале “Методы оптимизации и софтвер” [112]. Вычисление вторых производных в случае многошаговых процессов приведено в [50, 53, 106]. На основе этих формул в [50] методом Ньютона решена простейшая тестовая задача оптимального управления с фазовыми ограничениями.

§ 1.4. Теорема 1.4.1 взята из [112]. Из других публикаций о методе БАД укажем [94, 100, 104, 106]. Численная реализация БАД была проведена несколькими коллективами. Первый пакет, “Адольф-С”, был создан в Аргоннской Национальной Лаборатории (США) под руководством А. Гриванка. В Японии под руководством М. Ири и К. Кубота [117, 118] был разработан пакет PADRE. Решение практических задач с помощью этих пакетов показало, что основным их недостатком была необходимость использования большой компьютерной памяти. Поэтому в настоящее время ведутся работы по сокращению требуемой памяти.

§ 1.5. Теоретические и практические аспекты оценки погрешности с помощью БАД были исследованы М. Ири [117, 118].

§ 1.6. Идея использовать технику БАД для определения обратной матрицы принадлежит японскому ученому К. Кубота.

## Глава 2. Приложения к задачам оптимизации

### 2.1. Задачи безусловной минимизации

Пусть вектор  $z \in \mathbb{R}^n$ , определена скалярная функция  $W(z) : \mathbb{R}^n \rightarrow \mathbb{R}^1$ . Ставится задача безусловной минимизации (БМ) функции  $W(z)$  по  $z$ :

$$\min_{z \in \mathbb{R}^n} W(z). \quad (2.1.1)$$

Точка  $z_*$  называется *глобальным решением задачи* (2.1.1), если для каждого  $z \in \mathbb{R}^n$  имеет место неравенство

$$W(z_*) \leq W(z).$$

Пусть  $\Gamma(z_*)$  — некоторая окрестность  $z_*$ ;  $z_*$  называется *точкой локального минимума* в задаче (2.1.1), если для каждого  $z \in \Gamma(z_*)$  имеет место неравенство

$$W(z_*) \leq W(z). \quad (2.1.2)$$

Если в (2.1.2) выполнено строгое неравенство для всех  $z \in \Gamma(z_*)$ ,  $z \neq z_*$ , то будем говорить, что  $z_*$  — *точка строгого (изолированного) локального минимума* в задаче (2.1.1). Аналогично ставится задача поиска максимума функции  $W(z)$  по  $z$ .

Предполагаем, что функция  $W(z)$  дифференцируема в точке  $z_*$ . Тогда она непрерывна и имеет место формула Тейлора с остаточным членом в форме Пеано

$$W(z) = W(z_*) + (z - z_*)^\top W_z(z_*) + \alpha(z_*, \|z - z_*\|) \|z - z_*\|, \quad (2.1.3)$$

где  $W_z(z_*)$  — градиент функции  $W$  в точке  $z_*$ , функция  $\alpha(\cdot)$  удовлетворяет условию

$$\lim_{z \rightarrow z_*} \alpha(z_*, \|z - z_*\|) = 0. \quad (2.1.4)$$

**Теорема 2.1.1 (Ферма).** Пусть функция  $W(z)$  достигает своего глобального или локального минимума или максимума в точке  $z_*$ , где она дифференцируема. Тогда

$$W_z(z_*) = 0_n. \quad (2.1.5)$$

**Доказательство.** Пусть, например, функция  $W(z)$  имеет в точке  $z_*$  локальный минимум. Обозначим  $q = W_z(z_*)$ . Предположим от противного, что  $\|q\| \neq 0$ . Используя (2.1.3), получим

$$W(z) = W(z_*) + (z - z_*)^\top q + \alpha(z_*, \|z - z_*\|) \|z - z_*\|.$$

Возьмем  $z = z_* - \tau q$ , где  $\tau$  — достаточно малое положительное число. Тогда для всякого  $z \in \Gamma(z_*)$  имеем

$$W(z) = W(z_*) + \tau \|q\| [-\|q\| + \alpha(z_*, \tau \|q\|)]. \quad (2.1.6)$$

Учитывая свойство (2.1.4), можно взять  $\tau_*$  настолько малым, что  $|\alpha(z_*, \tau \|q\|)| \leq \|q\|/2$  для всех  $0 \leq \tau \leq \tau_*$ . В случае поиска максимума функции  $W(z)$  из (2.1.2) и (2.1.6) следует

$$0 \leq W(z) - W(z_*) \leq -\frac{1}{2} \|q\|^2 < 0.$$

Противоречие доказывает утверждение теоремы о том, что  $\|q\| = 0$ . Аналогично рассматривается задача поиска максимума функции  $W(z)$ .  $\square$



Точки, в которых градиент функции обращается в нуль, называются *стационарными*. Теорему 2.1.1 можно переформулировать так: если в точке  $z_*$  дифференцируемая функция  $W$  достигает своего максимума или минимума, то  $z_*$  является стационарной точкой этой функции. Теорема Ферма дает необходимое условие экстремума дифференцируемой функции. Согласно теории Ферма точки локального экстремума функции  $f(x)$  следует искать среди тех точек области ее определения, в которых производная функции либо равна нулю, либо не существует.

**Теорема 2.1.2.** Пусть функция  $W$  дважды непрерывно дифференцируема в окрестности стационарной точки  $z_*$ . Тогда

1) если  $z_*$  — точка локального минимума функции  $W(z)$ , то матрица вторых производных  $W_{zz^\top}(z_*)$  положительно полуопределена, т.е.  $\xi^\top W_{zz^\top}(z_*)\xi \geq 0$  для любых  $\xi \in \mathbb{R}^n$ ;

2) если  $z_*$  — точка локального максимума функции  $W(z)$ , то матрица  $W_{zz^\top}(z_*)$  отрицательно полуопределена, т.е.  $\xi^\top W_{zz^\top}(z_*)\xi \leq 0$  для любых  $\xi \in \mathbb{R}^n$ ;

3) если матрица  $W_{zz^\top}(z_*)$  положительно определена (т.е.  $\xi^\top W_{zz^\top}(z_*)\xi > 0$  для всех  $\xi \in \mathbb{R}^n$ ,  $\|\xi\| \neq 0$ ), то  $z_*$  — точка строгого локального минимума функции  $W(z)$ ;

4) если матрица  $W_{zz^\top}(z_*)$  отрицательно определена (т.е.  $\xi^\top W_{zz^\top}(z_*)\xi < 0$  для всех  $\xi \in \mathbb{R}^n$ ,  $\|\xi\| \neq 0$ ), то  $z_*$  — точка строгого локального максимума функции  $W(z)$ .

**Доказательство.** Воспользуемся формулой Тейлора

$$\begin{aligned} W(z) &= W(z_*) + (z - z_*)^\top W_z(z_*) + \\ &+ (z - z_*)^\top W_{zz^\top}(z_*) \frac{(z - z_*)}{2} + \|z - z_*\|^2 \alpha(z_*, \|z - z_*\|). \end{aligned} \quad (2.1.7)$$

Доказательство всех утверждений теоремы производится однотипно. Поэтому ограничимся доказательством третьего утверждения. Из положительной определенности матрицы  $W_{zz^\top}(z_*)$  следует, что существует число  $\lambda_* > 0$  такое, что  $\xi^\top W_{zz^\top}(z_*)\xi > \lambda_* \|\xi\|^2$  для всех  $\xi \in \mathbb{R}^n$ ,  $\|\xi\| \neq 0$ . Учитывая условия стационарности (2.1.5), из (2.1.7) получаем

$$W(z) > W(z_*) + \|z - z_*\|^2 [\lambda_*/2 + \alpha(z_*, \|z - z_*\|)].$$

Возьмем окрестность  $\Gamma(z_*)$  настолько малой, что  $|\alpha(z_*, \|z - z_*\|)| \leq \lambda_*/4$  для всех  $z \in \Gamma(z_*)$ . Тогда для всех  $z \in \Gamma(z_*)$ ,  $z \neq z_*$ , имеем

$$W(z) > W(z_*) + \lambda_* \|z - z_*\|^2 / 4,$$

т.е.  $z_*$  — точка строгого локального минимума функции  $W(z)$ .  $\square$

Утверждения 1 и 2 теоремы 2.1.2 дают необходимые условия экстремума, утверждения 3 и 4 — достаточные условия.

Задача безусловной минимизации (2.1.1) решается особенно просто, если  $W(z)$  — выпуклая функция. Напомним, что функция  $W$  называется *выпуклой*, если для любых  $z, y \in \mathbb{R}^n$  выполнено неравенство

$$\lambda W(z) + (1 - \lambda)W(y) \geq W(\lambda z + (1 - \lambda)y), \quad (2.1.8)$$

где  $0 \leq \lambda \leq 1$ . Если при  $0 < \lambda < 1$ ,  $z \neq y$  неравенство (2.1.8) строгое, то функция  $W(z)$  называется *строго выпуклой*.

Если функция  $W(z)$  дифференцируема на  $\mathbb{R}^n$ , то для ее выпуклости необходимо и достаточно, чтобы для любых  $z$  и  $y$  выполнялись условия

$$W(z) - W(y) \geq [W_y(y)]^\top (z - y). \quad (2.1.9)$$

Меняя местами  $z$  и  $y$ , получим эквивалентное неравенство

$$W(y) - W(z) \geq [W_z(z)]^\top (y - z). \quad (2.1.10)$$

Объединяя эти неравенства, запишем условие выпуклости в симметричной форме:

$$[W_z(z)]^\top (z - y) \geq W(z) - W(y) \geq [W_y(y)]^\top (z - y). \quad (2.1.11)$$

Если функция  $W$  выпуклая и дифференцируемая, то (2.1.5) — необходимое и достаточное условие глобального минимума. Действительно, если в точке  $y$  выполнено условие (2.1.5), то из приведенного условия выпуклости (2.1.9) следует, что  $W(z) \geq W(y)$  для всех  $z$ .

Если  $W$  дважды непрерывно дифференцируема, то для ее выпуклости на  $\mathbb{R}^n$  необходимо и достаточно, чтобы матрица вторых производных  $W_{zz^\top}(z)$  была всюду положительно полуопределена, т.е. при любых  $z, y$  имеет место неравенство  $y^\top W_{zz^\top}(z)y \geq 0$ . Всякий локальный минимум выпуклой функции является одновременно глобальным.

*Субдифференциалом* выпуклой функции  $W(z)$  в точке  $y$  называется множество

$$\partial W(y) = \{q \in \mathbb{R}^n : W(z) - W(y) \geq q^\top (z - y) \text{ для всех } z \in \mathbb{R}^n\}. \quad (2.1.12)$$

Множество  $\partial W(y)$  — выпуклое, замкнутое и ограниченное. Всякий вектор  $q \in \partial W(y)$  называется *субградиентом* функции  $W$  в точке  $y$ , если для всех  $z \in \mathbb{R}^n$  выполнено неравенство

$$W(z) - W(y) \geq q^\top (z - y).$$

Если выпуклая функция  $W$  дифференцируема в точке  $y$ , то субдифференциал функции  $W(y)$  в точке  $y$  совпадает с градиентом функции в этой точке, т.е.  $W_y(y) = \partial W(y)$ ; если субдифференциал  $\partial W(y)$  состоит из единственного вектора  $q$ , то функция  $W$  дифференцируема в точке  $y$  и  $q = W_y(y)$ .

Геометрический смысл субградиента следующий: аффинная функция  $\varphi(z) = W(y) + q^\top (z - y)$  задает гиперплоскость, опорную ко множеству  $\text{epi } W$  в точке  $[y, W(y)]$ . Здесь эпиграф (надграфик) выпуклой функции  $W(y)$  есть множество точек из  $\mathbb{R}^{n+1}$ , расположенных на графике и выше графика  $W$ :

$$\text{epi } W = \{y \in \mathbb{R}^n, \mu \in \mathbb{R} : \Omega(y) \leq \mu\}.$$

**Теорема 2.1.3 (аналог теоремы Ферма).** *Для того чтобы  $z_*$  была точкой минимума выпуклой функции  $W(z)$ , необходимо и достаточно, чтобы*

$$0_n \in \partial W(z_*). \quad (2.1.13)$$

Действительно. если выполнено включение (2.1.13), то из (2.1.12) следует, что  $W(z) \geq W(z_*)$  для всех  $z$ . Поэтому  $z_*$  есть точка глобального минимума функции  $W(z)$ . С другой стороны, если  $z_*$  — точка глобального минимума, то для всех  $z$  выполнено неравенство

$$W(z) - W(z_*) \geq 0 = \langle 0_n, z - z_* \rangle,$$

т.е. имеет место (2.1.13).

Если выпуклая функция  $W(z)$  дифференцируема в точке  $z$ , то  $W_z(z) = \partial W(z)$ , и условие (2.1.13) переходит в необходимое условие минимума Ферма (2.1.5).

Функция  $f(x)$  называется *вогнутой*, если  $-f(x)$  — выпуклая функция. В случае вогнутых функций знаки неравенства в формулах (2.1.9)–(2.1.12) меняются на противоположные.

Если функция  $W(y)$  выпукла и дважды дифференцируема на  $\mathbb{R}^n$ , то матрица вторых производных  $W_{yy^\top}(y)$  положительно полуопределена, т.е. для любых  $z \in \mathbb{R}^n$  имеет место неравенство

$$z^\top W_{yy^\top}(y)z \geq 0. \quad (2.1.14)$$

Если функция  $W(y)$  строго выпукла по  $y$  на  $\mathbb{R}^n$  и дважды дифференцируема, то для любых векторов  $z$ , имеющих ненулевую норму, выполняется строгое неравенство

$$z^\top W_{yy^\top}(y)z > 0. \quad (2.1.15)$$

В упрощенной записи неравенства (2.1.14) и (2.1.15) будем записывать соответственно в виде

$$W_{yy^\top}(y) \geq 0_{nn}, \quad W_{yy^\top}(y) > 0_{nn}.$$

Для вогнутых функций знаки неравенства в (2.1.14) и (2.1.15) меняются на противоположные.

**Пример 2.1.1.** Пусть  $z \in \mathbb{R}^1$ ,  $W(z) = z_+$ ; тогда

$$\partial W(z) = \begin{cases} 0, & \text{если } z < 0; \\ 1, & \text{если } z > 0; \\ [0, 1], & \text{если } z = 0. \end{cases}$$

Здесь при  $z = 0$  можно взять любое число из интервала  $[0, 1]$ . При численных расчетах обычно берется  $1/2$ . Включение (2.1.13) имеет место при  $z \leq 0$ , что совпадает с множеством минимумов функции  $z_+$ . ■

**Пример 2.1.2.** Пусть  $z \in \mathbb{R}^n$ ,  $W(z) = \|z\|_2$ ; тогда

$$\partial W(0_n) = \{z : \|z\|_2 \leq 1\},$$

если  $\|z\| \neq 0$ , то  $\partial W(z) = \{z : z/\|z\|_2\}$ . ■

**Пример 2.1.3.** Пусть  $z \in \mathbb{R}^n$ ,  $W(z) = \|z\|_1 = \sum_{i=1}^n |z^i|$ , тогда

$$\begin{aligned} \partial W(0_n) &= \{z : \|z\|_\infty \leq 1\}, \\ \partial W(z) &= \sum_{i \in I_+(z)} e_i - \sum_{i \in I_-(z)} e_i + \sum_{i \in I_0(z)} [-e_i, e_i], \quad z \neq 0, \end{aligned}$$

где  $I_+(z) = \{i \mid z^{(i)} > 0\}$ ,  $I_-(z) = \{i \mid z^{(i)} < 0\}$  и  $I_0(z) = \{i \mid z^{(i)} = 0\}$ . ■

**Пример 2.1.4.** Пусть  $z \in \mathbb{R}^n$ ,  $W(z) = \|z\|_\infty$ ; тогда

$$\partial W(0_n) = \text{conv} \{e_i : 1 \leq i \leq n\},$$

если  $\|z\| \neq 0$ , то  $\partial W(z) = \text{conv} \{e_i : i \in I(z)\}$ , где  $I(z) = \{i : z^i = W(z)\}$ . ■

**Пример 2.1.5.** Пусть  $W(z) = \|(Az - b)_+\|^2/2$ , где  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ . Функция  $W(z)$  выпуклая, дифференцируемая, ее градиент  $\nabla W = A^\top (Az - b)_+$  недифференцируем, но можно выписать субдифференциал

$$\partial \nabla W(z) = A^\top D^\#(Az - b)A.$$

Здесь  $D^\#(a)$  — диагональная матрица, у которой на диагонали находится вектор  $a$  такой, что  $i$ -я компонента вектора  $a$  равна нулю, если  $a^i < 0$ ; равна единице, если  $a^i > 0$ ; если  $a^i = 0$ , то в качестве  $i$ -й компоненты можно взять любое число из интервала  $[0, 1]$ . Обычно берется  $a = 1/2$ . ■

## 2.2. Обобщенный критерий Сильвестра

Предположим, что функция  $W(z)$  дважды непрерывно дифференцируема в окрестности стационарной точки  $z_*$ , в которой  $W_z(z_*) = 0_n$ . Пусть  $\delta_1, \delta_2, \dots, \delta_n$  — последовательные главные миноры матрицы  $W_{zz^\top}(z_*)$ , у которой  $ij$ -й элемент есть

$$\frac{\partial^2 W(z_*)}{\partial z^i \partial z^j} = \frac{\partial^2 W(z_*)}{\partial z^j \partial z^i}.$$

Матрице  $W_{zz^\top}(z_*)$  соответствует квадратичная форма

$$\Phi(z) = z^\top W_{zz^\top}(z_*)z, \quad z \in \mathbb{R}^n.$$

Квадратичная форма  $\Phi(z)$  называется

а) положительно определенной, если  $\Phi(z) > 0$  для любого вектора  $z \in \mathbb{R}^n$ , имеющего ненулевую норму;

б) отрицательно определенной, если  $\Phi(z) < 0$  для любого вектора  $z \in \mathbb{R}^n$ , имеющего ненулевую норму.

Из этих определений следует, что квадратичная форма  $\Phi(z)$  положительно определена в том и только том случае, когда квадратичная форма  $-\Phi(z)$  отрицательно определена.

**Теорема 2.2.1 (критерий Сильвестра).** *Для того чтобы матрица  $W_{zz^\top}(z_*)$  была положительно определена, необходимо и достаточно, чтобы все ее главные миноры были положительны, т.е.*

$$\delta_1 > 0, \quad \delta_2 > 0, \quad \dots, \quad \delta_n > 0. \quad (2.2.1)$$

*Для того чтобы матрица  $W_{zz^\top}(z_*)$  была отрицательно определена, необходимо и достаточно, чтобы все ее главные миноры удовлетворяли условиям*

$$\delta_1 < 0, \quad \delta_2 > 0, \quad \dots, \quad (-1)^n \delta_n > 0. \quad (2.2.2)$$

Доказательство критерия можно найти в многочисленных книгах по линейной алгебре. В курсах математического анализа (см., например, [89]) рассматриваются только те случаи, в которых выполнены либо условия (2.2.1), либо (2.2.2). Все другие случаи не рассматриваются, так как они не дают достаточных условий экстремума функций.

У любознательного читателя может возникнуть вопрос: неужели лишены смысла условия, подобные (2.2.1) и (2.2.2), но отличающиеся чередованием знаков неравенств, ведь основные закономерности природы имеют, как правило, замкнутый характер? Оказывается, что возможны многие другие правила чередования знаков главных миноров, соответствующие достаточным условиям последовательных минимаксов и максиминов в многошаговых процессах оптимизации.

Пусть непрерывная функция  $W(x, u)$  определена для всех  $x \in \mathbb{R}^m$  и  $u \in \mathbb{R}^r$ . Ставятся задачи нахождения безусловного максимина и минимакса

$$V_1 = \max_{x \in \mathbb{R}^m} \min_{u \in \mathbb{R}^r} W(x, u). \quad (2.2.3)$$

$$V_2 = \min_{u \in \mathbb{R}^r} \max_{x \in \mathbb{R}^m} W(x, u), \quad (2.2.4)$$

На примере второй задачи разъясним смысл, вкладываемый в приведенные формулы. В (2.2.4) при каждом фиксированном  $u \in \mathbb{R}^r$  определим множество решений задачи максимизации  $W(x, u)$  по  $x$ :

$$x(u) = \text{Arg} \max_{x \in \mathbb{R}^m} W(x, u). \quad (2.2.5)$$

Построим функцию максимума

$$\Omega(u) = W(x(u), u). \quad (2.2.6)$$

Здесь выражение  $W(x(u), u)$  обозначает, что вычисляются значения  $W(\bar{x}, u)$ , где  $\bar{x}$  — произвольные векторы из множества  $x(u)$ . Из (2.2.5) следует:  $W(x, u)$  принимает одно и то же значение при любых  $\bar{x} \in x(u)$ , что оправдывает такую форму записи. Определим далее множество решений задачи минимизации  $\Omega(u)$  по  $u$ :

$$\Omega_* = \text{Arg} \min_{u \in \mathbb{R}^r} \Omega(u).$$

Множество  $\Omega_*$  и точечно-множественное отображение  $x(u)$  назовем *синтезом задачи* (2.2.4); хотя бы одну пару точек  $[x_*, u_*]$  такую, что  $x_* \in x(u_*)$ ,  $u_* \in \Omega_*$ , назовем *решением (глобальным) минимаксной задачи* (2.2.4). Величину  $V_2 = W(x_*, u_*)$  назовем *минимаксной гарантированной оценкой*.

Нахождение для каждого  $u$  хотя бы одного элемента из множества  $x(u)$  будем называть *внутренней задачей*. Определение хотя бы одной точки из  $\Omega_*$  назовем *внешней задачей*.

Чтобы упростить изложение, в этом параграфе будем считать, что все встречающиеся операции поиска экстремумов имеют единственные решения.

**Теорема 2.2.2.** *Если задачи (2.2.4) и (2.2.3) имеют решения, то*

$$V_1 \leq V_2.$$

**Доказательство.** Для любого фиксированного  $x \in \mathbb{R}^m$  выполнено неравенство

$$\min_{u \in \mathbb{R}^r} W(x, u) \leq W(x, u).$$

Аналогично, для всякого фиксированного  $u \in \mathbb{R}^r$  имеем

$$W(x, u) \leq \max_{x \in \mathbb{R}^m} W(x, u),$$

поэтому для любых  $\tilde{x} \in \mathbb{R}^m$  и  $\tilde{u} \in \mathbb{R}^r$  выполнены неравенства

$$\min_{u \in \mathbb{R}^r} W(\tilde{x}, u) \leq W(\tilde{x}, \tilde{u}) \leq \max_{x \in \mathbb{R}^m} W(x, \tilde{u}).$$

В силу произвольности  $\tilde{u}$  и  $\tilde{x}$  окончательно получим

$$V_1 = \max_{x \in \mathbb{R}^m} \min_{u \in \mathbb{R}^r} W(x, u) \leq \min_{u \in \mathbb{R}^r} \max_{x \in \mathbb{R}^m} W(x, u) = V_2.$$

□

Если внутреннюю и внешнюю задачи трактовать как нахождение локальных экстремумов, то придем к понятию локальных решений минимаксных и максиминных задач. Такие решения получаются, если численно решать задачи (2.2.4) или (2.2.3), используя лишь методы локальной безусловной минимизации и максимизации. Пусть точки  $x_*$  и  $u_*$  обладают соответственно окрестностями  $\Gamma(x_*)$  и  $\Gamma(u_*)$  такими, что для любых векторов  $u \in \Gamma(u_*)$  условие

$$x(u) = \text{Arg} \max_{x \in \Gamma(x_*)} W(x, u)$$

определяет однозначную вектор-функцию  $x(u)$ , причем  $x(u_*) = x_*$ , точка  $u_*$  является точкой строгого минимума функции  $\Omega(u) = W(x(u), u)$  на  $\Gamma(u_*)$ . Тогда  $z_*^\top = [x_*^\top, u_*^\top]$  называется *точкой строгого локального минимакса* в задаче (2.2.4).

Если  $z_*$  — точка локального минимакса, то для любых  $x \in \Gamma(x_*)$ ,  $u \in \Gamma(u_*)$  выполнены следующие строгие неравенства:

$$W(x(u_*), u_*) \leq W(x(u), u) = \Omega(u) \geq W(x, u). \quad (2.2.7)$$

Аналогичным образом можно определить локальные или глобальные по  $u$ , локальные или глобальные по  $x$ , строгие или нестрогие по  $u$ , строгие или нестрогие по  $x$  решения задач (2.2.3) и (2.2.4). Приведенные определения локальных решений являются обобщением известных понятий о локальных экстремумах. Минимаксная задача (2.2.4) может иметь несколько локальных решений. Однако после их нахождения нельзя утверждать, что среди этих решений есть глобальное: необходимо еще рассмотреть все глобальные по  $x$ , локальные по  $u$  решения и выбрать из них то  $u$ , для которого значение  $W(x(u), u)$  наименьшее. Изучение и отыскание локальных решений важно для функций, у которых точки локального минимакса являются одновременно и точками глобального минимакса.

Для вычисления первых и вторых производных сложной функции  $\Omega(u)$  воспользуемся результатами § 1.1 и § 1.2. Будем считать, что функция  $W$  принадлежит классу дважды непрерывно дифференцируемых функций  $C^2$ . Согласно теореме Ферма необходимое условие максимума  $W(x, u)$  по  $x$  состоит в выполнении условия

$$W_x(x, u) = 0_m. \quad (2.2.8)$$

Это выражение можно рассматривать как аналог условия связи  $G(x, u) = 0_m$ , введенного в предыдущей главе. Вектор  $u$  возьмем в качестве независимой переменной, тогда  $x$  будет вектором зависимых переменных. Если  $W_x(x_*, u_*) = 0_m$  и матрица  $W_{xx^\top}(x_*, u_*)$  невырождена, то согласно теореме 1.1.1 существует окрестность  $\Gamma(u_*)$  точки  $u_*$  такая, что для всех  $u \in \Gamma(u_*)$  выполняются тождественно условия

$$\begin{aligned} W_x(x(u), u) &\equiv 0_m, \\ \frac{d}{du} W_{x^\top}(x(u), u) &= W_{ux^\top}(x(u), u) + x_u^\top(u) W_{xx^\top}(x(u), u) \equiv 0_{rm}. \end{aligned} \quad (2.2.9)$$

Эти формулы аналогичны (1.1.2) и (1.1.3) с очевидной заменой: вместо вектора  $G$  в них берется  $W_x$ . Умножим (2.2.9) справа на невырожденную матрицу  $[W_{xx^\top}(x(u), u)]^{-1}$ , получим

$$x_u^\top(u) = -W_{ux^\top}(x(u), u) [W_{xx^\top}(x(u), u)]^{-1}. \quad (2.2.10)$$

Первая производная функции максимума  $\Omega(u)$  есть

$$\nabla \Omega(u) = W_u(x(u), u) + x_u^\top(u) W_x(x(u), u).$$

С учетом первого тождества (2.2.9) получим

$$\nabla \Omega(u) = W_u(x(u), u). \quad (2.2.11)$$

Аналогично вычисляется матрица вторых производных функции  $\Omega(u)$ :

$$\Delta \Omega(u) = W_{uu^\top}(x(u), u) - W_{ux^\top}(x(u), u) [W_{xx^\top}(x(u), u)]^{-1} W_{xu^\top}(x(u), u). \quad (2.2.12)$$

Пусть  $z_*$  обозначает объединение векторов  $x_*$  и  $u_*$ . Из (2.2.8), (2.2.11) следует, что пара  $[x_*, u_*]$  является стационарной точкой функции  $W(z)$ , т.е.

$$W_x(x_*, u_*) = 0_m, \quad W_u(x_*, u_*) = \nabla \Omega(u_*) = 0_r. \quad (2.2.13)$$

Согласно (2.2.12) имеем

$$\Delta\Omega(u_*) = W_{uu^\top}(z_*) - W_{ux^\top}(z_*)[W_{xx^\top}(z_*)]^{-1}W_{xu^\top}(z_*). \quad (2.2.14)$$

Для того чтобы в задаче (2.2.4) стационарная точка  $z_*$  была точкой строгого локального минимакса, необходимо и достаточно, чтобы матрица  $W_{xx^\top}(z_*)$  была отрицательно определена, а матрица  $\Delta\Omega(z_*)$  — положительно определена, т.е.

$$W_{xx^\top}(z_*) < 0_{mm}, \quad 0_{rr} < \Delta\Omega(z_*). \quad (2.2.15)$$

Аналогичным образом формулируются условия максимина в задаче (2.2.3), в которой локальные решения определяются аналогично. Пусть для точек  $\bar{u}_*$  и  $\bar{x}_*$  можно указать соответственно окрестности  $\Gamma(\bar{u}_*)$  и  $\Gamma(\bar{x}_*)$  такие, что для любых векторов  $x \in \Gamma(\bar{x}_*)$  условие

$$u(x) = \text{Arg} \min_{u \in \Gamma(\bar{u}_*)} W(x, u)$$

определяет однозначную вектор-функцию  $u(x)$ , причем  $\bar{x}_*$  является точкой строгого максимума функции  $W(x, u(x))$  на  $\Gamma(\bar{x}_*)$  и  $u(\bar{x}_*) = \bar{u}_*$ . Тогда  $\bar{z}_* = [\bar{x}_*^\top, \bar{u}_*^\top]$  называется точкой *строгого локального максимина* в задаче (2.2.3).

Введем функцию минимума  $Q(x) = W(x, u(x))$  и определим множество решений внешней задачи в (2.2.3):  $Q_* = \text{Arg} \max_{x \in \mathbb{R}^m} Q(x)$ . Аналогом неравенств (2.2.7) в этом случае будет

$$W(x, u) \geq W(x, u(x)) \leq W(\bar{x}_*, u(\bar{x}_*))$$

Вместо соотношений (2.2.9)–(2.2.14) здесь имеют место формулы

$$\begin{aligned} W_u(x, u(x)) &\equiv 0_r, & \nabla Q(x) &= W_x(x, u(x)), \\ \frac{d}{dx} W_{u^\top}(x, u(x)) &= W_{xu^\top}(x, u(x)) + u^\top_x(x) W_{uu^\top}(x, u(x)) \equiv 0_{mr}, \\ \Delta Q(x) &= W_{xx^\top}(x, u(x)) - W_{xu^\top}(x, u(x)) [W_{uu^\top}(x, u(x))]^{-1} W_{ux^\top}(x, u(x)). \end{aligned}$$

В частности, в точках  $\bar{x}_*$  и  $\bar{z}_*$  имеем

$$\begin{aligned} \nabla Q(\bar{x}_*) &= W_x(\bar{z}_*) = 0_m, & W_u(\bar{z}_*) &= \nabla Q(\bar{x}_*) = 0_r, \\ \Delta Q(\bar{x}_*) &= W_{xx^\top}(\bar{z}_*) - W_{xu^\top}(\bar{z}_*) [W_{uu^\top}(\bar{z}_*)]^{-1} W_{ux^\top}(\bar{z}_*). \end{aligned} \quad (2.2.16)$$

В общем случае пары оптимальных точек  $[x_*, u_*]$  и  $[\bar{x}_*, \bar{u}_*]$ , удовлетворяющие соответственно (2.2.13) и (2.2.16), не совпадают.

Для того чтобы в задаче (2.2.3)  $\bar{z}_*$  была точкой строгого локального максимина, необходимо и достаточно, чтобы матрица  $W_{uu^\top}(\bar{z}_*)$  была положительно определенной, а матрица  $\Delta Q(\bar{z}_*)$  — отрицательно определенной, т.е.

$$W_{uu^\top}(\bar{z}_*) > 0_{rr}, \quad \Delta Q(\bar{z}_*) < 0_{mm}. \quad (2.2.17)$$

Введем квадратную симметрическую матрицу

$$W_{zz^\top}(z_*) = \left[ \begin{array}{c|c} W_{xx^\top}(z_*) & W_{xu^\top}(z_*) \\ \hline W_{ux^\top}(z_*) & W_{uu^\top}(z_*) \end{array} \right]. \quad (2.2.18)$$

Здесь использовано разбиение матрицы на четыре матрицы-клетки. Последовательные главные миноры матриц  $W_{zz^\top}(z_*)$  и  $\Delta\Omega(z_*)$  обозначим  $\delta_1, \delta_2, \dots, \delta_n$  и  $\beta_1, \beta_2, \dots, \beta_r$  соответственно. Аналогично обозначим главные миноры матриц  $W_{uu^\top}(\bar{z}_*)$  и  $\Delta Q(\bar{z}_*)$ :

$\bar{\delta}_1, \bar{\delta}_2, \dots, \bar{\delta}_r$  и  $\bar{\beta}_1, \bar{\beta}_2, \dots, \bar{\beta}_m$ . В частности,  $\delta_m = |W_{xx^\top}(z_*)|$ ,  $\beta_r = |\Delta\Omega(z_*)|$ ,  $\bar{\delta}_r = |W_{uu^\top}(\bar{z}_*)|$ ,  $\bar{\beta}_m = |\Delta Q(\bar{z}_*)|$ .

По аналогии с критерием Сильвестра выразим условия (2.2.15), (2.2.17) через главные миноры матриц  $W_{xx^\top}(z_*)$  и  $W_{uu^\top}(\bar{z}_*)$  соответственно и придем к следующему результату.

**Теорема 2.2.3 (обобщенный критерий Сильвестра).** *Для того чтобы стационарная точка  $z_*$  функции  $W(z)$  была точкой строгого локального минимакса в задаче (2.2.4), необходимо и достаточно, чтобы для последовательных главных миноров матрицы  $W_{zz^\top}(z_*)$  выполнялись неравенства*

$$(-1)^i \delta_i > 0 \quad \text{при } 1 \leq i \leq m, \quad (2.2.19)$$

$$(-1)^m \delta_{m+j} > 0 \quad \text{при } 1 \leq j \leq r. \quad (2.2.20)$$

Для того чтобы в задаче (2.2.3) стационарная точка  $\bar{z}_*$  функции  $W(z)$  была точкой строгого локального максимина, необходимо и достаточно, чтобы выполнялись условия

$$\bar{\delta}_j > 0 \quad \text{при } 1 \leq j \leq r, \quad (2.2.21)$$

$$(-1)^i \bar{\delta}_i > 0 \quad \text{при } 1 \leq i \leq m. \quad (2.2.22)$$

**Доказательство.** Пусть выполнены условия (2.2.15). Тогда при  $1 \leq i \leq m$  все главные миноры матрицы  $W_{xx^\top}(z_*)$  не вырождены. Вычислим определитель блочной матрицы  $W_{zz^\top}(z_*)$ . В матрице (2.2.18) умножим верхнюю блочную строку слева на матрицу  $W_{xx^\top}(z_*)$  и вычтем ее из нижней строки. Получим матрицу, имеющую тот же определитель, что и исходная матрица:

$$\delta_n = |W_{zz^\top}(z_*)| = \left| \begin{array}{c|c} W_{xx^\top}(z_*) & W_{xu^\top}(z_*) \\ \hline 0_{rm} & \Delta\Omega(z_*) \end{array} \right| = |W_{xx^\top}(z_*)| |\Delta\Omega(z_*)| = \delta_m \beta_r. \quad (2.2.23)$$

Главный  $(m+j)$ -й минор  $\delta_{m+j}$  матрицы  $W_{zz^\top}(z_*)$  имеет вид

$$\delta_{m+j} = \left| \begin{array}{c|c} W_{xx^\top}(z_*) & W_{xu_j^\top}(z_*) \\ \hline W_{u_j x^\top}(z_*) & W_{u_j u_j^\top}(z_*) \end{array} \right|. \quad (2.2.24)$$

Здесь  $1 \leq j \leq r$ , символ  $u_j$  обозначает, что дифференцирование ведется по компонентам вектора  $u^1, u^2, \dots, u^j$ ,  $W_{xu_j^\top}$  — квадратная матрица размера  $m \times j$ ,  $W_{u_j u_j^\top}$  — матрица  $j \times j$ ; на пересечении  $i$ -й строки и  $j$ -го столбца этих матриц находятся  $\frac{\partial^2 W}{\partial x^i \partial u^j}$  и  $\frac{\partial^2 W}{\partial u^i \partial u^j}$  соответственно.

Матрицу  $W_{u_j x^\top} W_{xx^\top}^{-1}$  размера  $j \times m$  умножим на верхнюю строку блочной матрицы (2.2.24) слева и вычтем ее из нижней строки; получим

$$\delta_{m+j} = |W_{xx^\top}(z_*)| \cdot |\Delta\Omega_j(u_*)| = \delta_m \gamma_j,$$

где  $|\Delta\Omega_j(u_*)| = |W_{u_j u_j^\top}(z_*) - W_{u_j x^\top}(z_*) W_{xx^\top}^{-1}(z_*) W_{xu_j^\top}(z_*)|$  есть  $j$ -й главный минор матрицы  $\Delta\Omega(u_*)$ ,  $1 \leq j \leq r$ .

По аналогии с критерием Сильвестра условия (2.2.15) отрицательной определенности матрицы  $W_{xx^\top}(z_*)$  и положительной определенности матрицы  $\Delta\Omega(z_*)$  можно записать в виде следующих неравенств:

$$(-1)^i \delta_i > 0 \quad \text{при } 1 \leq i \leq m, \quad \beta^j > 0 \quad \text{при } 1 \leq j \leq r.$$



Первая группа неравенств совпадает с (2.2.19).

Из (2.2.19) следует, что знак сомножителя  $\delta_m$  есть  $(-1)^m$  и он совпадает с сомножителем  $\delta_{m+j}$ ; это обеспечивает неравенство (2.2.20). Таким образом, мы пришли к условиям (2.2.15).

Покажем эквивалентность условий (2.2.17) и неравенств (2.2.21), (2.2.22). Если выполнены условия (2.2.17), то матрица  $W_{uu^\top}(\bar{z}_*)$  не вырождена, нижнюю строку блочной матрицы  $W_{zz^\top}(\bar{z}_*)$  можно умножить справа на матрицу  $W_{uu^\top}^{-1}(\bar{z}_*)W_{xu^\top}(\bar{z}_*)$  и вычесть из верхней строки. Получим следующие формулы для вычисления определителя:

$$\bar{\delta}_n = |W_{zz^\top}(\bar{z}_*)| = \left| \frac{\Delta Q(\bar{z}_*)}{W_{xu^\top}(\bar{z}_*)} \middle| \frac{0_{mr}}{W_{uu^\top}(\bar{z}_*)} \right| = |\Delta Q(\bar{x}_*)| |W_{uu^\top}(\bar{z}_*)| = \beta_m \delta_r.$$

Если матрица  $W_{uu}(\bar{z}_*)$  положительно определена, то не вырождены и положительно определены главные миноры  $W_{u_j u_j}(\bar{z}_*)$ , где  $1 \leq j \leq r$ , кроме того имеют место неравенства (2.2.21). Каждый  $i$ -й главный минор

$$\Delta Q_i(\bar{z}_*) = W_{x_i x_i^\top}(\bar{z}_*) - W_{x_i u}(\bar{z}_*) W_{uu}^{-1}(\bar{z}_*) W_{u x_i}(\bar{z}_*) = \bar{\delta}_i$$

удовлетворяет условию (2.2.20).  $\square$

Интересен и нагляден частный случай, когда  $x$  и  $u$  — скаляры ( $m = r = 1$ ). Рассмотрим четыре задачи:

$$\min_{x \in \mathbb{R}} \min_{u \in \mathbb{R}} W(x, u), \quad (2.2.25)$$

$$\max_{x \in \mathbb{R}} \max_{u \in \mathbb{R}} W(x, u), \quad (2.2.26)$$

$$V_1 = \max_{x \in \mathbb{R}} \min_{u \in \mathbb{R}} W(x, u). \quad (2.2.27)$$

$$V_2 = \min_{u \in \mathbb{R}} \max_{x \in \mathbb{R}} W(x, u), \quad (2.2.28)$$

Последовательными главными минорами матрицы  $W_{xu}(z_*)$  будут

$$\delta_1 = W_{xx}(z_*), \quad \delta_2 = W_{xx}(z_*)W_{uu}(z_*) - W_{ux}^2(z_*).$$

Для того чтобы в задаче (2.2.25) стационарная точка  $[x_*, u_*]$  функции  $W(z, u)$  была точкой строгого безусловного локального минимума по  $x$  и  $u$ , необходимо и достаточно, чтобы последовательные главные миноры матрицы  $W_{xu}(x_*, u_*)$  были строго положительны:

$$\delta_1 > 0, \quad \delta_2 > 0. \quad (2.2.29)$$

Для того чтобы в задаче (2.2.26) стационарная точка  $[x_*, u_*]$  была точкой безусловного строгого максимума функции  $W(x, u)$  по  $x$  и  $u$ , необходимо и достаточно выполнения условий

$$\delta_1 < 0, \quad \delta_2 > 0. \quad (2.2.30)$$

Условия строгого минимакса (2.2.19) и (2.2.20) в задаче (2.2.28) будут иметь вид

$$\delta_1 < 0, \quad \delta_2 < 0. \quad (2.2.31)$$

Оставшийся четвертый вариант условий

$$\delta_1 > 0, \quad \delta_2 < 0 \quad (2.2.32)$$

соответствует достаточным условиям строгого максимина в задаче (2.2.27).

Неравенства (2.2.29) и (2.2.30) являются соответственно частными случаями неравенств (2.2.1) и (2.2.2), использованных в критерии Сильвестра.

В учебниках по математическому анализу (см., например, [89]) наличие неравенств вида (2.2.31) и (2.2.32), не совпадающих со стандартными наборами неравенств (2.2.1) и (2.2.2), причисляют к “неопределенным случаям”. Выше было показано, что это не так и неравенства (2.2.31) и (2.2.32) дают вполне осмысленные результаты.

Приведенные достаточные условия имеют тесную связь с критерием Сильвестра. Эту связь можно развить, введя понятия, аналогичные положительной и отрицательной определенности. Например, матрицу  $W_{zz^\top}(z_*)$  и квадратичную форму

$$\Phi(z_*) = z^\top W_{zz^\top}(z_*)z = u^\top W_{uu^\top}(z_*)u + 2x^\top W_{xu^\top}(z_*)u + x^\top W_{xx^\top}(z_*)x$$

можно назвать *минимаксно определенными*, если для любых  $\|x\| \neq 0$ ,  $\|u\| \neq 0$  имеют место неравенства

$$x^\top W_{xx^\top}(z_*)x < 0 < \max_{x \in \mathbb{R}^m} [u^\top W_{uu^\top}(z_*)u + 2u^\top W_{ux^\top}(z_*)x + x^\top W_{xx^\top}(z_*)x].$$

Максимум правой части неравенства по  $x$  достигается при  $x = -W_{xx^\top}^{-1}(z_*)W_{xu^\top}(z_*)u$ . Подставляя эту формулу в выражение, стоящее в квадратных скобках, получим, что оно равно  $u^\top \Delta\Omega(z_*)u$ , и приходим к условиям

$$x^\top W_{xx^\top}(z_*)x < 0 < u^\top \Delta\Omega(z_*)u,$$

которые выполнены при любых ненулевых  $x$  и  $u$ . Таким образом, мы пришли к (2.2.15). Другими словами, матрица  $W_{xx^\top}(z_*)$  отрицательно определена, матрица  $\Delta\Omega(z_*)$  положительно определена, а матрица  $W_{zz^\top}(z_*)$  и квадратичная форма  $\Phi(z_*)$  минимаксно определены. Для таких матриц выполнены условия (2.2.19) и (2.2.20).

Аналогично матрицу  $W_{zz^\top}(z_*)$  и квадратичную форму  $\Phi(z_*)$  назовем *строго максиминно определенными*, если при любых ненулевых векторах  $x$  и  $u$  выполнены условия

$$x^\top \Delta Q(z_*)x < 0 < u^\top W_{uu^\top}(z_*)u.$$

Здесь мы пришли к (2.2.17).

Доказанный выше критерий дает необходимые и достаточные условия минимаксной и максиминной определенности матриц и квадратичных форм.

Функцию  $W(x, u)$  назовем минимаксной, если она при каждом фиксированном  $u$  вогнута по  $x$  и функция максимума  $W(x(u), u)$  выпукла по  $u$ . Для таких функций выполнены неравенства

$$\begin{aligned} W(x_2, u) - W(x_1, u) &\leq \langle W_x(x_1, u), x_2 - x_1 \rangle, \\ \Omega(u) = W(x(u_2), u_2) &\geq W(x(u_1), u_1) + \langle W_u(x(u_1), u_1), u_2 - u_1 \rangle. \end{aligned} \quad (2.2.33)$$

Функцию  $W(x, u)$  будем называть максиминной, если она при каждом фиксированном  $x$  выпукла по  $u$  и функция минимума  $W(x, u(x))$  вогнута по  $x$ . Для таких функций имеют место неравенства

$$\begin{aligned} W(x, u_2) - W(x, u_1) &\geq \langle W_u(x, u_1), u_2 - u_1 \rangle, \\ Q(x) = W(x, u(x_2)) &\leq W(x_1, u(x_1)) + \langle W_x(x_1, u(x_1)), x_2 - x_1 \rangle. \end{aligned} \quad (2.2.34)$$

Функция  $W(x, u)$  называется *выпукло-вогнутой* (выпукла по  $u$ , вогнута по  $x$ ), если она выпукла по  $u$  при всех фиксированных  $x \in \mathbb{R}^m$  и вогнута по  $x$  при всех фиксированных

$u \in \mathbb{R}^r$ . Функция  $W(x, u)$  называется *строго выпукло-вогнутой*, если при любом фиксированном  $x$  она строго выпукла по  $u$  и при любом фиксированном  $u$  строго вогнута по  $x$ . Если функция  $W(x, u)$  выпукло-вогнута и дифференцируема всюду, то при любых  $x$  и  $u$  имеют место неравенства (2.2.33), (2.2.34).

Напомним, что пара  $[x_*, u_*]$  называется *седловой точкой* функции  $W(x, u)$  в задачах (2.2.4) и (2.2.3), если для любых  $x$  и  $u$  выполнены неравенства

$$W(x, u_*) \leq W(x_*, u_*) \leq W(x_*, u).$$

В теории исследования операций задачи вида (2.2.4) и (2.2.3) называются *минимаксными и максиминными гарантированными оценками*. В них важно, чтобы экстремумы функций были глобальными. Поэтому для их нахождения приходится применять методы глобальной оптимизации. В работе [57], например, для этой цели предложено использовать метод неравномерных покрытий.

Если матрица  $W_{zz^T}(z)$  в стационарной точке  $z$  строго выпукло-вогнута (выпукла по  $u$ , вогнута по  $x$ ), то она строго минимаксно и максиминно определена.

Введенный таким образом класс функций можно поставить в один ряд с известными классами выпуклых и выпукло-вогнутых функций. Действительно, если у строго выпуклых функций существует локальный минимум, то он единственный и совпадает с глобальным. Если у строго выпукло-вогнутых функций существует локальное седло, то оно единственно и совпадает с глобальным. Если у строго минимаксных функций существует локальный минимакс, то он единственный и является вместе с тем глобальным. Важность введения такого класса функций связана с тем, что для них задачи (2.2.4) и (2.2.3) можно решать, используя итеративные локальные численные методы. Введенный класс богаче класса строго выпукло-вогнутых функций, так как содержит в себе последний.

В многошаговых задачах исследования операций ищутся оценки гарантированных результатов вида (2.2.28) и (2.2.27), решаются и более сложные задачи отыскания последовательных минимаксов и максиминов, например, задачи вида

$$\min_u \max_x \min_y \max_w \min_p W(u, x, y, w, p).$$

В таких задачах при анализе условий экстремумов встречаются самые разнообразные чередования знаков главных миноров матрицы вторых производных функции  $W(u, x, y, w, p)$ , обеспечивающих строгие локальные решения задач последовательного минимакса.

Рассмотрим четыре простейших примера, иллюстрирующих полученные свойства минимаксных задач. Будем считать, что  $x$  и  $u$  — скаляры.

**Пример 2.2.1.** Пусть  $W(x, u) = u^2 - x^2$ . В этом случае седловую точку образуют  $x_* = u_* = 0$ . Функция  $W$  строго выпукло-вогнута, матрица  $W_{zz^T}(z_*)$  минимаксно и максиминно определена. ■

**Пример 2.2.2.** Будем искать минимакс и максимин функции

$$W(x, u) = \sin 2\pi(x - u).$$

В силу периодичности этой функции по  $x$  и  $u$  ограничимся интервалами  $0 \leq x \leq 1, 0 \leq u \leq 1$ . Легко видеть, что здесь минимакс  $V_2 = 1$ , максимин  $V_1 = -1$ , внутренние задачи имеют очевидные решения  $x(u) = u + 1/4, u(x) = x + 1/4, \Omega(u) \equiv 1, Q(x) \equiv -1$ . Седловой точки у функции  $W(x, u)$  нет. ■

**Пример 2.2.3.** Будем искать минимакс и максимин функции

$$W(x, u) = e^{u^2} \sin 2\pi(x - u).$$

В силу периодичности функции  $W$  по  $x$  можно ограничиться интервалом  $0 \leq x \leq 1$ . Функция  $W(x, u)$  не имеет седла, для нее  $V_2 = 1$ ,  $V_1 = -1$ . Решением внутренней задачи максимизации по  $x$  будет функция  $x(u) = u + 1/4$ . Точку строгого локального и одновременно глобального минимакса образует пара  $x_* = 1/4$ ,  $u_* = 0$ . В ней выполнены достаточные условия (2.2.15), так как в ней  $W_{xx} = -4\pi^2 < 0$  и  $\Delta\Omega = 2 > 0$ . ■

**Пример 2.2.4.** В качестве  $W(x, u)$  возьмем квадратичную форму

$$-u^2 + 2kux - x^2, \quad u \in \mathbb{R}^1, \quad x \in \mathbb{R}^1, \quad k^2 > 1.$$

В результате максимизации  $W$  по  $x$  получим  $x = ku$ . В задаче отыскания минимакса (2.2.4) глобальным решением является  $u_* = x_* = 0$ ,  $V_2 = 0$ . Максимум в данной задаче не достигается на неограниченном множестве. ■

### 2.3. Нахождение локального минимакса и максимина

Предположим, что в задаче (2.2.3) функция  $W(x, u)$  определена и дифференцируема. Будем отыскивать предельные (при  $t \rightarrow \infty$ ) точки решения следующей задачи Коши:

$$\frac{dx}{dt} = W_x(x, u), \quad \frac{du}{dt} = -\varepsilon W_u(x, u). \quad (2.3.1)$$

Дискретный вариант этой системы имеет вид

$$x_{k+1} = x_k + \rho W_x(x_k, u_k), \quad u_{k+1} = u_k - \rho\varepsilon W_u(x_k, u_k). \quad (2.3.2)$$

Здесь  $0 < \varepsilon \ll 1$  — малый параметр, шаг интегрирования  $\rho > 0$ .

При  $\varepsilon = 1$  метод (2.3.1) переходит в хорошо известный метод отыскания седловой точки функции  $W(x, u)$ ; метод сходится к седловой точке, если функция  $W$  строго выпукловогнута. Ниже будет показано, что если функция  $W(x, u)$  минимаксно определенная, то метод (2.3.1) сходится к точке  $z_*$ , где достигается минимакс при  $0 < \varepsilon \ll 1$ , и к точке, где достигается максимум при  $\varepsilon \gg 1$ .

**Теорема 2.3.1.** Пусть функция  $W(x, u)$  дважды непрерывно дифференцируема в окрестности стационарной точки  $z_* = [x_*, u_*]$ , где выполнены достаточные условия локального минимакса, приведенные в теореме 2.2.3. Тогда существуют такие числа  $\bar{\varepsilon} > 0$ ,  $\bar{\rho} > 0$ , что при любых фиксированных значениях  $0 < \varepsilon < \bar{\varepsilon}$  и  $0 < \rho < \bar{\rho}$  решения  $x(\varepsilon, x_0, u_0, t)$ ,  $u(\varepsilon, x_0, u_0, t)$  системы (2.3.1) и итерации  $x(\varepsilon, x_0, u_0)$ ,  $u(\varepsilon, x_0, u_0)$  по схеме (2.3.2) локально сходятся к точке  $z_*$ .

Если  $0 < \varepsilon \ll 1$ , то переменные  $u$  называются *медленно меняющимися*, переменные  $x$  — *быстро меняющимися*.

Если в (2.3.1) сделать замену независимой переменной  $\tau = \varepsilon t$ , то получим

$$\frac{du}{d\tau} = -W_u(x, u), \quad (2.3.3)$$

$$\varepsilon \frac{dx}{d\tau} = W_x(x, u). \quad (2.3.4)$$

Системы такого типа изучаются в теории сингулярных возмущений обыкновенных дифференциальных уравнений. Положив  $\varepsilon = 0$ , получим так называемую *вырожденную систему*

$$\frac{du}{d\tau} = -W_u(x, u), \quad (2.3.5)$$

$$W_x(x, u) = 0_m. \quad (2.3.6)$$

Если выполнены условия теоремы 2.2.3, то из условия (2.3.6) определяется однозначная функция  $x(u)$  и строится функция  $\Omega(u)$ . Эти функции были введены выше формулами (2.2.5) и (2.2.6). Поэтому дифференциальное уравнение (2.3.5) можно записать в виде

$$\frac{du}{d\tau} = -\frac{d\Omega}{d\tau} = -W_u(x(u), u),$$

т.е. при  $\varepsilon = 0$  система (2.3.3), (2.3.4) переходит в простейший градиентный метод Коши безусловной минимизации функции  $\Omega(u)$ . Для локальной сходимости метода достаточно, чтобы матрица  $\Delta\Omega(u_*)$ , задаваемая (2.2.14), была положительно определена [36, 50].

Реализация метода (2.3.5), (2.3.6) затруднительна в связи с тем, что при всяком  $u$  из задачи максимизации следует определять значение функции  $x(u)$ . Метод (2.3.3), (2.3.4) намного проще, он лишен этого недостатка.

Перейдем к **доказательству** теоремы 2.3.1. Составим уравнение в вариациях для системы (2.3.1)

$$\delta\dot{z} = M(\varepsilon, z_*)\delta z;$$

здесь

$$M(\varepsilon, z_*) = \left[ \begin{array}{c|c} -\varepsilon W_{uu^\top}(z_*) & -\varepsilon W_{ux^\top}(z_*) \\ \hline W_{xu^\top}(z_*) & W_{xx^\top}(z_*) \end{array} \right],$$

где  $\delta z^\top = [\delta x^\top, \delta u^\top]$ ,  $\delta x = x(t) - x_*$ ,  $\delta u = u(t) - u_*$ ,  $z_*^\top = [x_*^\top, u_*^\top]$ . Непрерывная ограниченная (как функция  $\varepsilon$ ) матрица  $M(\varepsilon, z_*)$  удовлетворяет следующему условию: корни характеристического уравнения

$$|M(\varepsilon, z_*) - \lambda I_{m+r}| = 0 \quad (2.3.7)$$

при  $\varepsilon \rightarrow 0$  распадаются на две группы корней:

1-я группа из  $m$  корней — корни, близкие к корням уравнения

$$|W_{xx^\top}(z_*) - \beta I_m| = 0;$$

2-я группа из  $n$  “малых” корней порядка  $\varepsilon\mu$ , где  $\mu$  — корни, близкие к корням уравнения

$$\left| \begin{array}{c|c} W_{uu^\top}(z_*) + \mu I_r & W_{ux^\top}(z_*) \\ \hline W_{xu^\top}(z_*) & W_{xx^\top}(z_*) \end{array} \right| = 0.$$

Умножая нижнюю матрицу-строку слева на  $W_{ux^\top}(z_*)W_{xx^\top}^{-1}(z_*)$  и вычитая ее из верхней, получим, что  $\mu$  есть корень следующего уравнения:

$$|\Delta\Omega(u_*) + \mu I_r| = 0,$$

где матрица  $\Delta\Omega(u_*)$  находится из (2.2.14).

Согласно достаточным условиям локального минимакса матрица  $\Delta\Omega(u_*)$  положительно определена, поэтому все корни  $\mu$  действительные и строго отрицательные. Следовательно, при достаточно малых значениях  $\varepsilon$  все корни уравнения (2.3.7) имеют строго отрицательные действительные части. Согласно теореме А.М. Ляпунова об устойчивости по первому приближению стационарная точка  $z_*$  является асимптотически устойчивым положением равновесия для системы (2.3.1). Отсюда следует локальная экспоненциальная сходимость метода (2.3.1) к  $z_*$  и сходимость разностного варианта (2.3.2) при достаточно малых значениях  $\rho$ .  $\square$

Методы (2.3.1) и (2.3.2) можно использовать для нахождения локальных максимума в задаче (2.2.4). Однако в этом случае следует брать  $\varepsilon \gg 1$ , т.е.  $\varepsilon$  рассматривается как большой параметр. При выполнении достаточных условий локального максимума, приведенных

в теореме 2.2.3, решения (2.3.1) и (2.3.2) локально сходятся к локальному максимуму, если  $\varepsilon$  достаточно велико и  $\rho$  достаточно мало.

Введение малого параметра  $\varepsilon$  в (2.3.1) приводит к тому, что переменные  $u$  изменяются значительно медленнее, чем переменные  $x$ . Это усложняет интегрирование (2.3.1), но вместо с тем отпадает необходимость нахождения функции  $x(u)$ .

Численный метод (2.3.1) оказался весьма эффективным и успешно использовался для решения игровых задач. Ниже в § 3.8 приведены результаты решения этим методом задачи Р. Айзекса о долинхобрахистохроне.

## 2.4. Задачи нелинейного программирования с ограничениями типа равенства

Пусть вектор  $z \in \mathbb{R}^n$ . Заданы функции  $W(z) : \mathbb{R}^n \rightarrow \mathbb{R}^1$ ,  $G(z) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , где  $m \leq n$ . Ставится задача НЛП о минимизации целевой функции  $W(z)$  на допустимом множестве  $Z$ :

$$\min_{z \in Z} W(z), \quad Z = \{z \in \mathbb{R}^n : G(z) = 0_m\}. \quad (2.4.1)$$

Множество решений этой задачи обозначим через  $Z_*$ .

Ниже будет использоваться матрица Якоби  $G_{z^\top}$  размера  $m \times n$  и матрица транспонированная к ней  $G_z^\top$  размера  $n \times m$ .

В матрице  $G_{z^\top}$  символ  $G$  – есть вектор-столбец,  $z^\top$  – вектор-строка. В матрице  $G_z^\top$  наоборот  $G^\top$  – вектор-строка,  $z$  – вектор-столбец. Поэтому  $(i, j)$ -элементы матриц  $G_{z^\top}$  и  $G_z^\top$  будут соответственно  $\frac{\partial G^i}{\partial z^j}$ , где  $1 \leq i \leq m$ ,  $1 \leq j \leq n$  и  $\frac{\partial G^j}{\partial z^i}$ , где  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ .

Точка  $z_*$  называется *глобальным решением задачи* (2.4.1), если для всех  $z \in Z$  имеет место неравенство

$$W(z_*) \leq W(z).$$

Пусть  $\Gamma(z_*)$  – некоторая окрестность  $z_*$ . Точка  $z_*$  называется *точкой локального минимума* в задаче (2.4.1), если

$$W(z_*) \leq W(z) \quad (2.4.2)$$

для всех  $z \in \Gamma(z_*) \cap Z$ . Если в (2.4.2) имеет место строгое неравенство для всех  $z \in \Gamma(z_*) \cap Z$ ,  $z \neq z_*$ , то будем говорить, что  $z_*$  – *точка строгого (изолированного) локального минимума* в задаче (2.4.1).

Задача (2.4.1) уже упоминалась выше (см. (1.1.16)), для нее определялись множители и функция Лагранжа:

$$L(z, p) = W(z) + G^\top(z)p, \quad p \in \mathbb{R}^m. \quad (2.4.3)$$

Если градиенты  $G_z^i(z_*)$ ,  $1 \leq i \leq m$ , линейно независимы, то будем говорить, что точка  $z_*$  – *регулярная точка*, или “в точке  $z_*$  выполнено условие регулярности ограничений (у.р.о.)”. Эквивалентное условие, обеспечивающее выполнение у.р.о. в точке  $z_*$ , можно сформулировать так: ранг матрицы  $G_z^\top(z_*)$  максимальный и равен  $m$ , или, в символьной записи,  $\text{rang}[G_z^\top(z_*)] = m$ .

Аналогично (2.4.1) ставится задача поиска максимума функции  $W(z)$  на множестве  $Z$ :

$$\max_{z \in Z} W(z), \quad Z = \{z \in \mathbb{R}^n : G(z) = 0_m\}. \quad (2.4.4)$$

Необходимые условия экстремума для обеих задач формулируются одинаково. Имеет место

**Теорема 2.4.1 (Лагранжа).** Пусть в задачах (2.4.1) и (2.4.4) функции  $W$  и  $G$  непрерывно дифференцируемы; локальный минимум или максимум функции  $W(z)$  на множестве  $Z$

достигается в допустимой регулярной точке  $z_*$ . Тогда существует единственный вектор множителей Лагранжа  $p_* \in \mathbb{R}^m$ , удовлетворяющий условию

$$L_z(z_*, p_*) = W_z(z_*) + G_z^\top(z_*)p_* = 0_n \quad (2.4.5)$$

и определяемый по формуле

$$p_* = - [G_{z^\top}(z_*)G_z^\top(z_*)]^{-1} G_{z^\top}(z_*)W_z(z_*) = - (G_z(z_*))^+ W_z(z_*). \quad (2.4.6)$$

Здесь  $(G_z(z_*))^+$  – матрица  $m \times n$  псевдообратная (левая обратная) для матрицы  $G_z^\top(z_*)$ . (См. Приложение 3, формула (П.3.1)).

**Доказательство.** Вектор  $z$  представим как совокупность двух векторов  $x \in \mathbb{R}^m$  и  $u \in \mathbb{R}^r$ , т.е. имеют место представления (1.1.15),  $z_*^\top = [x_*^\top, u_*^\top]$ . Вектор  $u$  возьмем в качестве вектора независимых переменных. Из условия связи  $G(z) = 0_m$  определим зависимую переменную  $x$  как функцию от  $u$ . Подставив эту функцию в  $W(z)$ , получим сложную функцию  $\Omega(u) = W(x(u), u)$ , для которой

$$\min_{u \in \Gamma(u_*)} \Omega(u) = \min_{z \in Z} W(z) = \min_{u \in \Gamma(u_*)} W(x(u), u).$$

Из условия регулярности следует, что матрица Якоби  $G_{z^\top}(z_*)$  имеет ранг  $m$ , поэтому в ней можно выделить квадратную матрицу  $m \times m$ , определитель которой (минор) не равен нулю.

Будем предполагать, что выделенная матрица  $G_{x^\top}(z_*)$  состоит из первых  $m$  столбцов матрицы  $G_{z^\top}(z_*)$ . Определитель матрицы  $G_{x^\top}(z_*)$  отличен от нуля. Поэтому в силу теоремы 1.1.2 существует окрестность  $\Gamma(u_*)$  точки  $u_*$  такая, что для всех  $u \in \Gamma(u_*)$  существуют функции  $x(u)$ ,  $p(u)$ , удовлетворяющие следующим условиям:

$$\begin{aligned} x(u_*) &= x_*, & p(u_*) &= p_*, & G(x(u), u) &= 0_m, \\ L_x(x(u), u, p(u)) &= W_x(x(u), u) + G_x^\top(x(u), u)p(u) = 0_m, \\ \nabla \Omega(u) = L_u(x(u), u, p(u)) &= W_u(x(u), u) + G_u^\top(x(u), u)p(u). \end{aligned}$$

Так как  $u_*$  – точка безусловного экстремума функции  $\Omega(u)$ , то согласно теореме Ферма необходимо, чтобы

$$\nabla \Omega(u_*) = 0_r.$$

Отсюда следует

$$\begin{aligned} L_x(x_*, u_*, p_*) &= W_x(x_*, u_*) + G_x^\top(x_*, u_*)p_* = 0_m, \\ L_u(x_*, u_*, p_*) &= W_u(x_*, u_*) + G_u^\top(x_*, u_*)p_* = 0_r. \end{aligned} \quad (2.4.7)$$

Перейдем от пары  $[x_*, u_*]$  к вектору  $z_*$ ; объединив формулы (2.4.7), получим (2.4.5). Умножив (2.4.5) слева на матрицу  $G_{z^\top}(z_*)$ , получим систему

$$G_{z^\top}(z_*)W_z(z_*) + M(z_*)p_* = 0_m, \quad (2.4.8)$$

где введена квадратная порядка  $m$  матрица Грамма  $M(z_*) = G_{z^\top}(z_*)G_z^\top(z_*)$ . Из у.р.о. следует, что матрица  $M(z_*)$  не вырождена и единственное решение системы (2.4.8) определяется по формуле (2.4.6).  $\square$

В стационарной точке функции Лагранжа выполнены условия

$$L_z(z_*, p_*) = 0_n, \quad L_p(z_*, p_*) = G(z_*) = 0_m \quad (2.4.9)$$

и значение функции Лагранжа совпадает с оптимальным значением целевой функции, вычисленным в точке  $z_*$ , т.е.  $L(z_*, p_*) = W(z_*)$ .

Из того, что  $[z_*, p_*]$  — стационарная точка функции Лагранжа, совсем не следует, что  $z_*$  является точкой локального минимума или максимума функции  $W(z)$  на  $Z$ . Теорема Лагранжа, таким образом, дает только *необходимое* условие экстремума. Заметим, что задача (2.4.1) имеет, вообще говоря, неединственное решение. Однако при выполнении условий теоремы 2.4.1 каждой регулярной допустимой точке локального минимума или максимума  $z_*$  соответствует единственный вектор  $p_*$ .

Соотношения (2.4.9) задают систему из  $n + m$  уравнений с  $n + m$  неизвестными. В случае задач выпуклого программирования (когда функция  $W(z)$  выпуклая, а  $G(z)$  — линейная) эта система определяет не только необходимые условия экстремума, но и дает достаточные условия, т.е. если пара  $[z_*, p_*]$  удовлетворяет (2.4.9), то обязательно  $z_* \in Z_*$ . При этом знание вектора  $p_*$  не является излишним, так как он дает важную информацию о свойствах оптимального решения задачи (2.4.1) (см. следующий параграф).

Выражение  $(G_z^\top(z))^\parallel$  обозначает квадратную  $n \times n$  матрицу проектирования на пространство столбцов матрицы  $G_z^\top(z)$ . Аналогично  $(G_z^\top(z))^\perp$  — матрица проектирования на ортогональное дополнение к пространству столбцов, т.е. на нуль-пространство (ядро) транспонированной матрицы  $G_{z^\top}(z)$ . Согласно Приложению 3 имеем

$$(G_z^\top(z))^\parallel = G_z^\top(z)[G_{z^\top}(z)G_z^\top(z)]^{-1}G_{z^\top}(z), \quad (G_z^\top(z))^\perp = I_n - (G_z^\top(z))^\parallel.$$

Согласно формуле (П.3.7) из Приложения 3 проектирование вектора  $\Theta$  из  $\mathbb{R}^n$  на пространство столбцов матрицы  $G_{z^\top}$  записывается следующим образом:

$$\text{pr}(\Theta, \text{im } G_z^\top) = (G_z^\top)^\parallel \Theta = G_z^\top (G_z^\top)^\perp \Theta.$$

Подставив (2.4.6) в (2.4.5), получим

$$L_z(z_*, p_*) = (G_z^\top(z_*))^\perp W_z(z_*) = 0_n. \quad (2.4.10)$$

Геометрическая интерпретация теоремы Лагранжа состоит в том, что согласно (2.4.10) в точке  $z_*$  локального условного минимума или максимума функции  $W(z)$  вектор  $W_z(z_*)$  принадлежит множеству линейных комбинаций столбцов матрицы  $G_z^\top(z_*)$ , т.е. можно записать

$$W_z(z_*) \in \text{im } G_z^\top(z_*), \quad -W_z(z_*) = \sum_{i=1}^m G_z^i(z_*) p_*^i.$$

Последнее представление можно рассматривать как разложение вектора  $-W_z(z_*)$  по линейно независимым столбцам матрицы  $G_z^\top(z_*)$ . Компоненты вектора множителей Лагранжа  $p_*$  являются коэффициентами этого разложения.

Если в стационарной точке  $z_*$  ранг матрицы  $G_z(z_*)$  меньше  $m$ , то в ней не выполнено у.р.о. и утверждение теоремы Лагранжа может оказаться неверным. Возможны два случая: либо множители Лагранжа неединственны, либо они не существуют. Второй случай иллюстрирует приводимый ниже пример 2.4.1. Помимо приведенного условия регулярности ограничений есть и другие условия, обеспечивающие существование множителей Лагранжа (условие Карлина–Слейтера и т.д.).

**Пример 2.4.1.** Пусть  $z \in \mathbb{R}$ ,  $W(z) = z$ ,  $G(z) = z^2$ ,  $L(z, p) = z + pz^2$ ,  $L_z(z, p) = 1 + 2pz$ ,  $z_* = 0$ ,  $G_z(z_*) = 0$ ,  $W_z = 1$ ; при любом  $p$  имеем  $L_z(z_*, p) = 1 \neq 0$ . Здесь ранг матрицы  $G_z(z_*)$  равен нулю, поэтому условие (2.4.5) не выполнено ни при каких  $p$ . ■

**Пример 2.4.2.** Пусть  $z \in \mathbb{R}^2$ , вектор  $z$  состоит из двух скалярных компонент  $x$  и  $y$ ,  $f(z) = x$ ,  $m = 1$ ,  $G(z) = (x)^3 - (y)^2$ . Из условия  $G(z) = 0$  следует, что  $(x)^3 = (y)^2 \geq 0$ ,



допустимое множество  $Z$  состоит из пар точек  $z = [x, y]$ , у которых  $x \geq 0$ ,  $y$  — произвольно. Поэтому целевая функция принимает минимальное значение, равное нулю, при  $x = y = 0$ . В этой точке  $G_{z^\top} = [0 \mid 0]$ ,  $\text{rank}[G_{z^\top}] = 0 < 1$ ,  $W_{z^\top} = [1 \mid 0]$ . Отсюда видно, что не существует такого скаляра  $p$ , для которого  $W_{z^\top} + pG_{z^\top} = [0 \mid 0]$ . Свойство (2.4.5) в этом примере также не имеет места. ■

В функции Лагранжа, задаваемой формулой (2.4.3), вектор множителей  $p$  традиционно рассматривается как некоторый параметр, далее вводится максиминная задача

$$V = \max_{p \in \mathbb{R}^m} \min_{z \in \mathbb{R}^n} [W(z) + p^\top G(z)]. \quad (2.4.11)$$

Решив при фиксированном  $p$  внутреннюю задачу минимизации, определим возможно неоднозначное точечно-множественное отображение  $z(p)$ . Подставив его в функцию Лагранжа, получим так называемую *двойственную функцию*:

$$\varphi(p) = W(z(p)) + p^\top G(z(p)) = \min_{z \in \mathbb{R}^n} [W(z) + p^\top G(z)]. \quad (2.4.12)$$

Значение функции  $\varphi(p)$  не зависит от конкретного выбора вектора  $z$  из множества  $z(p)$ .

Внешняя задача максимизации в (2.4.11) обычно называется *двойственной* к (2.4.1) и состоит в нахождении

$$V = \max_{p \in \mathbb{R}^m} \varphi(p). \quad (2.4.13)$$

В ряде случаев удается найти аналитическое решение внутренней задачи, после этого решается внешняя задача, содержащая только множители Лагранжа. Приведем простейший пример, иллюстрирующий такую возможность.

**Пример 2.4.3.** Пусть непустое допустимое множество задается условием

$$Z = \{z \in \mathbb{R}^n : Az = b\}, \quad (2.4.14)$$

где  $A \in \mathbb{R}^{m \times n}$  — прямоугольная матрица,  $b \in \mathbb{R}^m$ ,  $m \leq n$ .

Задан вектор  $\bar{z} \in \mathbb{R}^n$ , требуется найти вектор  $z_*$  — проекцию  $\bar{z}$  на множество  $Z$ . В этой задаче квадратичного программирования ищется вектор  $z_*$  такой, что

$$\|z_* - \bar{z}\|^2 = \min_{z \in Z} \|z - \bar{z}\|^2.$$

Составим функцию Лагранжа  $L(z, p) = \|z - \bar{z}\|^2/2 + p^\top (b - Az)$ . Из условия минимума этой функции по  $z$  следует, что

$$z = \bar{z} + A^\top p. \quad (2.4.15)$$

Это выражение подставим в  $L(z, p)$ ; получим вогнутую двойственную функцию

$$\varphi(p) = p^\top \bar{b} - \|A^\top p\|^2/2,$$

где  $\bar{b} = b - A\bar{z}$ .

В точке  $p_*$ , где достигается максимум  $\varphi(p)$  по  $p$ , выполнены условия

$$\varphi_p(p_*) = \bar{b} - AA^\top p_* = 0_m, \quad z_* = \bar{z} + A^\top p_* \in Z. \quad (2.4.16)$$

Отсюда следует

$$\varphi(p_*) = p_*^\top \bar{b} - \frac{\|A^\top p_*\|^2}{2} = \frac{\|A^\top p_*\|^2}{2} = \frac{\|z_* - \bar{z}\|^2}{2} = \frac{p_*^\top \bar{b}}{2} = L(z_*, p_*).$$

Считая, что ранг матрицы  $A$  максимальный, равен  $m$ , получим

$$z_* = \text{pr}(\bar{z}, Z) = \bar{z} + A^\top p_*, \quad p_* = (AA^\top)^{-1}\bar{b}. \quad (2.4.17)$$

Если взять  $\bar{z} = 0_n$ , то  $b = \bar{b}$ , и решение задачи (2.4.15) определяет вектор  $z_*^n$  из  $Z$ , обладающий наименьшей нормой (нормальное решение системы  $Az = b$ ). Согласно (2.4.17)

$$z_*^n = A^\top(AA^\top)^{-1}b = \text{pr}(0_n, Z) = (A^\top)^\parallel z, \quad z \in Z.$$

Здесь и ниже использованы обозначения проектирования на пространство строк матрицы  $A$  (столбцов матрицы  $A^\top$ ) и на ортогональное дополнение (см. Приложение 3):

$$(A^\top)^\parallel = A^\top(AA^\top)^{-1}A, \quad (A^\top)^\perp = I_n - (A^\top)^\parallel.$$

Решение задачи (2.4.15) можно записать в виде

$$z_* = z_*^n + (A^\top)^\perp \bar{z},$$

т.е. проекция вектора  $\bar{z}$  на пространство решений системы (2.4.14) представима в виде суммы нормального решения и проекции вектора  $\bar{z}$  на нуль-пространство матрицы  $A$ .

Пусть  $\lambda$  — наименьшее собственное значение матрицы  $AA^\top$ . Так как  $\text{rang } A = m$ , то  $\lambda > 0$ , из (2.4.16) и (2.4.17) получим

$$\lambda \|p_*\|^2 \leq \|z_* - \bar{z}\|^2 = p_*^\top AA^\top p_* = p_*^\top \bar{b} \leq \|\bar{b}\| \|p_*\|, \quad \|p_*\| \leq \|\bar{b}\|/\lambda. \quad (2.4.18)$$

Отсюда находим расстояние от точки  $z_*$  до множества  $Z$ :

$$\text{dist}(\bar{z}, Z) = \|z_* - \bar{z}\| = (p_*^\top \bar{b})^{1/2} \leq \|b - A\bar{z}\|/\sqrt{\lambda}. \quad (2.4.19)$$

Следовательно, для нормального решения имеет место оценка  $\|z_*^n\| \leq \|b\|/\sqrt{\lambda}$ .

Приведенные формулы записываются особенно просто, если  $m = 1$ ; в этом случае  $b \in \mathbb{R}^1$ , матрица  $A$  является  $n$ -мерным вектором-строкой  $a^\top$ . Формулы (2.4.17) определяют проекцию вектора  $\bar{z}$  на плоскость  $a^\top z = b$ , формула (2.4.19) — расстояние от  $\bar{z}$  до этой плоскости, т.е.

$$z_* = \bar{z} + a(b - a^\top \bar{z})/\|a\|^2, \quad \text{dist}(\bar{z}, Z) = |b - a^\top \bar{z}|/\|a\|.$$

Если  $\bar{z} = 0_n$ , то  $z_* = ab/\|a\|^2$ ,  $\|z_*\| = |b|/\|a\|$ . ■

## 2.5. Функция чувствительности

Введем  $m$ -мерный вектор параметров  $\xi \in \mathbb{R}^m$ . Допустимое множество, зависящее от этого параметра, представим в виде

$$U(\xi) = \{z \in \mathbb{R}^n : G(z) = \xi\}. \quad (2.5.1)$$

Считаем, что вектор  $\xi$  таков, что множество  $U(\xi)$  не пусто при всех  $\xi$  из некоторой окрестности  $U(\xi_*)$  точки  $\xi_*$ . Минимальное значение целевой функции  $W(z)$  на множестве  $U(\xi)$  будет также зависеть от вектора  $\xi$ , поэтому обозначим

$$\Omega(\xi) = \min_{z \in U(\xi)} W(z). \quad (2.5.2)$$

В теории нелинейного программирования зависимость  $\Omega(\xi)$  называется *функцией чувствительности* (sensitivity function).

Для задачи (2.5.2) функция Лагранжа и соответствующая максиминная задача примут вид

$$\begin{aligned} L(z, p, \xi) &= W(z) + p^\top (G(z) - \xi), \\ \max_{p \in \mathbb{R}^m} \min_{z \in \mathbb{R}^n} L(z, p, \xi). \end{aligned} \quad (2.5.3)$$

Пусть  $z(p, \xi)$  — решение внутренней задачи в (2.5.3), т.е.

$$z(p, \xi) = \text{Arg} \min_{z \in \mathbb{R}^n} L(z, p, \xi).$$

Даже при условии дифференцируемости функций  $W(z)$  и  $G(z)$  функция минимума  $z(p, \xi)$  дифференцируема лишь по направлению. Приведем теорему, дающую достаточные условия дифференцируемости функции  $\Omega(\xi) = W(z(p, \xi))$ .

**Теорема 2.5.1.** Пусть функции  $W(z)$  и  $G(z)$  дважды непрерывно дифференцируемы, при  $\xi = \xi_*$  решением задачи (2.5.2) является регулярная точка  $z_*$ , ей соответствует двойственный вектор  $p_*$ , квадратичная форма  $\bar{z}^\top L_{z^\top z}(z_*, p_*, \xi_*) \bar{z} > 0$  при любых  $\bar{z}$  таких, что  $G_{z^\top}(z_*) \bar{z} = 0_m$ ,  $\|\bar{z}\| \neq 0$ . Тогда функция  $\Omega(\xi)$  дифференцируема по  $\xi$  в окрестности точки  $\xi_*$  и ее производная вычисляется по формуле

$$\frac{d\Omega(\xi_*)}{d\xi} = -p_*, \quad (2.5.4)$$

где множитель Лагранжа  $p_*$  находится из (2.4.6).

**Доказательство.** Согласно теореме 2.4.1 Лагранжа  $[x_*, p_*]$  является стационарной парой функции Лагранжа. В окрестности точки  $\xi_*$  существуют непрерывно дифференцируемые функции  $z(\xi)$  и  $p(\xi)$  такие, что  $z(\xi_*) = z_*$ ,  $p(\xi_*) = p_*$ , вектор  $z(\xi)$  является локальным решением задачи (2.5.1), а  $p(\xi)$  — соответствующим множителем Лагранжа; векторы  $z$ ,  $p$ ,  $\xi$  удовлетворяют системе

$$L_z(z, p, \xi) = W_z(z) + G_z^\top(z)p = 0_n, \quad G(z) - \xi = 0_m. \quad (2.5.5)$$

Решением этой системы является точка  $(z_*, p_*, \xi_*)$ . В ней квадратная матрица Якоби размера  $m + n$  имеет вид

$$J = \left[ \begin{array}{c|c} L_{z^\top z}(z_*, p_*, \xi_*) & G_z^\top(z_*) \\ \hline G_{z^\top}(z_*) & 0_{mm} \end{array} \right].$$

Если матрица  $J$  вырождена, то существуют векторы  $\bar{z}$ ,  $\bar{p}$ , не равные нулю одновременно и такие, что

$$L_{z^\top z}(z_*, p_*, \xi_*) \bar{z} + G_z^\top(z_*) \bar{p} = 0_n, \quad (2.5.6)$$

$$G_{z^\top}(z_*) \bar{z} = 0_m. \quad (2.5.7)$$

Умножим (2.5.6) слева на  $\bar{z}^\top$ , учтем (2.5.7), получим

$$\bar{z}^\top L_{z^\top z}(z_*, p_*, \xi_*) \bar{z} = 0. \quad (2.5.8)$$

Согласно достаточному условию минимума **б** из доказанной ниже теоремы 2.8.1 матрица  $L_{z^\top z}(z_*, p_*, \xi_*)$  положительно определена на ядре матрицы  $G_{z^\top}$ , и поэтому равенство (2.5.6) может иметь место только при  $\bar{z} = 0_n$ . Следовательно, из (2.5.6) вытекает

$$G_z^\top(z_*) \bar{p} = 0_n. \quad (2.5.9)$$

Но так как матрица  $G_z^\top(z_*)$  имеет ранг  $m$ , система (2.5.9) имеет только нулевое решение  $\bar{p} = 0_m$ . Приходим к выводу, что система (2.5.7), (2.5.8) имеет только нулевые решения  $\bar{z} = 0_n, \bar{p} = 0_m$ . Отсюда заключаем, что матрица  $J$  не вырождена и по теореме о неявной функции система (2.5.5) имеет решения  $z(\xi), p(\xi)$ , удовлетворяющие в окрестности точки  $\xi_*$  условиям

$$L_z(z, p, \xi) = W_z(z(\xi)) + G_z^\top(z(\xi))p(\xi) = 0_n, \quad (2.5.10)$$

$$G(z(\xi)) = \xi. \quad (2.5.11)$$

Из (2.5.10) следует

$$\frac{d\Omega}{d\xi} = \frac{\partial z^\top}{\partial \xi} W_z(z(\xi)) = -\frac{\partial z^\top}{\partial \xi} G_z^\top(z(\xi))p(\xi).$$

Дифференцируя (2.5.11) по  $\xi$ , получаем

$$\frac{\partial z^\top}{\partial \xi} G_z^\top(z(\xi)) = I_m.$$

Объединяя эти равенства, получим

$$\frac{d\Omega(\xi)}{d\xi} = -p(\xi), \quad (2.5.12)$$

откуда, в частности, при  $\xi = \xi_*$  следует формула (2.5.4).  $\square$

Можно получить формулу (2.5.4) проще, используя формулы БАД, приведенные в первой главе. Действительно, положим в формулах (1.3.18) – (1.3.20)  $u = \xi, x_j = z$ ; тогда

$$\Phi(j, X_j, U_j) = G(z) - \xi, \quad z = z + G(z) - \xi.$$

Из (1.3.10) и (1.3.11) следует

$$W_z + G_z(z)p = 0, \quad \frac{d\Omega(\xi)}{d\xi} = -p.$$

Таким образом, приходим к формуле (2.5.12).

Из дифференцируемости  $\Omega(\xi)$  по  $\xi$  следует

$$\Omega(\xi) = \Omega(\xi_*) - p_*^\top(\xi - \xi_*) + \|\xi - \xi_*\| \alpha(\xi_*, \|\xi - \xi_*\|),$$

где скалярная функция  $\alpha$  обладает свойствами (1.1.14).

Пусть в задаче нелинейного программирования допустимое множество задается с помощью векторного ограничения типа неравенства

$$Z(\xi) = \{z \in \mathbb{R}^n : h(z) \leq \xi\}, \quad (2.5.13)$$

где введен  $c$ -мерный вектор параметров  $\xi$  и дифференцируемая  $c$ -мерная вектор-функция  $h(z)$ .

Предполагаем, что существует вектор  $\bar{\xi} \in \mathbb{R}^c$  такой, что множество  $Z(\bar{\xi})$  не пусто. Ниже будем использовать векторы  $\xi$ , удовлетворяющие условию  $\bar{\xi} \leq \xi$ , что обеспечит непустоту множества  $Z(\xi)$ .

Ставится задача минимизации функции  $W(z)$  на множестве  $Z(\xi)$ , когда требуется найти

$$\Omega(\xi) = \min_{z \in Z(\xi)} W(z). \quad (2.5.14)$$

Будем говорить, что вектор  $\xi_2$  получен увеличением вектора  $\xi_1$ , если для всех его компонент выполнены неравенства  $\xi_2^i \geq \xi_1^i$ .

Из определений (2.5.14) и (2.5.13) следует, что множество  $Z(\xi)$  расширяется, а функция  $\Omega(\xi)$  монотонно убывает при увеличении вектора  $\xi$ , т.е. для  $\bar{\xi} \leq \xi_1 \leq \xi_2$  имеем

$$Z(\bar{\xi}) \subseteq Z(\xi_1) \subseteq Z(\xi_2), \quad (2.5.15)$$

$$\Omega(\bar{\xi}) \geq \Omega(\xi_1) \geq \Omega(\xi_2). \quad (2.5.16)$$

Действительно, пусть  $\xi_1 \leq \xi_2$ ; тогда для  $z_1 \in Z(\xi_1)$  имеем  $h(z_1) \leq \xi_1 \leq \xi_2$ , поэтому  $z_1 \in Z(\xi_2)$  и имеют место включения (2.5.15), из которых следуют неравенства (2.5.16). Если функция  $\Omega(\xi)$  дифференцируема в точке  $\xi$ , то из (2.5.16) следует

$$\frac{d\Omega(\xi)}{d\xi} \leq 0_c. \quad (2.5.17)$$

Согласно формуле (2.5.12) вектор множителей Лагранжа  $p$  определяет чувствительность оптимального значения целевой функции  $\Omega$  (в экономике обычно цены изделия) к изменению вектора ограничений  $\xi$ , поэтому в работах по моделированию экономических процессов множитель Лагранжа  $p$  часто называют *теневой ценой (shadow price)* данного вида товара.

Использование функции чувствительности открывает большие возможности решения сложных оптимизационных задач, в которых допустимое множество содержит изолированные точки. Поясним эту идею на простом примере, взятом из статьи [128].

**Пример 2.5.1.** Пусть в трехмерном пространстве ищется

$$W_* = \min_{z \in \mathbb{R}^3} W(z) = \min_{z \in \mathbb{R}^3} z^{(1)} \quad (2.5.18)$$

при наличии двух ограничений

$$G^{(1)}(z) = (z^{(1)} - 5)^2 + 2(z^{(2)} - 5)^2 + (z^{(3)} - 5)^2 - 18 \leq 0, \quad (2.5.19)$$

$$G^{(2)}(z) = 100 - (z^{(1)} + 7 - 2z^{(2)})^2 - 4(2z^{(1)} + z^{(2)} - 11)^2 - 5(z^{(3)} - 5)^2 \leq 0. \quad (2.5.20)$$

Особенностью задачи (2.5.18) является тот факт, что в ней глобальный минимум достигается в изолированной допустимой точке  $z_* = (1, 4, 5)$ , где  $W(z_*) = 1$ ,  $G^{(1)}(z_*) = G^{(2)}(z_*) = 0$  (в [128] ошибочно указана точка с координатами  $(1, 3, 5)$ ).

На рис. 2.5.1 представлена проекция допустимой области задачи на плоскость  $x^{(3)} = 5$ . Она представляет собой пересечение внутренней части эллипса, определяемой ограничением (2.5.19), и внешней части другого эллипса, соответствующей ограничению (2.5.20). Это пересечение состоит из заштрихованной области и точки  $z_*$ , где эллипсы касаются (см. случай I). На рисунке указана также точка  $\bar{z}$ , в которой достигается минимум функции  $W(z)$  на заштрихованной части. С помощью локальных методов поиска минимума получено, что  $\bar{z} = (3.747692, 7.171420, 2.362317)$ , где  $W(\bar{z}) = 3.747692$ ,  $G^{(1)}(\bar{z}) \leq 0$ ,  $G^{(2)}(\bar{z}) \leq 0_2$ .

При  $\xi > 0$  множество  $Z(\xi) = \{z \in \mathbb{R}^2 : G^{(1)}(z) \leq \xi, G^{(2)}(z) \leq \xi\}$  состоит из двух изолированных заштрихованных областей, одна из которых содержит точку  $z_*$  вместе с некоторой ее окрестностью (см. случай II).

При  $\xi < 0$  множество  $Z(\xi)$  представляет собой односвязное множество и точка  $z_*$  ему не принадлежит (см. случай III).

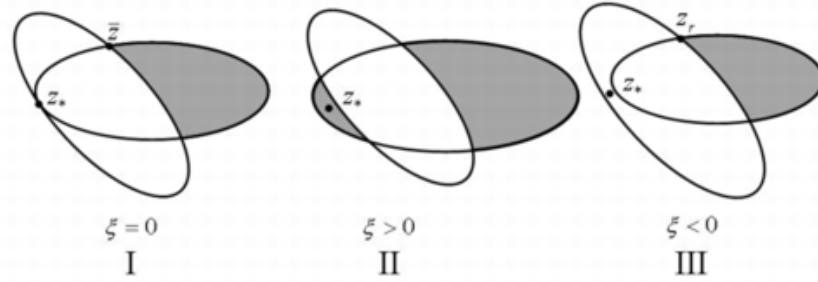


Рис. 2.5.1. Множества  $Z(\xi)$  (заштрихованы) при различных значениях  $\xi$

Расширение допустимого множества с помощью введения положительного параметра  $\xi$  привело к тому, что допустимое множество  $Z(\xi)$  стало состоять из двух областей, имеющих ненулевую меру. Это позволило находить точки  $z_*$  и  $\bar{z}$  с помощью локальных методов поиска экстремума функций. Так как в случае III допустимое множество  $Z(\xi)$  не односвязное, для расчетов следует использовать нелокальные методы оптимизации, например, метод неравномерных покрытий [57], предназначенный для поиска глобального минимума функций многих переменных.

Отличительным свойством задачи (2.5.18) является то, что все компоненты оптимального вектора  $z_*$  суть целые числа. Задачи, в которых требуется найти оптимальное решение, имеющее целочисленные компоненты, называются частично-целочисленными. Важно, что требование целочисленности существенно ускоряет вычислительный процесс решения задач нелинейного программирования, если для этой цели использовать метод неравномерных покрытий. Более подробно с этими результатами можно ознакомиться в работе [57].

Вторая особенность этой задачи состоит в том, что в ней при  $\xi = 0$  нарушено условие регулярности ограничений. Действительно, в точке  $z_*$  имеем

$$G_{z^*}^1 = [-8 \quad -4 \quad 0], \quad G_{z^*}^2 = [80 \quad 40 \quad 0],$$

т.е. эти векторы коллинеарны. Задачи такого типа обычно называют вырожденными (нерегулярными). Теории и методам решения таких задач посвящена обширная литература. Сошлемся, например, на работы [22, 58, 84–86]. ■

## 2.6. Теорема Куна–Таккера

Взяв в формулах (2.5.13) и (2.5.14)  $\xi = 0$ , приходим к следующей задаче нелинейного программирования:

$$\min_{z \in Z} W(z), \quad (2.6.1)$$

где допустимое множество определяется векторным условием типа неравенства

$$Z = \{z \in \mathbb{R}^n : h(z) \leq 0_m\}.$$

Составим для нее функцию Лагранжа

$$L^1(z, v) = W(z) + v^T h(z). \quad (2.6.2)$$

В отличие от задачи (2.4.1) здесь на множители Лагранжа наложено требование неотрицательности  $v \geq 0_m$ ; непустое множество решений задачи (2.6.1) по-прежнему обозначим  $Z_*$ .

Если  $h^i(z) = 0$ ,  $z \in Z$ , то  $i$ -е ограничение называется *активным* в точке  $z$ . Совокупность индексов всех активных ограничений в точке  $z$  обозначим через

$$\sigma(z) = \{i : h^i(z) = 0, 1 \leq i \leq m\}.$$

Число элементов, содержащихся в индексном множестве  $\sigma(z)$ , меньше или равно  $m$ . Будем говорить, что  $z$  — *регулярная точка* в задаче (2.6.1), или в точке  $z$  *выполнено у.р.о.*, если все векторы  $h_z^i(z)$ ,  $i \in \sigma(z)$ , линейно независимы. В регулярной точке  $z$  ранг прямоугольной  $n \times \ell$ -матрицы  $h_z^\top(z)$ , состоящей из градиентов активных ограничений, равен  $\ell$ , т.е. равен числу активных ограничений.

**Теорема 2.6.1 (Куна–Таккера).** Пусть  $W$  и  $h$  — непрерывно дифференцируемые функции, локальный минимум в задаче (2.6.1) достигается в регулярной допустимой точке  $z_*$ . Тогда существует единственный вектор множителей Лагранжа  $v_* \in \mathbb{R}_+^c$  такой, что

$$L_z^1(z_*, v_*) = W_z(z_*) + h_z^\top(z_*)v_* = 0_n, \quad (2.6.3)$$

$$v_*^j h^j(z_*) = 0, \quad 1 \leq j \leq m, \quad (2.6.4)$$

$$h(z_*) \leq 0_m, \quad (2.6.5)$$

$$v_* \geq 0_m. \quad (2.6.6)$$

Равенство (2.6.4) называется *условием дополняющей нежесткости*, совокупность всех условий (2.6.3) – (2.6.6) называются *условиями Куна–Таккера* для задачи (2.6.1).

**Доказательство.** Считаем для простоты, что в точке  $z_*$  активными являются первые  $\ell$  ограничений. Тогда вектор-функция  $h(z_*)$  расщепляется на две части:

$$h(z_*) = \begin{bmatrix} \hat{h}(z_*) \\ \bar{h}(z_*) \end{bmatrix},$$

где  $\hat{h}(z_*) = 0_\ell$ ,  $\bar{h}(z_*) < 0_s$ ,  $\ell + s = m$ . Определим множество из  $\mathbb{R}^m$ :

$$\hat{Z} = \{z \in Z : \hat{h}(z) = 0_\ell\}.$$

По построению множество  $\hat{Z} \subseteq Z$ , следовательно,

$$\min_{z \in \hat{Z}} W(z) \geq \min_{z \in Z} W(z) = W(z_*). \quad (2.6.7)$$

Точка  $z_* \in Z_*$  и, вместе с тем,  $z_* \in \hat{Z}$ , поэтому в (2.6.7) имеют место равенства, и вместо минимизации  $W(z)$  на множестве  $Z$  можно осуществлять минимизацию на множестве  $\hat{Z}$ .

По условию теоремы существует окрестность  $\Gamma$  точки  $z_*$  такая, что для любых  $z \in \Gamma \cap Z$  имеет место неравенство  $W(z) \geq W(z_*)$ . Возьмем окрестность  $\Gamma^*$  точки  $z_*$  настолько малой, что всюду на  $\Gamma^*$  неравенство  $\bar{h}(z) \leq 0_s$  будет автоматически выполнено; поэтому на множестве  $\Gamma^*$  оно несущественно. Вместо решения исходной задачи (2.6.1) можно перейти к нахождению локального минимума в более простой задаче

$$\min_{z \in \hat{Z}} W(z), \quad (2.6.8)$$

содержащей ограничения только типа равенства. К этой задаче применим теорему 2.4.1 Лагранжа, взяв в ней вместо  $Z$  множество  $\hat{Z}$ , вместо  $G(z)$  используем вектор-функцию  $\hat{h}(z)$ . Функцию Лагранжа для данной задачи запишем в виде

$$L^2(z, p) = W(z) + \sum_{i=1}^{\ell} p^i \hat{h}^i(z).$$

Считаем, что условия теоремы 2.4.1 здесь выполнены, поэтому из регулярности активных ограничений следует, что ранг матрицы  $\hat{h}_z(z_*)$  максимальный, равен  $m$ , следовательно, существует единственный вектор  $p_* \in \mathbb{R}^m$  такой, что

$$L_z^2(z_*, p_*) = W_z(z_*) + \sum_{i=1}^{\ell} p_*^i \hat{h}_z^i(z_*) = 0_n. \quad (2.6.9)$$

Определим вектор  $v_* \in \mathbb{R}^c$ , положив  $v_*^i = p_*^i$  для  $1 \leq i \leq \ell$  и  $v_*^j = 0$  при  $\ell + 1 \leq j \leq m$ . Тогда условие дополняющей нежесткости (2.6.4) будет выполнено, а из (2.6.9) следует (2.6.3). Условие допустимости (2.6.5) учтено в задаче (2.6.8)

Представим вектор  $z \in \mathbb{R}^n$  в виде  $z = z_* + \tau q$ , где  $q \in \mathbb{R}^n$ ,  $\tau \in \mathbb{R}_+^1$ . В силу дифференцируемости функций  $W$  и  $h$  имеем

$$W(z) = W(z_*) + \tau \cdot W_{z^\top}(z_*)q + o(\tau), \quad (2.6.10)$$

$$\hat{h}(z) = \hat{h}(z_*) + \tau \hat{h}_{z^\top}(z_*)q + o(\tau). \quad (2.6.11)$$

Пусть вектор  $q$  удовлетворяет следующей линейной системе из  $\ell$  уравнений:

$$\hat{h}_{z^\top}(z_*)q = b, \quad (2.6.12)$$

где  $b$  есть  $s$ -мерный вектор. В силу условий регулярности ограничений при всяком фиксированном векторе  $b$  система (2.6.12) имеет не более одного решения  $\bar{q}$ .

Учитывая (2.6.9) и (2.6.12), перепишем (2.6.10) и (2.6.11) в виде

$$\begin{aligned} W(z) &= W(z_*) - \tau \sum_{i=1}^{\ell} p_*^i b^i + o(\tau), \\ \hat{h}(z) &= \hat{h}(z_*) + \tau b + o(\tau). \end{aligned}$$

Оптимальный вектор множителей Лагранжа  $p_*$ , соответствующий задаче (2.6.8) с ограничениями только типа равенства, может иметь как положительные, так и отрицательные компоненты. Докажем, что в данном случае все компоненты вектора  $p_*$  неотрицательны.

Предположим от противного, что у вектора  $p_*$  первые  $k$  компонент отрицательные, остальные компоненты неотрицательные. В соответствии с этим определим компоненты вектора  $b$ , в результате получим, что

$$\begin{aligned} \text{при } 1 \leq i \leq k \text{ имеем} & \quad \hat{p}_*^i < 0, \quad b^i = -1, \quad \hat{h}^i(z_*) = 0, \\ \text{при } k+1 \leq i \leq \ell \text{ имеем} & \quad \begin{cases} \hat{p}_*^i \geq 0, \quad b^i = \varepsilon, \quad \hat{h}^i(z_*) = 0, \quad \varepsilon < 0, \\ \hat{h}^i(z) = \hat{h}^i(z_*) + \varepsilon\tau + o(\tau) \leq 0 \text{ при } \varepsilon \text{ достаточно малом.} \end{cases} \end{aligned}$$

Отсюда следует

$$W(z) - W(z_*) = \tau \sum_{i=1}^k p_*^i - \varepsilon\tau \sum_{i=k+1}^{\ell} p_*^i + o(\tau) \leq \frac{\tau}{2} \sum_{i=1}^k p_*^i < 0.$$

Таким образом, точка  $z_* + \tau\bar{q}$  — допустимая при  $\tau > 0$  и  $|\varepsilon|$  достаточно малых; кроме того,  $W(z_* + \tau\bar{q}) - W(z_*) < 0$ , что противоречит тому, что  $z_*$  — точка минимума в задаче (2.6.8). Поэтому  $p_* \geq 0_m$  и, следовательно,  $v_* \geq 0_c$ .  $\square$

Другой вариант доказательства состоит в использовании леммы Фаркаша. Проще, однако, воспользоваться формулами быстрого автоматического дифференцирования (2.5.12) и (2.5.17), откуда сразу следует, что  $p_* \geq 0_m$ .



С функцией Лагранжа связана *двойственная функция Лагранжа*

$$\psi(v) = \min_{z \in \mathbb{R}^n} L(z, v). \quad (2.6.13)$$

*Двойственная задача* состоит в нахождении

$$\max_{v \in \mathbb{R}_+^m} \psi(v).$$

Некоторые из ограничений  $h^j(x) \leq 0$  можно не включать в функцию Лагранжа (2.6.2), но тогда вместо безусловной минимизации в (2.6.13) следует искать минимум по  $z$  на множестве, определяемом неучтенными в (2.6.2) ограничениями  $h^j(x) \leq 0$ .

В качестве примера рассмотрим случай, когда в определении допустимого множества  $Z$  помимо ограничения  $h(z) \leq 0_c$  присутствует ограничение  $z \geq 0_n$ . Функцию Лагранжа определим по-прежнему формулой (2.6.2), не включая в нее ограничение  $z \geq 0_n$ , но при этом задачу безусловной минимизации (2.6.13) заменим на задачу минимизации по  $z$  на положительном ортанте:

$$\psi(v) = \inf_{z \in \mathbb{R}_+^n} L(z, v).$$

**Пример 2.6.1.** Пусть в задаче (2.6.1) допустимое множество имеет вид:  $Z = \{z \in \mathbb{R}^n : z \geq a, a \in \mathbb{R}^n\}$ . Составим функцию Лагранжа  $L(z, v) = W(z) + v^\top(a - z)$  где  $v$  —  $n$ -мерный вектор множителей Лагранжа, все компоненты которого неотрицательны.

Из условия (2.6.3), УДН (2.6.4) и допустимости (2.6.5), (2.6.6) получим условия Куна–Таккера

$$W_z(z_*) - v_* = 0_n, \quad z_* \geq a, \quad v_* \geq 0_n, \quad v_*^i(a^i - z_*^i) = 0, \quad 1 \leq i \leq n.$$

Эти условия можно записать в сокращенном виде, не используя вектор  $v_*$ :

$$W_z(z_*) \geq 0_n, \quad W_{z^i}(z_*)(a^i - z_*^i) = 0, \quad 1 \leq i \leq n.$$

В частности, если  $W(z) = \|z - \bar{z}\|^2/2$ , где  $\bar{z} \in \mathbb{R}^n$ , то  $W_z(z_*) = z_* - \bar{z}$ , и пример 2.6.1 превращается в задачу нахождения проекции точки  $\bar{z}$  на множество  $Z$ , при этом

$$W_z(z_*) = z_* - \bar{z} \geq 0_n, \quad (z_*^i - \bar{z}^i)(a^i - z_*^i) = 0, \quad z_*^i \geq a^i, \quad 1 \leq i \leq n.$$

Непосредственной проверкой можно убедиться, что все эти соотношения эквивалентны одному векторному выражению

$$z_* = a + (\bar{z} - a)_+. \quad (2.6.14)$$

Если  $a = 0_n$ , то отсюда следует  $z_* = \bar{z}_+ = \text{pr}(\bar{z}, \mathbb{R}_+^n)$ , и можно сказать, что оператор  $\bar{z}_+$  проектирует вектор  $\bar{z}$  на положительный ортант.

Если задан вектор  $z$ , то у вектора  $z_+$  равны нулю все компоненты, соответствующие отрицательным компонентам вектора  $z$ , все положительные и нулевые компоненты векторов  $z_+$  и  $z$  совпадают. ■

В примере 2.6.1 и в ряде других мест после проведения минимизации  $L$  по  $z$  (т.е. решив внутреннюю задачу (2.6.13) или выписав условия экстремума) мы будем к переменным, по которым ищется экстремум, добавлять звездочку. Точно так же после решения внешней задачи к вектору  $v$  будем приписывать звездочку.

Пример 2.6.1 показывает, что можно несколько видоизменить условия Куна–Таккера (2.6.3)–(2.6.6), записав их следующим образом:

$$\begin{aligned} L_z(z_*, v_*) &\geq 0_n, \quad D(v_*)h(z_*) = 0_c, \quad D(z_*)L_z(z_*, v_*) = 0_n, \\ h(z_*) &\leq 0_m, \quad z_* \geq 0_n, \quad v_* \geq 0_n. \end{aligned} \quad (2.6.15)$$

Здесь  $D(q)$  — диагональная матрица, у которой на главной диагонали стоит вектор  $q$ .

В условия Куна–Таккера (2.6.15) входят исходные переменные  $z$  и множители Лагранжа  $v$ . Таким образом, исходная задача (2.6.1) сводится к решению системы Куна–Таккера с неизвестными  $z$  и  $v$ . Если внутренняя задача (2.6.13) решается аналитически, то система редуцируется к двойственной задаче, содержащей только множители Лагранжа  $v$ . Это иллюстрирует следующий пример.

**Пример 2.6.2.** Пусть допустимое множество  $Z$  задается с помощью ограничений типа неравенства

$$Z = \{z \in \mathbb{R}^n : Az \geq b\}.$$

Здесь матрица  $A \in \mathbb{R}^{m \times n}$ ; предполагается, что  $n \geq m$  и множество  $Z$  не пусто. Ищется проекция  $z_*$  точки  $\bar{z}$  на множество  $Z$ . Функция Лагранжа для рассматриваемой задачи имеет вид

$$L(z, v) = \|z - \bar{z}\|^2/2 + v^\top(b - Az), \quad \text{где } v \in \mathbb{R}_+^m.$$

Из (2.6.3), (2.6.4) имеем

$$z_* - \bar{z} - A^\top v = 0_n, \quad D(v)(b - Az_*) = 0_m. \quad (2.6.16)$$

Умножив первое равенство слева на  $A$ , получим  $AA^\top v = \bar{b}$ , где  $\bar{b} = b - A\bar{z}$ . Отсюда следует

$$z_* = \bar{z} + A^\top v = \text{pr}(\bar{z}, Z), \quad D(v)(\bar{b} - AA^\top v) = 0_m. \quad (2.6.17)$$

Подставляя выражение для  $z_*$  из (2.6.17) в  $L(z, v)$ , получим вогнутую двойственную функцию

$$\psi(v) = v^\top \bar{b} - \|A^\top v\|^2/2.$$

Необходимые и достаточные условия максимума  $\psi(v)$  по  $v \in \mathbb{R}_+^m$  имеют вид

$$\psi_v(v_*) = \bar{b} - AA^\top v_* \leq 0_m, \quad D(v_*)(\bar{b} - AA^\top v_*) = 0_m. \quad (2.6.18)$$

Отсюда следует

$$\psi(v_*) = \|z_* - \bar{z}\|^2/2 = \|A^\top v_*\|^2/2 = v_*^\top \bar{b}/2 \geq 0.$$

Определяемая из (2.6.16) точка  $z_*$  принадлежит множеству  $Z$ . Действительно, с учетом (2.6.18) имеем

$$Az_* = A\bar{z} + AA^\top v_* \geq A\bar{z} + \bar{b} = b.$$

Следовательно,  $z_* = \text{pr}(\bar{z}, Z)$ ,  $\text{dist}(\bar{z}, Z) = \|A^\top v_*\|$ ,  $\|z_* - \bar{z}\|^2 = \|A^\top v_*\|^2 = v_*^\top \bar{b}_+$ , где

$$0 \leq v_*^\top \bar{b} \leq \|v_*\| \|\bar{b}_+\|, \quad \bar{b}_+ = (b - A\bar{z})_+. \quad (2.6.19)$$

Пусть  $\lambda$  — наименьшее собственное значение матрицы  $AA^\top$ . Тогда имеют место неравенства

$$\|v_*\| \leq \|\bar{b}_+\|/\lambda, \quad \text{dist}(\bar{z}, Z) = (v_*^\top \bar{b})^{1/2} \leq \|\bar{b}_+\|/\sqrt{\lambda}, \quad (2.6.20)$$

аналогичные (2.4.18) и (2.4.19). В (2.6.19) можно вместо евклидовой использовать гильбертовскую норму; тогда

$$v_*^\top \bar{b}_+ \leq \|v_*\|_\gamma \|\bar{b}_+\|_{\gamma_*},$$

где  $1/\gamma + 1/\gamma_* = 1$ . В частности, если  $\gamma = 1$ , то  $\gamma_* = \infty$  и

$$\text{dist}(\bar{z}, Z) \leq \left[ \|v_*\|_1 \cdot \max_{1 \leq i \leq m} (b - A\bar{z})_+^i \right]^{1/2} = [\|v_*\|_1 \cdot \|(b - A\bar{z})_+\|_\infty]^{1/2}.$$

Учитывая (2.6.20) и неравенства  $\|v_*\|_1 \leq \sqrt{m}\|v_*\|$  и  $\|\bar{b}\| \leq \sqrt{m}\|\bar{b}_+\|_\infty$  между гельдеровскими нормами, получим

$$\|v_*\|_1 \leq \sqrt{m}\|\bar{b}_+\|/\lambda \leq m\|\bar{b}_+\|_\infty/\lambda, \quad \text{dist}(\bar{z}, Z) \leq \sqrt{m/\lambda}\|(b - A\bar{z})_+\|_\infty.$$

При  $\bar{z} = 0_n$  вектор  $z_*$  совпадает с  $\tilde{z}$  — нормальным решением системы  $Az \geq b$ . Из найденных соотношений имеем

$$\|\tilde{z}\| = (v_*^\top b)^{1/2} \leq \|b\|/\sqrt{\lambda}, \quad \|v_*\| \leq \|b\|/\lambda.$$

Аналогичные результаты для равенств приведены выше в (2.4.19).

Оптимальный вектор  $v_*$  можно найти либо максимизируя вогнутую квадратичную функцию  $\psi(v)$  на  $\mathbb{R}_+^m$ , либо решая систему (2.6.18) на неотрицательном ортанте. Решив двойственную задачу, получим вектор  $v_*$  и по первой формуле в (2.6.16) определим проекцию вектора  $\bar{z}$  на множество  $Z$ , а при  $\bar{z} = 0_n$  — нормальное решение системы  $Az \geq b$ . Этот переход особенно удобен, когда  $n \gg m$ . Такой же прием использовался в примере 2.4.2, в котором вместо решения задачи квадратичной условной минимизации (2.4.15) в  $\mathbb{R}^n$  решалась задача безусловной минимизации квадратичной функции (2.4.16) в  $\mathbb{R}^m$  и после этого из формулы (2.4.17) определялась проекция  $z_*$ .

Двойственная задача может иметь неединственное решение, однако, подставляя любое из этих решений в (2.4.15), получаем единственную проекцию вектора  $\bar{z}$  на  $Z$ . ■

## 2.7. Общая задача нелинейного программирования

Теорема Лагранжа может быть объединена с теоремой Куна–Таккера для получения необходимых условий локального экстремума функций при наличии *смешанных* ограничений. В общей задаче НЛП допустимое множество определяется ограничениями типа равенства и неравенства:

$$\min_{z \in Z} W(z), \quad Z = \{z \in \mathbb{R}^n : G(z) = 0_m, h(z) \leq 0_c\}. \quad (2.7.1)$$

Предполагаем, что множество решений  $Z_*$  задачи (2.7.1) не пусто. Все те ограничения типа неравенства, у которых  $h^i(x) = 0$ , называются *активными* в точке  $x$  и объединяются одним вектором  $\sigma(x)$ . Будем говорить, что в задаче (2.7.1) точка  $z$  *регулярная*, если векторы  $G_z^i(z)$ ,  $1 \leq i \leq m$ , и векторы  $h_z^j(z)$  такие, что  $j \in \sigma(z)$  линейно независимы.

Составим функцию Лагранжа для задачи (2.7.1):

$$L(z, p, v) = W(z) + G^\top(z)p + h^\top(z)v. \quad (2.7.2)$$

Сформулируем и докажем ряд теорем, дающих достаточные условия минимума в задаче (2.7.1). Рассмотрим две вспомогательные задачи отыскания максимина и минимакса функции Лагранжа:

$$V_1 = \sup_{p \in \mathbb{R}^m} \sup_{v \in \mathbb{R}_+^c} \inf_{z \in \mathbb{R}^n} L(z, p, v), \quad (2.7.3)$$

$$V_2 = \inf_{z \in \mathbb{R}^n} \sup_{p \in \mathbb{R}^m} \sup_{v \in \mathbb{R}_+^c} L(z, p, v). \quad (2.7.4)$$

Тройка  $[z_*, p_*, v_*] \in \mathbb{R}^{n+s}$ , где  $v_* \geq 0_c$ ,  $s = m + c$ , будет *седловой точкой* функции  $L$ , если для любых  $z \in \mathbb{R}^n$ ,  $p \in \mathbb{R}^m$ ,  $v \in \mathbb{R}_+^c$  выполнены неравенства

$$L(z_*, p, v) \leq L(z_*, p_*, v_*) \leq L(z, p_*, v_*). \quad (2.7.5)$$

**Теорема 2.7.1.** Пусть  $[z_*, p_*, v_*]$  — седловая точка функции Лагранжа  $L$ . Тогда точка  $z_*$  является глобальным решением задачи (2.7.1) и в точке  $[z_*, v_*]$  имеет место условие дополняющей нежесткости

$$v_*^i h^i(z_*) = 0, \quad 1 \leq i \leq c. \quad (2.7.6)$$

**Доказательство.** Из левого неравенства в определении седловой точки (2.7.5) следует, что для любых  $p \in \mathbb{R}^m$ ,  $v \in \mathbb{R}_+^c$  выполнено условие

$$\sum_{i=1}^m G^i(z_*)[p^i - p_*^i] + \sum_{j=1}^c h^j(z_*)[v^j - v_*^j] \leq 0. \quad (2.7.7)$$

Положим  $v = v_*$ ,  $p^i = p_*^i$  для всех  $i \in [1 : m]$ , за исключением только одной компоненты  $p^k = p_*^k + 1$ . Подставив эти векторы в (2.7.7), получим  $G^k(z_*) \leq 0$ . Можно положить  $p^k = p_*^k - 1$ , и тогда придем к условию  $G^k(z_*) \geq 0$ . Остается одна возможность:  $G^k(z_*) = 0$ . В силу произвольности  $k$  приходим к выводу, что  $G(z_*) = 0$ . Взяв в (2.7.7)  $p = p_*$ ,  $v^j = v_*^j$  для всех  $j \in [1 : c]$ , за исключением  $v^\ell = v_*^\ell + 1$ , получим, что  $h^\ell(z_*) \leq 0$ . В силу произвольности  $\ell$  из  $[1 : c]$  приходим к выводу, что  $h(z_*) \leq 0_c$ . Таким образом, точка  $z_* \in Z$ , т.е. является допустимой.

Положим в (2.7.7)  $p = p_*$ ,  $v = 0_c$ ; тогда

$$0 \leq \langle h(z_*), v_* \rangle. \quad (2.7.8)$$

Вместе с тем  $h(z_*) \leq 0_c$  и  $0_c \leq v_*$ , поэтому для всех  $j \in [1 : c]$  выполнено неравенство  $h^j(z_*)v_*^j \leq 0$ . Сравнивая это неравенство с (2.7.8), приходим к выводу, что имеет место условие дополняющей нежесткости (2.7.6). Из правого неравенства (2.7.5) с учетом того, что  $z_* \in Z$  и имеет место (2.7.6), следует, что для всех  $z \in \mathbb{R}^n$  выполнены неравенства

$$L(z_*, p_*, v_*) = W(z_*) \leq W(z) + \langle p_*, G(z) \rangle + \langle v_*, h(z) \rangle.$$

В частности, для любых  $z \in Z$  имеем  $W(z_*) \leq W(z)$ . Отсюда заключаем, что  $z_* \in Z_*$  и  $z_*$  — глобальное решение общей задачи НЛП.  $\square$

Если в (2.7.3) внутренняя задача минимизации разрешима, то можно ввести двойственную функцию, обозначив

$$\tilde{L}(p, v) = \inf_{z \in \mathbb{R}^n} L(z, p, v).$$

Задача, двойственная к (2.7.1), состоит в нахождении

$$V_1 = \sup_{p \in \mathbb{R}^m} \sup_{v \in \mathbb{R}_+^c} \tilde{L}(p, v). \quad (2.7.9)$$

**Теорема 2.7.2.** Пусть множество решений задачи (2.7.1) не пусто; тогда справедливы оценки

$$V_1 \leq W(z_*) = V_2, \quad z_* \in Z_*. \quad (2.7.10)$$

**Доказательство.** Рассмотрим вначале задачу (2.7.4). В ней внутренняя задача максимизации решается просто:

$$\beta(z) = \sup_{p \in \mathbb{R}^m} \sup_{v \in \mathbb{R}_+^c} L(z, p, v) = \begin{cases} W(z), & \text{если } z \in Z, \\ +\infty, & \text{если } z \notin Z. \end{cases}$$

Решая внешнюю задачу, приходим к выводу, что минимум функции  $\beta(z)$  достигается на допустимом множестве и совпадает с минимумом функции  $W(z)$  на  $Z$ , поэтому

$$W(z_*) = \min_{z \in Z} W(z) = V_2.$$

Для любых  $z \in Z$ ,  $p \in \mathbb{R}^m$ ,  $v \in \mathbb{R}_+^c$  имеет место неравенство  $W(z) \geq L(z, p, v)$ . Проведем минимизацию левой и правой частей этого неравенства по  $z \in Z$ ; получим

$$W(z) \geq W(z_*) = \min_{z \in Z} W(z) \geq \min_{z \in Z} L(z, p, v) \geq \inf_{z \in \mathbb{R}^n} L(z, p, v).$$

В силу произвольности  $p \in \mathbb{R}^m$ ,  $v \in \mathbb{R}_+^c$  приходим к левому неравенству в (2.7.10).  $\square$

Если  $V_1 = V_2$ , то функция Лагранжа  $L(z, p, v)$  имеет седловую точку. Если  $V_1 < V_2$ , то говорят, что в задаче (2.7.1) имеет место *разрыв двойственности*.

**Теорема 2.7.3 (слабая теорема двойственности).** Пусть  $\bar{z} \in Z$ ,  $\bar{p}$ ,  $\bar{v}$  — допустимые точки в двойственной задаче (2.7.9). Тогда

$$W(\bar{z}) \geq \tilde{L}(\bar{p}, \bar{v}).$$

**Доказательство** следует из очевидных неравенств:

$$W(\bar{z}) \geq L(\bar{z}, \bar{p}, \bar{v}) \geq \inf_{z \in \mathbb{R}^n} L(z, \bar{p}, \bar{v}) = \tilde{L}(\bar{p}, \bar{v}).$$

$\square$

Теорема 2.7.1 дает достаточное условие минимума в задаче (2.7.1). Приведем другое достаточное условие.

**Теорема 2.7.4.** Пусть решением задачи (2.7.3) является тройка векторов  $[z_*, p_*, v_*] \in \mathbb{R}^{n+s}$ , у которой  $z_* \in Z$ ,  $v_* \geq 0_c$ , и выполнено условие дополняющей нежесткости. Тогда  $z_* \in Z_*$ .

**Доказательство.** При  $p = p_*$ ,  $v = v_*$  точка  $z_*$  должна быть решением внутренней задачи в (2.7.3), т.е. при любом  $z \in \mathbb{R}^n$  выполняется неравенство

$$W(z_*) + \langle p_*, G(z_*) \rangle + \langle v_*, h(z_*) \rangle = W(z_*) \leq W(z) + \langle p_*, G(z) \rangle + \langle v_*, h(z) \rangle.$$

Если  $z \in Z$ , то из этого неравенства следует, что

$$W(z_*) \leq L(z, p_*, v_*) \leq W(z),$$

откуда заключаем, что  $z_* \in Z_*$ .  $\square$

Имеет место теорема, объединяющая утверждения теорем Лагранжа и Куна–Таккера.

**Теорема 2.7.5.** Пусть  $W : \mathbb{R}^n \rightarrow \mathbb{R}^1$ ,  $G : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^c$ , функции  $W(z)$ ,  $G(z)$  дифференцируемы, в задаче (2.7.1) минимум достигается в регулярной допустимой точке  $z_*$ . Тогда существуют множители Лагранжа  $p_* \in \mathbb{R}^m$  и  $v_* \in \mathbb{R}_+^c$  такие, что

$$L_z(z_*, p_*, v_*) = W_z(z_*) + G_z^\top(z_*)p_* + h_z^\top(z_*)v_* = 0_n \quad (2.7.11)$$

и выполнены условия дополняющей нежесткости (2.6.4).

Теорема 2.7.5 является естественным обобщением теорем Лагранжа и Куна–Таккера. Здесь вместо функций Лагранжа, задаваемых формулами (2.4.3) и (2.6.2), берется функция (2.7.2). Аналогично обобщаются теоремы 2.4.1 и 2.6.1.

Векторы  $x_*$ ,  $u_*$ ,  $v_*$  удовлетворяют условиям Куна–Таккера в задаче (2.7.1), если выполнено условие (2.7.11), условие дополняющей нежесткости (2.7.6), условие допустимости прямой задачи:  $g(x_*) = 0_c$ ,  $h(x_*) \leq 0_c$  и допустимости двойственной задачи  $v_* \geq 0_c$ .

Будем говорить, что (2.7.1) — задача выпуклого программирования, если функции  $W(z)$  и  $h(z)$  выпуклые, а функция  $G(x)$  линейная. Покажем, что для задачи выпуклого программирования (2.7.1) условия Куна–Таккера являются не только необходимыми, но и достаточными условиями существования седловой точки  $[z_*, p_*, v_*]$ , удовлетворяющей неравенствам (2.7.5). Действительно, из выпуклости  $W(z)$  и  $h(z)$  и линейности  $G(z)$  следует, что для произвольного  $z \in \mathbb{R}^n$  имеют место соотношения

$$W(z) \geq W(z_*) + W_{z^\top}(z_*)(z - z_*), \quad (2.7.12)$$

$$G(z) = G(z_*) + G_{z^\top}(z_*)(z - z_*), \quad (2.7.13)$$

$$h(z) \geq h(z_*) + g_{z^\top}(z_*)(z - z_*). \quad (2.7.14)$$

Умножим (2.7.13) скалярно на  $p_*$ , (2.7.14) — скалярно на  $v_*$ , сложим полученные выражения с (2.7.12), учтем, что  $v_* \geq 0_c$ , воспользуемся равенством (2.7.11), получим

$$\begin{aligned} L(z, p_*, v_*) &= W(z) + \langle v_*, h(z) \rangle \geq L(z_*, p_*, v_*) + \\ &+ (W_{z^\top}(z_*) + p_*^\top G_{z^\top}(z_*) + v_*^\top h_{z^\top}(z_*))(z - z_*) = \\ &= L(z_*, p_*, v_*) = W(z_*) + \langle v_*, h(z_*) \rangle. \end{aligned}$$

Воспользуемся условием дополняющей нежесткости (2.7.6) и предположением, что  $h(z) \leq 0_c$ , получим

$$W(z) \geq v_*^\top h(z) + W(z_*) \geq W(z_*).$$

Отсюда заключаем, что  $W(z) \geq W(z_*)$  для всякого  $z \in Z$ , т.е.  $z_* \in Z_*$ .

Таким образом, правое неравенство в (2.7.5) выполнено. Левое неравенство имеет в данном случае вид

$$\sum_{i=1}^c v^i h^i(z_*) \leq \sum_{i=1}^c v_*^i h^i(z_*).$$

В силу условий дополняющей нежесткости правая часть этого неравенства равна нулю. Так как  $z_* \in Z$ , получаем, что  $\langle v, h(z_*) \rangle \leq 0$  при любых  $v \geq 0_c$ . Поэтому все неравенства (2.7.5) выполнены.

**Пример 2.7.1.** Пусть допустимое множество  $Z$  задается с помощью ограничений типа равенства и неравенства

$$Z = \{z \in \mathbb{R}^n : Az = b, z \geq 0_n\}.$$

Здесь  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $n \geq m$ .

Предполагается, что множество  $Z$  не пусто, ищется вектор  $z_*$  — проекция вектора  $\bar{z} \in \mathbb{R}^n$  на множество  $Z$ . Функция Лагранжа и двойственная функция имеют вид

$$L(z, p) = \|z - \bar{z}\|^2/2 + p^\top(b - Az), \quad \varphi(p) = \min_{z \in \mathbb{R}_+^n} L(z, p).$$

Необходимые условия минимума запишем в виде

$$L_z(z, p) = z - \bar{z} - A^\top p \geq 0_n, \quad D(z)(z - \bar{z} - A^\top p) = 0_n.$$

Отсюда, учитывая (2.6.14), получим  $z(p) = (\bar{z} + A^\top p)_+$  и

$$\varphi(p) = b^\top p - \|(\bar{z} + A^\top p)_+\|^2/2 + \|\bar{z}\|^2/2.$$

Двойственная задача оказалась задачей безусловной максимизации:

$$\varphi(p_*) = \max_{p \in \mathbb{R}^m} \varphi(p). \quad (2.7.15)$$

Согласно теореме Ферма вектор  $p_*$ , являющийся решением этой задачи, удовлетворяет условию

$$\varphi_p(p_*) = b - A(\bar{z} + A^\top p_*)_+ = 0_m.$$

Следовательно, искомый вектор  $z_* = (\bar{z} + A^\top p_*)_+ \in Z$  является проекцией вектора  $\bar{z}$  на  $Z$ . Отсюда следует, что вектор с наименьшей евклидовой нормой (нормальный вектор)  $z_*^n$  из  $Z$  определяется по формуле

$$z_*^n = (A^\top p_*)_+, \quad (2.7.16)$$

которая следует из (2.7.16), если в ней положить  $\bar{z} = 0_n$ . Расстояние от точки  $\bar{z}$  до множества  $Z$  есть

$$\text{dist}(\bar{z}, Z) = \|(\bar{z} + A^\top p_*)_+ - \bar{z}\| = \|z_* - \bar{z}\|.$$

Аналогичные результаты были получены в примере 2.4.1, где отсутствовало ограничение на знак вектора  $z$ . ■

## 2.8. Условия оптимальности второго порядка

Рассмотрим задачу (2.4.1). Предположим, что известны векторы  $z_* \in Z$  и  $p_*$  такие, что  $[z_*, p_*]$  есть стационарная точка функции Лагранжа  $L(z, p)$ , т.е. выполнены условия (2.4.9). Если в точке  $z_*$  выполнены условия регулярности ограничений, то можно считать, что матрица  $G_x^\top(z_*)$  не вырождена,  $z_*^\top = [x_*^\top, u_*^\top]$ . Воспользуемся теоремами 1.1.2 и 1.2.1; по формулам (1.1.9), (1.2.5), (1.2.10) определим соответственно  $\nabla\Omega(u)$ ,  $K(u)$  и  $\Delta\Omega(u)$ .

**Теорема 2.8.1 (достаточные условия минимума).** Пусть функции  $W(z)$  и  $G(z)$  дважды непрерывно дифференцируемы в некоторой окрестности допустимой регулярной точки  $z_* \in Z$  и пусть  $[z_*, p_*]$  есть стационарная точка функции Лагранжа  $L(z, p)$ . Тогда для того, чтобы  $z_*$  была точкой строгого локального минимума в задаче (2.4.1), достаточно выполнения какого-либо из следующих условий:

а)  $r \times r$  матрица  $\Delta\Omega(u_*) = K(u_*)L_{zz^\top}(z_*, p_*)K^\top(u_*)$  положительно определена;

б)  $n \times n$  матрица  $L_{zz^\top}(z_*, p_*)$  положительно определена на ядре матрицы  $G_{z^\top}(z_*)$ , т.е. для любых ненулевых векторов  $\xi_* \in \ker G_{z^\top}(z_*)$  выполнено условие  $\xi_*^\top L_{zz^\top}(z_*, p_*)\xi_* > 0$ .

**Доказательство.** Согласно (2.4.7) из условия стационарности функции Лагранжа следует, что  $\nabla\Omega(u_*) = 0_m$ . Поэтому формулу (1.2.11) с учетом (1.1.18) можно записать в виде

$$\begin{aligned} \Omega(u) - \Omega(u_*) &= (u - u_*)^\top K(u_*)L_{zz^\top}(z_*, p_*)K^\top(u_*)(u - u_*)/2 + \\ &+ \alpha(u_*, \|u - u_*\|) \|u - u_*\|^2. \end{aligned} \quad (2.8.1)$$

Пусть выполнено условие **а**; тогда, обозначив через  $\lambda_*$  наименьшее собственное значение матрицы  $\Delta\Omega(u_*)$ , получим

$$(u - u_*)^\top K(u_*)L_{zz^\top}(z_*, p_*)K^\top(u_*)(u - u_*) \geq \lambda_* \|u - u_*\|^2$$

при любых  $u - u_*$ . Из (2.8.1) следует

$$\Omega(u) - \Omega(u_*) \geq [\lambda_*/2 + \alpha(u_*, \|u - u_*\|)] \|u - u_*\|^2.$$

Окрестность  $\Gamma(u_*)$  можно взять настолько малой, что  $|\alpha(u_*, \|u - u_*\|)| \leq \lambda_*/4$  для всех  $u \in \Gamma(u_*)$ . Поэтому

$$\Omega(u) - \Omega(u_*) \geq \lambda_* \|u - u_*\|^2/4 \quad (2.8.2)$$

для всех  $u$ , достаточно близких к  $u_*$ . Отсюда заключаем, что  $u_*$  — точка строгого локального минимума функции  $\Omega(u)$  на  $\Gamma(u_*)$ , и, следовательно, точка локального минимума функции  $W(z)$  на допустимом множестве  $Z$ .

Пусть выполнено условие **б**. Введем вектор  $\eta(u) = K^\top(u_*)u \in \mathbb{R}^n$ . Из (1.2.6) следует

$$G_{z^\top}(z_*)K^\top(u_*) = 0_{mr}.$$

Матрица  $K^\top(u_*)$  размера  $n \times r$  имеет максимальный ранг, равный  $r$ . Поэтому столбцы этой матрицы образуют нуль-пространство матрицы  $G_{z^\top}(z_*)$ , вектор  $\eta(u) \in \ker[G_{z^\top}(z_*)]$  и существует число  $\mu_* > 0$  такое, что

$$[\eta(u) - \eta(u_*)]^\top L_{zz^\top}(z_*, p_*) [\eta(u) - \eta(u_*)] \geq \mu_* \|\eta(u) - \eta(u_*)\|^2$$

для всех  $\eta(u) - \eta(u_*)$ . Из определения  $\eta(u)$  следует

$$\|\eta(u) - \eta(u_*)\| \leq \|K(u_*)\| \|u - u_*\|,$$

поэтому  $\lim_{u \rightarrow u_*} \|\eta(u) - \eta(u_*)\| = 0$ . Взяв окрестность  $\Gamma(u_*)$  достаточно малой, получим неравенство, аналогичное (2.8.2):

$$\Omega(u) - \Omega(u_*) \geq \mu_* \|\eta(u) - \eta(u_*)\|^2/4,$$

откуда следует утверждение теоремы.  $\square$

Достаточное условие **б**, использованное в данной теореме, является общепринятым. Введение вместо него альтернативного (нетрадиционного) условия **а** позволяет вместо матрицы  $L_{zz^\top}(z_*, p_*)$  размера  $n \times n$  использовать приведенный гессиан  $\Delta L(u_*)$  — матрицу меньшей размерности  $r \times r$ . Такой переход к условию **а** особенно эффективен, когда числа  $m$  и  $r$  близки друг к другу, так как при этом разность  $r = n - m$  становится малым числом, что упрощает анализ достаточного условия **а**. В частности, если  $r = 1$ , то мы приходим к анализу вторых производных скалярной функции  $\Omega(u)$ .

Аналогично теореме 2.8.1 доказываются следующие результаты.

**Теорема 2.8.2.** Пусть функции  $W(z)$ ,  $G(z)$  принадлежат классу  $C^2$  и существуют векторы  $z_* \in Z$ ,  $p_* \in \mathbb{R}^m$ , удовлетворяющие (2.4.5) и (2.4.6). Тогда:

1) если  $z_*$  — точка локального минимума функции  $W$  на множестве  $Z$ , то

$$\eta^\top L_{zz^\top}(z_*, p_*) \eta \geq 0 \text{ для всех } \eta \in \ker G_z^\top(z_*);$$

2) если  $z_*$  — точка локального максимума функции  $W$  на множестве  $Z$ , то

$$\eta^\top L_{zz^\top}(z_*, p_*) \eta \leq 0 \text{ для всех } \eta \in \ker G_z^\top(z_*);$$

3) если  $\eta^\top L_{zz^\top}(z_*, p_*) \eta > 0$  для всех  $\eta \in \ker G_z^\top(z_*)$ ,  $\|\eta\| \neq 0$ , то  $z_*$  — точка строгого локального минимума функции  $W$  на множестве  $Z$ ;

4) если  $\eta^\top L_{zz^\top}(z_*, p_*) \eta < 0$  для всех  $\eta \in \ker G_z^\top(z_*)$ ,  $\|\eta\| \neq 0$ , то  $z_*$  — точка строгого локального максимума функции  $W$  на множестве  $Z$ .

Утверждения 1 и 2 дают необходимые условия оптимальности, утверждения 3 и 4 — достаточные условия экстремума. Утверждение 3 этой теоремы совпадает со случаем **б** в



теореме 2.8.1. Аналогично тому, как это делалось в теореме 2.8.1, в теореме 2.8.2 можно выразить необходимые и достаточные условия экстремума в терминах матрицы  $K(u_*)L_{zz^\top}(z_*, p_*)K^\top(u_*)$ . Все результаты формулируются аналогично теореме 2.1.2, роль матрицы  $W_{zz^\top}(z_*)$  здесь играет матрица  $L_{zz^\top}(z_*, p_*)$ , и вместо условий безусловного локального экстремума даются условия локального экстремума на множестве  $Z$ .

Рассмотрим локальный вариант максиминной задачи (2.4.13), т.е. в (2.4.12) ищется локальный минимум по  $z$  в окрестности точки  $z_*$ , а в (2.4.13) — локальный максимум сложной функции  $\varphi(p)$  по  $p$  в окрестности точки  $p_*$ . Согласно теореме Ферма необходимое условие минимума в задаче (2.4.12) запишется в виде

$$L_z(z, p) = W_z(z) + G_z^\top(z)p = 0_n. \quad (2.8.3)$$

Предположим, что матрица  $L_{zz^\top}(z_*, p_*)$  положительно определена. Тогда из условия (2.8.3) можно определить дифференцируемую функцию  $z(p)$  такую, что  $z(p_*) = z_*$ . Формулы для вычисления первых и вторых производных функции  $\Omega(p)$  имеют вид

$$\begin{aligned} \Omega(p) &= W(z(p)) + p^\top G(z(p)) = L(z(p), p), \\ \nabla \Omega(p) &= G(z(p)), \\ z_{p^\top}(p) &= -[L_{zz^\top}(z(p))]^{-1} G_z^\top(z(p)), \\ \Delta \Omega(p) &= -G_{z^\top}(z(p)) [L_{zz^\top}(z(p))]^{-1} G_z^\top(z(p)), \\ L_{zz^\top}(z, p) &= W_{zz^\top}(z) + \sum_{i=1}^m G_{zz^\top}^i(z)p^i. \end{aligned}$$

Если матрица  $L_{zz^\top}(p_*, x_*)$  положительно определена, то обратная матрица  $[L_{zz^\top}(p_*, x_*)]^{-1}$  также положительно определена. Отсюда следует

**Теорема 2.8.3 (о слабой двойственности).** Пусть в задаче (2.4.12) регулярная допустимая точка  $z_*$  является локальным решением; ей соответствует множитель Лагранжа  $p_*$  и положительно определенный гессиан  $L_{zz^\top}(z_*, p_*)$ . Тогда  $[p_*, z_*]$  является точкой строгого локального максимина в задаче (2.4.13).

**Пример 2.8.1.** Слегка модифицируем пример, взятый из книги [121, § 10.4]. Пусть  $n = 3$ ,  $m = 1$ ,  $r = 2$ ,  $W(z) = -z^1 z^2 - z^2 z^3 - z^1 z^3$ ,  $G(z) = z^1 + z^2 + z^3 - 3 = 0$ . Необходимое условие минимума (2.4.5) вместе с условием допустимости образуют систему из четырех уравнений с четырьмя неизвестными

$$\begin{aligned} -z_*^2 - z_*^3 + p_* &= 0, \\ -z_*^1 - z_*^3 + p_* &= 0, \\ -z_*^2 - z_*^1 + p_* &= 0, \\ z_*^1 + z_*^2 + z_*^3 &= 3. \end{aligned}$$

Решение этой системы следующее:  $z_*^1 = z_*^2 = z_*^3 = 1$ ,  $p_* = 2$ . Следовательно,  $x_* = 1$ ,  $u_*^1 = u_*^2 = 1$ . Матрица вторых производных функции Лагранжа

$$L_{zz^\top}(z_*, p_*) = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}$$

не является ни положительно, ни отрицательно определенной. Представим  $z^\top$  в виде  $z^\top = [x^1, u^1, u^2]$ . Тогда согласно (1.1.2), (1.2.5), (1.2.10) имеем  $x(u) = 1 - u^1 - u^2$ ,  $x_{u^\top} = [-1 | -1]$ ,

$$G_{z^\top}^1(z_*) = [1 \ 1 \ 1], \quad K^\top(u_*) = \begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\Delta\Omega(u_*) = K(u_*)L_{zz^\top}(z_*, p_*)K^\top(u_*) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad W(z_*) = \Omega(u_*) = -3.$$

Главные миноры матрицы  $\Delta\Omega(u_*)$  суть  $\delta_1 = 2 > 0$ ,  $\delta_2 = 3 > 0$ . Поэтому согласно критерию Сильвестра матрица  $\Delta\Omega(u_*)$  положительно определена и, следовательно, выполнено условие **a** теоремы 2.8.1. Таким образом, найденное решение  $z_*^1 = z_*^2 = z_*^3 = 1$  является точкой локального изолированного минимума в задаче (2.4.1). Собственные значения матрицы  $\Delta\Omega(u_*)$  суть 1 и 3, поэтому в оценке (2.8.2) можно взять  $\lambda_* = 1$ . В данном примере функция  $\Omega(u)$  квадратичная, поэтому в формуле Тэйлора (2.8.1) можно считать, что  $\alpha(\cdot) \equiv 0$ . Благодаря этому вместо (2.8.2) имеем  $\Omega(u) = \Omega(u_*) + \|u - u_*\|^2/2$ .

Рассмотрим условие **б**. Пусть  $\eta \in \mathbb{R}^3$ ; вычислим трехмерный вектор

$$L_{zz^\top}(z_*, p_*)\eta = \begin{bmatrix} -\eta^2 - \eta^3 \\ -\eta^1 - \eta^3 \\ -\eta^1 - \eta^2 \end{bmatrix}.$$

Умножая этот вектор скалярно на вектор  $\eta$ , получим

$$\eta^\top L_{zz^\top}(z_*, p_*)\eta = -\eta^1(\eta^2 + \eta^3) - \eta^2(\eta^1 + \eta^3) - \eta^3(\eta^1 + \eta^2).$$

На подпространстве  $G_{z^\top}(z_*)\eta = 0_1^\top$  имеем

$$\eta^\top L_{zz^\top}(z_*, p_*)\eta = (\eta^1)^2 + (\eta^2)^2 + (\eta^3)^2 > 0, \quad \text{если } \|\eta\| \neq 0.$$

Таким образом, в рассматриваемом примере достаточное условие **б** теоремы 2.8.1 также выполнено. Здесь  $r = n - m = 3 - 1 = 2$ , вместо традиционно используемой матрицы  $L_{zz^\top}$  размера  $3 \times 3$  можно взять приведенный гессиан размера  $2 \times 2$ . ■

## 2.9. Применение методов решения систем нелинейных уравнений к задачам НЛП

Выше было показано, что необходимые условия экстремума для задач НЛП представимы в виде систем нелинейных равенств и неравенств. Поэтому простейшие методы решения задач НЛП состоят в использовании численных методов, разработанных для решения нелинейных систем.

Наиболее универсальными и эффективными для оптимизации являются методы, основанные на использовании дифференциальных свойств функции, определяющих задачу. Эта группа методов наиболее подробно изучена, ей посвящены основные публикации, реализованы и практически используются многочисленные пакеты программ. К этой группе можно отнести градиентные методы, обладающие линейной локальной сходимостью (наискорейший спуск, методы сопряженных градиентов, квазиньютоновские методы), метод Ньютона и его многочисленные модификации, обладающие квадратичной скоростью сходимости, метод П.Л. Чебышева, имеющий более высокий порядок сходимости. Методы этой группы успешно используются для решения разнообразных задач, с их помощью часто удается добиться высокой точности расчетов и решать задачи большой размерности.

Ко второй группе можно отнести многочисленные эвристические методы, которые искусственно подбираются пользователем под класс решаемых задач. Такие методы не являются универсальными, их теоретическое обоснование обычно отсутствует. К этой группе можно отнести метод случайного поиска (самый простой для реализации), метод прыжков, объединяющий случайный поиск с локальными градиентными методами, генетические алгоритмы,

метод разогрева и т.д. Методы этой группы строятся по аналогии с некоторыми процессами, происходящими в живых системах. Успех их применения во многом зависит от того, насколько решаемая задача близка к аналогу, который используется для построения метода. Ниже ограничимся только методами первой группы.

Пусть задана нелинейная система из  $n$  уравнений с неизвестным  $n$ -мерным вектором  $z$ :

$$F(z) = 0_n, \quad (2.9.1)$$

где  $F(z)$  — дважды дифференцируемая функция,  $z \in \mathbb{R}^n$ . Для нахождения решений этой системы воспользуемся методом Ньютона

$$z_{k+1} = z_k - F_{zz}^{-1}(z_k)F_z(z_k). \quad (2.9.2)$$

При реализации метода необходимо либо обращать симметричную матрицу  $F_{zz}$  порядка  $n$ , либо решать систему из  $n$  линейных уравнений

$$F_{zz}(z_k)(z_{k+1} - z_k) = -F_z(z_k).$$

Существует обширная литература, посвященная модификациям и анализу сходимости метода Ньютона. При определенных условиях он сходится локально с квадратичной скоростью.

Задача безусловной минимизации функции многих переменных

$$\min_{z \in \mathbb{R}^n} \Phi(z)$$

сводится к задаче (2.9.1) при помощи теоремы Ферма. В этом случае следует положить  $F(z) = \Phi_z(z)$ . Общая задача НЛП (2.7.1) также сводится к (2.9.1). В простейшем случае для такой редукции в качестве  $\Phi$  берется гладкая невязка ограничений

$$\Phi(z) = \|G(z)\|^2 + \|h(z)_+\|^2.$$

Ее градиент имеет вид

$$\Phi_z(z) = 2 [G_z^\top(z)G(z) + h_z^\top(z)h_+(z)].$$

Из-за наличия оператора  $(\cdot)_+$  функция  $\Phi(z)$  не является дважды дифференцируемой. Поэтому применение метода Ньютона для минимизации  $\Phi(z)$  требует специального обоснования. Вместо градиента функции  $\Phi_z(z)$  здесь следует использовать субградиент, определенный в (2.1.12). Такой подход широко используется при решении задач линейного и нелинейного программирования [30–34, 105, 106, 119, 122].

С помощью теории двойственности в примерах 2.4.3, 2.6.2 и 2.7.1 задачи квадратичного программирования в пространстве  $\mathbb{R}^n$  свелись к максимизации квадратичных дифференцируемых двойственных функций в пространстве  $\mathbb{R}^m$ . Последние задачи можно решать с помощью различных методов безусловной минимизации. Рассмотрим для определенности задачу (2.7.15). Здесь  $\varphi(p)$  — вогнутая всюду дифференцируемая кусочно-квадратичная функция. Ее матрица Гессе не определена всюду, но для нее строится обобщенная матрица вторых производных. Воспользуемся примером 2.1.5, получим

$$\partial\varphi_p(p) = -AD^\#(z(p))A^\top.$$

Обобщенный метод Ньютона для решения задачи (2.7.15) запишется в виде

$$p_{k+1} = p_k - h_k[\partial\varphi_p(p_k)]^{-1}\varphi_p(p_k). \quad (2.9.3)$$

В простейшем случае шаг  $h_k$  берется постоянным и равным единице. Доказательство глобальной сходимости метода можно найти в работе О. Мангасарьяна [122].

Предполагаем, что есть точка  $z_*$ , являющаяся решением задачи (2.9.1). Возьмем точку  $z$ , принадлежащую некоторой окрестности точки  $z_*$ , расстояние между этими точками  $\|z - z_*\|$  считаем величиной первого порядка малости. Воспользуемся формулой Тейлора, получим разложение функции  $F(z)$  в окрестности точки  $z$

$$F(z_*) = F(z) + F_z^\top(z)(z_* - z) + \frac{1}{2}(z_* - z)^\top F_{zz}(z)(z_* - z) + O(\|z_* - z\|^3) = 0_m; \quad (2.9.4)$$

здесь  $O(\|\cdot\|^3)$  —  $m$ -мерный вектор, каждая компонента которого имеет 3-й порядок малости.

Предполагаем, что матрица  $F_{z^\top}(z_*)$  не вырождена. Тогда существует окрестность точки  $z_*$ , где матрица  $F_{z^\top}(z)$  также не вырождена и где можно определить матрицу  $N(z) = F_{z^\top}^{-1}(z)F(z)$ . Умножим (2.9.4) слева на матрицу  $F_{z^\top}^{-1}(z)$  и, отбрасывая члены третьего порядка малости, получим

$$N(z) = z - z_* - \frac{1}{2}F_{z^\top}^{-1}(z)(z_* - z)^\top F_{zz}(z)(z_* - z) + O(\|z - z_*\|^3). \quad (2.9.5)$$

Здесь, отбрасывая члены второго порядка малости, находим приближенную формулу для вектора  $N(z)$ :

$$N(z) \simeq z - z_* + O(\|z - z_*\|^2). \quad (2.9.6)$$

Подставим это выражение в оценку (2.9.5) в слагаемое, содержащее величины второго порядка малости, получим, что с погрешностью третьего порядка малости имеет место следующая оценка:

$$z_* - z = -N(z) - \frac{1}{2}F_{z^\top}^{-1}(z)N^\top(z)F_{zz^\top}(z)N(z) + O(\|z - z_*\|^3). \quad (2.9.7)$$

Найденные формулы позволяют определить два итеративных метода решения системы (2.9.1):

$$z_{i+1} = z_i - N(z_i), \quad N(z) = F_{z^\top}^{-1}(z)F(z), \quad (2.9.8)$$

$$z_{i+1} = z_i - N(z_i) - \frac{1}{2}F_{z^\top}^{-1}(z_i)N^\top(z_i)F_{zz^\top}(z_i)N(z_i). \quad (2.9.9)$$

Первый процесс есть метод Ньютона (2.9.2), для него согласно (2.9.8)

$$\|z_{i+1} - z_*\| = \|z_i - z_* - N(z_i)\|.$$

В формуле (2.9.6) в качестве  $z$  возьмем  $z_i$ , получим оценку скорости сходимости последовательности (2.9.8) к  $z_*$ :

$$\|z_{i+1} - z_*\| = O(\|z_i - z_*\|^2).$$

Мы пришли к хорошо известному утверждению о том, что метод Ньютона обладает локальной квадратичной скоростью сходимости. Второй итеративный процесс (2.9.9) имеет кубическую скорость сходимости, так как из (2.9.9) и (2.9.7) следует, что  $\|z_{i+1} - z_*\| = O(\|z_i - z_*\|^3)$ .

В частном случае, когда  $n = m = 1$ , из (2.9.9) получаем метод П.Л. Чебышева, в котором

$$z_{i+1} = z_i - N(z_i) - \frac{1}{2}[N(z_i)]^2 \frac{F_{zz}(z_i)}{F_z(z_i)}, \quad N(z_i) = \frac{F(z_i)}{F_z(z_i)}.$$

Для данного скалярного случая в работе П.Л. Чебышева [90] указаны еще два члена третьего и четвертого порядка малости:

$$-\frac{1}{2}[N(z_i)]^3 \left[ \frac{(F_{zz}(z_i))^2}{(F_z(z_i))^2} - \frac{F_{zzz}(z_i)}{3F_z(z_i)} \right] - [N(z_i)]^4 \left[ \frac{5(F_{zz}(z_i))^3}{8(F_z(z_i))^3} - \frac{5F_{zz}(z_i)F_{zzz}(z_i)}{12(F_z(z_i))^2} + \frac{F_{zzzz}(z_i)}{24F_z(z_i)} \right].$$

Несложно построить рекуррентные формулы для построения вариантов метода, обладающих  $p$ -й скоростью сходимости.

Применим метод (2.9.7) к решению задачи нелинейного программирования (2.4.1), в которой ищется

$$\min_{z \in Z} W(z), \quad Z = \{z \in \mathbb{R}^n : G(z) = 0_m\}. \quad (2.9.10)$$

Составим функцию Лагранжа  $L(z, p) = W(z) + G^\top(z)p$  и штрафную функцию

$$P(z, \varepsilon) = \varepsilon W(z) + \frac{1}{2} \|G(z)\|^2.$$

Необходимые условия минимума этих функций по  $z$  имеют вид

$$\begin{aligned} L_z(z, p) &= W_z(z) + G_z^\top(z)p = 0_n, & G(z) &= 0_m, \\ P_z(z, \varepsilon) &= \varepsilon W_z(z) + G_z^\top(z)G(z) = 0_n. \end{aligned}$$

Определим  $(n + m)$ -мерные вектор-функции

$$F(z, p) = \begin{bmatrix} W_z(z) + G_z^\top(z)p \\ G(z) \end{bmatrix}, \quad \Phi(\varepsilon, z, p) = \begin{bmatrix} W_z(z) + G_z^\top(z)p \\ G(z) - \varepsilon p \end{bmatrix}.$$

Продифференцируем эти функции по  $z$  и  $p$ , получим две матрицы Якоби

$$F'(z, p) = \begin{bmatrix} L_{zz}(z, p) & G_z^\top(z) \\ G_{z^\top}(z) & 0_{mm} \end{bmatrix}, \quad \Phi'(\varepsilon, z, p) = \begin{bmatrix} L_{zz} & G_z^\top p \\ G_{z^\top} & -\varepsilon J_m \end{bmatrix}.$$

Задача (2.9.10) сводится к решению какой-нибудь из двух систем:

$$F(z, p) = 0_{n+m}, \quad (2.9.11)$$

$$\Phi(\varepsilon, z, p) = 0_{n+m}. \quad (2.9.12)$$

Для решения системы (2.9.11) можно воспользоваться методом (2.9.8), считая, что в нем в качестве  $z$  взято объединение векторов  $z$  и  $p$ , тогда приходим к методу Ньютона. Если воспользоваться формулами (2.9.9), получим кубически сходящийся метод. Условие невырожденности матрицы  $F'(z_*, p_*)$  будет выполнено, если ранг матрицы  $G_{z^\top}(z_*)$  равен  $m$  и матрица  $L_{zz}(z_*, p_*)$  положительно определена на ядре матрицы  $G_z^\top(z_*)$  (см. теорему 2.5.1).

**Пример 2.9.1.** Рассмотрим простейший вариант задачи (2.9.8). Пусть  $n = m = 1$ ,  $W(z) = z$ ,  $G(z) = z + z^2$ . Тогда  $z_* = 0$ ,  $p_* = -1$ ,  $L(z, p) = z + p(z + z^2)$ . Задача сводится к решению системы вида (2.9.1), в которой  $n = m = 2$ , а в качестве вектора  $z$  берется пара  $z$  и  $p$ . Поэтому

$$F = \begin{bmatrix} 1 + p + 2pz \\ z + z^2 \end{bmatrix}, \quad F' = \begin{bmatrix} 2p & 1 + 2z \\ 1 + 2z & 0 \end{bmatrix}, \quad (F')^{-1} = \frac{1}{(1 + 2z)^2} \begin{bmatrix} 0 & 1 + 2z \\ 1 + 2z & -2p \end{bmatrix}.$$

Считаем  $|z_i - z_*|$ ,  $|p_i - p_*|$  величинами первого порядка малости (порядка  $\delta$ ). Тогда можно записать  $z_i - z_* = z_i \simeq O(\delta)$ ,  $p_i - p_* = p_i + 1 \simeq O(\delta)$ .

Воспользуемся стандартным методом Ньютона (2.9.8), получим

$$z_{i+1} = (z_i)^2 \simeq O(\delta^2), \quad p_{i+1} + 1 = 2z_i(1 + p_i) \simeq O(\delta^2). \quad (2.9.13)$$

Более точный результат следует из метода (2.9.9):

$$z_{i+1} \simeq O(\delta^3), \quad p_{i+1} + 1 \simeq O(\delta^3).$$

Согласно (2.9.13) метод Ньютона определения  $z_{i+1}$  и  $p_{i+1}$  с погрешностями соответственно  $(z_i)^2$  и  $2z_i(1+p_i)$ , имеющими второй порядок малости. В случае использования метода (2.9.9), напротив, получены точные решения  $z_{i+1} = 0$ ,  $p_{i+1} = -1$  с гарантией, что погрешность, если она есть, имеет третий порядок малости. ■

Метод внешних штрафных функций состоит в параметрическом представлении  $z(\varepsilon)$  и  $p(\varepsilon)$  решений системы из  $n + m$  уравнений с таким же числом неизвестных

$$\Phi(\varepsilon, z(\varepsilon), p(\varepsilon)) = 0_{n+m}. \quad (2.9.14)$$

Здесь решения  $z_*$  и  $p_*$  получаются асимптотически в пределе при  $\varepsilon \rightarrow 0$ .

Если размерность  $m$  невелика (порядка  $10^3$ ), то максимизацию двойственной функции можно проводить с помощью метода Ньютона (2.9.3) на обычном персональном компьютере типа Pentium IV. При этом размерность  $n$  может быть сравнительно большой (порядка  $10^6$ ).

В работах [28, 30–34, 119, 122] показано, что метод Ньютона может быть использован для решения задач линейного программирования большой размерности.

## 2.10. Некоторые свойства минимаксных задач

В задачах (2.2.4) и (2.2.3) никаких условий на  $x$  и  $u$  не накладывалось, поэтому их можно назвать задачами о поиске безусловного максимина и минимакса соответственно. Рассмотрим случай, когда на  $u$  и  $x$  наложены ограничения:

$$V_1 = \max_{x \in X} \min_{u \in U} W(x, u), \quad (2.10.1)$$

$$V_2 = \min_{u \in U} \max_{x \in X} W(x, u). \quad (2.10.2)$$

Определение решения этих задач дается аналогично тому, как это делалось выше (см. (2.2.5) и (2.2.6)):

$$B(u) = \text{Arg} \max_{x \in X} W(x, u), \quad \Omega(u) = W(u, B(u)), \quad U_* = \text{Arg} \min_{u \in U} \Omega(u). \quad (2.10.3)$$

Для приложений весьма важен вопрос: какие свойства функции  $W(x, u)$  изменяются и какие сохраняются после осуществления операции максимизации по  $x$ , т.е. при переходе от  $W(x, u)$  к  $\Omega(u)$ ? Чтобы ответить на этот вопрос, рассмотрим вначале свойства непрерывности, выпуклости и выполнимости условия Липшица. Для сокращения записи обозначим

$$z^\top = [x^\top, u^\top] \in \mathbb{R}^{r+m}, \quad Z = X \times U, \quad W(z) = W(x, u).$$

Будем говорить, что функция  $W(z)$  удовлетворяет условию Липшица на  $Z$  с константой  $\ell$  и функция  $\Omega(u)$  удовлетворяет условию Липшица на  $U$  с константой  $\ell$ , если для любых  $z_1, z_2 \in Z$  и  $u_1, u_2 \in U$  выполнены соответственно условия

$$\begin{aligned} |W(z_1) - W(z_2)| &\leq \ell \|z_1 - z_2\|, \\ |\Omega(u_1) - \Omega(u_2)| &\leq \ell \|u_1 - u_2\|. \end{aligned} \quad (2.10.4)$$

**Теорема 2.10.1.** *Если функция  $W(x, u)$  непрерывна на прямом произведении компактных множеств  $Z = X \times U$ , то функция  $\Omega(u)$  также непрерывна на  $U$ . Если функция  $W(z)$  удовлетворяет условию Липшица на  $Z$ , то функция  $\Omega(u)$  также удовлетворяет условию Липшица на  $U$  с той же константой. Если при каждом  $x \in X$  функция  $W(x, u)$  выпукла по  $u$  на выпуклом множестве  $U$ , то и  $\Omega(u)$  выпукла на  $U$ .*

**Доказательство.** Функция  $W(z)$ , будучи непрерывной на компактном множестве  $Z$ , равномерно непрерывна на этом множестве. Поэтому для любого  $\varepsilon > 0$  существует такое  $\delta(\varepsilon)$ , что если  $\|u_1 - u_2\| < \delta(\varepsilon)$ ,  $u_1, u_2 \in U$ , то

$$|W(x, u_1) - W(x, u_2)| < \varepsilon$$

для любого  $x \in X$ . Рассмотрим выражение

$$|\Omega(u_1) - \Omega(u_2)| = |W(x_1, u_1) - W(x_2, u_2)|.$$

Здесь обозначено  $x_1 \in B(u_1)$ ,  $x_2 \in B(u_2)$ . Рассмотрим два случая. В первом имеем

$$\Omega(u_1) - \Omega(u_2) \geq 0.$$

Учитывая определение множества  $B(u)$ , получим

$$0 \leq \Omega(u_1) - \Omega(u_2) \leq W(x_1, u_1) - W(x_2, u_1) \leq \varepsilon.$$

Если функция  $W(z)$  удовлетворяет условию Липшица по  $z$ , то она удовлетворяет условию Липшица по  $u$  и равномерно по  $x$ , поэтому

$$0 \leq \Omega(u_1) - \Omega(u_2) \leq \ell \|u_1 - u_2\|.$$

Во втором случае

$$\Omega(u_2) - \Omega(u_1) \geq 0.$$

Аналогичным образом получаем, что

$$0 \leq \Omega(u_2) - \Omega(u_1) \leq W(x_2, u_2) - W(x_1, u_2) \leq \varepsilon,$$

и если  $W$  удовлетворяет условию Липшица, то

$$0 \leq \Omega(u_2) - \Omega(u_1) \leq \ell \|u_2 - u_1\|.$$

Объединяя эти два случая, приходим к выводу, что при любых  $u_1$  и  $u_2$  таких, что  $\|u_1 - u_2\| < \delta(\varepsilon)$ , имеет место неравенство  $|\Omega(u_1) - \Omega(u_2)| < \varepsilon$ , т.е. функция  $\Omega(u)$  непрерывна на компактном множестве  $U$ . Если  $W$  удовлетворяет условию Липшица, то имеет место (2.10.4).

Если  $W$  выпукла по  $u$ , то для любых  $u_1, u_2 \in U$ ,  $0 \leq \lambda \leq 1$  имеют место оценки

$$\begin{aligned} \Omega(\lambda u_1 + (1 - \lambda)u_2) &= \max_{x \in X} F(x, \lambda u_1 + (1 - \lambda)u_2) \leq \\ &\leq \max_{x \in X} [\lambda W(x, u_1) + (1 - \lambda)W(x, u_2)] \leq \\ &\leq \lambda \Omega(u_1) + (1 - \lambda)\Omega(u_2), \end{aligned}$$

что доказывает выпуклость функции  $\Omega$ .  $\square$

Свойство дифференцируемости функции  $W$  не сохраняется при переходе к функции  $\Omega$ . Имеет место следующая теорема.

**Теорема 2.10.2 (Данскина–Демьянова).** Пусть  $U$  — открытое множество в  $\mathbb{R}^r$ ,  $X$  — компактное множество в  $\mathbb{R}^m$ , функция  $W$  непрерывна вместе с  $W_u(x, u)$  по совокупности переменных в  $X \times U$ . Тогда определяемая из (2.10.3) функция  $\Omega(u)$  имеет в каждой точке  $u \in U$  производную по любому направлению  $G \in \mathbb{R}^r$ ,  $\|G\| = 1$ , причем

$$\frac{\partial \Omega(u)}{\partial G} = \max_{x \in B(u)} \langle W_u(x, u), G \rangle.$$

Доказательство теоремы можно найти в [40–42]. В последней книге доказана также и следующая теорема.

**Теорема 2.10.3.** Пусть  $U$  — замкнутое выпуклое множество в  $\mathbb{R}^r$ ,  $X$  — компактное множество в  $\mathbb{R}^m$ , функция  $W$  непрерывна вместе с  $W_u(x, u)$  по совокупности переменных в  $X \times U$ . Тогда для того, чтобы функция  $\Omega(u)$  достигала минимального на  $U$  значения в точке  $u_* \in U$ , необходимо, а в случае выпуклости  $\Omega(u)$  на  $U$  и достаточно выполнение условия:

$$\inf_{u \in U} \max_{x \in B(u_*)} \langle W_u(x, u_*), u - u_* \rangle = 0.$$

В заключение приведем теорему о существовании решений задач (2.10.1) и (2.10.2).

**Теорема 2.10.4.** Если функция  $W(x, u)$  непрерывна на произведении компактных множеств  $X$  и  $U$ , то решения задач (2.10.1) и (2.10.2) существуют и имеет место неравенство

$$V_1 = \max_{x \in X} \min_{u \in U} W(x, u) \leq V_2 = \min_{u \in U} \max_{x \in X} W(x, u). \quad (2.10.5)$$

Доказательство теоремы почти дословно повторяет доказательство теоремы 2.2.2.

Точка  $z_* = [x_*, u_*]$  называется *седловой точкой* функции  $W(x, u)$  в задачах (2.10.1) и (2.10.2), если  $x_* \in X$ ,  $u_* \in U$  и для любых  $x \in X$ ,  $u \in U$  имеют место неравенства

$$W(x, u_*) \leq W(x_*, u_*) \leq W(x_*, u). \quad (2.10.6)$$

**Лемма 2.10.1.** Если задачи (2.10.1) и (2.10.2) имеют решения, то для того, чтобы имело место равенство

$$V_1 = W(x_*, u_*) = V_2, \quad (2.10.7)$$

необходимо и достаточно, чтобы  $z_*$  была седловой точкой функции  $W(x, u)$ .

**Доказательство.** Пусть  $[x_*, u_*]$  — седловая точка функции  $W(x, u)$ . Тогда

$$\max_{x \in X} W(x, u_*) \leq W(x_*, u_*) \leq \min_{u \in U} W(x_*, u).$$

Очевидны следующие неравенства:

$$\begin{aligned} \min_{u \in U} \max_{x \in X} W(x, u) &\leq \max_{x \in X} W(x, u_*), \\ \min_{u \in U} W(x_*, u) &\leq \max_{x \in X} \min_{u \in U} W(x, u). \end{aligned}$$

Объединяя последние три группы неравенств, получаем, что

$$\min_{u \in U} \max_{x \in X} W(x, u) \leq W(x_*, u_*) \leq \max_{x \in X} \min_{u \in U} W(x, u).$$

Сравнивая это неравенство с (2.10.5), приходим к выводу, что справедливо равенство (2.10.7).

Пусть имеет место (2.10.7). Покажем, что точка  $[x_*, u_*]$  — седловая. Из (2.10.7) следует

$$\begin{aligned} W(x_*, u_*) &= \min_{u \in U} \max_{x \in X} W(x, u) = \max_{x \in X} W(x, u_*) \geq W(x, u_*), \\ W(x_*, u_*) &= \max_{x \in X} \min_{u \in U} W(x, u) = \min_{u \in U} W(x_*, u) \leq W(x_*, u). \end{aligned}$$

Объединяя эти неравенства, получаем, что имеет место (2.10.6), т.е.  $[x_*, u_*]$  — седловая точка функции  $W(x, u)$ .  $\square$



## Комментарии к главе 2

§ 2.1. Впервые утверждение теоремы 2.1.1 для многочленов было получено в 1629 году выдающимся французским математиком П. де Ферма. Впоследствии, в конце XVII века, создатели дифференциального исчисления И. Ньютон и Г. Лейбниц получили необходимое условие экстремума в форме, близкой к приведенной в теореме 2.1.1, за которой в последние годы закрепилось имя Ферма.

Т.А. Бардадым — автор перевода книги Ж.-Б. Ириарт-Урурти [62] в разделе “Исторические замечания” пишет: “В 1629 г., за 13 лет до рождения Ньютона, Ферма создал свой метод “максимума и минимума”, воспользовавшись для определения значений, в которых функция принимает максимум или минимум, понятием касательной к кривой, которое стало играть фундаментальную роль в дифференциальном исчислении. В зале выдающихся личностей Капитолия в Тулузе (здания городского управления) находится статуя П. де Ферма с надписью: “Основоположнику дифференциального исчисления”, а Ги дю Рютар добавил к ней “Еще раз спасибо!” (в Бомонь-де-Ломань, родном городе Ферма). Итак, любезные студенты, если у вас есть проблемы с дифференцированием, теперь вы знаете, кого благодарить...”

§ 2.2. Понятие локальных решений минимаксных и максиминных задач было введено и исследовано автором в книге [50]. Там же дано обобщение критерия Сильвестра. Теории максимина и минимакса посвящены книги [40–42].

§ 2.3. Теорема 2.3.1 сформулирована в статье Н.И. Грачева и Ю.Г. Евтушенко [36], ее доказательство и приложение к простейшим игровым задачам дано в книге автора [50], где в качестве примера приведены результаты численного решения задачи Р. Айзекса о долихобрахистохроме.

§ 2.4. Ж. Лагранж опубликовал необходимые условия экстремума функций многих переменных в 1797 г. в книге “Теория аналитических функций”. Приведенная здесь формулировка теоремы 2.4.1 отличается от формулировки Лагранжа наличием утверждения о единственности множителей Лагранжа. Дополнительные сведения о теореме Лагранжа можно найти в книге [9].

§ 2.5. Свойства функции чувствительности исследованы в статьях В.Г. Жадана [63], А.И. Голикова и Г.Г. Коткина [29].

§ 2.6. Теорема 2.6.1 была доказана Куном и Таккером в 1951 г. Она вошла во многие статьи и книги под названием “теорема Куна–Таккера”. Однако впоследствии было обнаружено, что утверждение теоремы было дано в 1939 г. У. Карушом в его малоизвестной диссертационной работе, выполненной в университете г. Чикаго. Поэтому в последнее время теорему 2.6.1 некоторые авторы называют теоремой Каруша–Куна–Таккера.

Еще раз обратимся к “Историческим замечаниям”, опубликованным Т.А. Бардадым в книге [62]. Там по этому поводу написано следующее:

“Хотя правило Лагранжа (т.е. условия оптимальности в задачах с ограничениями типа равенств) были известны с начала XIX века, пришлось ожидать до середины XX века, чтобы хоть кто-нибудь заинтересовался (по необходимости) условиями оптимальности в задачах с ограничениями типа неравенств. А.У. Таккер (американский математик канадского происхождения, недавно умерший) и его ученик Г. Кун опубликовали в 1951 г. фундаментальную работу на эту тему. Так случилось, что их опередил У. Каруш, опубликовавший свою работу (которая была, по существу, его квалификационной работой, подготовленной в 1938 г.) в 1948 г. (уважаемые студенты, можете пометчать о подобном!). Сейчас эти три имени ассоциируются в основном с “условиями ККТ”. Прочитируем: “Теорема ККТ является единственным и самым главным оружием в современном экономическом анализе как с теоретической, так и с вычислительной точки зрения” (Д. Гейл, 1967).

Мы, справедливости ради, несколько снизим пафос этих слов, указав, что утвержде-

ние о неотрицательности множителей Лагранжа, соответствующих ограничениям типа неравенств, впервые в 1834 г. заметил выдающийся русский математик М.В. Остроградский в статьях [125, 126], перевод с французского языка на русский опубликован в [72, см. стр. 205–229]. На приоритет М.В. Остроградского указал венгерский математик А. Прекопа в статье [127]. Удивительно, что российские математики не замечают этого. По-видимому, правильно написано в Новом Завете: “Нет пророка в отечестве своем” (см. Евангелие от Матфея, 13.57, от Марка 6, 4, от Иоанна 4, 44). Поэтому если быть справедливым, то к названию теоремы 2.6.1 следует добавить в первую очередь имя М.В. Остроградского. Мы не добавляем ни имени М.В. Остроградского, ни имени У. Каруша, подчеркивая тем самым условность первоначального, широко распространенного названия теоремы 2.6.1 — теорема Куна–Таккера.

Михаил Васильевич Остроградский (1801–1862) после окончания Харьковского университета жил в 1822–1828 гг. в Париже, где посещал лекции О. Коши, Ж. Фурье, П. Лапласа и многих других видных ученых. В 1828 г. он вернулся в Россию и вскоре был избран академиком Петербургской академии наук. Его основные работы относятся к математическому анализу теории вероятностей, вариационному исчислению.

Нестандартное доказательство теоремы Куна–Таккера дано в статье А.А. Третьякова [86], материалы которой использовались при написании этого параграфа.

§ 2.9. Метод отыскания корней уравнения  $f(x) = 0$  был предложен в 1838 г. П.Л. Чебышевым в его студенческой работе [90], которую он написал при переходе с первого на второй курс университета. Итерационные методы, обладающие более высокой, чем квадратичная, скоростью сходимости, изучались в работе [43].

В настоящее время в печати Докладов Академии наук находится статья автора и А.А. Третьякова "Методы  $p$ -го порядка для решения систем нелинейных уравнений". Изложенный в ней подход открывает большие возможности для решения задач НЛП с высокой скоростью сходимости.

В [30–32, 105] даны методы нахождения нормальных решений систем линейных уравнений с помощью безусловной минимизации альтернативных несовместных систем с использованием результатов, приведенных в статье [58]. Метод Ньютона оказался весьма эффективным для решения задач линейного программирования. Укажем цикл работ, выполненных в этом направлении: [28, 30–34, 122].

Редукция оптимизационных задач к решению систем нелинейных уравнений осуществлялась многими авторами, сошлемся на [19, 22, 49, 50, 113].

В данном параграфе для простоты рассмотрен только случай, когда выполнено условие регулярности ограничений. Вырожденные случаи исследовались в работах А.А. Третьякова [61, 84–86], а также его учеников и последователей.

§ 2.10. Свойства минимаксных задач изучались в монографиях [40]– [42].

# Глава 3. Приложение к задачам оптимального управления

## 3.1. Производные по начальным условиям

Приведенные в первой главе формулы для “прямого” и “обратного” дифференцирования не являются чем-то стоящим в стороне от общеизвестных результатов. В данном разделе покажем, что существует тесная связь найденных формул с производными решений систем обыкновенных дифференциальных уравнений по начальным условиям [82]. Эти результаты существенно упрощают анализ задач оптимального управления, дают возможность строить эффективные численные методы их решения с использованием первых и вторых производных. Логика этих рассуждений приводит к новому взгляду на принцип максимума Л.С. Понтрягина.

Пусть некоторый процесс описывается следующей системой обыкновенных дифференциальных уравнений:

$$\frac{dx}{dt} = f(t, x). \quad (3.1.1)$$

Решение (интегральная кривая, траектория) этой системы есть вектор-функция  $x(t)$ ,  $x \in X \subset \mathbb{R}^n$ ,  $t$  — независимая переменная (время),  $t \in T = \{\tau \in \mathbb{R}^1 : T_1 \leq \tau \leq T_2\}$ . Пусть задана некоторая скалярная функция векторного аргумента  $W(x)$ . Ниже будет изучаться зависимость этой функции от начального фазового вектора.

В этом и следующем параграфе для простоты изложения в правую часть дифференциального уравнения (3.1.1) не включено управление. Только в третьем параграфе мы введем вектор управления  $u(t)$ , и тогда  $W(x(T_2))$  окажется функционалом от  $u$  (критерием качества). Такое изложение позволяет дать простую интерпретацию вектора импульсов, без привлечения вариационного исчисления, оставаясь в рамках теории дифференциальных уравнений.

Функции  $f : T \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $W(x) : \mathbb{R}^n \rightarrow \mathbb{R}^1$  непрерывны и непрерывно дифференцируемы по всем аргументам. Решение системы обыкновенных дифференциальных уравнений (3.1.1), принимающее в начальный момент времени  $t = T_1$  значение  $y \in \mathbb{R}^n$ , обозначим через  $x(t, T_1, y)$ . Совокупность  $[T_1, y] \subset G = T \times X$  будет называться *начальными условиями*. В ряде мест, где  $T_1$  и  $y$  фиксированы, для простоты изложения будем решение (3.1.1) обозначать  $x$  или  $x(t)$ , опуская зависимость от начальных условий.

Из определения функции  $x(t, T_1, y)$  следует, что должно выполняться равенство

$$x(s, s, y) = y, \quad T_1 \leq s \leq T_2, \quad (3.1.2)$$

которое называется *условием согласованности*.

Предполагаем, что выполнены условия существования, единственности, непрерывности и дифференцируемости решения (3.1.1) по начальным условиям. Тогда при  $T_1 \leq \tau \leq s \leq t \leq T_2$  выполняется следующее тождество:

$$x(t, \tau, y) = x(t, s, x(s, \tau, y)). \quad (3.1.3)$$

Оно автоматически вытекает из единственности решения (3.1.1), так как каждая из частей тождества является решением (3.1.1) и при  $t = s$  согласно (3.1.2) они совпадают. Строгое обоснование дифференцируемости функции  $x(t, \tau, y)$  дано, например, в учебнике В.В. Степанова [82] “Курс дифференциальных уравнений” (см. глава 7, § 3 “Существование производных по начальным значениям от решений системы”, с. 298–307).

Функция  $x(t, \tau, y)$  удовлетворяет дифференциальному уравнению (3.1.1), т.е.

$$\frac{dx(t, \tau, y)}{dt} = f(t, x(t, \tau, y)). \quad (3.1.4)$$

Воспользуемся следующим результатом, доказанным в [82].

**Теорема 3.1.1.** *Если правые части (3.1.1) имеют непрерывные частные производные по  $x$  в области  $G$ , то решение  $x(t, \tau, y)$ , содержащееся в  $G$ , допускает непрерывные производные по начальным данным.*

Для процесса, описываемого (3.1.1), определим матричную функцию  $n \times n$

$$N(t, \tau, y) = \frac{\partial x^\top(t, \tau, y)}{\partial y}. \quad (3.1.5)$$

Из условия согласованности следует, что  $N(\tau, \tau, y) = I_n$ . Продифференцировав левую и правую части (3.1.4) по  $y$ , учтем, что операции дифференцирования по  $y$  и  $t$  перестановочны, получим

$$\frac{dN(t, \tau, y)}{dt} = N(t, \tau, y) f_x^\top(t, x(t, \tau, y)). \quad (3.1.6)$$

Пусть начальные условия  $[\tau, y], [\bar{\tau}, \bar{y}] \in G$ . Обозначим

$$\Omega(\tau, y) = W(x(T_2, \tau, y)), \quad \bar{\Omega}(\bar{\tau}, \bar{y}) = W(x(T_2, \bar{\tau}, \bar{y})).$$

Изучим свойства введенной функции. Из непрерывности функций  $W(x)$ ,  $x(T_2, \tau, y)$  следует непрерывность сложной функции  $\Omega(\tau, y)$  по обоим аргументам. Будем говорить, что функция  $\Omega(\tau, y)$  дифференцируема по обоим аргументам, если существует скаляр  $q(\tau, y) \in \mathbb{R}^1$  и вектор  $w(\tau, y) \in \mathbb{R}^m$  такие, что

$$\lim_{[\delta, \eta] \rightarrow [0, 0]} \frac{1}{\|\eta\| + \delta} [\Omega(\bar{\tau}, \bar{y}) - \Omega(\tau, y) - \langle w(\tau, y), \eta \rangle - q(\tau, y)\delta] = 0.$$

Здесь  $\bar{\tau} = \tau + \delta$ ,  $\bar{y} = y + \eta$ .

Вектор  $w$  и скаляр  $q$  являются производными функции  $\Omega$ , поэтому можно записать

$$q(\tau, y) = \frac{\partial \Omega(\tau, y)}{\partial \tau}, \quad w(\tau, y) = \frac{\partial \Omega(\tau, y)}{\partial y}.$$

Для вычисления введенных производных воспользуемся правилом дифференцирования сложных функций

$$\frac{\partial \Omega(\tau, y)}{\partial y} = \frac{\partial x^\top(T_2, \tau, y)}{\partial y} W_x(x(T_2, \tau, y)). \quad (3.1.7)$$

Если в (3.1.7) взять  $\tau = T_2$ , то получим

$$x(T_2, T_2, y) = y, \quad w(T_2, y) = \frac{\partial \Omega(T_2, y)}{\partial y} = W_x(y).$$

Из (3.1.3) получим, что при всех  $\tau \leq t \leq T_2$  выполнены условия

$$x(T_2, \tau, y) = x(T_2, t, x(t, \tau, y)),$$

поэтому из определения функции  $\Omega(\tau, y)$  заключаем, что

$$\Omega(\tau, y) = \Omega(t, x(t, \tau, y)) \quad (3.1.8)$$

при всяком  $\tau \leq t \leq T_2$ . Функция  $\Omega(t, x(t, \tau, y))$  принимает постоянное значение при всех  $\tau \leq t \leq T_2$  и является первым интегралом системы (3.1.1), поэтому

$$\frac{d\Omega(t, x(t, \tau, y))}{dt} = q(t, x(t, \tau, y)) + \langle w(t, x(t, \tau, y)), f(t, x(t, \tau, y)) \rangle = 0.$$

Таким образом, если в качестве аргументов функции  $\Omega$  берется пара  $\tau$  и  $y$ , то имеем

$$\frac{\partial \Omega(\tau, y)}{\partial \tau} + \frac{\partial \Omega(\tau, y)}{\partial y^\top} f(\tau, y) = 0. \quad (3.1.9)$$

Следовательно, функция  $\Omega(\tau, y)$  удовлетворяет линейному однородному уравнению с частными производными (3.1.9) с граничным условием

$$\Omega(T_2, y) = W(y).$$

Зафиксируем начальные условия  $\tau, y$ . Введем  $n$ -мерную вектор-функцию

$$a(t) = N(t, \tau, y)p(t),$$

где  $p(t)$  удовлетворяет системе обыкновенных дифференциальных уравнений

$$\frac{dp(t)}{dt} = -f_x^\top(t, x(t, \tau, y))p(t) \quad (3.1.10)$$

с краевым условием на правом конце траектории

$$p(T_2) = W_x(x(T_2, \tau, y)). \quad (3.1.11)$$

Продифференцируем  $a(t)$  по  $t$  и воспользуемся формулами (3.1.6) и (3.1.10); получим

$$\dot{a}(t) = \dot{N}(t, \tau, y)p(t) + N(t, \tau, y)\dot{p}(t) = 0_n.$$

Из постоянства  $a(t)$  с учетом (3.1.7) следует

$$\begin{aligned} a(\tau) &= N(\tau, \tau, y)p(\tau) = p(\tau) = a(T_2) = \\ &= N(T_2, \tau, y)p(T_2) = \frac{\partial x^\top(T_2, \tau, y)}{\partial y} W_x(x(T_2, \tau, y)) = \frac{\partial \Omega(\tau, y)}{\partial y}. \end{aligned}$$

Используя (3.1.9), заключаем, что

$$\frac{\partial \Omega(\tau, y)}{\partial y} = p(\tau), \quad \frac{\partial \Omega(\tau, y)}{\partial \tau} = -f^\top(\tau, y)p(\tau). \quad (3.1.12)$$

Проинтегрировав системы (3.1.1) и (3.1.6) от  $t = T_1$  до  $t = T_2$ , получим функции  $x(t, T_1, y)$ ,  $N(t, T_1, y)$ . Интегрируя (3.1.10) в обратном направлении с учетом условия (3.1.11), получим функцию  $p(t)$ . Градиент функции  $\Omega(T_1, y) = W(x(T_2, T_1, y))$  может быть найден двумя способами:

$$\frac{\partial \Omega(T_1, y)}{\partial y} = N(T_2, T_1, y)W_x(x(T_2, T_1, y)), \quad \frac{\partial \Omega(T_1, y)}{\partial y} = p(T_1). \quad (3.1.13)$$

Оба выражения дают одинаковый результат, но в первом случае интегрируется матричная система (3.1.6), которая состоит из  $n^2$  скалярных дифференциальных уравнений. Во втором случае интегрируется векторная система (3.1.10), которая состоит только из  $n$  скалярных

дифференциальных уравнений. Таким образом, во втором случае для вычисления градиента требуется в  $n$  раз меньше вычислений, чем в первом.

Формулы (3.1.13) обычно применяются для численного решения краевых задач с помощью метода стрельб, основанного на редукции краевой задачи к решению последовательности задач Коши.

Каждую пару  $[\tau, y]$ , где  $T_1 \leq \tau \leq T_2$ ,  $y = x(\tau, T_1, x_1)$ , можно взять в качестве начальных условий для системы (3.1.1) и воспользоваться результатами этого параграфа; тогда в момент  $\tau$  фазовый вектор  $y$  и производная сложной функции  $W(x(T_2, \tau, y))$  по  $y$  вычисляются по формулам

$$y = x(\tau, T_1, x_1), \quad \frac{dW(x(T_2, \tau, y))}{dy} = p(\tau), \quad (3.1.14)$$

где  $p$  — решение системы (3.1.10) с краевым условием (3.1.11). Вектор  $p(t)$  обычно называют *импульсом* или *сопряженной переменной*.

Во многих публикациях по теории оптимального управления вектор-функция  $p(\tau)$  вводится формально, ее смысл остается необъясненным. Использование формул (3.1.14) позволяет получить простую интерпретацию вектора импульсов, давая наглядное толкование вектор-функциям  $p$  и  $y$ , а именно:

$y$  есть точка, принадлежащая в момент  $t = \tau$  фазовой траектории  $x(t, T_1, x_1)$ , стартующая из точки  $x_1$  в момент  $T_1$ , или, что то же самое в силу (3.1.2), (3.1.3), стартующая из точки  $y$  в момент  $\tau$ ;

$p(\tau)$  есть производная (градиент) сложной функции  $W(x(T_2, \tau, y))$  по фазовому вектору  $y$ , определяемая из второго условия в (3.1.14) (см. также формулу (1.3.15)).

При  $T_1 \leq \tau \leq t \leq T_2$  рассмотрим траекторию  $x(t, \tau, y + d)$  системы (3.1.1), исходящую в момент  $\tau$  из точки  $y + d$ , где  $y = x(\tau, T_1, x_1)$ ,  $d \in \mathbb{R}^n$ . Используя дифференцируемость  $W$ , из (3.1.14) получим, что импульс  $p(\tau)$  удовлетворяет условию

$$W(x(T_2, \tau, y + d)) - W(x(T_2, \tau, y)) = \langle p(\tau), d \rangle + O(\|d\|^2). \quad (3.1.15)$$

Другими словами, отображение  $W(x(\cdot, \cdot, y)) : \mathbb{R}^n \rightarrow \mathbb{R}^1$  дифференцируемо по Фреше в точке  $y$ , и его производной (градиентом) является вектор  $p(\tau)$ .

Полученная формула оказалась чрезвычайно полезной при отладке программ, интегрирующих систему (3.1.10). Дело в том, что при дифференцировании громоздких выражений, входящих в функцию  $f$ , и при их программной реализации иногда возникают ошибки в вычислении правых частей системы (3.1.10). В этом случае можно искусственно в момент  $\tau$  прервать интегрирование системы (3.1.1) и продолжить интегрирование от  $t = \tau$  до  $t = T_2$  из измененной точки  $y + d$ . В результате при  $t = T_2$  получится некоторое новое значение  $W(x(T_2, \tau, y + d))$ . Сравнивая его со значением, вычисленным по формуле (3.1.15), проверяем правильность вычисления вектора  $p(\tau)$  по формулам (3.1.10), (3.1.11).

Во многих задачах оптимизации скалярная функция  $W(x, \xi)$  и правые части системы (3.1.1) зависят от некоторого фиксированного вектора параметров  $\xi \in \mathbb{R}^s$ . По этому параметру производится оптимизация целевой функции  $W$ . Для этого важно знать производную  $W$  по  $\xi$ . Вместо (3.1.1) рассматривается следующая система дифференциальных уравнений:

$$\frac{dx}{dt} = f(t, x, \xi). \quad (3.1.16)$$

Для этой системы ставится задача Коши о нахождении решения  $x(t, \xi)$ , удовлетворяющего на левом конце траектории начальным условиям

$$x(T_1, \xi) = x_1, \quad x_\xi(T_1, \xi) = 0_n. \quad (3.1.17)$$

Предположим, что в рассматриваемой области функции  $f$  и  $W$  непрерывны вместе со своими частными производными  $f_x, f_\xi, W_x, W_\xi$ . Дифференцируемая функция  $p(t)$  удовлетворяет системе, аналогичной (3.1.10):

$$\frac{dp(t)}{dt} = -f_x^\top(t, x, \xi)p(t). \quad (3.1.18)$$

Введем функцию

$$\widetilde{W}(\xi) = W(x(T_2, \xi), \xi) + \int_{T_1}^{T_2} \left[ f^\top(t, x(t, \xi), \xi) - \frac{dx^\top(t, \xi)}{dt} \right] p(t) dt. \quad (3.1.19)$$

Обозначим  $z(t, \xi) = \partial x(t, \xi) / \partial \xi$ . Продифференцируем (3.1.19) по  $\xi$ , считая, что операции дифференцирования по  $t$  и  $\xi$  перестановочны, получим

$$\frac{d\widetilde{W}(\xi)}{d\xi} = \varphi(\xi) + \int_{T_1}^{T_2} \left[ f_\xi^\top(t, x(t, \xi), \xi) + z^\top(t, \xi) f_x^\top(t, x(t, \xi), \xi) - \frac{dz^\top(t, \xi)}{dt} \right] p(t) dt, \quad (3.1.20)$$

где  $\varphi(\xi) = W_\xi(x(T_2, \xi), \xi) + z^\top(T_2, \xi) W_x(x(T_2, \xi), \xi) \in \mathbb{R}^s$ .

Продифференцируем по  $t$  скалярное произведение  $z^\top(t, \xi)p(t)$ , имеем

$$\frac{d}{dt}(z^\top(t, \xi)p(t)) = \frac{dz^\top(t, \xi)}{dt}p(t) + z^\top(t, \xi)\frac{dp(t)}{dt}.$$

Проинтегрируем это выражение по частям от  $t = T_1$  до  $t = T_2$ , примем во внимание (3.1.17), находим

$$- \int_{T_1}^{T_2} \frac{dz^\top(t, \xi)}{dt} p(t) dt = \int_{T_1}^{T_2} z^\top(t, \xi) \frac{dp(t)}{dt} dt - z^\top(T_2, \xi) p(T_2).$$

В это выражение подставим (3.1.20), учтем (3.1.18), получим

$$\begin{aligned} \frac{d\widetilde{W}(\xi)}{d\xi} &= \int_{T_1}^{T_2} f_\xi^\top(t, x(t, \xi), \xi) p(t) dt + W_\xi(x(T_2, \xi), \xi) + \\ &+ z^\top(T_2, \xi) [W_x(x(T_2, \xi), \xi) - p(T_2)]. \end{aligned}$$

На правом конце траектории наложим на вектор  $p$  условие

$$p(T_2) = W_x(x(T_2, \xi), \xi). \quad (3.1.21)$$

Тогда производная функции  $\widetilde{W}(\xi)$  по  $\xi$  вычисляется по формуле

$$\frac{d\widetilde{W}(\xi)}{d\xi} = W_\xi(x(T_2), \xi) + \frac{d}{d\xi} \int_{T_1}^{T_2} f^\top(t, x(t, \xi), \xi) p(t) dt. \quad (3.1.22)$$

Приведем два примера использования оптимизации по параметрам.

**Пример 3.1.1.** Простейшая оптимизационная задача для системы (3.1.16) ставится следующим образом: для фиксированных  $T_1, T_2, x_1$  определить вектор параметров  $\xi_*$  так, чтобы целевая функция  $W(x(T_2, \xi), \xi)$  принимала наименьшее возможное значение при  $\xi = \xi_*$ .

Если на выбор вектора  $\xi$  не наложено никаких ограничений, то можно воспользоваться теоремой Ферма и из (3.1.22) получим необходимое условие минимума

$$W_\xi(x(T_2, \xi_*), \xi_*) + \int_{T_1}^{T_2} f_\xi^\top(t, x(t, \xi_*), \xi) p(t) dt = 0_s, \quad (3.1.23)$$

где  $p(t)$  – решение системы (3.1.18) с конечным условием (3.1.21), вычисленном при  $\xi = \xi_*$ .

Из необходимого условия (3.1.23) определяется искомый оптимальный вектор параметров  $\xi_*$ . Если на выбор вектора  $\xi$  наложены какие-либо ограничения, то применяется обычная техника нелинейного программирования, вводятся множители Лагранжа, используется либо теорема Лагранжа, либо Куна-Таккера, и вместо (3.1.23) записывается условие минимума с учетом ограничений. ■

**Пример 3.1.2.** Пусть  $x(t, \tau, \xi)$  – решение системы (3.1.4), удовлетворяющее условиям (3.1.2), (3.1.3). Требуется для системы (3.1.4) найти начальные условия  $\tau$  и  $y$  так, чтобы целевая функция  $W(x(T_2, \theta, \xi))$  в конце движения принимала наименьшее возможное значение. В этом случае начальный момент и начальный фазовый вектор будут рассматриваться как управляющие параметры  $\theta$  и  $\xi$ . Краевые условия для систем (3.1.4) и (3.1.10) имеют вид

$$x(\theta, \theta, \xi) = \xi, \quad p(T_2) = W_x(x(T_2, \theta, \xi)).$$

Из (3.1.12) следуют искомые производные

$$\frac{dW(x(\theta, \theta, \xi))}{d\xi} = p(T_1), \quad \frac{dW(\theta, \theta, \xi)}{d\theta} = -f^\top(\theta, \xi)p(\theta).$$

С помощью методов нелинейного программирования можно учесть, если это потребуется, дополнительные ограничения на выбор  $\xi$  и  $\theta$ . ■

Все приведенные в этом параграфе формулы можно вывести из результатов, полученных в первой главе. Действительно, разобьем интервал движения на  $k-1$  равных частей, положим  $t_1 = T_1, t_{i+1} = t_i + h, h = (T_2 - T_1)/(k-1), 1 \leq i \leq k-1, k \geq 2$ . Проинтегрировав (3.1.1) по схеме Эйлера, получим следующую дискретную аппроксимацию дифференциального уравнения (3.1.1):

$$x_{i+1} = x_i + hf(t_i, x_i), \quad 1 \leq i \leq k-1. \quad (3.1.24)$$

Формулы (3.1.24) – частный случай соотношений (1.3.2). Следовательно, все результаты, приведенные в § 1.3, могут быть использованы здесь, давая возможность вычислять первые и вторые производные для дискретных аппроксимаций задач ОУ.

Обозначим  $N_j = \partial x_j^\top / \partial u, u \in \mathbb{R}^n, 1 \leq j \leq k$ . Прямое дифференцирование (3.1.24) дает последовательность, аналогичную (1.3.4):

$$N_1 = I_n, \quad N_{j+1} = N_j + hN_j f_{x_j}(t_j, x_j), \quad 1 \leq j \leq k-1. \quad (3.1.25)$$

Введем вспомогательную функцию  $E$ , применим формулы (1.3.8)–(1.3.10), получим уравнения в конечных разностях для определения векторов  $p_i \in \mathbb{R}^n$ :

$$E(t, x, u, p) = W(x_k) + u^\top p_1 + \sum_{i=1}^{k-1} [x_i + hf(t_i, x_i)]^\top p_{i+1},$$



$$p_i = p_{i+1} + hf_{x_i}^\top(t_i, x_i)p_{i+1}, \quad p_k = W_x(x_k), \quad 1 \leq i \leq k-1. \quad (3.1.26)$$

Определим сложную функцию  $\Omega$  от начальных условий, положив  $\Omega(t_1, x_1) = W(x_k)$ . Эта функция постоянна на траектории (3.1.24), т.е.

$$\Omega(t_1, x_1) = \Omega(t_2, x_2) = \dots = \Omega(t_k, x_k) = W(x_k).$$

Искомая производная функции  $\Omega(T_1, x_1)$  по начальному условию  $x_1$  может быть вычислена по любой из следующих эквивалентных формул:

$$p_1 = E_u = \frac{\partial \Omega(T_1, x_1)}{\partial x_1} = N_k W_{x_k}(x_k).$$

В рекуррентных соотношениях (3.1.25) и (3.1.26) перейдем к пределу при  $h \rightarrow 0$  и  $k \rightarrow \infty$ . Легко видеть, что соотношения (3.1.25) переходят в дифференциальное уравнение (3.1.6), а соотношения (3.1.26) — в (3.1.10).

Всякую пару  $[t_i, x_i]$  можно взять в качестве начальных условий в (3.1.26); тогда получим

$$p_i = \frac{\partial \Omega(t_i, x_i)}{\partial x_i} = \frac{d\widehat{W}(x_k)}{dx_i}.$$

Здесь введена полная производная сложной функции  $\widehat{W}$  по  $x_i$ . Функциональная зависимость  $\widehat{W}(x_i)$  определяется следующим образом: из последовательности (3.1.26), взяв в качестве начальных условий  $[t_i, x_i]$ , определим  $x_k$  и положим  $\widehat{W}(x_i) = W(x_k)$ . Второй вариант задания этой функции приведен в § 1.3, где эта зависимость определялась как функция  $W^i$ .

Учитывая (3.1.26), получим

$$\begin{aligned} p_i &= \frac{d\widehat{W}(x_i)}{dx_i} = \frac{\partial x_{i+1}^\top}{\partial x_i} \frac{d\widehat{W}(x_{i+1})}{dx_{i+1}} = \\ &= (I_m + hf_{x_i}^\top(t_i, x_i)) \frac{d\widehat{W}(x_{i+1})}{dx_{i+1}} = p_{i+1} + hf_{x_i}^\top(t_i, x_i)p_{i+1}. \end{aligned} \quad (3.1.27)$$

Таким образом, несколько видоизменив рассуждения, мы снова пришли к формуле (3.1.26), являющейся дискретным аналогом (3.1.10). Такой же прием позволяет просто сосчитать матрицы вторых производных. Введем  $m \times m$  матрицу

$$\Delta \Omega_i = \frac{d^2 \widehat{W}(x_i)}{dx_i dx_i^\top}, \quad 1 \leq i < k.$$

Дифференцируя (3.1.27) по  $x_i$ , получим

$$\begin{aligned} \Delta \Omega_i &= (I_m + hf_{x_i}^\top(t_i, x_i)) \Delta \Omega_{i+1} (I_m + hf_{x_i}^\top(t_i, x_i)) + \\ &+ h \sum_{s=1}^n f_{x_i x_i^\top}^s(t_i, x_i) p_{i+1}^s, \quad \Delta \Omega_k = \frac{d^2 W(x_k)}{dx_k dx_k^\top}. \end{aligned}$$

Если в этом соотношении перейти к пределу при  $h \rightarrow 0$ ,  $k \rightarrow \infty$ , то приходим к следующему дифференциальному уравнению:

$$\frac{d\Delta \Omega(t)}{dt} = - \left[ f_{x_i}^\top(t, x(t, \tau, y)) \Delta \Omega(t) + \Delta \Omega(t) f_{x_i}^\top(t, x(t, \tau, y)) + \sum_{j=1}^m f_{x_i x_i^\top}^j(t, x(t, \tau, y)) p^j(t) \right].$$

Упомянем сравнительно редкий случай, когда прямое дифференцирование оказывается более экономным. Этот случай возникает, когда ищутся градиенты  $s$  функций  $\Omega^i(T_1, y) = W^i(x(T_2, T_1, y))$ , где  $1 \leq i \leq s$ . Определим матрицу  $N(T_2, T_1, y)$ ; после этого все градиенты находим следующим образом:

$$\frac{\partial \Omega^i(T_1, y)}{\partial y} = N(T_2, T_1, y) \frac{\partial W^i(x(T_2, T_1, y))}{\partial x(T_2, T_1, y)}.$$

Следовательно, если  $s > n$ , то прямой метод более эффективен, чем обратный.

### 3.2. Системы с разрывными правыми частями

Системы с разрывными правыми частями рассматривались в [3, 18, 26, 38, 87]. Пусть в  $(n + 1)$ -мерном пространстве  $G = \tilde{T} \times X$  задана гладкая поверхность

$$\mathcal{G}(t, x) = 0, \quad (3.2.1)$$

называемая в дальнейшем поверхностью переключений. Правая часть системы (3.1.1) терпит разрыв первого рода при пересечении траекториями системы (3.1.1) этой поверхности, причем

$$f(t, x) = \begin{cases} f_-(t, x), & \text{если } \mathcal{G}(t, x) \leq 0, \\ f_+(t, x), & \text{если } \mathcal{G}(t, x) > 0, \end{cases}$$

где  $f_-$  и  $f_+$  суть непрерывные вектор-функции своих аргументов, обладающие непрерывными частными производными по  $x$ .

Пусть  $\mathcal{G}(T_1, x_1) \leq 0$ , и траектория  $x(t, T_1, x_1)$  системы (3.1.1) попадает в первый раз на поверхность  $\mathcal{G}(t, x) = 0$  при  $t = \tilde{t}$ . Далее при  $\tilde{t} < t < T_2$  вдоль траектории, обозначаемой через  $x_+(t, T_1, x_1)$ , выполнено условие  $\mathcal{G}(t, x_+(t, T_1, x_1)) > 0$ . Траектория непрерывна, поэтому  $x_-(\tilde{t}, T_1, x_1) = x_+(\tilde{t}, T_1, x_1)$  и, как и в предыдущем параграфе,  $\Omega(T_1, x_1) = W(x_+(T_2, T_1, x_1))$ . Справа от поверхности переключения функция  $W(x(T_2, t, x_+(t, T_1, x_1)))$  дифференцируема по  $x_+$ , и при переходе через поверхность переключения ее производная меняется скачком. Предполагаем, что траектория “прошивает” поверхность без касания. В этом случае

$$\begin{aligned} \mathcal{G}_t(\tilde{t}, \tilde{x}) + \langle \mathcal{G}_x(\tilde{t}, \tilde{x}), f_-(\tilde{t}, \tilde{x}) \rangle &\neq 0, \\ \mathcal{G}_t(\tilde{t}, \tilde{x}) + \langle \mathcal{G}_x(\tilde{t}, \tilde{x}), f_+(\tilde{t}, \tilde{x}) \rangle &\neq 0. \end{aligned}$$

Наряду с траекторией  $x(t, T_1, x_1)$  рассмотрим близкую траекторию  $x(t, T_1, x_1 + \Delta)$ . Обозначим  $\delta t = t - \tilde{t}$ ,  $\delta x(t) = x(t, T_1, x_1 + \Delta) - x(t, T_1, x_1)$ . Вблизи точки  $[\tilde{t}, \tilde{x}]$  уравнение поверхности (3.2.1) можно записать в виде

$$\mathcal{G}_t(\tilde{t}, \tilde{x})\delta t + \langle \mathcal{G}_x(\tilde{t}, \tilde{x}), \delta x \rangle + O(|\delta t|^2 + \|\delta x(t)\|^2) = 0. \quad (3.2.2)$$

Траекторию до и после пересечения поверхности переключения можно соответственно представить в виде

$$\delta x_-(\tilde{t} + \delta t) = \delta x(\tilde{t}) + f_-(\tilde{t}, \tilde{x})\delta t + O(|\delta t|^2), \quad (3.2.3)$$

$$\delta x_+(\tilde{t} + \delta t) = \delta x(\tilde{t}) + f_+(\tilde{t}, \tilde{x})\delta t + O(|\delta t|^2). \quad (3.2.4)$$

Подставляя (3.2.3) в (3.2.2), находим

$$\delta t = \frac{-\langle \mathcal{G}_x(\tilde{t}, \tilde{x}), \delta x(t) \rangle}{\mathcal{G}_t(\tilde{t}, \tilde{x}) + \langle \mathcal{G}_x(\tilde{t}, \tilde{x}), f_-(\tilde{t}, \tilde{x}) \rangle}.$$

Следовательно, из (3.2.4) получаем

$$\delta x_+(\tilde{t} + \delta t) = \delta x_-(\tilde{t} + \delta t) + \frac{[f_+(\tilde{t}, \tilde{x}) - f_-(\tilde{t}, \tilde{x})] \langle \mathcal{G}_x(\tilde{t}, \tilde{x}), \delta x(t) \rangle}{\mathcal{G}_t(\tilde{t}, \tilde{x}) + \langle \mathcal{G}_x(\tilde{t}, \tilde{x}), f_-(\tilde{t}, \tilde{x}) \rangle}.$$

Отсюда имеем

$$\frac{\partial x_+^i(\tilde{t}, \tilde{x})}{\partial x_-^j(\tilde{t}, \tilde{x})} = \delta_{ij} + \frac{[f_+^i(\tilde{t}, \tilde{x}) - f_-^i(\tilde{t}, \tilde{x})] \partial \mathcal{G}(\tilde{t}, \tilde{x}) / \partial x^j}{\mathcal{G}_t(\tilde{t}, \tilde{x}) + \langle \mathcal{G}_x(\tilde{t}, \tilde{x}), f_-(\tilde{t}, \tilde{x}) \rangle}. \quad (3.2.5)$$

Предполагая, что на участке траектории при  $\tilde{t} \leq t \leq T_2$  выполнено условие

$$\mathcal{G}(t, x(t, T_1, x_1)) > 0,$$

воспользуемся формулой (3.1.14), обозначив

$$p_+(\tilde{t}) = \frac{dW(x(T_2, \tilde{t}, \tilde{x}))}{d\tilde{x}}.$$

Слева от поверхности переключения правых частей согласно (3.2.5) имеем

$$p_-(\tilde{t}) = \frac{\partial x_+^\top(\tilde{t}, \tilde{x})}{\partial x_-(\tilde{t}, \tilde{x})} p_+(\tilde{t}) = p_+(\tilde{t}) + \frac{\langle H_+(\tilde{t}, \tilde{x}) - H_-(\tilde{t}, \tilde{x}), \mathcal{G}_x(\tilde{t}, \tilde{x}) \rangle}{\mathcal{G}_t(\tilde{t}, \tilde{x}) + \langle \mathcal{G}_x(\tilde{t}, \tilde{x}), f_-(\tilde{t}, \tilde{x}) \rangle}, \quad (3.2.6)$$

где  $H_+(\tilde{t}, \tilde{x}) = \langle f_+(\tilde{t}, \tilde{x}), p_+(\tilde{t}) \rangle$ ,  $H_-(\tilde{t}, \tilde{x}) = \langle f_-(\tilde{t}, \tilde{x}), p_+(\tilde{t}) \rangle$ . Таким образом, вектор-функция  $p(t)$  терпит разрыв первого рода при пересечении поверхности переключения. Формула (3.2.6) дает условие скачка вектор-функции  $p(t)$ .

Определение градиентов в задачах с разрывными правыми частями дано в работах [3, 38].

### 3.3. Принцип максимума Л.С. Понтрягина

Пусть управляемый процесс описывается системой обыкновенных дифференциальных уравнений

$$\dot{x} = f(t, x, u), \quad T_1 \leq t \leq T_2, \quad (3.3.1)$$

где фазовый вектор  $x \in \mathbb{R}^n$ , вектор управления  $u \in \mathbb{R}^r$ .

Считаем, что начальный фазовый вектор  $x_1$ , начальный и конечный моменты времени  $T_1$  и  $T_2$  фиксированны. В этом параграфе решение (3.3.1) будем обозначать  $x(t)$ , поэтому для рассматриваемой задачи с закрепленным начальным условием (левым концом траектории) имеем

$$x(T_1) = x_1. \quad (3.3.2)$$

Интервал движения обозначим  $\tilde{T} = \{t : T_1 \leq t \leq T_2\}$ . Определенное на  $\tilde{T}$  управление  $u(t)$  называем *допустимым*, если  $u(t) \in U \in \mathbb{R}^r$  при всяком  $t$  из  $\tilde{T}$ . Считаем, что множество  $U$  компактное. Траектории системы (3.3.1) будут рассмотрены в предположении, что управление является кусочно-непрерывной функцией  $t$ , т.е.  $u(t)$  непрерывна во всех точках  $t \in \tilde{T}$  за исключением, быть может, конечного числа точек  $t_i$ , в которых функция  $u(t)$  может терпеть разрывы типа скачка. При этом существуют конечные пределы

$$\lim_{t \rightarrow t_i - 0} u(t) = u(t_i - 0), \quad \lim_{t \rightarrow t_i + 0} u(t) = u(t_i + 0),$$

но, вообще говоря,  $u(t_i - 0) \neq u(t_i + 0)$ . Управления  $u(t)$  в изолированных точках разрыва не влияют на решение системы (3.3.1) и поэтому могут быть доопределены произвольным

образом. Будем считать для определенности, что  $u(t_i) = u(t_i - 0) = \lim_{t \rightarrow t_i - 0} u(t)$  при всех  $t \in \tilde{T}$  и, в частности,  $u(T_2) = u(T_2 - 0)$ .

При любом  $t \in \tilde{T}$  во всех точках непрерывности допустимого управления  $u(t)$  выполнено равенство  $\dot{x}(t) = f(t, x(t), u(t))$ . Функция  $\dot{x}(t)$  является кусочно-дифференцируемой на отрезке  $\tilde{T}$ .

Чтобы не усложнять изложение, сделать его наиболее простым и исключить второстепенные детали, рассмотрим задачу Майера об отыскании на фиксированном интервале движения  $\tilde{T}$  допустимого кусочно-непрерывного управления  $u(t) \in U \subset \mathbb{R}^r$ , доставляющего минимум терминальному функционалу

$$J = \min_{u \in U} W(x(T_2)), \quad (3.3.3)$$

где  $W(x)$  – заданная функция.

Поставленную таким образом задачу называют *задачей с закрепленным левым концом, со свободным правым концом и с фиксированным интервалом движения*. Предполагается, что поставленная задача разрешима в классе кусочно-непрерывных функций  $u(t)$ . Найденное управление  $u_*(t)$  будем называть *оптимальным*.

После анализа свойств задачи (3.3.3) все найденные результаты легко переносятся на более общий класс задач, например на задачи с интегральными функционалами, со свободным (не фиксированным) временем движения, на задачи быстрогодействия, с подвижными концами, оптимизацию по вектору параметров и т.д. Приемы таких обобщений хорошо известны и осуществляются, главным образом, введением дополнительных фазовых координат.

Введем вектор  $p \in \mathbb{R}^n$ , удовлетворяющий системе обыкновенных дифференциальных уравнений (сопряженной системе)

$$\dot{p} = -f_x^\top(t, x, u)p \quad (3.3.4)$$

с граничным условием

$$p(T_2) = W_x(x(T_2)), \quad (3.3.5)$$

где  $p(t)$  – импульс (сопряженный вектор). Определим функцию Понтрягина

$$H(t, x, u, p) = \langle f(t, x, u), p \rangle.$$

В случае, когда  $f$  не зависит явно от  $t$  (система (3.1.1) автономная), будем писать  $H(x, u, p) = \langle f(x, u), p \rangle$ . В подробной форме записи приведенные формулы имеют вид

$$\begin{aligned} \frac{d}{dt} p^i(t) &= - \sum_{j=1}^n \frac{\partial f^j(t, x, u)}{\partial x^i} p^j(t), \quad 1 \leq i \leq n, \\ H(t, x, u, p) &= \sum_{j=1}^n f^j(t, x, u) p^j. \end{aligned}$$

Подставим оптимальное управление в систему (3.3.1) и проинтегрируем ее с начальным условием (3.3.2) от  $t = T_1$  до  $t = T_2$  (слева направо), получим соответствующую траекторию  $x_*(t)$  и оптимальное значение целевой функции  $W(x_*(T_2))$  в конце траектории (движения). Проинтегрировав (3.3.4) от  $t = T_2$  до  $t = T_1$  (справа налево) с краевым условием (3.3.5), находим функцию  $p_*(t)$ . Существует конечное число точек  $t$  разрыва функции  $u_*(t)$ , в которых значение  $u(t)$  определяется как предел слева.

Исключим из множества  $\tilde{T}$  точки, в которых управление  $u_*(t)$  имеет разрывы, обозначим полученное множество  $\tilde{T}'$ .

Имеет место

**Теорема 3.3.1 (принцип максимума Понтрягина).** Пусть вектор-функция  $f(t, x, u)$  непрерывна по совокупности аргументов и непрерывно дифференцируема по  $x$  на  $\tilde{T} \times \mathbb{R}^n \times U$ , функция  $W(x)$  непрерывно дифференцируема по  $x$  на  $\mathbb{R}^m$ . Пусть  $u_*(t)$  — оптимальное управление в задаче (3.3.3),  $x_*(t)$ ,  $p_*(t)$  — соответствующие траектория и импульс, удовлетворяющие дифференциальным уравнениям (3.3.1) и (3.3.4) и краевым условиям (3.3.2) и (3.3.5). Тогда при всех  $t \in \tilde{T}'$  выполнены следующие условия:

1) в точке  $u_*(t)$  достигается минимум функции  $H(t, x_*(t), w, p_*(t))$  по  $w \in U$ , т.е.

$$u_*(t) \in \text{Arg} \min_{w \in U} H(t, x_*(t), w, p_*(t)); \quad (3.3.6)$$

2) если система (3.3.1) автономная, то функция  $H(x_*(t), u_*(t), p_*(t))$  принимает постоянное значение при всех  $t \in \tilde{T}$ .

**Доказательство.** Условия теоремы обеспечивают единственность решения задачи Коши (3.3.1), (3.3.2) для любого допустимого управления. Возьмем произвольное  $\tau$ , лежащее внутри  $\tilde{T}'$ ; пусть оптимальное управление  $u_*(t)$  непрерывно в точке  $\tau$ . Введем новое, “проварьированное” управление

$$v(t) = \begin{cases} \bar{v} & \text{при } \tau - \varepsilon \leq t \leq \tau, \\ u_*(t) & \text{при } t \in [T_1, T_2] \setminus (\tau - \varepsilon, \tau), \end{cases} \quad (3.3.7)$$

где  $\bar{v} \in U$ , малая величина  $\varepsilon > 0$ .

Пусть  $x(t)$  — решение (3.3.1) с управлением  $v(t)$  и начальным условием (3.3.2). Тогда  $x(t) = x_*(t)$  при  $T_1 \leq t \leq \tau - \varepsilon$ . Обозначим  $\delta x(t) = x(t) - x_*(t)$ . Из (3.3.1) следует, что при  $\tau - \varepsilon \leq t \leq \tau$  вариация траектории удовлетворяет следующему дифференциальному уравнению:

$$\delta \dot{x}(t) = f(t, x_*(t) + \delta x(t), \bar{v}) - f(t, x_*(t), u_*(t)) \quad (3.3.8)$$

с начальным условием  $\delta x(\tau - \varepsilon) = 0$ . Проинтегрируем уравнение (3.3.8) на интервале  $\tau - \varepsilon \leq t \leq \tau$ , получим

$$\delta x(\tau) = \varepsilon [f(\tau - \varepsilon, x_*(\tau - \varepsilon), \bar{v}) - f(\tau - \varepsilon, x_*(\tau - \varepsilon), u_*(\tau - \varepsilon))] + O(\varepsilon^2).$$

Отсюда с погрешностью порядка  $\varepsilon^2$  можно записать

$$\delta x(\tau) = \varepsilon [f(\tau, x_*(\tau), \bar{v}) - f(\tau, x_*(\tau), u_*(\tau))] + O(\varepsilon^2).$$

Умножив скалярно левую и правую части этого равенства на  $p_*(\tau)$ , получим

$$\langle p_*(\tau), \delta x(\tau) \rangle = \varepsilon g(\tau) + O(\varepsilon^2), \quad (3.3.9)$$

где  $g(\tau) = H(\tau, x_*(\tau), \bar{v}, p_*(\tau)) - H(\tau, x_*(\tau), u_*(\tau), p_*(\tau))$ .

Управления  $v(t)$  и  $u_*(t)$  совпадают при  $t$ , принадлежащих интервалам  $T_1 \leq t \leq \tau - \varepsilon$  и при  $\tau \leq t \leq T_2$ . На первом интервале совпадают траектории  $x_*(t)$  и  $x(t)$ , на втором интервале траектории отличаются, так как у них взяты разные фазовые векторы при  $t = \tau$ . В первом случае на оптимальной траектории взято  $x_*(\tau)$ , во втором случае на проварьированной траектории берется  $x_*(\tau) + \delta x(\tau)$ . Согласно (3.1.15), (3.3.9) имеем

$$W(x(T_2)) - W(x_*(T_2)) = \langle p_*(\tau), \delta x(\tau) \rangle + O(\|\delta x\|^2) = \varepsilon g(\tau) + O(\varepsilon^2). \quad (3.3.10)$$

Управление  $u_*(t)$  допустимое и оптимальное, управление  $v(t)$  тоже допустимое, но не оптимальное, поэтому целевая функция, вычисленная в конце оптимальной траектории  $W(x_*(T_2))$ , не превосходит значения целевой функции в конце проварьированной траектории и

$$0 \leq W(x(T_2)) - W(x_*(T_2)) = \varepsilon [g(\tau) + \gamma(\varepsilon)] \geq 0, \quad (3.3.11)$$

где функция  $\gamma(\varepsilon)$  такова, что  $\lim_{\varepsilon \rightarrow 0} \gamma(\varepsilon) = 0$ .

Из (3.3.11) следует, что  $g(\tau) \geq 0$ . Действительно, если, от противного,  $g(\tau) = -c < 0$ , то из свойства функции  $\gamma(\tau)$  заключаем, что существует такое  $\bar{\varepsilon}$ , что при всех  $0 < \varepsilon < \bar{\varepsilon}$  выполнено неравенство  $\gamma(\varepsilon) < c/2$ . Но при этих же значениях  $\varepsilon$  выражение, стоящее в квадратных скобках в (3.3.11), отрицательно, меньше чем  $-c/2$ , что противоречит левому неравенству в (3.3.11). Следовательно, для всех достаточно малых  $\varepsilon$  выполнено неравенство  $g(\tau) \geq 0$ , откуда следует утверждение (3.3.6). Доказательство первого утверждения теоремы, таким образом, завершено.

Пусть система (3.3.1) автономна. Тогда

$$\dot{x} = f(x, u), \quad \dot{p} = -f^\top_x(x, u)p, \quad H(x, u, p) = \langle f(x, u), p \rangle. \quad (3.3.12)$$

Пусть  $\tau, t \in [T_1, T_2]$ , оптимальному управлению  $u_*(t)$  соответствуют в силу (3.3.12), (3.3.2), (3.3.5) функции  $x_*(t)$  и  $p_*(t)$ . Обозначим

$$h(t) = H(x_*(t), u_*(t), p_*(t)).$$

Функции  $x_*(t)$  и  $p_*(t)$  как решения системы (3.3.12) являются непрерывными функциями  $t$ . Функция  $h(t)$  из-за кусочной непрерывности  $u_*(t)$  будет, вообще говоря, уже кусочно-непрерывной. Тем не менее покажем, что определяемая на оптимальной траектории функция Понтрягина оказывается непрерывной функцией  $t$  всюду на  $[T_1, T_2]$  и, в частности, в точках разрыва оптимального управления  $u_*(t)$ .

Из (3.3.6) следует, что для любых  $\tau$  и  $t$  из интервала  $[T_1, T_2]$  имеют место неравенства

$$\begin{aligned} H(x_*(t), u_*(\tau), p_*(t)) &\geq h(t), \\ H(x_*(\tau), u_*(t), p_*(\tau)) &\geq h(\tau). \end{aligned}$$

После простейших преобразований получим

$$\begin{aligned} &H(x_*(t), u_*(\tau), p_*(t)) - H(x_*(\tau), u_*(\tau), p_*(\tau)) \geq \\ &\geq h(t) - h(\tau) \geq H(x_*(t), u_*(t), p_*(t)) - H(x_*(\tau), u_*(t), p_*(\tau)). \end{aligned} \quad (3.3.13)$$

При фиксированном значении  $u_*(\tau)$  функция  $H(x_*(t), u_*(t), p_*(t))$  непрерывно зависит от  $t$ . Поэтому, если  $t \rightarrow \tau$ , то левая часть неравенства (3.3.13) стремится к нулю. При фиксированном значении  $u_*(t)$  функция  $H(x_*(\tau), u_*(t), p_*(\tau))$  непрерывно зависит от  $\tau$ , следовательно, устремив  $\tau \rightarrow t$ , получим, что правая часть неравенства (3.3.13) также стремится к нулю. Отсюда заключаем, что вдоль оптимальной траектории функция  $h(t)$  непрерывно зависит от  $t$  в каждой точке  $t \in [T_1, T_2]$ .

Разделив левое неравенство (3.3.13) на  $\delta t = t - \tau > 0$ , преобразуем к виду

$$\begin{aligned} &[H(x_*(t), u_*(\tau), p_*(t)) - H(x_*(\tau), u_*(\tau), p_*(t))]/\delta t + \\ &+ [H(x_*(\tau), u_*(\tau), p_*(t)) - H(x_*(\tau), u_*(\tau), p_*(\tau))]/\delta t \geq [h(t) - h(\tau)]/\delta t. \end{aligned}$$

Функция  $H(x, u, p)$  дифференцируема по  $x$  и  $p$ , поэтому, переходя в полученном неравенстве к пределу при  $\delta t \rightarrow 0$ , получим

$$\langle \dot{x}, H_x(x_*(\tau), u_*(t), p_*(t)) \rangle + \langle \dot{p}, H_p(x_*(\tau), u_*(t), p_*(t)) \rangle \geq \lim_{\delta t \rightarrow 0} [h(t) - h(t - \delta t)]/\delta t.$$

Подставляя сюда из (3.3.12) выражения для  $\dot{x}$  и  $\dot{p}$ , находим

$$0 \geq \lim_{\delta t \rightarrow 0} [h(t) - h(t - \delta t)]/\delta t.$$

Выполняя аналогичные действия с правой частью (3.3.12), получим

$$\lim_{\delta t \rightarrow 0} [h(t) - h(t - \delta t)]/\delta t \geq 0,$$

откуда заключаем, что

$$\frac{d}{dt}H(x_*(t), u_*(t), p_*(t)) = 0.$$

Итак, вдоль оптимальной траектории функция Понтрягина сохраняет постоянное значение.  $\square$

Функция  $\delta u(t) = v(t) - u_*(t)$ , где  $v(t)$  определяется из условия (3.3.7), называется *игольчатой вариацией* управления, или *вариацией Макшейна*. Термин “игольчатая” подчеркивает тот факт, что, во-первых, вариация управления  $\delta u(t)$  отлична от нуля лишь на малом интервале  $[\tau - \varepsilon, \tau]$  и, во-вторых, на этом интервале вариация может принимать конечные значения и вовсе не должна быть малой. Такой “скачкообразный” характер игольчатой вариации отличен от обычно применяемых в классическом вариационном исчислении “гладких” вариаций управления. В этом состоит отличие теоремы 3.3.1 от условий экстремума, полученных в классическом вариационном исчислении. Вместе с тем утверждения, даваемые теоремой 3.3.1, как и в вариационном исчислении, являются следствиями локального варьирования траектории и основаны на анализе близких фазовых траекторий. Формально это использовано в основной формуле (3.3.10) для оценки приращения функционала, где сохранен лишь главный (линейный) член вариации траектории. В классическом вариационном исчислении варьирование проводилось в классе близких управлений, т.е. в более узком классе вариаций. Принцип максимума Л.С. Понтрягина существенно расширил возможности решения оптимизационных задач, однако он не создал теоретической базы для нахождения глобальных решений.

Обычно в теории оптимального управления вектор импульсов вводится несколько иначе. Обозначим  $\lambda(t) = -p(t)$ . Тогда система (3.3.4) и краевое условие (3.3.5) запишутся в виде

$$\dot{\lambda}(t) = -f_x^\top(t, x, u)\lambda, \quad \lambda(T_2) = -W_x(x(T_2)).$$

Роль функции Понтрягина будет играть функция

$$\hat{H}(t, x, u, \lambda) = \langle f(t, x, u), \lambda \rangle - H(t, x, u, p).$$

При использовании новой функции  $\tilde{H}$  имеем

$$\min_{w \in U} H(t, x_*(t), w, p_*(t)) = -\max_{w \in U} \hat{H}(t, x_*(t), w, \lambda_*(t))$$

и условие принципа максимума (3.3.6) запишется в виде

$$u_*(t) \in \text{Arg} \max_{w \in U} \hat{H}(t, x_*(t), w, \lambda_*(t)),$$

т.е. оптимальное управление доставляет максимум функции  $\hat{H}$  по  $w \in U$ . Отсюда возникло слово “максимум” в термине “принцип максимума Л.С. Понтрягина”. Утверждение (3.3.6) следовало, скорее, назвать принципом минимума, однако, учитывая широкую популярность термина “принцип максимума Л.С. Понтрягина”, мы сохраним его и за условием (3.3.6). Это тем

более оправдано, что отличие между обеими формулировками несущественно. Формулировка (3.3.6) представляется более удобной, чем традиционная, предложенная Л.С. Понтрягиным. Действительно, в задаче на минимум некоторого функционала проще запомнить, что оптимальное управление доставляет минимум (а не максимум) вспомогательной функции  $H$ .

В том случае, когда функция  $H(t, x_*(t), u, p_*(t))$  дифференцируема по  $u$  в точке  $u = u_*(t)$ , справедливы следующие утверждения, вытекающие из теоремы 3.3.1.

**Следствие 3.3.1.** *Пусть выполнены условия теоремы 3.3.1 и либо  $U$  — открытое множество, либо  $u_*(t) \in \text{int } U$ . Тогда необходимо, чтобы при всех  $t \in \tilde{T}'$  оптимальное управление  $u_*(t)$  являлось стационарной точкой функции Понтрягина, т.е.*

$$H_u(t, x_*(t), u_*(t), p_*(t)) = 0_r. \quad (3.3.14)$$

Согласно теореме Ферма это утверждение следует из (3.3.6). Полученное необходимое условие минимума эквивалентно уравнению Эйлера в классическом вариационном исчислении.

К основному утверждению (3.3.6) принципа максимума применим принцип линеаризации (см. Приложение 2), получим

**Следствие 3.3.2 (линеаризованный принцип максимума).** *Пусть выполнены условия теоремы 3.3.1; кроме того, функция  $f(t, x, u)$  непрерывно дифференцируема по  $u$ , множество  $U$  выпукло. Тогда условие (3.3.6) можно заменить следующим необходимым условием минимума:*

$$u_*(t) \in \text{Arg } \min_{u \in U} \langle H_u(t, x_*(t), u_*(t), p_*(t)), u - u_* \rangle. \quad (3.3.15)$$

**Следствие 3.3.3.** *Пусть помимо предположений, приведенных в следствии 3.3.2, функция  $f(t, x, u)$  дважды дифференцируема по  $u$ ; тогда необходимо, чтобы на оптимальной траектории при всех  $t$  из интервала  $\tilde{T}'$  и для любых векторов  $v \in \mathbb{R}^r$  выполнялось неравенство*

$$v^\top H_{uu^\top}(t, x_*(t), u_*(t), p_*(t))v \geq 0.$$

Это утверждение, обычно называемое *линеаризованным принципом минимума*, также следует из (3.3.6). Принцип максимума был изложен на примере задачи Майера по двум причинам. Во-первых, это простейшая задача оптимального управления, ее описание не усложнено второстепенными деталями. Во-вторых, многие задачи ОУ легко сводятся к этой задаче. Ниже приведем несколько примеров.

В задаче Больца ищется минимум функционала, содержащего интегральное слагаемое. Требуется найти

$$\min_{u \in U} \left[ W(x(T_2)) + \int_{T_1}^{T_2} \psi(t, x(t), u(t)) dt \right]. \quad (3.3.16)$$

Введем новую фазовую переменную  $x^{n+1}(t)$  и соответствующий импульс  $p^{n+1}(t)$ , удовлетворяющие следующим дифференциальным уравнениям:

$$\dot{x}^{n+1} = \psi(t, x(t), u(t)), \quad \dot{p}^{n+1} = 0. \quad (3.3.17)$$

Тогда задачу (3.3.16) можно заменить эквивалентной задачей Майера о минимизации терминального функционала  $W(x(T_2)) + x^{n+1}(T_2)$  для систем (3.3.1) и (3.3.17). Для этой задачи справедлив принцип максимума Л.С. Понтрягина с очевидными добавлениями: вместо (3.3.4) здесь используется система

$$\dot{p} = -f_x^\top(t, x, u)p - \psi_x(t, x, u),$$



к краевым условиям (3.3.2) и (3.3.5) добавляются следующие:  $x^{n+1}(T_1) = 0$ ,  $p^{n+1}(T_2) = 1$ . Для сопряженной системы  $p^{n+1}(t) \equiv 1$ .

Рассмотрим частный случай, когда система (3.3.1) линейна по  $x$  и имеет вид

$$\frac{dx}{dt} = A(t)x + \mathcal{R}(t, u), \quad x(T_1) = x_1, \quad (3.3.18)$$

где  $A(t)$  — матрица  $n \times n$ ,  $\mathcal{R}$  — непрерывная вектор-функция своих аргументов. Ставится задача нахождения оптимального управления  $u_*(t) \in U$ , доставляющего минимум линейному функционалу

$$W(x(T_2)) = \langle c, x(T_2) \rangle, \quad (3.3.19)$$

где  $x(T_2)$  определяется из (3.3.18), вектор  $c \in \mathbb{R}^n$ .

Предполагается, что матрица  $A(t)$  является непрерывной функцией  $t$ . Система (3.3.4) и условие (3.3.5) в данном случае не зависят от фазового вектора  $x$ , так как

$$\frac{dp}{dt} = -A^\top(t)p, \quad p(T_2) = c. \quad (3.3.20)$$

Продифференцируем скалярное произведение  $\langle x(t), p(t) \rangle$  по  $t$ . С учетом (3.3.18) и (3.3.20) имеем

$$\frac{d}{dt} \langle x(t), p(t) \rangle = p^\top(t) \mathcal{R}(t, u).$$

Отсюда легко получим, что

$$W(x(T_2)) = \langle x_1, c \rangle + \int_{T_1}^{T_2} p^\top(t) \mathcal{R}(t, u) dt. \quad (3.3.21)$$

Пусть  $u_*(t)$  — управление, доставляющее минимум функционала  $W(x)$ , тогда из формулы (3.3.21) следует

**Теорема 3.3.2.** *Для линейных систем (3.3.18) и (3.3.20) выполнение неравенства*

$$p^\top(t) \mathcal{R}(t, u(t)) \geq p^\top(t) \mathcal{R}(t, u_*(t)), \quad \forall t \in T, \quad \forall u \in U,$$

*дающее необходимое и достаточное условие оптимальности управления  $u_*(t)$  в задаче минимизации терминального функционала (3.3.19).*

Расширим класс управляемых процессов (3.3.1), добавив возможность одновременной минимизации целевого функционала не только по управлению  $u(t)$ , но и по управляющему параметру  $\xi \in \mathbb{R}^s$ , сохраняющему на интервале движения постоянное значение. При решении практических задач вектор параметров  $\xi$  имеет самое широкое применение: в него могут быть включены и недостающие краевые условия, и интервал движения, и конструктивные оптимизируемые параметры объекта и т.д.

Пусть управляемый процесс описывается системой обыкновенных дифференциальных уравнений:

$$\frac{dx}{dt} = f(t, x, u, \xi), \quad T_1 \leq t \leq T_2, \quad x(T_1) = x_1. \quad (3.3.22)$$

где вектор-функция  $x(t) \in \mathbb{R}^n$ , управлением является кусочно-непрерывная вектор-функция  $u(t)$ , принимающая значения из допустимого множества  $U \subset \mathbb{R}^r$ , вектор параметров  $\xi \in V \subset \mathbb{R}^s$ .

Считаем, что целевая функция  $W(x(T_2), \xi)$  и правые части (3.3.22) непрерывно дифференцируемы по  $x$  и по  $\xi$ . Тогда сопряженная система (3.3.4) и краевое условие (3.3.5) запишутся в виде

$$\frac{dp}{dt} = -f_x^\top(t, x, u, \xi)p, \quad T_1 \leq t \leq T_2, \quad (3.3.23)$$

$$p(T_2) = W_x(x_*(T_2), \xi_*). \quad (3.3.24)$$

Предположим, что числа  $T_1$  и  $T_2$  фиксированы, заданы начальное и конечное условия

$$x(T_1) = x_1, \quad x(T_2) = x_2. \quad (3.3.25)$$

Для системы (3.3.22) ставится задача отыскания допустимого оптимального управления  $u_*(t)$ , оптимального вектора  $\xi_*$ , доставляющих минимум функционала  $W(x(T_2), \xi)$ . Предполагается, что эти решения существуют, им соответствуют решения  $x_*(t)$ ,  $p_*(t)$  систем (3.3.22) и (3.3.23), удовлетворяющие краевым условиям (3.3.24) и (3.3.25).

Согласно формуле (3.1.22) производная целевой функции  $W$  вычисляется по формуле

$$\frac{dW(x_*(T_2), \xi_*)}{d\xi} = W_\xi(x_*(T_2), \xi_*) + \frac{d}{d\xi} \int_{T_1}^{T_2} f^\top(t, x_*(t), u_*(t), \xi_*) p_*(t) dt. \quad (3.3.26)$$

Для систем (3.3.22), (3.3.23) выполнены условия максимума Л.С. Понтрягина. Кроме того, в случае, если  $V = \mathbb{R}^s$ , то согласно теореме Ферма необходимое условие минимума  $W$  по  $\xi$  имеет вид

$$\frac{dW(x_*(T_2), \xi_*)}{d\xi} = 0_s.$$

Это равенство содержит  $s$  скалярных соотношений, необходимых для определения компонент оптимального вектора  $\xi_*$ .

Рассмотрим задачу с нефиксированным временем движения. Пусть требуется для системы (3.3.22) найти допустимое управление  $u(t)$ , конечный момент движения  $T_2$  так, чтобы функция  $W(x(T_2), T_2)$  принимала наименьшее возможное значение. Начальный момент  $T_1$  и начальная точка  $x(T_1)$  заданы и фиксированы. Вместо  $t$  введем новую независимую переменную  $\tau$  по формуле  $t = T_1 + \xi\tau$ , где  $\tau$  изменяется от 0 до 1, соответствующее  $t$  при этом изменяется от  $T_1$  до  $T_2 = T_1 + \xi$ . Таким образом здесь скаляр  $\xi$  – интервал движения по переменной  $t$ . Обратное преобразование есть  $\tau = (t - T_1)/\xi$ .

Определим вектор-функции  $\bar{x}(\tau)$ ,  $\bar{u}(\tau)$ ,  $\bar{p}(\tau)$ , положив

$$\bar{x}(\tau) = x(t(\tau)), \quad \bar{u}(\tau) = u(t(\tau)), \quad \bar{p}(\tau) = p(t(\tau)), \quad t(\tau) = T_1 + \xi\tau, \quad 0 \leq \tau \leq 1,$$

где  $x(t)$ ,  $p(t)$ ,  $u(t)$  – рассматриваемый процесс,  $u(t)$  – кусочно-непрерывная функция, принимающая значения из области управления  $U$ . Дифференцируя  $\bar{x}(\tau)$ ,  $\bar{p}(\tau)$  по  $\tau$  с учетом (3.3.22), (3.3.23), имеем

$$\frac{d\bar{x}}{d\tau} = \frac{dx(t)}{dt} \frac{dt}{d\tau} = \xi f(t(\tau), \bar{x}(\tau), \bar{u}(\tau)).$$

Следовательно, функции  $\bar{x}(\tau)$  и  $\bar{p}(\tau)$  удовлетворяют системам

$$\frac{d\bar{x}(\tau)}{d\tau} = \xi f(t(\tau), \bar{x}(\tau), \bar{u}(\tau), \xi), \quad (3.3.27)$$

$$\frac{d\bar{p}(\tau)}{d\tau} = -\xi f_x^\top(t(\tau), \bar{x}(\tau), \bar{u}(\tau), \xi) \bar{p}(\tau) \quad (3.3.28)$$

с краевыми условиями для  $\bar{x}(\tau)$  и  $\bar{p}(\tau)$ :

$$\begin{aligned} x(T_1) &= \bar{x}(0) = x_1, & \bar{p}(0) &= W_x(x_1), \\ x(T_2) &= \bar{x}(1) = x_2, & \bar{p}(1) &= W_x(x_2). \end{aligned}$$

Благодаря переходу к новой независимой переменной  $\tau$ , мы получили систему (3.3.28), в которой фиксировано время движения по переменной  $\tau$ , так как здесь  $0 \leq \tau \leq 1$ . К этой задаче можно применить принцип максимума Л.С. Понтрягина (теореме 3.3.1).

В качестве переменной интегрирования возьмем первоначальное время  $t = T_1 + \xi\tau$ , при этом изменяются пределы интегрирования, получим

$$\frac{dW(x(T_2), \xi)}{d\xi} = W_\xi(x_*(T_2), \xi) + \frac{d}{d\xi} \int_{T_1}^{T_1+\xi} f^\top(t, x(t), u(t), \xi)p(t)dt.$$

Отсюда следует

$$\frac{dW(x(T_2), \xi)}{d\xi} = W_\xi(x(T_2), \xi) + f^\top(T_2, x(T_2), u(T_2), \xi)p(T_2). \quad (3.3.29)$$

Вектор  $p(T_2)$  определяется из (3.3.23).

Скалярное соотношение (3.3.29) дает одно дополнительное условие для определения оптимальной величины интервала движения  $\xi = T_2 - T_1$ . Во всем остальном остаются справедливыми утверждения теоремы 3.3.1.

В задаче об оптимальном быстродействии на интервале  $\tilde{T} = T_2 - T_1$  ищется допустимое управление  $u(t)$ , переводящее систему (3.3.1) из точки  $x_1$  в точку  $x_2$  за минимально возможное время. Это значит, что решение  $x(t)$  системы (3.3.1) удовлетворяет следующим краевым условиям:

$$x(T_1) = x_1, \quad x(T_2) = x_2 \quad (3.3.30)$$

и не существует другого допустимого управления, удовлетворяющего (3.3.30), для которого интервал  $\tilde{T} = T_2 - T_1$  был бы меньшим. Сделаем преобразование независимой переменной, получим условие (3.3.29). В качестве целевой функции возьмем  $W = \xi$ . Тогда согласно (3.3.29) необходимое условие в задаче о быстродействии будет иметь вид

$$1 = f^\top(T_2, x(T_2), u(T_2))p(T_2).$$

Приведенный выше принцип максимума Л.С. Понтрягина был получен для задач, содержащих ограничения на выбор управления  $u \in U$ , но не учитывались какие-либо ограничения на фазовый вектор. Вместе с тем существует широкий класс практических задач оптимального управления, содержащих ограничения на фазовые переменные. К методам решения таких задач мы переходим.

**Основная задача оптимального управления** ставится следующим образом: найти управление  $u_*(t) \in U$  и вектор параметров  $\xi_* \in V$ , которые минимизируют функционал  $W(x(T_2), \xi)$ , при наличии “смешанных” ограничений на векторы состояния, управления и вектор параметров вдоль траектории системы (3.3.22):

$$h_1(t, x(t), u(t), \xi) = 0_\alpha, \quad g_1(t, x(t), u(t), \xi) \leq 0_\beta, \quad T_1 \leq t \leq T_2 \quad , - \quad (3.3.31)$$

и в конце движения:

$$h_2(T_2, x(T_2), \xi) = 0_\alpha, \quad g_2(T_2, x(T_2), \xi) \leq 0_\beta, \quad T_1 \leq t \leq T_2. \quad (3.3.32)$$

Для решения этой задачи используем ту же самую идею, разработанную и широко применяемую в нелинейном программировании. Составим функционал, штрафующий за нарушение ограничений (3.3.31) и (3.3.32):

$$P = \|h_2(T_2, x(T_2), \xi)\|^2 + \|(g_2(T_2, x(T_2), \xi))_+\|^2 + \int_{T_1}^{T_2} [\|h_1(t, x(t), u(t), \xi)\|^2 + \|(g_1(t, x(t), u(t), \xi))_+\|^2] dt.$$

В качестве целевой функции возьмем

$$J(\gamma) = W(x(T_2), \xi) + \gamma P.$$

При каждом фиксированном  $\gamma$  решаем задачу минимизации  $J(\gamma)$  по  $u \in U$ ,  $\xi \in V$ ; получив решение этих задач и устремив  $\gamma$  к бесконечности, находим решение задачи с фазовыми ограничениями. Для дискретного варианта задач это утверждение очевидно, оно следует из теории нелинейного программирования. Предельный переход от дискретного случая к непрерывному требует специального обоснования.

Выше, во введении, были указаны работы, выполненные А.А. Милютиним, А.Я. Дубовицким, их учениками и коллегами, в которых принцип максимума Л.С. Понтрягина был распространен на основную задачу оптимального управления. Ниже будет использоваться другой, более простой подход, основанный на решении дискретного варианта основной задачи ОУ. Вместо исходной динамической системы изучается ее дискретная аппроксимация вида (1.1.2). Редукция задач оптимального управления к задачам нелинейного программирования дает большие возможности для использования богатого набора методов решения задач НЛП.

**Каноническая форма задания задачи оптимального управления.** Системы (3.3.22) и (3.3.23) запишем в виде

$$\frac{dx}{dt} = H_p(t, x, u, p, \xi), \quad \frac{dp}{dt} = -H_x(t, x, u, p, \xi), \quad (3.3.33)$$

где  $H(t, x, u, p, \xi) = \langle f(t, x, u, \xi), p \rangle$ .

Используя принцип максимума на оптимальной траектории, определим оптимальное управление

$$u_*(t) \in \text{Arg} \min_{w \in U} H(t, x, w, p, \xi).$$

Воспользуемся теоремой 2.10.2 Данскина–Демьянова (см. [40, с. 73], а также [41, 42]). Считаем, что разрыв управления происходит только на множестве нулевой меры по  $t$ . Тогда систему (3.3.33) можно записать в виде

$$\frac{dx}{dt} = H_p(t, x, u_*(t), p, \xi), \quad \frac{dp}{dt} = -H_x(t, x, u_*(t), p, \xi). \quad (3.3.34)$$

Эта форма будет использована ниже в примере 3.8.1.

### 3.4. Дискретный принцип максимума

Рассмотрим дискретный вариант системы (3.3.1). Воспользуемся простейшей схемой интегрирования по методу Эйлера

$$x_{i+1} = x_i + hf(t_i, x_i, u_i). \quad (3.4.1)$$

Здесь  $1 \leq i \leq k-1$ ,  $h = (T_2 - T_1)/(k-1)$ ,  $x_i \in \mathbb{R}^m$ ,  $u_i \in \mathbb{R}^r$ ,  $x_1$  — задано, терминальный функционал  $W$  заменяется функцией  $W(x_k)$ .

Оптимальное управление для дискретного процесса (3.4.1) находится из условия минимума функции  $W(x_k)$  по полному вектору управлений  $u^\top = [u_1^\top, u_2^\top, \dots, u_{k-1}^\top]$ , где все  $u_i \in U$ . Если функции  $f(t, x, u)$  и  $W(x)$  непрерывны по  $x$ ,  $u$  и множество  $U$  компактное, тогда оптимальное управление существует для любого начального вектора  $x_1$ , при этом оптимальное значение целевой функции  $W_*$  конечно.

При исследовании дискретного процесса (3.4.1) идея использования игольчатой вариации может быть реализована особенно отчетливо. В отличие от стандартного варьирования, когда варьируется оптимальное управление, ниже проводится вариация произвольного допустимого управления. Изменим постановку задачи. Зафиксируем все векторы, составляющие полный вектор управлений  $u$ , за исключением  $u_i$ . Вариация вектора  $u_i$  приводит к изменению последующих фазовых векторов  $x_{i+1}, x_{i+2}, \dots, x_k$ . Целевая функция  $W(x_k)$  сложным образом зависит от вектора  $x_{i+1}$ . Для упрощения изложения зависимость вектора  $x_{i+1}$  от  $u_i$ , задаваемую (3.4.1), запишем в виде  $x_{i+1}(u_i) = x_i + hf(t_i, x_i, u_i)$ . Порождаемому процессом (3.4.1) сложную функцию  $W$ , зависящую от  $x_{i+1}$ , обозначим  $\widetilde{W}(x_{i+1})$  и рассмотрим задачу минимизации функции  $\widetilde{W}(x_{i+1}(u_i))$  только по  $u_i \in U$ . Через  $U_*$  обозначим множество решений этой задачи

$$U_* = \text{Arg} \min_{u_i \in U} \widetilde{W}(x_{i+1}(u_i)). \quad (3.4.2)$$

Образ множества  $U$  при отображении  $x_{i+1}(u_i)$  обозначим через  $X_{i+1}$ , т.е.

$$X_{i+1} = x_{i+1}(U) = \{x_{i+1} \in \mathbb{R}^n : x_{i+1} = x_i + hf(t_i, x_i, u_i), u_i \in U\}.$$

Здесь  $X_{i+1}$  — множество достижимости на  $(i+1)$ -м шаге процесса (3.4.1), т.е. совокупность всех состояний  $x_{i+1}$  в пространстве  $\mathbb{R}^n$ , в которое можно перевести точку  $x_i$  за один шаг с помощью всевозможных допустимых управлений  $u_i \in U$ .

Вместо минимизации функции  $\widetilde{W}(x_{i+1}(u_i))$  по вектору управлений  $u_i \in U$  рассмотрим эквивалентную задачу о минимизации  $\widetilde{W}(x_{i+1})$  по фазовому вектору  $x_{i+1} \in X_{i+1}$  (проведем минимизацию в пространстве состояний) и определим множество

$$X_{i+1}^* = \text{Arg} \min_{x_{i+1} \in X_{i+1}} \widetilde{W}(x_{i+1}). \quad (3.4.3)$$

Очевидно, что  $X_{i+1}^*$  является образом множества  $U_*$  при отображении  $x_{i+1}(u_i)$ , т.е.  $X_{i+1}^* = x_{i+1}(U_*)$ .

Если вектор-функция  $f$  и функция  $W$  дифференцируемы по  $x$ , то согласно формулам (1.3.9) — (1.3.11) имеем

$$p_i = p_{i+1} + hf_x^\top(t_i, x_i, u_i)p_{i+1}, \quad p_k = W_x(x_k), \quad 1 \leq i \leq k-1. \quad (3.4.4)$$

Теперь для всякого вектора  $u_i \in U$  можно вычислить не только векторы  $x_{i+1}, x_{i+2}, \dots, x_k$ , но и с помощью (3.4.4) последовательно найти  $p_k, p_{k-1}, \dots, p_{i+1}$ . Такая операция определяет однозначные отображения

$$p_{i+1}(x_{i+1}) = p_{i+1}(x_{i+1}(u_i)) = \frac{d\widetilde{W}(x_{i+1})}{dx_{i+1}}.$$

Если функция  $f$  дифференцируема по  $u$ , то по правилу дифференцирования сложной функции находим

$$\begin{aligned} \frac{d\widetilde{W}(x_{i+1}(u_i))}{du_i} &= \frac{\partial x_{i+1}^\top}{\partial u_i} \frac{d\widetilde{W}(x_{i+1}(u_i))}{dx_{i+1}} = hf_{u_i}^\top(t_i, x_i, u_i)p_{i+1}(x_{i+1}(u_i)) = \\ &= H_{u_i}(t_i, x_i, u_i, p_{i+1}(x_{i+1})), \end{aligned}$$

где  $H(t_i, x_i, u_i, p_{i+1}) = hf^\top(t_i, x_i, u_i)p_{i+1}$ .

**Теорема 3.4.1 (необходимые условия минимума).** Пусть функции  $W$  и  $f$  дифференцируемы по  $x$ , множество  $X_{i+1}$  выпукло. Тогда:

1) если  $x_{i+1}^* \in X_{i+1}^*$ , то

$$0 \leq \langle p_{i+1}(x_{i+1}^*), x_{i+1} - x_{i+1}^* \rangle \quad \forall x_{i+1} \in X_{i+1}; \quad (3.4.5)$$

2) если  $u_i^* \in U_*$ , то

$$0 \leq \langle p_{i+1}(x_{i+1}(u_i^*)), f(t_i, x_i, u_i) - f(t_i, x_i, u_i^*) \rangle \quad \forall u_i \in U; \quad (3.4.6)$$

3) если множество  $U$  выпуклое,  $u_i^* \in U_*$ , то

$$0 \leq \langle f_u^\top(t_i, x_i, u_i^*)p_{i+1}(x_{i+1}(u_i^*)), u_i - u_i^* \rangle \quad \forall u_i \in U. \quad (3.4.7)$$

**Доказательство.** Из дифференцируемости функции  $\widetilde{W}(x_{i+1})$  в точке  $x_{i+1}^*$  следует

$$\widetilde{W}(x_{i+1}) - \widetilde{W}(x_{i+1}^*) = \langle p_{i+1}(x_{i+1}^*), x_{i+1} - x_{i+1}^* \rangle + \|x_{i+1} - x_{i+1}^*\| \alpha(x_{i+1}^*, \|x_{i+1} - x_{i+1}^*\|), \quad (3.4.8)$$

где  $\lim_{x_{i+1} \rightarrow x_{i+1}^*} \alpha(x_{i+1}^*, \|x_{i+1} - x_{i+1}^*\|) = 0$ .

К задаче минимизации  $\widetilde{W}$  по  $x_{i+1}$  применим принцип линеаризации (см. Приложение 2). Тогда из  $x_{i+1}^* \in X_{i+1}^*$  следует, что имеет место (3.4.5). Переписав неравенство (3.4.5) в пространстве управлений, получим (3.4.6).

Если  $U$  выпукло, то к задаче

$$\min_{u \in U} \langle p_{i+1}(x_{i+1}(u^*)), f(t_i, x_i, u_i) \rangle$$

снова применим принцип линеаризации, приходим к (3.4.7).  $\square$

Если  $u_i^* \in U_*$ , то условие (3.4.6) является дискретным аналогом принципа максимума Л.С. Понтрягина. Условие (3.4.5) выражает это же свойство, но в пространстве состояний. Необходимое условие теоремы о справедливости (3.4.5) можно назвать *дискретным принципом максимума в пространстве состояний*, а утверждение о выполнимости (3.4.6) — *дискретным принципом максимума в пространстве управлений*. Условие (3.4.7) есть дискретный аналог линеаризованного принципа максимума (3.3.15).

Если функция  $\widetilde{W}(x_{i+1})$  дифференцируема в точке  $x_{i+1}^* \in x_{i+1}(U_*)$ , то можно ввести точно-множественное отображение

$$V_1(x_{i+1}^*) = \text{Arg} \min_{x_{i+1} \in X_{i+1}} \langle p_{i+1}(x_{i+1}^*), x_{i+1} - x_{i+1}^* \rangle$$

или, перейдя в пространство управлений, определить для  $u_i^* \in U$  многозначное отображение

$$V_2(u_i^*) = \text{Arg} \min_{u_i \in U} \langle p_{i+1}(x_{i+1}(u_i^*)), x_{i+1}(u_i) - x_{i+1}(u_i^*) \rangle.$$

Пусть  $x_{i+1}^* = x_{i+1}(u_i^*)$ ; тогда множество  $V_1(x_{i+1}^*)$  является образом множества  $V_2(u_i^*)$  при отображении  $x_{i+1}(u_i)$ :

$$V_1(x_{i+1}^*) = x_{i+1}(V_2(u_i^*)).$$

Для каждого вектора  $u_i^* \in U$  определим точечно-множественное отображение

$$V_3(u_i^*) = \text{Arg} \min_{u_i \in U} \left\langle \frac{d\widetilde{W}(x_{i+1}(u_i^*))}{du_i}, u_i - u_i^* \right\rangle.$$

Подобно тому, как это было сделано в Приложении 2, можно отыскивать неподвижные точки многозначных отображений  $V_1(x_{i+1})$ ,  $V_2(u_i)$ ,  $V_3(u_i)$ , т.е. точки, удовлетворяющие условиям

$$x_{i+1}^* \in V_1(x_{i+1}^*), \quad u_i^* \in V_2(u_i^*), \quad u_i^* \in V_3(u_i^*).$$

**Теорема 3.4.2 (достаточные условия минимума).** Пусть функция  $\widetilde{W}(x_{i+1})$  и множество  $X_{i+1}$  выпуклы; тогда из условий (3.4.5), (3.4.6) следует, что  $x_{i+1}^* \in X_{i+1}^*$ ,  $u_i^* \in U_*$  и для любых  $\hat{x}_{i+1} \in V_1(x_{i+1}^*)$ ,  $\hat{u}_i \in V_2(u_i^*)$  имеют место неравенства

$$0 \leq \widetilde{W}(x_{i+1}) - \widetilde{W}(x_{i+1}^*) \leq \langle p_{i+1}(x_{i+1}^*), x_{i+1}^* - \hat{x}_{i+1} \rangle, \quad (3.4.9)$$

$$0 \leq \widetilde{W}(x_{i+1}) - \widetilde{W}(x_{i+1}^*) \leq H(t_i, x_i, u_i^*, p_{i+1}(x_{i+1}^*)) - H(t_i, x_i, \hat{u}_i, p_{i+1}(x_{i+1}^*)). \quad (3.4.10)$$

**Доказательство.** Из выпуклости функции  $\widetilde{W}(x_{i+1})$  по  $x_{i+1}$  следует, что для любого  $x_{i+1} \in X_{i+1}$  имеет место неравенство

$$\widetilde{W}(x_{i+1}) - \widetilde{W}(x_{i+1}^*) \geq \langle p_{i+1}(x_{i+1}^*), x_{i+1} - x_{i+1}^* \rangle, \quad \text{где} \quad p_{i+1}(x_{i+1}^*) = \frac{d\widetilde{W}(x_{i+1}^*)}{dx_{i+1}}. \quad (3.4.11)$$

Если выполнено (3.4.5) или (3.4.6), то согласно (3.4.11) имеем  $\widetilde{W}(x_{i+1}) \geq \widetilde{W}(x_{i+1}^*)$ , и в силу того, что  $x_{i+1}$  — произвольный допустимый вектор, заключаем, что  $x_{i+1}^* \in X_{i+1}^*$ ,  $u_i^* \in U_*$ . Из условия  $\hat{x}_{i+1} \in V_1(x_{i+1}^*)$  следует

$$\langle p_{i+1}(x_{i+1}^*), x_{i+1}^* \rangle \leq \langle p_{i+1}(x_{i+1}^*), x_* \rangle. \quad (3.4.12)$$

Возьмем в формуле (3.4.11)  $x_{i+1} \in X_{i+1}^*$  и, воспользовавшись (3.4.12), получим

$$0 \leq \widetilde{W}(x_{i+1}^*) - \widetilde{W}(x_*) \leq \langle p_{i+1}(x_{i+1}^*), x_{i+1}^* - x_* \rangle \leq \langle p_{i+1}(x_{i+1}^*), x_{i+1}^* - \hat{x}_{i+1} \rangle.$$

Мы пришли, таким образом, к (3.4.9); если правую часть этой формулы выразить через функцию  $H$ , то получим (3.4.10).  $\square$

В отличие от принципа максимума для непрерывных систем здесь не постулируется, что управление  $u$  оптимально на всей траектории системы (3.4.1), так как оптимизация осуществляется только на  $i$ -м интервале. Аналогичные утверждения можно получить относительно условий (3.4.7). Для этого определим функцию

$$\psi(u_i^*, u_i) = hf^\top(t_i, x_i, u_i) p_{i+1}(x_{i+1}(u_i^*)) = H(t_i, x_i, u_i, p_{i+1}(u_i^*)).$$

**Теорема 3.4.3 (достаточные условия минимума).** Пусть сложная функция

$$\widetilde{W}(x_{i+1}(u_i))$$

определена на выпуклом множестве  $U$  и дифференцируема в  $u_i^*$ . Тогда:

- 1) если  $u_i^* \in U_*$ , то имеет место (3.4.7);
- 2) если  $u_i^* \in U_*$  и  $\psi(u_i^*, u_i)$  — псевдовыпуклая функция  $u_i$  в точке  $u_i^*$  относительно  $U$ , то в точке  $u_i^*$  выполнено условие (3.4.6), а в точке  $x_{i+1}^* = x_{i+1}(u_i^*)$  — условие (3.4.5);

3) если в точке  $u_i^*$  выполнено либо условие (3.4.6) и  $U$  выпукло, либо условие (3.4.7) и в  $u_i^*$  функция  $\widetilde{W}(x_{i+1}(u_i))$  псевдовыпукла по  $u_i$  относительно  $U$ , то  $u_i^* \in V_*$ ;

4) если  $\widetilde{W}(x_{i+1}(u_i))$  – выпуклая функция  $u_i$  на  $U$ , то для любых  $\tilde{u}_i \in V_3(u_i^*)$  имеет место оценка

$$0 \leq \widetilde{W}(x_{i+1}^*) - \widetilde{W}(x_*) \leq h \langle f_{u_i}^\top(t_i, x_i, u_i^*) p_{i+1}(x_{i+1}(u_i^*)), u_i^* - \tilde{u}_i \rangle. \quad (3.4.13)$$

**Доказательство.** Докажем первое утверждение. Из дифференцируемости сложной функции  $\widetilde{W}(x_{i+1}(u_i))$  следует, что

$$\widetilde{W}(x_{i+1}(u_i)) - \widetilde{W}(x_{i+1}(u_i^*)) = \langle H_u(t_i, x_i, u_i^*, p_{i+1}(x_{i+1}(u_i^*))), u_i - u_i^* \rangle + \|u_i - u_i^*\| \beta(u_i^*, u_i - u_i^*),$$

где  $\lim_{u_i \rightarrow u_i^*} \beta(u_i^*, u_i - u_i^*) = 0$ .

Условие  $u_i^* \in U_*$  означает, что функция  $\widetilde{W}(x_{i+1}(u_i))$  достигает в точке  $u_i^*$  минимума по  $u_i$  на  $U$ . Учтем выпуклость  $U$ , воспользуемся теоремой Т-П.2.1. Тогда приходим к выводу, что для любых  $u_i \in U$  должно иметь место неравенство

$$\left\langle \frac{d\widetilde{W}(x_{i+1}(u_i^*))}{du_i}, u_i - u_i^* \right\rangle \geq 0, \quad (3.4.14)$$

т.е. выполняется условие (3.4.7).

Докажем второе утверждение. Используя дифференцируемость функции  $\psi(u_i^*, u_i)$  по  $u_i$  в точке  $u_i = u_i^*$ , перепишем неравенство (3.4.14) следующим образом:

$$\left\langle \frac{d\psi(u_i^*, u_i^*)}{du_i}, u_i - u_i^* \right\rangle \geq 0.$$

Здесь дифференцирование  $\psi(u_i^*, u_i^*)$  производится по второму аргументу. Но тогда из свойства псевдовыпуклости функции  $\psi(u_i^*, u_i)$  следует, что

$$f^\top(t_i, x_i, u_i^*) p_{i+1}(x_{i+1}(u_i^*)) = \psi(u_i^*, u_i^*) \leq \psi(u_i^*, u_i). \quad (3.4.15)$$

В силу произвольности вектора  $u_i \in U$  приходим к выводу, что имеют место условия (3.4.5) и (3.4.6).

Докажем третье утверждение. Пусть  $u_i^*$  удовлетворяет (3.4.6); тогда для любых  $u_i \in U$  выполнено (3.4.15). Из условия минимума  $\psi(u_i^*, u_i)$  по  $u_i$  на выпуклом множестве  $U$  получим, что для всякого  $u_i \in U$  имеет место неравенство

$$0 \leq \left\langle \frac{d\psi(u_i^*, u_i^*)}{du_i}, u_i - u_i^* \right\rangle = \langle H_{u_i}(t_i, x_i, u_i, p_{i+1}(x_{i+1}(u_i^*))), u_i - u_i^* \rangle,$$

т.е. выполнено (3.4.7). В силу псевдовыпуклости  $\widetilde{W}(x_{i+1}(u_i))$  по  $u_i$  получаем, что  $u_i^* \in U_*$ .

Неравенство (3.4.13) получается точно так же, как было доказано (3.4.8).  $\square$

Опуская условия непрерывности и дифференцируемости, можно резюмировать утверждения, приведенные в теоремах 3.4.1 и 3.4.2, следующим образом:

- если  $u_i^* \in V_*$ , то для выполнимости в этой точке (3.4.6) достаточно либо выпуклости множества  $\Omega$ , либо псевдовыпуклости  $H(t_i, x_i, u_i, p_{i+1}(x_{i+1}(u_i^*)))$  по  $u_i$  в точке  $u_i = u_i^*$  относительно выпуклого множества  $U$ ;



- если в точке  $u_i^* \in U$  имеет место (3.4.6), то для выполнимости условия  $u_i^* \in V_*$  достаточно, чтобы либо функция  $\widetilde{W}(x_{i+1})$  была псевдовыпукла в точке  $x_{i+1}^* = x_{i+1}(u_i^*)$  относительно  $\Omega$ , либо функция  $\widetilde{W}(x_{i+1}(u_i))$  была псевдовыпуклой по  $u_i$  в точке  $u_i$  относительно выпуклого множества  $U$ .

При обосновании необходимых условий минимума требование о выпуклости  $\Omega$  можно не вводить, если вариационные неравенства трактовать локально. Например, вместо условия (3.4.6) потребовать выполнимости неравенства

$$[f^\top(t_i, x_i, u_i^*) - f^\top(t_i, x_i, u_i^*)] \geq 0$$

для всех векторов  $u_i$ , принадлежащих  $U$  и лежащих в достаточно малой окрестности точки  $u_i^*$ . Утверждение о том, что для дифференцируемой функции  $\widetilde{W}(x_{i+1})$  из условия  $u_i^* \in U_*$  следует приведенное неравенство, называется *локальным принципом максимума в пространстве управлений*.

### 3.5. Дискретный принцип квазимаксима

Сравним принцип максимума, сформулированный в § 3.3, с результатами § 3.4. Обоснование принципа максимума для системы дифференциальных уравнений (3.3.1) не потребовало введения условия выпуклости множества  $X_{i+1}$ , в то время как для дискретных задач ОУ это условие оказывается весьма существенным. Несложно привести примеры, в которых нарушение выпуклости множества  $X_{i+1}$  приводит к тому, что дискретные принципы максимума в пространстве управлений и состояний оказываются несправедливыми. Вместе с тем ясно, что, взяв в численных схемах интегрирования системы (3.3.1) достаточно малые шаги интегрирования, можно при весьма общих предположениях получить сколь угодно точную аппроксимацию решения исходной системы; незначительно при этом должны отличаться и свойства полученных решений. Кажущееся противоречие легко устраняется, если дискретным принципам давать иную трактовку, идея которой была высказана в работе Р. Габасова и Ф.М. Кирилловой [27]. Этот вопрос обсуждается ниже на стр. 112, 113.

Покажем, что приближенно выполняемые дискретные принципы максимума имеют место без предположения о выпуклости и они выполняются тем точнее, чем меньше шаг интегрирования исходной системы (3.3.1). При этом речь будет идти о принципе максимума в пространстве управлений и состояний, так как линеаризованный принцип максимума справедлив лишь при выполнении условия выпуклости множества  $U$ .

В § 3.4 были введены множества  $V_1(x_{i+1}^*)$ ,  $V_2(u_i^*)$ , которые можно представить следующим эквивалентным образом:

$$\begin{aligned} V_1(x_{i+1}^*) &= \{x_{i+1} \in X_{i+1} : \langle p_{i+1}(x_{i+1}^*), x_{i+1} \rangle = \min_{\hat{x}_{i+1} \in X_{i+1}} \langle p_{i+1}(x_{i+1}^*), \hat{x}_{i+1} \rangle\}, \\ V_2(u_i^*) &= \{u_i \in U : H(t_i, x_i, u_i, p_{i+1}(x_{i+1}(u_i^*))) = \min_{\hat{u}_i \in U} H(t_i, x_i, \hat{u}_i, p_{i+1}(x_{i+1}(u_i^*)))\}. \end{aligned}$$

Вместо этих множеств введем множества, которые получаются из решения аналогичных минимизационных задач, но в них поиск минимума осуществляется приближенно, с погрешностью, не превосходящей величины  $\varepsilon h_i$ :

$$\begin{aligned} V_1^\varepsilon(x_{i+1}^*) &= \{x_{i+1} \in X_{i+1} : \langle p_{i+1}(x_{i+1}^*), x_{i+1} \rangle \leq \\ &\leq \varepsilon h + \min_{\hat{x}_{i+1} \in X_{i+1}} \langle p_{i+1}(x_{i+1}^*), \hat{x}_{i+1} \rangle\}, \\ V_2^\varepsilon(u_i^*) &= \{u_i \in U : \langle f(t_i, x_i, u_i), p_{i+1}(x_{i+1}(u_i^*)) \rangle \leq \\ &\leq \varepsilon + \min_{\hat{u}_i \in U} \langle f(t_i, x_i, \hat{u}_i), p_{i+1}(x_{i+1}(u_i^*)) \rangle\}. \end{aligned}$$

Очевидно, что при любых  $\varepsilon > 0$  имеют место включения

$$V_1(x_{i+1}^*) \subset V_1^\varepsilon(x_{i+1}^*), \quad V_2(u_i^*) \subset V_2^\varepsilon(u_i^*).$$

$V_2^\varepsilon(u_i^*)$  есть множество  $\varepsilon$ -оптимальных точек в задаче о минимуме скалярного произведения  $\langle f(t_i, x_i, u_i), p_{i+1}(x_{i+1}(u_i^*)) \rangle$ .

**Теорема 3.5.1.** Пусть множество  $\tilde{X}_{i+1} = f(t_i, x_i, U) \subset \mathbb{R}^m$  ограничено, функция  $\tilde{W}(x_{i+1})$  определена на открытом множестве, содержащем множество  $\tilde{X}_{i+1}$ , и существует точка  $u_i^* \in U_*$  такая, что функция  $\tilde{W}(x_{i+1})$  дифференцируема в  $x_{i+1}^* = x_{i+1}(u_i^*)$ . Тогда для любого  $\varepsilon > 0$  существует  $\hat{h} > 0$  такое, что при любом шаге интегрирования, удовлетворяющем условию  $0 < h < \hat{h}$ , имеют место следующие утверждения:

$$\begin{aligned} -\varepsilon h &\leq \langle p_{i+1}(x_{i+1}^*), x_{i+1} - x_{i+1}^* \rangle && \forall x_{i+1} \in \tilde{W}(x_{i+1}), \\ \varepsilon &\leq \langle p_{i+1}(x_{i+1}(u_i^*)), f(t_i, x_i, u_i) - f(t_i, x_i, u_i^*) \rangle && \forall u_i \in U. \end{aligned}$$

### 3.6. Дискретный вариант задач оптимального управления с разрывными правыми частями

Управляемый процесс описывается системой обыкновенных дифференциальных уравнений

$$\frac{dx}{dt} = f(x(t), u(t)), \quad T_1 \leq t \leq T_2 \quad (3.6.1)$$

с фиксированным начальным условием

$$x(T_1) = x_1. \quad (3.6.2)$$

Правая часть (3.6.1) имеет разрывы на поверхности  $\mathcal{G}(x, u) = 0$ , т.е.

$$f(x(t), u(t)) = \begin{cases} f_-(x(t), u(t)), & \text{если } \mathcal{G}(x(t), u(t)) < 0, \\ f_+(x(t), u(t)), & \text{если } \mathcal{G}(x(t), u(t)) \geq 0. \end{cases}$$

Здесь  $x(\cdot) \in \mathbb{R}^m$ ,  $u(\cdot) \in \mathbb{R}^r$ , функции  $f_-(x, u)$  и  $f_+(x, u)$  — непрерывно дифференцируемые по совокупности аргументов.

Траектория и управление должны удовлетворять условиям

$$\Gamma_1(x(t), u(t)) = 0, \quad \Gamma_2(x(t), u(t)) \leq 0, \quad T_1 \leq t \leq T_2, \quad (3.6.3)$$

$$\Gamma_3(x(T_2)) = 0, \quad \Gamma_4(x(T_2)) \leq 0. \quad (3.6.4)$$

Здесь  $\Gamma_1(x, u)$ ,  $\Gamma_2(x, u)$ ,  $\Gamma_3(x(T_2))$ ,  $\Gamma_4(x(T_2))$  суть гладкие вектор-функции. Условия (3.6.3) называются *траекторными ограничениями типа равенства и неравенства*; условия (3.6.4) представляют собой *терминальные ограничения*, заданные на правый конец траектории.

Введем минимизируемый функционал:

$$I(x(t), u(t)) = W(x(T_2)) + \int_{T_1}^{T_2} \psi(x(t), u(t)) dt. \quad (3.6.5)$$

Функции  $W(x, u)$  и  $\psi(x, u)$  предполагаются дифференцируемыми по совокупности аргументов. Среди измеримых управлений требуется найти такую вектор-функцию  $u(t)$ , чтобы для

$u(t)$  и соответствующего решения  $x(t)$  системы (3.6.1) выполнялись условия (3.6.2) – (3.6.4) и функционал (3.6.5) принимал наименьшее возможное значение.

Траектория  $x(t)$  может либо протыкать поверхность разрыва, либо скользить по ней. Пусть  $\tilde{t}$  есть момент разрыва правой части системы (3.6.1). Траектория  $x(t)$  протыкает поверхность  $\mathcal{G}(x(t), u(t)) = 0$  в момент времени  $\tilde{t}$ , если в любой достаточно малой окрестности точки  $\tilde{t}$  функция  $\mathcal{G}(x(t), u(t))$  меняет знак. Если траектория  $x(t)$  скользит по поверхности  $\mathcal{G}(x(t), u(t)) = 0$ , то на некотором интервале  $[t_1, t_2] \subset [T_1, T_2]$  имеем  $\mathcal{G}(x(t), u(t)) \equiv 0$  ( $t_1$  — момент входа,  $t_2$  — момент схода траектории с поверхности  $\mathcal{G}(x, u) = 0$ ). Рассмотрим только те случаи, когда траектория протыкает поверхность разрыва.

Систему обыкновенных дифференциальных уравнений проинтегрируем по неявной схеме. Допустим, что на интервале времени  $[T_1, T_2]$  может быть несколько разрывов правой части. Разобьем интервал времени  $[T_1, T_2]$  на  $k - 1$  отрезков точками  $t_1 = T_1, \dots, t_k = T_2$ . Введем обозначения:  $x_i = x(t_i)$ ,  $u_i = u(t_i)$ , шаг интегрирования  $h_i = t_{i+1} - t_i$ . Предположим, что величины  $h_i$  достаточно малы, так что на каждом интервале  $[t_i, t_{i+1}]$  функция  $\mathcal{G}(x, u)$  меняет знак не более одного раза.

Формула интегрирования системы (3.6.1) на каждом  $i$ -м интервале, не содержащем моментов разрыва, имеет вид

$$x_{i+1} = x_i + h_i f(x_{i+1}, u_{i+1}). \quad (3.6.6)$$

При выбранной дискретной аппроксимации управление и фазовый вектор превращаются соответственно в последовательности векторов  $u = (u_1, \dots, u_k) \in \mathbb{R}^{rk}$  и  $x = (x_1, \dots, x_k) \in \mathbb{R}^{mk}$ . Компоненты векторов управления  $u_1, \dots, u_k$  подставляются в систему уравнений (3.6.6). Тогда относительно компонент векторов  $x_1, \dots, x_k$  получим нелинейную систему уравнений, которая решается каким-нибудь итерационным методом. Траекторные и терминальные ограничения (3.6.3), (3.6.4) заменяются их дискретными аналогами:

$$\begin{aligned} \Gamma_1(x_i, u_i) &= 0, & \Gamma_2(x_i, u_i) &\leq 0, & i &= 1, 2, \dots, k, \\ \Gamma_3(x_k) &= 0, & \Gamma_4(x_k) &\leq 0. \end{aligned}$$

Для вычисления приближенного значения функционала используется одна из квадратурных формул. Тогда задача ОУ сводится к задаче безусловной минимизации функции

$$J(x, u) = W(x_k) + \sum_{i=1}^k h_i \psi(x_i, u_i).$$

Задав вектор  $u$ , с помощью (3.6.6) однозначным образом определяем вектор  $x$ . Учитывая эту зависимость, будем писать  $x = x(u)$ ,  $\Omega(u) = J(x(u), u)$ .

Пусть известны компоненты вектора  $x$  в  $i$ -м узле. Если  $\mathcal{G}(x_i, u_i) < 0$ , то вектор  $x_{i+1}$  вычисляем по формуле  $x_{i+1} = x_i + h_i f_-(x_{i+1}, u_{i+1})$ , решая систему нелинейных алгебраических уравнений. Если же  $\mathcal{G}(x_i, u_i) \geq 0$ , то для определения  $x_{i+1}$  решаем систему  $x_{i+1} = x_i + h_i f_+(x_{i+1}, u_{i+1})$ . После этого проверяем знак выражения  $\mathcal{G}(x_{i+1}, u_{i+1})$ . Если знаки выражений  $\mathcal{G}(x_i, u_i)$  и  $\mathcal{G}(x_{i+1}, u_{i+1})$  совпадают, переходим к вычислению фазового вектора в следующем узле сетки по формулам, аналогичным указанным выше. Если знаки выражений  $\mathcal{G}(x_i, u_i)$  и  $\mathcal{G}(x_{i+1}, u_{i+1})$  различны, это означает, что на интервале  $[t_i, t_{i+1}]$  происходит разрыв правых частей. В этом случае вектор  $x_{i+1}$  в точке  $i + 1$  вычисляется заново по указанной ниже схеме.

Момент разрыва правой части на интервале  $[t_i, t_{i+1}]$  обозначим через  $\tilde{t}_i$ , а величину интервала  $[t_i, t_{i+1}]$  — через  $a_i$ . Следовательно, величина интервала  $[\tilde{t}_i, t_{i+1}]$  равна  $h_i - a_i$ . Учтем, что на шаге интегрирования не более одного разрыва. Тогда если  $\mathcal{G}(x_i, u_i) < 0$ , а  $\mathcal{G}(x_{i+1}, u_{i+1}) = 0$ , то будем считать, что  $a_i = h_i$ . Если же  $\mathcal{G}(x_i, u_i) = 0$ , а  $\mathcal{G}(x_{i+1}, u_{i+1}) < 0$ , тогда положим  $a_i = 0$ .

Введем обозначения  $\tilde{x}_i = x(\tilde{t}_i)$ ,  $\tilde{u}_i = u(\tilde{t}_i)$ . Тогда  $\mathcal{G}(\tilde{x}_i, \tilde{u}_i) = 0$ . В постановке исходной задачи ОУ последнее равенство занимает особое место и принципиально отличается от ограничений (3.6.3), (3.6.4). Именно оно вносит специфику разрывности в рассматриваемую задачу. В процессе оптимизации не допускается нарушения этого условия. Поэтому управление на шаге интегрирования будем считать постоянным и равным значению  $u_{i+1}$ . Таким образом, равенство  $\mathcal{G}(\tilde{x}_i, \tilde{u}_i) = 0$  можно переписать в следующем виде:  $\mathcal{G}(\tilde{x}_i, u_{i+1}) = 0$ . В этом случае на  $i$ -м шаге интегрирование дифференциальных уравнений ведется по следующим формулам.

**Случай А.** Если  $\mathcal{G}(x_i, u_i) < 0$ ,  $\mathcal{G}(x_{i+1}, u_{i+1}) \geq 0$ , то

$$\tilde{x}_i = x_i + a_i f_-(\tilde{x}_i, u_{i+1}), \quad x_{i+1} = \tilde{x}_i + (h_i - a_i) f_+(x_{i+1}, u_{i+1}). \quad (3.6.7)$$

**Случай Б.** Если  $\mathcal{G}(x_i, u_i) \geq 0$ ,  $\mathcal{G}(x_{i+1}, u_{i+1}) < 0$ , то

$$\tilde{x}_i = x_i + a_i f_+(\tilde{x}_i, u_{i+1}), \quad x_{i+1} = \tilde{x}_i + (h_i - a_i) f_-(x_{i+1}, u_{i+1}). \quad (3.6.8)$$

И в первом, и во втором случаях система уравнений ((3.6.7) или (3.6.8)) состоит из  $2m$  уравнений и содержит  $(2m + 1)$  неизвестных:  $x_{i+1}^1, \dots, x_{i+1}^m, \tilde{x}_{i+1}^1, \dots, \tilde{x}_{i+1}^m, a_i$ . Для их определения используется еще одно условие:

$$\mathcal{G}(\tilde{x}_i, u_{i+1}) = 0. \quad (3.6.9)$$

Допустим, что имеет место **Случай А.** Учитывая, что  $\tilde{x}_i = x_i + a_i f_-(\tilde{x}_i, u_{i+1})$ , условие (3.6.9) записываем в виде

$$\mathcal{G}(x_i + a_i f_-(\tilde{x}_i, u_{i+1}), u_{i+1}) = 0.$$

Из этого выражения видно, что величина  $a_i$  неявно зависит от векторов  $x_i$ ,  $\tilde{x}_i$  и  $u_{i+1}$ . В качестве начального приближения вектора  $\tilde{x}_i$  берем  $\tilde{x}_i = x_i + a_i f_-(x_i, u_{i+1})$ . Для получения начального приближения величины  $a_i$  можно воспользоваться разложением функции  $\mathcal{G}(x_i + a_i f_-(x_i, u_{i+1}), u_{i+1})$  в ряд по степеням  $a_i$ . В результате получим

$$a_i = -\frac{\mathcal{G}(x_i, u_{i+1})}{\mathcal{G}_x^\top(x_i, u_{i+1}) f_-(x_i, u_{i+1})}.$$

Исходя из этих начальных приближений решается нелинейная система из  $(m + 1)$  уравнений

$$\tilde{x}_i = x_i + a_i f_-(\tilde{x}_i, u_{i+1}), \quad \mathcal{G}(x_i + a_i f_-(\tilde{x}_i, u_{i+1}), u_{i+1}) = 0$$

и определяются компоненты вектора  $\tilde{x}_i$  и величина  $a_i$ .

Начальное приближение вектора  $x_{i+1}$  определяем аналогично, по схеме  $x_{i+1} = \tilde{x}_i + (h_i - a_i) f_+(\tilde{x}_i, u_{i+1})$ , после чего решаем нелинейную систему из  $m$  уравнений  $x_{i+1} = \tilde{x}_i + (h_i - a_i) f_+(x_{i+1}, u_{i+1})$  и определяем компоненты вектора  $x_{i+1}$ . Точно так же поступаем и во втором случае, когда решается система (3.6.8), (3.6.9).

В дальнейшем нам понадобятся производные  $\partial a_i / \partial x_i$ ,  $\partial a_i / \partial \tilde{x}_i$ ,  $\partial a_i / \partial u_{i+1}$ , которые вычисляются из условия  $\mathcal{G}(\tilde{x}_i, u_{i+1}) = 0$ . Пусть

$$\varphi(\tilde{x}_i, u_{i+1}) = \begin{cases} f_-(\tilde{x}_i, u_{i+1}), & \text{если } \mathcal{G}(x_i, u_i) < 0, \quad \mathcal{G}(x_{i+1}, u_{i+1}) \geq 0, \\ f_+(\tilde{x}_i, u_{i+1}), & \text{если } \mathcal{G}(x_i, u_i) \geq 0, \quad \mathcal{G}(x_{i+1}, u_{i+1}) < 0. \end{cases}$$

Учитывая, что  $\tilde{x}_i = x_i + a_i \varphi(\tilde{x}_i, u_{i+1})$ , имеем

$$\mathcal{G}(\tilde{x}_i, u_{i+1}) = \mathcal{G}(x_i + a_i \varphi(\tilde{x}_i, u_{i+1}), u_{i+1}) = 0.$$

Отсюда, воспользовавшись теоремой о неявной функции, получим

$$\begin{aligned}\frac{\partial a_i}{\partial x_i} &= -\frac{\mathcal{G}_x(\tilde{x}_i, u_{i+1})}{\mathcal{G}_x^\top(\tilde{x}_i, u_{i+1})\varphi(\tilde{x}_i, u_{i+1})}, \\ \frac{\partial a_i}{\partial \tilde{x}_i} &= -\frac{a_i\varphi_x^\top(\tilde{x}_i, u_{i+1})\mathcal{G}_x(\tilde{x}_i, u_{i+1})}{\mathcal{G}_x^\top(\tilde{x}_i, u_{i+1})\varphi(\tilde{x}_i, u_{i+1})}, \\ \frac{\partial a_i}{\partial u_{i+1}} &= -\frac{\mathcal{G}_u(\tilde{x}_i, u_{i+1}) + a_i\varphi_u^\top(\tilde{x}_i, u_{i+1})\mathcal{G}_x(\tilde{x}_i, u_{i+1})}{\mathcal{G}_x^\top(\tilde{x}_i, u_{i+1})\varphi(\tilde{x}_i, u_{i+1})},\end{aligned}$$

где  $i = 1, \dots, k-1$ .

В последних трех формулах выражение в знаменателе должно быть отлично от нуля. Это и означает, что траектория протыкает поверхность  $\mathcal{G}$ , а не касается ее.

В соответствии с указанным выше алгоритмом решения задачи (3.6.1) при  $i = 2, 3, \dots, k-1$  возможны следующие восемь случаев.

1. Если  $\mathcal{G}(x_{i-1}, u_{i-1}) < 0$ ,  $\mathcal{G}(x_i, u_i) < 0$ ,  $\mathcal{G}(x_{i+1}, u_{i+1}) < 0$ , то  $x_i$  и  $u_i$  входят в правую часть двух следующих уравнений:

$$x_i = x_{i-1} + h_{i-1}f_-(x_i, u_i) \quad \text{и} \quad x_{i+1} = x_i + h_i f_-(x_{i+1}, u_{i+1}).$$

Согласно формулам (1.3.10) и (1.3.11) имеем

$$p_i = p_{i+1} + h_{i-1} \frac{\partial f_-^\top(x_i, u_i)}{\partial x} p_i + J_{x_i}, \quad \frac{d\Omega}{du_i} = h_{i-1} \frac{\partial f_-^\top(x_i, u_i)}{\partial u} p_i + J_{u_i}.$$

2. Если  $\mathcal{G}(x_{i-1}, u_{i-1}) \geq 0$ ,  $\mathcal{G}(x_i, u_i) \geq 0$ ,  $\mathcal{G}(x_{i+1}, u_{i+1}) \geq 0$ , то имеем

$$p_i = p_{i+1} + h_{i-1} \frac{\partial f_+^\top(x_i, u_i)}{\partial x} p_i + J_{x_i}, \quad \frac{d\Omega}{du_i} = h_{i-1} \frac{\partial f_+^\top(x_i, u_i)}{\partial u} p_i + J_{u_i}.$$

3. Если  $\mathcal{G}(x_{i-1}, u_{i-1}) \geq 0$ ,  $\mathcal{G}(x_i, u_i) < 0$ ,  $\mathcal{G}(x_{i+1}, u_{i+1}) < 0$ , то  $x_i$  и  $u_i$  входят в правую часть следующих соотношений:

$$\begin{aligned}\tilde{x}_{i-1} &= x_{i-1} + a_{i-1}f_+(\tilde{x}_{i-1}, u_i), \\ x_i &= \tilde{x}_{i-1} + (h_{i-1} - a_{i-1})f_-(x_i, u_i), \\ x_{i+1} &= x_i + h_i f_-(x_{i+1}, u_{i+1}).\end{aligned}$$

Так как при вычислении фазового вектора в  $i$ -м узле введена дополнительная точка  $\tilde{x}_{i-1}$ , то в процессе расчетов импульсов введем дополнительный импульс  $\tilde{p}_{i-1}$ , соответствующий вектору  $\tilde{x}_{i-1}$  и используемый для вычисления компоненты  $d\Omega/du_i$ . Тогда

$$\begin{aligned}p_i &= p_{i+1} + (h_{i-1} - a_{i-1}) \frac{\partial f_-^\top(x_i, u_i)}{\partial x} p_i + J_{x_i}, \\ \tilde{p}_{i-1} &= f_+^\top(\tilde{x}_{i-1}, u_i) \tilde{p}_{i-1} \frac{\partial a_{i-1}}{\partial \tilde{x}_{i-1}} + a_{i-1} \frac{\partial f_+^\top(\tilde{x}_{i-1}, u_i)}{\partial x} \tilde{p}_{i-1} + p_i - f_-^\top(x_i, u_i) p_i \frac{\partial a_{i-1}}{\partial \tilde{x}_{i-1}}, \\ \frac{d\Omega}{du_i} &= f_+^\top(\tilde{x}_{i-1}, u_i) \tilde{p}_{i-1} \frac{\partial a_{i-1}}{\partial u_i} + a_{i-1} \frac{\partial f_+^\top(\tilde{x}_{i-1}, u_i)}{\partial u} \tilde{p}_{i-1} - \\ &- f_-^\top(x_i, u_i) p_i \frac{\partial a_{i-1}}{\partial u_i} + (h_{i-1} - a_{i-1}) \frac{\partial f_-^\top(x_i, u_i)}{\partial u} p_i + J_{u_i}.\end{aligned}$$

4. Если  $\mathcal{G}(x_{i-1}, u_{i-1}) < 0$ ,  $\mathcal{G}(x_i, u_i) \geq 0$ ,  $\mathcal{G}(x_{i+1}, u_{i+1}) \geq 0$ , то имеем

$$\begin{aligned} p_i &= p_{i+1} + (h_{i-1} - a_{i-1}) \frac{\partial f_+^\top(x_i, u_i)}{\partial x} p_i + J_{x_i}, \\ \tilde{p}_{i-1} &= f_-^\top(\tilde{x}_{i-1}, u_i) \tilde{p}_{i-1} \frac{\partial a_{i-1}}{\partial \tilde{x}_{i-1}} + a_{i-1} \frac{\partial f_-^\top(\tilde{x}_{i-1}, u_i)}{\partial x} \tilde{p}_{i-1} + p_i - f_+^\top(x_i, u_i) p_i \frac{\partial a_{i-1}}{\partial \tilde{x}_{i-1}}, \\ \frac{d\Omega}{du_i} &= f_-^\top(\tilde{x}_{i-1}, u_i) \tilde{p}_{i-1} \frac{\partial a_{i-1}}{\partial u_i} + a_{i-1} \frac{\partial f_-^\top(\tilde{x}_{i-1}, u_i)}{\partial u_i} \tilde{p}_{i-1} - \\ &\quad - f_+^\top(x_i, u_i) p_i \frac{\partial a_{i-1}}{\partial u_i} + (h_{i-1} - a_{i-1}) \frac{\partial f_+^\top(x_i, u_i)}{\partial u_i} p_i + J_{u_i}. \end{aligned}$$

5. Если  $\mathcal{G}(x_{i-1}, u_{i-1}) < 0$ ,  $\mathcal{G}(x_i, u_i) < 0$ ,  $\mathcal{G}(x_{i+1}, u_{i+1}) \geq 0$ , то

$$\begin{aligned} \tilde{p}_i &= p_{i+1} - f_+^\top(x_{i+1}, u_{i+1}) p_{i+1} \frac{\partial a_i}{\partial \tilde{x}_i} + f_-^\top(\tilde{x}_i, u_{i+1}) \tilde{p}_i \frac{\partial a_i}{\partial \tilde{x}_i} + a_i \frac{\partial f_-^\top(\tilde{x}_i, u_{i+1})}{\partial x} \tilde{p}_i, \\ p_i &= h_{i-1} \frac{\partial f_-^\top(x_i, u_i)}{\partial x} p_i + \tilde{p}_i + f_-^\top(\tilde{x}_i, u_{i+1}) \tilde{p}_i \frac{\partial a_i}{\partial x_i} - f_+^\top(x_{i+1}, u_{i+1}) p_{i+1} \frac{\partial a_i}{\partial x_i} + J_{x_i}, \\ \frac{d\Omega}{du_i} &= h_{i-1} \frac{\partial f_-^\top(x_i, u_i)}{\partial u_i} p_i + J_{u_i}. \end{aligned}$$

6. Если  $\mathcal{G}(x_{i-1}, u_{i-1}) \geq 0$ ,  $\mathcal{G}(x_i, u_i) \geq 0$ ,  $\mathcal{G}(x_{i+1}, u_{i+1}) < 0$ , то имеем

$$\begin{aligned} \tilde{p}_i &= p_{i+1} - f_-^\top(x_{i+1}, u_{i+1}) p_{i+1} \frac{\partial a_i}{\partial \tilde{x}_i} + f_+^\top(\tilde{x}_i, u_{i+1}) \tilde{p}_i \frac{\partial a_i}{\partial \tilde{x}_i} + a_i \frac{\partial f_+^\top(\tilde{x}_i, u_{i+1})}{\partial x} \tilde{p}_i, \\ p_i &= h_{i-1} \frac{\partial f_+^\top(x_i, u_i)}{\partial x} p_i + \tilde{p}_i + f_+^\top(\tilde{x}_i, u_{i+1}) \tilde{p}_i \frac{\partial a_i}{\partial x_i} - f_-^\top(x_{i+1}, u_{i+1}) p_{i+1} \frac{\partial a_i}{\partial x_i} + J_{x_i}, \\ \frac{d\Omega}{du_i} &= h_{i-1} \frac{\partial f_+^\top(x_i, u_i)}{\partial u_i} p_i + J_{u_i}. \end{aligned}$$

7. Пусть  $\mathcal{G}(x_{i-1}, u_{i-1}) < 0$ ,  $\mathcal{G}(x_i, u_i) \geq 0$ ,  $\mathcal{G}(x_{i+1}, u_{i+1}) < 0$ . Тогда при выводе формул для  $p_i$  и  $d\Omega/du_i$  используем соотношения

$$\begin{aligned} \tilde{x}_{i-1} &= x_{i-1} + a_{i-1} f_-(\tilde{x}_{i-1}, u_i), & x_i &= \tilde{x}_{i-1} + (h_{i-1} - a_{i-1}) f_+(x_i, u_i), \\ \tilde{x}_i &= x_i + a_i f_+(\tilde{x}_i, u_{i+1}), & x_{i+1} &= \tilde{x}_i + (h_i - a_i) f_-(x_{i+1}, u_{i+1}) \end{aligned}$$

и вычисляем значения двух вспомогательных импульсов  $\tilde{p}_{i-1}$  и  $\tilde{p}_i$ :

$$\begin{aligned} \tilde{p}_i &= f_+^\top(\tilde{x}_i, u_{i+1}) \tilde{p}_i \frac{\partial a_i}{\partial \tilde{x}_i} + a_i \frac{\partial f_+^\top(\tilde{x}_i, u_{i+1})}{\partial x} \tilde{p}_i + p_{i+1} - f_-^\top(x_{i+1}, u_{i+1}) p_{i+1} \frac{\partial a_i}{\partial \tilde{x}_i}, \\ p_i &= (h_{i-1} - a_{i-1}) \frac{\partial f_+^\top(x_i, u_i)}{\partial x} p_i + \tilde{p}_i + f_+^\top(\tilde{x}_i, u_{i+1}) \tilde{p}_i \frac{\partial a_i}{\partial x_i} - \\ &\quad - f_-^\top(x_{i+1}, u_{i+1}) p_{i+1} \frac{\partial a_i}{\partial x_i} + J_{x_i}, \\ \tilde{p}_{i-1} &= f_-^\top(\tilde{x}_{i-1}, u_i) \tilde{p}_{i-1} \frac{\partial a_{i-1}}{\partial \tilde{x}_{i-1}} + a_{i-1} \frac{\partial f_-^\top(\tilde{x}_{i-1}, u_i)}{\partial x} \tilde{p}_{i-1} - f_+^\top(x_i, u_i) p_i \frac{\partial a_{i-1}}{\partial \tilde{x}_{i-1}} + p_i, \\ \frac{d\Omega}{du_i} &= f_-^\top(\tilde{x}_{i-1}, u_i) \tilde{p}_{i-1} \frac{\partial a_{i-1}}{\partial u_i} + a_{i-1} \frac{\partial f_-^\top(\tilde{x}_{i-1}, u_i)}{\partial u_i} \tilde{p}_{i-1} - \\ &\quad - f_+^\top(x_i, u_i) p_i \frac{\partial a_{i-1}}{\partial u_i} + (h_{i-1} - a_{i-1}) \frac{\partial f_+^\top(x_i, u_i)}{\partial u_i} p_i + J_{u_i}. \end{aligned}$$

8. Если  $\mathcal{G}(x_{i-1}, u_{i-1}) \geq 0$ ,  $\mathcal{G}(x_i, u_i) < 0$ ,  $\mathcal{G}(x_{i+1}, u_{i+1}) \geq 0$ , то приходим к аналогичным формулам:

$$\begin{aligned}
\tilde{p}_{i-1} &= f_+^\top(\tilde{x}_{i-1}, u_i) \tilde{p}_{i-1} \frac{\partial a_{i-1}}{\partial \tilde{x}_{i-1}} + a_{i-1} \frac{\partial f_+^\top(\tilde{x}_{i-1}, u_i)}{\partial x} \tilde{p}_{i-1} - f_-^\top(x_i, u_i) p_i \frac{\partial a_{i-1}}{\partial \tilde{x}_{i-1}} + p_i, \\
\tilde{p}_i &= f_-^\top(\tilde{x}_i, u_{i+1}) \tilde{p}_i \frac{\partial a_i}{\partial \tilde{x}_i} + a_i \frac{\partial f_-^\top(\tilde{x}_i, u_{i+1})}{\partial x} \tilde{p}_i + p_{i+1} - f_+^\top(x_{i+1}, u_{i+1}) p_{i+1} \frac{\partial a_i}{\partial \tilde{x}_i}, \\
p_i &= (h_{i-1} - a_{i-1}) \frac{\partial f_-^\top(x_i, u_i)}{\partial x} p_i + \tilde{p}_i + f_-^\top(\tilde{x}_i, u_{i+1}) \tilde{p}_i \frac{\partial a_i}{\partial x_i} - \\
&\quad - f_+^\top(x_{i+1}, u_{i+1}) p_{i+1} \frac{\partial a_i}{\partial x_i} + J_{x_i}, \\
\frac{d\Omega}{du_i} &= f_+^\top(\tilde{x}_{i-1}, u_i) \tilde{p}_{i-1} \frac{\partial a_{i-1}}{\partial u_i} + a_{i-1} \frac{\partial f_+^\top(\tilde{x}_{i-1}, u_i)}{\partial u} \tilde{p}_{i-1} - \\
&\quad - f_-^\top(x_i, u_i) p_i \frac{\partial a_{i-1}}{\partial u_i} + (h_{i-1} - a_{i-1}) \frac{\partial f_+^\top(x_i, u_i)}{\partial u} p_i + J_{u_i}.
\end{aligned}$$

Пусть теперь  $i = k$ . Тогда возможны следующие четыре случая.

1. Если  $\mathcal{G}(x_{k-1}, u_{k-1}) < 0$ ,  $\mathcal{G}(x_k, u_k) < 0$ , то

$$p_k = h_{k-1} \frac{\partial f_-^\top(x_k, u_k)}{\partial x} p_k + J_{x_k}, \quad \frac{d\Omega}{du_k} = h_{k-1} \frac{\partial f_-^\top(x_k, u_k)}{\partial u} p_k + J_{u_k}.$$

2. Если  $\mathcal{G}(x_{k-1}, u_{k-1}) \geq 0$ ,  $\mathcal{G}(x_k, u_k) \geq 0$ , имеем

$$p_k = h_{k-1} \frac{\partial f_+^\top(x_k, u_k)}{\partial x} p_k + J_{x_k}, \quad \frac{d\Omega}{du_k} = h_{k-1} \frac{\partial f_+^\top(x_k, u_k)}{\partial u} p_k + J_{u_k}.$$

3. При  $\mathcal{G}(x_{k-1}, u_{k-1}) < 0$ ,  $\mathcal{G}(x_k, u_k) \geq 0$  получаем

$$\begin{aligned}
p_k &= (h_{k-1} - a_{k-1}) \frac{\partial f_+^\top(x_k, u_k)}{\partial x} p_k + J_{x_k}, \\
\tilde{p}_{k-1} &= p_k + f_-^\top(\tilde{x}_{k-1}, u_k) \tilde{p}_{k-1} \frac{\partial a_{k-1}}{\partial \tilde{x}_{k-1}} + a_{k-1} \frac{\partial f_-^\top(\tilde{x}_{k-1}, u_k)}{\partial x} \tilde{p}_{k-1} - \\
&\quad - f_+^\top(x_k, u_k) p_k \frac{\partial a_{k-1}}{\partial \tilde{x}_{k-1}}, \\
\frac{d\Omega}{du_k} &= f_-^\top(\tilde{x}_{k-1}, u_k) \tilde{p}_{k-1} \frac{\partial a_{k-1}}{\partial u_k} + a_{k-1} \frac{\partial f_-^\top(\tilde{x}_{k-1}, u_k)}{\partial u} \tilde{p}_{k-1} + \\
&\quad + (h_{k-1} - a_{k-1}) \frac{\partial f_+^\top(x_k, u_k)}{\partial u} p_k - f_+^\top(x_k, u_k) p_k \frac{\partial a_{k-1}}{\partial u_k} + J_{u_k}.
\end{aligned}$$

4. Если  $\mathcal{G}(x_{k-1}, u_{k-1}) \geq 0$ ,  $\mathcal{G}(x_k, u_k) < 0$ , то

$$\begin{aligned}
p_k &= (h_{k-1} - a_{k-1}) \frac{\partial f_-^\top(x_k, u_k)}{\partial x} p_k + J_{x_k}, \\
\tilde{p}_{k-1} &= p_k + f_+^\top(\tilde{x}_{k-1}, u_k) \tilde{p}_{k-1} \frac{\partial a_{k-1}}{\partial \tilde{x}_{k-1}} + a_{k-1} \frac{\partial f_+^\top(\tilde{x}_{k-1}, u_k)}{\partial x} \tilde{p}_{k-1} - \\
&\quad - f_-^\top(x_k, u_k) p_k \frac{\partial a_{k-1}}{\partial \tilde{x}_{k-1}}, \\
\frac{d\Omega}{du_k} &= f_+^\top(\tilde{x}_{k-1}, u_k) \tilde{p}_{k-1} \frac{\partial a_{k-1}}{\partial u_k} + a_{k-1} \frac{\partial f_+^\top(\tilde{x}_{k-1}, u_k)}{\partial u} \tilde{p}_{k-1} + \\
&\quad + (h_{k-1} - a_{k-1}) \frac{\partial f_-^\top(x_k, u_k)}{\partial u} p_k - f_-^\top(x_k, u_k) p_k \frac{\partial a_{k-1}}{\partial u_k} + J_{u_k}.
\end{aligned}$$

Все расчеты проводятся следующим образом. Вычисляются импульсы  $p_i$  и  $\tilde{p}_i$ ,  $i = k, \dots, 2$  (вектор  $p_1$  не используется при расчете градиента функционала). На тех интервалах, где нет разрыва правой части, считаем, что  $\tilde{p}_i = 0$ . Компоненты вектора градиента  $d\Omega/du$  вычисляются, используя уже известные значения для импульсов.

### 3.7. Численные методы решения задач оптимального управления

Численным методам решения задач оптимального управления посвящена обширная литература. Некоторые публикации были упомянуты выше во введении. Здесь мы кратко остановимся на подходе, основанном на использовании принципа максимума Л.С. Понтрягина, сводящего задачу ОУ к решению краевой задачи для систем обыкновенных дифференциальных уравнений. Более подробно будет рассмотрен основной подход к решению задач ОУ, основанный на редукции к задачам нелинейного программирования и использующий технику быстрого автоматического дифференцирования, упомянутый в первой главе.

Пусть управляемый процесс описывается системой (3.3.1). Составим сопряженную систему (3.3.4). Попытаемся на основе принципа максимума решить простейшую задачу Майера. При фиксированном наборе  $x_*(t)$ ,  $p_*(t)$  из условия (3.3.6) найдем точно-множественное отображение

$$u_* = \theta(t, x_*(t), p_*(t)), \quad (3.7.1)$$

на котором функция Понтрягина  $H(t, x_*, u, p_*)$  достигает минимума по  $u \in U$ . После подстановки (3.7.1) в уравнения (3.3.1) и (3.3.4) получим систему из  $2n$  обыкновенных дифференциальных уравнений для  $n$ -мерных вектор-функций  $x_*(t)$ ,  $p_*(t)$ :

$$\begin{aligned} \dot{x}_* &= f(t, x_*, \theta(x_*, t, p_*)), \\ \dot{p}_* &= -f_x^\top(t, x_*, \theta(x_*, t, p_*))p_*. \end{aligned} \quad (3.7.2)$$

Для этой системы  $n$  условий (3.3.2) задано в начальный момент  $T_1$  (на левом конце траектории) и  $n$  условий в конечный момент  $T_2$  (на правом конце траектории):

$$x_*(T_1) = x_1, \quad p_*(T_2) = W_x(x_*(T_2)). \quad (3.7.3)$$

Таким образом, решение исходной задачи формально сведено к решению краевой задачи для системы из  $2n$  дифференциальных уравнений. Если решить эту задачу, т.е. определить функции  $x_*(t)$ ,  $p_*(t)$ , удовлетворяющие системе (3.7.2) и условиям (3.7.3), то оптимальное управление  $u_*(t)$  для исходной задачи должно находиться среди функций, представимых в виде (3.7.1). Описанный прием весьма удобен, если делается попытка решить исходную задачу аналитически. Для численных расчетов он оказался приемлемым только на последнем, заключительном этапе расчетов. Дело в том, что краевая задача обычно оказывается существенно нелинейной. Отображения  $\theta$  часто бывают многозначными; это приводит к тому, что правые части системы (3.3.1) становятся негладкими либо даже неоднозначными. Краевая задача не всегда обладает единственным решением, стандартные процедуры ее решения часто оказываются расходящимися. В лучшем случае для такой задачи можно определить хорошее начальное приближение, и только после этого появляется реальная возможность решения поставленной краевой задачи. Поэтому на начальном этапе расчетов обычно используются методы, допускающие большой произвол в выборе начального управления (например, метод штрафов, метод параметризации целевой функции, метод модифицированных функций Лагранжа и т.д.).

Иногда указывается еще один недостаток, возникающий при применении принципа максимума для численного решения задач ОУ. Он состоит в том, что принцип максимума не



учитывает фазовые ограничения. Это возражение снимается в тех случаях, когда применяемый метод сводит решение задачи к последовательной безусловной минимизации вспомогательных функций (например, метод штрафов, модифицированных функций Лагранжа и т.д.). Тогда минимизацию вспомогательных функций можно проводить, используя принцип максимума Л.С. Понтрягина.

Базисная задача ОУ может быть записана следующим образом. Управляемый процесс описывается системой обыкновенных дифференциальных уравнений:

$$\frac{dx}{dt} = f(t, x, u, \xi), \quad T_1 \leq t \leq T_2, \quad (3.7.4)$$

где вектор состояния  $x \in \mathbb{R}^n$ , управлением является кусочно-непрерывная функция  $u$ , принимающая значения из допустимого множества  $U \subset \mathbb{R}^r$ .

Начальное условие в (3.7.4) фиксировано, поэтому  $x(t)$  будет обозначать решение дифференциального уравнения (3.7.4) с начальным условием  $x(T_1) = x_1$ . Считаем, что  $U$  — компактное множество в пространстве  $\mathbb{R}^r$ . Вектор параметров  $\xi \in V \subset \mathbb{R}^s$ . Если выбор  $T_1$ ,  $T_2$  и начального состояния  $x_1$  следует оптимизировать, то мы включаем их в вектор  $\xi$ . Задача ставится следующим образом: найти управление  $u(t) \in U$  и вектор параметров  $\xi \in V$ , которые минимизируют функционал  $W(x(T_2), \xi)$ , при наличии “смешанных” ограничений на векторы состояния, управления и вектор параметров:

$$g(t, x(t), u(t), \xi) = 0_\alpha, \quad q(t, x(t), u(t), \xi) \leq 0_\beta, \quad T_1 \leq t \leq T_2.$$

От непрерывной системы (3.7.4) перейдем к дискретной системе. Воспользуемся самой простой дискретной аппроксимацией системы (3.7.4), которая задается формулой Эйлера

$$x_{i+1} = x_i + h_i f(t_i, x_i, u_i, \xi) = F(t_i, x_i, u_i, \xi), \quad (3.7.5)$$

где  $\sum_{i=1}^{k-1} h_i = T_2 - T_1$ ,  $0 < h_i$ ,  $t_1 = T_1$ ,  $t_k = T_2$ ,  $t_{i+1} = t_i + h_i$ ,  $1 \leq i \leq k-1$ . При этом непрерывное управление  $u(t)$  заменяется кусочно-постоянным, ограничения вдоль траектории (3.7.5) учитываются только в точках дискретизации:

$$g(t_i, x_i, u_i, \xi) = 0_\alpha, \quad q(t_i, x_i, u_i, \xi) \leq 0_\beta, \quad 1 \leq i \leq k \quad (3.7.6)$$

и в конечный момент движения (3.7.6).

Полный фазовый вектор  $x^\top = [x_1^\top, \dots, x_k^\top]$  является функцией полного вектора управлений  $u^\top = [u_1^\top, u_2^\top, \dots, u_{k-1}^\top]$  и вектора параметров  $\xi$ . Эту зависимость, определяемую формулой (3.7.5), обозначим через  $x(u, \xi)$ . Функционал исходной задачи  $W(x(T_2))$  заменяется функцией  $W(x_k)$ , где  $x_k$  есть сложная функция  $u$  и  $\xi$ . Каждому  $i$ -му шагу процесса (3.7.5) ставим в соответствие вектор  $p_i$ . Вспомогательная функция  $E$  имеет вид

$$E(t, x, u, \xi, p) = W(x_k, \xi) + \sum_{i=1}^{k-1} F^\top(t_i, x_i, u_i, \xi) p_{i+1}.$$

Дискретный вариант задачи состоит в минимизации  $W(x_k, \xi)$  по  $u_i \in U$ ,  $1 \leq i < k$ , и  $\xi \in V$  при наличии смешанных ограничений (3.7.6). Чтобы применить градиентные методы НЛП, необходимо найти формулы дифференцирования целевой функции. Применяя результаты § 1.3, находим, что все векторы  $p_i \in \mathbb{R}^n$  и производные сложной функции  $\Omega(u, \xi) = W(x_k, \xi)$

вычисляются по формулам

$$\left\{ \begin{array}{l} p_i = p_{i+1} + h_i f_{x_i}^\top(t_i, x_i, u_i, \xi) p_{i+1}, \quad p_k = W_{x_k}(x_k, \xi), \quad 1 \leq i \leq k-1, \\ \frac{d\Omega}{du_i} = h_i f_{u_i}^\top(t_i, x_i, u_i, \xi) p_{i+1}, \quad 1 \leq i \leq k-1, \quad \frac{d\Omega}{du_k} = 0, \\ \frac{d\Omega}{d\xi} = W_\xi(x_k, \xi) + \sum_{i=1}^{k-1} h_i f_\xi^\top(t_i, x_i, u_i, \xi) p_{i+1}. \end{array} \right. \quad (3.7.7)$$

Описанный подход редукции исходной задачи ОУ к конечномерной дает возможность использовать разнообразные, хорошо изученные и отработанные численные методы НЛП. Среди многочисленных методов НЛП наиболее популярным, часто используемым для практических расчетов, является метод штрафных функций. Реже используется метод модифицированных функций Лагранжа, метод проекции градиента, линеаризации, метод возможных направлений, методы внутренней точки и т.д. Опыт численных расчетов показывает, что не существует универсальных методов оптимизации. Поэтому обычно используется многометодный режим расчетов [50], когда в процессе решения задачи выбирается наиболее подходящая последовательность методов. Вначале расчетов часто применяется метод внешних штрафных функций, так как этот метод имеет, как правило, наибольшую область сходимости, и только после нахождения хорошего приближения, определения множителей Лагранжа, производится переход к другим, более "тонким" методам расчета.

Для решения дискретных задач методами НЛП обычно вводятся дополнительные функции (например, штрафная функция, модифицированная функция Лагранжа и т.д.). Выражения для вычисления их производных подобны (3.7.7): вместо  $W$  следует подставить эти функции и учесть все отличные от нуля производные по  $u_i, u_k, x_i, x_k$ .

Обычно в методах НЛП задача сводится к последовательной безусловной минимизации нелинейных функций многих переменных. В методе штрафов, например, ищется минимум  $E$  по полному фазовому вектору  $u$  и по вектору параметров  $\xi$ . Если используется дифференцируемая внешняя квадратичная функция штрафа с коэффициентом штрафа  $\gamma$ , то функция  $E$  определяется следующим образом:

$$\begin{aligned} E = & W(x_k, \xi) + \sum_{i=1}^{k-1} F^\top(t_i, x_i, u_i, \xi) p_{i+1} + \\ & + \gamma \sum_{i=1}^k [\|g(t_i, x_i, u_i, \xi)\|^2 + \|q_+(t_i, x_i, u_i, \xi)\|^2]. \end{aligned} \quad (3.7.8)$$

Безусловную минимизацию функции  $E$  можно осуществлять методами, не использующими вычисления производных. Например, реализуя покоординатный спуск, мы приходим к методу локальных вариаций [91]. Более эффективными оказались градиентные методы. Дело в том, что в отличие от общей задачи НЛП здесь сравнительно просто вычисляются градиенты. Для этого достаточно просчитать один раз уравнения для импульсов и по формуле (3.7.7) определить все требуемые производные. Именно в этом смысле надо понимать утверждение о том, что задачи НЛП, полученные при дискретизации задач ОУ, намного проще задач НЛП общего вида.

С помощью описанного подхода были решены многочисленные практические задачи, созданы пакеты программ для решения задач ОУ [35, 38, 49, 54, 55, 107, 114]. По существу, это те же самые пакеты программ решения задач НЛП, но с добавлением процедур формирования вспомогательных функций типа (3.7.8) и вычисления их производных по формулам (3.7.7) или по более сложным формулам, разработанным для более точных схем интегрирования системы (3.7.4) (метод Рунге–Кутты, Адамса и т.д.).

Высокую скорость сходимости и большую точность расчетов можно получить, используя методы второго порядка. Например, в [50] описано решение простейшей модельной задачи ОУ с фазовыми ограничениями с помощью метода Ньютона. Там же приведены результаты расчетов боевого разворота самолета за кратчайшее время с одновременным изменением высоты полета с учетом ограничений на перегрузку и скорость изменения управлений (тяги двигателя, угла атаки и угла крена). В [50] также даны результаты проектирования шпангоута минимального веса, выдерживающего заданные нагрузки при напряжениях, не превосходящих допустимые.

Переходя к пределу при  $h_i \rightarrow 0$ ,  $k \rightarrow \infty$ , из (3.7.7) находим, что функция  $p(t)$  удовлетворяет следующим дифференциальным уравнениям:

$$\dot{p} = -f^T_x(t, x, u, \xi)p, \quad p(T_2) = W_x(x(T_2), \xi). \quad (3.7.9)$$

Мы приходим к сопряженному уравнению. Это уравнение интегрируется в обратном направлении от  $t = T_2$  до  $t = T_1$ .

Если система (3.7.4) жесткая, тогда целесообразно использовать неявные схемы интегрирования [80]. Например, применение неявной схемы Эйлера приводит к формуле

$$x_i = x_{i-1} + h_i f(t_i, x_i, u_i, \xi), \quad 2 \leq i \leq k. \quad (3.7.10)$$

Здесь приходится решать систему нелинейных уравнений на каждом шаге интегрирования, тем не менее шаг интегрирования  $h_i$  в (3.7.10) можно брать много больший, чем в (3.7.5). Из (1.3.10) получаются следующие дискретные сопряженные уравнения:

$$\begin{aligned} p_i &= p_{i+1} + h_i f^T_{x_i}(t_i, x_i, u_i, \xi)p_i, \quad 2 \leq i \leq k-1, \\ p_k &= W_{x_k}(x_k, \xi) + h_k f^T_{x_k}(t_k, x_k, u_k, \xi)p_k. \end{aligned}$$

Следовательно, все векторы  $p_i$  находятся из неявной системы линейных алгебраических уравнений, и формулы для вычисления градиента принимают вид

$$\frac{d\Omega}{du_i} = h_i f^T_{u_i}(t_i, x_i, u_i, \xi)p_i, \quad 2 \leq i \leq k, \quad \frac{d\Omega}{du_1} = 0.$$

В пределе при  $h_i \rightarrow 0$ ,  $k \rightarrow \infty$  мы получаем ту же сопряженную систему, что и (3.7.9).

В некоторых публикациях градиенты находятся из необходимых условий оптимальности путем дискретизации исходной и сопряженной систем. В этом случае, однако, могут возникнуть некоторые погрешности. Действительно, если мы одновременно по схеме Эйлера дискретизируем системы обыкновенных дифференциальных уравнений (3.7.4) и (3.7.9), то получим (3.7.5) и

$$p_{i+1} = p_i - h_i f^T_{x_i}(t_i, x_i, u_i, \xi)p_i, \quad 1 \leq i \leq k-1.$$

Тогда искомая производная будет иметь вид

$$\frac{d\Omega}{du_i} = h_i f^T_{u_i}(t_i, x_i, u_i, \xi)p_i.$$

Эти выражения не совпадают с (3.7.7) и, следовательно, градиент, полученный по этим формулам, неверен. Если шаг интегрирования  $h_i$  мал, то различие между этим и точным выражением (3.7.7) является величиной порядка  $O(h_i^2)$ . Но погрешность в вычислении градиента нежелательна, особенно в случаях, когда используются чувствительные алгоритмы минимизации или когда величина шага  $h_i$  недостаточно мала.

Подчеркнем следующий важный вывод: когда мы имеем дело с задачей ОУ, то дискретизация сопряженного уравнения должна соответствовать схеме интегрирования исходной

системы. Это не должно выполняться независимо. То же самое свойство имеет место для более сложных процессов, описываемых дифференциальными уравнениями с частными производными. Однако там это свойство не так очевидно, как в рассматриваемом случае обыкновенных дифференциальных уравнений. Мы остановимся на этом вопросе в следующей главе.

Применение современных методов глобальной оптимизации функций многих переменных открывает широкие возможности находить глобальные решения задач оптимального управления (см. [56, 108, 109]), учитывать частичную целочисленность (mixed-integer), как это сделано, например, в работе [57].

При решении линейных задач оптимального управления перспективным может оказаться метод ЛП-проекций, описанный в работах [28], [30–34] и предоставивший возможность решать задачи линейного программирования высокой размерности (с десятками миллионов неизвестных). Другой подход, основанный на итерационных расчетах, содержится в работах А.С. Антипина и его учеников (см. [13–15]), его также можно перенести на динамические задачи.

В [50, 52, 53] приведены формулы для вычисления градиентов целевой функции и ограничений для дискретизации по методам Эйлера и Рунге–Кутты. Там же приведены формулы для вторых производных и дан пример их применения для решения задачи ОУ со смешанными ограничениями методом Ньютона. Реализация формул (3.7.7), короткое описание пакета программ и некоторые численные примеры приведены в [52, 53].

Излагаемый в данной работе подход весьма универсальный. Он применяется для решения игровых задач, для нахождения синтеза управления. В последнем случае обычно строится приближенное синтезирующее управление  $u(t, x(t), \xi)$ , зависящее от вектора параметров  $\xi$ , по которому производится оптимизация.

### 3.8. Три простейших примера

Изложенный выше подход к решению задач оптимального управления использовался для решения многих практических задач. Результаты таких расчетов опубликованы во многочисленных статьях и отчетах. Упомянем книгу автора [50], где даны некоторые сведения об оптимизации динамики полета самолета и оптимизации его шпангоута. Здесь ограничимся лишь тремя простейшими примерами.

**Пример 3.8.1.** Задача об оптимальном параметрическом возбуждении.

Пусть управляемая система описывается следующим дифференциальным уравнением

$$d^2z/dt^2 + (1 - \varepsilon u)z = 0. \quad (3.8.1)$$

Здесь  $u$  — управление, подчиненное условию  $0 \leq u \leq 1$ . Из всех допустимых управлений  $u(t)$  на интервале  $T_1 \leq t \leq T_2$  следует найти такое, чтобы полная энергия  $W(t) = [z^2(t) + \dot{z}^2(t)]/2$  принимала в конце процесса при  $t = T_2$  наименьшее (наибольшее) значение.

Вместо уравнения второго порядка (3.8.1) введем два уравнения первого порядка для новых фазовых переменных  $x$  и  $y$ . С помощью стандартной замены переменных  $z = x \cos y$ ,  $dz/dt = -x \sin y$  сводим (3.8.1) к стандартному виду:

$$\dot{x} = -\varepsilon u x \cos y \sin y, \quad \dot{y} = 1 - \varepsilon u \cos^2 y, \quad W = x^2/2. \quad (3.8.2)$$

Отсюда следует, что исходная задача эквивалентна минимизации (максимизации) амплитуды  $x$ . Фазовым переменным  $x$  и  $y$  поставим в соответствие импульсы  $p$  и  $\lambda$ . Функция Понтрягина будет иметь вид

$$H(x, y, u, p, \lambda) = \lambda - \varepsilon u \cos^2 y [\lambda + x p \operatorname{tg} y]. \quad (3.8.3)$$

Выпишем уравнения для импульсов

$$\dot{p} = -H_x = \varepsilon u p \sin 2y/2, \quad (3.8.4)$$

$$\dot{\lambda} = -H_y = \varepsilon u [xp \cos 2y - \lambda \sin 2y]. \quad (3.8.5)$$

Краевые условия для системы (3.8.2), (3.8.4), (3.8.5) имеют вид

$$x(T_1) = x_1, \quad y(T_1) = y_1, \quad p(T_2) = 1, \quad \lambda(T_2) = 0. \quad (3.8.6)$$

Из (3.8.2) и (3.8.4) следует

$$\frac{dx}{dp} = -\frac{x}{p}, \quad x(t)p(t) = x(T_2), \quad T_1 \leq t \leq T_2.$$

Из принципа максимума получим, что при минимизации (максимизации) полной энергии  $W(x(T_2))$  управление определяется по формуле

$$u_* = \theta[\pm(\lambda + xp \operatorname{tg} y)],$$

где в выражении  $\pm$  верхний знак соответствует оптимальному увеличению энергии, а нижний – уменьшению. Здесь введена функция Хевисайда:  $\theta(\beta) = 0$  при  $\beta < 0$  и  $\theta(\beta) = 1$  при  $\beta > 0$ .

Подставим  $u_*$  в (3.8.3), воспользуемся (3.3.34), получим

$$H_* = \lambda - \varepsilon[\lambda \cos^2 y + xp \cos y \sin y] \theta[\pm(\lambda + xp \operatorname{tg} y)].$$

Считаем, что параметр  $\varepsilon$  неотрицателен и достаточно мал. Тогда в системах (3.8.2), (3.8.4) и (3.8.5) функции  $x(t)$ ,  $p(t)$ ,  $\lambda(t)$  будут медленно меняющимися, а  $y$  – быстро меняющейся. Для решения задачи используем асимптотический метод усреднения, согласно которому в первом приближении системы (3.8.2), (3.8.4), (3.8.5) можно заменить усредненной по  $y$  на периоде  $2\pi$ . Усредненное уравнение (3.8.5) имеет вид  $\dot{\lambda} = 0$ . Из (3.8.6) следует, что  $\lambda(t) = 0$ . Функция Понтрягина (3.8.3) упрощается:  $H_* = -\varepsilon u x p (\sin 2y)/2$ .

Для задачи минимизации  $W$  используем принцип максимума. При  $xp > 0$ , получим

$$u_* = \operatorname{Arg} \min_{0 \leq u \leq 1} (-u \sin 2y) = \theta(\sin 2y), \quad (3.8.7)$$

$$H_* = H(x, y, u_*, 0) = -\varepsilon x p \theta(\sin 2y)/2. \quad (3.8.8)$$

Другими словами, если решается задача уменьшения полной энергии, то  $H_* = u_* = 0$  при  $\sin 2y \leq 0$  и  $H_* = -\varepsilon x p/2$ ,  $u_* = 1$  при  $\sin 2y \geq 0$ . В задаче увеличения полной энергии имеем:  $H_* = \varepsilon x p/2$ ,  $u_* = 1$  при  $\sin 2y \leq 0$  и в противном случае  $H_* = u_* = 0$ .

Усредняя уравнения (3.8.2), (3.8.4), (3.8.5) по  $y$  от 0 до  $\pi$ , находим

$$\frac{dx}{dt} = -\frac{\varepsilon x}{2\pi}, \quad \frac{dy}{dt} = 1 - \frac{\varepsilon}{4}, \quad \frac{dp}{dt} = \frac{\varepsilon p}{2\pi}, \quad \frac{d\lambda}{dt} = 0 \quad x(t) > 0, \quad p(t) > 0.$$

Отсюда в первом приближении в задаче минимизации (максимизации) энергии имеем

$$x(t) = x_1 \exp \frac{\pm \varepsilon(t - T_1)}{2\pi}, \quad y(t) = y_1 + (1 - \frac{\varepsilon}{4})(t - T_1), \quad W(T_2) = W(T_1) \exp \frac{\pm \varepsilon(T_2 - T_1)}{\pi}.$$

Решение найдено с погрешностью  $\sim \varepsilon$  на интервале движения  $T_2 - T_1 \sim O(1/\varepsilon)$ .

Данный пример можно рассматривать как оптимальное гашение (раскачивание) на качелях. Сидящий на них человек, поджимая и выпрямляя ноги, изменяет момент инерции и тем самым гасит или увеличивает полную энергию качелей. ■

**Пример 3.8.2.** Коррекция вращения движения маятника. Изложенная схема может быть использована для расчета оптимального управления существенно нелинейными системами. В качестве примера рассмотрим задачу об оптимальном уменьшении энергии вращающегося маятника. Уравнение движения запишем в виде

$$d^2y/dt^2 + w \sin y = 0. \quad (3.8.9)$$

Требуется выбрать управление  $w(t)$  из интервала  $w_2 - w_1 > 0$  так, чтобы в конце движения угловая скорость была наименьшей. Считаем, что на протяжении всего процесса маятник совершает быстрое вращательное движение. Введем новое время  $\tau$ , фазовые переменные  $x$  и  $y$ , начальную угловую скорость  $\Omega$  с помощью условий

$$dy/dt = \Omega + x, \quad \tau = \varepsilon^{-1}t, \quad \Omega = \varepsilon^{-1},$$

где  $\varepsilon \leq 1$  – малый параметр, получим систему, эквивалентную уравнению (3.8.9):

$$dx/d\tau = -\varepsilon w \sin y, \quad dy/d\tau = 1 + \varepsilon x, \quad W = x. \quad (3.8.10)$$

Фазовым переменным  $x$  и  $y$  поставим в соответствие импульсы  $p$  и  $\lambda$ . Функция Понтрягина будет иметь вид  $H = -\varepsilon p w \sin y + \lambda + \varepsilon \lambda x$ .

Уравнение для импульсов следующее:

$$\frac{dp}{d\tau} = -\varepsilon \lambda, \quad \frac{d\lambda}{d\tau} = \varepsilon p w \cos y. \quad (3.8.11)$$

Приведем краевые условия  $x(\tau_1) = 0$ ,  $y(\tau_1) = \Omega$ ,  $p(\tau_2) = 1$ ,  $\lambda(\tau_2) = 0$ ,  $\tau_1 \leq \tau \leq \tau_2$ . Оптимальное управление запишется в виде

$$w_*(y) = w_2 \theta(p \sin y) + w_1 \theta(-p \sin y). \quad (3.8.12)$$

Для усреднения правых частей (3.8.10) и (3.8.11) надо посчитать два интеграла

$$J_1 = \int_0^{2\pi} w_*(y) \sin y \, dy \quad \text{и} \quad J_2 = \int_0^{2\pi} w_*(y) \cos y \, dy.$$

При  $p \geq 0$  получим  $w_*(y) = w_2 \theta(\sin y) + w_1 \theta(-\sin y)$ ,  $J_1 = 2(w_2 - w_1)$ ,  $J_2 = 0$ . Усредненная функция  $H$ , усредненные уравнения и оптимальное управление будут

$$\begin{aligned} H_* &= \lambda(1 + \varepsilon x) + \frac{\varepsilon p}{\pi}(w_2 - w_1), \\ \frac{dx}{d\tau} &= -\frac{\varepsilon}{\pi}(w_2 - w_1), \quad \frac{dy}{d\tau} = 1 + \varepsilon x, \quad \frac{dp}{d\tau} = -\varepsilon \lambda, \quad \frac{d\lambda}{d\tau} = 0, \quad p(t) \equiv 1, \\ w_*(y) &= w_2 \theta(\sin y) + w_1 \theta(-\sin y). \end{aligned} \quad (3.8.13)$$

Окончательное решение задачи имеет вид

$$x = -\frac{w_2 - w_1}{\pi}(\tau - \tau_1), \quad y = \Omega + \tau - \tau_1 = \frac{(\tau - \tau_1)^2}{2\pi}(w_2 - w_1), \quad p(\tau) = 1.$$

В этом примере решение найдено с погрешностью порядка  $\varepsilon$  на интервале  $\sim \varepsilon^{-1}$  ■

**Пример 3.8.3.** В качестве простейшего примера приведем игровую задачу о долинхобрахистохроне, взятую из книги Р. Айзекса [2]. Игра описывается системой из двух дифференциальных уравнений

$$\begin{aligned}\frac{dx}{dt} &= \sqrt{y} \cos u + \frac{1+v}{2}, \\ \frac{dy}{dt} &= \sqrt{y} \sin u + \frac{v-1}{2}.\end{aligned}$$

Здесь  $u$  и  $v$  – управления, удовлетворяющие ограничениям:  $0 \leq u \leq 2\pi$ ,  $-1 \leq v \leq 1$ .

Первый игрок, управляющий функцией  $u$ , стремится за кратчайшее время провести фазовый вектор на множество  $M = \{x, y : x = 0, y \geq 0\}$ .

Второй игрок, распоряжающийся функцией  $v$ , стремится не допустить попадания на  $M$  или по крайней мере отсрочить его. Численное решение задачи начиналось с исследования метода (2.3.2), далее продолжалось методом условного градиента.

Результаты расчетов приведены на рис. 3.8.1, где изображены фазовые траектории в плоскости  $[x, y]$ . Изломы траектории лежат на так называемой параболе переключений, где управление второго игрока релейно изменяется от  $-1$  до  $+1$ . Интересно сравнить этот рисунок с аналогичным рисунком, приведенным на с. 127 книги Р. Айзекса [2], где ищется решение игры в позиционных стратегиях. Фазовые траектории в области, лежащей выше параболы переключений, в точности соответствуют аналогичному решению, полученному Р. Айзексом. Однако фазовые траектории, лежащие ниже параболы, имеют иной вид. В статье С. А. Чигиря [92] было указано на неточности в качественном описании Р. Айзексом поля траекторий. Вместе с тем в этой статье неточно изображены фазовые траектории вблизи множества  $M$ , которые должны подходить к оси ординат под прямым углом. Это следует из аналитических формул и численного решения, приведенного на рис. 3.8.1.

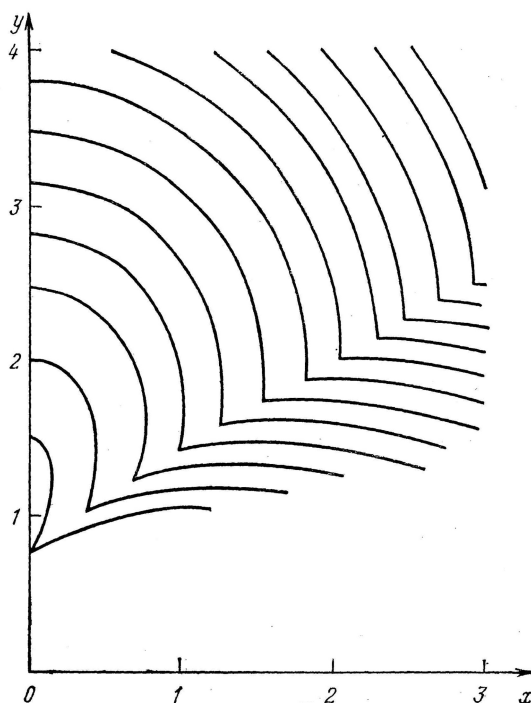


Рис. 3.8.1

■

### 3.9. Локально оптимальные управления

Для решения задач оптимального управления используются различные приближенные подходы. Укажем, например, метод усреднения [51], метод продолжения по параметру [93] и метод малого параметра. Здесь коротко остановимся на одном простейшем приеме нахождения приближенного синтеза управления, опираясь на идею локальной оптимизации.

Пусть управляемый процесс описывается системой (3.3.1). Требуется найти допустимый синтез управления, минимизирующий функционал  $W(x)$  в конечный момент движения. Упростим задачу, синтез управления будем искать из условия локальной оптимизации.

Продифференцируем  $W(x)$ , в силу системы (3.3.1) получим

$$\frac{dW(x(t))}{dt} = f^\top(t, x(t), u(t))W_x(x(t)).$$

При каждом  $t$  управление  $\tilde{u}(t)$  ищем из условия минимизации скорости изменения целевой функции на допустимом множестве, т.е.

$$\tilde{u}(t) \in \text{Arg} \min_{u \in U} f^\top(t, x(t), u)W_x(x(t)). \quad (3.9.1)$$

Так определенное управление называется локально оптимальным.

Пусть  $u_*(t)$  — управление, найденное из решения задачи оптимального управления (3.3.3), а  $x_*(t)$  и  $p_*(t)$  — соответствующие траектория и импульс, тогда согласно принципу максимума

$$u_*(t) \in \text{Arg} \min_{u \in U} f^\top(t, x_*(t), u)p_*(t). \quad (3.9.2)$$

Учитывая условие на правом конце (3.3.5), приходим к выводу: в конце движения при  $t = T_2$  обе формулы (3.9.1) и (3.9.2) совпадают. Встречаются случаи, когда эти управления совпадают на некотором участке траектории. В примерах 3.8.1 и 3.8.2 совпадение имеет место на всей траектории. Для подобных задач точный оптимальный синтез управления удастся находить из простейших соображений, не привлекая принцип максимума. Совпадение синтезов управления (3.8.7) и (3.8.13) произошло здесь благодаря тому, что в этих задачах импульс  $p_*(t)$  не изменял знак.



## Комментарии к главе 3

§ 3.1. Теоремы о дифференцировании по начальным условиям решений дифференциальных уравнений приведены во многих учебниках. Укажем, например, книгу В.В. Степанова [82].

§ 3.2. Общая теория анализа систем с разрывными правыми частями изложена в книге А.Ф. Филиппова [87]. Свойства оптимальных решений разрывных систем изучались в работах Р.В. Гамкрелидзе [75], В.В. Величенко [26], Л.Т. Ащепкова [18], А.Ф. Албу и В.И. Зубова [3–8].

§ 3.3. Опубликовано значительное количество статей и книг, посвященных доказательству принципа максимума. Укажем прежде всего книгу Л.С. Понтрягина и его учеников [75], а также [9, 12, 16, 81]. Из других подходов сошлемся на доказательства Л.И. Розоноэра [78], Н.Н. Моисеева [70], Р. Габасова и Ф.М. Кирилловой [27]. Приведенное в этом параграфе доказательство является простым следствием формулы (3.1.14) о производной решений обыкновенных дифференциальных уравнений по начальным условиям.

Формула (3.1.14) использовалась автором в 1980 г., когда он совместно с Н.И. Грачевым решал оптимизационные задачи проектируемого самолета по заказу конструкторского бюро П.О. Сухого. Было составлено математическое описание динамики самолета в виде системы (3.3.1), по ней построена система вычисления импульсов (3.3.4) и градиентов (3.7.7). Полученные расчеты давали явно неверные результаты, процедуры минимизации функции  $H$  не сходились. Поиск ошибок в программах, реализующих расчетные формулы, был не прост, так как описание самолета давалось весьма детальное: в него входили, например, дроссельные характеристики двигателя, зависевшие сложным образом от плотности атмосферы, скорости полета и т.д. Входящие в правые части системы (3.3.4) производные считались аналитически. Чтобы избежать ошибок, было решено отказаться от использования системы (3.3.4) и провести расчеты “напрямую”, как это делается в определении производной по Фреше (см. § 3.1, формула (3.1.15)) без привлечения системы (3.3.4). Благодаря этому ошибка в расчете правых частей системы для импульсов (3.3.4) была локализована и устранена.

Может создаться впечатление, что можно вообще отказаться от использования сопряженной системы (3.3.4). На самом деле это не так. Чтобы найти вектор  $p(\tau)$  по определению производной Фреше, пришлось дополнительно проинтегрировать систему (3.3.1) от  $t = \tau$  до  $t = T_2$ . Использование системы (3.3.4) дает возможность найти  $p(t)$  для всех  $T_1 \leq t \leq T_2$  в процессе только одного интегрирования сопряженной системы. Таким образом существенно упрощается определение функции  $p(t)$ .

§ 3.4. В книге [88] принцип максимума был перенесен с непрерывных систем на дискретные без введения дополнительного предположения о выпуклости годографа правых частей системы (3.3.1), что было очевидной ошибкой. В предисловии к русскому изданию книги переводчик А.И. Пропой указал на эту ошибку. Первая правильная формулировка и доказательство дискретного принципа максимума были выполнены А.И. Пропоём [76]. Впоследствии этой теме были посвящены книги А.И. Пропоёв [77] и В.Г. Болтянского [21]. Излагаемый здесь подход, основанный на принципе линеаризации, взят из книги автора [50].

§ 3.5. Различия в формулировках принципа максимума и дискретного принципа максимума вызывали в свое время острую полемику. Одна из точек зрения была изложена Н.Н. Моисеевым в его предисловии к русскому переводу книги Э. Поллака [73]: “Проблемы отыскания оптимальных управлений — не только проблемы вычислительной математики и построения конкретных программ. Это — источник глубоких идей и понимания принципов управления. Разве мог появиться принцип максимума при изучении лишь дискретных си-

стем управления? Конечно, нет. Ибо в теории дискретных систем принцип максимума верен только при специальных предположениях о выпуклости, а в непрерывном случае он верен всегда! Разве, не рассматривая континуальные системы, можно было догадаться о том, что при уменьшении шага дискретизации, т.е. с ростом размерности соответствующей задачи нелинейного программирования, неизбежно наступит такой момент, когда принцип максимума станет справедливым?”

Выход из положения был найден Р. Габасовым и Ф.М. Кирилловой [27], показавшими, что дискретный вариант принципа максимума Л.С. Понтрягина можно сформулировать специальным образом, названным ими принципом квазимаксимума. Этот принцип, дающий необходимые условия оптимальности для дискретных систем, не требует предположения о выпуклости годографа правых частей системы (3.3.1) и в пределе, при переходе от дискретных систем к непрерывным, из принципа квазимаксимума удастся получить принцип максимума Л.С. Понтрягина, не постулируя выпуклость годографа правых частей.

§ 3.8. Примеры 3.8.1 и 3.8.2 взяты из статьи автора [51].

Близкая к рассматриваемой в примере 3.8.1 задача, так называемая задача Фуллера, приведена в [16].

Решение задачи о долихобрахистохроне приведено в книге автора [50].

## Глава 4. Оптимизация систем, описываемых уравнениями с частными производными

### 4.1. Оптимизация решения задачи Бюргерса с помощью граничных условий

Теория оптимального управления системами с распределенными параметрами является предметом многочисленных исследований, укажем лишь некоторые книги и статьи по этой теме [3–8, 23, 50, 52, 67, 68]. В этом и в следующих разделах покажем, что наш подход дает новую методологию нахождения точных градиентов для сложных управляемых систем, описываемых уравнениями с частными производными. Здесь, так же как и в предыдущей главе, дискретизируя бесконечномерную оптимизационную задачу, получим конечномерную задачу нелинейного программирования и, используя технику БАД для аппроксимирующей задачи, вычислим точный градиент целевой функции. В качестве первого примера рассмотрим задачу Бюргерса.

При решении задач прибрежной физической океанографии часто используются специальные упрощенные формы уравнений Навье–Стокса. Одной из них является одномерная нестационарная нелинейная конвективно-диффузионная модель, в основе которой лежит уравнение Бюргерса. В данном параграфе рассматривается задача отыскания оптимального управления решениями уравнения Бюргерса. Под оптимальным управлением понимается решение задачи минимизации функционала, представляющего собой норму отклонения решения уравнения Бюргерса от некоторой предписанной функции, найденной экспериментальным путем. В качестве управления выбраны граничные условия. Применяя подход, описанный в главе 1, построим формулы для вычисления градиента дискретизированного функционала, после чего минимизация будет проведена численно с помощью метода сопряженных градиентов.

Постановка рассматриваемой в этом параграфе оптимизационной задачи и основная канва изложения материала взяты из статьи французских авторов [120]. Однако развиваемый ими подход представляется неудачным, так как авторы предлагают выбирать аппроксимацию исходной системы специальным образом из условия возможности вычисления точного градиента целевой функции. Здесь этому подходу противопоставляется методология БАД, в которой точное вычисление градиента обеспечивается автоматически при любой аппроксимации исходной задачи. Поэтому при выборе конкретной расчетной модели можно руководствоваться только требованиями учета точности аппроксимации и обеспечения устойчивости вычислительного процесса.

Рассмотрим смешанную задачу для одномерного нестационарного уравнения Бюргерса:

$$\begin{cases} \frac{\partial y}{\partial t} + \frac{\partial}{\partial x} \left( \frac{y^2}{2} \right) - \nu \frac{\partial^2 y}{\partial x^2} = 0, & (x, t) \in Q, \\ y(x, 0) = v(x), & x \in [0, \ell], \\ y(0, t) = \hat{u}(t), & t \in (0, T], \\ y(\ell, t) = \tilde{u}(t), & t \in (0, T]. \end{cases} \quad (4.1.1)$$

Здесь  $x$  — пространственная координата;  $t$  — время;  $y(x, t)$  — скорость жидкости в точке  $x$  в момент  $t$ ; прямоугольник  $Q = (0, \ell) \times (0, T)$ ;  $\nu$  — постоянный коэффициент вязкости;  $v(x)$  — заданная функция, определяющая распределение скорости в начальный момент времени;  $\hat{u}(t)$ ,  $\tilde{u}(t)$  — искомые функции, в дальнейшем называемые *управлениями на левом и правом конце* соответственно.

Соотношения (4.1.1) будем называть *исходной (прямой) системой*. Пусть задана функция  $\bar{y}(x, t)$ , где  $(x, t) \in Q$ . Ее можно рассматривать как распределение скорости жидкости, полученное в результате некоторых экспериментальных исследований. Поставим следующую задачу: подобрать граничные условия таким образом, чтобы решение  $y(x, t)$  системы (4.1.1) было как можно ближе к заданной функции  $\bar{y}(x, t)$ . Точнее, требуется определить оптимальное управление  $u_*(t) = (\hat{u}_*(t), \tilde{u}_*(t))$  и соответствующее решение  $y_*(x, t)$  системы (4.1.1), при котором интегральный функционал

$$W(u) = \frac{1}{2} \int_0^T \int_0^\ell [y(x, t) - \bar{y}(x, t)]^2 dx dt \quad (4.1.2)$$

достигал бы минимально возможного значения.

Задачи оптимального управления, подобные этой, обычно решаются численно с помощью некоторого метода спуска, который требует знания приведенного градиента функционала (4.1.2). Описание подхода начнем с использования классической техники вариационного исчисления. В соответствии с идеей метода множителей Лагранжа запишем функционал Лагранжа в виде

$$W = \int_0^T \int_0^\ell \left\{ \frac{1}{2} [y(x, t) - \bar{y}(x, t)]^2 + p(x, t) \left[ \frac{\partial y}{\partial t} + \frac{\partial}{\partial x} \left( \frac{y^2}{2} \right) - \nu \frac{\partial^2 y}{\partial x^2} \right] \right\} dx dt. \quad (4.1.3)$$

Здесь  $p(x, t)$  — некоторая функция,  $p(x, t) \in C_{t,x}^{1,2}(Q)$ ; ее определение приведем ниже.

Пусть  $\delta\hat{u}(t)$  и  $\delta\tilde{u}(t)$  — вариации управлений, а  $\delta y(x, t)$  — соответствующая им вариация решения системы (4.1.1). Первая вариация функционала (4.1.3) записывается следующим образом:

$$\delta W = \int_0^T \int_0^\ell \left\{ (y - \bar{y}) \delta y + p(x, t) \left[ \frac{\partial}{\partial t} (\delta y) + \frac{\partial}{\partial x} (y \delta y) - \nu \frac{\partial^2}{\partial x^2} (\delta y) \right] \right\} dx dt.$$

В силу условий на левом и правом концах (при  $x = 0$ ,  $x = \ell$ ) для вариации  $\delta y(x, t)$  выполнены условия  $\delta y(0, t) = \delta\hat{u}(t)$ ,  $\delta y(\ell, t) = \delta\tilde{u}(t)$ . Используя интегрирование по частям, получим

$$\begin{aligned} \int_0^T p(x, t) \frac{\partial}{\partial t} \delta y dt &= p(x, T) \delta y(x, T) - p(x, 0) \delta y(x, 0) - \int_0^T p_t(x, t) \delta y(x, t) dt, \\ \int_0^\ell p(x, t) \frac{\partial}{\partial x} (y(x, t) \delta y(x, t)) dx &= p(\ell, t) \tilde{u}(t) \delta \tilde{u}(t) - p(0, t) \hat{u}(t) \delta \hat{u}(t) - \\ &\quad - \int_0^\ell p_x(x, t) y(x, t) \delta y(x, t) dx, \\ -\nu \int_0^\ell p(x, t) \frac{\partial^2}{\partial x^2} (\delta y(x, t)) dx &= \nu \left[ p(0, t) \frac{\partial \delta y(0, t)}{\partial x} - p(\ell, t) \frac{\partial \delta y(\ell, t)}{\partial x} + \right. \\ &\quad \left. + \int_0^\ell p_x(x, t) \frac{\partial \delta y(x, t)}{\partial x} dx \right], \end{aligned}$$

$$\int_0^{\ell} p_x(x, t) \frac{\partial \delta y(x, t)}{\partial x} dx = p_x(\ell, t) \delta \tilde{u}(t) - p_x(0, t) \delta \hat{u}(t) - \int_0^{\ell} p_{xx}(x, t) \delta y(x, t) dx.$$

Объединяя эти результаты и учитывая, что  $\delta y(x, 0) \equiv 0$ , получим

$$\begin{aligned} \delta W &= \int_0^T \int_0^{\ell} \left[ (y - \bar{y}) - \frac{\partial p}{\partial t} - y \frac{\partial p}{\partial x} - \nu \frac{\partial^2 p}{\partial x^2} \right] \delta y(x, t) dx dt + \int_0^{\ell} p(x, T) \delta y(x, T) dx + \\ &+ \int_0^T \{ -[p(0, t) \hat{u}(t) + \nu p_x(0, t)] \delta \hat{u}(t) + [p(\ell, t) \tilde{u}(t) + \nu p_x(\ell, t)] \delta \tilde{u}(t) \} dt + \\ &- \nu \int_0^T \left[ p(0, t) \frac{\partial \delta y(0, t)}{\partial x} - p(\ell, t) \frac{\partial \delta y(\ell, t)}{\partial x} \right] dt. \end{aligned}$$

Наложим на функцию  $p(x, t)$  следующие условия:

$$\begin{cases} \frac{\partial p}{\partial t} + y \frac{\partial p}{\partial x} + \nu \frac{\partial^2 p}{\partial x^2} = y - \bar{y}, & (x, t) \in Q, \\ p(x, T) = 0, & x \in [0, \ell], \\ p(0, t) = p(\ell, t) = 0, & t \in [0, T]. \end{cases} \quad (4.1.4)$$

Систему (4.1.4) в теории вариационного исчисления обычно называют *сопряженной* к исходной системе (4.1.1). Используя (4.1.4), приходим к следующему выражению для первой вариации функционала:

$$\delta W = \nu \int_0^T [-p_x(0, t) \delta \hat{u}(t) + p_x(\ell, t) \delta \tilde{u}(t)] dt.$$

Отсюда находим искомый градиент функционала

$$\frac{\partial W}{\partial \hat{u}(t)} = -\nu p_x(0, t), \quad \frac{\partial W}{\partial \tilde{u}(t)} = \nu p_x(\ell, t). \quad (4.1.5)$$

Перейдем к анализу дискретного варианта системы (4.1.1). Разобьем временной интервал  $[0, T]$  на  $m$  равных подынтервалов, каждый из которых имеет длину  $h = T/m$ ; аналогично, пространственный интервал  $[0, \ell]$  разобьем на  $(k+1)$  подынтервалов длины  $q = \ell/(k+1)$ . Дискретная версия задачи оптимального управления следующая: найти дискретное оптимальное управление — векторы  $\hat{u}$  и  $\tilde{u}$  с компонентами  $\hat{u}^j, \tilde{u}^j, 1 \leq j < m$ , — при котором сложная дискретизированная функция  $\tilde{W}(u)$  принимает минимальное значение

$$\tilde{W}(u) = hq \sum_{i=1}^k \sum_{j=1}^m (y_i^j - \bar{y}_i^j)^2 / 2, \quad (4.1.6)$$

где использованы следующие обозначения:

$$\begin{aligned} y_i^j &= y(x_i, t_j), \quad \bar{y}_i^j = \bar{y}(x_i, t_j), \quad x_i = iq, \quad t_j = jh, \quad y_i^0 = v_i = v(x_i), \\ &0 \leq i \leq k+1, \quad 0 \leq j \leq m. \end{aligned}$$

Рассмотрим семейство конечно-разностных аппроксимаций исходной задачи (4.1.1)

$$\begin{aligned} (y_i^{j+1} - y_i^j)/h + \gamma \{ (1 - \beta_i) [(y_{i+1}^j)^2 - (y_i^j)^2] + \\ + \beta_i [(y_i^j)^2 - (y_{i-1}^j)^2] \} - \lambda (y_{i+1}^j - 2y_i^j + y_{i-1}^j) = 0, \end{aligned} \quad (4.1.7)$$

где  $\gamma = 1/(2q)$ ,  $\lambda = \nu/q^2$ ,  $2 \leq i \leq k-1$ ,  $0 \leq j < m$ . Здесь для аппроксимации конвективного члена  $\partial/\partial x(y^2/2)$  представлена возможность использовать две разные схемы путем специального выбора вектора  $\beta$  с компонентами  $\beta_i$ ,  $1 \leq i \leq k$ :

первая схема:  $\beta_1 = 0$ ,  $\beta_k = 1$ ,  $\beta_s = 0.5$  для  $1 < s < k$ ;

вторая схема:  $\beta_s = 0.5$  при всех  $1 \leq s \leq k$ .

Граничные и начальные условия имеют вид

$$\begin{aligned} y(x_i, 0) &= v_i, \quad 0 \leq i \leq k+1; \\ y(0, t_j) &= y_0^j = \hat{u}(t_j) = \hat{u}^j, \\ y(\ell, t_j) &= y_{k+1}^j = \tilde{u}(t_j) = \tilde{u}^j, \end{aligned} \quad (4.1.8)$$

где  $1 \leq j < m$ .

Несколько нестандартно записываются аппроксимации в случаях, когда  $i = 1$  и  $i = k$ . С учетом граничных условий (4.1.8) получим

$$\begin{aligned} (y_1^{j+1} - y_1^j)/h + \gamma \{ (1 - \beta_1) [(y_2^j)^2 - (y_1^j)^2] + \beta_1 [(y_1^j)^2 - (\hat{u}^j)^2] \} - \\ - \lambda (y_2^j - 2y_1^j + \hat{u}^j) = 0, \end{aligned} \quad (4.1.9)$$

$$\begin{aligned} (y_k^{j+1} - y_k^j)/h + \gamma \{ (1 - \beta_k) [(\tilde{u}^j)^2 - (y_k^j)^2] + \beta_k [(y_k^j)^2 - (y_{k-1}^j)^2] \} - \\ - \lambda (\tilde{u}^j - 2y_k^j + y_{k-1}^j) = 0. \end{aligned} \quad (4.1.10)$$

Функция  $E$ , определенная по формуле (1.3.8), в данном случае имеет вид

$$\begin{aligned} \tilde{E}(y, u, p)/(hq) &= \sum_{i=1}^k \sum_{j=1}^m \{ (y_i^j - \bar{y}_i^j)^2/2 + \sum_{i=2}^{k-1} \sum_{j=0}^{m-1} p_i^j \{ (y_i^{j+1} - y_i^j)/h + \\ &+ \gamma \{ (1 - \beta_i) [(y_{i+1}^j)^2 - (y_i^j)^2] + \\ &+ \beta_i [(y_i^j)^2 - (y_{i-1}^j)^2] \} - \lambda (y_{i+1}^j - 2y_i^j + y_{i-1}^j) \} + \\ &+ \sum_{j=0}^{m-1} p_1^j \{ (y_1^{j+1} - y_1^j)/h + \gamma \{ (1 - \beta_1) [(y_2^j)^2 - (y_1^j)^2] + \\ &+ \beta_1 [(y_1^j)^2 - (\hat{u}^j)^2] \} - \lambda (y_2^j - 2y_1^j + \hat{u}^j) \} + \\ &+ \sum_{j=0}^{m-1} p_k^j \{ (y_k^{j+1} - y_k^j)/h + \gamma \{ (1 - \beta_k) [(\tilde{u}^j)^2 - (y_k^j)^2] + \\ &+ \beta_{k-1} [(y_k^j)^2 - (y_{k-1}^j)^2] \} - \lambda (\tilde{u}^j - 2y_k^j + y_{k-1}^j) \} = 0. \end{aligned}$$

Из (1.3.19) находим формулы для определения импульсов

$$\begin{aligned} y_i^j - \bar{y}_i^j + (p_i^{j-1} - p_i^j)/h + 2\gamma y_i^j [(1 - \beta_{i-1})p_{i-1}^j + (2\beta_i - 1)p_i^j - \beta_{i+1}p_{i+1}^j] = \\ = \lambda (p_{i-1}^j - 2p_i^j + p_{i+1}^j), \quad 1 \leq i \leq k, \quad 1 \leq j < m. \end{aligned} \quad (4.1.11)$$

При всех  $0 \leq j \leq m$  составляющие  $y_0^j$  и  $y_{k+1}^j$  не входят в выражение для функции  $\tilde{E}$ . При  $0 \leq i \leq k$  функция  $\tilde{E}$  также явно не зависит от  $y_i^0$ . Поэтому можно положить

$$p_0^j = p_{k+1}^j = 0, \quad p_i^m = 0$$

для всех  $0 \leq j \leq m$ ,  $0 \leq i \leq k + 1$ .

Градиенты целевой функции при  $0 \leq j \leq m - 1$  вычисляются по формулам

$$\begin{aligned}\frac{d\Omega}{d\hat{u}^j} &= -\frac{1}{q}(\nu + q\beta_1\hat{u}^j)p_1^j, \\ \frac{d\Omega}{d\tilde{u}^j} &= \frac{1}{q}[-\nu + q(1 - \beta_k)\tilde{u}^j]p_{k-1}^j.\end{aligned}\tag{4.1.12}$$

Первая вариация дискретизированного функционала записывается следующим образом:

$$\delta\Omega = \frac{1}{q} \sum_{j=0}^{m-1} \{-(\nu + q\beta_1\hat{u}^j)p_1^j\Delta\hat{u}^j + [-\nu + q(1 - \beta_k)\tilde{u}^j]p_k^j\} \Delta\tilde{u}^j.\tag{4.1.13}$$

Учтем, что

$$\lim_{q \rightarrow 0} \frac{p_1^j}{q} = \lim_{q \rightarrow 0} \frac{p_1^j - p_0^j}{q} = \frac{\partial p(0, t)}{\partial x}, \quad \lim_{q \rightarrow 0} \frac{p_k^j}{q} = \lim_{q \rightarrow 0} \frac{-p_{k+1}^j + p_k^j}{q} = -\frac{\partial p(\ell, t)}{\partial x}.$$

Разделим (4.1.13) на  $h$ ; в пределе при  $k \rightarrow \infty$ ,  $m \rightarrow \infty$ ,  $h \rightarrow 0$ ,  $q \rightarrow 0$ ,  $\ell = q(k + 1)$ ,  $T = h\ell$  из (4.1.12) получим (4.1.5).

Излагаемый в данной работе подход к нахождению градиента не является общепринятым в обширной литературе по оптимизации систем с частными производными. В работе [120], например, ставится следующий вопрос: “Каким способом следует аппроксимировать прямую задачу (4.1.1), чтобы получить “хороший” дискретный градиент функционала?”. В этой же работе дается ответ на поставленный вопрос: “Аппроксимация прямой задачи должна быть такой, чтобы граничные значения не принимали бы участия в аппроксимации нелинейных конвективных членов”. Для конечно-разностной схемы (4.1.7) авторы работы [120] отдают предпочтение схеме 1 перед схемой 2. Обосновывается это тем, что выражения (4.1.12) для градиента дискретного функционала принимают в первом случае вид, “похожий” на выражения (4.1.5) для градиента непрерывного функционала. Вторым аргументом в пользу такого выбора аппроксимации прямой задачи (4.1.1) авторы работы [120] считают положительные результаты, полученные при расчете тестовой задачи, выразившиеся в проверке равенства

$$\lim_{q \rightarrow 0} \mathcal{R}(q) = 1,\tag{4.1.14}$$

где

$$\mathcal{R}(q) = \frac{\Omega(u + q\nabla\Omega(u)) - \Omega(u)}{q\|\nabla\Omega(u)\|^2}, \quad \nabla\Omega = \begin{bmatrix} \frac{d\Omega}{d\hat{u}} \\ \frac{d\Omega}{d\tilde{u}} \end{bmatrix}.$$

На наш взгляд, такой подход принципиально неверен. Выбор конечно-разностной схемы для аппроксимации исходной системы (4.1.1) должен диктоваться исключительно свойствами этой системы, а не свойствами получаемого затем градиента. Из изложенной выше методики БАД следует, что при выборе конечно-разностной аппроксимации исходной системы достаточно руководствоваться исключительно тем, насколько хорошо выбранная разностная схема воспроизводит точное решение, так как согласованность аппроксимации сопряженной задачи и градиента функционала с аппроксимацией исходной задачи обеспечивается при использовании подхода БАД автоматически. При этом для любой выбранной аппроксимации исходной задачи вычисленный градиент будет точным.

При решении краевых задач (4.1.1) для уравнения Бюргерса следует обращать внимание на два момента. Во-первых, в случае малых  $\nu$  (точнее, конвективный член сравним с вязким членом) аппроксимация конвективного члена проводится с учетом знака решения  $y(x, t)$

в данной точке. Это позволяет качественно правильно отражать область зависимости решения  $y(x, t)$  от начальных и граничных условий. Неучет этого факта приводит к более жестким ограничениям на допустимую величину шага по времени  $h$ . Во-вторых, уравнение Бюргера нелинейное, поэтому решение краевой задачи может содержать области больших градиентов (“размазанных разрывов”). Если не предпринимать специальных мер для сгущения сетки в областях больших градиентов, то при получении решения задачи необходимо использовать консервативные конечно-разностные схемы [79]. Использование неконсервативных схем может привести к искажению решения в окрестности областей больших градиентов (например, к неправильной величине скорости распространения “размазанного разрыва”).

Следует отметить, что в отличие от схемы 1 схема 2 является консервативной и обладает большим запасом устойчивости. Если расчет проводится на подробных сетках, то схемы 1 и 2 дают примерно одинаковые результаты. Если же сетки грубые (а именно с такими сетками часто приходится иметь дело в задачах физической метеорологии и прибрежной физической океанографии), то в случае  $y(x, y) > 0$  схема 1 часто оказывается неустойчивой, в то время как схема 2 позволяет получить решение задачи.

Что же касается теста, основанного на проверке равенства (4.1.14), то результаты всегда будут положительными, если для той или иной аппроксимации задачи (4.1.1) (в нашем случае для некоторого вектора  $\beta_i$ ) градиент целевой функции определять по формуле (4.1.12) с этим  $\beta_i$ .

Все численные результаты, описываемые ниже в этом параграфе, относятся к системе (4.1.1) со следующими входными параметрами:

$$\begin{aligned} \nu &= 10^{-2} \text{ м}^2/\text{сек}, & \ell &= 1 \text{ м}, & T &= 1 \text{ сек}, \\ v(x) &= 0, & x &\in [0, \ell], \\ \hat{u}(t) &= 1 - t, & t &\in [0, T), \\ \tilde{u}(t) &= t - 1, & t &\in [0, T). \end{aligned} \tag{4.1.15}$$

**Пример 4.1.1.** Сравним выражения для градиента целевой функции, получаемые с использованием схемы 1 и схемы 2. Сравнение проводится на основе проверки выполнения равенства (4.1.14). Пусть  $\bar{y}(x, t) \equiv 0$ ,  $(x, t) \in Q$ , параметры сетки следующие:  $k = 49$ ,  $m = 50$ .

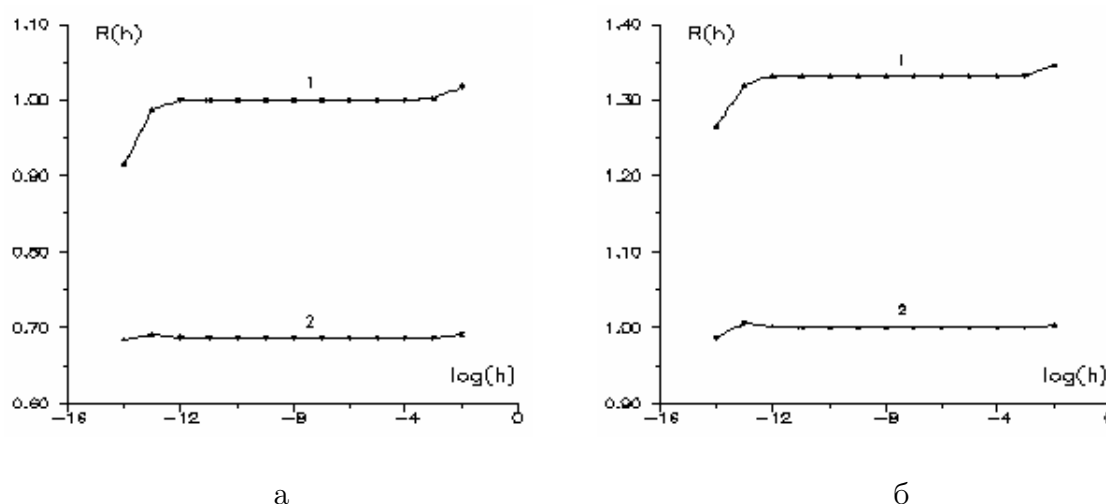


Рис. 4.1.1

На рис. 4.1.1 представлена зависимость величины  $\mathcal{R}(q)$  от десятичного логарифма  $q$ . Цифры около кривых указывают на номер схемы, которая использовалась при расчетах по формулам (4.1.7), (4.1.9), (4.1.10) и (4.1.11). На рис. 4.1.1,а градиент (4.1.12) вычислялся с векто-



ром  $\beta$ , соответствующим первой схеме, на рис. 4.1.1,б — с вектором  $\beta$ , соответствующим второй схеме. Таким образом, на кривой 1 (рис. 4.1.1,а) и кривой 2 (рис. 4.1.1,б) расчеты исходной и сопряженной систем были правильно согласованы с вычислениями градиентов. Только в этих двух случаях величина  $\mathcal{R}$  была близкой к единице.

На рис. 4.1.2 представлены результаты того же эксперимента, что и на рис. 4.1.1, с той лишь разницей, что расчеты проводились на более подробной сетке:  $k = 69$ ,  $m = 100$ . Сравнение результатов, представленных на рис. 4.1.1 и рис. 4.1.2, показывает, что различие в градиентах, рассчитанных с помощью схемы 1 и схемы 2, уменьшается на более подробных сетках. ■

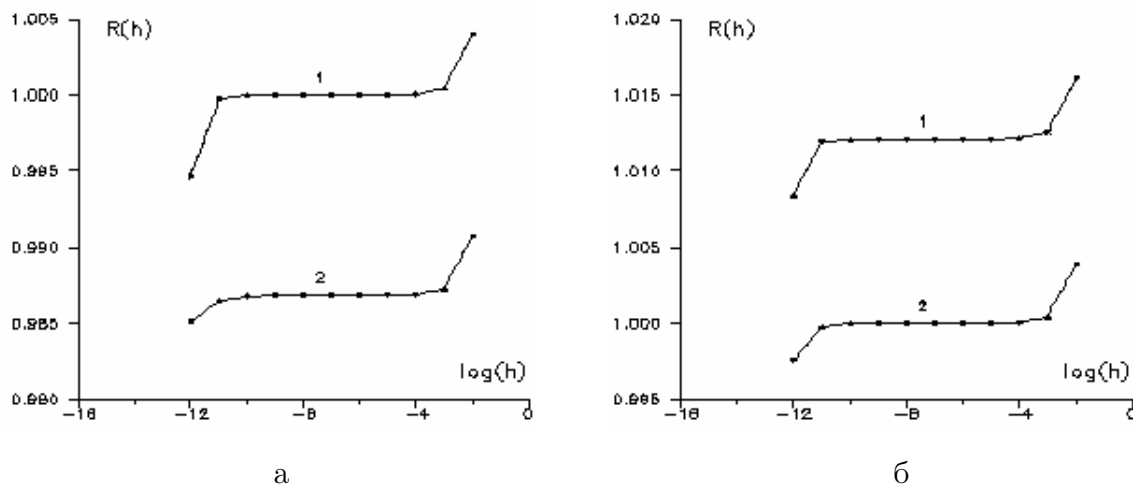


Рис. 4.1.2

**Пример 4.1.2.** Для аппроксимации исходной системы (4.1.1) используется схема 2, по формулам (4.1.15) определяются функции  $\hat{u}(t)$ ,  $\tilde{u}(t)$ , считается  $v(x) = 0$  для всех  $x \in [0, \ell]$ . В результате определяется функция  $\bar{y}(x, t)$ . Далее путем наложения на выбранные функции  $\hat{u}(t)$  и  $\tilde{u}(t)$  гауссовского белого шума со среднеквадратичным отклонением 0.3 вводится начальное управление. После этого с помощью метода сопряженных градиентов ищется оптимальное управление  $\hat{u}_*(t)$  и  $\tilde{u}_*(t)$ .

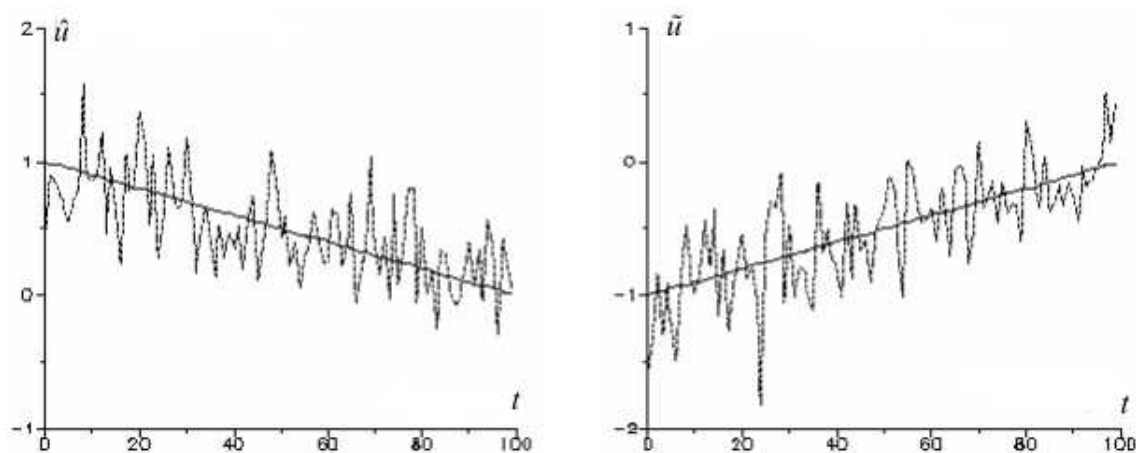


Рис. 4.1.3

На рис. 4.1.3 представлено начальное управление (пунктирная линия) и оптимальное управление (сплошная линия), полученные в результате численной минимизации функционала  $\tilde{W}$  с использованием градиентов, определяемых по формуле (4.1.12). Приближенное

управление отличается от истинного, задаваемого формулами (4.1.15), не более чем на  $10^{-9}$ . Расчеты проводились на сетке  $k = 69$ ,  $m = 100$ . ■

**Пример 4.1.3.** Этот пример полностью совпадает с примером 4.1.1. Единственное отличие состоит в том, что функция  $\bar{y}(x, t)$ , соответствующая условиям (4.1.15), определяется здесь с использованием схемы 1. На рис. 4.1.4 представлены начальное управление (пунктирная линия) и оптимальное управление (сплошная линия). Оптимальное управление  $u_*$ , полученное в результате численной оптимизации функционала  $W$ , отличается от управления (4.1.15). Наибольшее отличие (порядка 22%) наблюдается в окрестности начального момента времени  $t = 0$  и объясняется это следующим образом. Краевые условия (4.1.15), при которых определяются “экспериментальные” данные (функция  $\bar{y}(x, t)$ ), разрывны в точках  $(0, 0)$  и  $(\ell, 0)$ . Аналитическое решение краевой задачи (4.1.1), (4.1.15) имеет особенности в указанных точках. В окрестности этих точек решение будет аналитическим, но содержащим области больших градиентов (“размазанные” разрывы). По мере удаления от особых точек влияние особенностей на решение затухает, и решение становится все более плавным. Вдали от особых точек (там, где влияние особенностей мало) обе схемы (схема 1 и схема 2) могут быть использованы для получения решения исходной задачи, и решения, полученные с их помощью, будут мало отличаться. Мало отличаться вне окрестности особых точек должны и управления, что и наблюдается в расчетах. Что касается окрестности особых точек, то здесь схема 1 и схема 2 дают различающиеся решения. Предпочтение следует отдать той схеме, которая точнее воспроизводит поведение решения в окрестности особых точек. Схема 2 для краевых условий (4.1.15) более “физичная”, так как качественно правильно учитывает распространение и размазывание разрыва в краевых условиях. Ей, на наш взгляд, и следует отдать предпочтение. Схема 1 дает решение в окрестности особых точек, соответствующее граничным условиям, отличным от (4.1.15). Результатом этого и является отличие истинного управления (граничные условия (4.1.15)) от оптимального управления  $\bar{u}_*$  в окрестности особых точек. ■

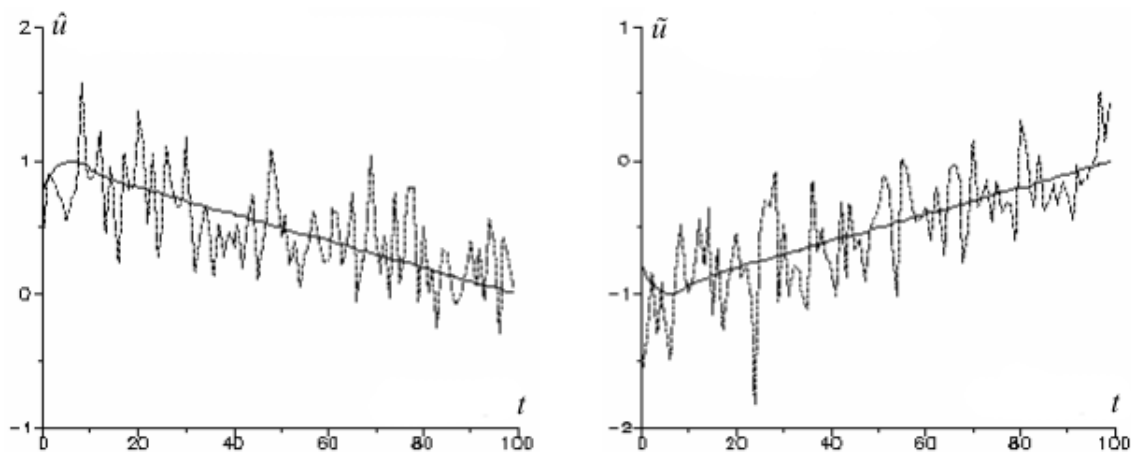


Рис. 4.1.4

**Пример 4.1.4.** Во всех приведенных выше примерах определялось оптимальное управление, которое имело весьма простой, линейный вид, задаваемый формулами (4.1.15). Интересно было попробовать “восстановить” более сложное управление — например, порождаемое использованием белого шума. С этой целью на управление, задаваемое (4.1.15), накладывался белый шум, и для этого управления  $u(t)$  (сплошные линии на рис. 4.1.5) по формулам (4.1.7)–(4.1.10) с использованием схемы 2 рассчитывалась исходная задача, определялась “экспериментальная” функция  $\bar{y}(x, t)$  и с ее помощью вычислялось значение целевой функции  $\tilde{W}$  (см.

формулу (4.1.6)). Затем в результате минимизации  $\widetilde{W}$  определялось оптимальное управление  $u_*(t)$ . В качестве начального управления в процессе минимизации  $\widetilde{W}$  брались линейные функции (4.1.15), представленные на рис. 4.1.5 пунктирной линией. Полученное оптимальное управление  $u_*(t)$  практически не отличается от истинного управления  $u(t)$ . Несмотря на весьма сложную структуру управления  $u(t)$  разность  $|u(t) - u_*(t)|$  не превышала  $10^{-8}$  при всех  $0 \leq t \leq 1$ . Следует отметить, что процесс нахождения оптимального управления в этом примере проходил гораздо медленнее, чем в предыдущих примерах. Отметим также, что решение разностной задачи при таком экзотическом оптимальном управлении весьма приближенно аппроксимирует точное решение системы (4.1.1). Для получения более точного решения исходной задачи необходимо значительно сгустить сетку — чтобы выполнить условия применимости аппроксимационных схем интегрирования системы (4.1.1). Поэтому смысл последнего эксперимента — продемонстрировать работоспособность предложенного подхода в “тяжелых” случаях. ■

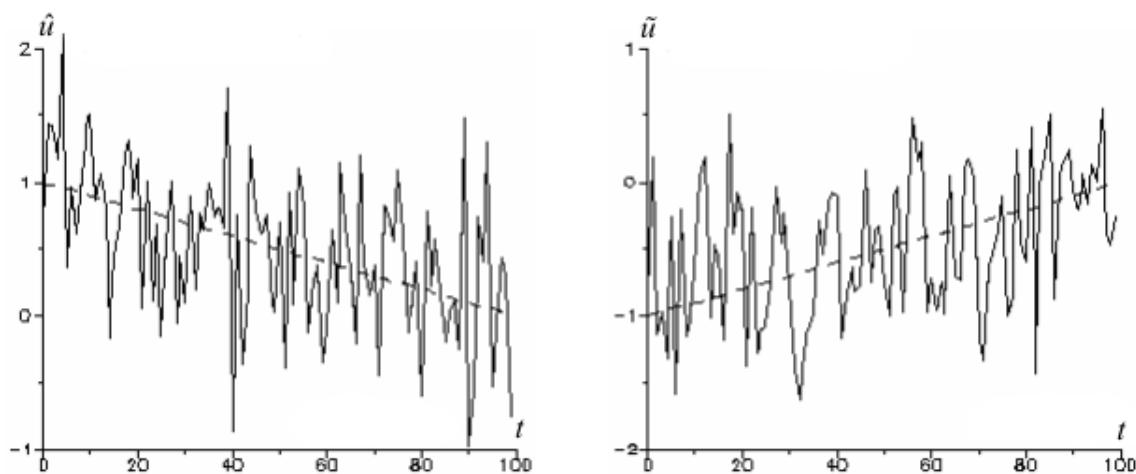


Рис. 4.1.5

Использованная методология БАД предоставила возможность сравнительно просто найти точный градиент минимизируемого функционала, давая таким образом возможность использовать для минимизации различные градиентные методы. Можно определять вторые производные, что прокладывает путь к использованию метода Ньютона для численной оптимизации.

Методологическая простота предлагаемого подхода открыла большие возможности построения библиотек программ для решения сложных практических оптимизационных задач. Описание блока ДИСО для решения задач оптимального управления с фазовыми ограничениями содержится в [35, 39].

## 4.2. Оптимальное управление параболической системой

Рассмотрим задачу определения оптимального режима нагрева стержня. Пусть управляемый процесс описывается параболическим уравнением теплопроводности порядка

$$\frac{\partial y(x, t)}{\partial t} = a^2 \frac{\partial^2 y(x, t)}{\partial x^2} + u(x, t), \quad 0 < x < \ell, \quad 0 < t \leq T, \quad (4.2.1)$$

где  $y(x, t)$  и  $u(x, t)$  — температура стержня и распределенное управление (подвод тепла) в точке  $x$  в момент  $t$  соответственно. Начальные и граничные условия следующие:

$$y(x, 0) = \varphi(x), \quad 0 \leq x \leq \ell, \quad (4.2.2)$$

$$\frac{\partial y(0, t)}{\partial x} = 0, \quad \frac{\partial y(\ell, t)}{\partial x} = \nu[g(t) - y(\ell, t)], \quad 0 < t \leq T, \quad (4.2.3)$$

где  $g(t)$  — граничное управление (сосредоточенный подвод тепла). В данной задаче ищутся управления  $g(t)$  и  $u(x, t)$ , минимизирующие функционал

$$W = \int_0^{\ell} \Psi(y(x, T)) dx, \quad (4.2.4)$$

где  $\Psi$  — некоторая заданная непрерывно дифференцируемая функция своего аргумента.

Нахождение градиента функционала  $W$  проведем сначала с помощью техники классического вариационного исчисления. Пусть  $\delta g(t)$ ,  $\delta u(x, t)$  — вариации управлений, а  $\delta y(x, t)$  — соответствующее решение следующей краевой задачи:

$$\frac{\partial \delta y(x, t)}{\partial t} = a^2 \frac{\partial^2 \delta y(x, t)}{\partial x^2} + \delta u(x, t), \quad 0 < x < t, \quad 0 < t \leq T, \quad (4.2.5)$$

$$\delta y(x, 0) = 0, \quad 0 \leq x \leq t, \quad (4.2.6)$$

$$\frac{\partial \delta y(0, t)}{\partial x} = 0, \quad \frac{\partial \delta y(\ell, t)}{\partial x} = \nu(\delta g(t) - \delta y(\ell, t)), \quad 0 \leq t \leq T. \quad (4.2.7)$$

В линейном приближении вариация функционала имеет вид

$$\delta W = \int_0^{\ell} \Psi_y(y(x, T)) \delta y(x, T) dx. \quad (4.2.8)$$

Введем дифференцируемую функцию  $p(x, t)$ , удовлетворяющую краевой задаче для сопряженной системы:

$$\frac{\partial p(x, t)}{\partial t} + a^2 \frac{\partial^2 p(x, t)}{\partial x^2} = 0, \quad 0 < x < t, \quad 0 \leq t \leq T, \quad (4.2.9)$$

$$p(x, T) = \Psi_y(y(x, T)), \quad 0 \leq x \leq T, \quad (4.2.10)$$

$$\frac{\partial p(0, t)}{\partial x} = 0, \quad \frac{\partial p(\ell, t)}{\partial x} - \nu p(\ell, t) = 0, \quad 0 < t < T. \quad (4.2.11)$$

Используя интегрирование по частям и учитывая условия (4.2.6) и (4.2.10), вариации функционала (4.2.8) представим в виде

$$\delta W = \int_0^{\ell} p(x, T) \delta y(x, T) dx = \int_0^{\ell} \left[ \int_0^T \frac{\partial}{\partial t} (p(x, t) \delta y(x, t)) \right] dt dx = \iint_{0,0}^{\ell, T} \left[ p_t \delta y + p \frac{\partial}{\partial t} \delta y \right] dt dx.$$

Учитывая (4.2.1) и (4.2.5), получим

$$\delta W = \iint_{0,0}^{\ell, T} \left[ a^2 \left( p \frac{\partial^2 \delta y}{\partial x^2} - p_{xx} \delta y \right) + p \delta u \right] dt dx = \iint_{0,0}^{\ell, T} \left[ a^2 \frac{\partial}{\partial x} \left( p \frac{\partial \delta y}{\partial x} - p_x \delta y \right) + p \delta u \right] dt dx.$$

Используя условия (4.2.5) и (4.2.7), находим

$$\delta W = \iint_{0,0}^{T, \ell} p(x, t) \delta u(x, t) dx dt + \nu a^2 \int_0^T p(\ell, t) \delta g(t) dt.$$

Эти формулы были получены ранее в монографии Ф.П. Васильева [23].

Градиенты функционала можно записать в виде

$$\frac{d\Omega}{du(x, t)} = p(x, t), \quad \frac{d\Omega}{dg(t)} = \nu a^2 p(\ell, t). \quad (4.2.12)$$

Перейдем к простейшей дискретной аппроксимации. Обозначим

$$\begin{aligned} x_i &= iq, & t_j &= jh, & i &= 0, \dots, k, & j &= 0, \dots, m, \\ q &= \ell/k, & h &= T/m, & y_i^j &= y(iq, jh), & u_i^j &= u(iq, jh), \\ \varphi_i &= \varphi(iq), & g^j &= g(jh), & i &= 0, \dots, k, & j &= 0, \dots, m. \end{aligned}$$

Аппроксимируя параболическое уравнение (4.2.1), применим явную схему Эйлера для интегрирования по времени, центрированную схему в пространстве и используем самую простую дискретизацию граничных условий. Функционал (4.2.4), дифференциальное уравнение (4.2.1), условия (4.2.2) и (4.2.3) заменяются следующими соотношениями:

$$\begin{aligned} \widetilde{W} &= q \sum_{i=0}^k \Psi(y_i^m), \\ y_i^j &= \begin{cases} (1 - 2\lambda)y_i^{j-1} + \lambda(y_{i-1}^{j-1} + y_{i+1}^{j-1}) + hu_i^{j-1}, & 1 \leq i \leq k-1, & 1 \leq j \leq m, \\ y_1^j, & i = 0, & 1 \leq j \leq m, \\ \mu y_{k-1}^j + \nu q g^j, & i = k, & 1 \leq j \leq m, \\ \varphi_i, & 0 \leq i \leq k, & j = 0, \end{cases} \end{aligned} \quad (4.2.13)$$

где  $\lambda = a^2 h / q^2$ ,  $\mu = 1 / (1 + \nu q)$ .

Введем сопряженные переменные  $p_i^j$  и вспомогательную функцию

$$\begin{aligned} E &= \widetilde{W} + \sum_{i=1}^{k-1} \sum_{j=1}^m [(1 - 2\lambda)y_i^{j-1} + \lambda(y_{i-1}^{j-1} + y_{i+1}^{j-1}) + hu_i^{j-1}] + \\ &+ \sum_{j=1}^m [y_1^j p_0^j + (\mu y_{k-1}^j + \nu q g^j) p_k^j] + \sum_{i=0}^k \varphi_i p_i^0. \end{aligned}$$

Применяя формулы (1.3.10), получим

$$p_i^j = \begin{cases} (1 - 2\lambda)p_i^{j+1} + \lambda(p_{i-1}^{j+1} + p_{i+1}^{j+1}), & 2 \leq i \leq k-2, & 0 \leq j \leq m-1, \\ (1 - 2\lambda)p_1^{j+1} + p_0^j + \lambda p_2^{j+1} & i = 1, & 1 \leq j \leq m-1, \\ (1 - 2\lambda)p_{k-1}^{j+1} + \mu p_k^j + \lambda p_{k-2}^{j+1}, & i = k-1, & 1 \leq j \leq m-1, \\ \Psi_{y_i^m} + p_0^m \delta_i^1 + \mu p_k^m \delta_i^{k-1}, & 0 \leq i \leq k, & j = m, \\ (1 - 2\lambda)p_i^1 + \lambda p_{i+1}^1, & i = 1, & j = 0, \\ (1 - 2\lambda)p_i^1 + \lambda p_{i-2}^1, & i = k-1, & j = 0, \\ \lambda p_1^{j+1}, & i = 0, & 0 \leq j \leq m-1, \\ \lambda p_{k-1}^{j+1}, & i = k, & 0 \leq j \leq m-1, \end{cases}$$

где  $\delta_\alpha^\beta = \begin{cases} 1, & \text{если } \alpha = \beta, \\ 0, & \text{если } \alpha \neq \beta. \end{cases}$

Используя (1.3.11), находим искомые производные

$$\begin{aligned}
\frac{d\Omega}{du_i^j} &= hp_i^{j+1}, & 1 \leq i \leq k-1, & 0 \leq j \leq m-1, \\
\frac{d\Omega}{du_0^j} &= \frac{d\Omega}{du_k^j} = 0, & & 0 \leq j \leq m-1, \\
\frac{d\Omega}{du_i^m} &= 0, & 0 \leq i \leq k, & \\
\frac{d\Omega}{dg^j} &= \mu\nu qp_k^j, & & 1 \leq j \leq m. \\
\frac{d\Omega}{dg^0} &= 0, & & j = 0.
\end{aligned} \tag{4.2.14}$$

В пределе при  $k \rightarrow \infty$ ,  $m \rightarrow \infty$ ,  $h \rightarrow 0$ ,  $q \rightarrow 0$ ,  $\ell = qk$ ,  $T = h\ell$  получаем (4.2.12).

Согласно [80] дискретная аппроксимация (4.2.13) устойчива, только если  $\lambda \leq 1/2$ . Мы можем избавиться от этого ограничения, используя неявную схему, которая является устойчивой для всех  $\lambda$ :

$$y_i^{j+1} = y_i^j + \lambda[y_{i-1}^{j+1} - 2y_i^{j+1} + y_{i+1}^{j+1}] + \Delta t u_i^{j+1}.$$

При этом во внутренней области вектор  $p$  определяется из неявного соотношения

$$p_i^j = p_i^{j+1} + \lambda[p_{i+1}^j - 2p_i^j + p_{i-1}^j].$$

Если  $k \rightarrow \infty$ ,  $h \rightarrow 0$ ,  $q \rightarrow 0$ , то получим, что в обоих случаях функция  $p(x, t)$  удовлетворяет условиям (4.2.9)–(4.2.11).

### 4.3. Оптимальное управление гиперболической системой

Рассмотрим динамическую систему, описываемую линейным гиперболическим дифференциальным уравнением с частными производными

$$\frac{\partial^2 y(x, t)}{\partial t^2} = a^2 \frac{\partial^2 y(x, t)}{\partial x^2} + u(x, t), \quad 0 < x < \ell, \quad 0 < t \leq T, \tag{4.3.1}$$

с начальными и граничными условиями

$$\begin{aligned}
y(x, 0) &= \varphi(x), & y_t(x, 0) &= \psi(x), & 0 \leq x \leq \ell, \\
y_x(0, t) &= g_1(t), & y_x(\ell, t) &= g_2(t), & 0 < t \leq T.
\end{aligned}$$

Управление  $u(x, t)$ ,  $g_1(t)$ ,  $g_2(t)$  должно быть выбрано так, чтобы минимизировать следующий функционал:

$$W = \int_0^\ell \Psi(y(x, T), y_t(x, T)) dx,$$

где функция  $\Psi$  непрерывно дифференцируема по всем своим аргументам.

Заменяя непрерывную задачу управления на дискретную, воспользуемся представлением дискретного процесса вида (1.3.1); получим

$$\widetilde{W} = \sum_{i=0}^k \Psi(y_i^m, (y_i^m - y_i^{m-1})/h),$$

$$y_i^j = \begin{cases} 2(1-\lambda)y_i^{j-1} - y_i^{j-2} + \lambda[y_{i+1}^{j-1} + y_{i-1}^{j-1}] + h^2 u_i^{j-1}, & 1 \leq i \leq k-1, \quad 2 \leq j \leq m, \\ y_1^j - qg_1^j, & i = 0, \quad 1 \leq j \leq m, \\ y_{k-1}^j + qg_2^j, & i = k, \quad 1 \leq j \leq m, \\ \varphi_i, & 0 \leq i \leq k, \quad j = 0, \\ y_i^0 + h\psi_i, & 1 \leq i \leq k-1, \quad j = 1, \end{cases}$$

где  $\lambda = a^2 h^2 / q^2$ ,  $\varphi_i = \varphi(iq)$ ,  $\psi_i = \psi(iq)$ ,  $g_1^j = g_1(jh)$ ,  $g_2^j = g_2(jh)$ , вычисление интеграла ведется по формуле трапеций.

Из формул (1.3.8)–(1.3.11) следует

$$E = \widetilde{W} + \sum_{i=1}^{k-1} \sum_{j=2}^m \{2(1-\lambda)y_i^{j-1} - y_i^{j-2} + \lambda[y_{i+1}^{j-1} + y_{i-1}^{j-1}] + h^2 u_i^{j-1}\} p_i^j +$$

$$+ \sum_{j=1}^m \{(y_1^j - qg_1^j)p_0^j + (y_{k-1}^j + qg_2^j)p_k^j\} + \sum_{i=0}^k \varphi_i p_i^0 + \sum_{i=1}^{k-1} (y_i^0 + h\psi_i)p_i^1,$$

$$p_i^j = \begin{cases} 2(1-\lambda)p_i^{j+1} - p_i^{j+2} + \lambda[p_{i-1}^{j+1} + p_{i+1}^{j+1}], & 2 \leq i \leq k-2, \quad 1 \leq j \leq m-2, \\ 2(1-\lambda)p_1^{j+1} - p_1^{j+2} + \lambda p_2^{j+1} + p_0^j, & i = 1, \quad 1 \leq j \leq m-2, \\ 2(1-\lambda)p_{k-1}^{j+1} - p_{k-1}^{j+2} + \lambda p_{k-2}^{j+1} + p_k^j, & i = k-1, \quad 1 \leq j \leq m-2, \\ \lambda p_1^{j+1} \delta_i^0 + \lambda p_{k-1}^{j+1} \delta_i^k, & i = 0, i = k, \quad 1 \leq j \leq m-2, \\ \widetilde{W}_{y_i^j} + 2(1-\lambda)p_i^m + \lambda(p_{i-1}^m + p_{i+1}^m), & 2 \leq i \leq k-2, \quad j = m-1, \\ \widetilde{W}_{y_1^j} + p_0^j + 2(1-\lambda)p_1^m + \lambda p_2^m, & i = 1, \quad j = m-1, \\ \widetilde{W}_{y_i^j} + p_k^j + 2(1-\lambda)p_i^m + \lambda p_{i-1}^m, & i = k-1, \quad j = m-1, \\ \widetilde{W}_{y_i^j} + \lambda p_1^m \delta_i^0 + \lambda p_{k-1}^m \delta_i^k, & i = 0, i = k, \quad j = m-1, \\ \widetilde{W}_{y_i^m} + p_0^m \delta_i^1 + p_k^m \delta_i^{k-1}, & 0 \leq i \leq k, \quad j = m, \\ p_i^1 - p_i^2, & 1 \leq i \leq k-1, \quad j = 0, \\ 0, & i = 0, i = k, \quad j = 0, \end{cases}$$

Искомый точный градиент (для выбранной дискретной аппроксимации) вычисляется по формулам

$$\frac{d\Omega}{du_i^j} = h^2 p_i^{j+1}, \quad 1 \leq i \leq k-1, \quad 1 \leq j \leq m-1,$$

$$\frac{d\Omega}{du_0^j} = \frac{d\Omega}{du_k^j} = 0, \quad 1 \leq j \leq m-1,$$

$$\frac{d\Omega}{du_i^0} = \frac{d\Omega}{du_i^m} = 0, \quad 0 \leq i \leq k,$$

$$\frac{d\Omega}{dg_1^j} = -qp_0^j, \quad \frac{d\Omega}{dg_2^j} = qp_k^j, \quad 1 \leq j \leq m,$$

$$\frac{d\Omega}{dg_1^0} = \frac{d\Omega}{dg_2^0} = 0.$$

Если  $k \rightarrow \infty$ ,  $m \rightarrow \infty$ ,  $h \rightarrow 0$ ,  $q \rightarrow 0$  и  $\ell = qk$ , то сопряженное уравнение и граничные условия запишутся в виде

$$\begin{aligned} \frac{\partial^2 p(x, t)}{\partial t^2} &= a^2 \frac{\partial^2 p(x, t)}{\partial x^2}, \quad 0 < x < \ell, \quad 0 < t < T, \\ \frac{\partial p(0, t)}{\partial x} &= 0, \quad \frac{\partial p(\ell, t)}{\partial x} = 0, \quad 0 < t < T. \end{aligned} \tag{4.3.2}$$

Здесь автоматически по формулам БАД получены дискретные соотношения для сопряженного уравнения (4.3.2), а также для начальных и граничных условий. Применение других аппроксимаций системы (4.3.2) приведет к тому, что градиент функционала  $W$  для выбранной дискретизации исходной системы (4.3.1) будет вычислен неточно.



## Комментарии к главе 4

§ 4.1. Материал параграфа взят из статьи [52]. Постановка оптимизационной задачи для уравнения Бюргера, основная канва изложения материала заимствованы из статьи [119]. Техника вычисления градиента, использованная в этой статье, не совпадает с техникой БАД. Численные расчеты наглядно иллюстрируют ошибочность такого подхода.

§ 4.2. Подробное изложение отыскания градиента для непрерывной задачи теплопроводности можно найти в книге Ф.П. Васильева [23]. Оптимизация сложных процессов распространения тепла в жидких и кристаллических средах, учет фазовых переходов содержатся в работе В.И. Зубова, А.Ф. Албу и А.В. Албу [3–8]. Материал этого параграфа взят из работы автора [111].

§ 4.3. Нахождение градиента функционала  $W$  для непрерывного варианта дано в книге Ф.П. Васильева [23]. Материал этого параграфа взят из работы автора [111]. В статьях [3–8] опубликованы результаты определения оптимальных режимов кристаллизации материалов.

## Заключение

Теоретические исследования, проведенные в области нелинейного программирования, численные методы решения задач НЛП были и остаются плодотворным полем для решения задач оптимального управления. В данной работе мы привели лишь некоторые аспекты такого подхода. Укажем направления, в которых перенос теоретических результатов и численных методов решения со статических задач на динамические представляется особенно перспективным:

- 1) решение многокритериальных задач оптимального управления, построение Паретовских множеств на базе методов, разработанных для статических задач (см., например [57]);
- 2) решение задач глобальной оптимизации динамических управляемых систем на базе наиболее эффективных методов поиска глобального оптимума; здесь может оказаться плодотворным хорошо зарекомендовавший себя метод неравномерных покрытий (см. [56, 57, 108, 115]);
- 3) решение линейных задач большой размерности на базе метода ЛП-проекций [28, 30–34], а также использование градиентных и экстраградиентных методов [13–15].

Результаты таких обобщений планируется опубликовать в будущем.

## Приложение 1. Нормы векторов

Отображение  $\|\cdot\|$  из  $\mathbb{R}^n$  в  $\mathbb{R}^1$  называется *нормой*, если для любого  $x \in \mathbb{R}^n$  выполнены следующие три условия:

- 1)  $0 \leq \|x\|$ , причем  $\|x\| = 0$  тогда и только тогда, когда  $x = 0$ ;
- 2)  $\|\alpha x\| = |\alpha| \cdot \|x\|$ ,  $\alpha \in \mathbb{R}^1$ ;
- 3)  $\|x + y\| \leq \|x\| + \|y\|$  для любых  $x, y \in \mathbb{R}^n$ .

Для векторной нормы  $\|\cdot\|_p$  в  $\mathbb{R}^n$  существует векторная норма  $\|\cdot\|_q$ , называемая *сопряженной* к  $\|\cdot\|_p$ , которая определяется из условия

$$\|x\|_q = \sup_{\|y\|_p=1} \langle x, y \rangle.$$

Если в качестве  $\|\cdot\|_p$  взять норму Гёльдера с показателем  $p$ :

$$\|x\|_p = \left[ \sum_{i=1}^n |x^i|^p \right]^{1/p}, \quad 1 \leq p \leq \infty, \quad (\text{П.1.1})$$

то сопряженной к ней будет норма с показателем  $q$ :

$$\|x\|_q = \left[ \sum_{i=1}^n |x^i|^q \right]^{1/q}, \quad \text{где } \frac{1}{p} + \frac{1}{q} = 1.$$

Имеет место следующее неравенство Гёльдера:

$$|\langle x, y \rangle| \leq \|x\|_p \cdot \|y\|_q.$$

В частности, для так называемой *октаэдрической* нормы

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

сопряженной является норма

$$\|x\|_\infty = \max_{i \in [1:n]} |x^i|,$$

называемая обычно *чебышёвской* или *кубической*. Исходная и сопряженная нормы совпадают, когда  $p = q = 2$ :

$$\|x\|_2 = \left[ \sum_{i=1}^n (x^i)^2 \right]^{1/2}.$$

Такая норма называется *евклидовой* или *сферической*.

Для скалярного произведения в  $\mathbb{R}^n$  из неравенства Гёльдера следует классическое неравенство Коши–Буняковского

$$|\langle x, y \rangle| \leq \|x\|_2 \cdot \|y\|_2.$$

## Приложение 2. Принцип линеаризации

Пусть на  $\mathbb{R}^m$  задана скалярная функция векторного аргумента  $f(x)$ , определено множество  $X \subset \mathbb{R}^m$ . Ставится следующая оптимизационная задача:

$$\min_{x \in X} f(x).$$

Предполагаем, что множество решений этой задачи  $X^*$  не пусто. В тех точках  $x \in X$ , где функция  $f$  дифференцируема, можно определить точечно-множественное отображение

$$V(x) = \text{Arg} \min_{y \in X} \langle f_x(x), y - x \rangle. \quad (\text{П.2.1})$$

Если функция  $f$  дифференцируема всюду на  $X$ , то условие (П.2.1) каждой точке  $x \in X$  ставит в соответствие множество  $V(x) \subset X$ . Таким образом,  $V$  осуществляет точечно-множественное отображение множества  $X$  в себя. Говоря о неподвижных точках отображения  $V(x)$ , считаем, что

$$x \in X, \quad x \in V(x). \quad (\text{П.2.2})$$

Если  $x \notin V(x)$ , то существует точка  $y \in X$  такая, что

$$\langle f_x(x), y - x \rangle < 0.$$

Если  $x \in V(x)$ , то имеет место неравенство

$$0 \leq \langle f_x(x), y - x \rangle \quad \forall y \in X. \quad (\text{П.2.3})$$

Это условие иногда называют *вариационным неравенством*; всякую точку  $x \in X$ , в которой выполнено (П.2.3), называют *решением вариационного неравенства*. Для того чтобы  $x$  была неподвижной точкой отображения  $V$ , необходимо и достаточно, чтобы она была решением вариационного неравенства (П.2.3).

В каждой фиксированной точке  $x \in \mathbb{R}^n$ , где  $\|f_x(x)\| \neq 0$ , множество

$$Q(x) = \{y \in \mathbb{R}^n : \langle f_x(x), y - x \rangle = 0\}$$

представляет собой гиперплоскость, проходящую через точку  $y$ , с вектором нормали  $f_x(x)$ . Вместе с тем  $Q(x)$  является касательной гиперплоскостью к гиперповерхности

$$\Gamma(x) = \{y \in \mathbb{R}^n : f(y) = f(x)\}$$

в точке  $x$ . Множество  $\Gamma(x)$  является линией уровня функции  $f(x)$ , градиент  $f_x(x)$  направлен по нормали к линии уровня, внутрь множества

$$\Gamma_1(x) = \{y \in \mathbb{R}^n : f(y) \geq f(x)\}.$$

Гиперплоскость  $Q(x)$  порождает два полупространства

$$0 \leq \langle f_x(x), y - x \rangle, \quad \langle f_x(x), y - x \rangle \leq 0.$$

Пусть точка  $x \in Q$  такова, что в ней выполнено условие (П.2.3). Тогда в первом из приведенных полупространств содержится множество  $X$ . Поэтому (П.2.3) можно интерпретировать как условие того, что множество  $X$  целиком содержится в одном из двух замкнутых полупространств, определяемых касательной гиперплоскостью  $Q(x)$ . Касательная гиперплоскость  $Q(x)$  является опорной к множеству  $X$  в точке  $x$ , вектор  $f_x(x)$  есть опорный вектор множества  $X$  в точке  $x$ .

**Теорема Т-П.2.1.** Пусть функция  $f(x)$  непрерывна на  $\mathbb{R}^m$  и дифференцируема в точке  $x_*$ ; допустимое множество  $X$  выпуклое. Тогда для того, чтобы  $x_* \in X^*$ , необходимо, а в случае выпуклости функции  $f(x)$  и достаточно, чтобы для любых  $x \in X$  выполнялось неравенство

$$\langle f_x(x_*), x - x_* \rangle \geq 0. \quad (\text{П.2.4})$$

**Доказательство необходимости.** Пусть  $x$  — произвольная точка из множества  $X$ ,  $0 \leq \lambda \leq 1$ . Из выпуклости  $X$  следует, что каждая точка отрезка, соединяющего точки  $x_*$  и  $x$ , также принадлежит  $X$ , поэтому из условия  $x_* \in X^*$  получаем, что

$$f(x_* + \lambda(x - x_*)) \geq f(x_*).$$

Используя дифференцируемость  $f$  в точке  $x_*$ , это неравенство запишем в виде

$$0 \leq f(x_* + \lambda(x - x_*)) - f(x_*) = \lambda \langle f_x(x_*), x - x_* \rangle + \lambda \|x - x_*\| \alpha(x_*, \lambda \|x - x_*\|), \quad (\text{П.2.5})$$

где функция  $\alpha(x_*, \lambda \|x - x_*\|)$  такова, что  $\lim_{\lambda \rightarrow 0} \alpha(x_*, \lambda \|x - x_*\|) = 0$ . Разделив (П.2.5) на  $\lambda$  и устремляя  $\lambda$  к нулю, получим требуемое неравенство (П.2.4).

**Доказательство достаточности** следует из свойства выпуклых функций, согласно которому для любых  $x \in X$  имеет место неравенство

$$f(x) - f(x_*) \geq \langle f_x(x_*), x - x_* \rangle.$$

Учитывая (П.2.4), получим, что  $f(x) \geq f(x_*)$  в силу произвольности точки  $x$  из  $X$ ; заключаем, что  $x_* \in X^*$ .  $\square$

Введенное в теореме Т-П.2.1 требование выпуклости множества  $X$  существенно. Легко привести примеры, где нарушение этого требования делает утверждение теоремы неверным.

Пусть функция  $f$  определена на некотором открытом множестве из  $\mathbb{R}^n$ , содержащем множество  $X$ ; функция  $f$  называется *псевдовыпуклой в точке  $\bar{x}$  относительно множества  $X$* , если  $f$  дифференцируема в точке  $\bar{x}$  и из выполнимости для любых  $x \in X$  условия

$$\langle f_x(\bar{x}), x - \bar{x} \rangle \geq 0 \quad (\text{П.2.6})$$

следует, что  $f(x) \geq f(\bar{x})$  для всех  $x \in X$ .

Если функция  $f$  определена всюду на  $\mathbb{R}^n$ , дифференцируема в точке  $\bar{x}$  и из выполнимости неравенства (П.2.6) для всех  $x \in \mathbb{R}^n$  следует, что  $f(\bar{x}) \leq f(x)$  для любых  $x$ , то функцию  $f$  будем называть *псевдовыпуклой в точке  $\bar{x}$* .

Предположим, что функция  $f$  выпукла на выпуклом множестве  $X$  и дифференцируема в точке  $\bar{x} \in X$ , в которой функция  $f$  принимает наименьшее на  $X$  значение. Тогда функция  $f$  псевдовыпукла в точке  $\bar{x}$  относительно множества  $X$ .

Из определения псевдовыпуклости вытекает следующая теорема, дающая достаточные условия глобального минимума функции  $f$  на множестве  $X$ .

**Теорема Т-П.2.2.** Пусть функция  $f$  псевдовыпукла в точке  $\bar{x}$  относительно множества  $X$  и  $\bar{x}$  является неподвижной точкой отображения  $V$ ; тогда  $\bar{x} \in X^*$ .

Пусть  $g \in \mathbb{R}^m$ ,  $\|g\| = 1$ . Определим производную функции  $f$  по направлению  $g$ :

$$\partial f(x_*) / \partial g = \lim_{\lambda \rightarrow 0} [f(x_* + \lambda g) - f(x_*)] / \lambda = \langle f_x(x_*), g \rangle.$$

В утверждении теоремы Т-П.2.1, таким образом, (П.2.4) можно заменить следующим условием:

$$\frac{\partial f(x_*)}{\partial g} \geq 0_m.$$

## Приложение 3. Матрицы проектирования

Обозначим через  $\ker A$  нуль-пространство (ядро) произвольной матрицы  $A$  размера  $m \times n$ :

$$\ker A = \{x \in \mathbb{R}^n : Ax = 0_m\}.$$

Через  $\operatorname{im} A^\top$  обозначим пространство строк матрицы  $A$  (образ матрицы  $A^\top$ ):

$$\operatorname{im} A^\top = \{\xi \in \mathbb{R}^n : \xi = A^\top u, \quad u \in \mathbb{R}^m\}.$$

Размерность линейного пространства  $\ker A$  равна  $d$  — дефекту матрицы  $A$ ; если ранг  $A$  равен  $m$ , то  $d = n - m$ .

Нуль-пространство и пространство строк матрицы  $A$  являются ортогональными дополнениями друг к другу, что записывается в виде  $\ker A \perp \operatorname{im} A^\top$ . Действительно, для всяких векторов  $x \in \ker A$  и  $\xi \in \operatorname{im} A^\top$  имеем  $x^\top \xi = x^\top A^\top u = u^\top Ax = 0$ . Аналогично  $\operatorname{im} A \perp \ker A^\top$ .

Матрица  $A^+$  размера  $n \times m$  называется псевдообратной для  $(m \times n)$ -матрицы  $A$ , если выполнены равенства

$$\begin{aligned} A^+ &= UA^\top = A^\top V, \\ AA^+A &= A, \end{aligned}$$

где  $U$  и  $V$  — некоторые матрицы. Для данной матрицы  $A$  не может существовать двух различных псевдообратных матриц.

В зависимости от соотношений между  $m$  и  $n$  возможны три случая.

1) если ранг  $A$  равен  $m$ ,  $m \leq n$ , то строки  $A$  линейно независимы, квадратная матрица  $AA^\top$  размера  $m^2$  обратима, матрица  $A$  имеет правую обратную  $A^+$ :

$$A^+ = A^\top (AA^\top)^{-1}, \quad AA^+ = I_m, \quad (A^+)^\top = (AA^\top)^{-1}A; \quad (\text{П.3.1})$$

2) если ранг  $A$  равен  $n$ ,  $n \leq m$ , то столбцы матрицы  $A$  линейно независимы, квадратная матрица  $A^\top A$  размера  $n^2$  обратима, матрица  $A$  имеет левую обратную  $A^+$ :

$$A^+ = (A^\top A)^{-1}A^\top, \quad A^+A = I_n, \quad (A^+)^\top = A(A^\top A)^{-1}; \quad (\text{П.3.2})$$

3) если  $n = m$  и строки матрицы  $A$  линейно независимы, то  $A^+ = A^{-1}$ ,  $A^\parallel = I_n$ ,  $A^\perp = 0_{mm}$ . Укажем некоторые свойства псевдообратных матриц:

$$\begin{aligned} (A^\top)^+ &= (A^+)^\top, \quad (A^+A)^\top = A^+A, \quad (A^+A)^2 = A^+A, \\ AA^+A &= A, \quad A^\top AA^+ = A^\top. \end{aligned} \quad (\text{П.3.3})$$

Для любого числа  $\alpha$ , отличного от нуля,  $(\alpha A)^+ = \alpha^{-1}A^+$ .

Введем систему из  $m$  линейных уравнений с  $n$  неизвестными:

$$Ax = b. \quad (\text{П.3.4})$$

*Псевдорешением* (или обобщенным решением) системы (П.3.4) называется решение следующей *нормальной системы*:

$$A^\top(Ax - b) = 0_n. \quad (\text{П.3.5})$$

Нормальная система всегда совместна и имеет единственное решение тогда и только тогда, когда столбцы матрицы  $A$  линейно независимы (другими словами, система  $Ax = 0$  имеет только нулевое решение).

Если система (П.3.4) совместна, то она эквивалентна системе (П.3.5). Псевдорешение  $x_0$  может быть найдено из задачи безусловной минимизации

$$x_0 = \arg \min_{x \in \mathbb{R}^n} \|Ax - b\|^2.$$

Действительно, применяя теорему Ферма к этой задаче, приходим к системе (П.3.5).

Общее решение нормальной системы (П.3.5) для системы (П.3.4) записывается в виде

$$x = A^+b + (A^\top)^\perp z = A^+b + [J_n - (A^\top)^\parallel]z = A^+b + (J_n - A^+A)z.$$

где  $z$  — произвольный вектор из  $\mathbb{R}^n$ . В частности, если система (П.3.4) совместна, то эта формула дает ее общее решение. Выражение  $(A^\top)^\perp z$  при произвольном  $z \in \mathbb{R}^n$  определяет общее решение нормальной однородной системы  $A^\top Ax = 0_n$ .

Если система (П.3.4) совместна, то среди ее решений найдется решение, имеющее наименьшую евклидову норму — так называемое *нормальное решение*; оно задается формулой  $x_0 = A^+b$ . Это утверждение с учетом ортогональности матриц  $A^+$  и  $(A^\top)^\perp$  следует из простейших рассуждений:

$$\|x\| = \|A^+b + (A^\top)^\perp z\| \geq \|A^+b\| = \|x_0\|.$$

Если  $b \in \text{im } A$ , то система (П.3.4) имеет единственное решение, задаваемое следующими эквивалентными формулами:

$$x = A^+b = (A^\top A)^{-1}A^\top b. \quad (\text{П.3.6})$$

Если  $b \in \text{im } A$ , то  $Ax = AA^+b = A^\parallel b = \text{pr}(b, \text{im } A) = b$ . Если  $b \notin \text{im } A$ , то система (П.3.4) не имеет решения, формула (П.3.6) определяет ее псевдорешение.

Нормальное решение системы (П.3.4) задается формулами

$$x_0 = A^+b = A^\top(AA^\top)^{-1}b = A^\top u,$$

где  $u = (AA^\top)^{-1}b$ ,  $Ax_0 = b$ .

Введем  $(m \times m)$ -матрицу ортогонального проектирования  $A^\parallel = AA^+$  на пространство столбцов матрицы  $A$  и матрицу проектирования  $A^\perp = I_m - A^\parallel$  — на нуль-пространство матрицы  $A^\top$ .

Матрица  $(A^\top)^\parallel = A^+A$  есть  $(n \times n)$ -матрица проектирования на пространство столбцов матрицы  $A^\top$  (строк матрицы  $A$ ),  $(A^\top)^\perp = I_n - (A^\top)^\parallel$  — матрица проектирования на нуль-пространство матрицы  $A$ ,  $A^\parallel = A^\top(A^+)^\top$ .

Матрицы проектирования  $A^\parallel$  и  $A^\perp$  симметричные и идемпотентные, т.е.

$$A^\parallel A^\parallel = A^\parallel, \quad A^\perp A^\perp = A^\perp.$$

Вектор  $A^\parallel b$  является компонентой вектора  $b$  в пространстве столбцов матрицы  $A$ , а невязка  $b - A^\parallel b = A^\perp b$  является вектором, лежащим в ортогональном дополнении, т.е.  $b \in \text{ker } A^\top$ .

Для квадратной матрицы  $A$  запись  $A > 0$  обозначает, что матрица  $A$  положительно определена, т.е.  $z^\top Az > 0$  для любых  $z$ ,  $\|z\| \neq 0$ .

Используя оператор проектирования  $\text{pr}(\cdot, \cdot)$ , можно записать:

$$\text{pr}(q, \text{im } A) = A^\parallel q = AA^+q, \quad \text{pr}(z, \text{im } A^\top) = (A^\top)^\parallel z = A^\top(A^\top)^\perp z = A^+Az,$$

где  $q \in \mathbb{R}^m$ ,  $z \in \mathbb{R}^n$ .

Если ранг  $A$  равен  $m$ , то из (П.3.1) следует

$$Ax_0 = AA^\top(AA^\top)^{-1}b = b,$$

т.е.  $x_0$  — по меньшей мере одно из решений системы (П.3.4).

Если ранг  $A$  равен  $n$ , то из (П.3.2) получим

$$Ax_0 = AA^+b = A(A^\top A)^{-1}A^\top b = A\|b;$$

здесь если решение (П.3.4) существует, то оно единственно и представимо в виде

$$x_0 = (A^\top A)^{-1}A^\top b = \text{pr}(b, \text{im } A).$$



## Литература

1. *Айда-Заде К.Р., Евтушенко Ю.Г.* Быстрое автоматическое дифференцирование на ЭВМ // Математическое Моделирование. 1989. Т. 1. С. 121–139.
2. *Айзекс Р.* Дифференциальные игры. М.: Мир, 1967.
3. *Албу А.Ф.* Вычисление градиента в задачах оптимального управления с разрывной правой частью // Ж. вычисл. матем. и матем. физ. 1995. Т. 35. С. 1058–1066.
4. *Албу А.Ф., Zubov В.И.* Оптимальное управление процессом кристаллизации вещества // Ж. вычисл. матем. и матем. физ. 2004. Т. 44. № 1. С. 38–50.
5. *Албу А.Ф., Zubov В.И.* Оптимизация процесса плавления и кристаллизации вещества. М.: ВЦ РАН, 2004.
6. *Албу А.Ф., Zubov В.И.* О выборе функционала и разностной схемы при решении задачи оптимального управления процессом кристаллизации металла // Ж. вычисл. матем. и матем. физ. 2011. Т. 51. № 1. С. 24–38.
7. *Албу А.В., Албу А.Ф., Zubov В.И.* Вычисление градиента функционала в одной задаче оптимального управления сложной динамической системой // Ж. вычисл. матем. и матем. физ. 2011. Т. 51. № 5. С. 814–833.
8. *Албу А.В., Албу А.Ф., Zubov В.И.* Математическое моделирование и управление процессом кристаллизации металла в литейном деле. М.: ВЦ РАН, 2012.
9. *Алексеев В.М., Тихомиров В.М., Фомин С.В.* Оптимальное управление. М.: Физматгиз, 1979.
10. *Андреева Е.А.* Оптимальное управление системами с запаздывающим аргументом. М.: ВЦ АН СССР, 1987.
11. *Андреева Е.А., Евтушенко Ю.Г.* Численные методы решения задач оптимального управления для систем, описываемых интегро-дифференциальными уравнениями типа Фредгольма / Сборник трудов ВНИИ системных исследований. М., 1989. Вып. 1. С. 4–13.
12. *Андреева Е.А., Цирулева В.М.* Вариационное исчисление и методы оптимизации. Тверь, 2001.
13. *Антипин А.С.* Методы решения систем задач выпуклого программирования // Ж. вычисл. матем. и матем. физ. 1987. Т. 27. № 3. С. 368–376.
14. *Антипин А.С.* О сходимости и оценках скорости сходимости проксимальных методов к неподвижным точкам экстремальных отображений // Ж. вычисл. матем. и матем. физ. 1995. Т. 35. № 5. С. 688–704.
15. *Антипин А.С., Артемьева Л.А., Васильев Ф.П.* Регуляризованный экстраградиентный метод решения параметрической многокритериальной задачи равновесного программирования // Ж. вычисл. матем. и матем. физ. 2010. Т. 50. № 12. С. 1–16.
16. *Арутюнов А.В., Магарил-Ильяев Г.Г., Тихомиров В.М.* Принцип максимума Понтрягина / Методы современной математики. Вып. 1. М.: Факториал Пресс, 2006.
17. *Афанасьев А.П., Дикусар В.В., Милютин А.А., Чуканов С.А.* Необходимое условие в оптимальном управлении. М.: Наука, 1990. С. 7–157.
18. *Ащепков Л.Т.* Оптимальное управление разрывными системами. Новосибирск: Наука, 1987.
19. *Бертсекас Д.* Условная оптимизация и методы множителей Лагранжа. М.: Радио и связь, 1987.

20. *Болтянский В.Г.* Математические методы оптимального управления. М.: Наука, 1969.
21. *Болтянский В.Г.* Оптимальное управление дискретными системами. М.: Наука, 1973.
22. *Брежнева О.А., Евтушенко Ю.Г., Третьяков А.А.* Новые численные методы и некоторые прикладные аспекты теории  $p$ -регулярности // *Ж. вычисл. матем. и матем. физ.* 2006. Т. 46. № 11. С. 1987–2000.
23. *Васильев Ф.П.* Лекции по методам решения экстремальных задач. М.: изд. МГУ, 1974.
24. *Васильев Ф.П.* Методы оптимизации. М.: Факториал Пресс, 2002.
25. *Васильев Ф.П., Ивануцкий А.Ю.* Линейное программирование. М.: Факториал, 2003.
26. *Величенко В.В.* О задачах оптимального управления для уравнений с разрывными правыми частями // *Автоматика и телемеханика.* 1966. Т. 7. С. 20–30.
27. *Габасов Р., Кириллова Ф.М.* К вопросу о распространении принципа максимума Л.С. Понтрягина на дискретные системы // *Автоматика и телемеханика.* 1966. Т. 11. С. 46–51.
28. *Гаранжа В.А., Голиков А.И., Евтушенко Ю.Г., Нгуен М.Х.* Параллельная реализация метода Ньютона для решения больших задач линейного программирования // *Ж. вычисл. матем. и матем. физ.* 2009. Т. 49. № 8. С. 1–17.
29. *Голиков А.И., Коткин Г.Г.* Характеристика множества оптимальных оценок задачи многокритериальной оптимизации // *Ж. вычисл. матем. и матем. физ.* 1988. Т. 28. № 10. С. 1461–1474.
30. *Голиков А.И., Евтушенко Ю.Г.* Отыскание нормальных решений в задачах линейного программирования // *Ж. вычисл. матем. и матем. физ.* 2000. Т. 40. № 12. С. 1766–1786.
31. *Голиков А.И., Евтушенко Ю.Г.* Теоремы об альтернативах и их применение в численных методах // *Ж. вычисл. матем. и матем. физ.* 2003. Т. 43. № 3. С. 354–375.
32. *Голиков А.И., Евтушенко Ю.Г.* Метод решения задач линейного программирования большой размерности // *Докл. РАН.* 2004. Т. 397. № 6. С. 727–732.
33. *Голиков А.И., Евтушенко Ю.Г.* Нахождение проекции заданной точки на множество решений задачи линейного программирования // *Тр. ин-та матем. и механики УрО РАН.* 2008. Т. 14. № 2. С. 33–47.
34. *Голиков А.И., Евтушенко Ю.Г., Моллаверди Н.* Применение метода Ньютона к решению задач линейного программирования большой размерности // *Ж. вычисл. матем. и матем. физ.* 2004. Т. 44. № 9. С. 1564–1573.
35. *Грачев Н.И., Евтушенко Ю.Г.* Библиотека программ для решения задач оптимального управления // *Ж. вычисл. матем. и матем. физ.* 1979. Т. 19. № 2. С. 367–387.
36. *Грачев Н.И., Евтушенко Ю.Г.* Применение метода сингулярных возмущений для решения минимаксных задач // *Докл. АН СССР.* 1977. Т. 233. № 3. С. 277–280.
37. *Грачев Н.И., Евтушенко Ю.Г.* Пакет программ для решения задач оптимального управления // М.: ВЦ АН СССР, 1978.
38. *Грачев Н.И., Фильков А.Н.* Алгоритмические основы оптимизации управляемых систем с разрывной правой частью. М.: ВЦ АН СССР, 1988.
39. *Грачев Н.И., Фильков А.Н.* Решение задач оптимального управления в системе ДИСО. М.: ВЦ АН СССР, 1986.

40. *Данскин Дж.* Теория максимина. М.: Советское радио, 1970.
41. *Демьянов В.Ф.* Условия экстремума и вариационное исчисление. М.: Высшая школа, 2005.
42. *Демьянов В.Ф., Малоземов В.Н.* Введение в минимакс. М.: Наука, 1972.
43. *Денисов А.А., Карманов В.Г., Третьяков А.А.* Ускоренный метод Ньютона для решения функциональных уравнений // ДАН СССР. 1985. Т. 281. № 6. С. 1293–1297.
44. *Дикусар В.В., Милютин А.А.* Качественные и численные методы в принципе максимума. М.: Наука, 1989. С. 11–114.
45. *Дмитрук А.В.* К обновлению метода скользящих режимов для задачи оптимального управления со смешанными ограничениями // Функциональный анализ и его приложения. 1976. Т. 10, 3, С. 39–44.
46. *Дубовицкий А.Я., Милютин А.А.* Задачи на экстремум при наличии ограничений. // Ж. вычисл. матем. и матем. физ. 1965. Т. 5, 3. С. 395–453.
47. *Дубовицкий А.Я., Милютин А.А.* Необходимые условия слабого экстремума в общей задаче оптимального управления. М.: Наука, 1971.
48. *Дубовицкий А.Я., Милютин А.А.* Необходимые условия экстремума в некоторых линейных задачах со смешанными ограничениями. // Вероятностные процессы и управление. М.: Наука, 1978. С. 42–74.
49. *Дубовицкий А.Я., Милютин А.А.* Теория принципа максимума // Методы теории экстремальных задач в экономике. М.: Наука, 1981. С. 6–47.
50. *Евтушенко Ю.Г.* Методы решения экстремальных задач и их применение в системах оптимизации. М.: Наука, 1982.
51. *Евтушенко Ю.Г.* Приближенный расчет оптимального управления // Прикладная математика и механика. 1970. Т. 34. № 1. С. 95–104.
52. *Евтушенко Ю.Г., Засухина Е.С., Зубов В.И.* О численном подходе к оптимизации решения задачи Бюргерса с помощью граничных условий // Ж. вычисл. матем. и матем. физ. 1997. Т. 37. № 12. С. 1449–1458.
53. *Евтушенко Ю.Г., Засухина Е.С., Зубов В.И.* Вычисление вторых производных сложной функции с помощью обобщенной БАД-методологии. М.: ВЦ РАН, 2005.
54. *Евтушенко Ю.Г., Веселов Е.Н., Мазурик В.П.* Пакет прикладных программ для решения оптимизационных задач / Численные методы механики сплошной среды. Т. 12. № 3. С. 7–23. М.: Знание, 1989.
55. *Евтушенко Ю.Г., Мазурик В.П.* Программное обеспечение систем оптимизации. М.: Знание, 1989.
56. *Евтушенко Ю.Г., Малкова В.У., Станевичюс А.А.* Параллельный поиск глобального экстремума функций многих переменных // Ж. вычисл. матем. и матем. физ. 2009. Т. 49. № 2. С. 255–269.
57. *Евтушенко Ю.Г., Посыпкин М.А.* Метод неравномерных покрытий для решения задач многокритериальной оптимизации с гарантированной точностью // Ж. вычисл. матем. и матем. физ. 2013. Т. 53. № 2. С. 39–54.

58. *Евтушенко Ю.Г., Третьяков А.А.* Элементарное доказательство конструктивного варианта теоремы о касательных направлениях и теоремы о неявной функции // ДАН. 2012. Т. 442. № 2. С. 156–161.
59. *Егоров А.И.* Оптимальное управление тепловыми и диффузионными процессами. М.: Наука, 1978.
60. *Еремин И.И.* Теория линейной оптимизации. Екатеринбург: НИСО УрО РАН, 1998.
61. *Измаилов А.Ф., Третьяков А.А.* Факторанализ нелинейных отображений. М.: Наука, 1994.
62. *Ириарт-Урурти Ж.Б.* Оптимизация и выпуклый анализ. Сборник задач и упражнений. Киев: КИТ, 2004.
63. *Жадан В.Г.* Метод модифицированной функции Лагранжа для задач многокритериальной оптимизации // Ж. вычисл. матем. и матем. физ. 1988. Т. 28. № 11. С. 1603–1618.
64. *Карманов В.Г.* Математическое программирование. М.: Физматлит, 2000.
65. *Ким К., Нестеров Ю., Скоков В., Черкасский Б.* Эффективный алгоритм для дифференцирования и задачи экстремали // Экономика и математические методы, 1984. Т. 20. С. 309–318.
66. *Крылов И.А., Черноусько Ф.Л.* О методе последовательных приближений для решения задач оптимального управления // Ж. вычисл. матем. и матем. физ. 1962. Т. 2. № 6. С. 669–683.
67. *Лэддон Л.С.* Оптимизация больших систем. М.: Наука, 1975.
68. *Марчук Г.И.* Сопряженные уравнения и анализ сложных систем. М.: Наука, 1992.
69. *Милютин А.А.* Принцип максимума в общей задаче оптимального управления. М.: ФизМат-Лит, 2001.
70. *Моисеев Н.Н.* Численные методы в теории оптимальных систем. М.: Наука, 1971.
71. *Моисеев Н.Н.* Как далеко до завтрашнего дня... Свободные размышления. 1917–1993. М.: АО Аспект пресс, 1994.
72. *Остроградский М.В.* Избранные труды. М.: Изд-во Академии наук СССР, 1958. С. 129–150.
73. *Поллак Э.* Численные методы оптимизации. Единый подход. М.: Мир, 1974.
74. *Поляк Б.Т.* Введение в оптимизацию. М.: Наука, 1983.
75. *Понтрягин Л.С., Болтянский В.Г., Гамкрелидзе Р.В., Мищенко Е.Ф.* Математическая теория оптимального процесса. М.: Наука, 1961.
76. *Пропой А.И.* О принципе максимума для дискретных систем управления // Автоматика и телемеханика, 1965. № 7. С. 1177–1187.
77. *Пропой А.И.* Элементы теории оптимальных дискретных процессов. М.: Наука, 1973.
78. *Розоноэр Л.И.* Принцип максимума Л.С. Понтрягина в теории оптимальных систем // Автоматика и телемеханика, 1959. Т. 20. № 10. С. 1320–1324; № 11. С. 1441–1458; № 12. С. 1561–1578.
79. *Роч П.* Вычислительная гидродинамика. М.: Мир, 1980.
80. *Самарский А.А.* Введение в теорию разностных схем. М.: Наука, 1971.

81. *Срочко В.А.* Итерационные методы решения задач оптимального управления. М.: Физматлит, 2000.
82. *Степанов В.В.* Теория обыкновенных дифференциальных уравнений. М.: Физматгиз, 1958.
83. *Тер-Крикоров А.М., Шабунин М.И.* Курс математического анализа. М.: Издательство МФТИ, 2000.
84. *Третьяков А.А.* Нелинейные методы решения экстремальных задач. Дисс. на соиск. уч. степ. канд. физ.-мат. наук. Москва, 1983.
85. *Третьяков А.А.* Две схемы нелинейного метода оптимизации в экстремальных задачах // Журн. вычисл. мат. и матем. физ. 1984. Т. 24. № 7. С. 986–992.
86. *Третьяков А.А.* Элементарное доказательство классических необходимых условий оптимальности для задачи математического программирования с ограничениями-неравенствами // ДАН. 2009. Т. 425. № 4. С. 459–461.
87. *Филлипов А.Ф.* Дифференциальные уравнения с разрывной правой частью. М.: Наука, 1985.
88. *Фан Л.Т., Вань Ц.С.* Дискретный принцип максимума. М.: Мир, 1967.
89. *Фихтенгольц Г.М.* Курс дифференциального и интегрального исчисления. М.: Наука, 1959.
90. *Чебышев П.Л.* Собрание сочинений. М.-Л.: Изд-во АН СССР, 1951. Т. 5. С. 7–25.
91. *Черноусько Ф.Л., Баничук Н.В.* Вариационные задачи механики и управления. М.: Наука, 1973.
92. *Чигурь С. А.* Об игровой задаче о долинхобрахистохроне. // Прикладная математика и механика. 1976. Т. 40. № 6. С. 1003–1012.
93. *Шалашилин В. И., Кузнецов Е. Б.* Метод продолжения решения по параметру и наилучшая параметризация. М.: Эдиториал УРСС, 1999.
94. *Baur W. and Strassen V.* The complexity of partial derivatives // Theoretical Computer Sciences, 1983. V. 22. Pp. 317–320.
95. *Bertsekas D.* Constrained optimization and Lagrange multiplier methods. New York: Academic Press, 1982.
96. *Birgin E.G., Evtushenko Y.G.* Automatic differentiation and spectral projected gradient methods for optimal control problem // Optimization methods and software. 1998. V. 10. Pp. 125–146.
97. *Brodal G.S.* Cache-oblivious algorithms and data structures // Proc. 9th Scandinavian Workshop on Algorithm Theory. Vol. 3111 of Lect. Not. in Comput. Sci. Berlin: Springer, 2004. P. 3–13.
98. *Caratheodory C.* Variationsrechnung und partielle Differentialgleichungen erster Ordnung. Band 1. Leipzig, 1956.
99. *Coleman T.F., Czyzyk J., Sun C., Wagner M., Wright S.J.* pPCx: Parallel Software for Linear programming // Proc. Eighth SIAM Conf. Parallel Proc. Sci. Comput., PPSC 1997, March 14–17, 1997. Minneapolis, Minnesota, USA: SIAM, 1997.
100. *Frigo M., Leiserson C.E., Prokop H., Ramachandran S.* Cache-oblivious algorithms // Proc. 40th Ann. Symposium on Foundations of Computer Sci. New York, October, 1999. P. 285–297

101. *Griewank A.* Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation // Optimization Methods and Software, 1992. V. 1. Pp. 35–54.
102. *Griewank A.* On automatic differentiation / In: M. Iri and K. Tanabe (Eds.) Mathematical Programming: Recent Developments and Applications. Kluwer Academic Publishers, 1989. Pp. 83–108.
103. *Griewank A.* Evaluating Derivatives. Philadelphia: SIAM, 2000.
104. *Griewank A. and Corliss G.F., eds.* Automatic Differentiation of Algorithms. Theory, Implementation and Application. Philadelphia: SIAM, 1991.
105. *Evtushenko Y.G., Golikov A.I.* New perspective on the theorems of alternative / High Performance algorithms and software for Nonlinear Optimization. Eds.: G. Di Pillo and A. Murli. Kluwer Academic Publishers B.V. Pp. 227–241.
106. *Evtushenko Y.G., Zasukhina E.S. and Zubov V.I.* FAD Method to Compute Second Order Derivatives / Automatic Differentiation of Algorithms. From Simulation to Optimization. Springer-Verlag, New York, Inc., 2002. Pp. 327–333.
107. *Evtushenko Yu., Mazourik V. and Ratkin V.* Multicriteria optimization in the DISO system /In: Optimization, Parallel Processing and Applications / Lecture Notes in Control and Information Sciences. Springer-Verlag. V. 113. Pp. 231–240.
108. *Evtushenko Yu., Posypkin M., Sigal I.* A framework for parallel large-scale global optimization // Computer Science – Research and Development. 23(3). Pp. 211–215.
109. *Evtushenko Yu., Posypkin M.* A deterministic approach to global box-constrained optimization // Optimization Letters, DOI: 10.1007/s11590-012-0452-1, 2012.
110. *Evtushenko Yu.G., Rubinov A.M. and Zhadan V.G.* General Lagrange-type functions in constrained global optimization. Part II: Exact auxiliary functions // Opt. Meth. and Software. V. 16. Pp. 231–256.
111. *Evtushenko Y.G.* Automatic differentiation viewed from optimal control theory / In [102]. 1991. Pp. 25–30.
112. *Evtushenko Y.* Computation of exact gradients in distributed dynamic systems for optimal control problem // Optimization methods and software, 1998. V. 9. Pp. 45–75.
113. *Evtushenko Y.* Generalized Lagrange Multiplier Technique for Nonlinear Programming // J. of Optim. Theory and Appl. 1977. V. 21. № 2. Pp. 121–135.
114. *Evtushenko Y.G.* Numerical Optimization Techniques. New York: Optimization Software, Inc., Publications Division, 1985.
115. *Evtushenko Y.G., Posypkin M.* A deterministic approach to global box-constrained optimization // Optimization Letters. 2013. V 7. №4. Pp. 819–829.
116. *Iri M.* History of Automatic differentiation and rounding error estimation / In [102], 1991. Pp. 3–16.
117. *Iri M.* Simultaneous computation of functions, partial derivatives and estimates of rounding errors – Complexity and practicality // Japan Journal of Applied Mathematics. 1984. V. 1. Pp. 223–252.
118. *Iri M. u Kubota K.* Methods of fast automatic differentiation and applications // Research memorandum RMI 87-02. Department of Mathematical Engineering and Instrumental Physics, Faculty of Engineering, University of Tokyo, 1987.

119. *Kanzow C., Qi H., Qi L.* On the Minimum Norm Solution of Linear Programs // J. Optimizat. Theory and Appl. 2003. V. 116. P. 333–345.
120. *Lellouche J.M., Devenon J.L., Dekeyser I.* Boundary control of Burger's equation — A numerical approach // Computers and Mathematics with Applications, 1994. V. 28. № 5. Pp. 33–44.
121. *Luenberger D.G.* Linear and nonlinear programming. Addison-Wesley Publ. Comp., 1984.
122. *Mangasarian O.L.* A Newton Method for Linear Programming // J. of Optimizat. Theory and Appl. 2004. V. 121. P. 1-18.
123. *Mangasarian O.L.* Nonlinear programming. Philadelphia: SIAM, 1994.
124. *Meszaros Cs.* The BPMPD interior point solver for convex quadratic programming problems // Optimizat. Meth. and Software. 1999. V. 11&12. P. 431–449.
125. *Ostrogradski M.* Considérations générales sur les momen(t)s des forces. [Общие соображения о моментах сил] /Mém. de l'Acad. des sc. de St.-Pbg., VI sér., sc. math. et phys. V. 1. 1835–1838. Pp. 129–150.
126. *Ostrogradski M.* Memoire sur les Deplacements Instatanes des Systemes Assujettis a des Conditions Variables. Memoiros des L'academie Imperiable des Sciences de Saint-Petersburg, Sixieme Serie. 1838. V. 1. Pp. 565–600.
127. *Precopa Andras.* On the development of optimization theory. Amer. Math. Monthly. 1980. V. 87.
128. *Tuy H.* DC-Optimization and Robust Global Optimization // J. Global Optim. 2010. V. 47. Pp. 435–501. Academic Press, Inc., 1984.

## Оглавление

<b>Основные обозначения и аббревиатуры</b>	<b>3</b>
<b>Введение</b>	<b>5</b>
<b>Глава 1. Дифференцирование сложных функций</b>	<b>11</b>
1.1. Вычисление первых производных . . . . .	11
1.2. Вычисление вторых производных . . . . .	15
1.3. Вычисление первых производных в случае многошаговых процессов . . . . .	17
1.4. Дифференцирование элементарных функций . . . . .	23
1.5. Оценка погрешностей . . . . .	29
1.6. Приложение к линейной алгебре . . . . .	30
Комментарии к главе 1 . . . . .	31
<b>Глава 2. Приложения к задачам оптимизации</b>	<b>32</b>
2.1. Задачи безусловной минимизации . . . . .	32
2.2. Обобщенный критерий Сильвестра . . . . .	36
2.3. Нахождение локального минимакса и максимина . . . . .	44
2.4. Задачи нелинейного программирования с ограничениями типа равенства . . . . .	46
2.5. Функция чувствительности . . . . .	50
2.6. Теорема Куна–Таккера . . . . .	54
2.7. Общая задача нелинейного программирования . . . . .	59
2.8. Условия оптимальности второго порядка . . . . .	63
2.9. Применение методов решения систем нелинейных уравнений к задачам НЛП . . . . .	66
2.10. Некоторые свойства минимаксных задач . . . . .	70
Комментарии к главе 2 . . . . .	73
<b>Глава 3. Приложение к задачам оптимального управления</b>	<b>75</b>
3.1. Производные по начальным условиям . . . . .	75
3.2. Системы с разрывными правыми частями . . . . .	82
3.3. Принцип максимума Л.С. Понтрягина . . . . .	83
3.4. Дискретный принцип максимума . . . . .	92
3.5. Дискретный принцип квазимаксимума . . . . .	97
3.6. Дискретный вариант задач оптимального управления с разрывными правыми частями . . . . .	98
3.7. Численные методы решения задач оптимального управления . . . . .	104
3.8. Три простейших примера . . . . .	108
3.9. Локально оптимальные управления . . . . .	112
Комментарии к главе 3 . . . . .	113
<b>Глава 4. Оптимизация систем, описываемых уравнениями с частными производными</b>	<b>115</b>
4.1. Оптимизация решения задачи Бюргерса с помощью граничных условий . . . . .	115
4.2. Оптимальное управление параболической системой . . . . .	123
4.3. Оптимальное управление гиперболической системой . . . . .	126
Комментарии к главе 4 . . . . .	129
<b>Заключение</b>	<b>130</b>
<b>Приложение 1</b>	<b>131</b>
<b>Приложение 2</b>	<b>132</b>
<b>Приложение 3</b>	<b>134</b>
<b>Литература</b>	<b>137</b>