

В

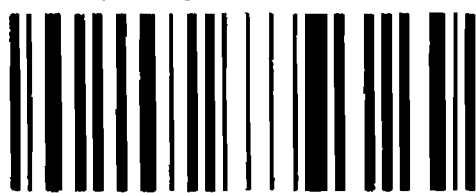
ДЛЯ ВУЗОВ

Р.К.Чуян

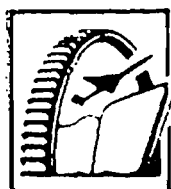
МЕТОДЫ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ДВИГАТЕЛЕЙ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

*Допущено Министерством высшего
и среднего специального образования СССР
в качестве учебного пособия
для студентов авиадвигателестроительных
специальностей вузов*

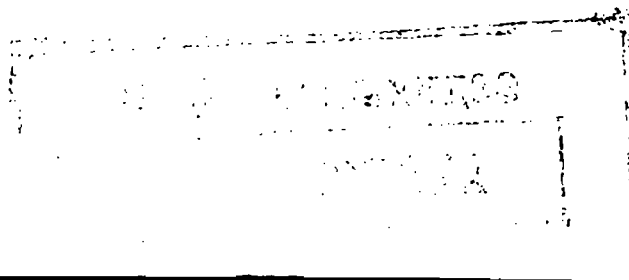
МАИ (ГТУ) НТБ



00263994



130343



МОСКВА
«МАШИНОСТРОЕНИЕ»
1988

ББК 39.55

Ч-81

УДК 629.7.03(0.75.8)

Рецензенты: кафедра энергоустановок и двигателей летательных аппаратов Харьковского авиационного института и
д-р техн. наук проф. *А. А. ПОРОТНИКОВ*

Чуян Р. К.

Ч-81 Методы математического моделирования двигателей летательных аппаратов. Учеб. пособие для студентов авиадвигательностроительных специальностей вузов. — М.: Машиностроение, 1988. — 288 с.: ил.

ISBN 5-217-00131-3

Рассмотрены методы математического моделирования в оптимизации двигателей летательных аппаратов и их элементов на современных электронных цифровых и аналоговых вычислительных машинах.

Последовательно изложены принципы построения математических моделей и приведены основные их типы, используемые при изучении и проектировании двигателей летательных аппаратов.

Ч $\frac{3606030000-123}{038(01)-88}$ 123-88

ББК 39.55

ISBN 5-217-00131-3 © Издательство «Машиностроение», 1988

Вторая половина XX века связана с появлением и широким распространением новой методологии исследования сложных объектов и систем, методологии, краеугольным камнем которой стал математический эксперимент, метод математического моделирования. Математические модели использовались и раньше. В частности, они достаточно давно и широко применяются в механике, астрономии, физике. Однако считалось, что методы математического моделирования не пригодны для исследования сложных технических, экономических, биологических и социальных систем. В области техники отсутствие объективных математических методов привело, с одной стороны, к созданию многочисленных частных, так называемых инженерных методик расчета, носивших рецептурный характер, а с другой — к почти безраздельному господству эмпирики. Недостаточно полная проработка вариантов проектируемых систем и отсутствие четких критериев для их сравнения часто приводили к субъективным решениям.

Положение начало меняться во второй половине нашего столетия. С одной стороны, непрерывное усложнение и увеличение стоимости социальных, экономических и технических программ практически исключает возможность экспериментирования, делает слишком большой цену возможной ошибки. С другой стороны, развитие средств вычислительной техники, в частности, современных электронных цифровых вычислительных машин, дает в руки исследователей новое эффективное средство моделирования сложных систем. Можно с уверенностью утверждать, что в настоящее время не существует объектов, при изучении которых не применялись бы методы математического моделирования. Разработаны и используются математические модели разнообразных технологических процессов, технических устройств, физические модели и экономические модели предприятий, регионов и целых государств, экологические модели, модели геологических и геофизических процессов, модели социальных систем, биологические и медицинские модели.

Математическое моделирование позволяет до создания реальной системы (объекта) или возникновения реальной ситуации рассмотреть возможные режимы работы, выбрать оптимальные управляющие воздействия, составить объективный прогноз будущих состояний системы. Математический эксперимент помогает увидеть за частным общее, развить универсальные методы управления объектами различной физической природы, познать общие законы развития материи. Наконец, математические модели являются центральным звеном интенсивно разрабатываемых автоматизированных систем проектирования, управления и обработки данных.

Несмотря на наличие хорошо разработанных и достаточно точных методов математическое моделирование в значительной

мере является искусством, а его результаты зависят от интуиции, таланта и просто удачи исследователя.

Предлагаемая читателю книга представляет собой учебное пособие по методам математического моделирования двигателей и двигательных установок летательных аппаратов и основана на курсе лекций, в течение ряда лет читавшихся студентам соответствующего профиля. Поэтому при отборе материала и методов изложения учитывались особенности учебных планов и подготовка студентов, полученная ими при изучении смежных дисциплин, в частности, математики.

При подготовке пособия использовались многочисленные литературные источники. Краткая библиография, приведенная в конце книги, охватывает лишь некоторые из них и предназначена для студентов, желающих получить дополнительную информацию по отдельным вопросам, затронутым в пособии. Автор считает необходимым отметить, что при подготовке ряда разделов пособия широко использовались работы В. В. Налимова, Л. А. Растригина, А. Г. Ивахненко, Д. Дж. Уайлда и М. Месаровича.

Предполагается, что читатели владеют основами высшей математики в объеме программы технических вузов и, в частности, имеют некоторое представление о векторном исчислении, элементах теории матриц, основах теории вероятностей и математической статистики, обыкновенных дифференциальных уравнениях и дифференциальных уравнениях в частных производных.

Автор считает своим приятным долгом выразить благодарность рецензентам пособия профессорам А. А. Поротникову и Н. В. Белану за рекомендации, способствовавшие улучшению книги. Автор будет благодарен читателям за замечания и пожелания, относящиеся к содержанию пособия, которые следует направлять по адресу: издательство «Машиностроение», 107076, Москва, Б-76, Стромынский пер., 4.

ОБЩИЕ ПРИНЦИПЫ ПОСТРОЕНИЯ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ДВИГАТЕЛЕЙ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

Непрерывное усложнение двигателей и двигательных установок летательных аппаратов, развитие математических методов исследования и современной вычислительной техники привели к возникновению новой методологии проектирования, получившей свое воплощение в так называемых системах автоматизированного проектирования двигательных установок. Центральным звеном систем автоматизированного проектирования является математическая модель двигательной установки, а методом исследования — математический эксперимент на этой модели. В настоящем разделе рассмотрены общие принципы построения и использования математических моделей сложных систем, а также методы проверки их адекватности действительности.

ГЛАВА 1. МАТЕМАТИЧЕСКИЕ МОДЕЛИ В СИСТЕМАХ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

1.1. СИСТЕМА АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ

Непрерывное совершенствование двигательных установок (ДУ) летательных аппаратов (ЛА), все более высокие требования, предъявляемые к их характеристикам, возрастающие объемы используемой информации приводят к непрерывному усложнению процесса их проектирования. Резко возрастают сроки и стоимость проектирования, а также число участвующих в нем специалистов. Этот процесс можно проиллюстрировать данными о разработке самолетов США и Франции (рис. 1.1, табл. 1.1) ¹.

За последние 30 лет сроки проектирования возросли примерно в 10 раз, стоимость разработки — более чем в 100 раз, затраты человеческого труда — почти в 1000 раз. Резко возросли также объем и стоимость стендовых и летных испытаний как отдельных узлов, так и летательного аппарата в целом. Такая же картина наблюдается при проектировании ракетно-космической техники и

¹ Машинное проектирование летательных аппаратов. «Итоги науки и техники» серия «Авиационное» том 3, ВИНТИ, М.: 1976. 204 с.

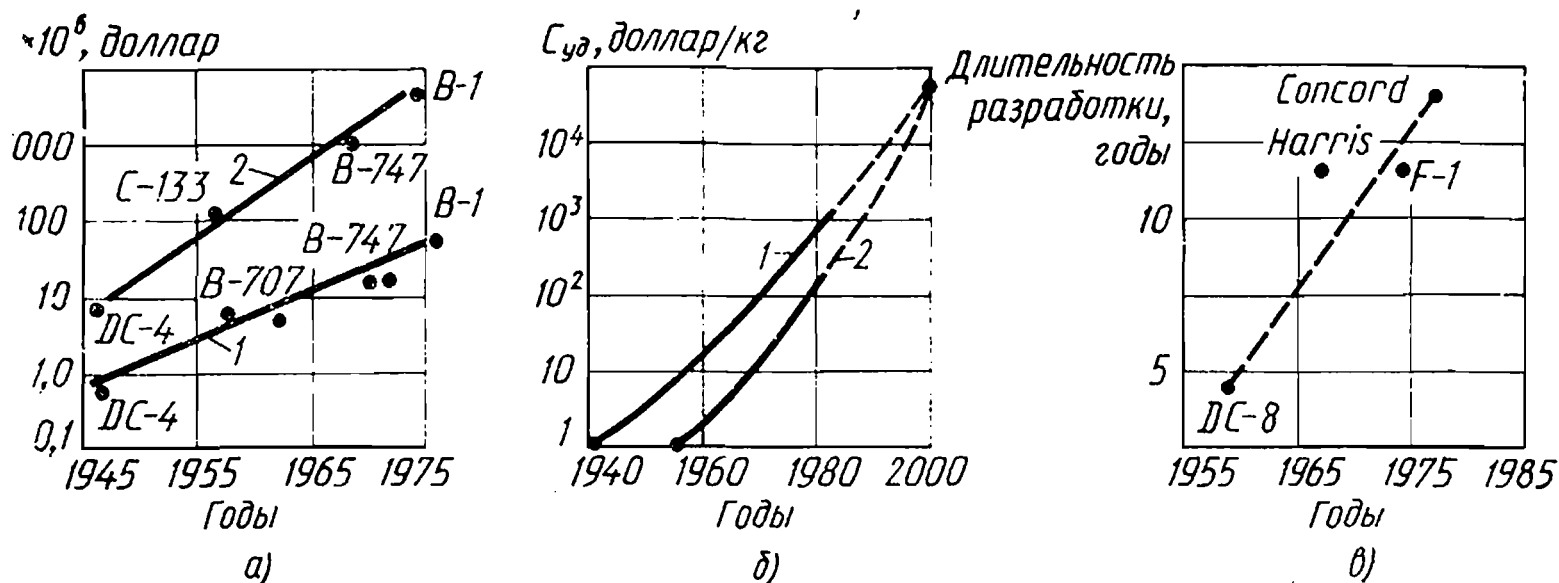


Рис. 1.1. Изменение характеристик самолетов:
 а — стоимость серийного самолета (1); стоимость разработки самолета (2); б — затраты на проектирование, отнесенные к 1 кг массы конструкции; 1 — общие; 2 — связанные с применением ЭВМ; в — сроки проектирования

двигателей летательных аппаратов. Все это приводит, с одной стороны, к моральному старению техники еще до ее эксплуатации, а с другой,— заставляет проектировщиков ограничиваться проработкой небольшого числа вариантов. Это, разумеется, затрудняет выбор оптимальных решений, особенно на стадии предэскизного проектирования при определении общей концепции, схемы, основных параметров и режимов работы изделия. Между тем, ошибки, допущенные на этой стадии, зачастую определяют неудачу проекта в целом. Возрастание общего объема информации, используемой при проектировании летательного аппарата и двигательной установки, усложняет проблемы хранения, обработки и передачи информации в ходе проектирования и ставит почти непреодолимые преграды на пути создания новой техники в рамках старой методологии проектирования.

В последние годы как в СССР, так и за рубежом интенсивно развиваются новые методы проектирования сложных технических объектов; в частности, двигателей летательных аппаратов, основанные на широком использовании современной вычислительной техники. Эти методы получили уже ставшее традиционным общее название методов автоматизированного проектирования. Системы,

Таблица 1.1

Год	Тип самолета	Цена серийного самолета $C \times 10^6$, долл.	Стоимость разработки самолета $C_{разр} \times 10^6$, долл.	Отношение стоимости разработки к цене $C_{разр}/C$
1936	DC-3	0,1	0,3	3 : 1
1947	DC-4	0,4	4	10 : 1
1958	DC-8	5	100	20 : 1
1969	B-747	17	1000	59 : 1
1972	DC-10	12	1050	87 : 1
1975	Concorde	40—47	2700—3000	65 : 1
1978	B-1	56	3300—4100	73 : 1

в которых реализуются эти методы, называются *системами автоматизированного проектирования* (САПР) [15].

Анализ основных затрат на проектирование двигательных установок и двигателей ЛА показал, что наиболее трудоемкими являются следующие этапы:

- выбор общей схемы и основных параметров двигателя;
- детальная разработка конструкции двигателя;
- испытание и доводка двигателя и его отдельных элементов.

Системы автоматизированного проектирования могут быть успешно использованы на всех этих этапах.

С помощью вычислительных машин можно относительно быстро и полно исследовать различные варианты схем двигателя и двигательной установки, подробно рассчитать и оценить наиболее перспективные из них, выбрать оптимальные режимы работы двигателя. Наконец, можно отказаться от проведения ряда испытаний, заменив реальный эксперимент численным математическим экспериментом, а в тех случаях, когда такая замена невозможна, ускорить проведение испытаний и снизить их стоимость за счет целенаправленного планирования эксперимента, автоматизации обработки экспериментальных данных.

При создании систем автоматизированного проектирования ставятся две задачи:

1. Улучшение качества проектирования за счет:

- расширения количества анализируемых вариантов возможных технических решений;
- детальной проработки рассматриваемых решений;
- изучения работы двигателя в предельных и запредельных ситуациях.

2. Снижение сроков, стоимости и трудоемкости разработки изделия за счет:

ускорения вычислительных работ на всех этапах проектирования;

передачи машине ряда рутинных операций, связанных, в частности, с выпуском технической документации, выбором стандартных элементов и т. д.;

сокращения объема экспериментальных работ;

создания емких и обеспечивающих легкий доступ хранилищ информации (банков данных).

Создание эффективных САПР представляет собой сложную задачу и требует решения ряда проблем. К ним относятся:

- совершенствование и формализация методов проектирования;
- разработка сложных и многообразных сервисных средств и необходимого для их работы математического обеспечения;
- разработка математических описаний (математических моделей) двигателя и его элементов.

Эффективные САПР могут быть созданы на базе больших вычислительных машин (ЕС-10-45, ЕС-10-46, ЕС-10-55, ЕС-10-61, ЕС-10-65, ЕС-10-66, «Эльбрус»), обладающих высоким быстродействием, большой памятью и развитой периферией. Такие системы

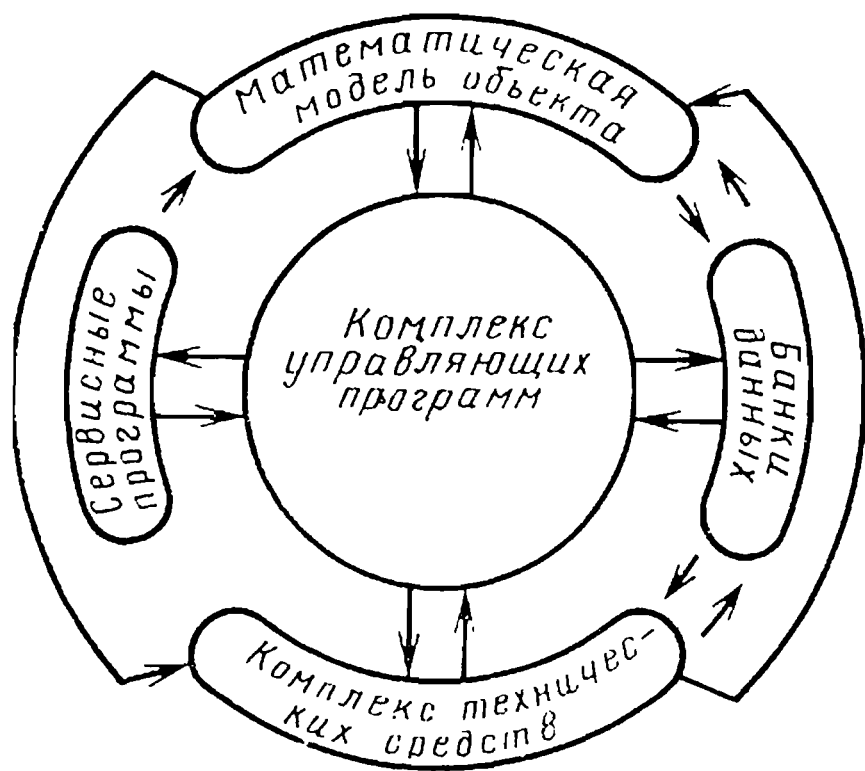


Рис. 1.2. Принципиальная схема САПР

(рис. 1.2) включают в себя устройства прямого доступа к ЭВМ, желательно с непосредственным вводом графической информации, устройства документирования (графопостроители, различные устройства печати), банки данных для хранения информации, многочисленные пакеты прикладных программ (оптимизации, статистической обработки и уплотнения информации и т. д.), комплексы

управляющих программ. Однако центральным звеном любой САПР является математическая модель (на самом деле речь идет о системе математических моделей) проектируемого изделия — ДЛА. Качество этой модели предопределяет качество проектирования, а следовательно, и эффективность методологии автоматизированного проектирования в целом. При этом, если сервисные системы, пакеты прикладных программ, некоторые управляющие программы универсальны, математическая модель объекта всегда уникальна, а ее разработка требует глубоких знаний его природы.

Вопросы для самоконтроля

1. Чем объясняется необходимость создания САПР?
2. Какие задачи должны решать САПР?
3. Какую роль в САПР играет математическая модель?

1.2. МАТЕМАТИЧЕСКИЕ МОДЕЛИ ДВИГАТЕЛЬНОЙ УСТАНОВКИ И ЕЕ ЭЛЕМЕНТОВ

С точки зрения математического описания двигательная установка современного ЛА представляет собой *сложную* или *большую* систему. В настоящее время это понятие не имеет четкого определения. Видимо, наиболее точным является определение сложной системы как объекта, для которого отсутствует *исчерпывающее априорное описание* и, следовательно, не может быть составлен точный прогноз дальнейших состояний. При этом подчеркивается не полное отсутствие описаний и прогнозов, а отсутствие исчерпывающей полноты этих описаний и необходимость принятия решений в условиях неопределенности.

В качестве других примеров сложных систем, изучаемых современной наукой, можно назвать человеческое общество, народное хозяйство страны или региона, человеческий мозг, экологические системы и т. д. Задачей исследования сложной системы в первую очередь является разработка более или менее адекватного, т. е.

соответствующего действительности, описания этой системы, выбор оптимальных условий ее функционирования и методов управления ею. Несмотря на глубокие физические различия все перечисленные выше сложные системы как *объекты управления* имеют общие закономерности. Эти закономерности являются предметом науки, получившей название *кибернетики*. В учебном пособии рассмотрен ряд вопросов составления и изучения математических моделей двигательных установок ЛА и поэтому оно может рассматриваться как некоторая частная глава технической кибернетики.

Сложность системы, в частности ДУ ЛА, является следствием ряда причин, к числу которых можно отнести:

множественность элементов и сложность структуры системы;
сложность и непредсказуемость взаимодействия системы с внешней средой;

неопределенность целевой функции, связанную с трудностями определения эффективности системы;

наличие конфликтов либо с другой сложной системой, либо внутри самой системы;

изменчивость характеристик системы, связанную как с изменениями условий внешней среды, так и со старением элементов системы.

Важно подчеркнуть, что оценка сложности системы в значительной мере зависит от точки зрения исследователя и задач исследования. Так, например, ДЛА несомненно является сложной системой с точки зрения инженера-двигателиста. В то же время для организатора производства, например, в рамках АСУ завода или отрасли, отдельная двигательная установка представляет собой простой элемент.

Графически любые, в том числе и сложные системы могут быть представлены в терминах взаимодействия типа вход — выход (рис. 1.3), позволяющих перейти от некоторого вектора входных параметров $x(x_1, x_2 \dots x_k)$ к вектору выходных характеристик $W(W_1, W_2, \dots W_l)$.

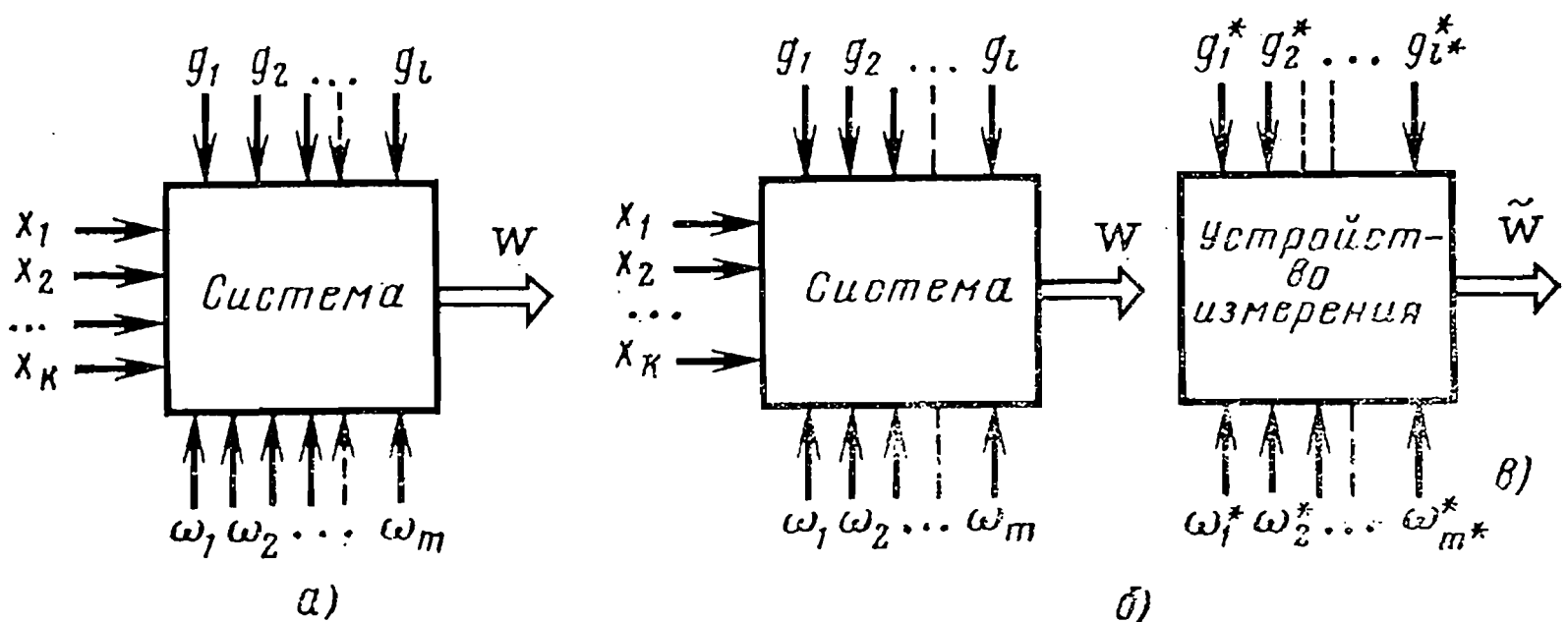


Рис. 1.3. Схема сложной системы:

а — принципиальная схема; б — сложная система с устройством измерения; x_i — управляющие переменные; W — характеристическая функция; ω_i — неопределенности (шум); g_i — ограничения

Существует ряд подходов к классификации сложных систем. Достаточно широкое распространение получила классификация систем по характеру их поведения, предусматривающая деление систем на *детерминированные, стохастические и целенаправленные*.

К детерминированным относятся системы с четко выраженными причинно-следственными связями, предопределяющими последующие состояния системы. Поведение стохастических систем носит вероятностный характер и определяется законами распределения случайных величин. К целенаправленным относятся системы, способные устанавливать свои цели и вырабатывать методы их достижения.

Сложные системы по их повторяемости делят на *массовые* (человек, ДЛА) и *уникальные* (человеческое общество, биосфера) системы. Наконец, сложные системы можно классифицировать по типу (языку) их описания.

С точки зрения разработчика двигательная установка (ДУ) современного ЛА несомненно представляет собой сложную систему. Работа ДУ определяется рядом сложнейших физических процессов, тесно связанных и взаимодействующих между собой. Современные ДУ включают в себя множество элементов и обладают сложной структурой. Их работа протекает в непрерывном взаимодействии с внешней средой, а также с другими подсистемами летательного аппарата. Противоречивые требования, предъявляемые к различным подсистемам ЛА, осложняют выбор целевой функции, приводят к конфликтным ситуациям и требуют принятия компромиссных решений. По своему характеру ДУ ЛА может рассматриваться как детерминированная (а при учете вероятностных характеристик, связанных, в частности, с условиями функционирования,— как стохастическая) массовая сложная система.

Изучение сложных систем требует последовательного комплексного системного подхода, предусматривающего не отдельное, а совместное исследование различных элементов, узлов и процессов в рамках системы в целом. Наиболее эффективным, а часто и единственным средством изучения сложных систем является метод их моделирования. *Моделью* называется любое представление объекта, системы или понятия в форме, отличной от формы их реального существования. Различают геометрические, физические и математические модели реальных систем.

Геометрические модели предназначены для отображения внешнего вида, конфигурации и взаимного расположения элементов изучаемых систем (макеты архитектурных сооружений и машин, машиностроительные чертежи и т. д.). *Физические* и *математические* модели предназначены для определения количественных характеристик системы. При физическом моделировании сохраняются основные физические процессы и условия функционирования системы (изучение обтекания самолетов в аэродинамических трубах, исследование двигателей и их элементов при стендовых испытаниях). Математической моделью называется математическое описание системы, позволяющее изучить эту систему средствами

математики. Именно математические модели ДЛА, их узлов и элементов рассматриваются в данном учебном пособии.

- Можно выделить четыре основные функции моделей [23]:
- модели как *средство познания* реального мира, выявляют характеристики объекта, его внутренние и внешние связи. Уже сама попытка составления модели позволяет обнаружить и устранить нелогичности и противоречия наших представлений, наметить пути дальнейших исследований;
 - модели как *средство передачи информации* и средство общения. Никакое вербальное описание не может сравниться по точности, компактности и объективности с моделью;
 - модели как *средство обучения и тренажа*, обеспечивающее приобретение надежных профессиональных навыков без риска возникновения критической ситуации (в частности, при подготовке шоферов, пилотов, операторов ядерных реакторов и т. д.);

Т а б л и ц а 1.2

Множество	ДУ с ВРД	ДУ с ЖРД	ДУ с ЭРД
Переменные X	Схема двигателя Наличие двухконтурности Степень форсажа Количество ступеней компрессора Расход рабочего тела	Схема двигателя Тип системы подачи рабочего тела Род рабочего тела Расход рабочего тела Давление в камере сгорания Степень расширения сопла	Количество тяговых блоков Тип ЭРД Ускоряющее напряжение Расход рабочего тела Напряженность магнитного поля Род рабочего тела
Промежуточные функции Y	Ресурс ДУ и ее элементов Характеристики компрессора КПД Вероятность безотказной работы	Ресурс ДУ и ее элементов Напорная характеристика турбонасосного агрегата Характеристика смесеобразования Вероятность безотказной работы	Ресурс ДУ и ее элементов Вероятность безотказной работы КПД Коэффициент использования рабочего тела
Шумы Ω	Разброс характеристик материалов и рабочих тел Разброс условий функционирования ДУ Разброс условий хранения и подготовки ДУ		
Результирующие функции W	Стоимость ДУ Стоимость пассажирокилометра (тоннокилометра) Масса ДУ	Стоимость достигнутого эффекта Массогабаритные характеристики	Стоимость программы (стоимость полезного эффекта) Масса (удельная масса) ДУ Относительная полезная нагрузка ЛА

— модели как средство *прогнозирования* состояния объекта и постановки математических экспериментов. Важнейшим достоинством математического эксперимента является возможность исследования множества вариантов при выборе оптимальной конструкции и условий работы изделия, а также изучения предельных (запредельных) режимов работы, реализация которых или невозможна, или крайне нежелательна (аварийные режимы работы ДУ, взрыв ядерной энергетической установки, космические катастрофы и т. д.).

В ряде случаев моделирование позволяет выбрать стратегию поведения в неожиданных ситуациях. Так, например, при полете космического корабля «Аполлон-13» математическая модель была использована для анализа возможных управлений в аварийной ситуации, возникшей в результате взрыва баллона с кислородом, и позволила выработать систему команд, обеспечивающих благополучное возвращение астронавтов на Землю.

Математически модель может быть представлена как некоторая функция

$$W = W(x, \omega); \omega : X \times \Omega \Rightarrow W, \quad (1.1)$$

или чаще как функционал

$$W = W(x, y(x), \omega); \omega : X \times Y \times \Omega \Rightarrow W; \quad (1.2)$$

$$x \in X; y \in Y; \omega \in \Omega,$$

осуществляющий отображение множества управляющих параметров X , множества промежуточных функций Y и множества случайных факторов (шумов) Ω на множество характеристических функций системы W .

Примеры элементов множеств X , Y , Ω , W для ДУ ЛА нескольких типов приведены в табл. 1.2.

Вопросы для самоконтроля

1. В чем причина сложности сложных систем?
2. Дайте классификацию сложных систем.
3. В чем заключается системный комплексный подход при изучении ДЛА?

1.3. КЛАССИФИКАЦИЯ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

Существует несколько схем классификации математических моделей. Все они достаточно условны. Одна из таких схем, используемая при классификации моделей ДЛА, показана на рис. 1.4.

Все математические модели можно разбить на *регулярные* и *модели идентификации*. Регулярные модели отражают реальные структуры и процессы в исследуемых объектах, т. е. опираются на теорию их работы. Модели идентификации строятся на основе изучения реакций объекта на изменение условий окружающей среды. При этом теория работы объекта не рассматривается, сам объект

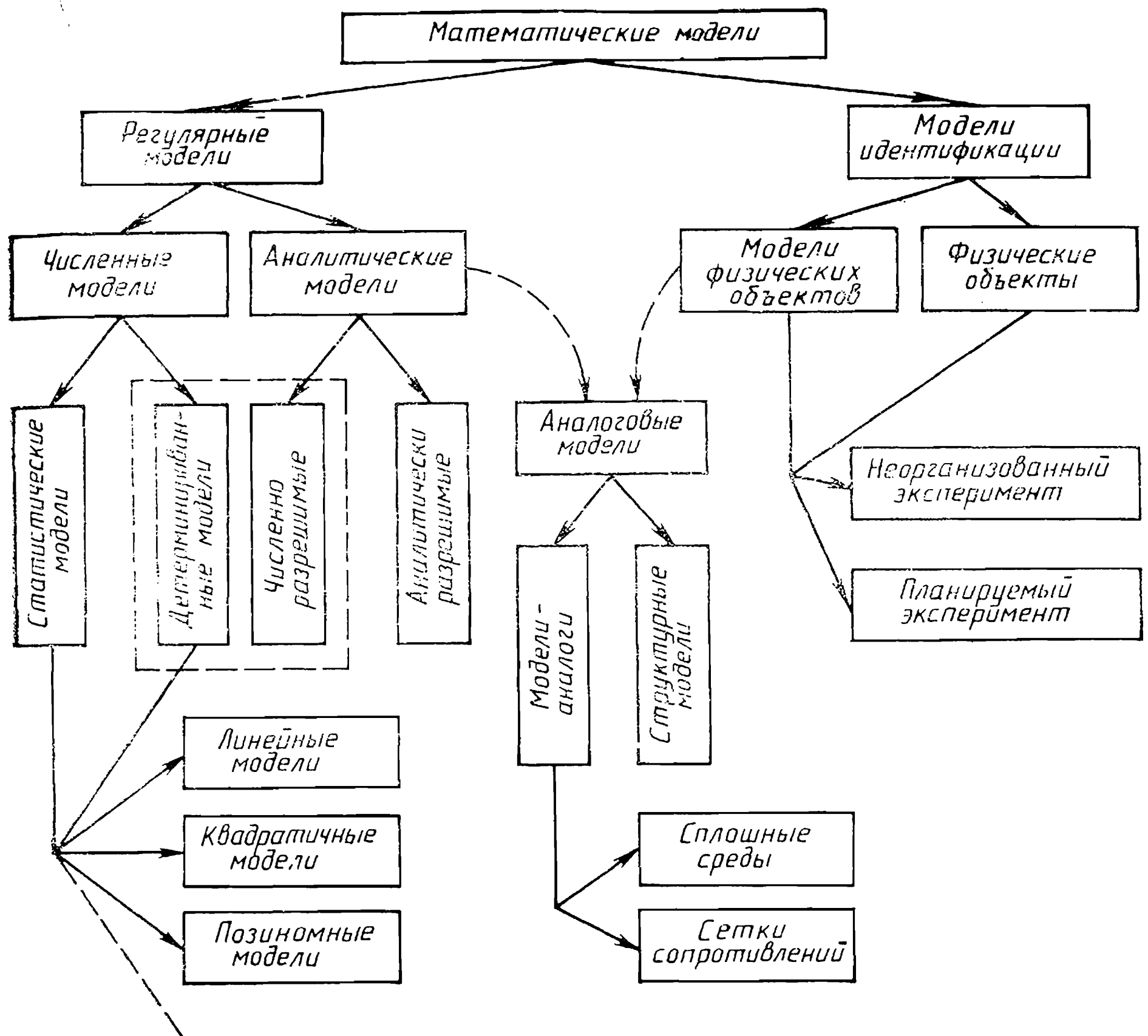


Рис. 1.4. Классификация моделей

представляет собой так называемый «черный ящик», а модель — некоторую интерполяционную зависимость.

Регулярные модели делят на *аналитические*, математическое описание которых допускает аналитическое исследование, и *численные*, исследования которых осуществляется численными методами. Если аналитическое исследование может быть доведено до конца, модели называются *аналитически разрешимыми*. В противном случае говорят о *численно разрешимых* аналитических моделях. Численные модели, в свою очередь, делятся на *детерминированные* и *статистические*. В детерминированных моделях все взаимосвязи, переменные и константы заданы точно, что приводит к однозначному определению результирующей функции. В статистических моделях задаются соответствующие законы распределения случайных величин, что приводит к вероятностной оценке этой функции. Наконец, по форме описания как детерминированные, так и статистические модели можно разделить на *линейные*, *квадратичные*, *нелинейные*, *поэтаные* и некоторые другие модели.

Модели идентификации могут быть построены на основе экспериментальных или численных данных. Эти данные получают непосредственно на исследуемых объектах или с помощью их физических и математических моделей. В САПР ДЛА модели идентификации играют исключительно важную роль для уплотнения информации при ее передаче от одного уровня иерархии к другому.

Наконец, несколько особняком стоят *аналоговые* модели, построенные на основе формальной общности математических описаний объектов различной физической природы. По характеру построения эти модели делят на *модели-аналоги* и *структурные аналоговые модели*.

Реальные модели, как правило, включают в себя различные элементы, а тип модели определяется или преобладающим типом зависимостей, или видом определяющих соотношений.

Важнейшей характеристикой моделей является их точность, *адекватность* действительности. При этом важно иметь в виду, что все модели представляют собой приближенные описания реальных объектов (процессов) и поэтому принципиально неточны. Интегральная оценка модели может быть получена путем сравнения результатов моделирования и экспериментальных данных для конкретных объектов или режимов. При этом часто используется так называемое ретроспективное моделирование уже имевших место состояний системы.

Для оценки *значимости* совпадения или несовпадения модельных и экспериментальных результатов широко используются методы математической статистики (см. гл. 3). Вместе с тем не следует переоценивать результаты такой проверки. Хорошее совпадение модельных и экспериментальных данных, вообще говоря, не доказывает точности модели, а лишь подтверждает ее функциональную пригодность для моделирования. Всегда может быть предложена модель, обеспечивающая лучшее совпадение с экспериментом. С другой стороны, всегда можно ввести в модель необходимые поправки и уточнения (например, за счет нормировки свободных коэффициентов), обеспечивающие совпадение экспериментальных и модельных данных.

Наконец, существуют (и имеют право на существование!) модели никогда не реализовавшихся явлений (межпланетные катастрофы, отравление мирового океана), вообще не допускающие прямой экспериментальной проверки.

Широкое применение математического моделирования приводит к определенному изменению методологии науки. Приближенный характер математических моделей изменяет требования к математическому описанию реального мира. Можно говорить о более и менее точных моделях, о моделях, отражающих те или иные свойства объекта, о развивающихся, эволюционных моделях. Возможно существование параллельных неконкурирующих между собой моделей одного и того же объекта. С другой стороны, когда истинная природа объекта неизвестна, для его описания могут быть

предложены *конкурирующие* модели, основанные на различных гипотезах, а сравнение результатов моделирования на качественном уровне позволяет отдать предпочтение одной из гипотез. Таким образом, переход к использованию моделей в известной мере приводит к ослаблению требований к математическому описанию, делает его более гибким, сближает с вербальным описанием объекта.

До сих пор часто приходится слышать споры о возможности математического описания и исследования таких сложных объектов, как двигательная установка, летательный аппарат или их основные узлы. Эти споры основаны главным образом на разнице методологических концепций. Противники математического моделирования исходят из невозможности исчерпывающего математического описания и получения точных решений для сложных систем, природа которых в настоящее время не может считаться до конца ясной. Сторонники подразумевают возможность и полезность построения приближенных моделей, описывающих хотя бы основные характеристики объекта.

Вопросы для самоконтроля

1. Дайте классификацию и сравнительную характеристику математических моделей.

1.4. ГЕОМЕТРИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

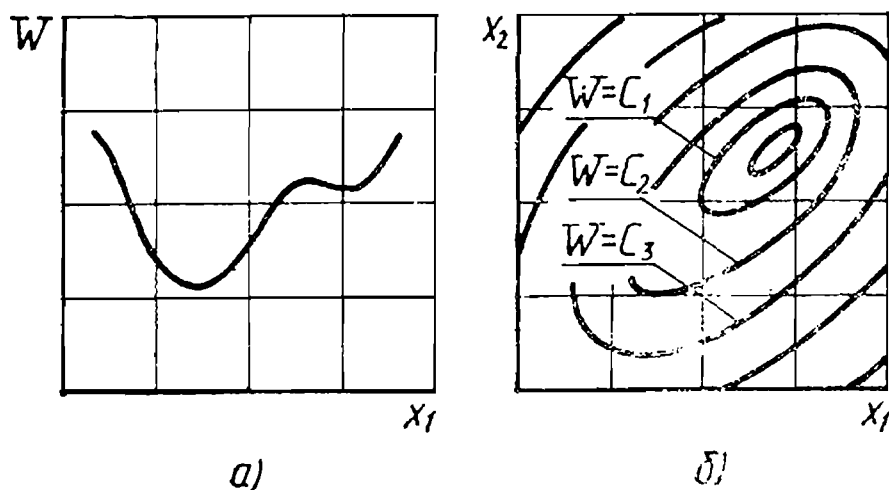
Геометрически модель может быть представлена как некоторая *поверхность отклика*, соответствующая расположению точек $W = W(x)$ в k -мерном факторном пространстве X . Такое представление оказывается чрезвычайно полезным, в частности, при решении задач оптимизации. К сожалению, наглядно можно представить себе только одномерную (рис. 1.5, а) и двумерную (рис. 1.5, б) поверхности отклика, причем в последнем случае удобно использовать *топографический* способ изображения рельефа поверхности с помощью линий уровня. Однако все представления, введенные для двумерных поверхностей отклика, действительны и в многомерном случае.

Область, в которой определена поверхность отклика, называется областью определения X^* . Эта область составляет, как правило, лишь часть полного факторного пространства X ($X^* \subset X$) и выделяется с помощью явных

$$\begin{aligned} x_i &= C_i; \quad i = 1, 2, \dots, m; \\ f_j(x) &= C_j; \quad j = 1, 2, \dots, l \end{aligned} \quad (1.3)$$

Рис. 1.5. Графическое изображение модели (поверхность отклика):

а — одномерная поверхность отклика; б — двумерная поверхность отклика



или неявных

$$x_{i\min} \leq x_i \leq x_{i\max}; i=1, 2, \dots, k; f_j(x) \leq C_j; j=1, 2, \dots, q \quad (1.4)$$

ограничений, наложенных на управляющие переменные. При этом функции $f_j(x)$ могут зависеть как одновременно от всех переменных, так и от некоторой их части. Ограничения типа равенств (1.3) или фиксируют одну из переменных на заданном уровне, или отражают наличие зависимости между переменными. При этом фактически понижается размерность факторного пространства.

Ограничения типа неравенств (1.4) характеризуют или физические ограничения на процессы в изучаемом объекте (например, необходимость поддержания заданного агрегатного состояния вещества приводит к ограничениям температуры), или технические ограничения, связанные с условиями работы объекта (например, могут быть заданы предельные длина разбега и взлетная масса самолета).

Возможности исследования моделей существенно зависят от свойств (рельефа) поверхности отклика, в частности, от количества имеющихся на ней «вершин» и ее контрастности. Количество вершин (впадин) определяет модальность поверхности отклика. Если в области определения на поверхности отклика имеется одна вершина (впадина), модель называется *унимодальной*. Характер изменения функции при этом может быть различным. На рис. 1.6 показаны три одномерные унимодальные модели. В общем случае (рис. 1.6, а) модель может иметь разрывы 1-го рода. *Непрерывная* унимодальная модель (рис. 1.6, б) может иметь точки разрыва производной. Наконец, *выпуклая* унимодальная модель (рис. 1.6, в) для произвольной секущей в области сечения целиком лежит над (под) секущей. Однако для всех трех случаев выполняется общее требование унимодальности, которое может быть сформулировано следующим образом: если x^* соответствует экстремальному значению поверхности отклика $W(x^*) = \text{extr } W$, то из условия $x_1 < x_2 < x^*$ ($x_1 > x_2 > x^*$) следует $W(x_1) < W(x_2) < W(x^*)$ (если экстремум — максимум) или $W(x_1) > W(x_2) > W(x^*)$ (если экстремум — минимум), т. е. по мере удаления от экстремальной

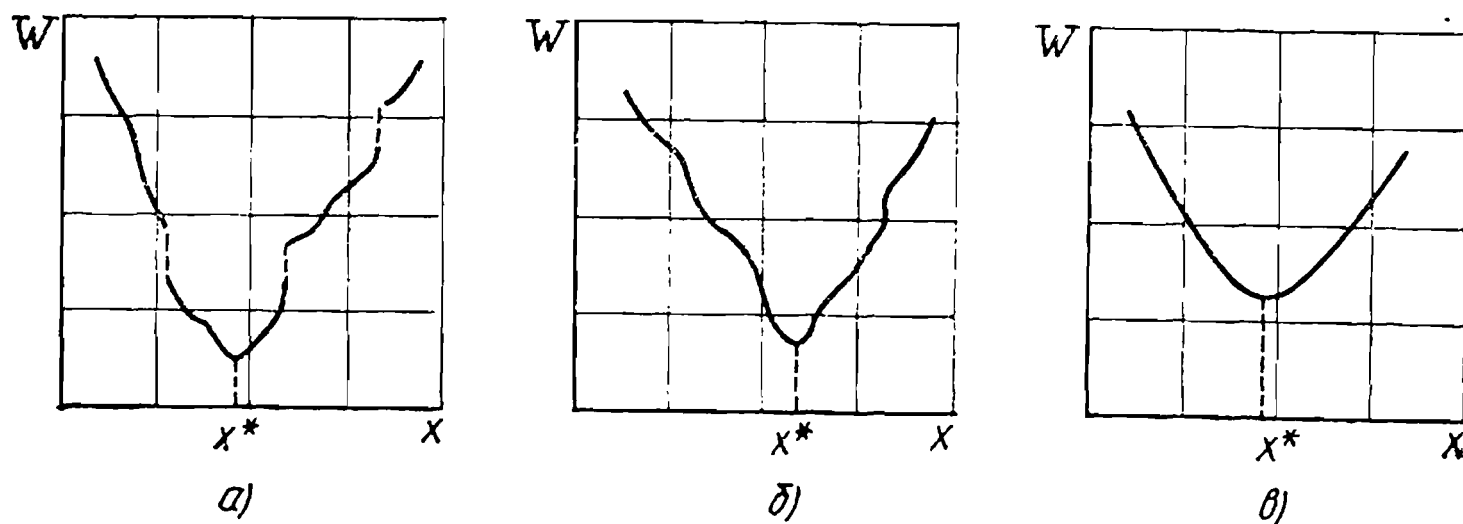


Рис. 1.6. Одномерные унимодальные поверхности отклика:

а — унимодальная поверхность отклика; б — непрерывная унимодальная поверхность отклика; в — выпуклая (вниз) поверхность отклика

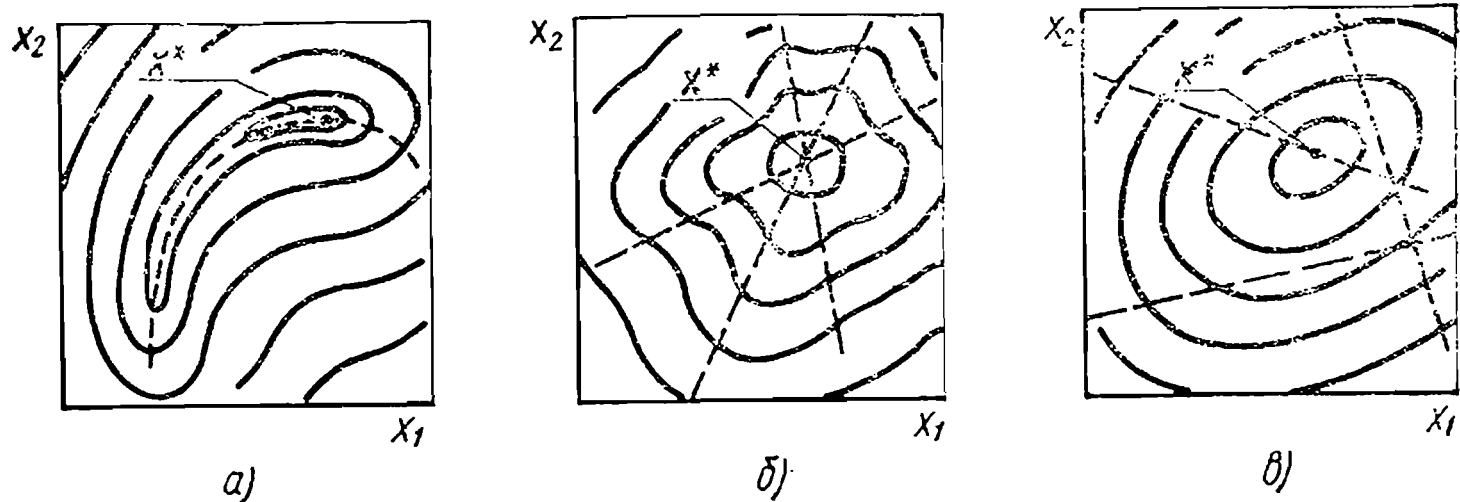


Рис. 1.7. Двумерные унимодальные поверхности отклика:

а — унимодальная поверхность отклика; *б* — строго унимодальная поверхность отклика; *в* — линейно унимодальная поверхность отклика

точки значение функции *непрерывно* падает (растет). Унимодальная функция не может включать в себя участки постоянства (т. е. при $x_1 < x_2 < x^*$ невозможно $W(x_1) = W(x_2)$). Свойство унимодальности для многомерных функций может быть продемонстрировано на примере двумерных функций (рис. 1.7). Двумерная (многомерная) модель называется *унимодальной*, если существует хотя бы одна кривая, проходящая через точку x^* , вдоль которой функция унимодальна в одномерном смысле (рис. 1.7, *а*). Модель называется *строго унимодальной* (рис. 1.7, *б*), если она унимодальна в одномерном смысле вдоль любой прямой, проведенной через точку x^* . Наконец, модель называется *линейно унимодальной* (рис. 1.7, *в*), если она унимодальна в одномерном смысле вдоль любой прямой в области определения X^* . Наряду с унимодальными могут существовать *полимодальные* модели (рис. 1.8).

Другим важным свойством поверхности отклика является ее контрастность, показывающая чувствительность результирующей функции к изменению факторов. Контрастность характеризуется величинами производных и может быть различной для разных факторов и разных участков области определения. Примерно равная контрастность по всем переменным x_i характеризует ситуацию «склона» на поверхности отклика (точка *а* на рис. 1.9). Сущест-

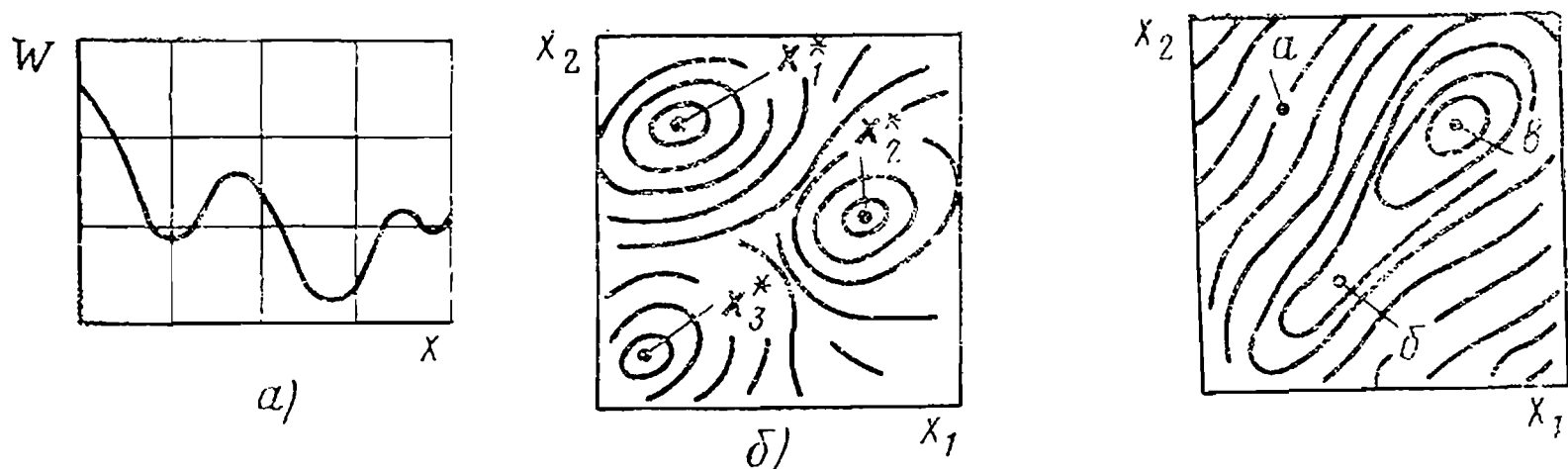


Рис. 1.8. Полимодальные поверхности отклика:

а — одномерная; *б* — двумерная

Рис. 1.9. Характеристики контрастности поверхности отклика:

а — ситуация «склона»; *б* — «овражная» ситуация; *в* — стационарная область

венно различная контрастность по различным переменным (плохая обусловленность функции) является признаком «овражной» ситуации (точка б). Наконец, общая низкая контрастность (точка в) говорит или о близости экстремума, или о своеобразном «плато» на поверхности отклика.

Вопросы для самоконтроля

1. Что называется поверхностью отклика? В чем заключается свойство уни-модальности поверхности отклика?

ГЛАВА 2. ИЕРАРХИЧЕСКИЕ ПРИНЦИПЫ ПОСТРОЕНИЯ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

2.1. ИЕРАРХИЧЕСКИЙ ПОДХОД К ОПИСАНИЮ ДЛА

Основная трудность, с которой приходится сталкиваться при разработке математических моделей сложных систем, в частности математических моделей ДУ и ДЛА, заключается в преодолении противоречия между требованием возможно большей простоты описания (без него трудно понять и решить задачу) и необходимостью учета многочисленных параметров, закономерностей, ограничений и взаимосвязей, определяющих работу системы. Сложную систему практически невозможно полно и детально описать в рамках единой математической модели. Поэтому при описании таких систем обычно используется *модульный* принцип, предусматривающий разработку не одной, а семейства взаимосвязанных и взаимодействующих между собой моделей. Использование единой модели сложной системы нецелесообразно и по методическим соображениям. Различные части системы (подсистемы) изучены в разной степени (например, с разной степенью достоверности могут быть описаны процессы горения, охлаждения стенок и расширения газа в сопле при изучении жидкостного ракетного двигателя), поэтому их математические описания обладают разной точностью, а непрерывное накопление физической информации приводит к систематическому совершенствованию описаний отдельных элементов. При использовании единой модели точность моделирования целиком определяется точностью описания наименее изученных подсистем, а любая новая информация требует полной переделки модели. Разумеется, переход к модульному описанию сам по себе не приведет к увеличению точности. Однако в этом случае появляется возможность более достоверного моделирования по крайней мере некоторых из подсистем. Изменение представлений об отдельных процессах или об отдельных элементах системы при модульном описании не требует полной переделки модели и затрагивает лишь отдельные модели — модули.

Таким образом, применение модульного принципа при описании сложных систем позволяет:

— разрешить проблему больших объемов информации за счет распределения этой информации между отдельными блоками модели;

— «развязать» изучение отдельных подсистем между собой, так как каждый из модулей модели может быть исследован в квази-автономном режиме.

Вместе с тем переход к системе моделей приводит к необходимости исследования связывающих их взаимодействий. Возникает задача координации подсистем и их моделей между собой.

Наиболее эффективные результаты при описании и изучении сложных систем дает *иерархический* подход, предусматривающий разбиение системы на вертикально соподчиненные подсистемы разных уровней, разработку модульных моделей каждой из подсистем, введение приоритетов (т. е. права преимущества) для подсистем старших уровней по отношению к подсистемам младших уровней, известную автономность каждой из подсистем. Поскольку каждая из подсистем в свою очередь может быть разбита на новые подсистемы, возникает *многоуровневая иерархическая система* моделей.

Для сложных систем различной физической природы разработаны различные концепции иерархического описания: концепции *страт, слоев и эшелонов* [12]. Все эти концепции предусматривают введение ряда уровней иерархии (рис. 2.1), отличающихся различной степенью абстрагирования от реальной картины явления. На каждом уровне описания вводится свой набор переменных, принципов и представлений, позволяющий в значительной степени ограничиться изучением только этого уровня.

Наблюдается известная асимметрия представлений. Понимание физических процессов и точность описания объекта растет с переходом к нижним уровням иерархии. Общие же цели и задачи функционирования системы, ее место в материальном мире более полно раскрываются на верхних уровнях.

Для описания физических объектов используется концепция страт, и принято говорить о *стратифицированной* иерархической системе. При иерархическом описании ДУ ЛА на верхних уровнях рассматриваются общие характеристики ДУ, такие как тяга, мощность, суммарный и удельный импульс тяги, массогабаритные характеристики, коэффициенты полезного действия, ресурс и

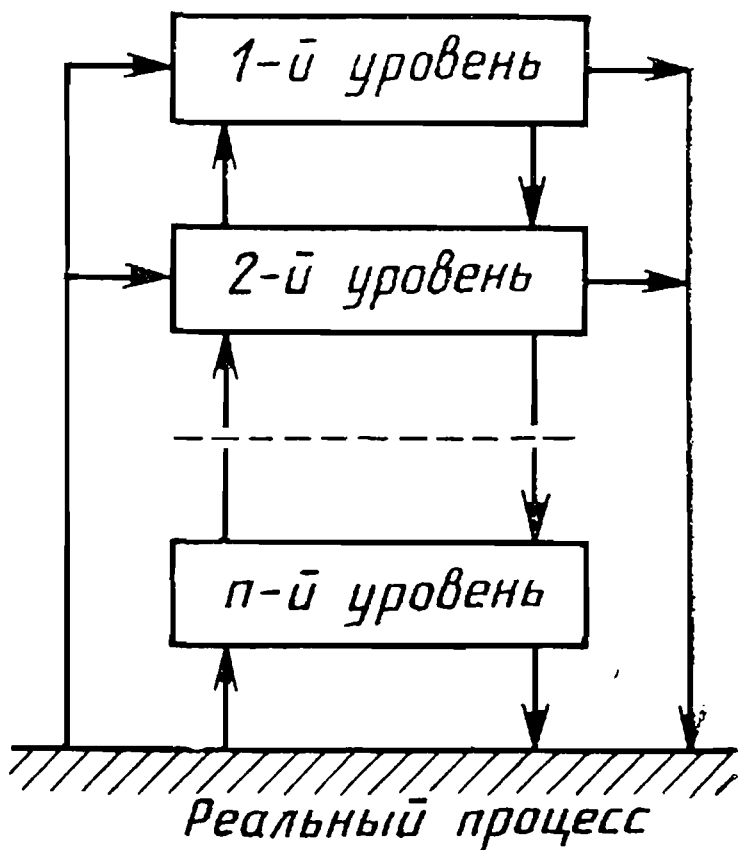


Рис. 2.1. Уровни иерархического описания сложной системы

надежность, количество параллельно работающих и резервных блоков (например, количество двигателей или баков топлива), а также взаимодействие ДУ с другими подсистемами ЛА. На нижних уровнях рассматриваются отдельные узлы и элементы ДЛА. При этом может стоять задача изучения, например, процессов воспламенения, прочности отдельных элементов или узлов, ускорения заряженных частиц, температурных, гидродинамических, магнитных и иных полей в элементах конструкции и т. д. Очевидно, изучение ДЛА на каждом из уровней требует специфических знаний и методов и зачастую ведется специалистами различного профиля.

Можно привести и другие примеры стратифицированных технических и вообще научно-естественных систем. Так, например, ЭВМ может быть рассмотрена на страте физических законов, управляющих работой ее элементов (электроника), на страте обработки информации (программирование), на страте логического анализа и семантики (теория алгоритмических языков). Исследование на каждой из страт осуществляется независимо (программист может и не знать законов электроники, а также принципов разработки алгоритмических языков). Однако изучение ЭВМ как целого требует совместного использования результатов, полученных на каждой из страт.

Человек как сложная система может быть изучен на страте биохимических процессов (биохимия), на страте физиологии (медицина), на страте общественных наук (социология), изучающих поведение человека в коллективе.

Концепция слоев относится к процессам принятия сложных решений, в частности, в условиях неопределенности. В этих случаях определяют иерархически соподчиненное семейство проблем, решение которых осуществляют последовательно, причем решение каждой из проблем фиксирует определенные условия для решения следующей проблемы. Примером послойной системы принятия решений может служить комплексная система выбора методов оптимизации.

Наконец, концепция эшелонов вводится для описания иерархических структур, в качестве элементов которых выступают люди или коллективы, в частности при исследовании организаций. Идеальной иллюстрацией многоуровневой эшелонированной системы могла бы служить модель функционирования большого предприятия, министерства.

Вопросы для самоконтроля

1. Почему используется иерархическая структура построения моделей?
2. Как изменяются объемы информации при переходе от уровня к уровню иерархической системы моделей?

2.2. ДЕКОМПОЗИЦИЯ МАТЕМАТИЧЕСКОГО ОПИСАНИЯ

Математически стратификация системы основана на возможности представления множеств входных переменных X , шумов Ω и выходов W в виде декартовых произведений:

$$\begin{aligned} X &= X_1 \times X_2 \times \dots \times X_n; \quad \Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n; \\ W &= W_1 \times W_2 \times \dots \times W_n, \end{aligned} \quad (2.1)$$

причем каждая тройка (X_i, Ω_i, W_i) $i=1, 2, \dots, n$ приписывается определенной страте (рис. 2.1). Для каждой страты задается самостоятельная модель, соответствующая некоторому отображению w_i :

$$\begin{aligned} w_i : X_i \times \Omega_i \times U_i &\Rightarrow W_i, \quad i=n; \\ w_i : X_i \times \Omega_i \times Z_i \times U_i &= W_i, \quad 1 < i < n; \\ w_i : X_i \times \Omega_i \times Z_i &\Rightarrow W_i, \quad i=1, \end{aligned} \quad (2.2)$$

где U_i и Z_i — множества связей i -й страты с соседними стратами:

$$u_i : W_i \Rightarrow U_{i+1}; \quad z_i : W_i \Rightarrow Z_{i-1}. \quad (2.3)$$

Соотношения (2.2) и (2.3) показывают связь моделей между собой и фактически определяют границы их независимости.

Требования, предъявляемые к работе системы на вышележащих стратах, выступают как условия или ограничения на нижележащих. Так, требования определенной тяги или ресурса ДЛА накладывают определенные ограничения на выбор параметров рабочего процесса и конструкцию двигателя.

При переходе на нижние страты (уровни) растет объем перерабатываемой информации и число переменных, что приводит к *декомпозиции* описания системы на нижней страте на описания подсистем одного уровня. Для каждой из подсистем может быть сформулирована самостоятельная модель:

$$w_{ik} : X_{ik} \times \Omega_{ik} \times H_{ik} \times U_{ik} \times Z_{ik} \Rightarrow W_{ik}, \quad (2.4)$$

где H_{ik} — множество внутриуровневых связей:

$$h_{ik} : W_{ij} \Rightarrow W_{ik}, \quad j=1, 2, \dots, n, \quad j \neq k. \quad (2.5)$$

Структура двигательной установки ЛА может включать в себя различное число страт и строиться по различным принципам. Наиболее очевидным и распространенным является *морфологический* принцип декомпозиции, при котором структура модели соответствует физической структуре ДУ. При этом отдельные подсистемы соответствуют отдельным узлам и элементам установки. Однако возможны и другие подходы, когда в основу декомпозиции положены, например, процессы энергообмена или передачи информации. Число страт зависит от целей исследования и объема имеющейся информации. Однако почти всегда можно выделить три основные группы страт (три *обобщенные* страты), соответствующие трем *горизонтам* описания ДУ.

Примеры возможной декомпозиции ДУ ЛА разных типов приведены на рис. 2.2—2.4. Первый горизонт описания включает в себя одну или две страты, на которых рассматривается ДУ в целом и ее основные подсистемы: двигатель, система подачи и хранения топлива, система энергоснабжения и автоматического управления установкой. Задача координации подсистем заключается в согласовании их характеристик с целью оптимизации параметров ДУ. Математические модели, используемые для исследо-

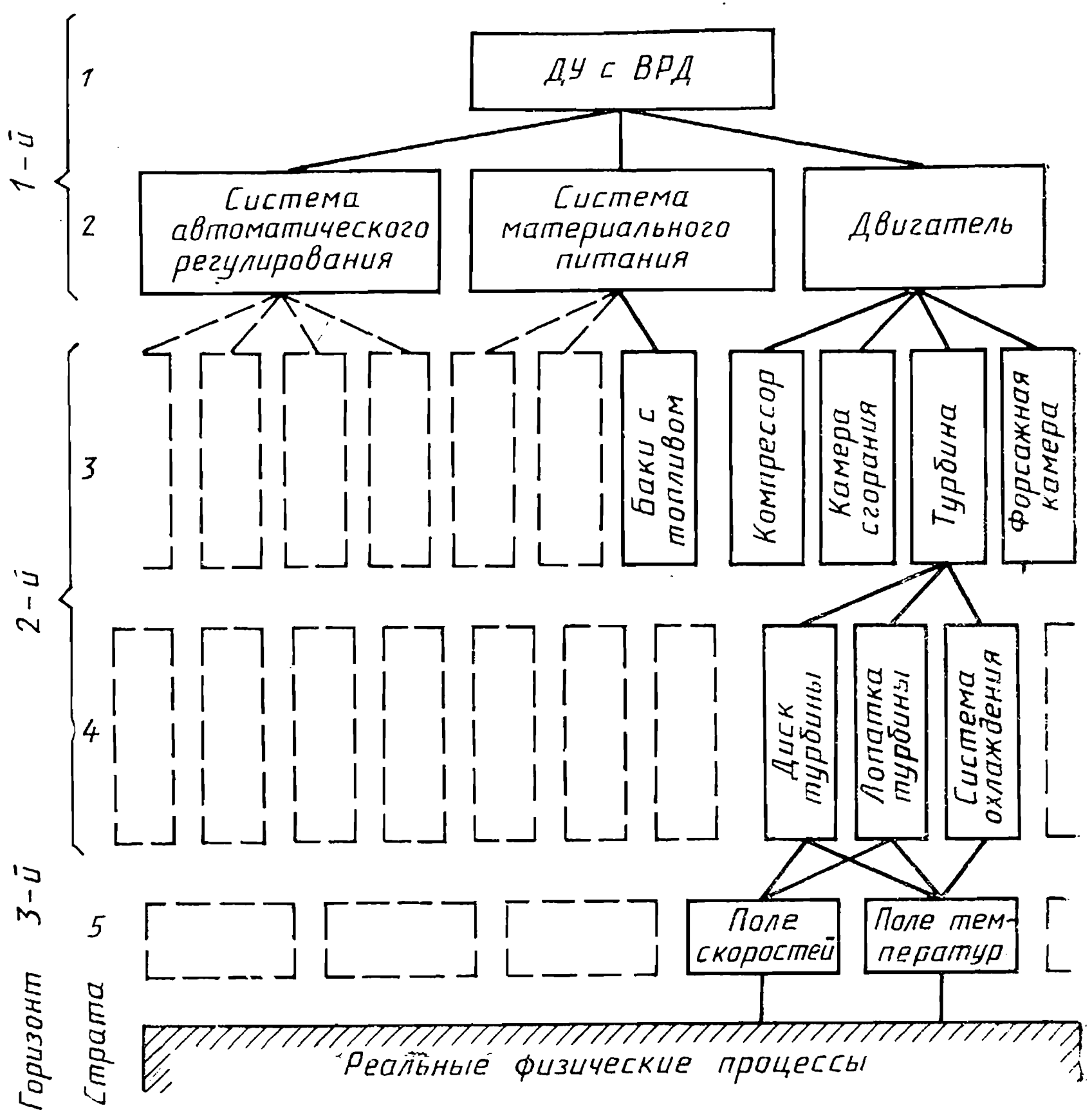


Рис. 2.2. Иерархическое описание ДУ с ВРД

вания задач первого горизонта, отличаются общностью и, как правило, не включают в себя описания отдельных процессов в ДЛА.

Ко второму горизонту относятся задачи исследования (модели) достаточно сложных комплексных узлов ДУ (например турбина ГТД, камера сгорания ЖРД, ускоритель ионного двигателя). Функционирование таких узлов определяется совокупностью взаимодействующих реальных физических процессов, а исследование соответствующих моделей позволяет уточнить рекомендации, полученные при изучении задач первого горизонта. Задачей координации предусматривается согласование характеристик процессов, определяющих работу узла.

Наконец, к третьему горизонту относятся подсистемы (модели), для которых может быть выделен один доминирующий физический процесс, определяющий работу подсистемы (гидродинамика течения газа в межлопаточном канале турбины, поля напряжений в диске или лопатке турбины, температурные, электростатические, магнитные поля и т. д.). Координация в данном

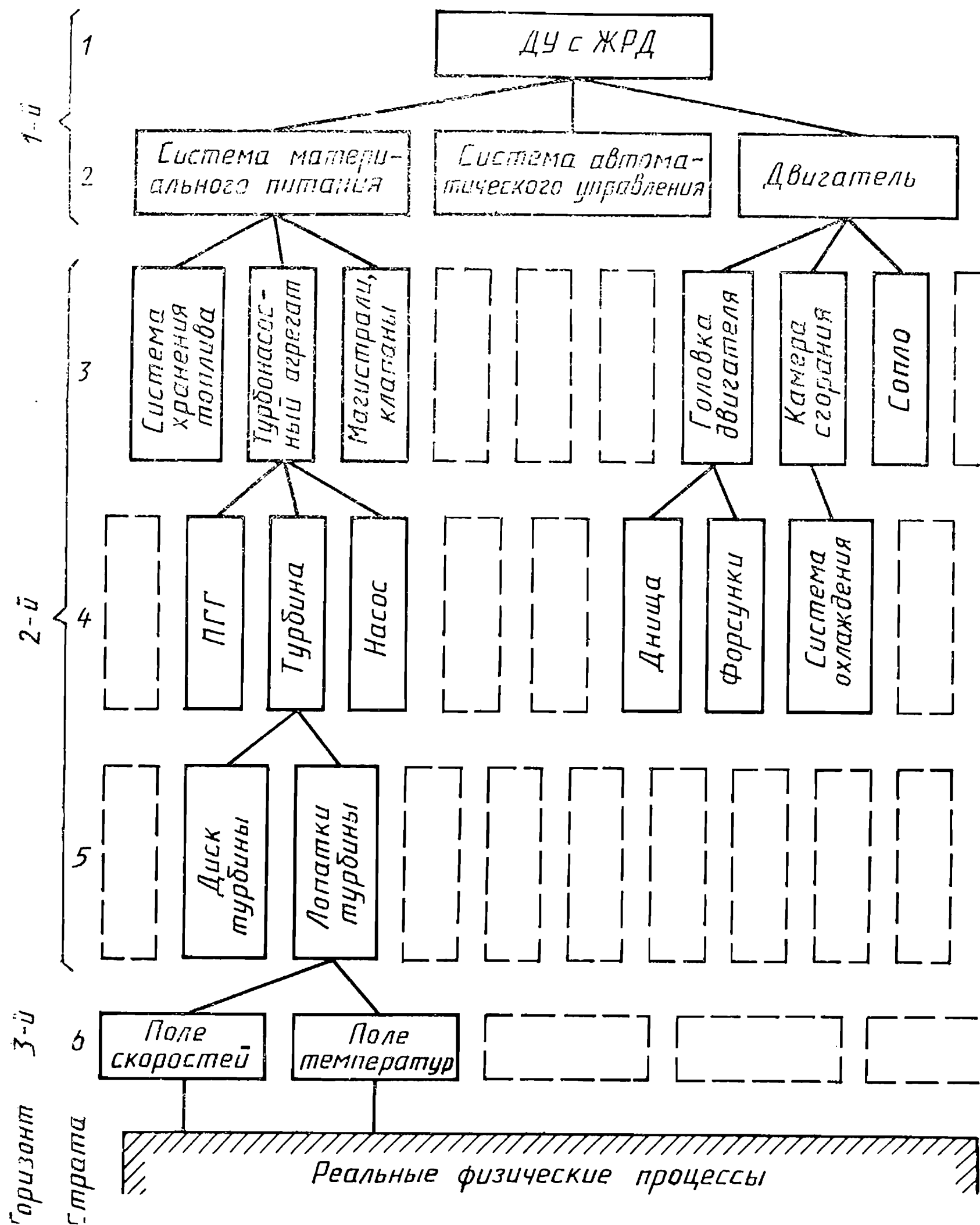


Рис. 2.3. Иерархическое описание ДУ с ЖРД

случае сводится к согласованию характеристик элементов, определяющих течение процесса, например, элементов магнитной системы, возбуждающей поле. Очевидно, модели этого уровня обладают наибольшей точностью и достоверностью. Третий горизонт включает в себя одну страту и, как правило, не допускает дальнейшей декомпозиции.

Важнейшее значение имеет характер связей подсистем в структуре декомпозиции. Наилучшей с точки зрения численного исследования является структура, при которой существенны только вертикальные (восходящие и нисходящие) межуровневые связи ($H_{ik} = \emptyset$). Такие структуры называются *сепарабельными* и обес-

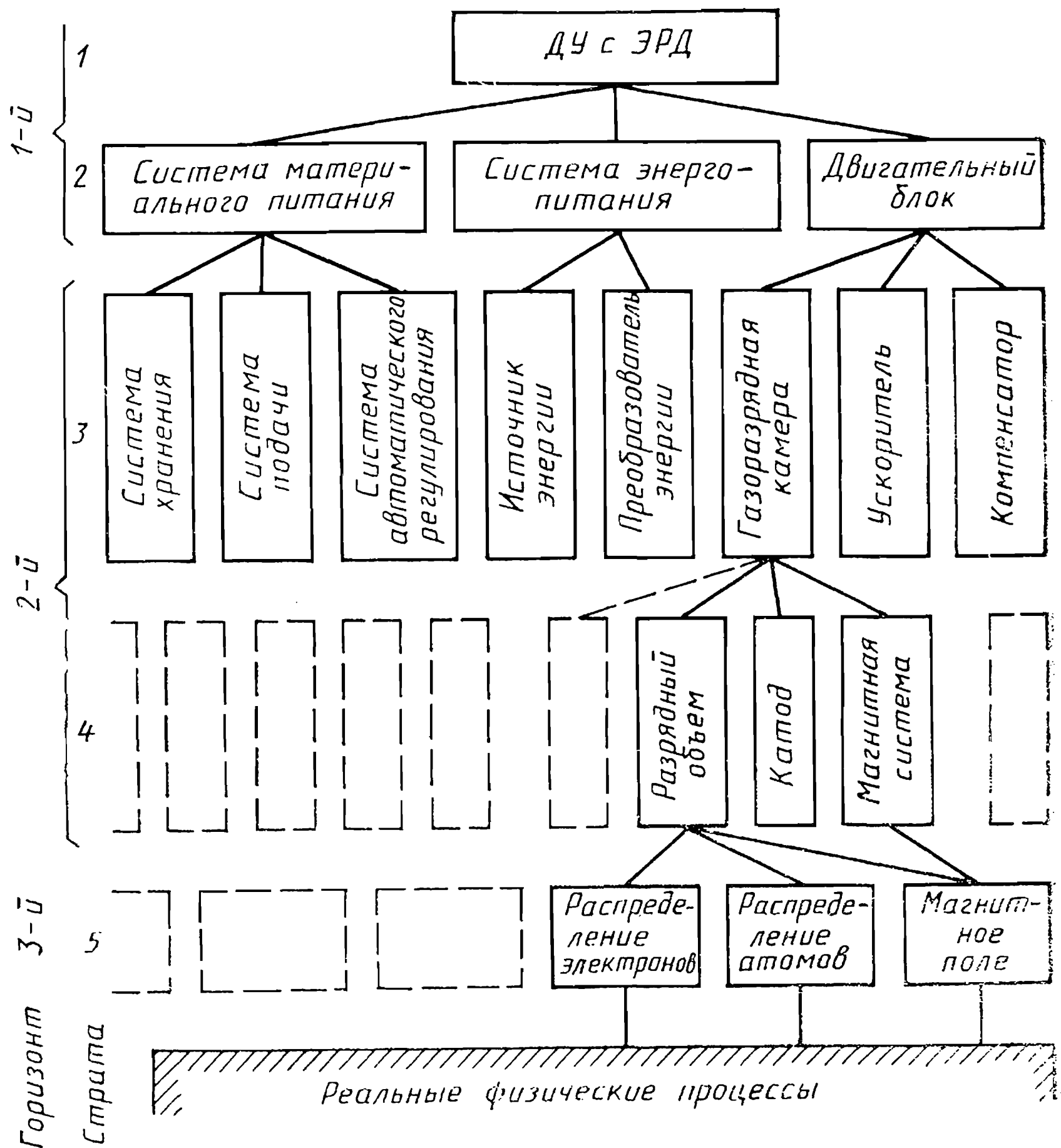


Рис. 2.4. Иерархическое описание ДУ с ЭРД

печивают возможность автономного исследования подсистем. К сожалению, описание ДУ ЛА, впрочем как и большинства других реальных систем, как правило, не может быть сведено к сепарабельным структурам ($H_{ik} \neq \emptyset$).

Так, например, режимы работы камеры сгорания воздушно-реактивного двигателя влияют на массогабаритные характеристики ДУ не только (и не столько!) через массу камеры сгорания, а главным образом — через массу сопряженных с камерой сгорания подсистем: от температуры газов зависят характеристики и масса турбины, надежность двигателя и, следовательно, его резервирование, эффективность использования топлива, а значит, масса топливных баков и т. д. (рис. 2.3). Точно также режимы работы газоразрядной камеры ионного двигателя оказывают влияние на массогабаритные характеристики электроракетной двигательной установки через характеристики систем энергетического и матери-

Таблица 2.1

Тип модели				Уровень иерархии				
				Первый горизонт		Второй горизонт		Третий горизонт
				1-я страта	2-я страта	3-я страта	4-я страта	5-я страта
Регуляр- ные	Аналитические							
	Числен- ные	Детерми- нирован- ные	Линейные					
			Квадратич- ные					
			Нелинейные					
			Статистические					
Аналоговые								
Идентификации		Натурные объекты						
		Физические модели						

ального питания двигателя. При этом эти побочные опосредствованные влияния часто оказываются значительно существеннее непосредственных восходяще-нисходящих связей. Наличие внутри-уровневых связей существенно усложняет исследование отдельных подсистем.

Не все рассмотренные выше модели в равной мере пригодны для описания подсистем различных уровней. На первом горизонте, как правило, используются численные детерминированные модели. Аналитические модели позволяют исследовать отдельные аспекты функционирования ДУ и ЛА.

На втором горизонте описания ДУ также используются численные детерминированные модели. Однако наряду с ними, особенно на нижних стратах этого горизонта, находят применение модели идентификации, опирающиеся на экспериментальные исследования отдельных узлов ДЛА, и статистические модели.

Наконец, на третьем горизонте при изучении отдельных элементов ДЛА и происходящих в них процессов используются все типы моделей, в том числе и аналитические. Области использования моделей различных типов на разных уровнях иерархии отражены в табл. 2.1 (сплошными линиями показаны области широкого применения моделей, пунктирными — области условного применения).

Вопросы для самоконтроля

- 1. В чем суть декомпозиции сложной системы?
- 2. Какие варианты связей в иерархической структуре моделей Вам известны?
- 3. В чем заключается свойство сепарабельности моделей?

2.3 ВЗАИМОСВЯЗИ МОДЕЛЕЙ. ЦИКЛ ПРОЕКТИРОВАНИЯ

Исследование ДЛА методами моделирования включает в себя этапы *анализа* и *синтеза*. Этап анализа тесно связан с декомпозицией системы и предусматривает выделение и исследование отдельных подсистем и происходящих в них процессов. Результаты такого исследования представляют самостоятельный интерес, однако еще не позволяют сделать какие-либо заключения о системе в целом. Поэтому за этапом анализа должен следовать этап синтеза, предусматривающий создание и изучение общего описания системы, исследование ее интегральных характеристик. По существу, процесс синтеза — это процесс конструирования общего описания системы из отдельных относительно простых частей (блоков). На этапе синтеза и исследования общих свойств системы возникает задача оптимального выбора ее характеристик, задача *оптимального проектирования*.

Модели разных уровней взаимосвязаны. Изучение моделей старших уровней позволяет найти определяющие параметры для нижележащих моделей, а те, в свою очередь, позволяют уточнить условия решения общих задач. Так, на старших уровнях определяются общие параметры ДЛА: количество двигателей, их тяга, надежность, ресурс, общие массогабаритные характеристики. Переход на средние уровни позволяет на основании этих данных найти степень сжатия, расход рабочего тела, температуру газа перед турбиной, характеристики смесеобразования, параметры разряда в газоразрядной камере и т. д. Наконец, на нижних уровнях осуществляется переход к исследованию отдельных процессов течения и теплообмена газа в межлопаточном канале или сопле двигателя, ускорения заряженных частиц, эрозии электродных поверхностей и т. д.

Результаты изучения элементов и подсистем позволяют уточнить характеристики системы в целом. Например, после изучения температурных полей и прочности турбины может возникнуть необходимость пересмотра общих параметров двигателя, последующего уточнения температурных полей и т. д.

Проектирование сложной системы осуществляется методом последовательных приближений. Схема процесса последовательных приближений показана на рис. 2.5. На старших уровнях вырабатывается управляющий сигнал *директива*, определяющий условия исследования на нижних уровнях. При этом возможен переход как непосредственно от уровня к уровню, так и через уровень. При изучении подсистем на основании полученных директив формируются

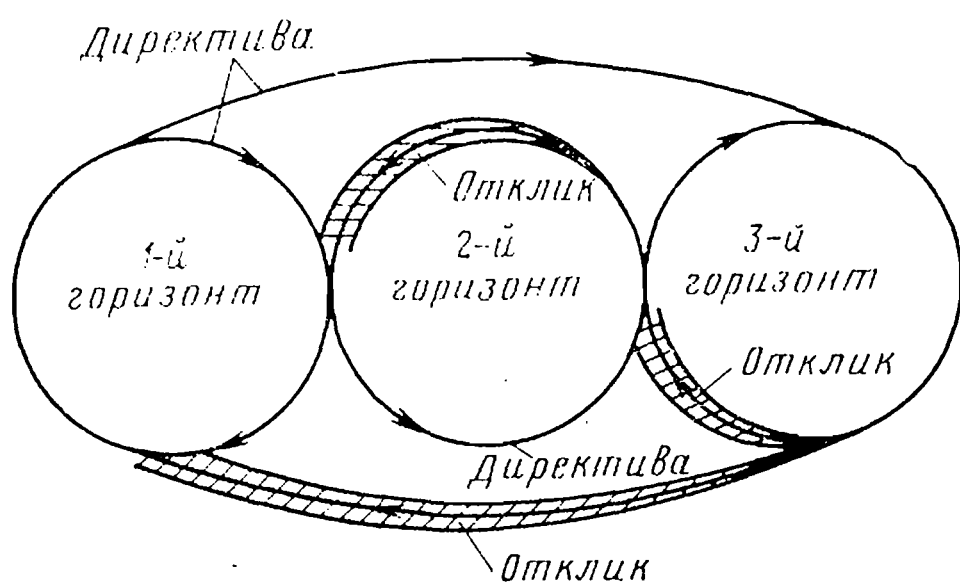
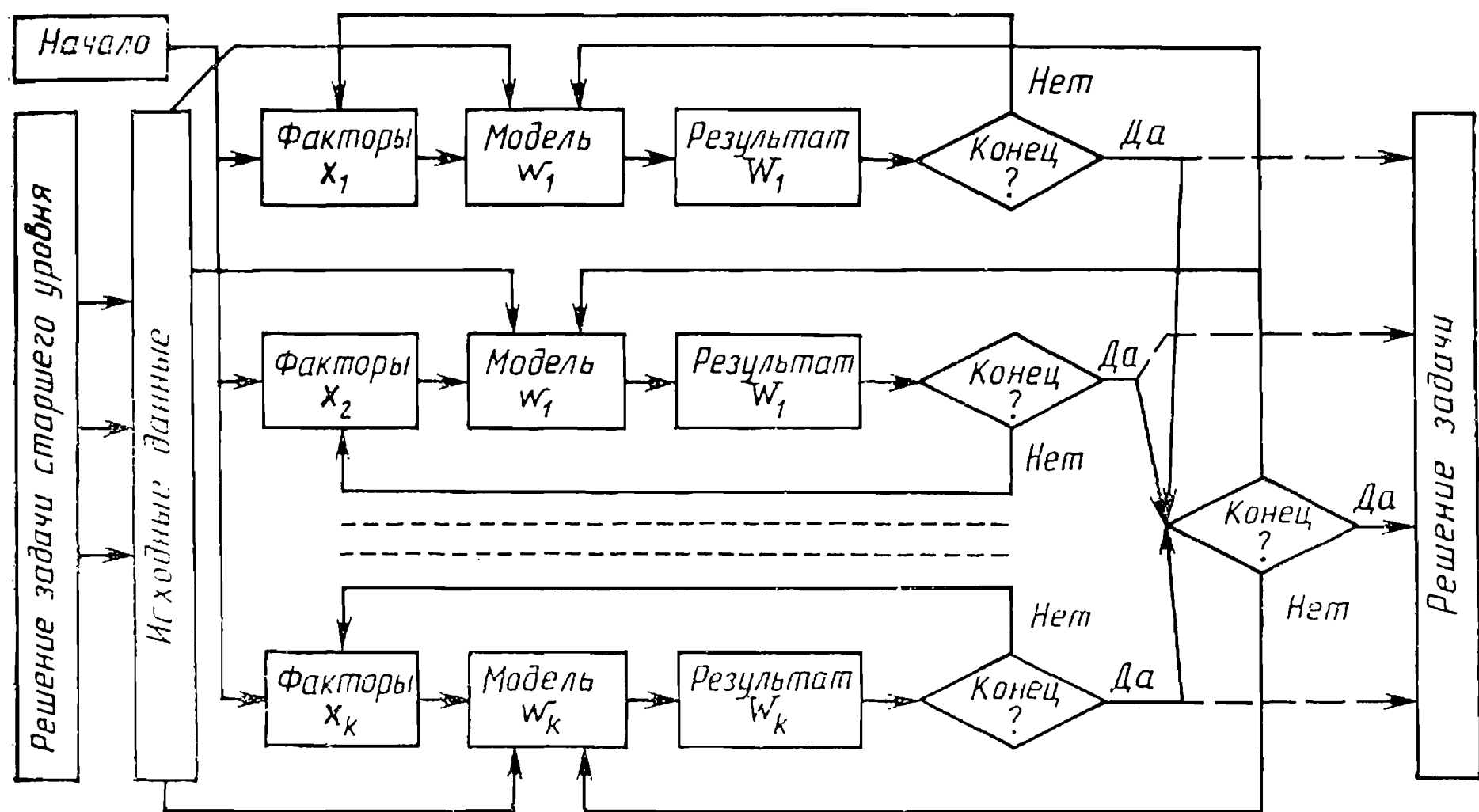


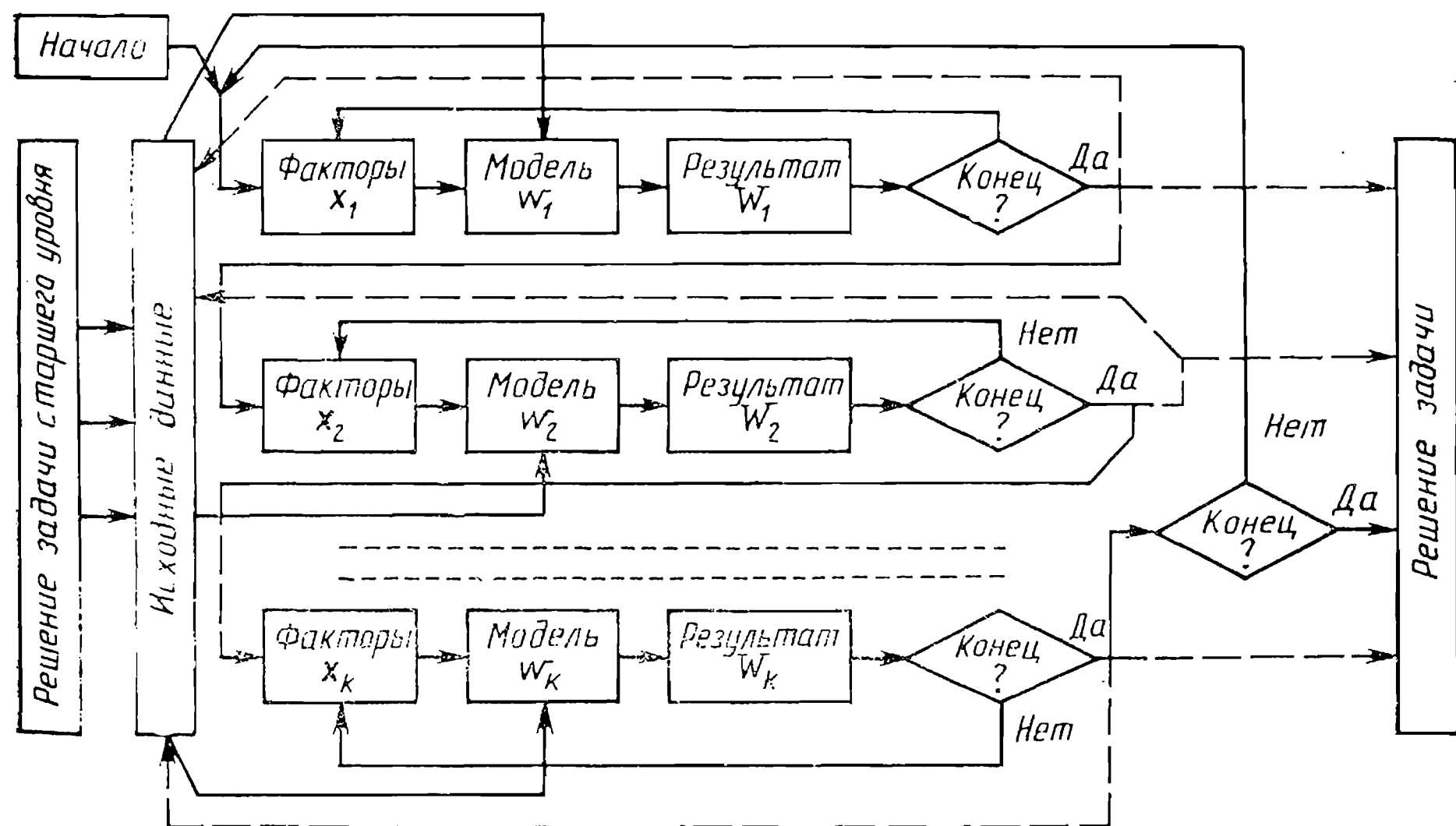
Рис. 2.5. Цикл последовательных приближений при проектировании ДУ

ответные сигналы — *отклики*, которые также могут передаваться как непосредственно на соседний, так и на другие старшие уровни. Процесс повторяется до тех пор, пока для контрольного звена результаты двух последовательных расчетов не совпадут с заданной степенью точности.

Для ускорения сходимости итераций целесообразно при формировании отклика передавать на старший уровень не одно



а)



б)

Рис. 2.6. Процедура последовательных приближений:
стратегия согласования (а) и прогнозирования и оценки (б) взаимодействий

решение, а некоторый комплекс близких решений (так называемую *трубку решений*), что обеспечивает старшему звену известную свободу выбора.

Поскольку рассмотренный процесс является итеративным и всегда начинается при наличии некоторой исходной информации, он, вообще говоря, может начинаться с любого уровня.

Конкретные процедуры последовательных приближений могут быть построены на основе стратегий *развязывания, прогнозирования и оценки взаимодействий* подсистем [12].

Стратегия развязывания взаимодействий предусматривает автономное решение задач оптимизации подсистем, причем функции связи подсистем выступают в роли дополнительных варьируемых параметров. В соответствии с принципом согласования взаимодействий

$$(\forall \gamma) (\forall x) \{ [p(x, D(\gamma)) \text{ и } k(\alpha_m(x)) = \alpha_n(x)] \Rightarrow p(\pi_m(x), D) \}, \quad (2.6)$$

где D, D — соответственно задачи исследования подсистемы и системы в целом; γ — координирующий сигнал; α_m, α_n — функции связи подсистем.

Локально оптимальное управление (x) является также глобально оптимальным $\pi_m(x)$, если связующие сигналы подсистем совпадают. Очевидно, несовпадение этих сигналов создает предпосылки для процедуры последовательных приближений, основанной на модификации функций оптимальности подсистем (рис. 2.6, а).

Стратегия прогнозирования взаимодействий предусматривает задание функции связи подсистем на приближенно оптимальном уровне, причем оптимизируемая подсистема получает некий образ сопряженных с ней подсистем. Из принципа прогнозирования взаимодействий

$$(\forall \gamma) (\forall x) \{ [p(x, D(\gamma)) \text{ и } \pi_m(x) = \pi_\alpha(x)] \Rightarrow p(x, D) \} \quad (2.7)$$

следует, что локально оптимальное управление (x) является также глобально оптимальным, когда фактические взаимодействия $\pi_m(x)$ совпадают с их прогнозными значениями $\pi_\alpha(x)$. При оптимизации предусматривается последовательная модификация образов сопряженных подсистем (рис. 2.6, б).

Наконец, стратегия оценки взаимодействий по существу представляет собой ослабленную стратегию прогнозирования взаимодействий и предусматривает задание вместо точных прогнозных значений функций связи некоторой области их возможного изменения. При этом в модель подсистемы включаются некоторые упрощенные модели-образы сопряженных подсистем.

В настоящее время исследование и оптимизация сложных систем происходит в интерактивном режиме, предусматривающем вмешательство человека на каждом из этапов выбора. Однако уже сделаны первые шаги на пути создания соответствующих автоматических программ.

Вопросы для самоконтроля

1. Какие задачи решаются на различных горизонтах описания ДЛА. В чем заключается свойство сепарабельности структуры декомпозиции сложной системы?
2. Опишите цикл последовательных приближений при проектировании ДЛА.

ГЛАВА 3. СТАТИСТИЧЕСКИЕ МЕТОДЫ ПРОВЕРКИ АДЕКВАТНОСТИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

3.1. ОСНОВНЫЕ ПОЛОЖЕНИЯ ТЕОРИИ ПРОВЕРКИ МОДЕЛЕЙ

Если имеются или могут быть получены необходимые и достоверные экспериментальные данные, для проверки адекватности моделей действительности можно использовать методы математической статистики [9, 22]. Однако возможности статистической проверки моделей нельзя переоценивать. Часто (например, при исследовании предельных и запредельных режимов работы ДЛА) экспериментальная проверка моделей невозможна. В этих случаях обоснование достоверности модели может опираться на ее качественный анализ, а в отдельных случаях — даже на ее эстетические достоинства.

Математически задача проверки адекватности модели действительности формулируется как задача проверки предположения о том, что значение отклика модели W_m отличается от реального отклика системы W не более чем на заданную величину ε^* :

$$| W(x) - W_m(x) | = \varepsilon \leq \varepsilon^*, \quad \forall x \in X^*. \quad (3.1)$$

Однако истинное значение отклика системы никогда неизвестно. Полученный в результате эксперимента отклик $\tilde{W}$ в силу неконтролируемого дрейфа системы, разброса характеристик ее элементов и, наконец, просто ошибок измерения представляет собой случайную величину, отличающуюся от W . Поэтому при сравнении результатов математического и физического экспериментов $(W_{mi}, \tilde{W}_i)$ будет получена совокупность случайных величин $\{\varepsilon_i\}$:

$$| \tilde{W}_i(x) - W_{mi}(x) | = \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (3.2)$$

среди которых могут оказаться как величины, удовлетворяющие ($\varepsilon_i \leq \varepsilon^*$), так и не удовлетворяющие ($\varepsilon_i > \varepsilon^*$) условию (3.1).

Вопрос сводится к тому, можно ли считать, что полученные значительные отклонения ($\varepsilon_i > \varepsilon^*$) объясняются случайными причинами или наличие этих отклонений должно быть признано существенным и в конечном счете привести к отказу от проверяемой модели. Для решения этого вопроса на основе выборки случайных величин $\{\varepsilon_i\}$ строят те или иные функции (статистические критерии) (см. разд. 3.2, 3.3) и сравнивают их с соответствующими характеристиками теоретических распределений случайных величин.

Как известно, для характеристики дискретных случайных величин $\{\xi\}$ вводят *математическое ожидание*, представляющее собой среднее значение случайной величины:

$$M(\xi) = \sum_{\xi} \xi p(\xi), \quad \sum_{\xi} p(\xi) = 1, \quad (3.3)$$

где $p(\xi)$ — вероятность реализации конкретной величины ξ .

Суммирование осуществляется по всем возможным значениям ξ . Для характеристики разброса случайной величины относительно ее среднего значения вводят *дисперсию*

$$\sigma^2(\xi) = \sum_{\xi} (\xi - M(\xi))^2 p(\xi). \quad (3.4)$$

Величина $\sigma(\xi)$ — среднее квадратическое (стандартное) отклонение величины ξ от $M(\xi)$.

Для характеристики *конечной* выборки случайных величин $\{\varepsilon_i\}$ вводятся *выборочные статистики*:

выборочное среднее

$$\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i \quad (3.3a)$$

и выборочная дисперсия

$$s^2(\varepsilon) = \frac{1}{n-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2, \quad (3.4a)$$

представляющие собой приближенные оценки математического ожидания $M(\varepsilon)$ и дисперсии $\sigma^2(\varepsilon)$. Оценки (3.3, a), (3.4, a) являются состоятельными и несмещенными, т. е. сходятся по вероятности к соответствующим величинам:

$$n \rightarrow \infty, \quad p(|\bar{\varepsilon} - M(\varepsilon)| < \delta) \Rightarrow 1; \quad p(|s^2(\varepsilon) - \sigma^2(\varepsilon)| < \delta) \Rightarrow 1, \quad (3.5)$$

где δ — любое наперед заданное положительное число.

Обычно предполагается, что совокупность отклонений $\{\varepsilon_i\}$ отвечает следующим допущениям.

1. Все отклонения ε_i независимы.
2. Все отклонения ε_i имеют нормальное распределение.
3. Математические ожидания всех отклонений ε_i равны нулю.
4. Дисперсии всех отклонений ε_i одинаковы.

Первое допущение требует независимого выбора опорных точек факторного пространства и независимости результатов последующих измерений от предыдущих. Это допущение обычно выполняется, так как результаты численных исследований модели очевидно не связаны между собой. Допущение о нормальном распределении ошибок может быть обосновано случайным характером воздействия многих факторов (центральная предельная теорема теории вероятностей). Допущение о равенстве нулю математических ожиданий естественно вытекает из предположения об адекватности модели и случайном характере ошибки. Допущение о равенстве всех дисперсий предполагает одинаковую точность модели во всей рассматриваемой области ($\forall x \in X^*$) и представляется наиболее спорным.

В ряде случаев для приведения распределения $\{\varepsilon_i\}$ к нормальному виду и выравнивания дисперсий можно использовать те или иные преобразования, например перейти от величин ε к их логарифмам $\ln \varepsilon$.

При соблюдении сформулированных допущений рассмотренные ниже статистические методы строго обоснованы. Если проверить выполнение этих допущений невозможно, те же методы можно использовать как эвристические.

Гипотеза об адекватности модели действительности (гипотеза H_0) может быть сформулирована как предположение о том, что полученная совокупность $\{\varepsilon_i\}$ не дает оснований отказаться от рассматриваемой модели. Альтернативная гипотеза H_1 состоит в том, что модель не отвечает заданным требованиям (3.1) и, следовательно, должна быть отвергнута. Так как выборка $\{\varepsilon_i\}$ случайна, решение о выборе одной из гипотез H_0 , H_1 носит вероятностный характер. При этом может быть допущена *ошибка первого рода*, состоящая в отказе от правильной модели (принимается H_1 , когда верна H_0), или *ошибка второго рода*, состоящая в принятии ошибочной модели (принимается H_0 , когда верна H_1). Вероятность ошибки первого рода обозначается α , второго β . Принято называть α риском разработчика (в системах обслуживания — риском продавца), β — риском потребителя (покупателя). Разумеется, желательно минимизировать как α , так и β . Однако при заданном объеме экспериментальной выборки и заданном виде статистического критерия уменьшение α влечет за собой увеличение β и наоборот.

На практике вероятность ошибки первого рода задается на определенном уровне (используются уровни $\alpha=0,05$; $0,01$; $0,005$; $0,001$; при этом в $100\alpha\%$ случаев правильная модель отвергается), а вероятность ошибки второго рода зависит от характера распределения, величины выборки и вида статистических критериев. Величина $1-\beta$, характеризующая *вероятность отказа от ошибочной модели*, называется *мощностью* критерия и является мерой его эффективности (рис. 3.1).

Выбор вероятностей ошибок α и β при проверке конкретной модели зависит от ответственности решений, принимаемых на основе моделирования. Так, например, если модель предназначена для управления двигателем летательного аппарата, необходимо в первую очередь минимизировать β , так как в данном случае принятие неверной модели, а значит, возможность ошибочных решений при управлении представляет неизмеримо больший вред, чем отказ от правильной модели.

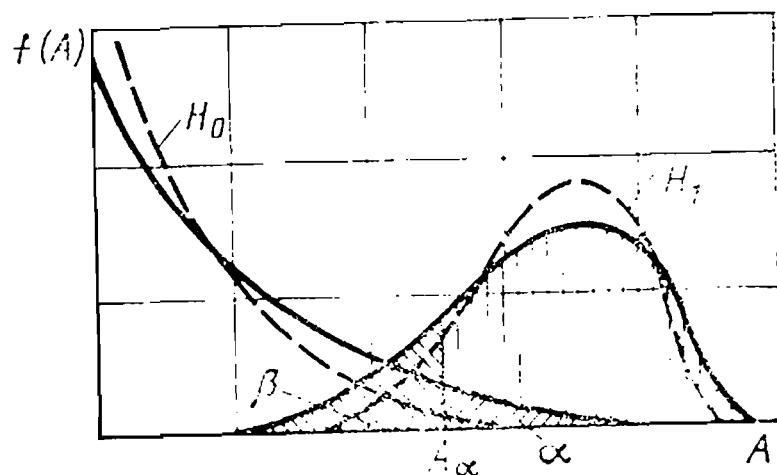


Рис. 3.1. Оценка гипотезы об адекватности модели действительности:

H_0 — распределение, соответствующее гипотезе H_0 для статистики A ; H_1 — распределение, соответствующее гипотезе H_1 для статистики A ; A_α — пороговое значение статистики

При заданном α мощность критерия может быть повышена за счет увеличения объема выборки. В силу неравенства Чебышева при произвольном законе распределения $\{\varepsilon_i\}$ и известной дисперсии $\sigma^2\{\varepsilon_i\}$ вероятность отклонения $\bar{\varepsilon}$ от $M(\varepsilon)$ задается соотношением

$$p \left(|\bar{\varepsilon} - M(\varepsilon)| \geq \frac{k\sigma(\varepsilon)}{\sqrt{n}} \right) \leq \frac{1}{k^2}, \quad (3.6)$$

где k — произвольная наперед заданная константа.

Для нормально распределенной совокупности $\{\varepsilon_i\}$ может быть получен более сильный результат:

$$M(\varepsilon) = \bar{\varepsilon} \pm p^* \frac{\sigma(\varepsilon)}{\sqrt{n}}, \quad (3.7)$$

где величина p^* характеризует заданную *доверительную вероятность* выполнения (3.7).

Если величина $\sigma(\varepsilon)$ заранее неизвестна, то для вычисления доверительного интервала используется t -распределение Стьюдента и соотношение (3.7) принимает вид

$$M(\varepsilon) = \bar{\varepsilon} \pm t_{\nu, n-1} \frac{s(\varepsilon)}{\sqrt{n}}, \quad (3.7a)$$

где $t_{\nu, n-1}$ — критерий Стьюдента с $(n-1)$ степенями свободы¹ при $100(1-\nu)$ % доверительной вероятности.

Характер распределений при $n > 3$ показан на рис. 3.1 пунктиром. Соотношения (3.6) — (3.7) показывают, что точность определения $M(\varepsilon)$ растет с ростом выборки пропорционально $\sqrt{n}$.

На основе (3.2) можно построить ряд критериев адекватности. В качестве меры совпадения результатов моделирования и экспериментов можно использовать, например, сумму отклонений

$$S_1 = \sum_{i=1}^n \varepsilon_i = \sum_{i=1}^n |W_i - W_{mi}|. \quad (3.8)$$

При этом лучшей будет модель, для которой величина S_1 минимальна ($S_{1\text{опт}} = S_{1\text{мин}}$). Недостатком меры является одинаковый вес малых и больших отклонений.

Можно воспользоваться величиной максимального отклонения

$$S_2 = \sup \{\varepsilon_i\} = \sup |\tilde{W}_i - W_{mi}|, \quad S_{2\text{опт}} = S_{2\text{мин}}, \quad (3.9)$$

приводящий к принципу минимакса.

Однако наибольшее распространение при оценке моделей получила величина суммы квадратов отклонений, называемая также *остаточной суммой квадратов*:

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (\tilde{W}_i - W_{im})^2, \quad S_{\text{опт}} = S_{\text{мин}}, \quad (3.10)$$

приводящая к принципу наименьших квадратов Гаусса — Лапласа. Достоинством меры (3.10) является преимущественный учет больших, т. е. наиболее существенных отклонений.

¹ Число степеней свободы f_s некоторой выборочной статистики равно числу независимых случайных величин минус число определенных по ним независимых параметров. Например, число степеней свободы $s^2\{\varepsilon_i\}$ (3.4a) для выборки из n величин ε_i равно $n-1$, так как по этой же выборке определено выборочное среднее $\bar{\varepsilon}$.

1. Как формулируются гипотезы об адекватности модели действительности?
2. Как оцениваются математическое ожидание и дисперсия случайной величины на основе экспериментальных данных?
3. Что такое риск α и риск β ? Как оценивается мощность критерия?
4. Что такое остаточная сумма квадратов?

3.2. КРИТЕРИИ СОГЛАСИЯ (χ^2 И СМИРНОВА — КОЛМОГОРОВА)

Оценка гипотезы об адекватности модели действительности (гипотезы H_0) может быть проверена путем сравнения экспериментального распределения отклонений $\{\varepsilon_i\}$ с тем или иным теоретическим распределением. Для проверки вводятся так называемые статистические критерии. Наибольшее распространение получили критерий согласия χ^2 , введенный Пирсоном, и критерий Смирнова — Колмогорова.

Использование критерия χ^2 основано на методе наименьших квадратов и предусматривает ответ на вопрос, можно ли считать полученную величину остаточной суммы квадратов S (3.10) случайной величиной.

Из статистики известно, что теоретическое распределение χ^2 представляет собой распределение случайной величины U , являющейся суммой квадратов нормально распределенных случайных величин ξ_N с нулевым математическим ожиданием и единичной дисперсией:

$$U = \sum_{i=1}^n \xi_{Ni}^2; \quad M(\xi_N) = 0; \quad \sigma^2(\xi_N) = 1. \quad (3.11)$$

Величина U имеет n степеней свободы ($f_s = n$). Поэтому U распределена по закону χ_n^2 , где индекс характеризует число степеней свободы. Имеются статистические таблицы [16], в которых для различных уровней значимости ($\alpha = 0,05; 0,01; 0,05; 0,001$) приведены пороговые константы $\chi_{n,\alpha}^2$, вероятность превзойти которые в случайном эксперименте соответствует величине α . Другими словами, событие $U > \chi_{n,\alpha}^2$ будет наблюдаться лишь для $100\alpha\%$ случайных выборок.

Пусть проверке подлежит гипотеза о том, что рассматриваемая модель адекватна исследуемой системе с вероятностью p (например $p = 0,95$). Это значит, что при n независимых испытаний pn значений ε_i должно удовлетворять условию (3.1) ($\varepsilon_i < \varepsilon^*$) и лишь в $(1-p)n$ случаях это условие может быть нарушено. В результате случайного эксперимента для этих событий будут получены частоты v_1 и v_2 :

$$v_1 \cong pn; \quad v_2 \cong (1-p)n; \quad v_1 + v_2 = n. \quad (3.12)$$

Частоты v_1, v_2 отличаются от точных вероятностных оценок или из-за несоответствия модели действительности (заданная вероятность p не соблюдается), или из-за случайных отклонений.

Для оценки предположения о том, что отклонения v_1, v_2 от соответствующих вероятностей случайны, строится функция

$$U^* = \frac{(v_1 - pn)^2}{pn} + \frac{(v_2 - (1-p)n)^2}{(1-p)n}, \quad (3.13)$$

представляющая собой сумму квадратов отклонений, нормированных на соответствующие вероятности. Можно показать, что при достаточно больших n (а также v_1 и v_2) величины $\frac{v_1 - pn}{\sqrt{pn}}$, $\frac{v_2 - (1-p)n}{\sqrt{(1-p)n}}$ распределены по нормальному закону с нулевым математическим ожиданием и единичной дисперсией и, следовательно, величина U^* имеет распределение χ^2 с одной степенью свободы χ_1^2 . Требования к математическому ожиданию и дисперсии (3.11) выполняются в силу предположения об адекватности модели и нормировки, а число степеней свободы равно 1, так как соотношение (3.13) содержит две случайных величины, связанные одним дополнительным условием $v_1 + v_2 = n$. Полученное значение U^* можно сравнить с табличным значением при заданном уровне значимости $\chi_{1,\alpha}^2$. Если U^* превышает пороговое значение ($U^* > \chi_{1,\alpha}^2$) модель должна быть отвергнута¹, и принимается гипотеза H_1 . Если $U^* \leq \chi_{1,\alpha}^2$, экспериментальные данные не противоречат гипотезе об адекватности модели, и принимается гипотеза H_0 . Очевидно, один такой результат еще не доказывает адекватности модели. Более сильный результат, указывающий на адекватность модели, может быть получен при многократных удачных испытаниях модели с помощью критерия χ^2 . При этом, конечно, резко возрастает трудоемкость проверки.

В общем случае проверки статистических гипотез для групп данных (например, область определения модели X может быть разбита на k подобластей и в каждой из них может быть поставлена серия экспериментов) выражение (3.13) принимает вид

$$U^* = \sum_{i=1}^k \frac{(v_i - np_i)^2}{np_i}; \quad \sum_{i=1}^k v_i = n, \quad (3.13a)$$

и имеет $f_S = k - 1$ степеней свободы. В этом случае U^* нужно сравнивать с пороговым значением $\chi_{f_S,\alpha}^2$.

Необходимым условием использования критерия χ^2 является *многочисленность* экспериментальных данных для каждой из групп данных ($v_1, v_2 \geq 10$), так как только в этом случае справедливо предположение о распределении U^* по закону χ^2 .

Если модель полностью задана (нет необходимости определять те или иные параметры модели по экспериментальным данным), то для проверки адекватности модели может быть использован дру-

¹ Модель можно «спасти» или увеличив допустимую ошибку ε^* , или уменьшив уровень значимости α ($\alpha = 0,05 \rightarrow \alpha = 0,01$). В последнем случае, однако, возрастает вероятность ошибки второго рода β и снижается мощность критерия.

гой критерий согласия — критерий Смирнова — Колмогорова, опирающийся на максимальное значение отклонения S_2 (3.9). При этом так же, как и при использовании критерия χ^2 , для заданной экспериментальной выборки строится вспомогательная функция

$$\lambda_n^* = S_2 \sqrt{n}, \quad (3.14)$$

которая и сравнивается с пороговым значением $\lambda_{n\alpha}$, определенным по таблицам распределения функции Смирнова — Колмогорова [16]. При $\lambda_n^* > \lambda_{n\alpha}$ модель должна быть отвергнута, а при $\lambda_n^* \leq \lambda_{n\alpha}$ экспериментальные данные не противоречат гипотезе об адекватности модели.

Критерий Смирнова — Колмогорова целесообразно использовать при относительно малых выборках, когда критерий χ^2 оказывается неэффективным. Могут быть построены и другие статистические критерии согласия.

Вопросы для самоконтроля

1. Чем отличаются критерии χ^2 и Смирнова — Колмогорова?
2. Запишите соотношения для этих критериев.

3.3. АНАЛИЗ ДИСПЕРСИЙ. КРИТЕРИЙ ФИШЕРА

Адекватность модели исследуемой системе можно проверить путем анализа дисперсий. При этом проверяется предположение о несмещенности $M(\varepsilon) = \bar{\varepsilon} = 0$ и приемлемом разбросе $s^2(\varepsilon) \cong \sigma^2(W)$ результатов моделирования. Выборочная оценка дисперсии $s^2(\varepsilon)$ дается выражением (3.4а). В общем случае, когда по экспериментальным данным кроме ε определяется m параметров модели, число степеней свободы остаточной суммы квадратов (3.10) составляет $f_s = n - m - 1$. Для корректной модели математическое ожидание ошибки $\varepsilon = 0$ и (3.4а) принимает вид

$$s^2(\varepsilon) = \frac{\sum_{i=1}^n (W_i - W_{iM})^2}{n - m - 1} = \frac{\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2}{n - m - 1}. \quad (3.15)$$

Если дисперсия, характеризующая ошибку эксперимента $\sigma^2(W)$, известна, то величина $s^2(\varepsilon)$ может быть использована для оценки адекватности модели действительности. Для этого нужно составить F -отношение:

$$F_{f_s, \infty} = \frac{s^2(\varepsilon)}{\sigma^2(W)}. \quad (3.16)$$

F -отношение характеризуется двумя последовательно записанными значениями степеней свободы числителя и знаменателя. При этом, так как точность статистических оценок возрастает с ростом числа степеней свободы, число степеней свободы *точной* величины $\sigma^2(W)$ принимается равным ∞ .

Полученную величину F -отношения ($F_{f_S, \infty}$) нужно сравнить с критическим (пороговым) значением критерия Фишера $F_{f_S, \infty, \alpha}$ при соответствующих числах степеней свободы и заданном уровне значимости α . При $F_{f_S, \infty} \leq F_{f_S, \infty, \alpha}$ полученная величина s^2 может быть объяснена случайным разбросом экспериментальных данных и, следовательно, нет оснований для отказа от проверяемой модели. Если $F_{f_S, \infty} \geq F_{f_S, \infty, \alpha}$, полученное расхождение результатов моделирования и экспериментальных данных *значимо* и, следовательно, модель должна быть отвергнута как недостаточно точная.

Критерий Фишера может быть использован для сравнения точности двух (или нескольких) конкурирующих моделей. Пусть рассматриваются две модели изучаемой системы (w_1, w_2), приводящие к двум различным множествам значений функций отклика:

$$w_1 : X \times \Omega \Rightarrow W_{m1} \quad w_2 : X \times \Omega \Rightarrow W_{m2}. \quad (3.17)$$

Для определенности можно считать, что модель w_2 более подробна и предположительно более точна, чем w_1 . Именно это предположение и должно быть проверено на основе экспериментальных данных, образующих выборку n значений функции отклика системы $\{\tilde{W}_i\}$ для n векторов управляющих переменных $x \in X^*$. Как и ранее, каждому из этих значений можно поставить в соответствие значения функций отклика моделей W_{mi1}, W_{mi2} . При этом векторы управляющих переменных моделей могут включать в себя лишь *часть* компонент реального вектора управляющих переменных ($x_1 \neq x_2 \neq x$), что отражает очевидное право исследователя пренебречь при описании системы некоторыми из действующих факторов. Для каждой из моделей может быть составлена остаточная сумма квадратов

$$S_1 = \sum_{i=1}^n (\tilde{W}_i - W_{mi1})^2; \quad f_{S1} = n - m_1 - 1; \quad (3.18)$$

$$S_2 = \sum_{i=1}^n (\tilde{W}_i - W_{mi2})^2; \quad f_{S2} = n - m_2 - 1,$$

и подсчитаны средние квадраты этих сумм:

$$s_1^2 = \frac{S_1}{f_{S1}} = \frac{\sum_{i=1}^n (\tilde{W}_i - W_{mi1})^2}{n - m_1 - 1}; \quad (3.19)$$

$$s_2^2 = \frac{S_2}{f_{S2}} = \frac{\sum_{i=1}^n (\tilde{W}_i - W_{mi2})^2}{n - m_2 - 1}.$$

Для сравнения моделей подсчитывается так называемая *дополнительная сумма квадратов SS*, связанная с дополнительными

данными, введенными в модель ω_2 , и характеризующая внесенные в нее уточнения, а также число степеней свободы этой дополнительной суммы квадратов:

$$SS = S_1 - S_2; f_{SS} = f_{S1} - f_{S2} = m_2 - m_1. \quad (3.20)$$

Средний квадрат дополнительной суммы определяется соотношением

$$s_{SS}^2 = \frac{SS}{f_{SS}} = \frac{\sum_{i=1}^n (\tilde{W}_i - W_{Mi1})^2 - \sum_{i=1}^n (\tilde{W}_i - W_{Mi2})^2}{m_2 - m_1}. \quad (3.21)$$

Роль дополнительной информации, содержащейся в модели ω_2 , оценивается путем сравнения s_{SS}^2 с известной дисперсией $\sigma^2(W)$ с помощью критерия Фишера:

$$F_{m_2 - m_1, \infty} = \frac{s_{SS}^2}{\sigma^2(W)}. \quad (3.22)$$

Если дисперсия экспериментальных данных $\sigma^2(W)$ неизвестна, сравнение проводится с оценкой дисперсии для упрощенной модели s_1^2 :

$$F_{m_2 - m_1, n - m_1 - 1} = \frac{s_{SS}^2}{s_1^2}. \quad (3.22a)$$

Если полученное значение критерия Фишера значимо

$$(F_{m_2 - m_1, \infty, \alpha} > F_{m_2 - m_1, \infty, \alpha}; F_{m_2 - m_1, n - m_1 - 1} > F_{m_2 - m_1, n - m_1 - 1, \alpha}),$$

дополнительная информация, заложенная в модели ω_2 , существенна, и модель ω_2 действительно отличается от модели ω_1 . В противном случае уточнения, вносимые моделью ω_2 , неразличимы на фоне шума; с точки зрения точности модели равноценны и предпочтение должно быть отдано более простой модели ω_1 .

В частном случае полиномиальных моделей, представляющих собой конечные отрезки бесконечных рядов, например, при разработке моделей идентификации (см. часть 4), этим методом можно проверить целесообразность включения (выключения) в модель отдельных членов.

Вопросы для самоконтроля

1. Сформулируйте критерий Фишера.
2. Как используется критерий Фишера для проверки адекватности модели действительности?
3. Как проверить, какая из двух конкурирующих моделей лучше?

3.4. ОЦЕНКА РОЛИ ПЕРЕМЕННЫХ. КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Для оценки влияния отдельных переменных или элементов (блоков, подсистем) на общие характеристики системы используются методы корреляционного анализа. При этом ставит-

ся задача выявления связи между изменением некоторого параметра x_i (или некоторой промежуточной функции — отклика подсистемы $F(x)$) и значением общего отклика системы $W(x)$. Например, может быть поставлена задача определения связи между степенью сжатия воздуха в компрессоре и общими массогабаритными характеристиками ДЛА. При анализе могут быть использованы как экспериментальные данные (например, результаты стендовых испытаний двигателя), так и результаты численного моделирования. При этом и исследуемый параметр (частная функция отклика подсистемы), и общий отклик системы представляют собой, вообще говоря, некоторые случайные величины.

Степень связи двух случайных величин (F, W) характеризуется *ковариацией* этих величин $\text{cov}\{F, W\}$, представляющей собой математическое ожидание произведения отклонений этих величин от их средних значений:

$$\text{cov}\{F, W\} = M\{(W - M(W))(F - M(F))\}. \quad (3.23)$$

Ковариация может быть непосредственно использована для оценки связи F, W . Однако обычно удобнее использовать нормированную величину, называемую *коэффициентом корреляции* (связи) $\rho_{F, W}$, в виде

$$\rho_{F, W} = \frac{\text{cov}\{F, W\}}{\sqrt{\sigma^2(F) \sigma^2(W)}} = M\left[\frac{F - M(F)}{\sigma(F)} \frac{W - M(W)}{\sigma(W)}\right]. \quad (3.24)$$

Коэффициент корреляции может изменяться в пределах

$$-1 \leq \rho_{F, W} \leq 1. \quad (3.25)$$

Для оценки коэффициента корреляции по экспериментальным или численным выборочным данным используется *выборочный коэффициент корреляции* в виде

$$r_{FW} = \frac{\sum_{i=1}^n (W_i - \bar{W})(F_i - \bar{F})}{\left[\sum_{i=1}^n (W_i - \bar{W})^2 \sum_{i=1}^n (F_i - \bar{F})^2 \right]^{1/2}}, \quad (3.26)$$

где $\bar{W}$ и $\bar{F}$ — средние значения W и F для выборки:

$$\bar{W} = \frac{1}{n} \sum_{i=1}^n W_i; \quad \bar{F} = \frac{1}{n} \sum_{i=1}^n F_i. \quad (3.27)$$

Значение $\rho_{F, W} = +1$ соответствует максимальной положительной корреляции (рис. 3.2, а), а $\rho_{F, W} = -1$ — максимальной отрицательной корреляции (рис. 3.2, б). При этом точки идеально ложатся на некоторую прямую. Значение $\rho_{F, W} = 0$ (рис. 3.2, в) соответствует отсутствию корреляции (связи) между F и W . Все эти предельные случаи встречаются очень редко. Обычно коэффициент корреляции имеет дробное значение и, вообще говоря, должен быть еще проверен на статистическую значимость.

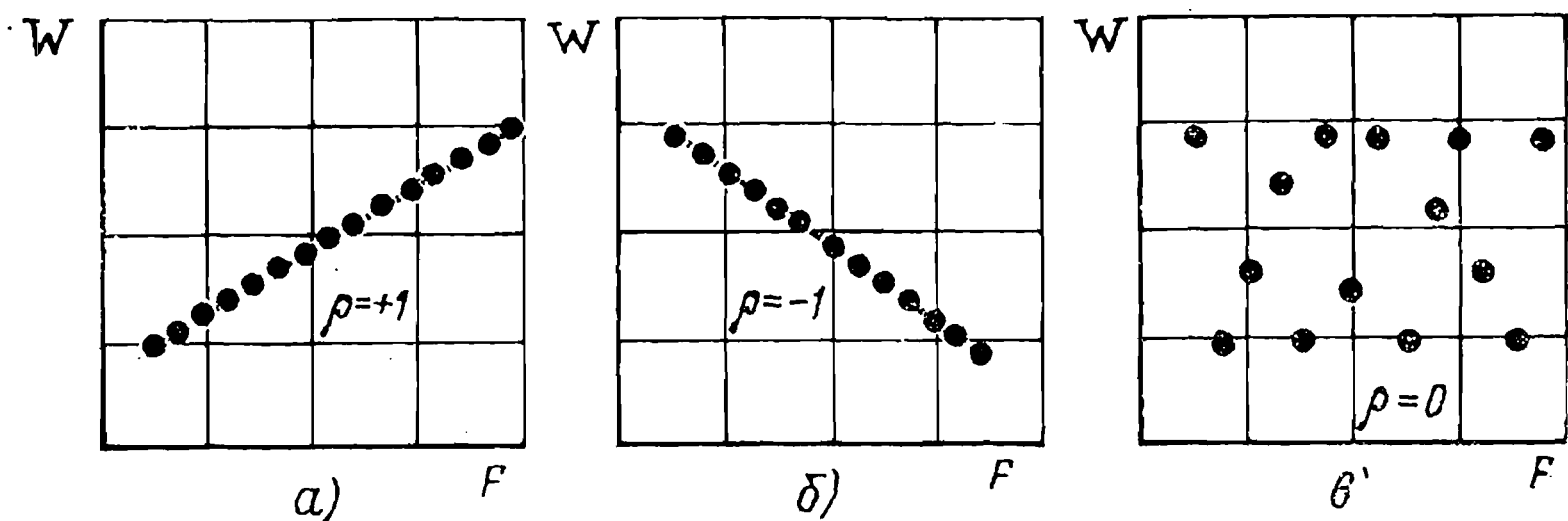


Рис. 3.2. Характер изменения коэффициента корреляции

Корреляционный анализ позволяет определить вклад той или иной переменной (подсистемы) в общую функцию отклика системы. По его результатам можно:

— провести ранжирование переменных (подсистем) по их влиянию на систему и может быть отказаться от учета или иных переменных;

— оценить целесообразность дальнейшего исследования и доработки (дальнейшего совершенствования математических моделей) отдельных подсистем;

— оценить влияние альтернативных описаний (конкурирующих моделей), возникающих при неоднозначности физических представлений о работе тех или иных подсистем.

Вместе с тем при использовании результатов корреляционного анализа необходимо иметь в виду, что:

— во-первых, эти результаты существенно зависят от выбора функции отклика системы. Поэтому изменение представлений о функциональном назначении системы может привести и к изменению корреляционных связей;

— во-вторых, наличие сильной корреляционной связи между двумя переменными (функциями, подсистемами) хотя и показывает взаимосвязанность их изменений, но еще не доказывает наличия причинно-следственной связи между ними.

Вопросы для самоконтроля

1. Что такое степени свободы в математической статистике?
2. Что такое ковариация и коэффициент корреляции?
3. Запишите основные соотношения для коэффициента корреляции?

3.5. ОЦЕНКА АДЕКВАТНОСТИ МОДЕЛИ СЛОЖНОЙ СИСТЕМЫ ПО ДАННЫМ ОБ АДЕКВАТНОСТИ ПОДСИСТЕМ

При разработке математических моделей вновь создаваемых сложных систем, в частности, ДЛА, статистические данные, необходимые для проверки адекватности модели действительности, отсутствуют. Поэтому единственным возможным путем оценки модели является анализ самой модели. Если создаваемая система синтезируется из уже существующих, а значит, изученных узлов и элементов, то оценка точности ее модели может быть полу-

чена как функция оценок точности моделей этих узлов и элементов.

При иерархическом описании системы ее модель представляет собой функционал W , выраженный через функционалы подсистем ближайшего нижнего уровня $(X_1, X_2, \dots, X_k)$, играющие роль агрегированных переменных:

$$W = f(X_1, X_2, \dots, X_k). \quad (3.28)$$

Сами эти функционалы X_i в свою очередь могут быть выражены через функционалы моделей следующего ниже лежащего уровня и т. д.

Если оценки точности моделей нижнего уровня известны, то можно построить и общую оценку точности модели системы. При этом, разумеется, можно принять в качестве такой оценки максимальную погрешность, возможную для моделей элементов. Однако полученный при этом «гарантированный» результат будет излишне пессимистичным.

Правильнее оценивать точность модели системы с учетом роли каждой из подсистем.

Пусть $W_n, X_{1n}, X_{2n} \dots X_{kn}$ — точные (идеальные) значения функционалов для модели системы в целом и моделей подсистем, а $w, x_1, x_2, \dots, x_k$ — отклонения от этих значений, т. е. погрешности моделей. Тогда

$$W_n + w = f(X_{1n} + x_1, X_{2n} + x_2, \dots, X_{kn} + x_k). \quad (3.29)$$

Разложив (3.29) в ряд Тейлора и сохранив в разложении только линейные члены, можно получить

$$W_n + w = f(X_{1n}, X_{2n}, \dots, X_{kn}) + \sum_{i=1}^k \frac{\partial W_n}{\partial X_i} x_i; \quad (3.30)$$

$$w = \sum_{i=1}^k \frac{\partial W_n}{\partial X_i} x_i.$$

Второе соотношение (3.30) после возведения в квадрат приводит к уравнению

$$w^2 = \sum_{i=1}^k \left(\frac{\partial W_n}{\partial X_i} \right)^2 x_i^2 + 2 \sum_{\substack{i=1 \\ j>i}}^k \frac{\partial W_n}{\partial X_i} \frac{\partial W_n}{\partial X_j} x_i x_j. \quad (3.31)$$

При многократных независимых исследованиях модели системы оценки (3.31) для каждого из таких исследований будут независимы. Суммирование этих оценок приводит к соотношению

$$\sum_n w^2 = \sum_{i=1}^k \left(\frac{\partial W_n}{\partial X_i} \right)^2 \sum_n x_i^2 + 2 \sum_{\substack{i=1 \\ j>i}}^k \frac{\partial W_n}{\partial X_i} \frac{\partial W_n}{\partial X_j} \sum_n x_i x_j. \quad (3.32)$$

Естественно предположить, что ошибки агрегированных переменных x_i описываются симметричным, в частности, нормальным распределением с нулевым математическим ожиданием $M(x_i) = 0$. В этом случае при достаточно больших n

$$\sum_n x_i x_j = 0 \text{ и } \sum_n w^2 = \sum_{i=1}^k \left(\frac{\partial W_{\text{и}}}{\partial X_i} \right)^2 \sum_n x_i^2. \quad (3.33)$$

Поделив (3.33) на $(n-1)$, можно получить оценки для дисперсии и среднеквадратического отклонения модели системы через соответствующие оценки агрегированных переменных (3.4а):

$$s^2(W) = \sum_{i=1}^k \left(\frac{\partial W_{\text{и}}}{\partial X_i} \right)^2 s^2(X_i); \quad (3.34)$$

$$s(W) = \sqrt{\sum_{i=1}^k \left(\frac{\partial W_{\text{и}}}{\partial X_i} \right)^2 s^2(X_i)}.$$

Аналогично могут быть получены и интервальные оценки l (l — интервал, в котором содержится один и тот же процент элементов генеральной совокупности результатов моделирования при многократном исследовании модели):

$$l_W^2 = \sum_{i=1}^k \left(\frac{\partial W_{\text{и}}}{\partial X_i} \right)^2 l_{X_i}^2; \quad (3.35)$$

$$l_W = \sqrt{\sum_{i=1}^k \left(\frac{\partial W_{\text{и}}}{\partial X_i} \right)^2 l_{X_i}^2}.$$

Соотношения (3.34), (3.35) содержат производные функционала модели системы по функционалам моделей подсистем. Эти производные характеризуют вклад каждой из подсистем в общие результаты моделирования системы и могут быть определены непосредственно на модели системы (как правило, численными методами).

Разумеется, необходимым условием использования оценок (3.34), (3.35) является учет всей совокупности подсистем, определяющих работу системы, т. е. полнота модели системы.

РЕГУЛЯРНЫЕ МАТЕМАТИЧЕСКИЕ МОДЕЛИ ДВИГАТЕЛЕЙ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

Регулярные математические модели включают в себя зависимости, отражающие физические закономерности рабочего процесса и характеристики конструкции ДЛА. Эти модели могут быть использованы тогда, когда имеются точные или приближенные, но достаточно подробные и достоверные данные о работе ДЛА или их элементов.

ГЛАВА 4. МАТЕМАТИЧЕСКИЕ МОДЕЛИ АНАЛИТИЧЕСКОГО ТИПА

4.1. ФУНКЦИЯ И ФУНКЦИОНАЛ

Простейшие математические модели двигательной установки и ее элементов могут быть заданы явно в виде функции одной или нескольких переменных. Обычно в виде функций задаются общие законы природы или общие закономерности, полученные в результате интегрирования дифференциальных уравнений. Примером такой модели может служить знаменитая формула К. Э. Циолковского

$$\Delta v_{\text{ЛА}} = v \ln \frac{M_0}{M_k}, \quad (4.1)$$

определяющая приращение скорости ракеты ($\Delta v_{\text{ЛА}}$) при импульсном сжигании топлива через скорость истечения рабочего тела (v) и отношение начальной (M_0) и конечной (M_k) масс ракеты.

Тяга ракетного двигателя определяется соотношением

$$P = \dot{m}v, \quad (4.2)$$

где $\dot{m}$ — расход; v — эффективная скорость истечения рабочего тела.

Тепловой поток q , излучаемый абсолютно черным телом, определяется законом Стефана — Больцмана:

$$q = C_s S \left(\frac{T}{100} \right)^4, \quad (4.3)$$

где C_s — постоянная Стефана; S — излучающая поверхность; T — температура.

Скорость ионов v в ускорителе ионного двигателя в соответствии с законом сохранения энергии определяется соотношением

$$v = \sqrt{\frac{2e\Delta U}{m}}, \quad (4.4)$$

где e , m — заряд и масса иона; ΔU — ускоряющая разность электростатических потенциалов.

Модель, заданная в явном виде, дает исчерпывающее описание исследуемого объекта. Она позволяет построить зависимость его характеристик от управляющих параметров, взять производные и найти экстремумы модели, определить характеристики модели в окрестности экстремумов и т. д. Очень удобна графическая интерпретация таких моделей. Однако модели в виде формул могут быть разработаны только для очень простых объектов. Поэтому они используются или на самых нижних иерархических уровнях описания ДУ при описании отдельных процессов (4.3), (4.4) или, наоборот, на самых верхних уровнях при описании общих свойств ДУ и ЛА (4.1), (4.2).

Характеристики изучаемой системы могут зависеть не непосредственно от переменных, а от выбора тех или иных функций переменных ($y(x)$):

$$W = W(y(x)); \quad W = W(x, y(x)); \quad y(x) \in Y. \quad (4.5)$$

В этом случае говорят, что модель задана в виде функционала (функции от функции). При этом функции $y(x)$ представляют собой специфические переменные, которые заданы на некотором классе функций Y .

К функционалам сводятся описания многочисленных баллистических и транспортных задач, задач, связанных с распределением ресурсов и капиталовложений, с заменой оборудования и т. д. Многие законы механики и физики могут быть сформулированы в виде требования максимизации или минимизации некоторого функционала (принцип наименьшего действия, закон сохранения энергии, принцип Ферма и т. д.). Эти законы носят название вариационных принципов механики или физики.

Например, при ряде достаточно естественных предположений¹ дальность полета ракеты L может быть описана функционалом

$$L = \int_0^t (\varphi dv_{\text{ЛА}} + \psi dm); \quad \varphi = -\frac{m}{kv_{\text{ЛА}}}; \quad \psi = -\frac{v}{kv_{\text{ЛА}}}, \quad (4.6)$$

где m — текущая масса ракеты.

Таким образом, дальность определяется выбором закона сжигания топлива $\dot{m}(v_{\text{ЛА}})$. Функции $\dot{m}(v_{\text{ЛА}})$ должны принадлежать к классу функций, удовлетворяющих заданным граничным услови-

¹ Миеле А. Определение экстремумов криволинейных интегралов по теореме Грина. В кн.: Методы оптимизации с приложениями к механике космического полета/Под ред. Д. Лейтмана. М.: Наука, 1965, с. 100—129.

ям (начальная M_0 и конечная M_k массы ракеты) и ограничениям на расход топлива ($0 \leq \dot{m} \leq m_{\max}$). При мгновенном сжигании топлива формула (4.6) преобразуется в выражение (4.1). При тех же предположениях время полета ракеты описывается функционалом

$$t_{\Sigma} = \int_0^t (\varphi_1 dv_{\text{ЛА}} + \psi_1 dm); \quad \varphi_1 = -\frac{m}{kv_{\text{ЛА}}^2}; \quad \psi_1 = -\frac{v}{kv_{\text{ЛА}}}. \quad (4.7)$$

Вероятность безотказной работы $\mathcal{P}$ ДЛА в течение интервала времени $(0, t_{\Sigma})$ может быть выражена через интенсивность потока отказов λ , зависящую от фазовых координат системы $x_j(t)$, управляющих функций $u_m(t)$, постоянных параметров k_l и времени t :

$$\mathcal{P} = \exp \left(- \int_0^{t_{\Sigma}} \lambda(x_j(t), u_m(t), k_l, t) dt \right). \quad (4.8)$$

В виде функционала, зависящего от функций, характеризующих поток заказов, может быть представлена стоимость функционирования системы массового обслуживания, многочисленные транспортные задачи, связанные с доставкой грузов со складов к пунктам потребления, и т. д.

В некоторых случаях к функционалам сводится описание отдельных рабочих процессов или элементов конструкции ДЛА. Например, потери импульса в сопле двигателя (Δp) могут быть выражены через показатель политропы n , газодинамическую функцию полного импульса $z(v^*)$, зависящую от приведенной скорости истечения v^* , и безразмерный интервал $\bar{l}$, характеризующий геометрический контур сопла [1]:

$$\Delta p = \frac{\left(\frac{2}{n+1} \right)^{1/n-1} [z(v^*) - 1] \bar{l}}{\left(\frac{2}{n+1} \right)^{1/n-1} z(v^*)}. \quad (4.9)$$

Модели в виде функционалов, так же как и модели в виде функций, используются для описания ДУ или на самых старших (баллистические задачи, распределение ресурсов), или на самых младших (выбор геометрии контура сопла) иерархических уровнях. При этом всегда подразумевается возможность более или менее полного аналитического исследования задачи.

Вопросы для самоконтроля

1. Чем отличаются модели, заданные в виде функции и функционала?
2. В чем заключаются преимущества моделей, заданных в виде функции и функционала?
3. Придумайте самостоятельно несколько моделей, заданных в виде функций.

4.2. МАТЕМАТИЧЕСКИЕ МОДЕЛИ В ВИДЕ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Математические модели в виде одного или нескольких обыкновенных дифференциальных уравнений широко используются при изучении переходных процессов в системах автоматического регулирования (САР) ДЛА, при описании баллистики ЛА, а также при описании процессов движения (потoki, частицы, механические элементы) в двигателе. В простейшем случае модель может иметь вид линейного дифференциального уравнения n -го порядка

$$a_n \frac{d^n x(t)}{dt^n} + a_{n-1} \frac{d^{n-1} x(t)}{dt^{n-1}} + \dots + a_1 \frac{dx(t)}{dt} + a_0 x(t) = f(t) \quad (4.10)$$

или системы дифференциальных уравнений первого порядка

[illegible]

Часто встречаются смешанные задачи, а также нелинейные обыкновенные дифференциальные уравнения.

Математическая модель, заданная в виде дифференциального уравнения (системы дифференциальных уравнений), должна включать в себя необходимый набор начальных условий. Так, для дифференциального уравнения (4.10) в начальный момент времени ($t=0$) должны быть заданы условия

$$x(0) = C_0; \quad \frac{dx(0)}{dt} = C_1, \dots, \frac{d^{n-1}x(0)}{dt^{n-1}} = C_{n-1}. \quad (4.12)$$

В качестве простейшего примера математической модели механической системы может быть рассмотрена модель движения груза массой m , закрепленного на вертикальной стенке с помощью пружины жесткостью C и совершающего колебательное движение вдоль оси x в среде с вязкостью ν (рис. 4.1). Возмущающая сила, вызывающая колебания, зависит от времени $f(t)$.

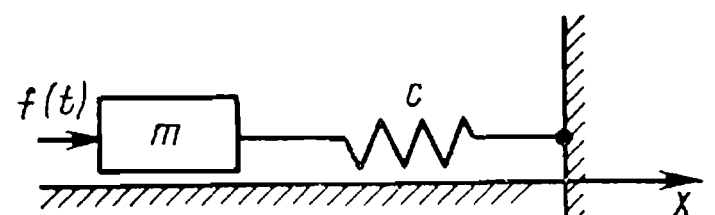
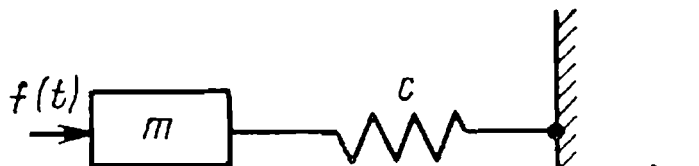


Рис. 4.1. Пример механической системы

$m \frac{d^2 x(t)}{dt^2}$, сила вязкого трения $\gamma \frac{dx(t)}{dt}$; усилие пружины $\frac{1}{C} x(t)$.

Все эти силы тормозят движение груза.

Согласно принципу Д'Аламбера сумма всех сил, действующих на груз, должна равняться нулю:

$$m \frac{d^2 x(t)}{dt^2} + \gamma \frac{dx(t)}{dt} + \frac{1}{C} x(t) - f(t) = 0. \quad (4.13)$$

Уравнение (4.13) совместно с начальными условиями, характеризующими начальное положение и начальную скорость груза,

$$x(0) = x_0; \quad \frac{dx(0)}{dt} = v_0 \quad (4.14)$$

представляет собой математическую модель рассмотренной механической системы.

К дифференциальным уравнениям сводится математическая модель динамического маневра ЛА при переводе ЛА из некоторого начального многообразия S_0 в конечное многообразие S_k [5]. Перевод осуществляется в пространстве координат — скоростей — времени

$$(\mathbf{r}(t_0), \mathbf{v}(t_0), t_0) \in S_0 \Rightarrow (\mathbf{r}(t_k), \mathbf{v}(t_k), t_k) \in S_k. \quad (4.15)$$

Система дифференциальных уравнений и граничных условий имеет вид

$$\begin{aligned} \dot{M} &= -\dot{m}, \quad \dot{\mathbf{r}} = \mathbf{v}, \quad \dot{\mathbf{v}} = \frac{\mathbf{P}(\mathbf{r}, t)}{M} + \mathbf{g}(\mathbf{r}, t); \\ M(t_0) &= M_0; \quad \mathbf{r}(t_0) = \mathbf{r}_0; \quad \mathbf{v}(t_0) = \mathbf{v}_0; \end{aligned} \quad (4.16)$$

$$M(t_k) = M_k; \quad \mathbf{r}(t_k) = \mathbf{r}_k; \quad \mathbf{v}(t_k) = \mathbf{v}_k,$$

где $\mathbf{P}(\mathbf{r}, t)$ — вектор тяги; $\mathbf{g}(\mathbf{r}, t)$ — вектор ускорения гравитационных сил.

Система (4.16) включает в себя уравнение изменения массы ЛА во время маневра и уравнения, характеризующие движение центра масс. В зависимости от конкретного выражения $\mathbf{P}$, $\mathbf{g}$, $\dot{m}$ модель (4.16) может принимать различные формы [5].

Дифференциальными уравнениями описывается модель раскрутки ротора одновального газотурбинного двигателя (ГТД) [24]. При ряде существенных упрощений эта модель может быть записана в виде

$$\begin{aligned} \frac{dn(t)}{dt} &= \frac{f(n, \dot{m})}{2\pi J}; \quad \frac{d\dot{m}}{dn} = \frac{2\pi J M}{f(n, \dot{m})}; \\ M &= \frac{d\dot{m}}{dt}; \quad n(t_0) = n_0; \quad M(t_0) = M_0, \end{aligned} \quad (4.17)$$

где n — число оборотов ротора; $\dot{m}$ — расход топлива; J — момент инерции ротора.

Функция $f(n, \dot{m})$ аналитически не выражается, поэтому (4.17) представляет собой аналитическую численно разрешимую модель. Попытка учесть реальные характеристики процесса раскрутки ротора (наличие автомата приемистости и т. д.) или характеристики сопряженных с ротором элементов ГТД приводит к существенному усложнению модели.

Переходные процессы при регулировании ядерного реактора могут быть описаны моделью [24]

$$\frac{d^2 N(t)}{dt^2} + \frac{\beta + \lambda t_H - \Delta k}{t_H} \frac{dN(t)}{dt} - \frac{1}{t_H} \left(\frac{d(\Delta k)}{dt} - \lambda \Delta k \right) N(t) = 0, \quad (4.18)$$

где $N(t)$ — плотность потока нейтронов; Δk — возмущение по стержневой реактивности, t_H — период жизни одного поколения нейтронов; β — доля запаздывающих нейтронов; λ — осредненная постоянная распада.

Движение иона в поле электромагнитных сил (при отсутствии столкновений и без учета релятивистских эффектов) в ускорителях электроракетных двигателей (ЭРД) может быть описано уравнением движения в форме Ньютона — Лоренца [13]:

$$\frac{m d \mathbf{v}(t)}{dt} = e \left[\mathbf{E} + \frac{1}{c} [\mathbf{v} \mathbf{H}] \right]; \quad \mathbf{v} = \frac{d \mathbf{r}}{dt}; \quad (4.19)$$

$$\mathbf{r}(0) = \mathbf{r}_0, \quad \mathbf{v}(0) = \mathbf{v}_0,$$

где m , $\mathbf{v}$, e — масса, скорость и заряд иона; c — скорость света; $\mathbf{E}$, $\mathbf{H}$ — напряженности электрического и магнитного полей; $\mathbf{r}$ — радиус-вектор.

В ускорителе ионного двигателя, где ускорение осуществляется исключительно за счет электростатических сил, уравнение (4.19) принимает вид

$$m \frac{d \mathbf{v}}{dt} = e \mathbf{E}; \quad \mathbf{r}(0) = \mathbf{r}_0, \quad \mathbf{v}(0) = \mathbf{v}_0, \quad (4.19a)$$

может быть проинтегрировано и приводит к модели (4.4).

Так же как и модели, заданные в виде явной функции или функционала, модели в виде обыкновенных дифференциальных уравнений очень просты и могут быть использованы на самых нижних уровнях иерархии или, наоборот, при описании ЛА или ДУ в целом.

Исследование моделей, заданных в виде обыкновенных дифференциальных уравнений, осуществляется *аналитическими, численными, детерминированными или статистическими*, а также *аналоговыми* методами. Безусловно, наиболее желательными и полными являются аналитические решения, обеспечивающие всесторонний анализ полученных результатов. К сожалению, такие решения

получены лишь для ограниченного числа дифференциальных уравнений. Численные и аналоговые методы решения позволяют найти лишь *конкретное* значение изучаемой функции при заданной комбинации исходных данных. Для анализа модели можно использовать некоторую совокупность решений. Однако очевидно, что результаты анализа в этом случае могут зависеть от выбора этой совокупности.

Методы аналитического и численного решения обыкновенных дифференциальных уравнений подробно рассматриваются в соответствующих разделах курса высшей математики. Некоторые примеры использования статистических и аналоговых решений обыкновенных дифференциальных уравнений приведены в разд. 6.2 и 9.2.

Вопросы для самоконтроля

1. Какие компоненты включает в себя модель, заданная в виде обыкновенного дифференциального уравнения?
2. Какую роль играют начальные условия при исследовании модели, заданной в виде обыкновенного дифференциального уравнения?
3. Приведите примеры моделей в виде обыкновенных дифференциальных уравнений.

4.3. МОДЕЛИ, ЗАДАННЫЕ В ВИДЕ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ

Ряд задач, связанных с исследованием физических полей в элементах ДЛА, а также в системах автоматического регулирования, приводит к моделям в виде дифференциальных уравнений в частных производных. Особенностью таких задач является зависимость изучаемых параметров от координат пространства. К моделям в виде уравнений в частных производных сводятся описания полей температур в элементах конструкции двигателя и полей скоростей при течении жидкости (газа). Уравнениями в частных производных описываются колебания элементов конструкции и поля напряжений, возникающих при работе этих элементов. К уравнениям в частных производных приводит описание электростатических и магнитных полей в ускорителях электро-ракетных двигателей. В зависимости от типа исследуемой величины поля могут быть или скалярными (поле температур) или векторными (поле скоростей). При этом каждому скалярному полю Φ может быть сопоставлено сопряженное векторное поле Ψ , связанное с ним соотношением

$$\Psi = - \text{grad } \Phi. \quad (4.20)$$

Так, скалярному полю температур соответствует векторное поле потоков тепла, а векторному полю скоростей — скалярное поле гидродинамического потенциала.

Наибольшее распространение на практике получили уравнения в частных производных второго порядка. Для описания стационар-

ных полей используются однородные и неоднородные уравнения в частных производных *эллиптического* типа

$$\nabla (\omega (\xi^r) \nabla \Phi (\xi^r)) = 0; \quad (4.21)$$

$$\nabla (\omega (\xi^r) \nabla \Phi (\xi^r)) = f (\xi^r, \Phi (\xi^r)); \quad (4.22)$$

включающие в себя потенциал — скалярную функцию $\Phi (\xi^r)$, зависящую от координат $\xi^r (r=1, 2, 3)^1$, характеристику среды $\omega (\xi^r)$ и мощность распределенных источников поля $f (\xi^r, \Phi (\xi^r))$. При $\omega (\xi^r) = \text{const}$ уравнения (4.21), (4.22) переходят в хорошо известные уравнения Лапласа и Пуассона:

$$\Delta \Phi (\xi^r) = 0, \quad (4.21a)$$

$$\Delta \Phi (\xi^r) = \frac{1}{\omega} f (\xi^r, \Phi (\xi^r)). \quad (4.22a)$$

Для описания нестационарных полей используются уравнения *параболического* и *гиперболического* типов, содержащие производные по времени:

$$\nabla (\omega (\xi^r, t) \nabla \Phi (\xi^r, t)) = a (\xi^r, t) \frac{\partial \Phi (\xi^r, t)}{\partial t}; \quad (4.23)$$

$$\nabla (\omega (\xi^r, t) \nabla \Phi (\xi^r, t)) = b (\xi^r, t) \frac{\partial^2 \Phi (\xi^r, t)}{\partial t^2}. \quad (4.24)$$

В общем случае поле может описываться общим уравнением второго порядка

$$\begin{aligned} \nabla (\omega (\xi^r, t, \Phi) \nabla \Phi (\xi^r, t)) = & f (\xi^r, t, \Phi) + a (\xi^r, t, \Phi) \frac{\partial \Phi (\xi^r, t)}{\partial t} + \\ & + b (\xi^r, t, \Phi) \frac{\partial^2 \Phi (\xi^r, t)}{\partial t^2}. \end{aligned} \quad (4.25)$$

Математическая модель, описанная уравнениями в частных производных, должна включать в себя также необходимые для решения задачи краевые условия. Во-первых, должна быть задана область D , ограниченная поверхностью (кривой) Γ , в которой определяется решение (рис. 4.2). Во-вторых, должны быть заданы условия на границе Γ этой области. В случае нестационарного поля эти граничные условия, так же как и сама область, могут меняться по времени.

Могут быть заданы граничные условия первого, второго и третьего рода. Граничные условия первого рода предусматривают задание на границе величины искомой функции

$$\Phi |_{\Gamma} = f_1 (\Gamma); \quad \Phi (t) |_{\Gamma} = f_1 (\Gamma, t), \quad (4.26)$$

¹ r — индекс. Для прямоугольных координат $\xi^1 = x$, $\xi^2 = y$, $\xi^3 = z$, для цилиндрических $\xi^1 = z$, $\xi^2 = r$, $\xi^3 = \alpha$ и т. д.

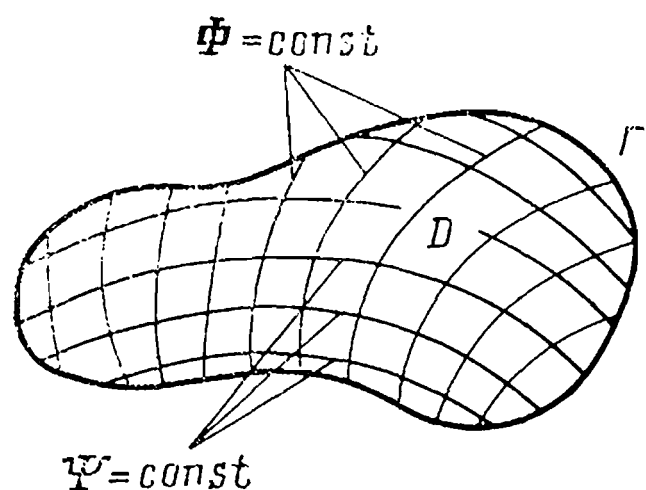


Рис. 4.2. Пример физического поля

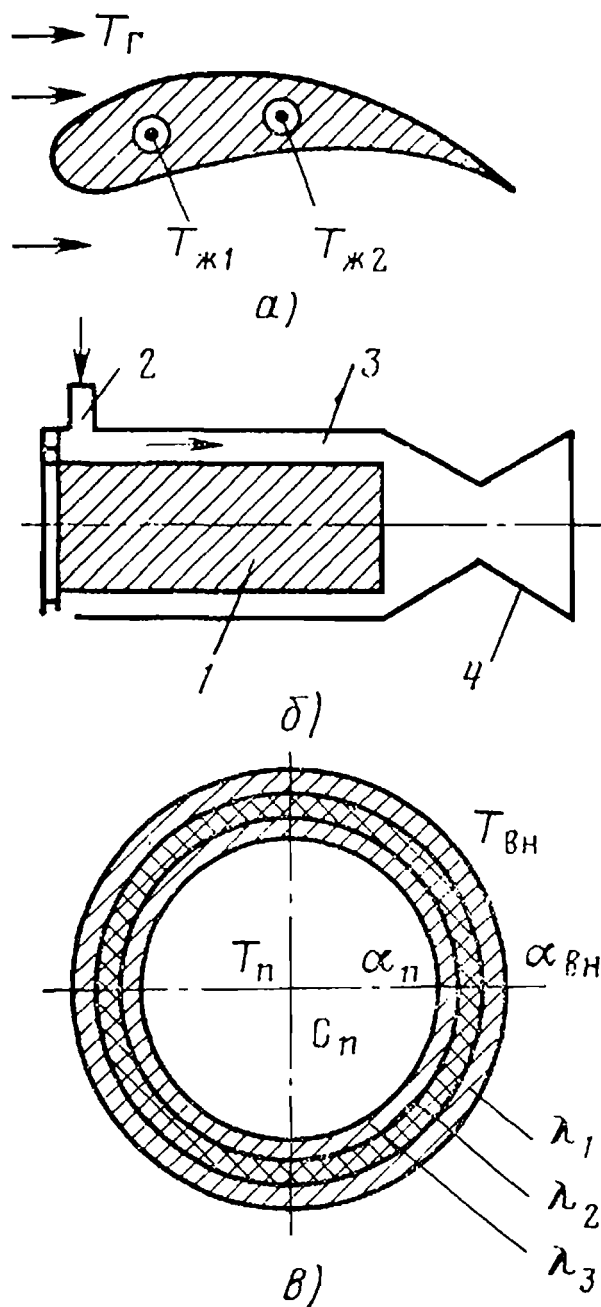


Рис. 4.3. Схемы объектов:

а — охлаждаемая лопатка газовой турбины; б — радиоизотопный двигатель; в — стенка сферического спутника; 1 — делящееся вещество; 2 — подача газа; 3 — газовый тракт; 4 — сопло

граничные условия второго рода — задание потока (производной искомой функции)

$$\frac{\partial \Phi}{\partial n} \Big|_{\Gamma} = f_2(\Gamma); \quad \frac{\partial \Phi}{\partial n}(t) \Big|_{\Gamma} = f_2(\Gamma, t), \quad (4.27)$$

граничные условия третьего рода — комбинации функции и ее производной

$$\left(a\Phi + b \frac{\partial \Phi}{\partial n} \right) \Big|_{\Gamma} = f_3(\Gamma); \quad \left(a(t)\Phi(t) + b(t) \frac{\partial \Phi}{\partial n}(t) \right) \Big|_{\Gamma} = f_3(\Gamma, t). \quad (4.28)$$

В-третьих, для нестационарных полей, кроме того, должны быть заданы одно (4.23) или два (4.24) начальных условия, характеризующие состояние поля в начальный момент времени:

$$\Phi(\xi^r) \Big|_{t=0} = f_4(\xi^r), \quad \frac{\partial \Phi}{\partial t}(\xi^r) \Big|_{t=0} = f_5(\xi^r). \quad (4.29)$$

Совокупность уравнений и краевых условий полностью определяет модель и позволяет провести ее исследование. Решение обычно задается в виде семейств изолиний $\Phi = \text{const}$, $\Psi = \text{const}$ (рис. 4.2).

Например, общее уравнение теплопроводности, определяющее распределение температур в твердом теле, может быть записано в виде

$$\nabla(\lambda(\xi^r, t) \nabla T(\xi^r, t)) = C(\xi^r, t) \rho(\xi^r, t) \frac{\partial T(\xi^r, t)}{\partial t} + w(\xi^r, t), \quad (4.30)$$

где T — температура; λ , C , ρ — соответственно теплопроводность,

теплоемкость и плотность вещества; ω — мощность тепловыделения в твердом теле.

Для стационарного температурного поля уравнение (4.30) принимает вид

$$\nabla (\lambda (\xi^r) \nabla T (\xi^r)) = \omega (\xi^r). \quad (4.30a)$$

В нестационарном случае, но при отсутствии распределенных источников тепла, уравнение (4.30) преобразуется к виду

$$\nabla (\lambda (\xi^r, t) \nabla T (\xi^r, t)) = C (\xi^r, t) \rho (\xi^r, t) \frac{\partial T (\xi^r, t)}{\partial t}. \quad (4.30б)$$

Наконец, при отсутствии источников стационарное температурное поле описывается уравнением

$$\nabla (\lambda (\xi^r) \nabla T (\xi^r)) = 0, \quad (4.30в)$$

а при постоянной теплопроводности ($\lambda = \text{const}$)

$$\Delta T = 0. \quad (4.30г)$$

Для температурного поля могут быть заданы граничные и начальные условия (4.26) — (4.29).

Стационарное температурное поле в охлаждаемой лопатке газовой турбины. Рассматривается температурное поле в лопатке газовой турбины с петлевым каналом охлаждения, по которому осуществляется принудительная циркуляция теплоносителя (рис. 4.3, а). Предполагается, что:

— изменением температуры по высоте лопатки можно пренебречь и, следовательно, достаточно рассмотреть температурное поле в одном ее сечении (поле плоскопараллельно);

— закон изменения температуры газа в межлопаточном канале, а следовательно, и у поверхности лопатки известен $T_r(\Gamma) = f_1(\Gamma)$;

— теплопроводность лопатки и температура теплоносителя также известны и постоянны ($\lambda = \text{const}$, $T_{ж1} = \text{const}$, $T_{ж2} = \text{const}$);

— тепловыделение в теле лопатки отсутствует.

При сделанных предположениях поле температур в лопатке двумерно ($T = T(x, y)$) и описывается уравнением Лапласа (4.30г).

Граничные условия определяются конвективным теплообменом на границе твердой и газовой (жидкой) сред. Тепловой поток q от газа к лопатке определяется соотношением

$$q = \alpha_r (T_r - T), \quad (4.31)$$

где α_r — коэффициент теплоотдачи от газа к лопатке.

С другой стороны, тепловой поток в твердом теле по нормали n к любой поверхности определяется соотношением

$$q = -\lambda \frac{\partial T}{\partial n}. \quad (4.32)$$

В силу стационарности поля на границе лопатки эти потоки равны, откуда следует условие

$$\alpha_r (T_r - T) = -\lambda \frac{\partial T}{\partial n} ; \left(T - \frac{\lambda}{\alpha_r} \frac{\partial T}{\partial n} \right) \Big|_r = T_r(\Gamma). \quad (4.33)$$

Аналогично для границы лопатки с теплоносителем легко получить

$$\left(T + \frac{\lambda}{\alpha_{ж}} \frac{\partial T}{\partial n} \right) \Big|_r = T_{ж}, \quad (4.34)$$

где $\alpha_{ж}$ — коэффициент теплоотдачи от лопатки к теплоносителю.

Таким образом, математическая модель температурного поля в лопатке задается уравнением (4.30г) с граничными условиями (4.33) и (4.34).

Стационарное температурное поле в тепловыделяющем элементе радиоизотопного ракетного двигателя. Радиоизотопный ракетный двигатель [5] представляет собой капсулу с радиоактивным изотопом, омываемую потоком рабочего вещества (рис. 4.3б). Капсула нагревается за счет радиоактивного распада. Рассматриваемая задача отличается от предыдущей наличием распределенных источников тепла (тепловыделение в капсуле), а также отсутствием заранее заданного закона распределения температуры в теплоносителе (рабочем веществе). Кроме того, строго говоря, нужно учесть различие теплопроводности в делящемся веществе и оболочке капсулы. В данном случае температурное поле двумерно ($T = T(z, r)$), описывается уравнением (4.30а) при граничных условиях (4.34) с заменой $\alpha_{ж}$ и $T_{ж}$ на α_r , T_r . В связи с тем, что распределение T_r заранее неизвестно, исследование модели должно осуществляться методом последовательных приближений (сначала при приближенном распределении $T_r = f(z)$ определяются температурное поле в капсуле и температура на ее границах, потом на основе этой температуры при заданном расходе теплоносителя уточняется распределение $T_r = f(z)$ и т. д.).

Нестационарное температурное поле в сложной стенке. Рассматривается нестационарное температурное поле в стенке космического аппарата при его входе в атмосферу. С целью упрощения задачи предполагается, что космический аппарат имеет сферическую форму, а его стенка представляет собой трехслойную оболочку из чередующихся конструкционных и теплоизолирующих материалов (рис. 4.3, в). Характеристики материалов (λ , C , ρ) заданы. При движении аппарата в атмосфере температура окружающей среды (пограничного слоя) изменяется по заданному закону $T_{вн} = f(t)$, определяемому траекторией и скоростью движения. Задано начальное распределение температуры в стенке аппарата $T_0 = f(r)$ и интегральная теплоемкость его внутренней полости $C_{пол}$. Распределение температурного поля в стенке описывается уравнением (4.30б). При сделанных предположениях это распределение одномерно $T = f(r, t)$. На внешней границе стенки выполняется условие (4.33), на внутренней (4.34). Однако изме-

нение температуры во внутренней полости $T_{\text{н}}$ должно определяться в процессе решения задачи.

Дифференциальные уравнения в частных производных второго порядка удовлетворительно характеризуют распределение потенциала для большинства реальных полей. Однако в некоторых случаях возникает необходимость использования уравнений более высокого порядка. Так, при анализе напряжений и изучении колебаний упругих твердых тел необходимо исследовать уравнения четвертого порядка. Стационарное поле напряжений в твердом теле описывается так называемым бигармоническим уравнением

$$\nabla^4 \Phi = 0, \tag{4.35}$$

которое в двумерном случае (например, для плоской мембраны) может быть записано в виде

$$\frac{\partial^4 \Phi}{\partial x^4} + 2 \frac{\partial^4 \Phi}{\partial x^2 \partial y^2} + \frac{\partial^4 \Phi}{\partial y^4} = 0; \tag{4.35a}$$

$$\frac{\partial^2 \Phi}{\partial x^2} = \sigma_y; \quad \frac{\partial^2 \Phi}{\partial y^2} = \sigma_x; \quad \frac{\partial^2 \Phi}{\partial x \partial y} = \rho_{xy},$$

где Φ — потенциал напряжений; σ_x , σ_y — нормальные напряжения в направлениях x , y соответственно; ρ_{xy} — касательное напряжение.

Модель должна включать в себя также граничные условия — напряжения на границах изучаемого тела.

Если напряжения меняются по времени, уравнение (4.35) должно быть заменено уравнением

$$\nabla^4 \Phi = k \frac{\partial^2 \Phi}{\partial t^2}, \tag{4.36}$$

а модель дополнена начальными условиями, характеризующими начальное положение и скорости всех точек тела.

Объекты, описываемые моделями в виде уравнений в частных производных, расположены, как правило, на самых нижних уровнях иерархии.

Исследование моделей, заданных в виде уравнений в частных производных, так же как и моделей в виде обыкновенных дифференциальных уравнений, осуществляется аналитическими, численными и аналоговыми методами. Оценки для каждого из этих методов полностью совпадают с оценками, приведенными в разд. 4.2. Нужно подчеркнуть, однако, что получить аналитические решения в данном случае удастся редко. Некоторые примеры статистических решений, а так-

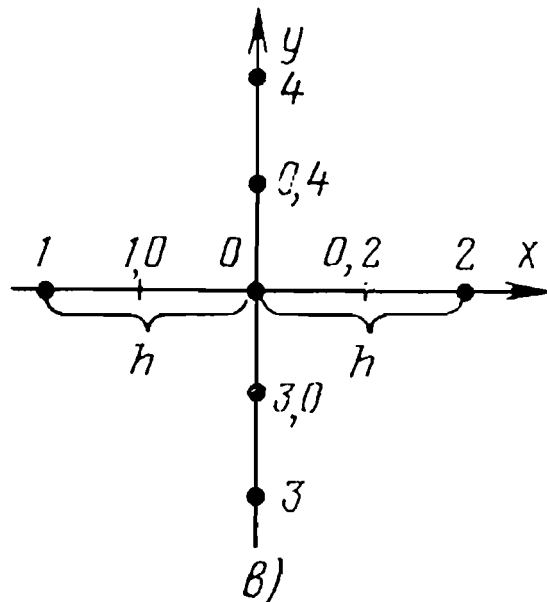
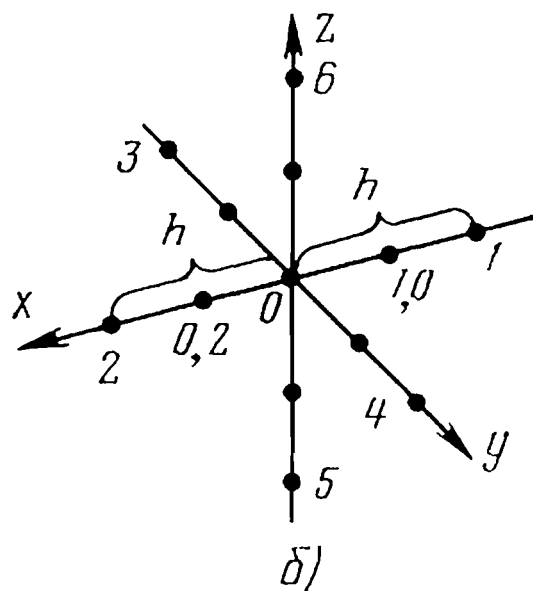
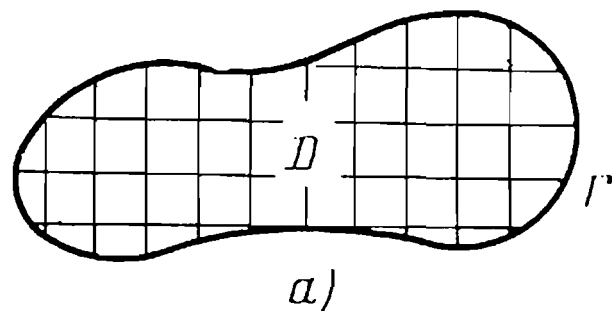


Рис. 4.4. Дискретное задание поля:

а — конечноразностная аппроксимация области; б — узловая точка в трехмерном случае; в — узловая точка в двумерном случае

же возможности построения аналоговых моделей для изучения полей рассмотрены в разд. 6.2. и 8.4.

Большинство численных решений, а также некоторые методы аналогового моделирования основаны на конечно-разностной аппроксимации уравнений в частных производных. При этом вся рассматриваемая область D покрывается некоторой (обычно прямоугольной) конечно-разностной сеткой (рис. 4.4, а), а исследуемая функция задается (определяется) лишь в узловых точках сетки. В промежуточных точках величина функции определяется путем интерполяции. Уравнения математической физики могут быть сведены к конечно-разностному виду заменой производных функций их конечно-разностными выражениями.

Для любой внутренней узловой точки сетки O (рис. 4.4, б)¹ при постоянном шаге вдоль каждой из осей ($h_x = h_y = h_z$) первые и вторые производные функции Φ имеют вид

$$\begin{aligned} \frac{\partial \Phi}{\partial x} \Big|_{1,0} &\approx \frac{\Phi_0 - \Phi_1}{h}; \quad \frac{\partial \Phi}{\partial x} \Big|_{0,2} \approx \frac{\Phi_2 - \Phi_0}{h}; \\ \frac{\partial^2 \Phi}{\partial x^2} \Big|_0 &\approx \frac{\frac{\partial \Phi}{\partial x} \Big|_{0,2} - \frac{\partial \Phi}{\partial x} \Big|_{1,0}}{h} = \frac{\Phi_2 - 2\Phi_0 + \Phi_1}{h^2}; \\ \frac{\partial \Phi}{\partial y} \Big|_{3,0} &\approx \frac{\Phi_0 - \Phi_3}{h}; \quad \frac{\partial \Phi}{\partial y} \Big|_{0,4} \approx \frac{\Phi_4 - \Phi_0}{h}; \\ \frac{\partial^2 \Phi}{\partial y^2} \Big|_0 &\approx \frac{\frac{\partial \Phi}{\partial y} \Big|_{0,4} - \frac{\partial \Phi}{\partial y} \Big|_{3,0}}{h} = \frac{\Phi_4 - 2\Phi_0 + \Phi_3}{h^2}; \\ \frac{\partial \Phi}{\partial z} \Big|_{5,0} &\approx \frac{\Phi_0 - \Phi_5}{h}; \quad \frac{\partial \Phi}{\partial z} \Big|_{0,6} \approx \frac{\Phi_6 - \Phi_0}{h}; \\ \frac{\partial^2 \Phi}{\partial z^2} \Big|_0 &\approx \frac{\frac{\partial \Phi}{\partial z} \Big|_{0,6} - \frac{\partial \Phi}{\partial z} \Big|_{5,0}}{h} = \frac{\Phi_6 - 2\Phi_0 + \Phi_5}{h^2}, \end{aligned} \quad (4.37)$$

где индексы $0, i; i, 0$ показывают, что соответствующая производная берется в середине отрезка между точками O и i .

Конечно-разностное выражение для лапласиана в точке O может быть записано в виде

$$\Delta \Phi = \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2} \approx \frac{1}{h^2} \left(\sum_{i=1}^6 \Phi_i - 6\Phi_0 \right), \quad (4.38)$$

а уравнение Лапласа (4.21а) — в виде

$$\sum_{i=1}^6 \Phi_i - 6\Phi_0 = 0; \quad \Phi_0 = \frac{1}{6} \sum_{i=1}^6 \Phi_i, \quad (4.39)$$

¹ На рис. 4.4, б представлен общий случай трехмерного поля ($\Phi = \Phi(x, y, z)$).

т. е. функция Φ в любой внутренней точке области D может быть выражена через значения функции в окружающих точках. Уравнения (4.38), (4.39) записаны для трехмерного поля, когда каждая внутренняя точка O сопряжена с шестью точками. В двумерном случае (рис. 4.4, в) таких точек четыре, поэтому в (4.38) и (4.39) все цифры 6 замещаются цифрой 4. В одномерном случае используется цифра 2.

Можно показать, что погрешность Λ конечно-разностной аппроксимации лапласиана пропорциональна квадрату шага сетки и четвертой производной функции Φ :

$$\Lambda = -\frac{h^2}{24} \sum_{i=1}^6 \Phi_{k_4}^{IV}(\xi^r). \quad (4.40)$$

В граничных точках области значения функции или заданы непосредственно (4.26), или могут быть выражены через граничные условия. Таким образом, конечно-разностная аппроксимация приводит к замене уравнения в частных производных (4.21а) системой алгебраических уравнений (4.39). Аналогично могут быть записаны и другие уравнения математической физики. Так, конечно-разностная аппроксимация поля, описываемого уравнением Пуассона (4.22а), может быть задана в виде

$$\Delta\Phi = f(\xi^r, \Phi) \Rightarrow \frac{1}{h^2} \left(\sum_{i=1}^6 \Phi_i - 6\Phi_0 \right) = A(\xi_0^r, \Phi_0); \quad (4.41)$$

$$A(\xi_0^r, \Phi_0) = h^3 f(\xi_0^r, \Phi_0),$$

где правая часть уравнения ($A(\xi_0^r, \Phi_0)$) характеризует суммарную мощность источников поля в элементарном объеме h^3 , соответствующем рассматриваемой точке (ξ_0^r).

При конечно-разностной аппроксимации нестационарных полей может быть использовано конечно-разностное представление времени (вводится интервал Δt). В этом случае уравнения (4.23), (4.24) могут быть представлены в так называемой *полной* конечно-разностной форме:

$$\frac{1}{h^2} \left(\sum_{i=1}^6 \Phi_i - 6\Phi_0 \right)_{t_i} = A(\xi_0^r, \Phi_0)_{t_i} \frac{\Phi_0 t_i - \Phi_0(t_i - \Delta t)}{\Delta t}; \quad (4.42)$$

$$A(\xi_0^r, \Phi_0) = h^3 a(\xi_0^r, \Phi_0);$$

$$\frac{1}{h^2} \left(\sum_{i=1}^6 \Phi_i - 6\Phi_0 \right)_{t_i} = B(\xi_0^r, \Phi_0)_{t_i} \frac{\Phi_0 t_i - 2\Phi_0(t_i - \Delta t) + \Phi_0(t_i - 2\Delta t)}{\Delta t^2}; \quad (4.43)$$

$$B(\xi_0^r, \Phi_0) = h^3 \bar{b}(\xi_0^r, \Phi_0).$$

Однако в ряде случаев конечно-разностная аппроксимация времени не производится и уравнения (4.23), (4.24) представляются в так называемой *неполной* конечно-разностной форме:

$$\frac{1}{h^2} \left(\sum_{i=1}^6 \Phi_i - 6\Phi_0 \right) = A(\xi_0^r, \Phi_0) \frac{\partial \Phi_0}{\partial t}; \quad (4.44)$$

$$\frac{1}{h^2} \left(\sum_{i=1}^6 \Phi_i - 6\Phi_0 \right) = B(\xi_0^r, \Phi_0) \frac{\partial^2 \Phi_0}{\partial t^2}. \quad (4.45)$$

Неполная конечно-разностная форма общего уравнения второго порядка (4.25) записывается в виде

$$\begin{aligned} \frac{1}{h^2} \left(\sum_{i=1}^6 \Phi_i - 6\Phi_0 \right) = & A(\xi_0^r, \Phi_0) + B(\xi_0^r, \Phi_0) \frac{\partial \Phi_0}{\partial t} + \\ & + D(\xi_0^r, \Phi_0) \frac{\partial^2 \Phi_0}{\partial t^2}. \end{aligned} \quad (4.46)$$

Так же как и в случае уравнения Лапласа, конечно-разностные представления (4.41) — (4.46) являются приближенными и сводят задачу решения уравнения в частных производных к задаче решения системы алгебраических уравнений.

Вопросы для самоконтроля

1. Приведите примеры математических моделей стационарных и нестационарных физических полей.
2. В чем заключается роль краевых условий? Какие краевые условия Вы знаете?
3. Что такое начальные условия? Когда они используются?
4. Запишите конечно-разностное выражение для уравнения Лапласа.
5. Чем отличаются полные и неполные конечно-разностные формы уравнений в частных производных?

ГЛАВА 5. ЧИСЛЕННЫЕ ДЕТЕРМИНИРОВАННЫЕ МОДЕЛИ

5.1 ЛИНЕЙНЫЕ ДЕТЕРМИНИРОВАННЫЕ МОДЕЛИ

Переход к численным моделям предполагает заведомый отказ от аналитических методов исследования. Поэтому численные модели более разнообразны и включают в себя различные по форме математические зависимости. Численные модели широко применяются практически на всех уровнях иерархического описания ДЛА. Вместе с тем наиболее простые численные модели допускают достаточно полное аналитическое исследование. Граница между аналитическими, особенно аналитическими численно разрешимыми,

и численными моделями достаточно условна. Детерминированные численные модели оперируют с точными значениями величин и однозначно заданными связями.

Наиболее простыми являются так называемые линейные детерминированные модели. Такие модели задаются в виде линейной формы управляющих переменных (x)

$$W = a_0 + a_1x_1 + a_2x_2 + \dots + a_kx_k \quad \left(W = a_0 + \sum_{i=1}^k a_ix_i \right) \quad (5.1)$$

при линейных ограничениях вида

$$b_{1j}x_1 + b_{2j}x_2 + \dots + b_{kj}x_k \geq b_j \quad \left(\sum_{i=1}^k b_{ij}x_i \geq b_j \right),$$

$$j = 1, 2, \dots, q_1;$$

$$c_{1j}x_1 + c_{2j}x_2 + \dots + c_{kj}x_k = c_j \quad \left(\sum_{i=1}^k c_{ij}x_i = c_j \right), \quad (5.2)$$

$$j = 1, 2, \dots, q_2;$$

$$d_{1j}x_1 + d_{2j}x_2 + \dots + d_{kj}x_k \leq d_j \quad \left(\sum_{i=1}^k d_{ij}x_i \leq d_j \right),$$

$$j = 1, 2, \dots, q_3.$$

причем общее число ограничений m может превосходить общее число переменных ($m = q_1 + q_2 + q_3 > k$). Кроме того, обычно вводится условие положительности переменных ($x_i \geq 0$).

Поверхность отклика для линейной модели представляет собой гиперплоскость. На рис. 5.1, а представлена графическая интер-

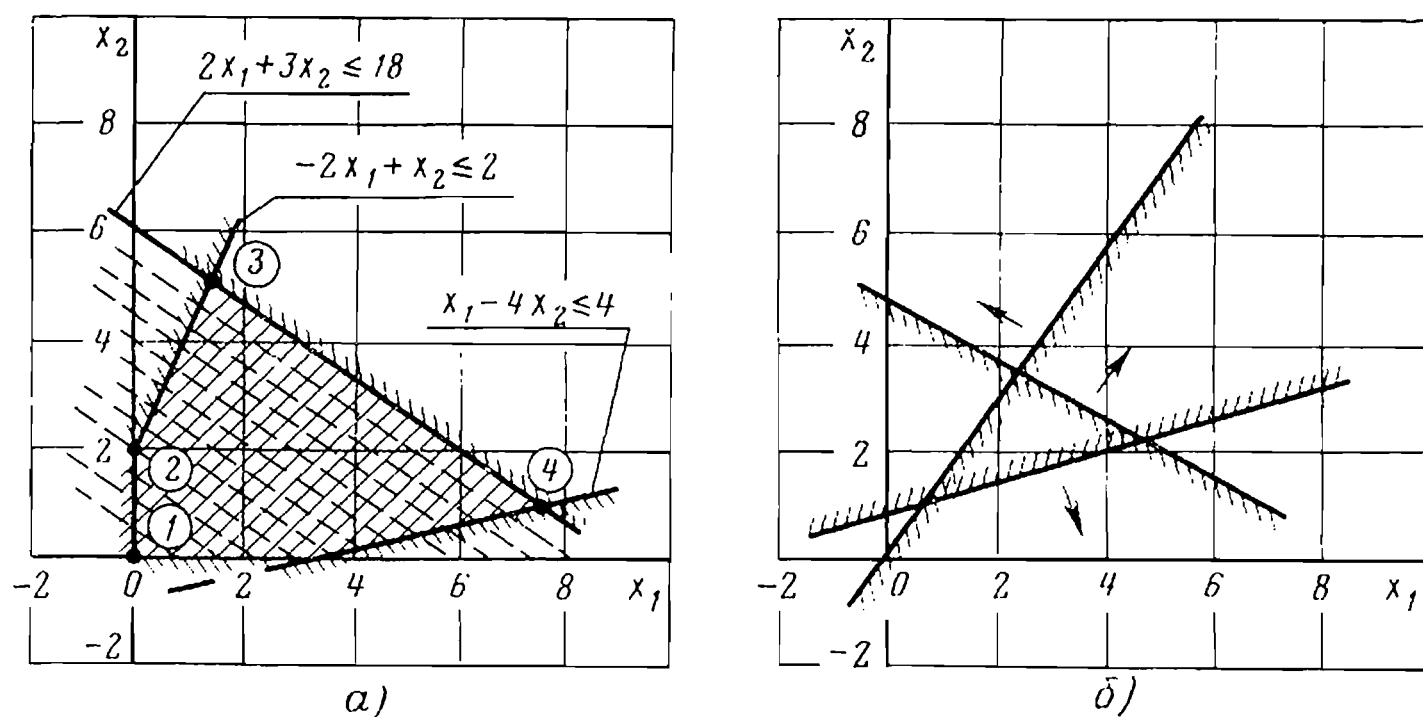


Рис. 5.1. Линейная модель:

а — двумерная задача; б — пустое множество возможных решений

претация (топографическая проекция) линейной модели для случая двух переменных при трех ограничениях:

$$W = -2x_1 - 3x_2; \quad (5.3)$$

$$2x_1 + 3x_2 \leq 18; \quad x_1 - 4x_2 \leq 4; \quad -2x_1 + x_2 \leq 2,$$

причем ограничения показаны сплошными линиями, а уровни поверхности отклика — пунктиром. Цифрами в кружках отмечены угловые точки границы области. Область допустимых решений заштрихована. При определенном соотношении ограничений множество допустимых решений может отсутствовать (множество допустимых решений *пусто*) (рис. 5.1, б).

Можно показать, что путем введения дополнительных свободных переменных ограничения типа неравенств могут быть сведены к ограничениям типа равенств. Действительно, неравенства, входящие в (5.2), можно записать в виде

$$\begin{aligned} \sum_{i=1}^k b_{ij}x_i \geq b_j &\Rightarrow \sum_{i=1}^k b_{ij}x_i - s_j = b_j, \quad s_j \geq 0; \\ \sum_{i=1}^k d_{ij}x_i \leq d_j &\Rightarrow \sum_{i=1}^k d_{ij}x_i + t_j = d_j, \quad t_j \geq 0. \end{aligned} \quad (5.4)$$

Иногда переменные s_j называют *остаточными* переменными, t_j — *избыточными* переменными. С учетом (5.4) линейная модель может быть записана в так называемой канонической форме [3]

$$W = a_0 + \sum_{i=1}^k a_i x_i; \quad \sum_{i=1}^k b_{ij} x_i = b_j; \quad (5.5)$$

$$i = 1, 2, \dots, k; \quad j = 1, 2, \dots, q; \quad b_j \geq 0, \quad x_i \geq 0.$$

Линейные модели достаточно просты и поэтому, с одной стороны, предполагают существенное упрощение задачи, а с другой — допускают разработку простых и эффективных методов решения. При исследовании ДЛА линейные модели используются редко и почти исключительно при приближенном описании задач старших

иерархических уровней. К линейным иногда сводятся простейшие модели стоимости, рассматриваемые как совокупность производимых затрат. Примером такой модели является классическая **модель стоимости перевозок** (рис. 5.2).

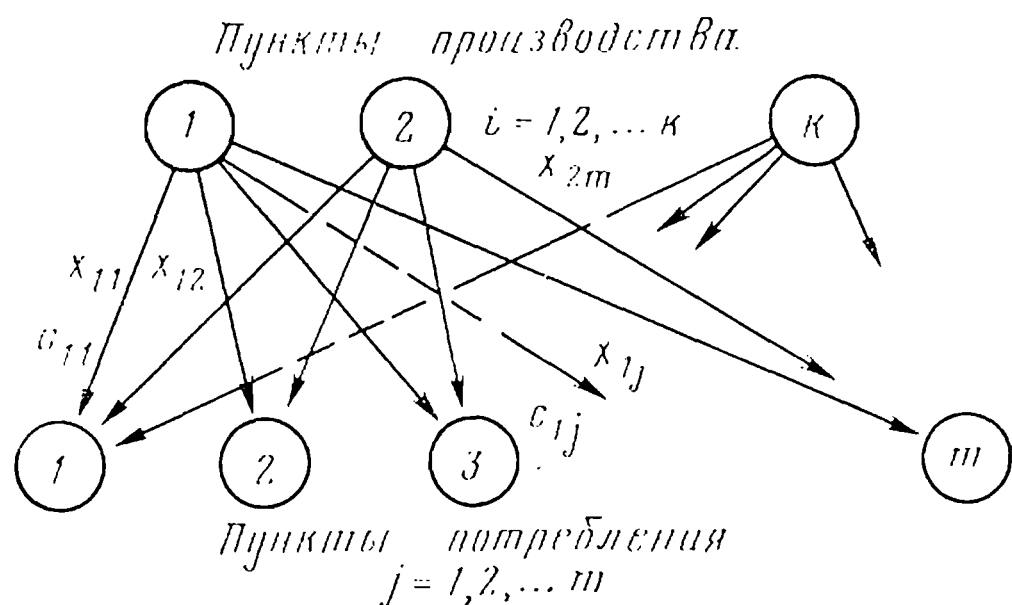


Рис. 5.2. Схема для решения транспортной задачи

Пусть имеется k ($i=1, 2, \dots, k$) пунктов производства и m ($j=1, 2, \dots, m$) пунктов потребления некоторого продукта. Количество продукта, произведенного в каждом из k пунктов, равно a_i , а количество продукта, необходимого в каждом из пунктов потребления, b_j . Обычно предполагается равенство общего производства и потребления $\left(\sum_{i=1}^k a_i = \sum_{j=1}^m b_j \right)$. Количество продукта, перевозимого из i пункта производства в j пункт потребления, равно x_{ij} , стоимость перевозки единицы продукта из i -го пункта производства в j -й пункт потребления c_{ij} . При сделанных предположениях суммарная стоимость перевозок c_Σ задается линейной моделью

$$c_\Sigma = \sum_{i=1}^k \sum_{j=1}^m c_{ij} x_{ij} \quad (5.6)$$

при естественных ограничениях

$$\sum_{j=1}^m x_{ij} = a_i; \quad \sum_{i=1}^k x_{ij} = b_j; \quad x_{ij} \geq 0, \quad (5.7)$$

учитывающих реальные возможности производства и потребления продукта в каждом из пунктов.

В ряде случаев линейные модели используются для простого и компактного описания экспериментальных данных и результатов численного моделирования в рамках простейших моделей идентификации (*линейная регрессия*). Они могут использоваться также при последовательной (поэтапной) аппроксимации более общих нелинейных моделей (*линеаризация задачи*). Особенно эффективен этот прием при изучении небольших областей исследуемого пространства (небольших участков поверхности отклика). Представление отдельных участков нелинейной поверхности отклика линейной моделью лежит в основе большой группы методов оптимизации, так называемых методов с линейной тактикой.

Исследование линейных моделей не представляет труда. В частности, влияние каждой из переменных на характеристики модели задается коэффициентами $a_i \left(\frac{\partial W}{\partial x_i} = a_i \right)$. Для нахождения оптимума линейной модели ($W_{\text{опт}}$) разработан эффективный симплекс-метод (см. разд. 14.4).

Вопросы для самоконтроля

1. Дайте общую характеристику численных детерминированных моделей.
2. Запишите каноническую форму линейной модели. Определите ее составные элементы.
3. Придумайте пример линейной модели.
4. Для чего используются линейные модели при описании ДЛА?

5.2. НЕЛИНЕЙНЫЕ ДЕТЕРМИНИРОВАННЫЕ МОДЕЛИ

Принятое в линейных моделях предположение о линейной зависимости изучаемых свойств объекта от всех действующих факторов несомненно является очень грубым. Большой точностью и гибкостью обладают так называемые нелинейные модели, не связанные с этим предположением. Определение нелинейной модели, вообще говоря, совпадает с общим определением модели (1.1), (1.2). Такая модель может быть записана в виде функционала, зависящего от управляющих переменных $(\mathbf{x})$ и некоторых функций $(\mathbf{f}(\mathbf{x}))$ всех или части этих переменных:

$$W = W(\mathbf{x}, \mathbf{f}(\mathbf{x})). \quad (5.8)$$

При этом функции $\mathbf{f}(\mathbf{x})$ могут представлять собой функционалы, зависящие от промежуточных функций $\mathbf{f}^*(\mathbf{x})$ и т. д. На класс функций $\mathbf{f}(\mathbf{x})$, $\mathbf{f}^*(\mathbf{x})$ не накладывается никаких ограничений, однако предполагается возможность однозначного перехода от вектора управляющих параметров $\mathbf{x}$ к общей характеристике модели W . Область существования модели может быть ограничена с помощью равенств или неравенств (1.3), (1.4). По существу под определение нелинейной модели подпадает любое математическое описание ДУ и ДЛА, не укладывающееся в стандартные рамки более простых моделей. Модели этого типа широко используются на всех уровнях иерархического описания ДЛА. Нелинейные модели ДЛА и их узлов обычно очень громоздки. Примеры таких моделей широко представлены в специальных дисциплинах и в настоящем пособии не приводятся. При поиске оптимума нелинейных моделей используются наиболее общие, так называемые прямые методы оптимизации. Целесообразно рассмотреть некоторые частные виды нелинейных моделей.

Полиномиальные модели основаны на идее приближенного представления модели конечным отрезком ее разложения в ряд Тейлора:

$$\begin{aligned} W(\mathbf{x}) = & W(\mathbf{x}_0) + \sum_{i=1}^k \frac{\partial W(\mathbf{x}_0)}{\partial x_i} (x_i - x_{i0}) + \\ & + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \frac{\partial^2 W(\mathbf{x}_0)}{\partial x_i \partial x_j} (x_i - x_{i0})(x_j - x_{j0}) + \dots \end{aligned} \quad (5.9)$$

Наиболее простой из моделей этого класса является **квадратичная модель**, задаваемая некоторой квадратичной формой

$$\begin{aligned} W(\mathbf{x}) = & a_0 + \sum_{i=1}^k a_i x_i + \sum_{\substack{i=1 \\ j=i}}^k a_{ij} x_i x_j; \\ & i, j = 1, 2, \dots, k \end{aligned} \quad (5.10)$$

при ограничениях типа неравенств или равенств (5.2).

Наряду с полными (5.10) часто используются неполные квадратичные модели, например, не содержащие квадратов переменных (x_i^2):

$$W(\mathbf{x}) = a_0 + \sum_{i=1}^k a_i x_i + \sum_{\substack{i=1 \\ j>i}}^k a_{ij} x_i x_j; \\ i, j = 1, 2, \dots, k. \quad (5.11)$$

Квадратичные модели широко используются для представления экспериментальных данных при идентификации ДЛА и их элементов. Такие модели могут быть использованы в САПР для представления результатов численного моделирования отдельных процессов и подсистем при передаче этих результатов от нижних иерархических уровней описания ДЛА к верхним.

Квадратичные модели, так же как и линейные, используются для аппроксимации отдельных участков поверхности отклика, когда линейное приближение оказывается недостаточно точным. В частности, квадратичная аппроксимация используется для описания поверхности отклика в окрестности экстремума и лежит в основе нелинейных методов оптимизации (см. разд. 16.4). Если квадратичная модель также оказывается недостаточно точной, используются полиномиальные модели более высоких порядков.

Исследование полиномиальных моделей частично можно осуществить аналитическими методами. Так, аналитически можно определить степень влияния отдельных переменных или комбинаций переменных (взаимодействий) на характеристики модели. При исследовании модели на оптимум обычно эффективны простейшие методы поиска экстремума.

Позиномные модели основаны на представлении модели в виде суммы произведений степенных функций. Модели этого типа в последнее время получили большое распространение при изучении сложных систем. В общем случае позиномная модель может быть записана в виде

$$W(\mathbf{x}) = \sum_{j=1}^m C_j x_1^{\alpha_{1j}} x_2^{\alpha_{2j}} \dots x_k^{\alpha_{kj}} = \sum_{j=1}^m C_j \prod_{i=1}^k x_i^{\alpha_{ij}}, \quad (5.12)$$

где x_i — управляющие переменные; α_{ij} — произвольные положительные числа. Обычно предполагается $C_j \geq 0$, что обеспечивает выпуклость модели.

Позиномные модели можно использовать для описания стоимости сложных систем. Так, можно показать, что суммарная стоимость C_Σ сложной системы, состоящей из m ($j=1, 2, \dots, m$) узлов, может быть записана в виде (5.12), если x_i — составляющие вектора технических параметров ($x_i \in \mathbf{X}$), α_{ij} — константы, характеризующие затраты, необходимые для достижения заданного уровня x_i , а C_j — константы, характеризующие вклад j узла в общие затраты. Величины α_{ij} и C_j рассчитываются на основе статистических данных, от-

ражающих опыт производства соответствующих узлов и систем. Важным достоинством позиционной модели стоимости является учет изменения стоимости при изменении технического совершенства системы. К позиционным моделям сводится задача выбора геометрических характеристик ряда технических устройств, в том числе элементов ДЛА, например, электромагнитов, силовых ферм и т. д.

Исследование позиционных моделей существенно сложнее, чем моделей полиномиального типа, и осуществляется в основном численными методами. Для поиска оптимальных решений на основе позиционных моделей разработан специальный аппарат — так называемое *геометрическое программирование*. Для этих же целей могут быть использованы общие методы поиска экстремума.

Сепарабельные нелинейные модели. Как уже отмечалось (см. разд. 2.2), одним из важнейших свойств многоуровневых иерархических структур сложных систем является их сепарабельность. Системы, обладающие свойством сепарабельности, могут быть описаны сепарабельными моделями, которые представляют собой линейную или нелинейную комбинацию функций (функционалов), определенных на независимых *непересекающихся* множествах переменных:

$$W(x) = \sum_{j=1}^N C_j f_j(x_j); \quad W(x) = \sum_{j=1}^N C_j f_j(x_j) + \\ + \sum_{\substack{j=1 \\ i \geq j}}^N C_{ij} f_i(x_i) f_j(x_j); \quad x_i \cap x_j = \emptyset; \quad i \neq j. \quad (5.13)$$

Результат функционирования сепарабельной системы представляет собой сумму эффектов функционирования подсистем (или сумму произведений этих эффектов). Соответственно сепарабельная модель представляет собой сумму (или сумму произведений) моделей нижнего уровня, что, разумеется, существенно упрощает исследование модели.

Вопросы для самоконтроля

1. Дайте общую характеристику нелинейных детерминированных моделей.
2. Что называется полиномиальной моделью? Для чего применяются полиномиальные модели?
3. Приведите пример позиционной модели.
4. В чем выражается свойство сепарабельности модели?

ГЛАВА 6. СТАТИСТИЧЕСКИЕ МОДЕЛИ. МЕТОД СТАТИСТИЧЕСКИХ ИСПЫТАНИЙ

6.1. ОБЩИЕ ПРИНЦИПЫ ПОСТРОЕНИЯ И ИССЛЕДОВАНИЯ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ

Точные величины и зависимости, используемые в детерминированных моделях, представляют собой лишь некоторые средние значения (математические ожидания) реальных случайных

величин (зависимостей). Так, физические константы, характеризующие материалы и рабочие тела (предел прочности материала σ , теплопроводность λ , плотность ρ и т. д.) меняются в зависимости от партии материала и условий окружающей среды. Всегда имеется определенный разброс размеров деталей (l), ресурсов элементов и узлов ДЛА (τ_p), гидравлических сопротивлений магистралей и расходов топлива в системах подачи. В ряде случаев возможны различные варианты протекания процесса (например, различные механизмы кипения жидкости при нагреве, различный характер взаимодействия частицы с твердой стенкой и т. д.). Все это приводит к тому, что и результирующие функции, характеризующие ДЛА или ее отдельные узлы, также носят случайный характер (например, имеется определенный разброс тяг двигателей, ресурсов и т. д.), а результаты, полученные с помощью детерминированной модели, представляют собой математическое ожидание этих характеристик. При этом конкретные данные для конкретной системы могут существенно отличаться от этих математических ожиданий. Например, ресурс конкретного двигателя может существенно отличаться от среднего ресурса двигателей данного типа. Для учета таких отличий вводятся всевозможные «запасы прочности», призванные гарантировать работоспособность реальных объектов при неблагоприятном стечении обстоятельств.

Значительно более полные и объективные результаты можно получить при переходе от детерминированных к *статистическим* моделям, т. е. при переходе от детерминированно заданных величин и связей к соответствующим случайным величинам. При этом константы (σ , λ , ρ , l , ...), а также коэффициенты зависимостей (a_0 , a_1 , a_2 , ...), учитывающих связи элементов, заменяются *случайными величинами* (ξ_σ , ξ_λ , ξ_ρ , ξ_l , ξ_{a_0} , ξ_{a_1} , ...), подчиненными определенным законам распределения. В ряде случаев в соответствии с заданным законом распределения может изменяться и тип используемых зависимостей. Конкретное значение той или иной величины (зависимости) для конкретного расчета определяется с помощью датчика случайных чисел.

Однократное исследование статистической модели приведет к некоторой случайной величине функции отклика ξ_w , представляющей собой, вообще говоря, ограниченную ценность. Для получения значимых результатов необходимо провести *многократное* исследование модели и получить распределение результирующей характеристики в интересующем исследователя диапазоне. Поверхность отклика в этом случае представляет собой некий размытый слой переменной плотности. Такой метод исследования статистической модели получил название метода *статистических испытаний* или *метода Монте-Карло*.

Трудоемкость исследования статистических моделей существенно выше, чем моделей детерминированных. Во-первых, значительно возрастает объем исходной информации. Замена констант случайными величинами, введение законов распределения этих величин, необходимость систематического обращения к специальным про-

граммам для их определения усложняют модель и увеличивают время ее исследования.

Во-вторых, для получения распределения результирующих функций необходимо многократное (обычно исчисляемое сотнями и тысячами) исследование модели. Поэтому статистические модели используются главным образом на нижних иерархических уровнях описания ДЛА. Вместе с тем их можно использовать для описания достаточно сложных *однородных* систем, например, многоэлементных систем с резервированием, систем электропитания и т. д.

С другой стороны, полученное при статистическом моделировании распределение характеристик системы дает в руки исследователя чрезвычайно ценную информацию. Такое распределение позволяет не только оценить среднее значение изучаемой величины, но и разброс этих значений, вероятности появления тех или иных значений при конкретном испытании (например, вероятность выхода из строя ДЛА через тот или иной промежуток времени) и их зависимость от различных факторов.

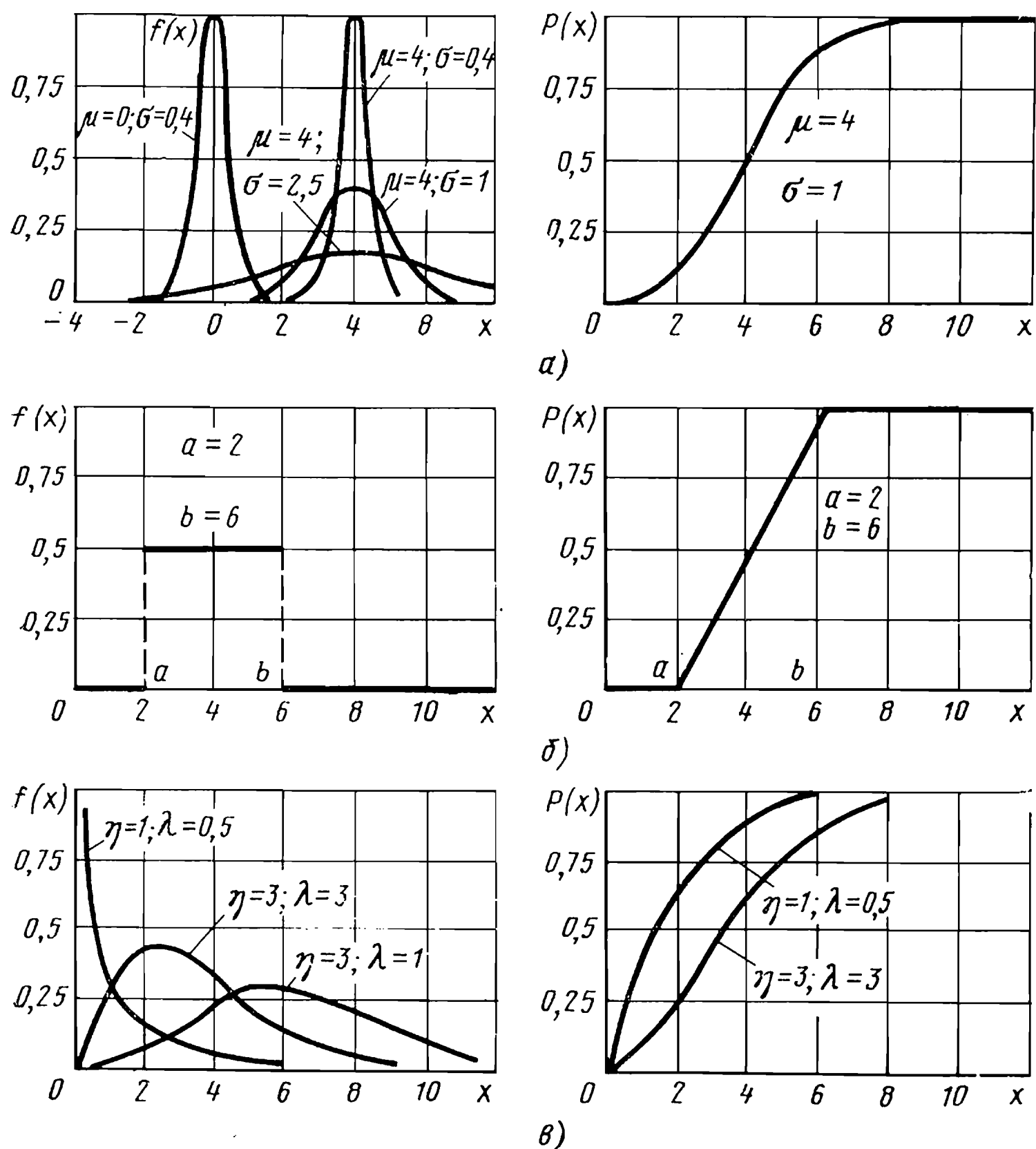


Рис. 6.1. Законы распределения случайных величин: нормальное (а), равномерное (б) и гамма-распределение (в)

Для построения статистической модели необходимо прежде всего задать законы распределения входящих в нее случайных величин.

Очень часто используется *нормальный* или *гауссов* закон распределения, для которого плотность вероятности $f(x)$ и распределение накопленной вероятности $\mathcal{P}(x)$ задаются соотношениями (рис. 6.1, а)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right); \quad (6.1)$$

$$\mathcal{P}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt.$$

Нормальное распределение определяется двумя параметрами: μ и σ . Для случайной величины ξ , распределенной по нормальному закону, $\mu = M(\xi)$; $\sigma = \sigma(\xi)$ (3.3), (3.4).

Популярность нормального распределения объясняется тем, что в соответствии с центральной предельной теоремой теории вероятностей оно описывает случайные величины, зависящие от большого числа различных и примерно равнозначных случайных воздействий. Так, по нормальному закону распределены характеристики материалов (предел прочности, теплопроводность), размеры деталей, ресурсы элементов ДЛА. Однако наряду с нормальным часто используются и некоторые другие законы распределения случайных величин.

Широко используется *равномерное* распределение, задающее равновероятные на отрезке $[a, b]$ оси x случайные величины (рис. 6.1, б):

$$f(x) = \begin{cases} 0, & -\infty < x < a \\ \frac{1}{b-a}, & a \leq x \leq b \\ 0, & b < x < \infty, \end{cases}; \quad (6.2)$$

$$\mathcal{P}(x) = \begin{cases} 0, & -\infty < x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b < x < \infty \end{cases}.$$

Случайные величины, связанные с процессами постоянной интенсивности (поток метеоритов в космическом пространстве, потоки вызовов в системах массового обслуживания, потоки отказов в сложных системах, включающих в себя большое количество неза-

висимых элементов и т. д.), описываются *гамма-распределением* (рис. 6.1, в):

$$\begin{aligned} f(x, \eta, \lambda) &= \frac{\lambda^\eta}{\Gamma(\eta)} x^{\eta-1} \exp(-\lambda x); \\ \mathcal{P}(x, \eta, \lambda) &= \frac{\lambda^\eta}{\Gamma(\eta)} \int_0^x t^{\eta-1} \exp(-\lambda t) dt; \\ x &\geq 0, \lambda > 0, \eta > 0, \end{aligned} \tag{6.3}$$

где $\Gamma(\eta)$ — гамма-функция; η — параметр формы; λ — параметр масштаба.

Частным случаем гамма-распределения ($\eta=1$) является часто используемое при расчетах надежности ДЛА *экспоненциальное распределение*

$$\begin{aligned} f(x, \lambda) &= \lambda \exp(-\lambda x); \quad \mathcal{P}(x, \lambda) = \int_0^x \lambda \exp(-\lambda t) dt = \\ &= 1 - \exp(-\lambda x); \quad x \geq 0; \lambda > 0, \end{aligned} \tag{6.4}$$

описывающее, в частности, время работы системы до первого отказа, если интенсивность потока отказов λ постоянна.

Используются также β -распределение, t -распределение Стьюдента, распределение Вейбулла и некоторые другие.

Для задания дискретных случайных величин чаще других используются биномиальное распределение

$$\begin{aligned} \mathcal{P}(x) &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}; \quad 0 \leq p \leq 1; \\ x &= 0, 1, \dots, \end{aligned} \tag{6.5}$$

характеризующее вероятность появления x событий в n испытаниях, если вероятность появления события в отдельном испытании равна p , и распределение Пуассона

$$\mathcal{P}(x) = \frac{\lambda^x}{x!} \exp(-\lambda); \quad \lambda \geq 0, \tag{6.6}$$

характеризующее вероятность появления x событий за определенный период времени или в определенном интервале пространства при интенсивности появления событий λ . В пределе биномиальное распределение и распределение Пуассона переходят в нормальное и гамма-распределение.

На основе экспериментальных данных могут быть построены также эмпирические законы распределения случайных величин.

Выбор закона распределения для конкретной случайной величины, входящей в статистическую модель, может быть обоснован или экспериментально, или теоретически. Конкретные же параметры распределения ($\mu, \sigma, \lambda, \eta, \dots$) всегда определяются на основе

экспериментальных данных. Так, в силу общих соотношений математической статистики оценка параметров нормального распределения на основе выборки $\{x_i\}$ из n случайных значений величины x дается соотношениями

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n};$$

$$\hat{\sigma} = s = \left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \right]^{1/2} = \left[\frac{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}{n(n-1)} \right]^{1/2}. \quad (6.7)$$

Параметры Γ -распределения оцениваются соотношениями

$$\hat{\lambda} = \frac{\bar{x}}{s^2} = (n-1) \left[\frac{\sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \right];$$

$$\hat{\eta} = \frac{\bar{x}^2(n-1)}{\sum_{i=1}^n (x_i - \bar{x})^2} = \bar{\lambda} \bar{x}.$$

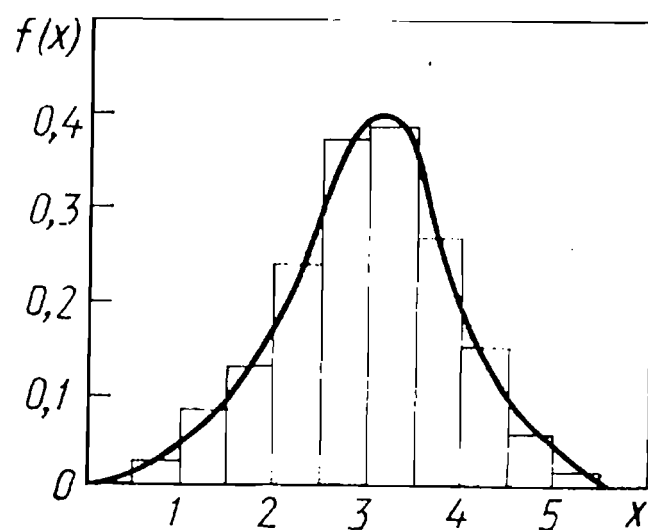
Оценка величины $\hat{\lambda}$ для экспоненциального распределения дается соотношением

$$\hat{\lambda} = \frac{1}{\bar{x}} = \frac{n}{\sum_{i=1}^n x_i}. \quad (6.9)$$

Выбор конкретной случайной величины при исследовании модели (при использовании метода статистических испытаний отдельный расчет характеристик модели принято называть *реализацией*) осуществляется с помощью *датчика случайных чисел*. Датчик случайных чисел в соответствии с принятым законом распределения вырабатывает последовательность случайных чисел, совокупность которых задает гистограмму данного закона распределения (рис. 6.2). При этом основную роль играет последовательность равномерно распределенных на отрезке $[0, 1]$ случайных чисел ξ_p .

Равномерно распределенные на отрезке $[0, 1]$ случайные числа можно получить с помощью таблиц, специальных радиотехнических устройств или рассчитать на ЭВМ по специальным программам.

Рис. 6.2. Гистограмма нормально распределенной случайной величины $f(x)$



Таблицы случайных чисел приводятся в справочниках и иногда используются при ручном счете. Ввод таких таблиц в память ЭВМ хотя принципиально и возможен, но практически нерентабелен, так как приводит к перегрузке запоминающих устройств. В качестве датчика случайных чисел может быть использовано специальное устройство, преобразующее результаты случайного физического процесса (обычно шум электронных ламп) в машинный код. Этот код непрерывно поступает в определенную ячейку памяти ЭВМ, и обращение к датчику эквивалентно обращению к памяти машины.

Широко распространен программный способ расчета последовательности случайных чисел на ЭВМ. Случайные числа рассчитываются с помощью того или иного рекуррентного соотношения, т. е. очередное случайное число ξ_r^n определяется по предыдущему числу ξ_r^{n-1} или группе предыдущих чисел. Такая последовательность, строго говоря, не является случайной (если при неизменных исходных данных повторить вычисления, последовательность повторяется), но может удовлетворять статистическим критериям случайности. В этом случае принято говорить о псевдослучайных числах. В пределах определенного отрезка последовательности (для современных программ этот отрезок составляет 10^6 — 10^7 чисел) свойства псевдослучайных чисел не отличаются от свойств подлинно случайных. Программы определения равномерно распределенных на отрезке $[0, 1]$ случайных чисел входят в библиотеки стандартных программ всех современных ЭВМ и приведены в справочниках по программированию.

С помощью датчика равномерно распределенных на отрезке $[0, 1]$ случайных чисел ξ_r можно построить последовательности случайных чисел ξ , соответствующих другим законам распределения. Ряд формул, позволяющих осуществить такой переход для некоторых, часто встречающихся законов распределения, представлен в табл. 6.1.

В общем случае последовательность случайных чисел ξ , распределенных на интервале $a < \xi < b$ с плотностью вероятности $f(x)$, может быть получена с помощью соотношения

$$\int_a^{\xi} f(x) dx = \xi_r. \quad (6.10)$$

Допустим случай $a = -\infty$, $b = \infty$. Если (6.10) разрешим относительно ξ , можно получить явную форму для «разыгрывания» ξ . В противном случае уравнение (6.10) приходится решать численным методом, что, конечно, увеличивает затраты машинного времени на моделирование.

При использовании метода статистических испытаний характеристики изучаемой системы оцениваются на основе некоторой ограниченной выборки реализаций. Поэтому важно определить достоверность этой оценки.

Вероятность p пребывания системы в некотором состоянии (например вероятность того, что время работы элемента ДЛА до пер-

Тип распределения	Плотность распределения вероятности	Способ получения случайной величины
Нормированное нормальное	$f(x, 0, 1) =$ $= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ $-\infty < x < +\infty$	$\xi_N = \frac{\sum_{i=1}^k \xi_{pi} - \frac{k}{2}}{\sqrt{k/12}}$ $k \rightarrow \infty (k \geq 12)$
Произвольное нормальное	$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \times$ $\times \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$ $-\infty < x < \infty;$ $-\infty < \mu < \infty; \sigma > 0$	$\xi = \sigma\xi_N + \mu$
Логарифмическое нормальное	$f(x, \mu, \sigma) = \frac{1}{\sigma x\sqrt{2\pi}} \times$ $\times \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right]$ $x \geq 0; -\infty < \mu < \infty; \sigma > 0$	$\xi = \exp(\sigma\xi_N + \mu)$
Гамма-распре- деление	$f(x, \eta, \lambda) = \frac{\lambda^\eta}{\Gamma(\eta)} \times$ $\times x^{\eta-1} \exp(-\lambda x)$ $x \geq 0; \lambda > 0; \eta > 0$	$\xi = -\frac{1}{\lambda} \sum_{i=1}^{\eta} \ln(1 - \xi_{pi})$
Экспоненциаль- ное	$f(x, \lambda) = \lambda \exp(-\lambda x)$ $x \geq 0; \lambda > 0$	$\xi = -\frac{1}{\lambda} \ln(1 - \xi_p)$
Распределение χ^2	$f(x, \gamma) = \frac{1}{2^{\gamma/2} \Gamma\left(\frac{\gamma}{2}\right)} \times$ $\times x^{\frac{\gamma}{2}-1} \exp\left(-\frac{x}{2}\right)$ $0 \leq x < \infty$	$\xi = \sum_{i=1}^{\gamma} \xi_{Ni}^2$

Тип распределения	Плотность распределения вероятности	Способ получения случайной величины
Бэ́та-распреде- ление	$f(x, \gamma, \eta) = \frac{\Gamma(\gamma + \eta)}{\Gamma(\gamma) \Gamma(\eta)} \times$ $\times x^{\gamma-1} (1-x)^{\eta-1}$ $0 \leq x \leq 1; \eta > 0; \gamma > 0$	$\xi = \frac{\sum_{i=1}^{2\gamma} \xi_{Ni}^2}{\sum_{i=1}^{2\gamma} \xi_{Ni}^2 + \sum_{i=2\gamma+1}^{2\gamma+2\eta} \xi_{Ni}^2}$
Распределение Вейбу́лла	$f(x, \sigma, \eta) = \frac{\eta}{\sigma} \left(\frac{x}{\sigma}\right)^{\eta-1} \times$ $\times \exp\left[-\left(\frac{x}{\sigma}\right)^\eta\right]$ $x \geq 0; \sigma > 0; \eta > 0$	$\xi = -\sigma [\ln(1 - \xi_p)]^{1/\eta}$
Распределение <i>S</i> _B -Джонсона	$f(x) = \frac{\eta}{\sqrt{2\pi}} \times$ $\times \frac{\lambda}{(x - \varepsilon)(\lambda - x + \varepsilon)} \times$ $\times \exp\left\{-\frac{1}{2} \left[\gamma + \eta \times\right.\right.$ $\times \left.\ln\left(\frac{x - \varepsilon}{\lambda - x + \varepsilon}\right)\right]^2\}$ $\varepsilon \leq x \leq \varepsilon + \lambda$	$\xi = \frac{\varepsilon + (\varepsilon + \lambda) \exp\left(\frac{\xi_N - \gamma}{\eta}\right)}{1 + \exp\left(\frac{\xi_N - \gamma}{\eta}\right)}$

вого отказа составит не менее *t* часов), определяется частотой реализации этого события при моделировании:

$$p \simeq \frac{n_+}{n}, \tag{6.11}$$

где *n*₊ — число реализаций, при которых наблюдалось изучаемое состояние системы (время работы ДЛА до первого отказа превысило *t*); *n* — общее число реализаций.

Оценка (6.11) является приближенной, так как определяется на основе ограниченной выборки. Отношение *n*₊/*n* часто называют *выборочной статистикой*.

Ошибка моделирования определяется отклонением выборочной статистики от вероятности:

$$\delta = \left| \frac{n_+}{n} - p \right|. \tag{6.12}$$

Можно показать [10], что при самых общих предположениях (при выводе используется неравенство Чебышева, т. е. так называемый

слабый закон больших чисел) эта ошибка удовлетворяет неравенству

$$\delta = \left| \frac{n_+}{n} - p \right| \leq \sqrt{\frac{p(1-p)}{an}}, \quad (6.13)$$

откуда следует оценка необходимого числа реализаций:

$$n \geq \frac{p(1-p)}{a\delta^2}. \quad (6.14)$$

В соотношениях (6.13), (6.14) p — вероятность рассматриваемого состояния; a — вероятность невыполнения оценки (6.13). Соответственно доверительная вероятность выполнения этой оценки равна $1-a$.

Из (6.13) следует, что погрешность статистического моделирования обратно пропорциональна $\sqrt{n}$:

$$\delta \sim \frac{1}{\sqrt{n}}. \quad (6.15)$$

Соотношение (6.15) имеет фундаментальное значение. С его помощью определяют точность метода статистических испытаний.

Из (6.15), в частности, следует, что увеличение точности при статистическом моделировании требует значительного увеличения числа реализаций. Так, для уменьшения погрешности в 10 раз необходимо увеличить число реализаций (а значит, и время счета!) в 100 раз. Поэтому метод статистических испытаний не может дать решения с очень высокой степенью точности. Обычно считается, что допустимая ошибка может составлять 1—5% максимальной величины, получаемой при моделировании.

Важно подчеркнуть, что хотя оценка (6.13) при некоторых дополнительных предположениях может быть заметно улучшена, зависимость (6.15) сохраняется всегда.

Величина ошибки зависит также от вероятности p оцениваемого состояния и допустимого уровня риска a . В частности, если перейти от абсолютной ошибки δ в (6.13) к относительной ошибке $\bar{\delta}$:

$$\bar{\delta} = \frac{\delta}{p} \leq \sqrt{\frac{(1-p)}{anp}}, \quad (6.13a)$$

легко видеть, что необходимое число испытаний

$$n \geq \frac{1-p}{ap\bar{\delta}^2} \quad (6.14a)$$

при постоянной величине относительной ошибки ($\bar{\delta} = \text{const}$) обратно пропорционально вероятности p . Поэтому количество испытаний, необходимых для оценки малых вероятностей, оказывается очень большим. Обычно в таких случаях рекомендуют преобразовать задачу в другую с приемлемыми значениями вероятности p .

Уменьшение допустимого уровня риска ($\alpha=0,05 \Rightarrow \alpha=0,01$) также приводит к увеличению необходимого числа испытаний. Обычно α задают на одном из фиксированных уровней ($\alpha=0,1; 0,05; 0,01, \dots$).

Вопросы для самоконтроля

1. Чем отличается статистическая модель от детерминированной?
2. В чем суть метода статистических испытаний?
3. Как осуществляется выбор случайных величин при статистическом моделировании?
4. Перечислите известные Вам законы распределения случайных величин.
5. Для каких величин может быть использован нормальный закон распределения?
6. Как оценивается точность статистического моделирования?

6.2. СТАТИСТИЧЕСКИЕ МОДЕЛИ ЭЛЕМЕНТОВ ДЛЯ

Статистические модели предназначены в первую очередь для изучения объектов и процессов стохастической природы. Однако в ряде случаев статистическое моделирование используется и при изучении детерминированных задач.

Модель накопления рабочего вещества при орбитальном полете летательного аппарата в верхних слоях атмосферы. На рис. 6.3, а показана принципиальная схема двигательной установки орбитального летательного аппарата с использованием рабочего вещества, накапливаемого во время полета [5]. Поток разреженного газа поступает в диффузор, собирается, охлаждается, конденсируется и накапливается в баке. Из бака рабочее тело подается к двигателям, предназначенным для компенсации аэродинамического сопротивления аппарата, поддержания орбиты или совершения заданных маневров. Целью моделирования является определение количества рабочего вещества, собранного в единицу времени.

Задача решается при ряде дополнительных предположений¹. Во-первых, предполагается, что длина свободного пробега молекул газа на высоте полета ЛА существенно больше характерных размеров диффузора. Поэтому взаимодействие молекул не учитывается (движение является свободномолекулярным) и каждая из них движется прямолинейно и равномерно до момента столкновения со стенкой.

Во-вторых, предполагается, что плотность вещества на орбите и скорость ЛА заданы, а значит, известно общее число молекул, попадающих на срез диффузора (сечение $I—I$) в единицу времени.

В-третьих, предполагается, что молекулы, попавшие в радиатор, не могут вылететь обратно, т. е. невозможно движение справа налево через сечение $II—II$. Молекулы, попавшие на стенку диффузора, могут упруго отразиться от нее или, отдав энергию стенке, испариться. Вероятности каждого из взаимодействий известны.

¹ Кузнецов Ю. Е., Флакман Я. Ш. Влияние формы поперечного сечения канала на характеристики свободномолекулярного воздухозаборника. — Тр. ЦАГИ, вып. 1863, 1977, с. 25—29.

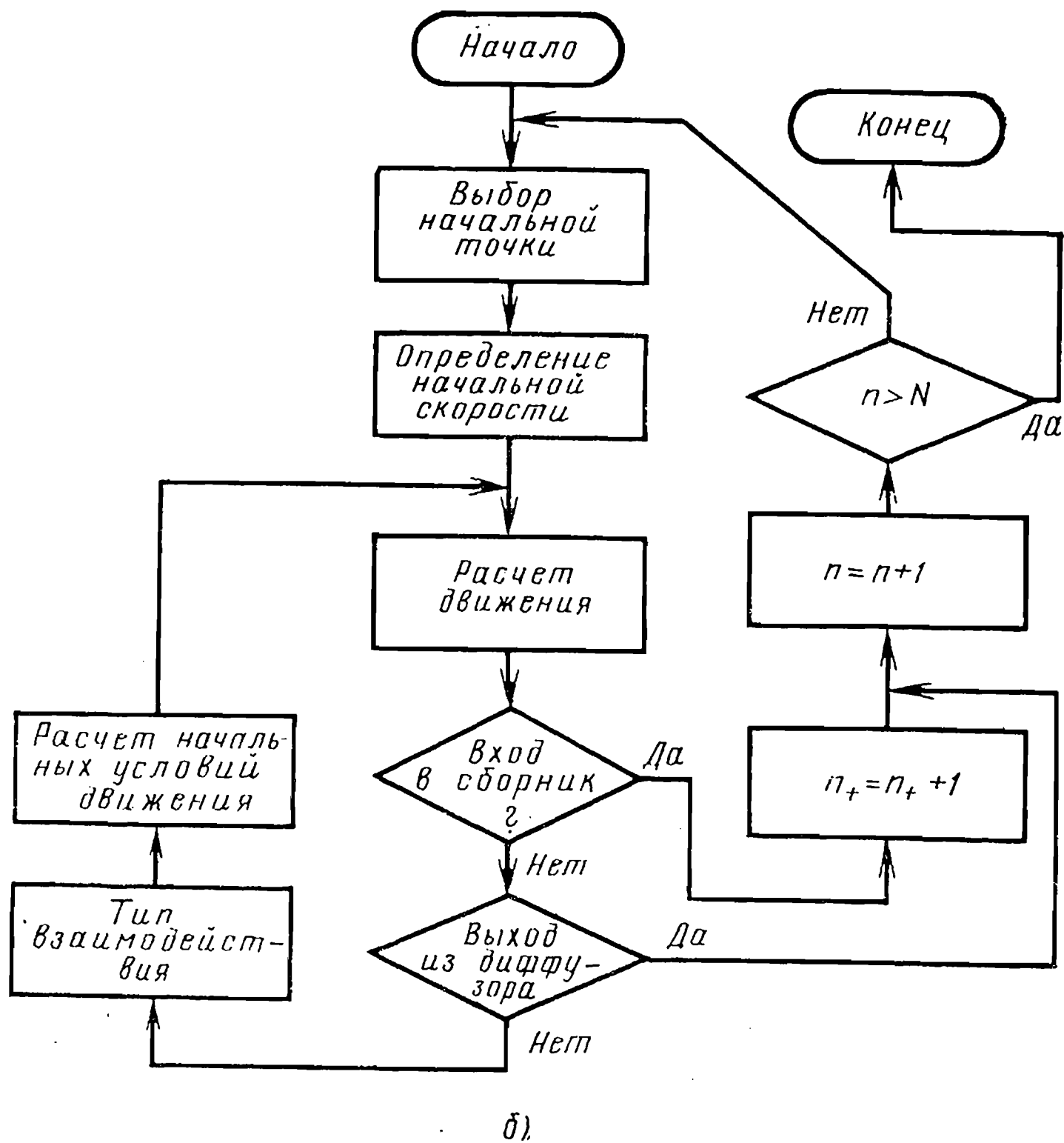
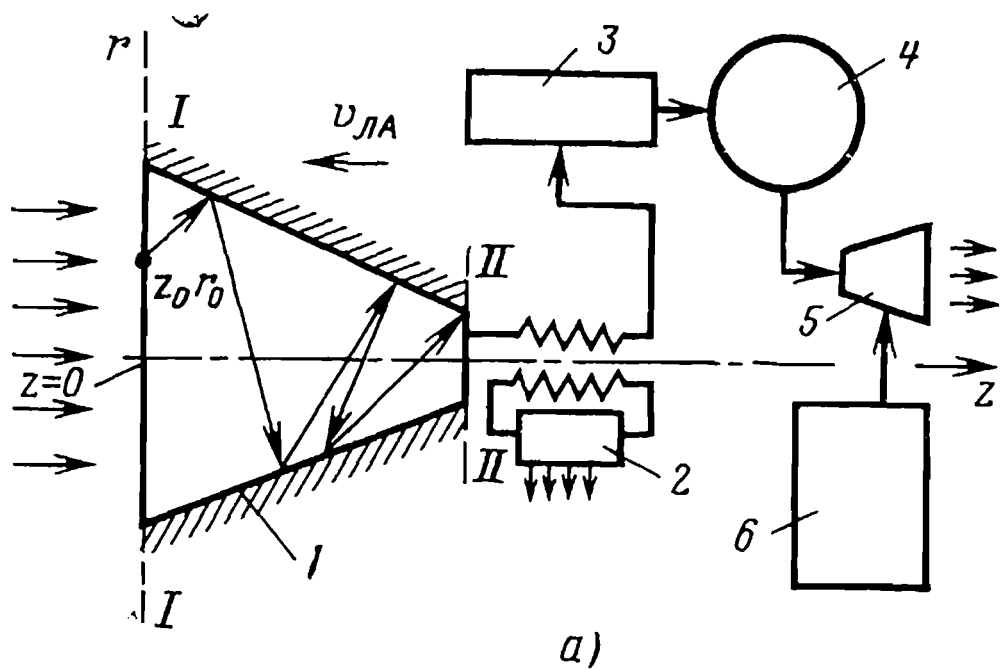


Рис. 6.3. Моделирование свободномолекулярного воздухозаборника:
 а — схема воздухозаборника; б — схема моделирования; 1 — воздухозаборник; 2 — радиатор;
 3 — сжижающая установка; 4 — бак рабочего тела; 5 — двигатель; 6 — энергоустановка

Движение молекул в пределах диффузора определяется их собственной тепловой скоростью (v_t) и скоростью ЛА ($v_{\text{ЛА}}$). Статистическая модель сбора рабочего вещества предусматривает последовательное изучение движения некоторого случайного набора молекул. Моделирование предусматривает ряд последовательных этапов (рис. 6.3, б).

1. Определяется («разыгрывается») начальное положение молекулы на срезе диффузора (сечение I—I): z_0, r_0, φ_0 . При выбран-

ной системе координат (рис. 6.3, а) $z_0=0$. При определении r_0 нужно учитывать увеличение площади сечения диффузора, а значит, и вероятности появления молекулы с увеличением r . Плотность вероятности $p(r)$ и накопленная вероятность $\mathcal{P}(<r)$ появления молекулы на радиусе r для диффузора круглого сечения имеют вид

$$p(r) = \frac{2\pi r dr}{\pi R^2} = \frac{2}{R^2} r dr;$$

$$\mathcal{P}(<r) = \frac{2}{R^2} \int_0^r t dt = \frac{r^2}{R^2}. \quad (6.16)$$

Следовательно, в силу (6.10) величина r_0 связана с величиной равномерно распределенного на отрезке $[0, 1]$ случайного числа ξ_p соотношением

$$r_0 = R \sqrt{\xi_p^1}; \quad 0 \leq \xi_p^1 \leq 1. \quad (6.17)$$

Наконец, все возможные значения φ_0 равновероятны и φ_0 определяется соотношением (верхний индекс у ξ_p означает порядковый номер случайного числа)

$$\varphi_0 = 2\pi \xi_p^2; \quad 0 \leq \xi_p^2 \leq 1. \quad (6.18)$$

Таким образом, определение случайного начального положения молекулы требует двукратного обращения к датчику $\xi_p [0, 1]$.

2. Определяется («разыгрывается») начальное состояние молекулы, т. е. величина и направление ее тепловой скорости. Абсолютная величина тепловой скорости молекулы определяется процессами в окружающем ЛА безграничном пространстве и подчиняется распределению Максвелла:

$$p(v) = \frac{dn_v}{n} = 4\pi \left(\frac{m}{2\pi kT} \right)^{3/2} \exp \left(-\frac{mv^2}{2kT} \right) v^2 dv; \quad (6.19)$$

$$\mathcal{P}(<v) = 4\pi \left(\frac{m}{2\pi kT} \right)^{3/2} \int_0^v \exp \left(-\frac{mt^2}{2kT} \right) t^2 dt,$$

где n — общее число молекул; dn_v — число молекул, имеющих скорость в интервале $(v, v+dv)$; m — масса молекулы; k — постоянная Больцмана; T — абсолютная температура газа.

Согласно (6.10) случайная абсолютная скорость молекулы $v_T \xi$ определяется соотношением

$$4\pi \left(\frac{m}{2\pi kT} \right)^{3/2} \int_0^{v_T \xi} \exp \left(-\frac{mt^2}{2kT} \right) t^2 dt = \xi_p^3; \quad (6.20)$$

$$0 \leq \xi_p^3 \leq 1.$$

Уравнение (6.20) решается относительно $v_T \xi$ численными методами.

Составляющие скорости по координатам (v_z, v_r, v_φ) связаны соотношением

$$v_{T\xi} = \sqrt{v_{Tz\xi}^2 + v_{Tr\xi}^2 + v_{T\varphi\xi}^2} \quad (6.21)$$

и могут быть определены последовательно.

Для определения составляющей $v_{Tz\xi}$ необходимо двукратное обращение к датчику ξ_p $[0, 1]$, причем первое случайное число ξ_p^4 определяет абсолютное значение $v_{Tz\xi}$, а второе — знак (направление) $v_{Tz\xi}$. Очевидно, абсолютная величина $v_{Tz\xi}$ не может превышать абсолютную скорость молекулы $v_{T\xi}$. Тогда

$$v_{Tz\xi} = \begin{cases} v_{T\xi}\xi_p^4, & \xi_p^5 \leq 0,5; \\ -v_{T\xi}\xi_p^4, & \xi_p^5 > 0,5. \end{cases} \quad (6.22)$$

Аналогично определяется составляющая v_r , причем абсолютное значение $v_{Tr\xi}$ должно удовлетворять очевидному неравенству

$$v_{Tr\xi} \leq \sqrt{v_{T\xi}^2 - v_{Tz\xi}^2};$$

$$v_{Tr\xi} = \begin{cases} \xi_p^6 \sqrt{v_{T\xi}^2 - v_{Tz\xi}^2}, & \xi_p^7 \leq 0,5; \\ -\xi_p^6 \sqrt{v_{T\xi}^2 - v_{Tz\xi}^2}, & \xi_p^7 > 0,5. \end{cases} \quad (6.23)$$

Наконец, абсолютное значение $v_{T\varphi\xi}$ определяется из соотношения (6.21), а для определения направления снова используется датчик случайных чисел:

$$v_{T\varphi\xi} = \begin{cases} \sqrt{v_{T\xi}^2 - v_{Tz\xi}^2 - v_{Tr\xi}^2}, & \xi_p^8 \leq 0,5; \\ -\sqrt{v_{T\xi}^2 - v_{Tz\xi}^2 - v_{Tr\xi}^2}, & \xi_p^8 > 0,5. \end{cases} \quad (6.24)$$

3. Рассчитывается движение молекулы в диффузоре как сумма случайного теплового движения молекулы и заданной скорости ЛА. При этом по определению (рис. 6.3, а) — $v_{ЛА} \equiv v_{zЛА}$.

В силу сделанных предположений взаимодействие молекул в пределах диффузора можно не учитывать, движение происходит равномерно и прямолинейно и заканчивается одним из трех возможных исходов:

— молекула пересекает плоскость $II-II$. В этом случае изучение движения данной молекулы заканчивается, а в счетчики n_+ и n прибавляется по единице;

— молекула выходит за пределы диффузора (за сечение $I-I$). В этом случае изучение движения данной молекулы также заканчивается и добавляется единица в счетчик общего числа молекул n ;

— молекула попадает на стенку диффузора. В этом случае изучение движения молекулы должно быть продолжено в соответствии с п. 4.

4. Определяется («разыгрывается») характер взаимодействия молекулы со стенкой. Если вероятность упругого взаимодействия

равна p , а неупругого соответственно $1-p$, то после обращения к датчику $\xi_p [0, 1]$ взаимодействие признается упругим при $\xi_p^9 \leq p$, и неупругим — при $\xi_p^9 > p$.

При упругом взаимодействии абсолютная величина скорости сохраняется, а ее компоненты пересчитываются по известным правилам (угол падения равен углу отражения). При неупругом взаимодействии молекула как бы оседает на стенке, а затем испаряется с нее. Абсолютная величина скорости в этом случае определяется температурой стенки, а ее направление может быть вычислено с помощью датчика случайных чисел из условия равной вероятности всех возможных направлений в полупространстве, ограниченном стенкой.

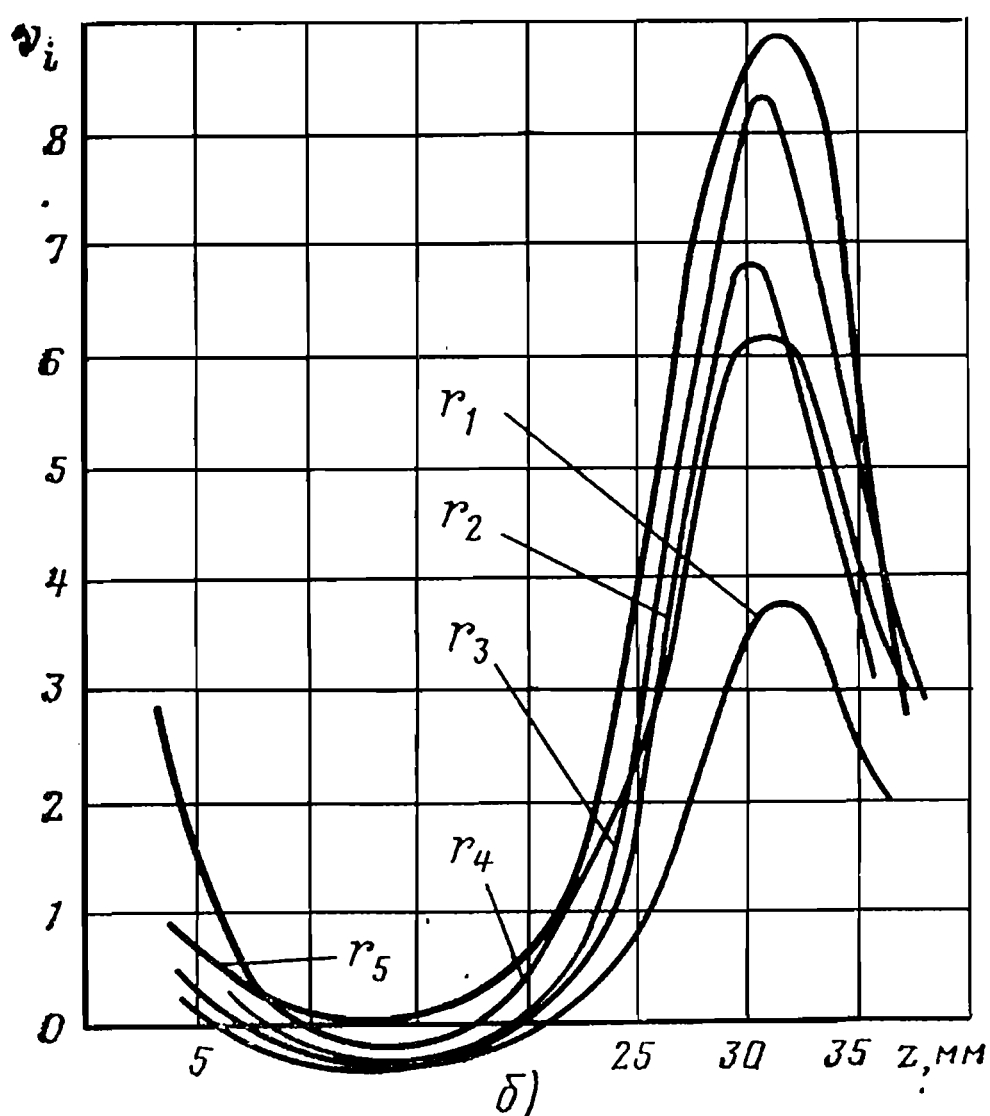
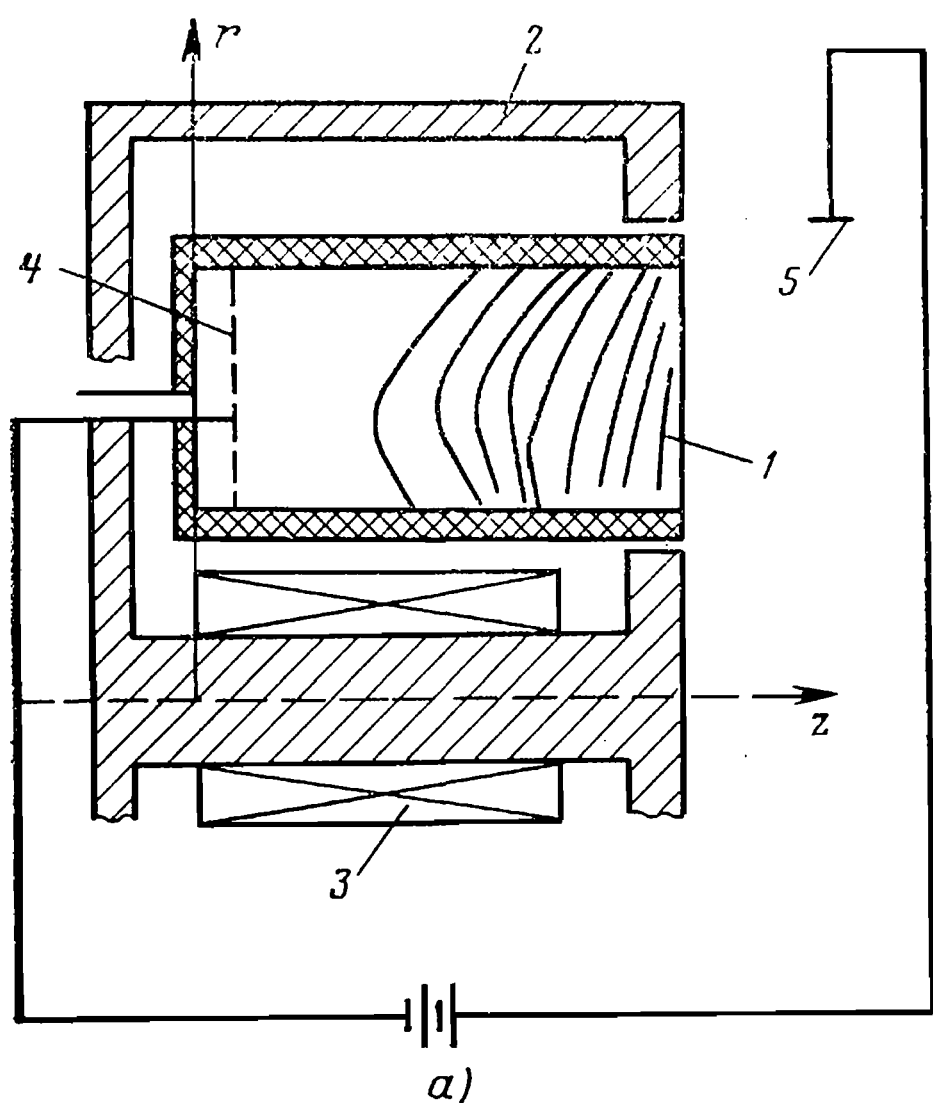
5. После определения новой скорости вновь рассчитывается движение молекулы до следующего соударения со стенкой и т. д. Процедура оканчивается после выхода молекулы за пределы диффузора (за сечения I—I или II—II).

Различные молекулы совершают разное число шагов, что предопределяет различное время движения (и счета) для различных реализаций. Однако при изучении стационарных процессов это несущественно, так как всегда можно говорить об изучении судьбы некоторой «секундной» порции частиц. Изучение достаточно большой выборки реализаций позволяет определить статистику, приближенно равную вероятности попадания молекулы в накопитель рабочего тела p_n , и общее количество собранного рабочего тела $M_{p.t.}$:

$$p_n \approx \frac{n_+}{n}; M_{p.t.} \approx p_n M t, \quad (6.25)$$

Рис. 6.4. Моделирование движения ионов в ускорителе с замкнутым дрейфом электронов

а — схема ускорителя; б — характеристики ионообразования в канале ускорителя; 1 — канал ускорителя; 2 — магнитопровод; 3 — катушка возбуждения магнита; 4 — анод; 5 — катод



где M — общее количество молекул, появляющихся на срезе диффузора в единицу времени.

Рассмотренная модель может быть использована, например, для оптимизации конфигурации диффузора.

Модель для изучения динамики ионной компоненты в канале ускорителя с замкнутым дрейфом электронов. Характер движения ионов в канале ускорителя (рис. 6.4, а) играет первостепенную роль при оценке рабочего процесса ускорителя, так как в значительной мере определяет его тяговые характеристики, энергетический баланс, возможности регулирования, а также темпы износа, ресурс и надежность стенок канала. Задача изучения динамики ионной компоненты распадается на две части: определение характеристик процесса ионообразования и расчет движения частиц и их характеристик на границах канала. Первая из этих задач требует исследования сложных закономерностей рабочего процесса, а ее решение опирается на результаты локальных измерений в канале. Вторая задача легче поддается формализации и приводит к описанной ниже модели¹.

Обработка результатов локальных зондовых измерений в канале ускорителя позволяет построить зависимости, характеризующие интенсивность ионообразования в различных точках канала $v_i(z, r)$.

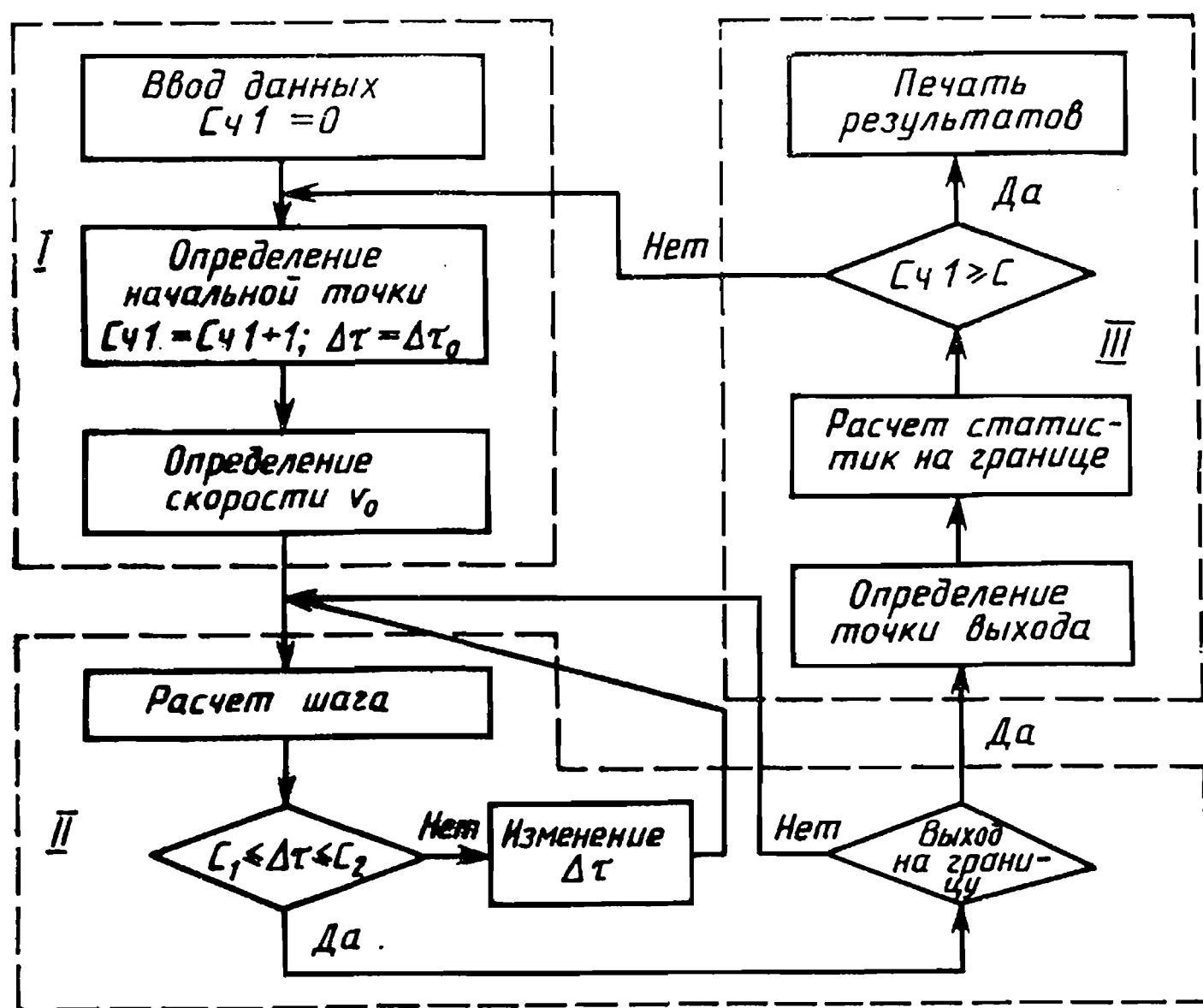


Рис. 6.5. Схема моделирования движения ионов в ускорителе с замкнутым дрейфом электронов

¹ Бишаев А. М., Ким В., Чуян Р. К. Исследование динамики ионов в канале УЗДП. — Материалы II Всесоюзной конференции по плазменным ускорителям, Минск, 1973, с. 83—84.

На рис. 6.4, б приведена серия таких кривых для одного из режимов работы ускорителя. Распределение потенциала в канале определяется в первую очередь магнитным полем и может быть рассчитано или определено экспериментально.

Модель состоит из трех больших блоков (рис. 6.5): блока определения случайной начальной точки и начальных параметров движения частицы (начальной случайной скорости v_0), блока расчета траектории движения частицы до границы рассматриваемой области и определения статистик, характеризующих модель. Из рис. 6.4, б видно, что функция ионообразования описывается весьма сложным двумерным законом распределения случайных величин, практически не поддающимся непосредственной интерпретации и алгоритмизации. В связи с этим координаты начальной точки траектории (z_0, r_0) определяются последовательно. Кривые $v_i = f(z)$ на рисунке представляют собой функции распределения случайных координат z при фиксированных радиусах $r = \text{const}$, которые естественно распространить на кольцевой слой $r - \frac{\Delta r}{2} \leq r \leq r + \frac{\Delta r}{2}$. Площадь под этими кривыми соответствует накопленной вероятности рождения иона в этом слое. Точнее, вероятность рождения иона $\mathcal{P}_j$ в слое $r_j \pm \frac{\Delta r}{2}$ определяется выражением

$$\mathcal{P}_j \left(\left| r - r_j \right| \leq \frac{\Delta r}{2} \right) = \frac{r_j \int_{z_{0j}}^{z_k} v_j(z, r_j) dz}{\sum_{i=1}^n r_i \int_{z_{0i}}^{z_k} v_i(z, r_i) dz} = \frac{r_j S_j}{\sum_{i=1}^n r_i S_i}, \quad (6.26)$$

где z_{0i} — координата, характеризующая начало зоны ионизации при $r = r_i$; S_i — площадь под i -й кривой (рис. 6.4, б).

Зависимость (6.26) характеризует распределение плотности вероятности $f(r)$ рождения иона на радиусе r . В частности, для функций ионообразования, представленных на рис. 6.4, б, график такого распределения, нормированного в соответствии с условием $f_{\max}(r) = 1$, показан на рис. 6.6, а. Случайную начальную координату можно определить с помощью простой процедуры, предусматривающей двукратное обращение к датчику $\xi_p [0, 1]$:

$$r_0 = \xi_p, \text{ если } \xi_p^2 \leq f(\xi_p^1). \quad (6.27)$$

Для найденного значения r_0 распределение вероятности случайного значения координаты z_0 описывается той из функций $v_i(r_i, z)$, в зоне действия которой лежит r_0 . Зависимость $v_i(r_i, z)$ может быть представлена как сумма двух распределений (рис. 6.6, б): треугольного и S_B -Джонсона с параметрами γ_i, η_i . Вес каждого из этих распределений задается соотношением площадей S_{1i}, S_{2i} под кривыми соответствующих распределений. При определении z_0 сначала с

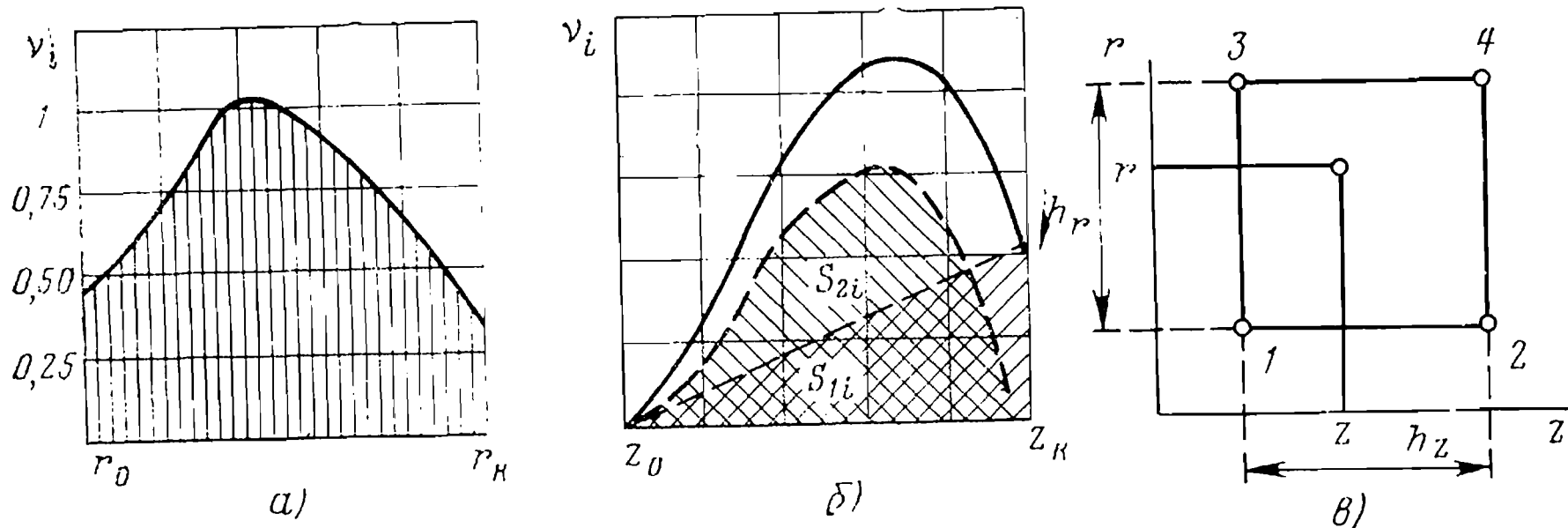


Рис. 6.6 Моделирование движения ионов в канале ускорителя с замкнутым дрейфом электронов:

a — зависимость интенсивности ионообразования от радиуса; *б* — выбор начального значения z_0 ; *б* — аппроксимация распределения потенциала

помощью датчика ξ_p $[0, 1]$ выбирается закон распределения, а затем и сама координата, причем может потребоваться обращение к датчику нормально распределенных случайных чисел ξ_N $(0, 1)$:

$$z_0 = \sqrt[4]{\xi_p^4} \text{ при } \xi_p^3 \leq \frac{S_{1i}}{S_{1i} + S_{2i}}; \quad (6.28)$$

$$z_0 = \frac{\exp[(\xi_N^1 - \gamma_i)/\eta_i]}{1 + \exp[(\xi_N^1 - \gamma_i)/\eta_i]} \text{ при}$$

$$\xi_p^3 > \frac{S_{1i}}{S_{1i} + S_{2i}}.$$

При нормировке стандартных распределений необходимо учитывать различие точек отрыва распределений z_{0i} для различных значений r_i . В момент образования ион обладает начальной (тепловой) скоростью. Величина и направление этой скорости также могут быть определены с помощью датчика случайных чисел: $(\xi_N$ $(0, 1)$ для величины скорости и ξ_p $[0, 1]$ — для направления). Однако расчеты показали, что в характерных для ускорителя условиях начальные параметры движения не оказывают существенного влияния на траекторию. В этом случае естественно предположить, что начальная скорость равна нулю, а движение ионов в силу азимутальной симметрии поля происходит исключительно в меридиональной плоскости (r, z) , т. е. двумерно.

Существует ряд методов расчета траектории заряженных частиц. Однако при использовании статистической модели, принципиально связанной с обработкой больших выборок, важнейшим требованием к методу является быстрота вычислений. Интегрирова-

ние уравнения движения с погрешностью $O(\Delta t)$ осуществляется по формулам

$$\dot{\mathbf{r}}_{n+1} = \dot{\mathbf{r}}_n + \Delta t_n B \nabla U \left(\mathbf{r}_n + \frac{\Delta t}{2} \dot{\mathbf{r}}_n \right); \quad (6.29)$$

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \frac{\Delta t_n}{2} (\dot{\mathbf{r}}_n + \dot{\mathbf{r}}_{n+1}).$$

где B — инерция частицы; Δt_n — n -й интервал времени интегрирования; ∇U — производная потенциала с погрешностью $O(h)$, аппроксимируемая выражением

$$\begin{aligned} \frac{\partial U}{\partial r} &\simeq \frac{1}{h_r} \left[U_2 - U_1 + \frac{1}{h_z} (U_4 - U_3 - U_2 + U_1)(r - r_i) \right]; \\ \frac{\partial U}{\partial z} &\simeq \frac{1}{h_z} \left[U_3 - U_1 + \frac{1}{h_r} (U_4 - U_3 - U_2 + U_1)(z - z_i) \right]. \end{aligned} \quad (6.30)$$

Индексы U в (6.30) соответствуют номерам точек на рис. 6.6, в. Использование более точных аппроксимационных формул для производной потенциала приводит к существенному увеличению времени расчета траектории.

Скорость иона меняется в процессе его движения, поэтому должна изменяться и величина шага интегрирования Δt_n . Для изменения Δt на каждом шаге интегрирования проверяется условие

$$0,1h \leq |\mathbf{r}_{n+1} - \mathbf{r}_n| \leq 0,4h; \quad h = \frac{h_z + h_r}{2}, \quad (6.31)$$

позволяющее корректировать величину Δt .

При достижении частицей границы рассматриваемой области определяется участок границы, на который вышла частица, ее энергия (скорость), направление движения (угол падения) и ускорение. Результаты расчета для всей совокупности траекторий представляются в виде статистик — массивов счетчиков, накапливающих данные о количестве, средних энергиях, углах падения и ускорениях частиц для каждого из участков границы. Введение двумерных массивов счетчиков позволяет получить распределения этих величин. Например, при расчете эрозии изолятора канала необходимо отдельно учитывать частицы с энергией ниже и выше порога распыливания материала.

На рис. 6.7 приведены некоторые результаты расчетов, полученные на основе обработки выборки из $3 \cdot 10^4$ случайных траекторий (реализаций). Рассмотренная модель позволяет определить распределение токов, величину и направление вектора тяги, а также сравнить между собой с точки зрения динамики ионной компоненты магнитные поля различных конфигураций. С помощью распределения частиц и их энергий на границах канала можно рассчитать процессы эрозии его стенок и ресурс ускорителя.

Иногда статистические модели используются для изучения детерминированных объектов и систем. В этом случае вводится ис-

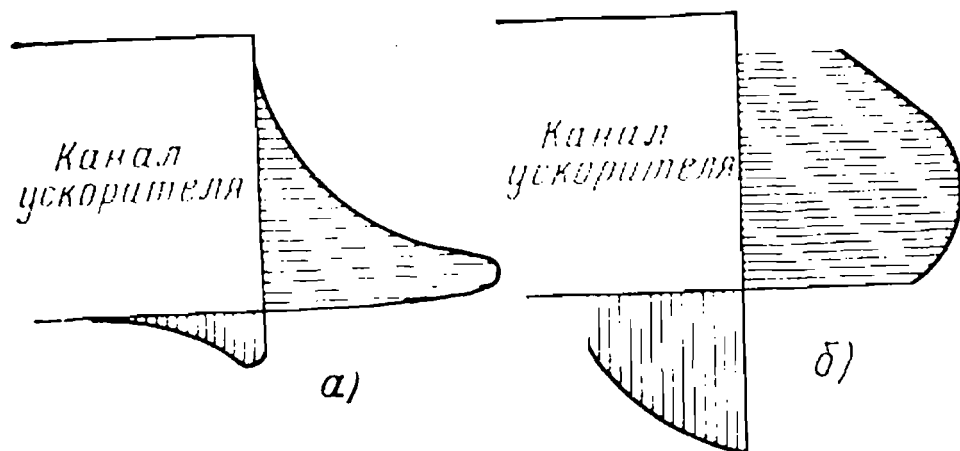


Рис. 6.7. Некоторые результаты моделирования движения ионов в канале ускорителя с замкнутым дрейфом электронов:

a — эпюры распределения ионов на границе канала; b — эпюры распределения скоростей ионов на границе канала

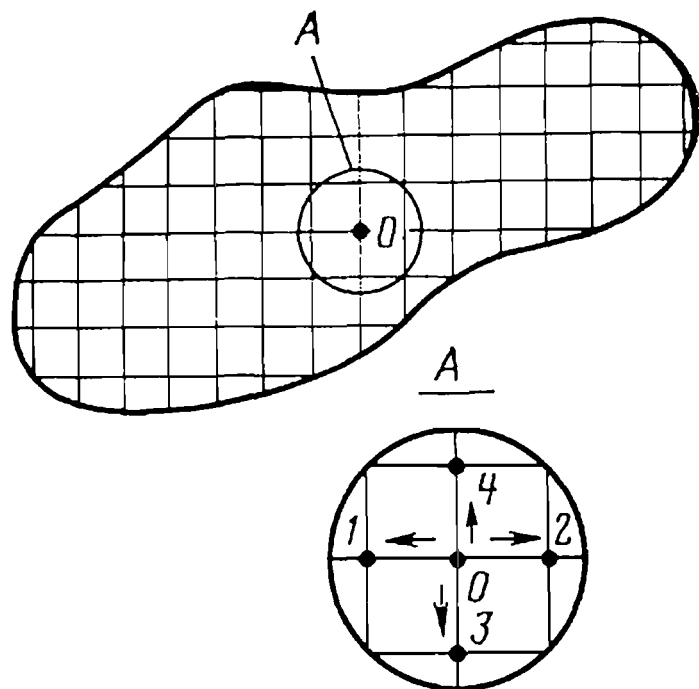


Рис. 6.8. Движение пробной частицы при моделировании поля методом Монте-Карло

кусственный случайный процесс с необходимыми характеристиками. Так, например, на основе анализа случайных ситуаций можно построить модель для исследования полей, описываемых уравнениями в частных производных (см. разд. 4.3).

В простейшем случае первая краевая задача (задача Дирихле) для уравнения Лапласа формулируется как задача определения функции $\Phi(\xi^r)$, удовлетворяющей в области D уравнению (4.21a) при граничных условиях (4.26). С помощью конечно-разностной аппроксимации эта задача сводится к решению системы алгебраических уравнений, при этом в силу (4.39) для внутренних узлов сетки (рис. 6.8) в двумерном случае выполняется соотношение

$$\Phi_0 = \frac{1}{4} [\Phi_1 + \Phi_2 + \Phi_3 + \Phi_4], \quad (6.32)$$

а значения функции Φ в граничных узлах заданы.

Еще в 1928 г. в работах Р. Куранта была установлена связь задачи решения системы разностных уравнений (6.32) с процессом случайного блуждания по конечно-разностной сетке. Классический процесс блуждания осуществляется следующим образом (рис. 6.8). Из некоторой внутренней точки сетки (i_0, j_0) последовательно выпускается n частиц. Частицы блуждают по сетке, причем в результате каждого шага блуждания частица должна переместиться в один из соседних узлов, а каждый из возможных переходов равновероятен. Реальный шаг определяется с помощью датчика $\xi_p [0, 1]$ по схеме

$$(ij) \rightarrow \begin{cases} i+1, j; & 0 \leq \xi_p < 0,25; \\ i-1, j; & 0,25 \leq \xi_p < 0,5; \\ i, j+1; & 0,5 \leq \xi_p < 0,75; \\ i, j-1; & 0,75 \leq \xi_p \leq 1. \end{cases} \quad (6.33)$$

Очевидно, в процессе блуждания частица может (в том числе и неоднократно) возвращаться в тот или иной узел сетки. Блуждание заканчивается при выходе частицы на границу, т. е. при попадании в один из граничных узлов $(i, j) \in \Gamma$. При этом запоминается значение функции в данном граничном узле $\Phi_k(Q) = \Phi(i, j) \in \Gamma$ и осуществляется переход к следующей частице. Очевидно, траектории частиц случайны и оканчиваются, вообще говоря, в различных граничных узлах. Можно показать, что при достаточно больших n функция Φ в исследуемой точке определяется соотношением

$$\Phi(i_0, j_0) \approx \frac{1}{n} \sum_{u=1}^n \Phi_{ku}(Q). \quad (6.34)$$

Действительно, вероятность $p(P_0, Q)$ того, что частица в результате случайного блуждания попадает из точки P_0 в граничный узел Q , представляет собой сумму вероятностей попадания частицы в Q из соседних с P_0 узлов P_1, P_2, P_3, P_4 (частица может попасть из P_0 в P_1 , а потом в Q , или из P_0 в P_2 , а потом в Q и т. д.):

$$p(P_0, Q) = \frac{1}{4} [p(P_1, Q) + p(P_2, Q) + p(P_3, Q) + p(P_4, Q)]. \quad (6.35)$$

Уравнения (6.35) эквивалентны (6.32). Кроме того, вероятность $p(P_0, Q)$ удовлетворяет граничным условиям

$$p(Q, Q) = 1; \quad p(Q, Q^1) = 0; \quad Q \neq Q^1. \quad (6.36)$$

Если m из n случайных траекторий заканчиваются в точке Q , вероятность попадания из P_0 в Q определяется статистикой

$$p(P_0, Q) \simeq \frac{m}{n}, \quad (6.37)$$

а статистика в правой части (6.34) представляет собой математическое ожидание значения функции в граничном узле, где оканчивается случайная траектория $\xi(P)$:

$$\Phi(i_0, j_0) = M\xi(P) = \sum_{t=1}^N \Phi_t(Q) p_t(P_0, Q), \quad (6.38)$$

где N — общее число граничных узлов области D . Очевидно, что (6.38) эквивалентно (6.34).

Величина $\Phi(i_0, j_0)$ в силу (6.35) удовлетворяет системе конечно-разностных уравнений (6.32) и принимает на границах заданные граничные значения, а следовательно, является решением задачи Дирихле для уравнения Лапласа.

Для исследования полного поля значений функции $\Phi(P)$ необходимо последовательно рассмотреть все внутренние узлы сетки.

Метод случайных блужданий может быть легко обобщен на ряд более сложных задач. Так, например, для полей с источниками, описываемых уравнением Пуассона (4.22а), система конечно-разностных уравнений (4.41) в двумерном случае может быть записана в виде

$$\Phi(P) = \frac{1}{4} [\Phi(P_1) + \Phi(P_2) + \Phi(P_3) + \Phi(P_4) + h^2 f(P)], \quad (6.39)$$

где h — шаг конечно-разностной сетки.

Величина $h^2 f(P)$ представляет собой мощность источников в элементарном объеме, связанном с узлом P . Для трехмерного случая, очевидно, должно быть $h^3 f(P)$. В случае неравномерной сетки ($h_1 \neq h_2$) $h^2 f(P) \rightarrow h_1 h_2 f(P)$.

При случайном блуждании запоминается не просто значение функции в конечной точке (граничном узле) случайной траектории, а величина $\Phi_k(Q) + m \left(\frac{1}{4} h^2 f(P) \right)$, где m — количество шагов, совершенных в процессе блуждания. Значение функции в исследуемой точке определится соотношением:

$$\Phi(i_0, j_0) = \frac{1}{n} \sum_{u=1}^n \left[\Phi_{ku}(Q) + \frac{1}{4} m_u h^2 f(P) \right]. \quad (6.40)$$

При исследовании краевых задач с другими типами граничных условий изменяется поведение частицы в граничном узле. Например, при задании на границе условия $\left. \frac{\partial \Phi}{\partial n} \right|_{\Gamma} = 0$ движение частицы в граничном узле не заканчивается, а происходит «отражение», т. е. частица отскакивает внутрь области на один шаг. При задании граничных условий 3-го рода (конвективный теплообмен) необходимо учитывать вероятность p остановки процесса блуждания и вероятность $1-p$ отражения частицы и т. д.

Необходимость последовательного изучения всех внутренних узлов области D при достаточно большом числе испытаний для каждого узла приводит к большим объемам вычислений, что, конечно, ограничивает эффективность метода. Однако часто нет необходимости в определении функции $\Phi(P)$ во всей области, а достаточно найти ее значение лишь в некоторых критических точках. Если же все-таки необходимо провести полное исследование области D , могут быть использованы различные приемы ускорения процесса. В частности, все уже исследованные узлы могут быть отнесены к категории граничных, что существенно сокращает время последующих блужданий.

Одним из вариантов случайного блуждания является так называемый метод массового блуждания (ММБ), предусматривающий одновременный запуск из изучаемой точки P большого (например, 10^6) количества частиц и распределение их на каждом шаге между узлами сетки в соответствии с заданными вероятностями. При этом отпадает необходимость обращения на каждом шаге к датчику случайных чисел, однако существенно возрастают объемы используемой машинной памяти, так как необходимо одновременно «помнить» количество частиц во всех узлах сетки. Процесс блуждания заканчивается, когда подавляющее количество частиц (например, 99,9%) закончит блуждание, т. е. выйдет в граничные узлы сетки.

Разработан также ряд эффективных методов блуждания с переменным шагом, позволяющих сократить время расчета за счет использования больших шагов на первых этапах вдали от границы («блуждание по сферам»).

Интересно отметить, что статистические модели оказываются наиболее эффективными при решении детерминированных задач большой размерности, когда методы линейной алгебры приводят к очень большим объемам вычислений и требуют большой машинной памяти.

В частности, рассмотренные выше статистические модели для решения краевых задач успешно используются при изучении тепловых полей в элементах ДЛА при сложной конфигурации изучаемых областей и нелинейных граничных условиях, а также при изучении электростатических полей в ускорителях ионных двигателей и других электрофизических установок.

Вопросы для самоконтроля

1. Каковы особенности статистических моделей, предназначенных для изучения движения частиц?
2. Каковы особенности метода статистических испытаний, используемого для решения краевых задач?
3. Сформулируйте основные принципы и последовательность действий при случайном блуждании.

ЧАСТЬ III

АНАЛОГОВЫЕ МОДЕЛИ

И АНАЛОГОВОЕ МОДЕЛИРОВАНИЕ ЭЛЕМЕНТОВ ДВИГАТЕЛЕЙ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

Аналоговые модели и методы аналогового моделирования занимают особое место в системе математических моделей ДЛА. По сути дела, обращаясь к аналоговым моделям, сталкиваемся не с новыми типами математических описаний, а с новыми методами исследования уже известных математических моделей и, в первую очередь, обыкновенных дифференциальных уравнений и дифференциальных уравнений в частных производных. В рамках САПР аналоговые модели представляют собой эффективное средство изучения элементов ДЛА. Изучение методов аналогового моделирования трудно отделить от рассмотрения конкретных моделирующих устройств — аналоговых вычислительных машин (АВМ), поэтому в данном разделе будут затронуты некоторые вопросы, связанные с созданием таких машин, а от читателя потребуются знания основ электротехники.

ГЛАВА 7. ОБЩИЕ ПРИНЦИПЫ АНАЛОГОВОГО МОДЕЛИРОВАНИЯ

Для любого физического объекта возможно два пути исследования (рис. 7.1): во-первых, можно поставить эксперимент и непосредственно измерить характеристики объекта; во-вторых, можно составить математическую модель (математическое описание) объекта и рассчитать его характеристики. Если при этом не допущено каких-либо ошибок, результаты экспериментального и численного исследования объекта должны совпадать между собой. Предположим, что существует два объекта (или группа объектов), у которых полностью совпадают математические модели (рис. 7.1). Как будет показано далее, такие объекты встречаются в природе довольно часто. Очевидно, в результате расчета для этих объектов будут получены одинаковые численные характеристики: математика игнорирует физическую суть явлений и оперирует лишь с математическими зависимостями, которые для наших объектов одинаковы. Это значит, что совпадут и экспериментальные характеристики таких объектов (ведь они должны совпадать с численными характеристиками!). В результате в правой части схемы рисунка возникает сплошная вертикальная цепочка равенств. *Объекты с совпада-*



Рис. 7.1 Принцип аналогии

ющими математическими моделями называются аналогичными объектами.

Идея аналогового моделирования на первый взгляд (только на первый!) кажется несколько парадоксальной. После того как аналогия (т. е. тождественность математических моделей) объектов доказана, предлагается экспериментально изучить один из них, а потом интерпретировать полученные зависимости как характеристики второго объекта (в силу аналогии они должны совпадать). При этом для второго (натурного) объекта такое исследование по существу является вычислением, так как общность объектов основана лишь на общности их математических моделей и, следовательно, первый (модельный) объект является по сути дела своеобразной вычислительной машиной для исследования этой модели. Разумеется, аналоговое моделирование имеет практический смысл лишь в том случае, когда модельный объект проще натурного.

Аналоговое моделирование применимо к относительно простым объектам (системам), так как трудно представить себе сложные системы различной физической природы с одинаковыми математическими моделями. Поэтому аналоговые модели используются для подсистем, лежащих на нижних уровнях иерархического описания ДЛА (см. табл. 2.1). Практически при изучении ДЛА аналоговое моделирование используется для изучения объектов двух классов: *физических полей* и *физических систем*.

Физические поля описывают непрерывное распределение скалярных или векторных величин в пространстве. Математические модели физических полей сводятся к дифференциальным уравнениям в частных производных (см. разд. 4.3). Физические системы характеризуются конечным числом величин, так как включают в себя конечное число элементов с конечным числом возможных состояний. Математические модели физических систем описываются

обыкновенными дифференциальными уравнениями, причем в качестве аргумента обычно выступает время (см. разд. 4.2). Примером физических систем являются механические системы и системы автоматического регулирования.

Возможны два подхода к построению аналоговых моделей. Можно рассмотреть математические модели изучаемого и модельного объекта в целом и показать тождественность этих моделей. Можно поступить иначе: разбить математическую модель изучаемого объекта на блоки, построить аналоговые модели этих блоков, а потом из них, как из деталей детского конструктора, собрать общую аналоговую модель объекта.

Аналоговые модели, построенные для изучаемого объекта в целом, называют *моделями-аналогами*, модели с блочной структурой — *структурными моделями*. Модели-аналоги используются главным образом для изучения физических полей. Структурные модели получили преимущественное распространение при изучении физических систем. Важно подчеркнуть, что разница между моделями-аналогами и структурными моделями носит методологический характер и существенна лишь на этапе построения модели. Когда модель уже построена, она всегда работает как модель-аналог.

Построение аналоговой модели всегда включает в себя следующие этапы:

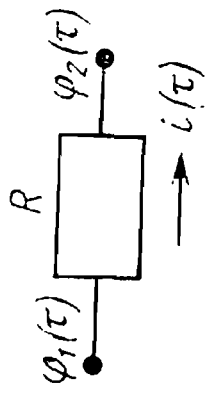
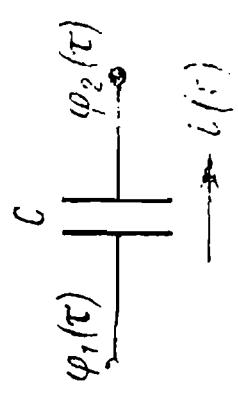
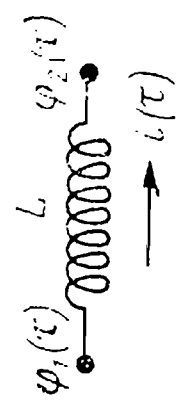
- 1) изучение математического описания объекта и выбор аналога, обеспечивающего наиболее простое его моделирование;
- 2) построение на основе выбранного аналога конкретной модели, задание краевых условий, доказательство аналогии;
- 3) экспериментальное определение характеристик модели;
- 4) интерпретацию полученных результатов как характеристик исследуемого объекта.

При доказательстве аналогии объектов вводятся понятия *сходственных величин, коэффициентов аналогии, индикаторов аналогии*. Аппарат аналогии представляет собой обобщение аппарата теории подобия. Конкретные примеры аналоговых моделей и анализ аналогий для физических полей и физических систем приведены в гл. 8 и 9.

Аналогичные объекты полностью равноправны между собой и любой из них может быть выбран в качестве модельного. На практике наиболее распространенным (но не единственным!) методом аналогового моделирования является *электрическое моделирование*, основанное на изучении электрических полей или цепей. Широкое распространение электрических моделей объясняется их относительной простотой, универсальностью и достаточно высокой точностью, связанной, в частности, с высокой точностью электрических измерений. Вместе с тем в ряде случаев для аналогового моделирования создаются и используются неэлектрические модели (см. разд. 8.5).

Некоторые характеристики электрических элементов и методы их описания представлены в табл. 7.1. В таблице приведены схемы

Таблица 7.1

Элемент	$i(\tau)$	$\Delta\varphi(\tau) = \varphi_1(\tau) - \varphi_2(\tau)$	$I(p)$	$\Delta\Phi(p)$	$Z(p)$
	$i(\tau) = \frac{\varphi_1(\tau) - \varphi_2(\tau)}{R}$	$\Delta\varphi(\tau) = Ri(\tau)$	$I(p) = \frac{1}{R} \Delta\Phi(p)$	$\Delta\Phi(p) = RI(p)$	$Z_R(p) = R$
	$i(\tau) = C \frac{\partial (\varphi_1(\tau) - \varphi_2(\tau))}{\partial \tau}$	$\Delta\varphi(\tau) = \frac{1}{C} \int_0^\tau i(\tau) d\tau + \Delta\varphi_0$	$I(p) = Cp\Delta\Phi(p)$	$\Delta\Phi(p) = \frac{1}{Cp} I(p)$	$Z_C(p) = \frac{1}{Cp}$
	$i(\tau) = \frac{1}{L} \int_0^\tau \Delta\varphi(\tau) d\tau + i(0)$	$\Delta\varphi(\tau) = L \frac{\partial i(\tau)}{\partial \tau}$	$I(p) = \frac{1}{Lp} \Delta\Phi(p)$	$\Delta\Phi(p) = LpI(p)$	$Z_L(p) = Lp$

электрических элементов, выражения для тока и изменения напряжения в этих элементах, записанные в оригиналах, и их «изображения» по Лапласу, используемые для описания электрических цепей методами операционного исчисления.

Операционное исчисление позволяет «понизить» порядок математических операций, в частности, заменить операции дифференцирования и интегрирования операциями умножения и деления на символический оператор p :

$$\begin{aligned} \frac{d\varphi}{dt} \Rightarrow p\Phi; \quad \frac{d^n \varphi}{dt^n} \Rightarrow p^n \Phi; \quad \int_0^t \varphi dt \Rightarrow \frac{\Phi}{p}; \\ \underbrace{\int_0^t \int_0^t \varphi dt}_{n} \Rightarrow \frac{\Phi}{p^n}. \end{aligned} \quad (7.1)$$

При этом осуществляется переход от реальных функций «оригиналов» ($\varphi(t)$, $i(t)$) к их «изображениям» ($\Phi(p)$, $I(p)$), проводятся необходимые операции над изображениями, а затем совершается обратный переход от изображения-результата к результату-оригиналу. Операционное исчисление позволяет существенно упростить вычисления при расчетах электрических цепей.

Переход к изображениям приводит к *однотипным* математическим выражениям. Если ввести обобщенные операторные сопротивления $Z(p)$ — импедансы (табл. 7.1), эти выражения можно записать в виде операторного закона Ома:

$$\Delta\Phi(p) = Z(p) I(p); \quad I(p) = \frac{\Delta\Phi(p)}{Z(p)}. \quad (7.2)$$

С операторными сопротивлениями $Z(p)$ при расчетах можно обращаться как с обычными активными сопротивлениями R . В частности, при последовательном соединении элементов суммарное сопротивление $Z_{\Sigma}(p)$ имеет вид

$$Z_{\Sigma}(p) = Z_1(p) + Z_2(p) + \dots + Z_n(p), \quad (7.3)$$

а для параллельного соединения можно записать:

$$\frac{1}{Z_{\Sigma}(p)} = \frac{1}{Z_1(p)} + \frac{1}{Z_2(p)} + \dots + \frac{1}{Z_n(p)}. \quad (7.4)$$

Соотношения, представленные в табл. 7.1, используются ниже при описании аналоговых моделей.

Вопросы для самоконтроля

1. Сформулируйте принцип аналогии.
2. Для каких объектов преимущественно используются методы аналогового моделирования?
3. В чем заключаются преимущества электрического моделирования?

4. Что такое модели-аналоги и структурные аналоговые модели?
5. Дайте характеристику электрических элементов R , C , L . Запишите зависимости для $i(t)$ и $\Delta\varphi(t)$, а также изображений $I(p)$ и $\Phi(p)$.
6. Что такое операторный закон Ома?
7. Перечислите основные этапы аналогового моделирования.

ГЛАВА 8. АНАЛОГОВОЕ МОДЕЛИРОВАНИЕ ФИЗИЧЕСКИХ ПОЛЕЙ В ЭЛЕМЕНТАХ ДВИГАТЕЛЕЙ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

8.1. ОБЩИЕ ПРИНЦИПЫ МОДЕЛИРОВАНИЯ ПОЛЕЙ

При аналоговом моделировании физических полей, как правило, используются модели-аналоги, предусматривающие моделирование изучаемого поля в целом. Аппарат аналогии можно рассмотреть на примере любых физических полей. Однако наиболее просто это сделать на примере потенциальных стационарных полей, описывающих установившееся состояние объекта.

Стационарные физические поля описываются однородными или неоднородными эллиптическими дифференциальными уравнениями в частных производных (4.21), (4.22):

$$\nabla(\omega(\xi^r, \Phi) \nabla \Phi) = f(\xi^r, \Phi). \quad (8.1)$$

Эти уравнения носят весьма общий характер и «работают» для полей различной физической природы. Некоторые из этих полей с указанием конкретных физических характеристик приведены в табл. 8.1. Общность математической формы описания различных физических полей является следствием общности фундаментальных законов сохранения материи и энергии и служит блестящей иллюстрацией философского принципа единства материального мира. В соответствии с общим принципом аналогового моделирования *одинаковость* уравнений, приведенных в табл. 8.1, может служить основанием для построения аналоговых моделей.

Каждое из уравнений, приведенных в табл. 8.1, включает в себя четыре величины, занимающие в этих уравнениях одинаковые позиции. Это потенциальная функция (T , U , U_r , U_m , φ ...), характеристика среды (λ , ε , ρ , μ , σ ...), мощность распределенных источников поля (ω , q , q_r , i ...) и, наконец, геометрический масштаб модели (L , l). Величины, занимающие в уравнениях одинаковые позиции, называются *сходственными* величинами, а для характеристики их отношений вводятся *коэффициенты аналогии*:

$$M_\varphi = \frac{T}{\varphi}, \frac{U}{\varphi}, \frac{U_r}{\varphi} \dots; \quad M_\sigma = \frac{\lambda}{\sigma}, \frac{\rho}{\sigma}, \frac{\varepsilon}{\sigma} \dots; \quad (8.2)$$

$$M_i = \frac{\omega}{i}, \frac{q}{i}, \frac{q_r}{i} \dots; \quad M_l = \frac{L}{l}.$$

Область физики	Уравнение (закон)	Форма уравнения	Условные обозначения
Теплопередача (поле температур в твердом теле)	Уравнение теплопроводности	$\nabla (\lambda (\xi^r, T) \nabla T) =$ $= w (\xi^r, T)$ (8.1a)	ξ^r — координаты T — температура, К λ — теплопроводность w — интенсивность объемного тепловыделения
Гидродинамика (течение идеальной жидкости)	Закон Бернулли	$\nabla (\rho (\xi^r, U_r) \times$ $\times \nabla U_r) =$ $= q (\xi^r, U_r)$ (8.1б)	ξ^r — координаты U_r — гидродинамический потенциал скоростей ρ — плотность среды q — мощность распределенного источника жидкости
Электростатика (электростатическое поле)	Теорема Гаусса	$\nabla (\epsilon (\xi^r, U) \times$ $\times \nabla U) =$ $= q (\xi^r, U)$ (8.1в)	ξ^r — координаты U — электростатический потенциал ϵ — диэлектрическая постоянная среды q — плотность распределения объемных зарядов
Магнитостатика (магнитное поле)	Закон Фарадея	$\nabla (\mu (\xi^r, U_m) \times$ $\times \nabla U_m) = 0$ (8.1г)	ξ^r — координаты U_m — магнитный скалярный потенциал μ — магнитная проницаемость среды
Фильтрация жидкости в пористом теле	Закон Дарси	$\nabla (\chi (\xi^r, H) \times$ $\times \nabla H) =$ $= q (\xi^r, H)$ (8.1д)	ξ^r — координаты H — напор χ — постоянная фильтрации q — мощность распределенных источников жидкости
Распределение нейтронов в делящемся материале	Закон Фика	$\nabla (\delta (\xi^r, F) \times$ $\times \nabla F) =$ $= \omega (\xi^r, F)$ (8.1е)	ξ^r — координаты F — поток нейтронов δ — характеристика поглощения нейтронов средой ω — мощность выделения нейтронов в объеме

Область физики	Уравнение (закон)	Форма уравнения	Условные обозначения
Гравитационное поле в пространстве	Закон Ньютона	$\nabla (\chi^{-1} (\xi^r, U_r) \times \times \nabla U_r) = = m (\xi^r, U_r) \quad (8.1\text{ж})$	ξ^r — координаты U_r — гравитационный потенциал χ — гравитационная постоянная m — распределенная в пространстве масса
Электрическое поле в проводниках	Закон Кирхгофа	$\nabla (\sigma (\xi^r, \varphi) \nabla \varphi) = = i (\xi^r, \varphi) \quad (8.1\text{з})$	ξ^r — координаты φ — потенциал электрического поля σ — электропроводность среды i — распределенные токи, вводимые в среду

С помощью коэффициентов аналогии можно выразить и заменить величины изучаемого (например, температурного) поля величинами модельного (электрического) поля:

$$T = M_\varphi \varphi; \quad \omega = M_i i; \quad L = M_l l; \quad \lambda = M_\sigma \sigma;$$
$$\frac{\partial T}{\partial X} = \frac{M_\varphi}{M_l} \frac{\partial \varphi}{\partial x}; \quad \frac{\partial T}{\partial Y} = \frac{M_\varphi}{M_l} \frac{\partial \varphi}{\partial y}; \quad \frac{\partial T}{\partial Z} = \frac{M_\varphi}{M_l} \frac{\partial \varphi}{\partial z}; \quad (8.3)$$
$$\nabla T = \frac{M_\varphi}{M_l} \nabla \varphi.$$

Тогда уравнение для стационарного температурного поля (8.1а) (табл. 8.1) можно записать в виде

$$\frac{M_\varphi M_\sigma}{M_i M_l^2} \nabla (\sigma (\xi^r, \varphi) \nabla \varphi) = i (\xi^r, \varphi). \quad (8.4)$$

Важно подчеркнуть, что уравнение (8.4) по-прежнему остается *уравнением теплопроводности*: ведь изменилась только форма записи тепловых величин (T, λ, ω, L). Вместе с тем уравнение теплопроводности в форме (8.4) полностью совпадает с уравнением электрического поля (8.1, з) (табл. 8.1) при условии

$$\frac{M_\varphi M_\sigma}{M_i M_l^2} \equiv 1. \quad (8.5)$$

Соотношение (8.5) выражает условие аналогии натурального и модельного (в рассматриваемом примере температурного и электрического) полей и носит название *индикатора аналогии*.

При практических расчетах неудобно пользоваться величиной точечной мощности распределенных источников поля $w(\xi^r, T)$, $i(\xi^r, \varphi)$. Целесообразно заменить эту величину некоторой сосредоточенной величиной (W, I), характеризующей мощность источников (тепло- и токовыведение) в некотором элементарном объеме (V, v):

$$W = wV, \quad I = iv. \quad (8.3a)$$

Отношение величин W и I характеризуется коэффициентом аналогии M_I :

$$M_I = \frac{W}{I} = \frac{wV}{iv} = M_i^3 M_i; \quad M_i = \frac{M_I}{M_i^3}, \quad (8.6)$$

а индикатор аналогии (8.5) записывается в виде

$$\frac{M_\varphi M_I M_\sigma}{M_I} \equiv 1. \quad (8.5a)$$

Соотношения (8.5) и (8.5a), в частности, показывают, что три коэффициента аналогии из четырех могут быть выбраны произвольно, а четвертый должен быть определен из этих соотношений. В рассмотренном случае имеется один индикатор аналогии (одно соотношение, связывающее коэффициенты аналогии). Вообще же количество индикаторов аналогии в силу известной π -теоремы Кирпичева на единицу меньше числа членов изучаемого уравнения.

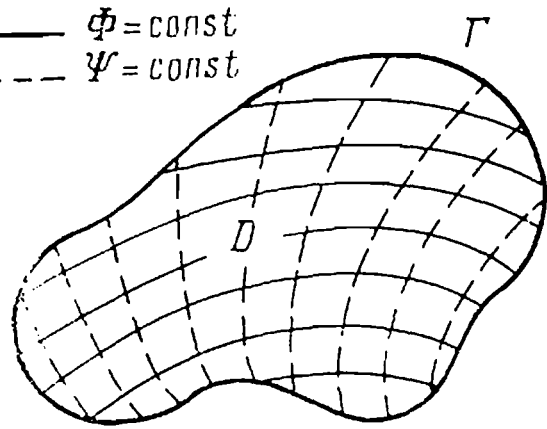
Для однородного эллиптического уравнения коэффициенты аналогии (8.4) могут быть просто сокращены. Это значит, что для полей без распределенных источников аналогия соблюдается *автоматически*, а коэффициенты аналогии не связаны между собой. В этом случае достаточно просто выбрать эти коэффициенты.

Аппарат аналогии является непосредственным обобщением аппарата теории подобия. Действительно, теория подобия также оперирует понятиями сходственных величин, коэффициентов (констант), индикаторов и критериев подобия. Однако в отличие от метода аналогии метод подобия всегда рассматривает объекты одной и той же физической природы. Поэтому коэффициенты подобия всегда *безразмерны*. Коэффициенты аналогии, наоборот, как правило, характеризуют отношение величин различной физической природы и имеют ту или иную, порой достаточно сложную, размерность.

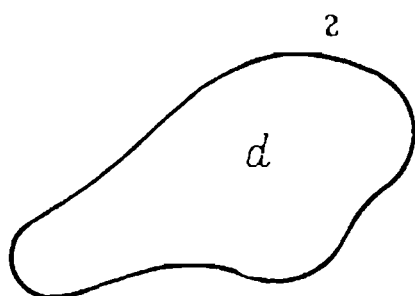
Для получения однозначного решения задачи кроме уравнений должны быть заданы граничные условия (4.26), (4.27), (4.28). При аналоговом моделировании граничные условия преобразовываются с помощью коэффициентов аналогии. При этом сама область исследования также изменяется в соответствии с масштабом моделирования (рис. 8.1):

$$d = \frac{1}{M_I^2} D; \quad r = \frac{1}{M_I} \Gamma; \quad v = \frac{1}{M_I^3} V, \quad (8.7)$$

— $\Phi = \text{const}$
 --- $\Psi = \text{const}$



a)



б)

Рис. 8.1. Задание модельной области при изучении поля:

a — натурная область; б — модельная область

что приводит к пропорциональному сжатию или расширению соответствующих участков границы. Поэтому граничные условия, заданные на некотором участке границы Γ , должны быть отнесены к соответствующему участку границы модели γ ($\gamma = \frac{1}{M_l} \Gamma$). Так, граничные условия первого рода (4.26) для исследуемой функции должны быть заменены граничными условиями первого рода для модельной функции:

$$\Phi|_{\Gamma} = f_1(\Gamma) \Rightarrow \varphi|_{\gamma} = \frac{1}{M_\varphi} f_1\left(\frac{\Gamma}{M_l}\right). \quad (8.8)$$

При этом на модели в каждой точке границы должен быть задан электрический потенциал φ . Точно также граничные условия второго рода задаются соотношением

$$\frac{\partial \Phi}{\partial n} \Big|_{\Gamma} = f_2(\Gamma) \Rightarrow \frac{\partial \varphi}{\partial n} \Big|_{\gamma} = \frac{M_l}{M_\varphi} f_2\left(\frac{\Gamma}{M_l}\right). \quad (8.9)$$

На модели граничные условия второго рода могут быть заданы с помощью токов, подведенных к заданным участкам границы. В силу очевидных соотношений

$$j = -\sigma \frac{\partial \varphi}{\partial n}; \quad I = j \delta_\gamma = -\sigma \delta_\gamma \frac{\partial \varphi}{\partial n} \quad (8.10)$$

(j — плотность тока, I — ток) эти токи определяются выражением

$$I = -\sigma \delta_\gamma \frac{M_l}{M_\varphi} f_2\left(\frac{\Gamma}{M_l}\right), \quad (8.11)$$

где знак учитывает направление тока.

Наконец, граничные условия третьего рода могут быть реализованы при подключении к границе заданных напряжений через заданные сопротивления. В общем виде эти условия могут быть записаны так:

$$\left(a\Phi + b \frac{\partial \Phi}{\partial n}\right) \Big|_{\Gamma} = f_3(\Gamma) \Rightarrow \left(a^*\varphi + b^* \frac{\partial \varphi}{\partial n}\right) \Big|_{\gamma} = f_3\left(\frac{\Gamma}{M_l}\right); \quad (8.12)$$

$$a^* = M_\varphi a; \quad b^* = \frac{M_\varphi}{M_l} b.$$

Условие (8.12) в силу (8.10) может быть записано в виде

$$\left(\varphi - \frac{b}{aM_l \sigma \delta \Gamma} I \right) \Big|_{\Gamma = \frac{\Gamma}{M_l}} = \frac{1}{aM_\varphi} f_3 \left(\frac{\Gamma}{M_l} \right). \quad (8.13)$$

Если положить

$$R = \frac{b}{aM_l \sigma \delta \Gamma}; \quad \varphi_0 = \frac{1}{aM_\varphi} f_3 \left(\frac{\Gamma}{M_l} \right),$$

то условие (8.13) сведется к виду

$$\varphi - RI = \varphi_0; \quad \varphi - \varphi_0 = RI, \quad (8.14)$$

эквивалентному закону Ома. Соотношение (8.14) удовлетворяется при подключении напряжения φ_0 к участку границы $\delta\Gamma$ с напряжением φ через сопротивление R .

Как уже отмечалось ранее, физические поля характеризуются двумя сопряженными семействами изолиний. В частном случае плоскопараллельных полей, описываемых уравнением Лапласа (4.21а), существует чрезвычайно простой и эффективный метод исследования сопряженных полей. Как известно, функция Φ , удовлетворяющая на плоскости уравнению Лапласа, является гармонической. Сопряженная с ней функция Ψ также является гармонической и, следовательно, в свою очередь, тоже удовлетворяет уравнению Лапласа. Граничные условия для функций Φ и Ψ связаны соотношениями Коши — Римана, которые на любой аналитической кривой Γ имеют вид.

$$\frac{\partial \Phi}{\partial n} = - \frac{\partial \Psi}{\partial \Gamma}; \quad \frac{\partial \Phi}{\partial \Gamma} = \frac{\partial \Psi}{\partial n}, \quad (8.15)$$

причем производные берутся по нормали n и касательной Γ к кривой.

В общем случае выражения (8.15) приводят к соотношениям

$$\begin{aligned} \Phi|_{\Gamma} = f_1(\Gamma) &\Rightarrow \frac{\partial \Psi}{\partial n} = \frac{\partial \Phi}{\partial \Gamma} = \frac{\partial f_1(\Gamma)}{\partial \Gamma} = f_1^*(\Gamma) \Rightarrow \frac{\partial \Psi}{\partial n} \Big|_{\Gamma} = f_1^*(\Gamma); \\ \frac{\partial \Phi}{\partial n} \Big|_{\Gamma} = f_2(\Gamma) &\Rightarrow \frac{\partial \Psi}{\partial \Gamma} = - \frac{\partial \Phi}{\partial n} = -f_2(\Gamma) \Rightarrow \Psi(\Gamma) = \\ &= - \int_{\Gamma_0}^{\Gamma} f_2(\Gamma) d\Gamma + \Psi(\Gamma_0) = f_2^*(\Gamma) \Rightarrow \Psi(\Gamma)|_{\Gamma} = f_2^*(\Gamma), \end{aligned} \quad (8.16)$$

т. е. в итоге к замене граничных условий первого рода граничными условиями второго рода и наоборот. Граничные условия третьего рода при переходе к сопряженной функции сохраняются, но входящие в них коэффициенты должны быть пересчитаны:

$$\left(a\Phi + b \frac{\partial \Phi}{\partial n} \right) \Big|_{\Gamma} = f_3(\Gamma) \Rightarrow \left(a^* \frac{\partial \Psi}{\partial n} + b^* \Psi \right) \Big|_{\Gamma} = f_3^*(\Gamma). \quad (8.17)$$

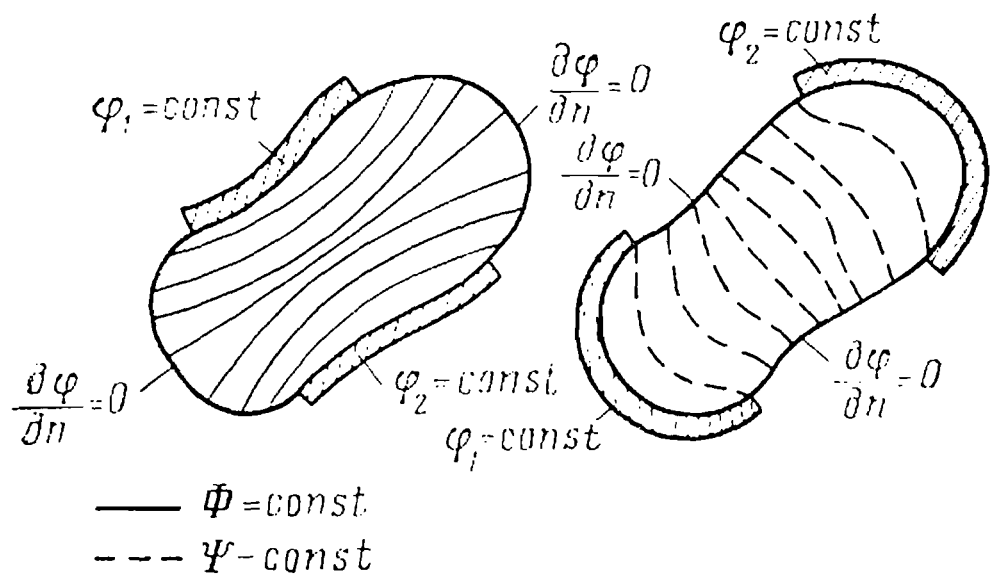


Рис. 8.2. Обращение модели плоско-параллельного однородного стационарного поля

Если на границах области заданы постоянные потенциалы и потоки, равные 0, то условия (8.16) упрощаются:

$$\Phi|_{\Gamma=\text{const}} \Rightarrow \frac{\partial \varphi}{\partial n} \Big|_{\Gamma} = 0; \quad \frac{\partial \Phi}{\partial n} \Big|_{\Gamma} = 0 \Rightarrow \Psi|_{\Gamma=\text{const}}. \quad (8.16a)$$

Потенциал и функция потока удовлетворяют одному и тому же уравнению (4.21a) и, следовательно, могут быть изучены методами аналогового моделирования на основе одного и того же модельного (например, электрического) поля. При этом конфигурация и размеры модельной области не меняются, но требуется изменить граничные условия в соответствии с соотношениями (8.16). В качестве примера на рис. 8.2 показаны модели для моделирования Φ и Ψ . Таким образом, для плоскопараллельных полей, описываемых уравнением Лапласа, может быть построено две системы аналогий, называемых обычно *прямой* (анalogией A) и *обратной* аналогией (анalogией B). При прямой аналогии электрический потен-

Таблица 8.2

Тип аналогии	Прямая аналогия (тип A)	Обратная аналогия (тип B)	Электрический эквивалент
Исследуемая функция	Потенциал поля Φ	Функция потока Ψ	Электрический потенциал φ
Крайевые условия	1-го рода $\Phi _{\Gamma} = f_1(\Gamma)$	2-го рода $\frac{\partial \psi}{\partial n} \Big _{\Gamma} = f_2^*(\Gamma)$	В соответствии с задачей $\varphi _{\Gamma} = f_1(\Gamma)$ $\frac{\partial \varphi}{\partial n} \Big _{\Gamma} = f_2(\Gamma)$
	2-го рода $\frac{\partial \Phi}{\partial n} \Big _{\Gamma} = f_2(\Gamma)$	1-го рода $\psi _{\Gamma} = f_1^*(\Gamma)$	
Изолинии	Эквипотенциали $\Phi = \text{const}$	Силовые линии $\Psi = \text{const}$	Эквипотенциали $\varphi = \text{const}$
	Силовые линии $\Psi = \text{const}$	Эквипотенциали $\Phi = \text{const}$	Силовые линии $\Psi = \text{const} \perp \varphi = \text{const}$

циал φ моделирует натуральный потенциал Φ , при обратной — сопряженную с потенциалом функцию потока Ψ . Некоторые сравнительные характеристики прямых и обратных моделей приведены в табл. 8.2. Рассмотренное свойство общности описания потенциальной функции и функции потока часто называют свойством *дуальности* плоскопараллельных полей.

К сожалению, такой простой подход к исследованию сопряженных полей реализуется лишь для плоскопараллельных полей, описываемых уравнением Лапласа. Уже в случае обобщенного однородного уравнения эллиптического типа переход к сопряженным функциям требует существенного изменения характеристик изучаемой области:

$$\nabla (\omega(\xi') \nabla \Phi) = 0 \Rightarrow \nabla \left(\frac{1}{\omega(\xi')} \nabla \Psi \right) = 0. \quad (8.18)$$

В частности, изучение сопряженной функции для осесимметричного поля требует перехода к области с *гиперболическим* изменением характеристик среды:

$$\nabla (r \nabla \Phi) = 0 \Rightarrow \nabla \left(\frac{1}{r} \nabla \Psi \right) = 0. \quad (8.18a)$$

Вопросы для самоконтроля

1. Что такое сходственные величины, коэффициенты аналогии, индикаторы аналогии?
2. Прodelайте преобразования, сводящие одно из полей, приведенных в табл. 8.1, к модельному полю.
3. Запишите модельные граничные условия 1-, 2- и 3-го рода.
4. В чем заключается свойство дуальности плоскопараллельных полей?
5. Как преобразуются граничные условия при переходе от исследования потенциала Φ к исследованию функции потока Ψ ?

8.2. МОДЕЛИ СТАЦИОНАРНЫХ ПОЛЕЙ

Существуют два способа построения электрических моделей физических полей в элементах ДЛА. Модели могут быть построены из *сплошных проводящих сред*. В этом случае модельное поле характеризуется непрерывным распределением параметров, в частности, электрического потенциала. Можно построить также *сеточные электрические модели*, основанные на конечно-разностной аппроксимации непрерывного поля. В этом случае модельное поле задается дискретно в узлах сетки.

При использовании сплошных проводящих сред модель исследуемого объекта выполняется из сплошного проводящего материала. В качестве такого материала могут быть использованы любые среды, слабо проводящие электрический ток (металлизированные ткани, проводящие пластины и сыпучие материалы, тончайшая фольга и т. д.). Однако доминирующее распространение получили два материала: электропроводная (насыщенная графитом) бумага и слабые электролиты, например, водопроводная вода. Материал

модели должен обладать возможно большей однородностью свойств. Желательно также, чтобы его проводимость возможно меньше зависела от внешних факторов (например, температуры). Наряду с основной проводящей средой для изготовления моделей используются еще две среды: так называемые идеальные проводники (металлы, графит), проводимость которых в 10^3 и более раз превосходит проводимость основной среды, и идеальные изоляторы (оргстекло, парафин, воздух), проводимость которых в 10^3 и более раз хуже проводимости основной среды.

Для исследования полей в элементах ДЛА могут быть созданы полноразмерные модели. Например, для изучения температурного поля в лопатке газовой турбины можно изготовить полноразмерную модель лопатки из твердого или жидкого (в последнем случае для сохранения формы модели необходим специальный профилированный сосуд — ванна) проводящего материала, задать на границах этой модели необходимые граничные условия, а затем измерить распределение потенциала в модели. Однако такие модели очень сложны и, главное, сугубо индивидуальны.

К счастью, большинство реальных полей обладает *плоскостной* или *осевой* симметрией, позволяющей ограничиться изучением поля в одной характерной плоскости. Так, если пренебречь концевыми эффектами, можно считать, что распределение температур идентично в ряде последовательных сечений лопатки турбины. Точно также идентичны в этих сечениях и поля скоростей обтекающего лопатку газа. Не зависят от азимутальной координаты и повторяются во всех меридиональных плоскостях распределение температур в диске газовой турбины, распределение скоростей при ламинарном течении в трубе круглого сечения, распределение магнитных силовых линий в ЭРД, возбуждаемых кольцевым электрическим током.

Во всех этих случаях достаточно изучить рассматриваемое поле или в *тонком плоском* (рис. 8.3, а), или в некотором *клиновом слое* (рис. 8.3, б), вырезанном плоскостями симметрии $I—I$, $II—II$.

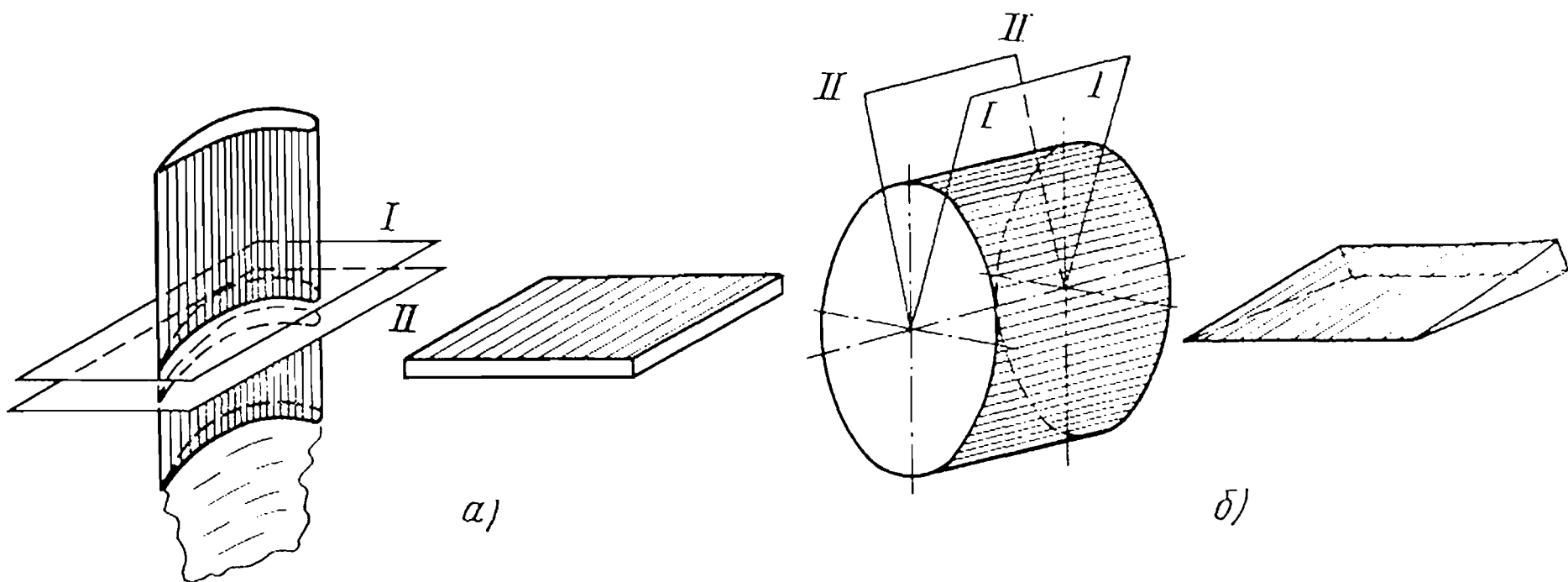


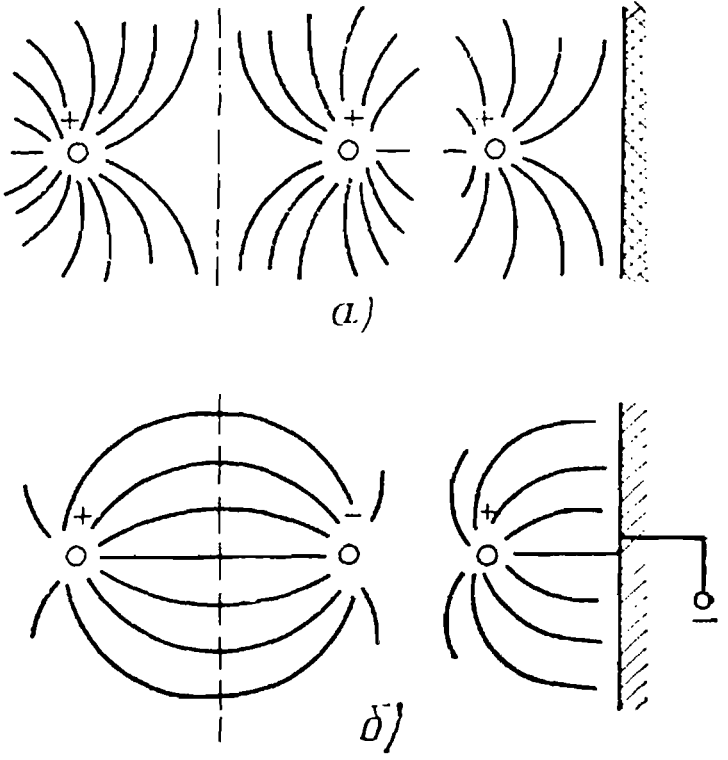
Рис. 8.3. Использование свойства симметрии при моделировании полей:
а — плоскопараллельное поле; б — осесимметричное поле

Рис. 8.4. Влияние изолирующей (а) и проводящей (б) стенок на характеристики поля

Возможность использования свойств симметрии и тонких слоев упрощает построение моделей, делает их универсальными.

В электрическом поле плоскости симметрии могут быть выполнены из изолирующих материалов. Действительно, влияние границы из изолятора на электрическое поле точечного источника в проводящей среде эквивалентно влиянию точно такого же источника, расположенного симметрично относительно границы (рис. 8.4, а). Справедливость этого утверждения следует из того, что электрический ток не идет в изолятор, на границе которого, следовательно, выполняется условие $\partial\phi/\partial n=0$. В модели изолирующая граница может быть выполнена, например, в виде пластины из оргстекла, установленной в электролите. Такую же роль играет граница проводящего слоя с непроводящей средой (роль изолирующей границы играет воздух). В силу таких же рассуждений (рис. 8.4, б) можно утверждать, что идеально проводящая поверхность играет роль поверхности антисимметрии.

Модель плоского слоя может быть выполнена из листа проводящей бумаги или в виде плоского слоя электролита в специальном сосуде — электролитической ванне (рис. 8.5, а). Имитация клинового слоя требует увеличения проводимости проводящего слоя пропорционально радиусу модели. Такой эффект можно получить или в плоском слое за счет соответствующего увеличения



Закон изменения электропроводности

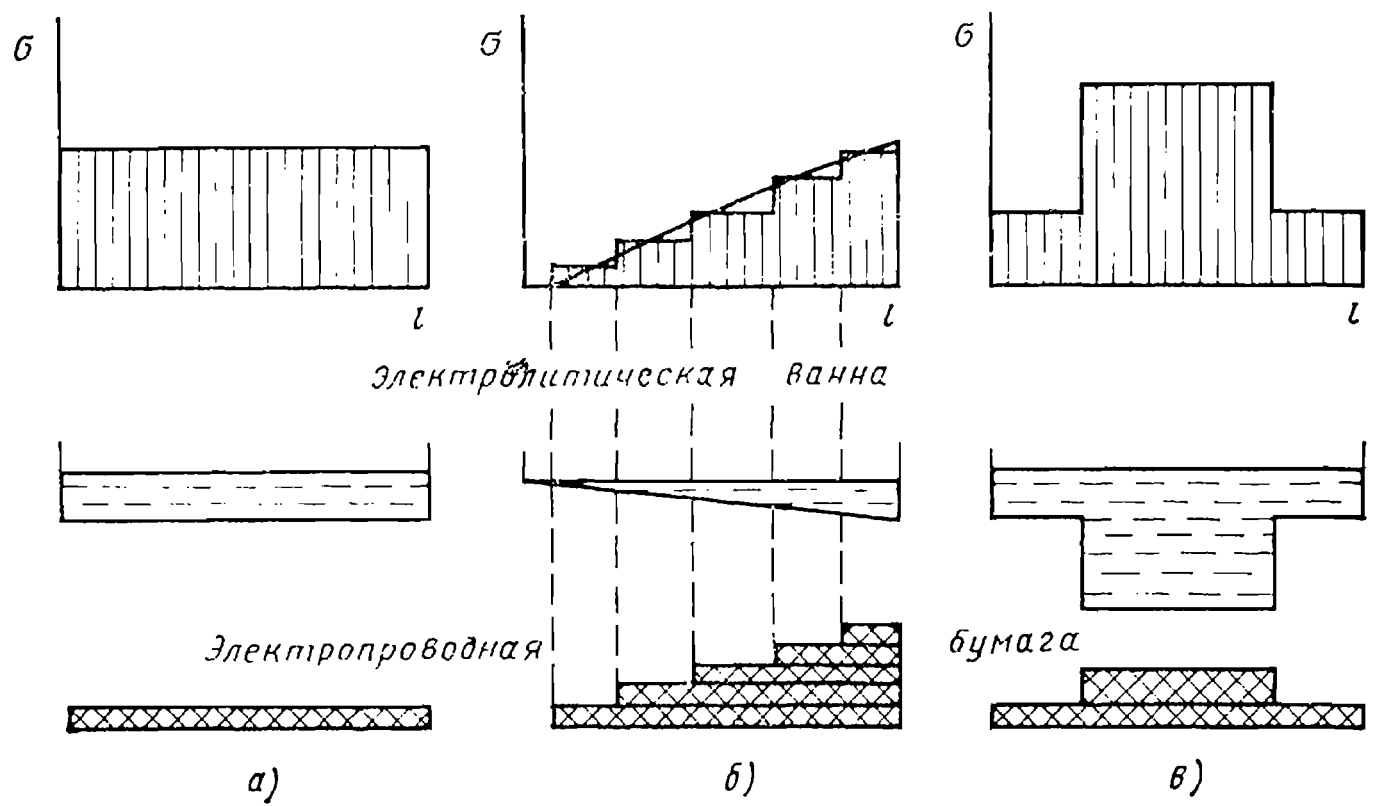


Рис. 8.5. Задание области моделирования поля в электролитической ванне и на электропроводной бумаге

а — плоскопараллельное поле; б — осесимметричное поле; в — поле при ступенчатом изменении характеристик среды

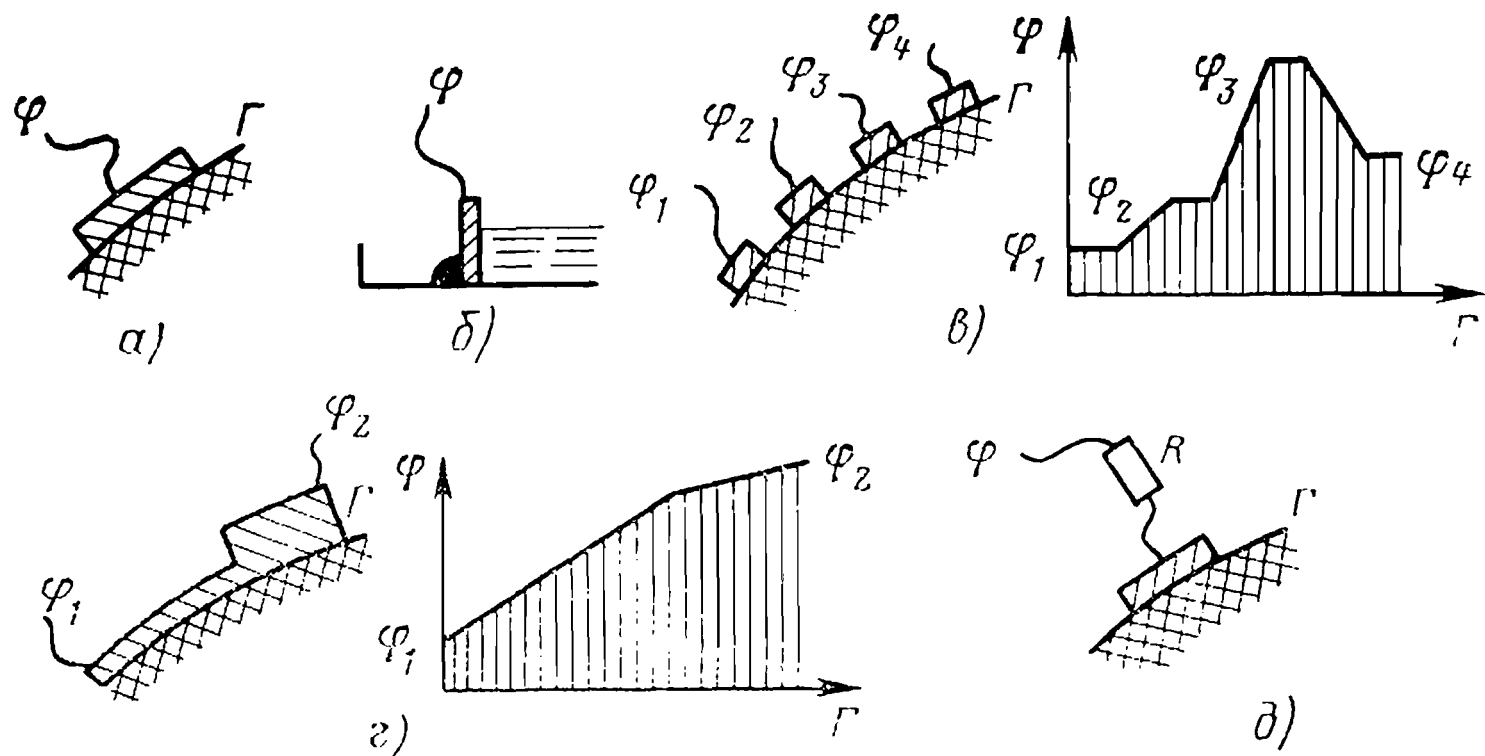


Рис. 8.6. Задание граничных условий при моделировании в сплошной проводящей среде

а, б — задание граничного потенциала на электропроводной бумаге и на электролитической ванне соответственно; в, г — задание потенциала, изменяющегося вдоль границы поля; д — задание граничных условий 3-го рода

концентрации проводящего наполнителя, или за счет изменения толщины проводящего слоя из однородного материала. При использовании жидких электролитов этот эффект достигается с помощью *наклонной* ванны (рис. 8.5, б).

При исследовании полей в объектах со ступенчатым изменением свойств (например, теплопроводности) можно использовать модели со ступенчатым изменением проводимости (рис. 8.5, в).

Для задания граничных условий при моделировании в сплошных проводящих средах используются электроды и изоляторы. Так, для задания на некотором участке границы модели постоянного потенциала ($\varphi|_{\Gamma} = \text{const}$) необходимо установить на этом участке металлический электрод и подать на него соответствующее напряжение. В моделях из электропроводной бумаги такой электрод выполняется в виде металлической шины, прижатой или приклеенной специальным электропроводным клеем к бумаге (рис. 8.6, а). В электролитической ванне металлический электрод устанавливается непосредственно на дно ванны (рис. 8.6, б). Для задания некоторого распределения потенциала может быть использована ступенчатая аппроксимация этого распределения с помощью цепочки электродов (рис. 8.6, в) или установленная на границе специальная шина, в которой обеспечивается заданное монотонное изменение потенциала (рис. 8.6, г). Точно также можно задать граничные условия 2-го рода, только вместо потенциалов к границе должны быть подведены токи. Граничные условия 3-го рода могут быть заданы с помощью потенциалов, поданных к границам через заданные сопротивления (рис. 8.6, д). Распределенные источники поля моделируются с помощью дополнительных электродов, подводящих токи во внутренние точки модели. Схемы моделей показаны на рис. 8.7, а, б.

Результаты моделирования представляются сеткой эквипотенциальных линий ($\varphi = \text{const}$), соответствующих изолиниям изучае-

мой функции. Получение такой сетки связано с решением двух задач:

- задачи определения потенциала φ в заданных точках области (или выделения точек с заданным потенциалом);
- задачи фиксации найденных точек.

Измерить потенциал в заданной точке можно непосредственно с помощью гальванометра (рис. 8.7, в). Однако такой способ не достаточно точен, так как, во-первых, при измерении абсолютной величины потенциала возможны погрешности из-за колебаний напряжения в источнике питания, а, во-вторых, такая схема измерения вносит в модель погрешность из-за утечек тока через измерительную цепь. Обычно при измерении потенциала в моделях используют *компенсационный* способ измерения, а отсчет ведут в процентах или долях полного, приложенного к модели напряжения. Схема измерения потенциала в модели компенсационным способом представлена на рис. 8.7, а, б, эквивалентная схема измерительно-го равноплечного моста — на рис. 8.7, г.

Всмотритесь внимательно в обе схемы! Обе эти схемы включают в себя измерительную цепь, состоящую из мостов сопротивлений R_1 и R_2 , исследуемую цепь (сопротивления R_3 и R_4) и измерительную диагональ с нульиндикатором (гальванометром). Точки на концах диагонали имеют потенциалы φ_1 и φ_2 . Положение равновесия моста (отсутствие тока в диагонали) соответствует условию

$$\varphi_1 = \varphi_2 \tag{8.19}$$

и достигается при

$$\frac{R_1}{R_1 + R_2} = \frac{R_3}{R_3 + R_4} . \tag{8.20}$$

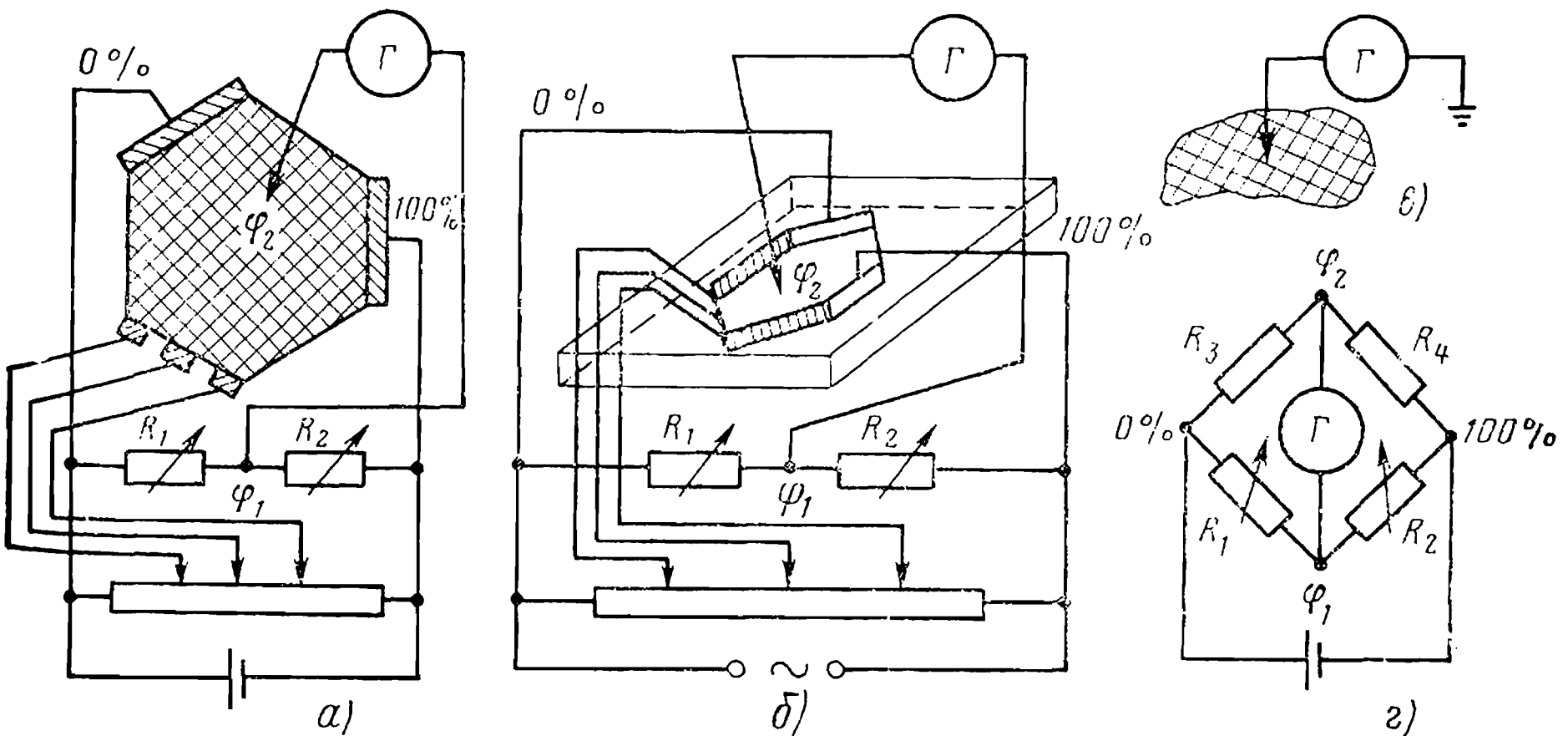


Рис. 8.7. Схемы моделирующих установок:
 а — моделирование на электропроводной бумаге; б — моделирование на электролитической ванне; в — непосредственное измерение потенциала в модели; г — компенсационная схема измерения потенциала с помощью равноплечного моста

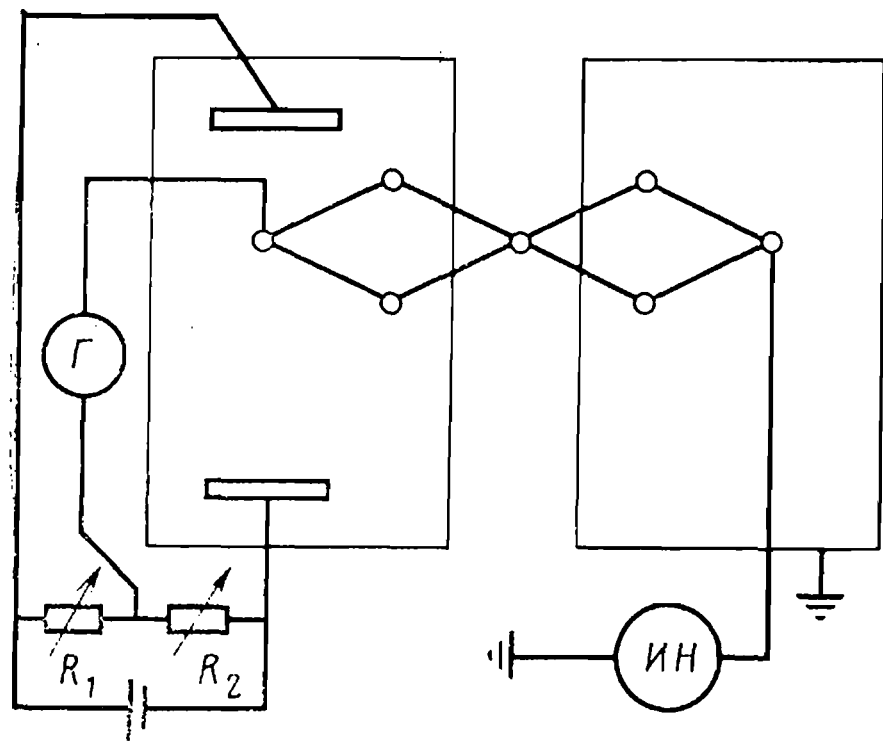
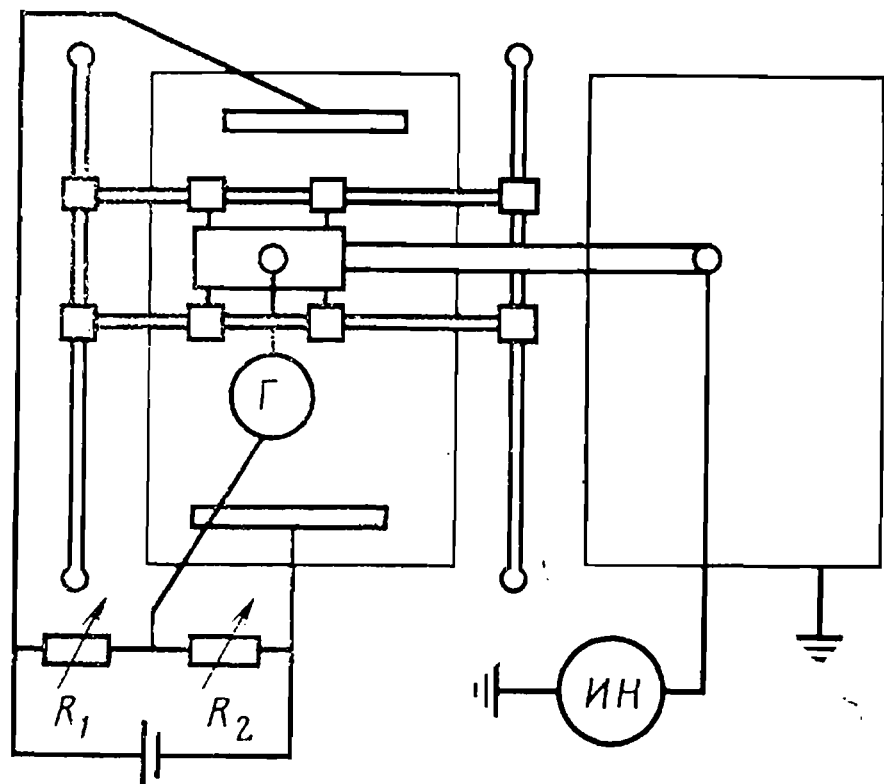


Рис. 8.8. Системы для записи результатов измерения в электролитической ванне

При этом потенциал φ_1 , выраженный в процентах, рассчитывается по величинам сопротивлений R_1 и R_2 :

$$\varphi_1 = 100 \frac{R_1}{R_1 + R_2} \% \quad (8.21)$$

Расчеты дополнительно упрощаются, если сумма $R_1 + R_2$ постоянна и равна 100 единицам (100 Ом, кОм, МОм и т. д.). Тогда $\varphi_1 = R_1 \%$. Для измерения потенциала в заданной точке необходимо установить в эту точку измерительный зонд, а потом, изменяя R_1 и R_2 , добиться равновесия моста. Потенциал исследуемой точки φ_2 определяется соотношениями (8.19) — (8.21). Для определения точек с *заданным* потенциалом ($\varphi_2 = \text{const}$) необходимо задать этот потенциал с помощью сопротивлений R_1 и R_2 в



точке φ_1 , а затем, плавно перемещая зонд по модели, выделить необходимое количество точек, для которых соблюдается равновесие моста ($\varphi_1 = \varphi_2$).

Фиксация эквипотенциалей в случае использования для моделирования электропроводной бумаги осуществляется непосредственно на модели с помощью цветных карандашей. В электролитических ваннах для фиксации эквипотенциалей используется система взаимоднозначного соответствия, связывающая точки модели с точками ее чертежа (рис. 8.8). При этом положение зонда в модели соответствует положению специального отметчика на чертеже.

Точность моделирования полей в проводящих средах зависит от ряда факторов и, в первую очередь, от однородности и изотропности среды, от отсутствия контактных сопротивлений между электродами модели и средой, от точности геометрического воспроизведения исследуемой области. При решении внешних задач, связанных с исследованием неограниченных областей, возникают ошибки из-за ограниченных размеров модели. Подробный анализ ошибок, возникающих при моделировании в сплошных средах, и путей их уменьшения можно найти в литературе¹. Принято счи-

¹ Ницецкий А. В. Аналоговые и разностные методы решения внешних краевых задач. Ученые записки РПИ, т. XI, Рига, Звайгзне, 1965. С. 5—430.

тать, что допустимые погрешности при моделировании полей на электропроводной бумаге составляют в среднем 5—10% измеряемой величины, на электролитических ваннах — 2—5%.

Использование сеточных моделей основано на конечно-разностной аппроксимации непрерывных полей (см. разд. 4.3).

Для электрического моделирования поля, описываемого уравнением Лапласа (точнее, конечно-разностным выражением для уравнения Лапласа (4.39)), может быть использована сетка резисторов R (R -сетка). Общий вид сетки показан на рис. 8.9, а, схема узловой точки — на рис. 8.10, а. Согласно первому закону Кирхгофа сумма токов, втекающих в узловую точку сетки, равна нулю:

$$\sum_{i=1}^6 I_i = 0; I_i = \frac{\varphi_i - \varphi_0}{R}; \sum_{i=1}^6 \frac{\varphi_i - \varphi_0}{R} = 0; \quad (8.22)$$

$$\frac{1}{R} \left(\sum_{i=1}^6 \varphi_i - 6\varphi_0 \right) = 0; \varphi_0 = \frac{1}{6} \sum_{i=1}^6 \varphi_i.$$

Распределение потенциала φ на R -сетке аналогично распределению потенциала, заданному соотношением (4.39), причем соот-

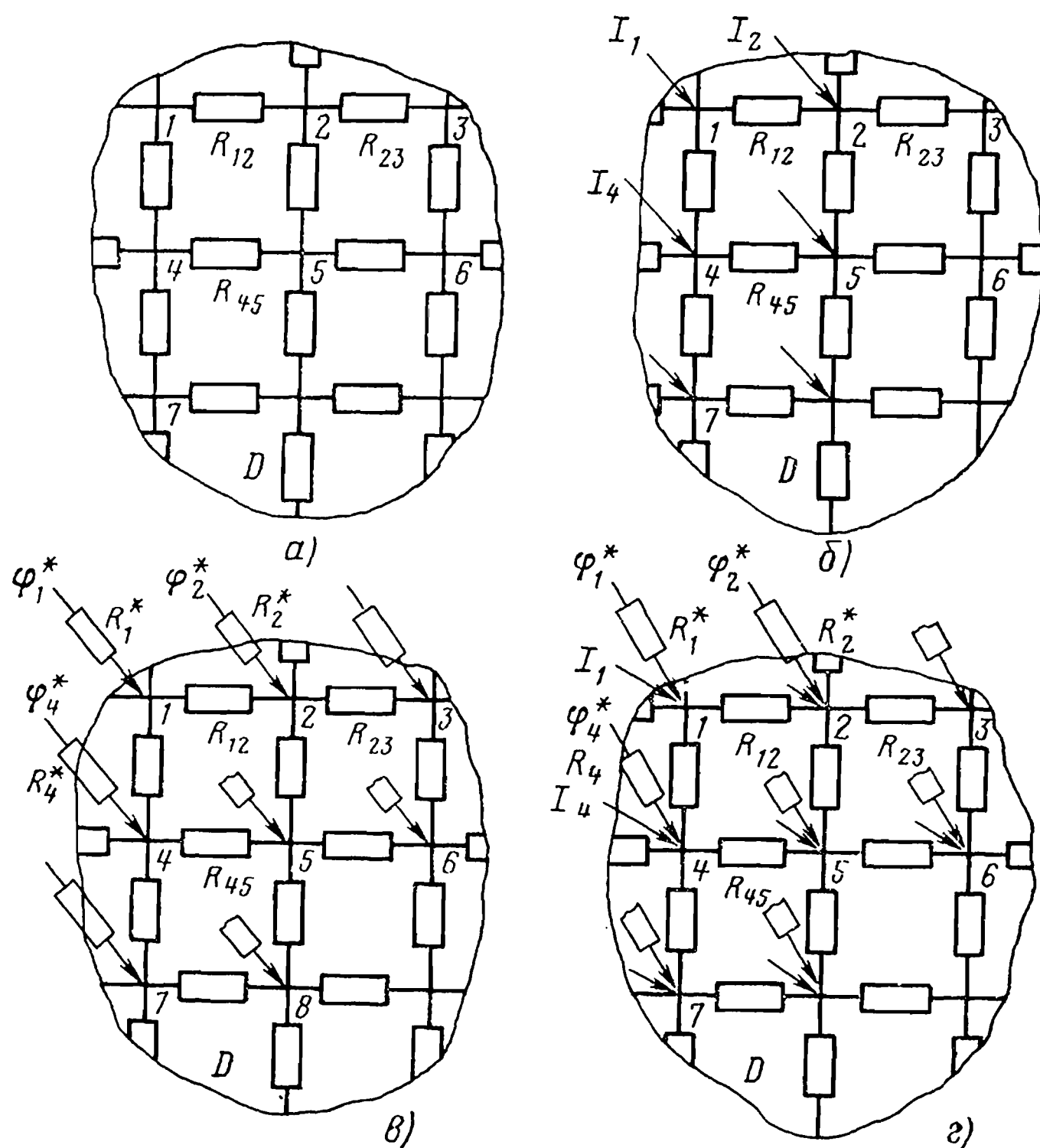


Рис. 8.9. Общий вид сеток резисторов для моделирования полей

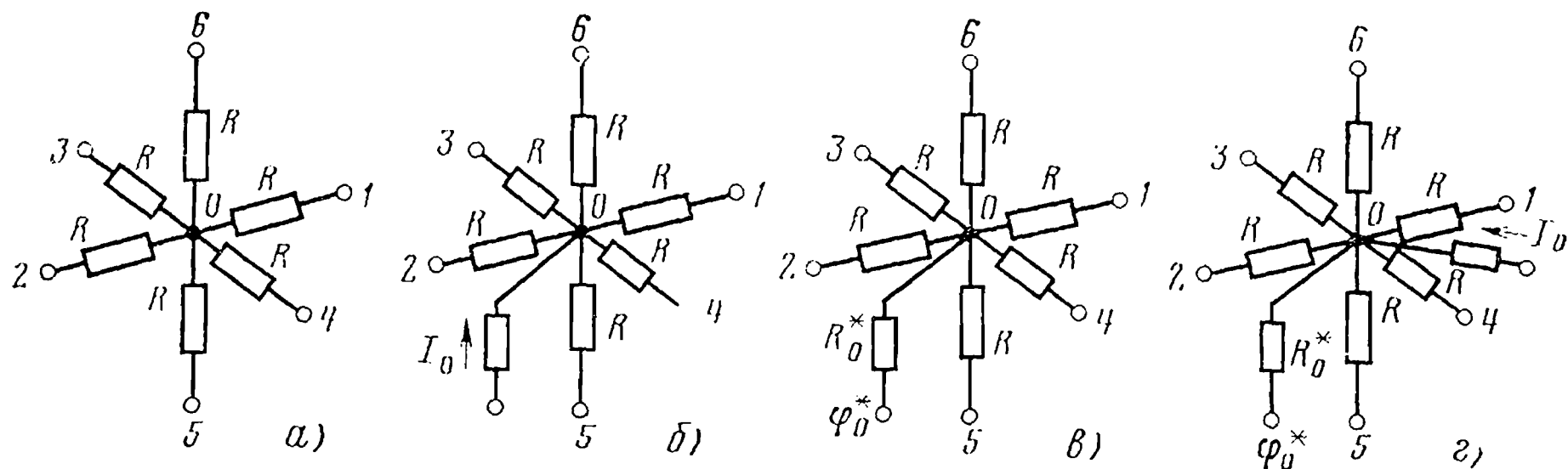


Рис. 8.10. Узлы сеток резисторов для моделирования полей

ветствие в данном случае устанавливается с помощью одного коэффициента аналогии ($M_\varphi = \Phi/\varphi$). Следовательно, R -сетка может быть использована для моделирования поля, описываемого уравнением Лапласа.

Для моделирования поля с распределенными источниками, описываемого уравнением Пуассона (конечно-разностная аппроксимация уравнения Пуассона задается соотношением (4.41)), может быть использована R -сетка с дополнительными источниками тока (рис. 8.9, б и 8.10, б). Для узла такой сетки первый закон Кирхгофа имеет вид

$$\sum_{i=1}^6 I_i = I_0^*; \quad \frac{1}{R} \left(\sum_{i=1}^6 \varphi_i - 6\varphi_0 \right) = I_0^*. \quad (8.23)$$

Аналогия уравнений (4.41) и (8.23) достигается при условии

$$\frac{M_\varphi}{M_R M_I} \equiv 1; \quad M_\varphi = \frac{\Phi}{\varphi}; \quad M_I = \frac{A(\xi_0^r, \Phi_0)}{I_0^*}; \quad (8.24)$$

$$M_R = \frac{h^2}{R},$$

причем легко показать, что (8.24) эквивалентно (8.5а), так как

$$M_R = \frac{1}{M_I M_\sigma}. \quad (8.25)$$

При создании конкретных моделей на основе сеток сопротивлений используются те же принципы, что и при использовании сплошных проводящих сред. Так, в силу симметрии моделирование плоскопараллельных и осесимметричных полей осуществляется на двумерной сетке сопротивлений, причем в последнем случае имитация клинового слоя достигается соответствующим выбором сопротивлений (рис. 8.11).

Для изучения трехмерных полей могут быть использованы объемные сетки сопротивлений. Граничные условия задаются с по-

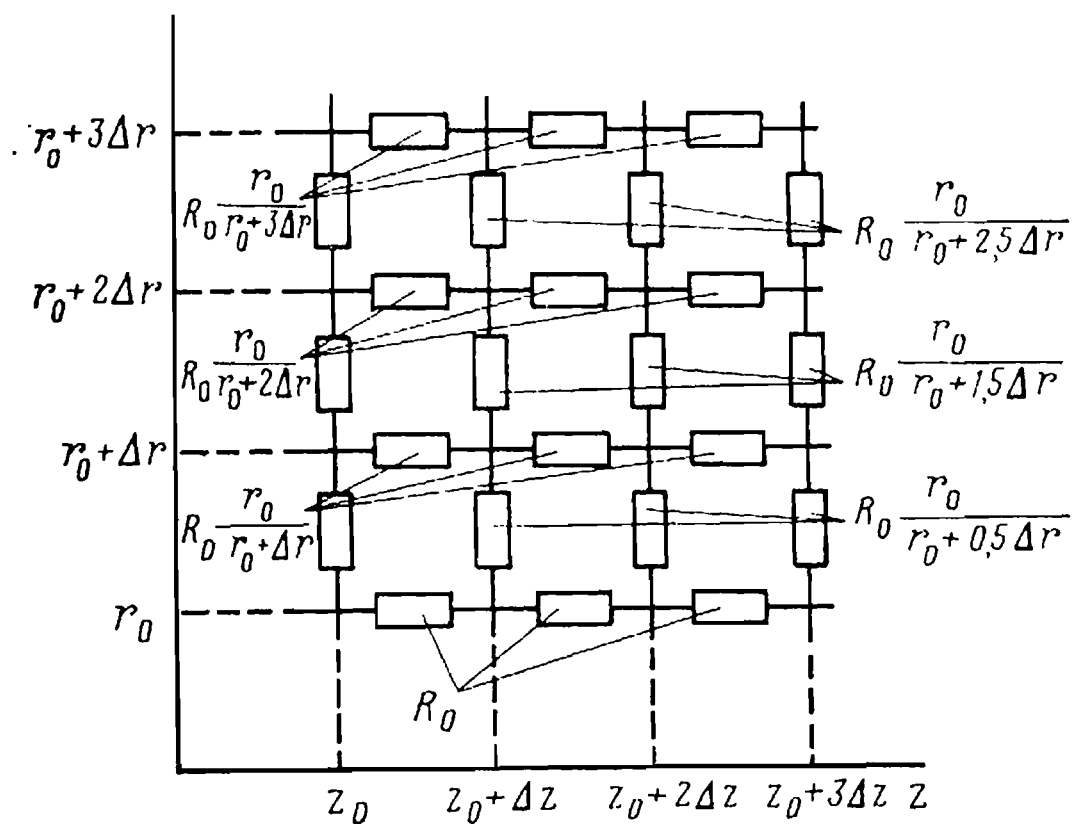
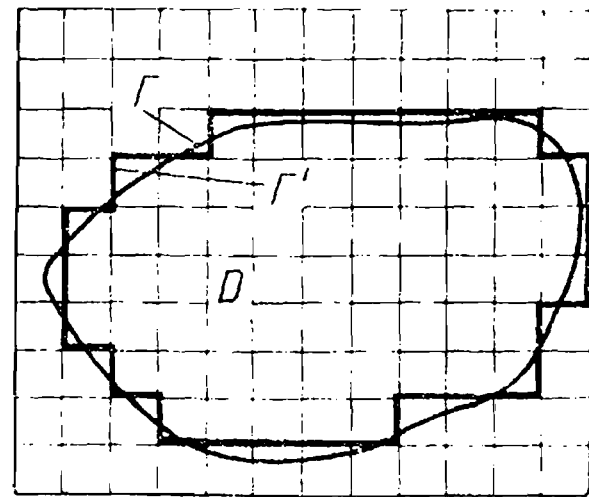
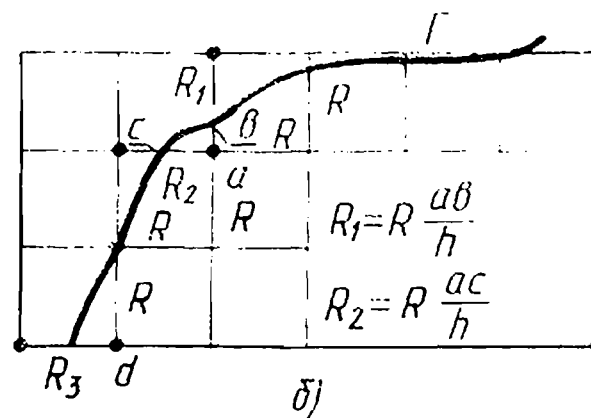


Рис. 8.11. Сетка для моделирования осесимметричного поля



а)



б)

Рис. 8.12. Задание границ криволинейной области на сетке сопротивлений:

а — конечно-разностная аппроксимация границы; б — метод «притягивания» узлов на границу

мощью потенциалов, токов или потенциалов, подключенных через заданные сопротивления. Однако эти потенциалы или токи подключаются к конкретным граничным узлам сетки. Задание постоянного потенциала осуществляется закорачиванием граничных узлов. Разрыв цепей на границах соответствует условию $\frac{\partial \varphi}{\partial n} = 0$ и т. д.

Специфическим является вопрос задания конфигурации границ области на сеточных моделях. Дело в том, что прямоугольные границы сетки могут не совпадать с реальными криволинейными границами области. Простейшее решение этого вопроса связано с заменой криволинейных границ их конечно-разностной аппроксимацией (рис. 8.12, а). Для сеток с переменными сопротивлениями может быть использован прием «притягивания» узлов на границу, заключающийся в изменении (уменьшении) соответствующих сопротивлений (рис. 8.12, б). При этом фактически происходит изменение «электрической» конфигурации сетки.

При размещении модели на сетке сопротивлений возможно разделение исследуемой области на элементы при соответствующей коммутации граничных точек (рис. 8.13), позволяющее лучше использовать стандартные, обычно прямоугольные сетки.

Сравнивая сеточные модели с моделями со сплошной проводящей средой, нужно учитывать следующие факторы:

с одной стороны, использование конечно-разностной аппроксимации и, следовательно, сеток сопротивлений всегда связано с внесением погрешности за счет перехода от производных к их конеч-

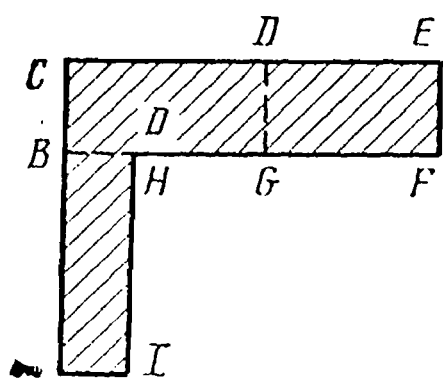
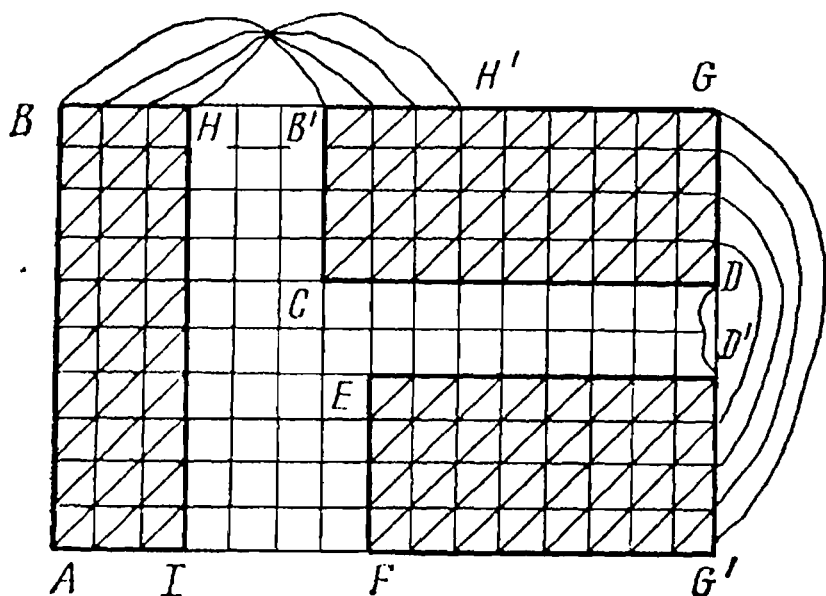


Рис. 8.13. Метод «кройки и шитья» при размещении модели на сетке:

$A, B, C, \dots, I$ — узловые точки границы области D и D' .



но-разностным выражениям. Поэтому сеточные модели в принципе менее точны, чем модели из сплошной проводящей среды;

с другой стороны, сеточные модели существенно более удобны и универсальны. Они позволяют легко моделировать поля в неоднородных и анизотропных средах. Так, при моделировании поля температур в конструкциях, состоящих из материалов с различными λ , достаточно просто изменить сопротивления между соответствующими узловыми точками сетки. Сеточные модели легко сопрягаются с цифровыми вы-

числительными машинами (при конечном числе точек удобно организовать их поочередный опрос), и поэтому их легче использовать в составе цифроаналоговых комплексов.

Вопросы для самоконтроля

1. Как используются при построении моделей свойства симметрии полей?
2. Как задать на модели граничные условия 1-, 2- и 3-го рода?
3. Запишите конечно-разностные выражения первых и вторых производных, лапласиана, уравнения Лапласа, уравнения Пуассона.
4. Приведите схемы сеток для моделирования полей, описываемых уравнениями Лапласа и Пуассона. Докажите аналогию этих сеток и соответствующих полей.
5. В чем заключаются преимущества компенсационного способа измерения потенциала поля?

8.3. МОДЕЛИ НЕСТАЦИОНАРНЫХ ПОЛЕЙ

Методы конечно-разностной аппроксимации успешно применяются для изучения как эллиптических, так и других уравнений математической физики. Поэтому естественно возникает желание построить сеточные модели для изучения *нестационарных* физических полей в элементах ДЛА. Такие сетки могут быть построены или из активных сопротивлений R , или включать в себя реактивные элементы C, L [8]. В зависимости от состава различают R -, RC -, LC -, RLC -сетки. Во всех рассматриваемых ниже случаях исследования нестационарных полей используется конечно-разностное выражение лапласиана (4.38). Модели из сплошных проводящих сред для моделирования нестационарных полей применяются относительно редко.

Использование резисторных сеток (R -сеток) основано на дискретном (ступенчатом) представлении времени. Так, например, при моделировании полей, описываемых уравнением Фурье (4.23), используется полная конечно-разностная форма (4.42). Эта форма напоминает соответствующую форму уравнения Пуассона (4.41), и поэтому для его исследования можно использовать сетку, подобную сетке, представленной на рис. 8.9, б, 8.10, б. Узел такой сетки показан на рис. 8.9, в, 8.10, в. Цепи R^* , φ^* включены в *каждую* узловую точку сеточной модели, что и отражено их индексами. Распределение потенциала на такой сетке в силу первого закона Кирхгофа описывается уравнениями

$$\sum_{i=1}^7 I_i = 0; \quad \sum_{i=1}^6 \frac{\varphi_i - \varphi_0}{R} + \frac{\varphi_0^* - \varphi_0}{R_0^*} = 0; \quad (8.26)$$

$$\frac{1}{R} \left(\sum_{i=1}^6 \varphi_i - 6\varphi_0 \right) = \frac{\varphi_0 - \varphi_0^*}{R_0^*}.$$

Последнее из этих уравнений полностью соответствует уравнению (4.42) при условии

$$\frac{M_R}{M_t^*} \equiv 1; \quad M_R = \frac{h^2 A(\xi^r, \Phi_0)}{R}; \quad M_t^* = \frac{\Delta t}{R_0^*}. \quad (8.27)$$

Коэффициент M_t^* включает в себя интервал времени Δt , поэтому сопротивление R_0^* называется «временным» сопротивлением точки 0.

Из (4.42) и (8.26) видно, что в то время как потенциал φ_0 соответствует исследуемой функции Φ_0 в момент t ($\Phi_0, t = M_\varphi \varphi_0$), потенциал φ_0^* соответствует этой же функции в предыдущий момент $t - \Delta t$ ($\Phi_0, t - \Delta t = M_\varphi \varphi_0^*$). Для решения задачи должны быть заданы граничные условия, в общем случае изменяющиеся по времени, и начальные условия — распределение $\Phi(\xi^r)$ в начальный момент t_0 .

Моделирование осуществляется методом последовательных приближений. Для начального момента t_0 распределение функции во всех точках поля известно, следовательно, известно и распределение потенциала φ . Это распределение может быть использовано в качестве распределения φ^* для определения распределения φ в момент $t_0 + \Delta t$. Полученное при этом распределение в свою очередь позволяет найти распределение в момент $t_0 + 2\Delta t$ и т. д. В процессе моделирования от шага к шагу могут меняться характеристики сетки (сопротивления R) и граничные условия. Если граничные условия в каждый момент времени сами определяются результатами решения задачи (например, при излучении с границы), возникает *двойной* цикл последовательных приближений (рис. 8.14).

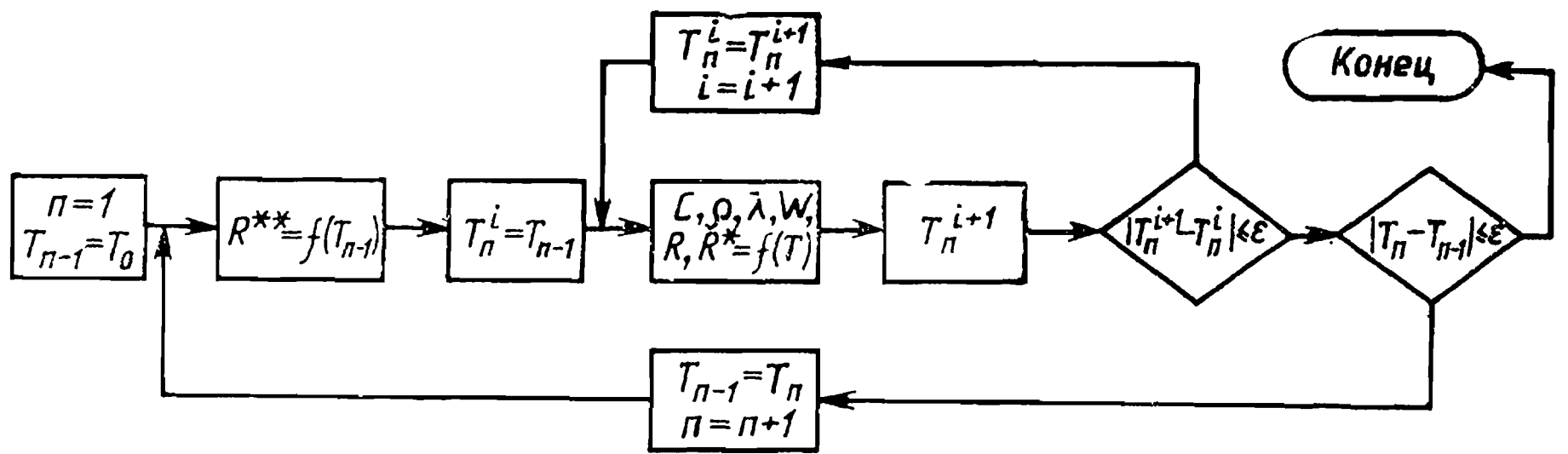


Рис. 8.14. Метод последовательных приближений при изучении температурных полей

Для общего уравнения теплопроводности с учетом изменения температуры по времени и объемного тепловыделения (4.30) конечно-разностная аппроксимация может быть задана в виде

$$\frac{1}{R_{T_0}} \left(\sum_{i=1}^6 T_i - 6T_0 \right) = C_0 \frac{T_{0t} - T_{0t-\Delta t}}{\Delta t} - W_0, \quad (8.28)$$

где R_{T_0} , C_0 , W_0 — тепловое сопротивление, теплоемкость и мощность тепловыделения в элементарном объеме, соответствующем точке O :

$$R_{T_0} = \frac{1}{h\lambda(\xi_0^r, T_0)}; \quad (8.29)$$

$$C_0 = h^3 C(\xi_0^r, T_0) \rho(\xi_0^r, T_0); \quad W_0 = h^3 w(\xi_0^r, T_0).$$

Для моделирования этого поля может быть использована сетка, узел которой показан на рис. 8.9, з, 8.10, з. Для потенциала в узлах такой сетки выполняются соотношения

$$\sum_{i=1}^8 I_i = 0; \quad \sum_{i=1}^6 \frac{\varphi_i - \varphi_0}{R} + \frac{\varphi_0^* - \varphi_0}{R_0^*} + I_0^* = 0; \quad (8.30)$$

$$\frac{1}{R} \left(\sum_{i=1}^{2m} \varphi_i - 6\varphi_0 \right) = \frac{\varphi_0 - \varphi_0^*}{R_0^*} - I_0^*.$$

Последнее из соотношений (8.30) полностью совпадает с (8.28) при условии

$$\begin{aligned} \frac{M_R}{M_t^*} &\equiv 1; \quad \frac{M_R M_I}{M_\varphi} \equiv 1; \quad M_R = \frac{RT'}{R}; \\ M_t^* &= \frac{\Delta t}{C_0 R_0^*}; \quad M_\varphi = \frac{T}{\varphi}; \quad M_I = \frac{W}{I_0^*}. \end{aligned} \quad (8.31)$$

В данном случае впервые имеем дело с двумя индикаторами аналогии (см. стр. 93).

Сетки резисторов можно использовать также для изучения полей, описываемых волновым уравнением (4.24). Использование R -ких сеток для изучения нестационарных полей всегда связано с громоздким циклом последовательных приближений из-за необходимости разбиения времени на интервалы Δt . Вместе с тем такая последовательная и, вообще говоря, длительная процедура решения позволяет относительно просто учитывать изменения свойств материала и граничных условий во время нестационарного процесса.

Необходимость в дискретном представлении времени отпадает, если в сетку сопротивлений ввести реактивные электрические элементы (C, L). В этом случае нестационарное физическое поле моделируется нестационарным полем электрического тока. Процедура моделирования при этом упрощается, однако сама модель становится значительно сложнее.

Для реактивных сеток отпадает необходимость конечно-разностной аппроксимации временных производных. Поэтому для обоснования аналогии используются так называемые *неполные* конечно-разностные формы уравнений математической физики. Так, уравнение Фурье может быть представлено в виде (4.44).

Для моделирования поля, описываемого уравнением Фурье, может быть использована реактивная RC -сетка (общий вид модели, выполненной на RC -сетке, при $m=2$ показан на рис. 8.15, *а*, схема узловой точки при $m=3$ — на рис. 8.16, *а*). Первый закон Кирхгофа для узловой точки RC -сетки (см. табл. 7.1) имеет вид

$$\sum_{i=1}^7 I_i = 0; \quad \sum_{i=1}^6 \frac{\varphi_i - \varphi_0}{R} + C_0 \frac{\partial (0 - \varphi_0)}{\partial \tau} = 0; \quad (8.32)$$

$$\frac{1}{R} \left(\sum_{i=1}^6 \varphi_i - 6\varphi_0 \right) = C_0 \frac{\partial \varphi_0}{\partial \tau}.$$

В отличие от «натурного» времени t время в модели обозначено τ .

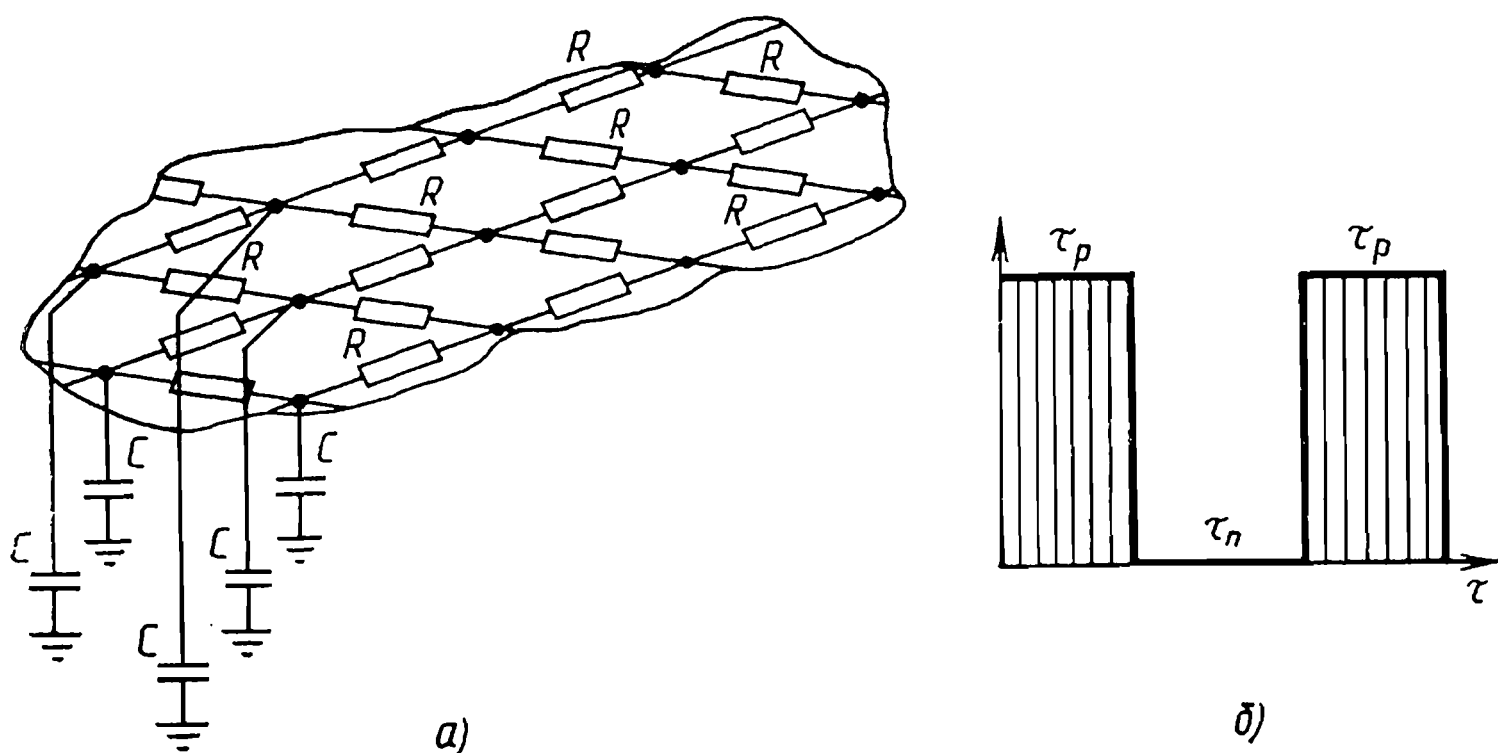


Рис. 8.15. Моделирование уравнения Фурье на RC -сетке:

а — схема RC -сетки; *б* — последовательность работы RC -сетки в режиме с периодизацией

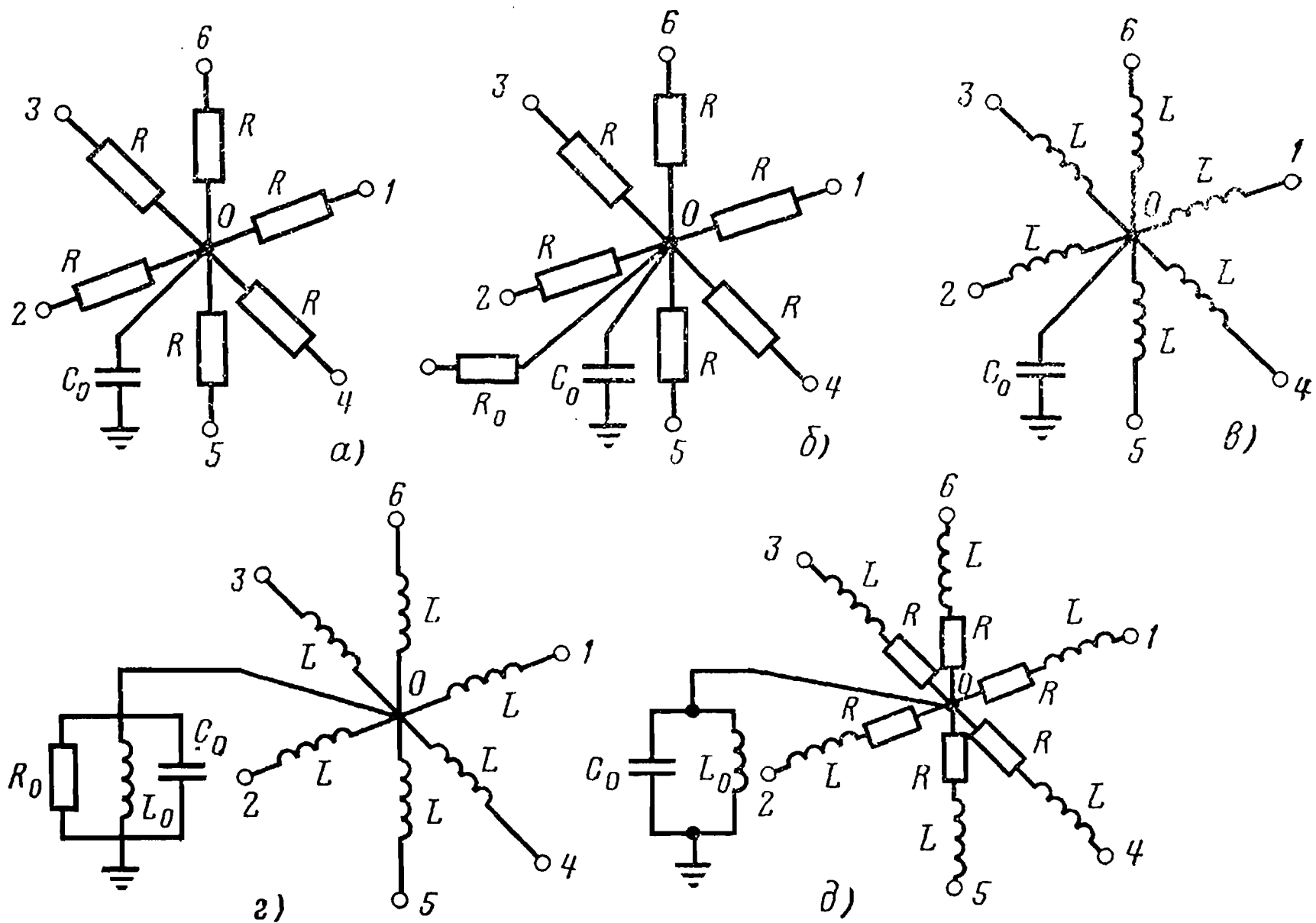


Рис. 8.16. Схемы узлов реактивных сеток сопротивлений:

а — RC -сетка; б — сетка для моделирования полного уравнения теплопроводности; в — LC -сетка; г, д — RLC -сетка

Уравнение (8.32) совпадает с уравнением (4.44) при условии

$$\frac{M_C M_R}{M_\tau} \equiv 1; \quad M_C = \frac{A(\xi_0^r, \Phi_0)}{C_0}; \quad M_R = \frac{h^2}{R}; \quad M_\tau = \frac{t}{\tau}. \quad (8.33)$$

В отличие от коэффициента M_t^* (8.31) коэффициент аналогии M_τ представляет собой истинный масштаб времени. Моделирование может осуществляться в ускоренном ($M_\tau > 1$), замедленном ($M_\tau < 1$), или натуральном ($M_\tau = 1$) масштабе времени.

Для задания начальных условий (потенциалов) во всех узловых точках модели в начальный момент времени используются дополнительные цепи «начальных» потенциалов. С помощью этих цепей обеспечивается начальный заряд конденсаторов. При этом может потребоваться предварительное (до начала моделирования) отключение цепей граничных условий. В момент включения модели строго одновременно должны быть выключены цепи начальных потенциалов и включены цепи граничных условий. Начальные условия могут быть заданы и с помощью подключения конденсаторных цепей к делителю напряжения. Нестационарный процесс изменения потенциала в узлах сетки записывается на шлейфном или электронном осциллографе.

Возможны два режима работы RC -сетки: одноктактный, при котором моделирование осуществляется относительно медленно (цикл моделирования продолжается несколько минут), и с периодизацией решения, при котором моделирование

осуществляется быстро (десятые доли секунды) и повторяется с определенной частотой. В первом случае сетка включает в себя конденсаторы большой емкости и оказывается довольно дорогой и громоздкой. Запись относительно медленно меняющихся напряжений может быть осуществлена с помощью самописцев и шлейфных осциллографов. Во втором случае в сетке используются малые емкости. Многократное повторение решений требует автоматического восстановления граничных и начальных условий, в связи с чем режим работы сетки включает в себя чередующиеся рабочие (τ_p) и подготовительные (τ_n) этапы (рис. 8.15, б). При этом сетка получается простой и дешевой, но существенно возрастает сложность и стоимость вспомогательного оборудования. Запись изменения потенциала осуществляется с помощью электронного осциллографа, причем при использовании приборов с послесвечением на экране может быть получено устойчивое изображение изучаемого процесса.

Для моделирования общего уравнения теплопроводности (4.30) может быть использована RC -сетка с дополнительной цепью для подачи в узловую точку тока I_0^* (рис. 8.16, б). Вывод необходимых уравнений и индикаторов аналогии предоставляется читателю.

Для изучения полей, описываемых волновым уравнением (4.24), может быть использована реактивная LC -сетка (рис. 8.16, в). Конечно-разностная аппроксимация волнового уравнения задается соотношением (4.45). Для узла LC -сетки в силу первого закона Кирхгофа выполняется соотношение

$$\sum_{i=1}^7 I_i = 0; \quad \sum_{i=1}^6 \left(\frac{1}{L} \int_0^{\tau} (\varphi_i - \varphi_0) d\tau + i_0 \right) + C \frac{\partial (0 - \varphi_0)}{\partial \tau} = 0. \quad (8.34)$$

Дифференцируя (8.34) по τ и перенося второй член в правую часть уравнения, легко получить

$$\frac{1}{L} \left(\sum_{i=1}^6 \varphi_i - 6\varphi_0 \right) = C_0 \frac{\partial^2 \varphi_0}{\partial \tau^2}. \quad (8.35)$$

Уравнение (8.35) совпадает с уравнением (4.45) при условии

$$\frac{M_C M_L}{M_\tau^2} \equiv 1; \quad M_L = \frac{\hbar^2}{L}; \quad M_C = \frac{B(\xi_0^r, \Phi_0)}{C_0};$$

$$M_\tau = \frac{t}{\tau}. \quad (8.36)$$

Задание начальных условий, так же как и на RC -сетке, осуществляется начальной подзарядкой конденсаторов сетки, причем в индуктивностях наводятся начальные токи i_{0i} .

Наконец, для изучения поля, описываемого общим уравнением второго порядка (4.25), (4.46), могут быть использованы так называемые RLC -сетки (схемы возможных вариантов RLC -сетки приведены на рис. 8.16, г и 8.16, д).

Для узла сетки, представленного на рис. 8.16, з, выполняются соотношения

$$\sum_{i=1}^9 I_i = 0;$$

$$\sum_{i=1}^6 \left(\frac{1}{L} \int_0^{\tau} (\varphi_i - \varphi_0) d\tau + i_0 \right) + \frac{1}{L_0} \int_0^{\tau} (0 - \varphi_0) d\tau +$$

$$+ i_{00} + \frac{0 - \varphi_0}{R} + C \frac{\partial (0 - \varphi_0)}{\partial \tau} = 0. \quad (8.37)$$

После дифференцирования по τ и группировки членов (8.37) можно переписать в виде

$$\frac{1}{L} \left(\sum_{i=1}^6 \varphi_i - 6\varphi_0 \right) = \frac{1}{L_0} \varphi_0 + \frac{1}{R} \frac{\partial \varphi_0}{\partial \tau} + C \frac{\partial^2 \varphi_0}{\partial \tau^2}. \quad (8.38)$$

Уравнение (8.38) совпадает с уравнением (4.46) при условии

$$\frac{M_L}{M_{L_0}} \equiv 1; \quad \frac{M_L}{M_R M_{\tau}} \equiv 1; \quad \frac{M_L M_C}{M_{\tau}^2} \equiv 1;$$

$$M_L = \frac{L}{h^2}; \quad M_{L_0} = \frac{A(\xi_0^r, \Phi_0)}{1/L_0}; \quad M_R = \frac{B(\xi_0^r, \Phi_0)}{1/R}; \quad (8.39)$$

$$M_C = \frac{D(\xi_0^r, \Phi_0)}{C}; \quad M_{\tau} = \frac{t}{\tau}.$$

Для узла сетки, показанной на рис. 8.16, д, может быть получено соотношение

$$\frac{1}{L} \left(\sum_{i=1}^6 \varphi_i - 6\varphi_0 \right) = \frac{R}{LR_0} \varphi_0 + \left(\frac{1}{R_0} + \frac{C_0 R}{L} \right) \frac{\partial \varphi_0}{\partial \tau} + C_0 \frac{\partial^2 \varphi_0}{\partial \tau^2}, \quad (8.40)$$

совпадающее с (4.46) при условиях

$$M_1 \equiv 1; \quad \frac{M_2}{M_{\tau}} \equiv 1; \quad \frac{M_3}{M_{\tau}^2} \equiv 1;$$

$$M_1 = \frac{h^2 D}{R/R_0}; \quad M_2 = \frac{h^2 B}{L/R_0 + C_0 R};$$

$$M_3 = \frac{h^2 A}{LC_0}; \quad M_{\tau} = \frac{t}{\tau}. \quad (8.41)$$

С помощью RLC -сеток можно исследовать все рассмотренные выше поля. Из (8.38) непосредственно следует, что для моделирования поля, описываемого уравнением Лапласа, достаточно отключить цепи R_0 , L_0 , C_0 (L -сетка). Для моделирования поля, описываемого уравнением Фурье, необходимо отключить цепи L_0 , C_0

и т. д. При этом необходимо помнить, что использование RLC -сеток для моделирования относительно простых полей, описываемых уравнением Лапласа, Пуассона, Фурье, всегда приводит к более сложным и менее точным моделям, чем использование специальных R -, RC -сеток. Это объясняется тем, что индуктивности более громоздки, дороги и неудобны в работе, чем активные сопротивления. Однако преимуществом RLC -сеток является их универсальность.

Можно построить также сеточные аналоговые модели для моделирования полей, описываемых более сложными уравнениями в частных производных. В частности, построены как активные, так и реактивные сетки [8] для исследования полей, описываемых бигармоническим уравнением (4.35).

Вопросы для самоконтроля

1. Чем принципиально отличаются реактивные сетки от резисторных сеток?
2. Нарисуйте схему RC -сетки. Выведите уравнение, описывающее потенциал на RC -сетке. Для исследования каких полей используются RC -сетки?
3. Нарисуйте схему LC -сетки. Выведите уравнение, описывающее потенциал на LC -сетке. Для исследования каких полей используется LC -сетка?
4. Что такое RLC -сетка? Для чего она используется?
5. Рассмотрите различные варианты включения RLC -сетки для моделирования физических полей различной природы.

8.4. ПРИМЕРЫ МОДЕЛИРОВАНИЯ ФИЗИЧЕСКИХ ПОЛЕЙ В ЭЛЕМЕНТАХ ДЛА

Рассмотрим два примера аналоговых моделей физических полей в элементах ДЛА. В качестве изучаемого выберем поле температур как одно из наиболее универсальных и часто встречающихся.

Стационарное температурное поле в охлаждаемой лопатке газовой турбины (сплошная среда). Моделируется температурное поле в лопатке газовой турбины с петлевым каналом охлаждения (см. разд. 4.3). Для его изучения используется модель, выполненная из электропроводной бумаги (рис. 8.17, а). Предполагается постоянство температуры газа ($T_r = \text{const}$). Модель представляет собой лист электропроводной бумаги, вырезанный по форме

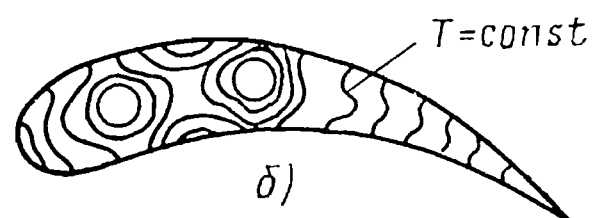
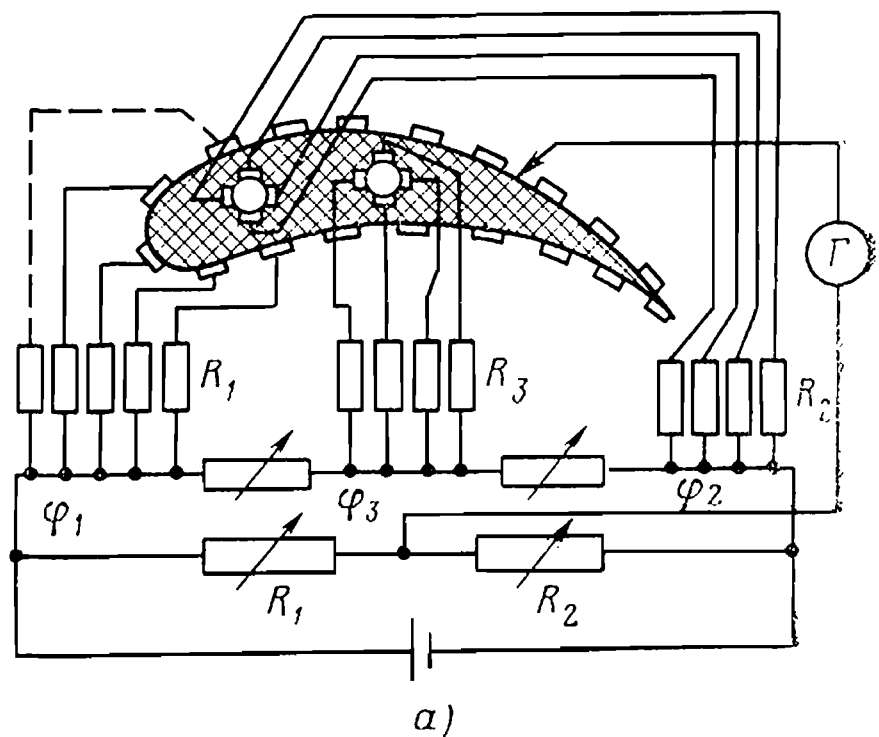


Рис. 8.17. Моделирование температурного поля в охлаждаемой лопатке газовой турбины:

а — схема модели; б — результаты моделирования

сечения лопатки, на границах которого установлены короткие пластинчатые шины. Шины присоединены к системе электропитания через сопротивления R_1, R_2, R_3 . При этом электрод, установленный на границе с промежуточной граничной температурой $T_{ж2}$ ($T_{ж1} < T_{ж2} < T_r$) (рис. 4.3, а), присоединен к промежуточному потенциалу φ_3 ($\varphi_2 < \varphi_3 < \varphi_1$). Распределенные источники тока в модели отсутствуют, поэтому электрический потенциал описывается уравнением Лапласа

$$\Delta\varphi=0. \quad (8.42)$$

На границах модели в силу закона Кирхгофа выполняется равенство подведенных и отведенных токов. Первые из них определяются законом Ома для соответствующих цепей:

$$I_1 = \frac{\varphi_1 - \varphi}{R_1}; \quad I_2 = \frac{\varphi - \varphi_2}{R_2}; \quad I_3 = \frac{\varphi - \varphi_3}{R_3}, \quad (8.43)$$

вторые — соотношениями

$$I_1 = -\sigma\delta\Gamma_1 \frac{\partial\varphi}{\partial n}; \quad I_2 = -\sigma\delta\Gamma_2 \frac{\partial\varphi}{\partial n}; \quad I_3 = -\sigma\delta\Gamma_3 \frac{\partial\varphi}{\partial n}, \quad (8.44)$$

где $\delta\Gamma_i$ — площадь участка границы модели, соответствующего отдельной цепи.

Из соотношений (8.43) и (8.44) следуют выражения для граничных условий:

$$\begin{aligned} \varphi - \sigma\delta\Gamma_1 R_1 \frac{\partial\varphi}{\partial n} &= \varphi_1; \\ \varphi + \sigma\delta\Gamma_2 R_2 \frac{\partial\varphi}{\partial n} &= \varphi_2; \\ \varphi + \sigma\delta\Gamma_3 R_3 \frac{\partial\varphi}{\partial n} &= \varphi_3. \end{aligned} \quad (8.45)$$

Математические описания исследуемого (4.30г), (4.33), (4.34) и модельного (8.42), (8.45) полей идентичны при условии

$$R_1 = \frac{\lambda}{\alpha_r \sigma \delta\Gamma_1}; \quad R_2 = \frac{\lambda}{\alpha_{ж} \sigma \delta\Gamma_2}; \quad R_3 = \frac{\lambda}{\alpha_{ж} \sigma \delta\Gamma_3}. \quad (8.46)$$

На рис. 8.17, б представлены результаты моделирования температурного поля в лопатке газовой турбины.

Рассмотренная модель может быть легко распространена на ряд более общих случаев. Так, например, учет изменения температуры газа по длине межлопаточного канала может быть осуществлен путем замены постоянного потенциала φ_1 набором потенциалов φ_{1i} при подключении цепей R_1 к различным точкам делителя напряжения. Для учета изменения величины коэффициента теплоотдачи α_r вдоль стенки лопатки можно использовать переменные сопротивления R_{1i} .

Нестационарное температурное поле в сложной стенке. Рассматривается температурное поле в стенке ЛА при его входе в атмосферу (см. разд. 4.3, рис. 4.3, в).

В силу сделанных предположений поле в стенке ЛА одномерно

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial T}{\partial r} \right) = \frac{C_p}{\lambda} \frac{\partial T}{\partial t} \quad (8.47)$$

и для i -го узла сетки может быть представлено в виде

$$\sum_j \frac{T_j - T_i}{R_{Tij}} = C_{T_i} \frac{\partial T_i}{\partial t}, \quad j = i - 1, i + 1;$$

$$R_{T_{i,i-1}} = \frac{\Delta r}{4\pi \left(r_i - \frac{\Delta r}{2} \right)^2 \lambda}; \quad (8.48)$$

$$R_{T_{i,i+1}} = \frac{\Delta r}{4\pi \left(r_i + \frac{\Delta r}{2} \right)^2 \lambda};$$

$$C_{T_i} = \pi r_i^2 \Delta r C_p.$$

На границах стенки выполняются условия

$$\begin{aligned} q_{\text{вн}} &= \alpha_{\text{вн}} S_{\text{вн}} (T_{\text{вн}} - T); \\ q_{\text{н}} &= \alpha_{\text{н}} S_{\text{н}} (T - T_{\text{н}}), \end{aligned} \quad (8.49)$$

которые можно записать через эквивалентные сопротивления

$$\begin{aligned} q_{\text{вн}} &= \frac{T_{\text{вн}} - T}{R_{T_{\text{вн}}}}; \quad R_{T_{\text{вн}}} = \frac{1}{\alpha_{\text{вн}} S_{\text{вн}}}; \\ q_{\text{н}} &= \frac{T - T_{\text{н}}}{R_{T_{\text{н}}}}; \quad R_{T_{\text{н}}} = \frac{1}{\alpha_{\text{н}} S_{\text{н}}}. \end{aligned} \quad (8.50)$$

Для моделирования температурного поля может быть предложена модель, выполненная в виде одномерной RC -сетки (рис. 8.18, а). Для определенности каждый из конструктивных слоев стенки разбит на два сферических слоя (т. е. представлен двумя узлами сетки), теплоизолирующий слой — на четыре сферических слоя. Такое разбиение отражает большее тепловое сопротивление изолятора. Каждый сферический слой представлен в модели так называемой T -образной схемой (рис. 8.18, б). При этом сопротивления R_{i1} и R_{i2} совместно представляют собой тепловое сопротивление слоя (в силу сферичности стенки $R_{i1} < R_{i2}$), а C_i — интегральную теплоемкость слоя. В принципе возможно объединение R_{i2} $R_{i+1,1}$, однако при этом утрачивается возможность определения температуры на границе слоев. Сопротивления $R_{\text{вн}}$ и $R_{\text{н}}$ соответствуют эквивалентным тепловым сопротивлениям на границах стенки, напряжение источника $\varphi_{\text{вн}}$ — температуре пограничного слоя

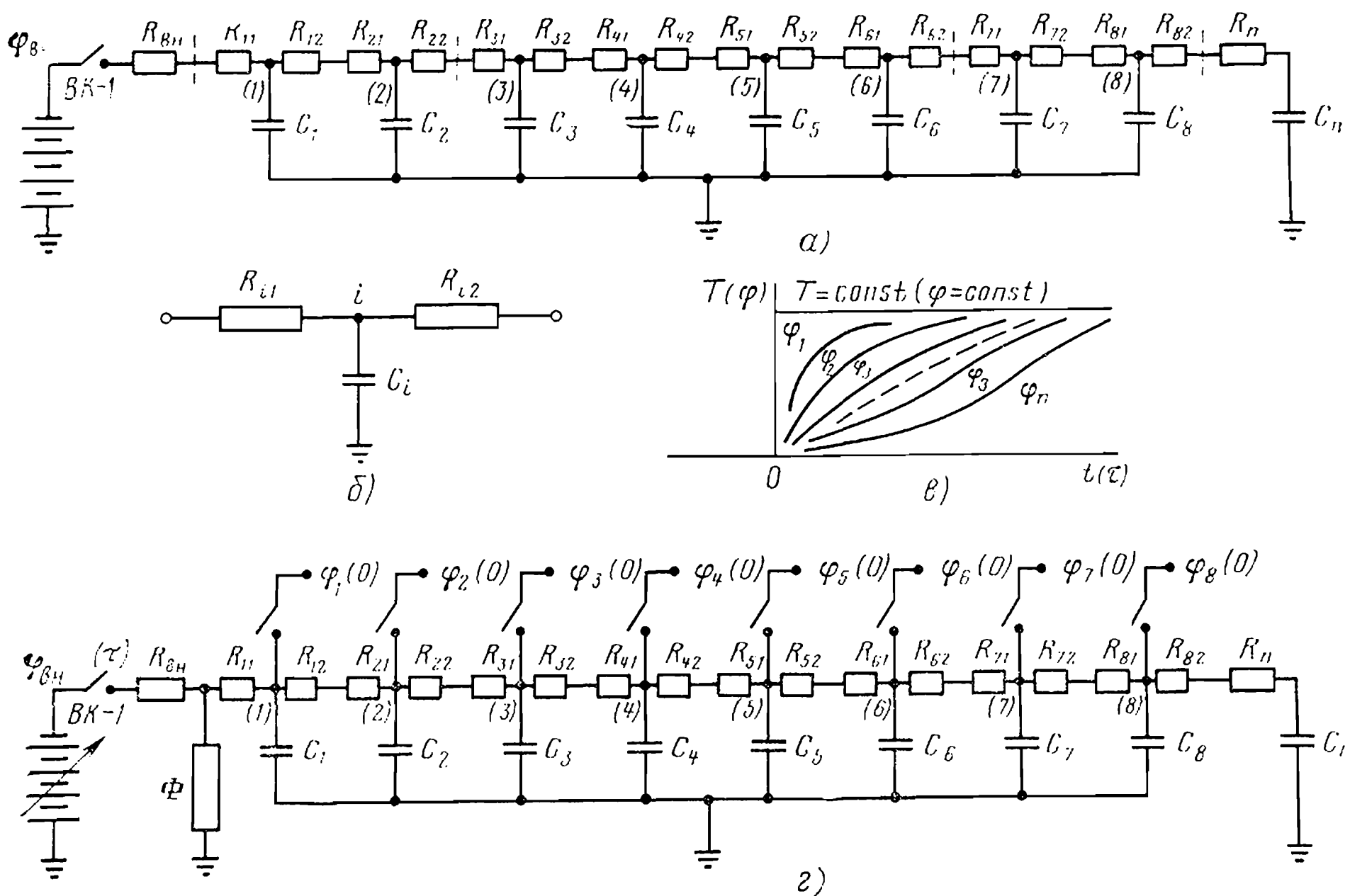


Рис. 8.18. Моделирование нестационарного температурного поля в стенке ЛА при входе в атмосферу:

а — схема модели; б — схема модели для отдельного сферического слоя; в — характер изменения температуры в стенке и потенциала в модели; г — усовершенствованная модель

после скачка температуры $T_{вн}$, емкость $C_{п}$ — интегральной теплоемкости полости. Для учета контактных тепловых сопротивлений между слоями в случае необходимости могут быть введены дополнительные сопротивления на границах конструкционного материала и изолятора или пропорционально увеличены уже имеющиеся сопротивления.

До начала моделирования выключатель ВК-1 разомкнут, все конденсаторы разряжены, и потенциалы во всех узлах сетки равны нулю. После замыкания выключателя ВК-1 потенциалы в сетке начинают меняться, причем темп изменения потенциала в каждой точке определяется емкостью конденсаторов. При достаточной длительности процесса все конденсаторы приобретают номинальные заряды, а потенциалы в узлах сетки выравниваются (рис. 8.18, в). Характер изменения потенциалов соответствует характеру изменения температур в стенке спутника. Рассмотренная модель позволяет решить ряд задач, например, определить время, в течение которого температура внутренней полости не превышает критических значений.

Модель допускает ряд обобщений. Во-первых, вовсе необязательно считать задачу одномерной. Многомерная модель не требует идеальной симметрии и позволяет учесть наличие разнообразных конструктивных элементов и изменение условий теплообмена на

границах. При этом, разумеется, требуется использование двумерных и трехмерных сеток.

Во-вторых, необязательно использовать описанный выше закон изменения внешней температуры и считать начальную температуру в стенке постоянной. Переменная, изменяющаяся по определенному закону внешняя температура может быть задана с помощью программно изменяемого источника напряжения, а начальное распределение температур в стенке — с помощью начальных цепей потенциалов или параллельно включенного источника начального напряжения (рис. 8.18, з). Отключение источников, задающих начальное распределение потенциалов, осуществляется синхронно с включением выключателя ВК-1.

Наконец, можно учесть такой эффект, как излучение части тепла, полученного стенкой, в окружающее пространство. Так как излучаемый тепловой поток пропорционален T^4 , то необходимо включить в соответствующую точку (рис. 8.18, з) нелинейное сопротивление Φ , обеспечивающее утечку тока в соответствии с законом $I = A\Phi^4$.

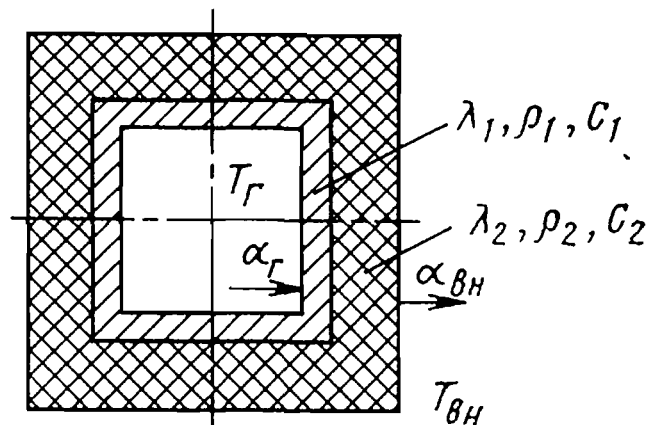


Рис. 8.19. Схема газопровода

Вопросы для самоконтроля

Разработайте аналоговые сеточные модели для моделирования температурного поля в стенке газопровода, показанного на рис. 8.19:

- при условии, что теплоотдача от наружной границы стенки к окружающей среде осуществляется излучением;
- при учете контактных тепловых сопротивлений между слоями стенки.

8.5. НЕЭЛЕКТРИЧЕСКИЕ АНАЛОГОВЫЕ МОДЕЛИ ФИЗИЧЕСКИХ ПОЛЕЙ

Как уже отмечалось, наибольшее распространение получили электрические модели физических полей. Однако в ряде случаев возможно и целесообразно использование аналоговых моделей другой физической природы. Так, наряду с электрическими используются электромагнитные, электростатические, гидродинамические и даже гравитационные модели.

Использование *электромагнитных* моделей основано на изучении поля индуцированных (наведенных) токов в проводящей среде. Поле индуцированных токов в отличие от поля, вызываемого разностью потенциалов на электродах, возбуждается переменным магнитным потоком, пересекающим область моделирования. Силовые линии этого поля не начинаются и не заканчиваются на электродах, а образуют замкнутые контуры или уходят в бесконечность. Эта особенность и определяет преимущества поля индуцированных токов при моделировании *вихревых* (соленоидальных) полей, в частности, магнитных полей и циркуляционных течений жидкости.

Использование *электростатических* моделей основано на изучении наведенных токов в системе электродов при пролете через эту систему точечного зонда с единичным зарядом. По этим токам может быть рассчитано изучаемое поле. Электро-

статические модели можно использовать при исследовании стационарных физических полей различной природы.

Для изучения полей, описываемых уравнением Лапласа, может быть использована *гравитационная модель*, основанная на аналогии изучаемой функции и гравитационного потенциала точек профилированной поверхности. В частности, очень остроумной и эффективной оказалась гравитационная модель для изучения движения заряженных частиц в электростатическом поле (ускорители электронов и ионов). В данном случае аналогия может быть распространена не только на описание поля, но и на описание движения частиц в поле. Можно показать, что тяжелый (например, металлический) шарик на профилированной поверхности без учета трения движется так же, как частица в электростатическом поле при условии, что профиль поверхности в заданном масштабе соответствует распределению потенциала в изучаемом пространстве.

Модель представляет собой растянутую резиновую мембрану, профиль которой изменяется с помощью различных подпорок или оттяжек. Движение блестящего шарика на поверхности мембраны фотографируется в прерывистом (стробоскопическом) свете, что позволяет получить на фотографии пунктирную линию, соответствующую траектории частицы.

Более подробно необходимо остановиться на *гидротепловой* аналогии, предназначенной для исследования нестационарных температурных полей и основанной на изучении переходного процесса распределения жидкости в системе сообщающихся сосудов. В качестве примера гидротепловой модели можно рассмотреть модель температурного поля в стенке ЛА при его входе в атмосферу (см. разд. 4.3, 8.4). Такая модель (рис. 8.20, а) представляет собой ряд последовательно соединенных между собой мензурок (число мензурок соответствует числу сферических слоев, на которые разбита стенка ЛА). В капиллярах, соединяющих мензурки, установлены автоматические краны и дроссели с заданными гидравлическими сопротивлениями R_r . Предполагается, что гидравлическое сопротивление капилляра по сравнению с сопротивлением дросселя пренебрежимо мало.

Вся система мензурок соединена с базовой емкостью, наполненной жидкостью. При моделировании задачи в рассмотренной в разд. 8.4 постановке в начальный момент все мензурки пусты, а краны закрыты. Включение модели осуществляется одновременным открытием всех кранов. При этом начинается перетекание жид-

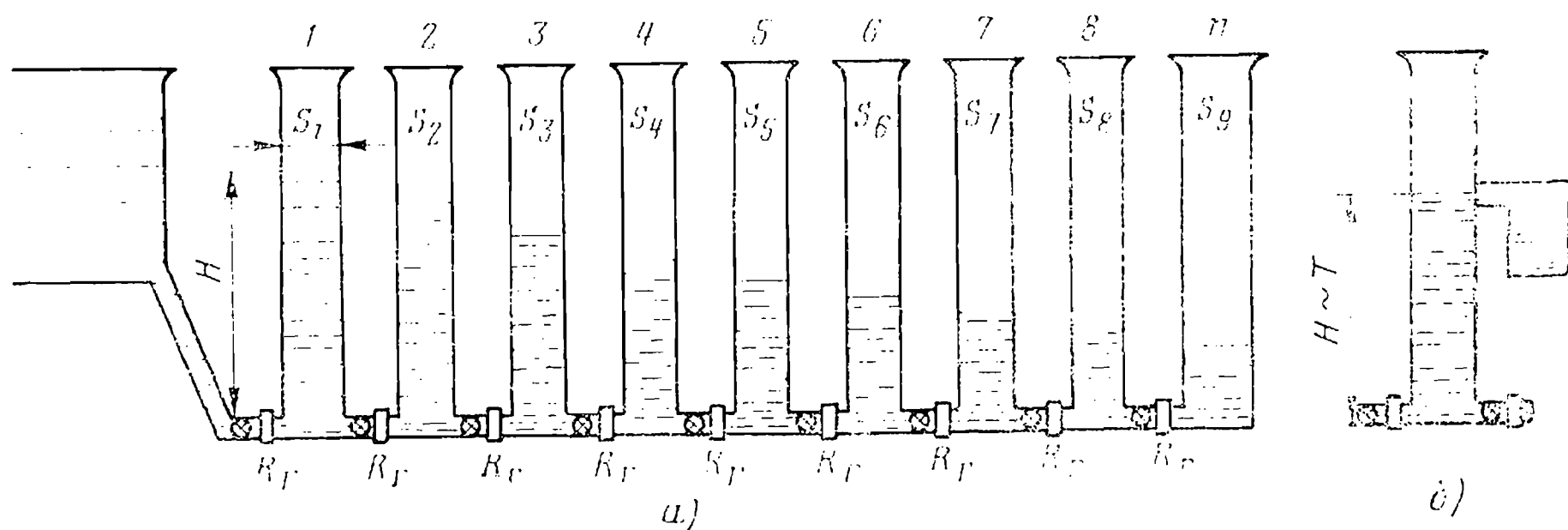


Рис. 8.20. Гидротепловая аналогия:

а — модель для изучения температурного поля в стенке ЛА при входе в атмосферу; б — узел гидротепловой модели для учета теплотрат при фазовом переходе

кости и постепенное заполнение мензурок. Уровень жидкости в базовой емкости остается постоянным.

Приток жидкости Q_{i+} в i -ю мензурку за время $\Delta\tau$ пропорционален напору жидкости $H_{i-1} - H_i$ (за время $\Delta\tau$ этот напор остается практически постоянным), длительности интервала $\Delta\tau$ и обратно пропорционален гидравлическому сопротивлению $R_{r(i-1, i)}$:

$$Q_{i+} = \frac{H_{i-1} - H_i}{R_{r(i-1, i)}} \Delta\tau. \quad (8.51)$$

Отток жидкости Q_{i-} из i -й мензурки

$$Q_{i-} = \frac{H_i - H_{i+1}}{R_{r, (i, i+1)}} \Delta\tau. \quad (8.52)$$

Изменение уровня (высоты) жидкости H_i в i -й мензурке может быть записано в виде

$$\begin{aligned} \Delta H_i &= \frac{Q_{i+} - Q_{i-}}{S_i} = \frac{\Delta\tau}{S_i} \sum_j \frac{H_j - H_i}{R_{rij}}; \\ S_i \frac{\partial H}{\partial \tau} &\simeq S_i \frac{\Delta H_i}{\Delta\tau} = \sum_j \frac{H_j - H_i}{R_{rij}}; \\ j &= i-1, \quad i+1. \end{aligned} \quad (8.53)$$

Уравнение (8.53) аналогично уравнению (8.48). Эти уравнения включают в себя четыре пары сходственных величин:

$$T \rightarrow H, \quad R_T \rightarrow R_r, \quad C_T \rightarrow S, \quad t \rightarrow \tau. \quad (8.54)$$

При этом сечение мензурки S пропорционально интегральной теплоемкости соответствующего сферического слоя стенки, а последняя мензурка моделирует внутреннюю полость ЛА. Распределение температур в стенке моделируется распределением уровней жидкости в мензурках. Читатели сами легко запишут необходимые коэффициенты и индикатор аналогии.

Построенная гидравлическая сетка эквивалентна RC -сетке, показанной на рис. 8.18, *а* (при условии объединения соседних электрических сопротивлений в одно сопротивление). Могут быть построены двумерные и трехмерные гидравлические сетки для моделирования двумерных и трехмерных полей.

Граничные условия первого рода задаются уровнем жидкости в граничных мензурках, второго рода — потоком жидкости, третьего рода — уровнем жидкости в сосуде, соединенном с граничными мензурками через заданное гидравлическое сопротивление (рис. 8.20, *а*). Начальное распределение температур задается начальным распределением жидкости во всех мензурках. Распределенные источники тепла задаются дополнительной подачей жидкости в мензурки.

Процесс моделирования на гидравлической сетке протекает медленно, что позволяет производить отсчет уровней визуально.

Кроме того, моделирование может быть в любой момент остановлено (одновременно автоматически закрываются все краны) для проведения наблюдений или изменения состояния модели и продолжено вновь при открытии кранов.

Гидродинамические модели особенно удобны при решении *нелинейных* задач теплопроводности. Так, нелинейная зависимость теплоемкости от температуры легко моделируется при применении мензурок переменного по высоте сечения или введения вытесняющих тел заданного профиля. Моделирование теплотрат (или тепловыделения) при *фазовом* переходе осуществляется добавлением или сливом необходимого количества жидкости при остановке модели. Может быть построена и непрерывно действующая гидравлическая сетка, позволяющая учесть теплотраты на фазовый переход. Ячейка такой сетки (рис. 8.20, б) должна включать дополнительный сосуд с емкостью, соответствующей затратам тепла на фазовый переход в элементе изучаемого объекта, присоединенный к основной мензурке на высоте, соответствующей температуре фазового перехода. При достижении температуры фазового перехода (соответствующей высоты уровня жидкости) дальнейшее поступление тепла (жидкости) не приводит к повышению температуры (высоты), пока не будет полностью осуществлен фазовый переход (не произойдет заполнение жидкостью дополнительного сосуда).

Недостатком гидравлических сеток является их громоздкость, в силу чего при моделировании приходится ограничиваться небольшим количеством ячеек.

Вопросы для самоконтроля

1. Что является аналогом количества тепла при использовании гидротепловой аналогии?
2. Как учесть при гидротепловой аналогии контактные тепловые сопротивления?
3. Постройте гидротепловые модели для решения задачи, показанной на рис. 8.19.

ГЛАВА 9. АНАЛОГОВОЕ МОДЕЛИРОВАНИЕ ФИЗИЧЕСКИХ СИСТЕМ

9.1 РЕШАЮЩИЕ ЭЛЕМЕНТЫ

Физическими системами (в отличие от физических полей) принято называть объекты, характеризуемые ограниченным набором параметров и описываемые обыкновенными дифференциальными уравнениями. Модели этого типа рассмотрены в разд. 4.2 и связаны в первую очередь с системами автоматического регулирования ДЛА и переходными процессами в их элементах. При аналоговом моделировании физических систем широкое распростране-

ние получили так называемые *структурные* модели, предусматривающие расчленение математического описания на блоки (вплоть до отдельных математических операций) и разработку аналоговых моделей для каждого из них. Соединяя модели блоков между собой, можно собрать единую структурную модель системы. Преимуществом структурного подхода является *универсальность* элементарных блоков, вытекающая из универсальности отдельных операций. Ограниченная номенклатура таких блоков позволяет разработать для их исследования достаточно совершенные (пусть даже сложные) решающие элементы.

Для реализации структурных моделей разработаны и выпускаются промышленностью серийные аналоговые вычислительные машины МН-7, МН-10, МН-14, МН-17, ЭМУ-10, «Аналог», «Электрон», АВК-31 и др.

Все аналоговые вычислительные машины (АВМ) включают в себя *решающие элементы* (блоки) для выполнения математических операций, *блоки питания*, обеспечивающие функционирование машины, *блоки измерения и управления*, а также элементы коммутации решающих блоков между собой.

Основную роль в работе машины играют решающие элементы. Каждый такой элемент совершает над *машинной* переменной (при электрическом моделировании в качестве машинной переменной выступает электрический потенциал) некоторую операцию (преобразование), приводящую к изменению этой переменной. Эта операция характеризуется *передаточной функцией* элемента (рис. 9.1, а):

$$\Phi_{\text{вых}} = F(\Phi_{\text{вх}}). \quad (9.1)$$

Машинные переменные соответствуют исследуемым *натурным* переменным и связаны с ними коэффициентами аналогии. Очевидно, в рамках одной модели эти коэффициенты должны быть одинаковы для всех блоков.

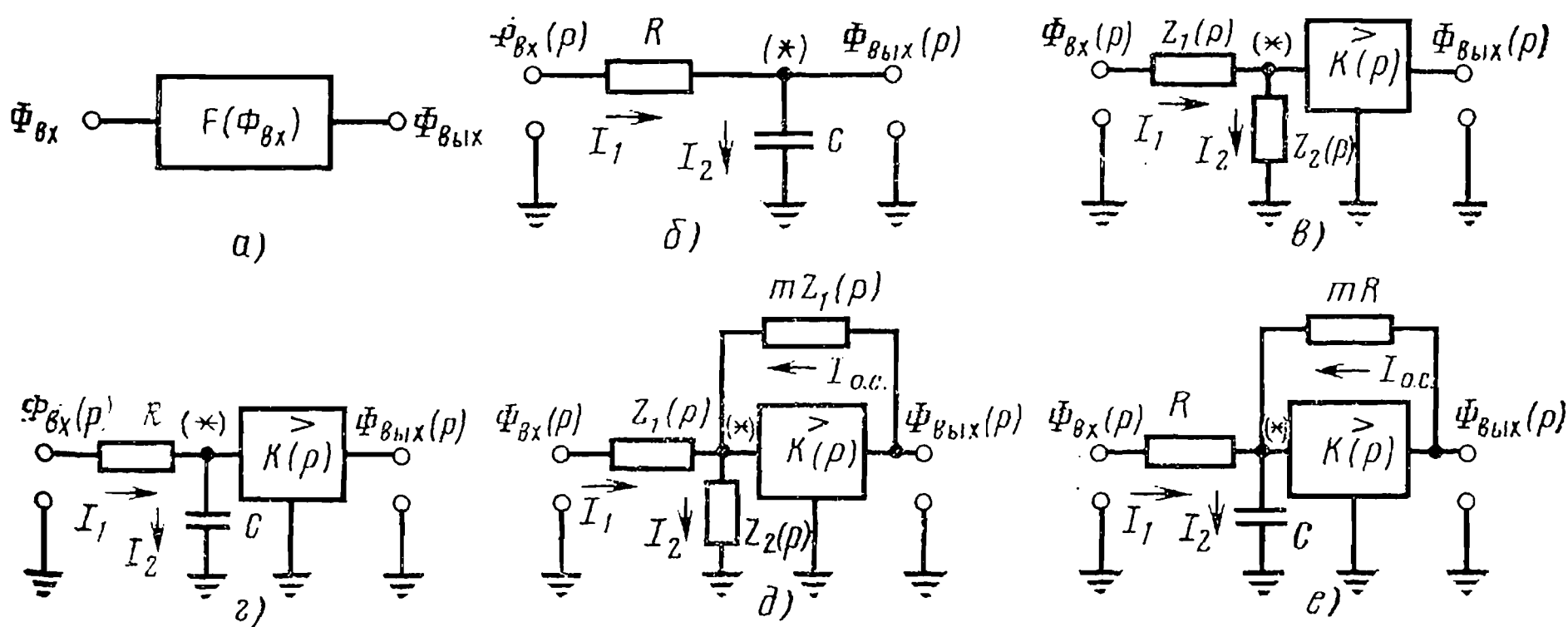


Рис. 9.1. Решающие элементы структурных моделей:

а — общая схема решающего звена; б — интегрирующая цепочка; в, г — решающее звено с операционным усилителем и пассивной цепью на входе; д, е — решающее звено с операционным усилителем, пассивной цепью на входе и положительной обратной связью

Изучение решающих элементов целесообразно начать с изучения работы чисто электрического решающего элемента — интегрирующей цепочки (рис. 9.1, б). Интегрирующая цепочка включает в себя резистор R и конденсатор C . Сопротивление проводов по сравнению с сопротивлением элементов цепи считается пренебрежимо малым. Точка, обозначенная на схеме (*), называется точкой суммирования. Для записи основных соотношений здесь и ниже используется операционное исчисление (см. табл. 7.1). Для точки суммирования в силу законов Ома и Кирхгофа, а также сделанных предположений могут быть записаны соотношения

$$I_1(p) = \frac{\Phi_{\text{вх}}(p) - \Phi^*(p)}{R}; \quad I_2(p) = Cp\Phi^*(p);$$

$$\Phi^*(p) = \Phi_{\text{вых}}(p);$$

$$Cp\Phi_{\text{вых}}(p) = \frac{\Phi_{\text{вх}}(p) - \Phi_{\text{вых}}(p)}{R}.$$
(9.2)

При дополнительном условии

$$\Phi_{\text{вх}}(p) \gg \Phi_{\text{вых}}(p)$$
(9.3)

последнее из соотношений (9.2) принимает вид

$$\Phi_{\text{вых}}(p) = \frac{1}{RCp} \Phi_{\text{вх}}(p)$$
(9.4)

и при переходе к оригиналам (см. табл. 7.1) преобразуется к виду

$$\varphi_{\text{вых}}(\tau) = \frac{1}{RC} \int_0^{\tau} \varphi_{\text{вх}}(\tau) d\tau + \varphi_{\text{вых}}(0).$$
(9.5)

Получен чрезвычайно важный результат: простая электрическая схема осуществляет сложнейшую математическую операцию. Для того чтобы получить (измерить на выходе схемы) любой (в том числе «неберущийся») интеграл, достаточно подать на вход схемы напряжение, моделирующее в некотором масштабе подынтегральную функцию. Существенным недостатком рассмотренной схемы является необходимость значительного понижения напряжения на выходе решающего блока (9.3). Предельные напряжения в АВМ ограничены соображениями техники безопасности. В современных АВМ напряжение может меняться в пределах или ± 100 , или ± 10 В. Поэтому выходное напряжение должно быть очень мало. При этом снижается точность измерений и, что самое главное, практически исключается возможность повторных и тем более многократных операций над машинной переменной (ведь условие (9.3) должно соблюдаться для каждого из последовательно соединенных звеньев). Этот недостаток может быть устранен введением в решающую (интегрирующую) схему *усилителя постоянного тока*, обеспечивающего повышение выходного напряжения и разделяющего вход и выход схемы.

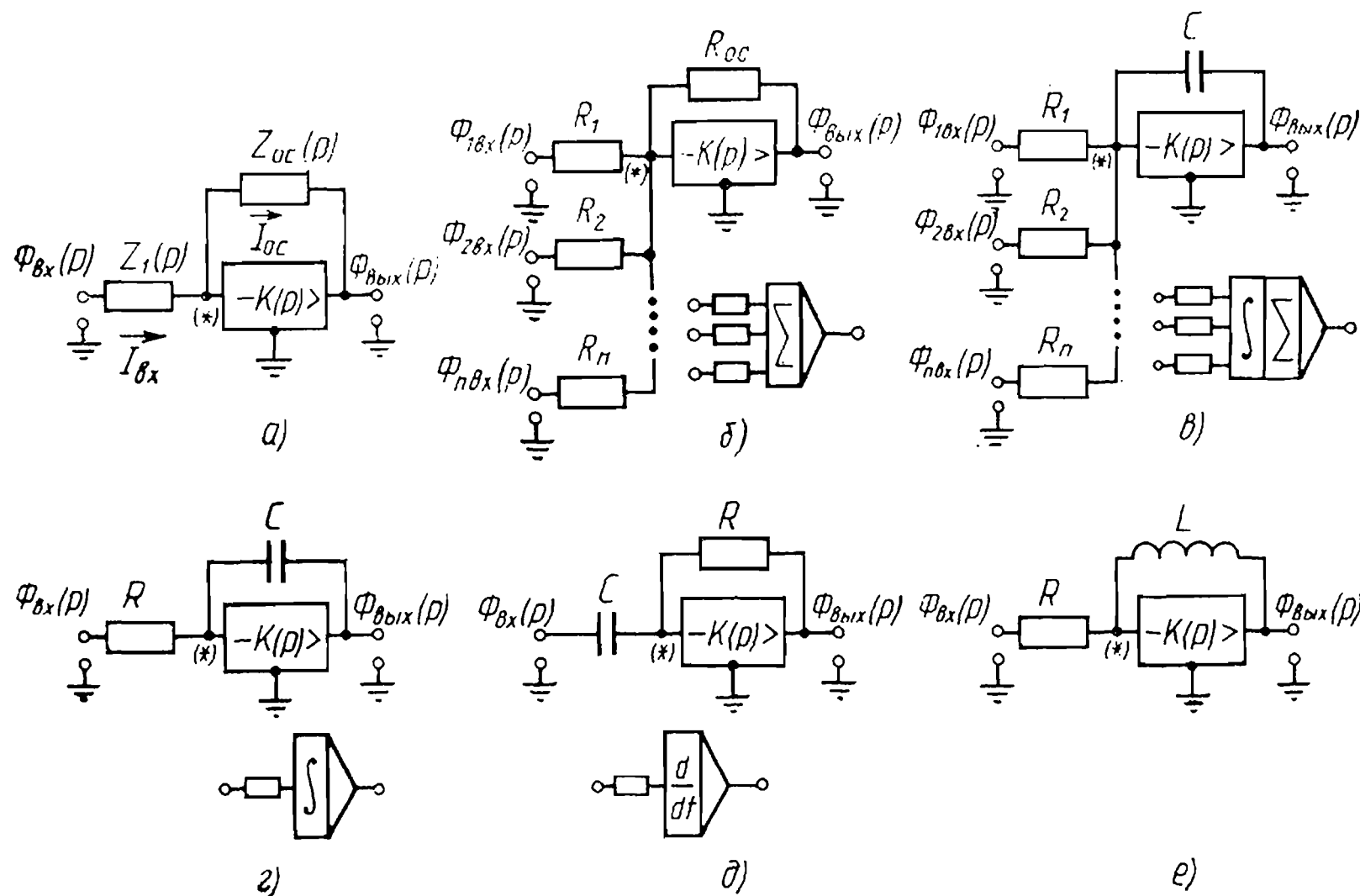


Рис. 9.2. Решающие звенья с операционным усилителем и отрицательной обратной связью:

a — общая схема звена; *б* — суммирующее звено; *в* — интегросуммирующее звено; *г* — интегрирующее звено; *д*, *е* — дифференцирующее звено

В настоящее время существует три типа решающих звеньев с усилителем: звенья с пассивной цепью на входе (рис. 9.1, *в*, *г*), звенья с пассивной цепью на входе и положительной обратной связью (рис. 9.1, *д*, *е*), звенья с отрицательной обратной связью (рис. 9.2).

Решающие звенья с пассивной цепью на входе отличаются от интегрирующей цепочки тем, что точка выхода отделена от точки суммирования (*) усилителем. Передаточная функция такого звена может быть получена с помощью соотношений

$$I_1(p) = \frac{\Phi_{\text{вх}}(p) - \Phi^*(p)}{Z_1(p)}; \quad I_2 = \frac{\Phi^*(p)}{Z_2(p)};$$

$$\Phi^*(p) = \frac{1}{K} \Phi_{\text{вых}}(p); \quad \frac{\Phi_{\text{вх}}(p) - \frac{1}{K} \Phi_{\text{вых}}(p)}{Z_1(p)} = \frac{1}{K} \frac{\Phi_{\text{вых}}(p)}{Z_2(p)}; \quad (9.6)$$

$$\Phi_{\text{вых}}(p) = K \frac{Z_2(p)}{Z_1(p)} \Phi_{\text{вх}}(p) \frac{1}{1 + \frac{Z_2(p)}{Z_1(p)}},$$

где

$\Phi_{\text{вых}}(p) = K \frac{Z_2(p)}{Z_1(p)} \Phi_{\text{вх}}(p)$ — собственно передаточная функция, а множитель

$$\frac{1}{1 + \frac{Z_2(p)}{Z_1(p)}}$$

характеризует погрешность выполнения операции ($Z_1(p) \gg Z_2(p)$). В частности,

если $Z_1(p) = R$, $Z_2(p) = \frac{1}{Cp}$, блок совершает операцию интегрирования (рис. 9.1, з).

$$\Phi_{\text{вых}}(p) \simeq K \frac{1}{RCp} \Phi_{\text{вх}}(p);$$

$$\varphi_{\text{вых}}(\tau) = \frac{1}{RC} \int_0^{\tau} \varphi_{\text{вх}}(\tau) d\tau + \varphi_{\text{вых}}(0). \quad (9.7)$$

Введение положительной обратной связи (рис. 9.1, д, е) позволяет уменьшить погрешность операции. При этом сопротивление обратной цепи выбирается кратным $Z_1(p)$. Для звеньев этого типа выполняются соотношения

$$I_1(p) = \frac{\Phi_{\text{вх}}(p) - \Phi^*(p)}{Z_1(p)}; \quad I_{\text{о.с}}(p) = \frac{\Phi_{\text{вых}}(p) - \Phi^*(p)}{Z_{\text{о.с}}(p)}; \quad I_2(p) = \frac{\Phi^*(p)}{Z_2(p)};$$

$$\Phi^*(p) = \frac{1}{K} \Phi_{\text{вых}}(p);$$

$$Z_{\text{о.с}}(p) = mZ_1(p); \quad (9.8)$$

$$\frac{\Phi_{\text{вх}}(p) - \frac{1}{K} \Phi_{\text{вых}}(p)}{Z_2(p)} + \frac{\Phi_{\text{вых}}(p) - \frac{1}{K} \Phi_{\text{вых}}(p)}{mZ_1(p)} = \frac{1}{K} \frac{\Phi_{\text{вых}}(p)}{Z_2(p)};$$

$$\Phi_{\text{вых}}(p) = K \frac{Z_2(p)}{Z_1(p)} \Phi_{\text{вх}}(p) \frac{1}{\left(1 + \frac{1}{m} - \frac{K}{m}\right) \frac{Z_2(p)}{Z_1(p)} + 1}.$$

При $K = m + 1 = \text{const}$ погрешность операции становится равной нулю. На практике полностью компенсировать погрешность невозможно. Однако введение положительной обратной связи позволяет уменьшить погрешность в десятки раз.

Наибольшее распространение получили решающие блоки (элементы) с операционными усилителями, охваченными *отрицательной обратной связью*. Общая схема такого решающего блока показана на рис. 9.2, а и отличается от всех рассмотренных ранее схем отсутствием пассивной цепи на входе. Направления токов в цепях показаны стрелками. Операционный усилитель имеет нечетное число каскадов усиления, а напряжение на выходе из блока — знак, обратный знаку напряжения на входе (отрицательная обратная связь). Потенциал точки суммирования (*) близок к нулю (точка называется *квазизаземленной*). В силу законов Ома и Кирхгофа можно записать соотношения

$$I_1(p) = \frac{\Phi_{\text{вх}}(p) - \Phi^*(p)}{Z_1(p)}; \quad I_{\text{о.с}}(p) = \frac{\Phi^*(p) - \Phi_{\text{вых}}(p)}{Z_{\text{о.с}}(p)};$$

$$\Phi^*(p) = -\frac{1}{K} \Phi_{\text{вых}}(p); \quad (9.9)$$

$$\Phi_{\text{вых}}(p) \left[\frac{1}{K Z_{\text{о.с}}(p)} + \frac{1}{Z_{\text{о.с}}(p)} + \frac{1}{K Z_1(p)} \right] = -\frac{\Phi_{\text{вх}}(p)}{Z_1(p)};$$

$$\Phi_{\text{вых}}(p) = - \frac{Z_{o.c}(p)}{Z_1(p)} \Phi_{\text{вх}}(p) \frac{1}{1 + \frac{1}{K} \left(1 + \frac{Z_{o.c}(p)}{Z_1(p)} \right)}.$$

При этом собственно передаточная функция определяется выражением

$$\Phi_{\text{вых}}(p) = - \frac{Z_{o.c}(p)}{Z_1(p)} \Phi_{\text{вх}}(p), \quad (9.10)$$

а дополнительная дробь в (9.9) представляет собой погрешность выполнения операции. Выражение (9.10) может быть получено непосредственно, если положить $\Phi^*(p) = 0$.

Звенья с отрицательной обратной связью устойчивы в работе, просты в наладке и регулировке. Важным достоинством этих звеньев является то, что величина коэффициента усиления K входит лишь в выражение погрешности, а не непосредственно в передаточную функцию [(сравнить (9.6), (9.8), (9.9)]. Поэтому специальной стабилизации величины K не требуется. При больших K погрешность выполнения операции невелика.

Решающие элементы различных типов могут быть получены изменением входного сопротивления $Z_1(p)$ и сопротивления цепи обратной связи $Z_{o.c}(p)$.

Суммирование нескольких переменных с постоянными коэффициентами. Схема суммирующего блока представлена на рис. 9.2, б. На том же рисунке показано упрощенное изображение блока. В данном случае

$$\frac{1}{Z_1(p)} = \sum_{i=1}^n \frac{1}{R_i}; \quad Z_{o.c}(p) = R_{o.c} \quad (9.11)$$

и уравнение (9.10) принимает вид

$$\Phi_{\text{вых}}(p) = - \sum_{i=1}^n \frac{R_{o.c}}{R_i} \Phi_{\text{вх}i}(p) \quad (9.12)$$

или после перехода к оригиналам

$$\varphi_{\text{вых}}(\tau) = - \sum_{i=1}^n \frac{R_{o.c}}{R_i} \varphi_{\text{вх}i}(\tau). \quad (9.13)$$

Таким образом, данный блок осуществляет суммирование напряжений с постоянными коэффициентами $R_{o.c}/R_i$. Поскольку погрешность операции растет с увеличением числа входных цепей, на одном сумматоре можно суммировать не более 6—7 величин.

Частными случаями рассмотренной операции являются:

суммирование величин без коэффициентов ($R_{o.c} = R_1 = R_2 = \dots = R_n$);

умножение величины на константу ($n=1, R_{o.c} \neq R_1$);

изменение знака величины ($n=1, R_{o.c} = -R_1$).

Интегрирование суммы переменных с постоянными коэффициентами. Схема интегросуммирующего (суммирующеинтегрирующего) блока и его упрощенное изображение показаны на рис. 9.2, в и отличаются от схемы суммирующего блока заменой резистора $R_{o.c}$ конденсатором C . В данном случае

$$\frac{1}{Z_1(p)} = \sum_{i=1}^n \frac{1}{R_i}; \quad Z_{o.c}(p) = \frac{1}{Cp}. \quad (9.11a)$$

После подстановки (9.11a) уравнение (9.10) принимает вид

$$\Phi_{\text{вых}}(p) = - \sum_{i=1}^n \frac{1}{R_i C p} \Phi_{\text{вх}i}(p) \quad (9.12a)$$

или после перехода к оригиналам

$$\varphi_{\text{вых}}(\tau) = - \sum_{i=1}^n \frac{1}{R_i C} \int_0^{\tau} \varphi_{\text{вх}i}(\tau) d\tau + \varphi_{\text{вых}}(0). \quad (9.13a)$$

Так же как и в случае суммирующего звена, $n \leq 7$. Частным случаем рассмотренной операции является простое интегрирование ($n=1$) (рис. 9.2, з).

Набор суммирующих и интегрирующих блоков составляет минимальный набор, входящий в состав каждой структурной аналоговой машины.

Дифференцирование переменной. Схема дифференцирующего блока и его упрощенное изображение показаны на рис. 9.2, д. Для данного блока

$$Z_1(p) = \frac{1}{Cp}; \quad Z_{o.c}(p) = R; \quad (9.11б)$$

$$\Phi_{\text{вых}}(p) = -RCp\Phi_{\text{вх}}(p); \quad (9.12б)$$

$$\varphi_{\text{вых}}(\tau) = -RC \frac{d\varphi_{\text{вх}}(\tau)}{d\tau}. \quad (9.13б)$$

Приведенные на рис. 9.2, б, в, з схемы решающих элементов наиболее рациональны. Однако могут быть предложены и иные схемы. Так, например, дифференцирующее звено может быть выполнено с индуктивностью (рис. 9.2, е)

$$Z_1(p) = R; \quad Z_{o.c}(p) = Lp; \quad \Phi_{\text{вых}}(p) = -\frac{Lp}{R} \Phi_{\text{вх}}(p);$$

$$\varphi_{\text{вых}}(\tau) = -\frac{L}{R} \frac{d\varphi_{\text{вх}}(\tau)}{d\tau}. \quad (9.14)$$

Усложняя входные цепи и цепи обратных связей, можно получить решающие элементы для осуществления более сложных операций. Ряд таких звеньев показан в табл. 9.1. Так, для инерционного звена выполняются соотношения

$$Z_1(p) = R_1; \quad \frac{1}{Z_{o.c}(p)} = \frac{1}{R_2} + C_2 p; \quad (9.11в)$$

Тип звена	Дифференциальное уравнение	Схема звена	Передаточная функция	Вид передаточной функции
Аперийное (инерционное) звено	$a \frac{dx_2(t)}{dt} + x_2(t) = kx_1(t)$		$\frac{\Phi_{BIX}(p)}{\Phi_{BX}(p)} = - \frac{R_2}{R_1 R_2 C_2 p + 1}$	
Колебательное звено	$a_2 \frac{d^2 x_2(t)}{dt^2} + a_1 \frac{dx_2(t)}{dt} + x_2(t) = kx_1(t)$		$\frac{\Phi_{BIX}(p)}{\Phi_{BX}(p)} = - \frac{C_1}{C_2} \times \frac{1}{L_1 C_1 p^2 + R_1 C_1 p + 1}$	
Форсирующее (усиливающее) звено	$x_2(t) = k \left(a \frac{dx_1(t)}{dt} + x_1(t) \right)$		$\frac{\Phi_{BIX}(p)}{\Phi_{BX}(p)} = - \frac{R_2}{R_1} (R_1 C_1 p + 1)$	

$$\Phi_{\text{ВЫХ}}(p) = - \frac{1}{R_1 \left(\frac{1}{R_2} + C_2 p \right)} \Phi_{\text{ВХ}}(p); \quad (9.12\text{В})$$

$$\frac{R_1}{R_2} (R_2 C_2 p + 1) \Phi_{\text{ВЫХ}}(p) = - \Phi_{\text{ВХ}}(p);$$

$$R_1 C_2 \frac{d\varphi_{\text{ВЫХ}}(\tau)}{d\tau} + \frac{R_1}{R_2} \varphi_{\text{ВЫХ}}(\tau) = - \varphi_{\text{ВХ}}(\tau). \quad (9.13\text{В})$$

Данное звено решает дифференциальное уравнение первого порядка относительно $\varphi_{\text{ВЫХ}}(\tau)$. Связь $\varphi_{\text{ВХ}}(\tau)$ и $\varphi_{\text{ВЫХ}}(\tau)$, соответствующая (9.13, в), характеризует работу любого элемента ДЛА, обладающего инерцией (привод створок регулируемого сопла, насос топлива и т. д.).

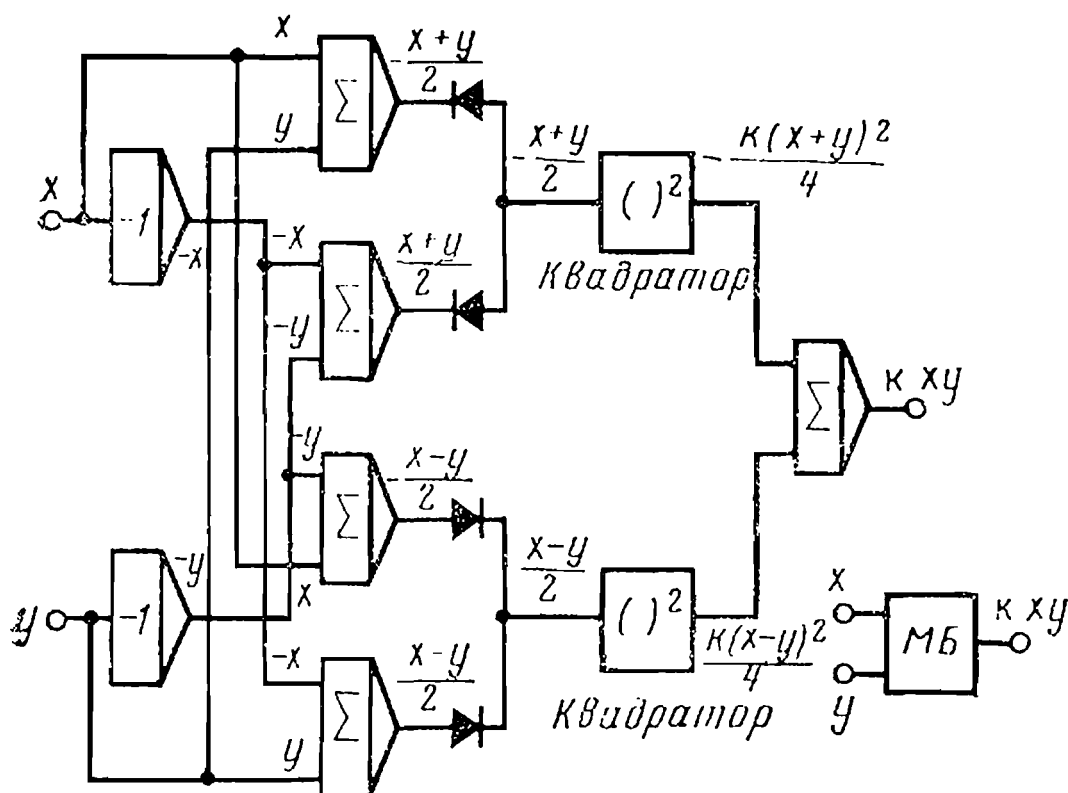


Рис. 9.3. Схема множительного блока

Обычно аналоговые машины включают в себя ряд элементов для моделирования нелинейных процессов, в том числе стандартных нелинейностей САР («зона нечувствительности», «ограничения», «люфт» и т. д.). Нелинейными для этих машин являются также такие простые математические операции, как деление и умножение. При построении множительных устройств, как правило, используется соотношение

$$z = xy = \frac{1}{4} [(x+y)^2 - (x-y)^2]. \quad (9.15)$$

На рис. 9.3 представлена структура множительного блока (МБ) для реализации соотношения (9.15), включающего в себя ряд сумматоров, функциональные преобразователи (квадраторы) и диоды, обеспечивающие выбор полусуммы и полуразности умножаемых величин с соответствующим знаком.

Вопросы для самоконтроля

1. На чем основана идея создания структурных моделей?
2. Что называется решающим элементом?
3. Какие типы решающих элементов Вам известны?
4. Нарисуйте схему интегрирующей цепочки. Напишите для нее основные соотношения.
5. Для чего используются усилители постоянного тока?
6. Нарисуйте схемы сумматора, интегратора и интегросумматора с отрицательной обратной связью и запишите передаточные функции для этих звеньев.

9.2. РЕШЕНИЕ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ НА АНАЛоговых ВЫЧИСЛИТЕЛЬНЫХ МАШИНАХ

Математическая модель физической системы может быть задана в виде одного дифференциального уравнения n -го порядка (4.10) или в виде системы дифференциальных уравнений первого порядка (4.11). Обе формулировки эквивалентны друг другу. Ниже для определенности рассматриваются преимущественно модели в виде одного дифференциального уравнения n -го порядка.

При аналоговом решении обыкновенных дифференциальных уравнений используется метод *понижения порядка производной*, предполагающий разрешение уравнения относительно старшей производной и последовательное интегрирование получаемых результатов. Процесс моделирования состоит из следующих этапов:

- составление схемы структурной модели;
- составление машинных уравнений, выбор масштабов моделирования, параметров решающих блоков;
- задание начальных условий;
- собственно процесс моделирования, измерение и запись потенциалов в модели;
- интерпретация зависимостей для потенциалов как зависимостей для исследуемых величин.

Процедуру решения дифференциального уравнения удобно рассмотреть на конкретном примере уравнения 4-го порядка:

$$a_4 \frac{d^4 x(t)}{dt^4} + a_3 \frac{d^3 x(t)}{dt^3} + a_2 \frac{d^2 x(t)}{dt^2} + a_1 \frac{dx(t)}{dt} + a_0 x(t) = b f(t). \quad (9.16)$$

Уравнение (9.16) может быть разрешено относительно старшей (4-й производной)

$$\frac{d^4 x(t)}{dt^4} = \beta f(t) - a_0 x(t) - a_1 \frac{dx(t)}{dt} - a_2 \frac{d^2 x(t)}{dt^2} - a_3 \frac{d^3 x(t)}{dt^3}; \quad (9.17)$$

$$\beta = \frac{b}{a_4}; \quad \alpha_0 = \frac{a_0}{a_4}; \quad \alpha_1 = \frac{a_1}{a_4}; \quad \alpha_2 = \frac{a_2}{a_4}; \quad \alpha_3 = \frac{a_3}{a_4}.$$

В АВМ абстрактная переменная $x(t)$ моделируется электрическим напряжением: *машинной переменной* $\varphi(\tau)$. Если напряжение, моделирующее старшую производную ($\varphi_{xIV}(\tau)$), известно, напряжения, моделирующие младшие производные и функцию, могут быть получены интегрированием. При этом переход от напряжения, моделирующего старшую производную, к функции осуществляется цепочкой четырех интегрирующих блоков (рис. 9.4, а).

Значение старшей производной может быть получено суммированием всех величин, входящих в правую часть уравнения (9.17). В АВМ эта операция осуществляется с помощью суммирующего блока (рис. 9.4, а), к которому подаются потенциалы, моделирующие соответствующие производные, и напряжение $\varphi_f(\tau)$, моделирующее возмущающую функцию $f(\tau)$.

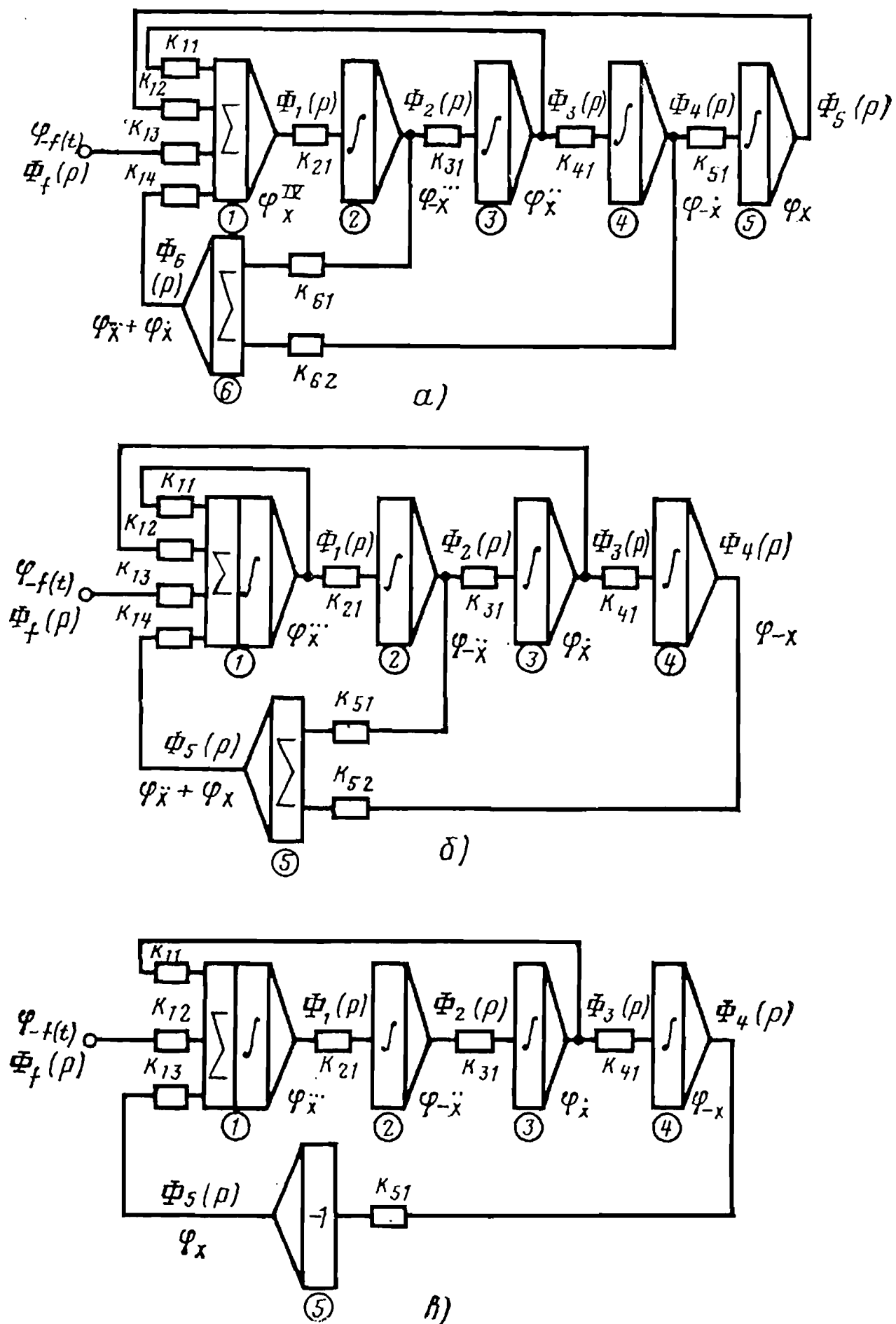


Рис. 9.4. Структурная модель для решения обыкновенного дифференциального уравнения 4-го порядка:

а — при использовании суммирующего блока; б — при использовании интегросуммирующего блока; в — при наличии в уравнении нулевых коэффициентов

При составлении структурной схемы необходимо помнить, что каждый решающий блок меняет знак машинной переменной на обратный. Поэтому, если напряжение ($\varphi_{xiv}(\tau)$), моделирующее старшую производную, положительно, то полученное в результате его интегрирования напряжение $\varphi_{\ddot{x}}(\tau)$ отрицательно (является аналогом $-\ddot{x}(t)$) и т. д. Это обстоятельство необходимо учитывать при суммировании потенциалов. Суммирующий блок также меняет знак напряжений и, следовательно, к сумматору нужно подать потенциалы, имеющие знаки, *обратные* знакам соответствующих членов уравнения (9.17). Поэтому в схеме на рис. 9.4, а предусмотрен промежуточный сумматор, играющий роль инвертора.

Суммирующий и первый интегрирующий блоки могут быть заменены одним интегросуммирующим блоком (рис. 9.4, б). Такая замена позволяет сократить общее число решающих блоков, правда при этом теряется возможность измерения величины потенциала, моделирующего четвертую производную $\varphi_{xIV}(\tau)$,

При исследовании дифференциальных уравнений, содержащих нулевые коэффициенты при некоторых членах, суммирование осуществляется только для членов с ненулевыми коэффициентами. В качестве примера на рис. 9.4, в приведена структурная схема для решения уравнения

$$a_4 \frac{d^4 x(t)}{dt^4} + a_1 \frac{dx(t)}{dt} + a_0 x(t) = b f(t). \quad (9.18)$$

Читателю рекомендуется внимательно изучить эту схему.

Величины k_{ij} представляют собой передаточные функции соответствующих блоков. Индекс i характеризует номер блока, j — номер входа в блок. Для интегратора и сумматора эти функции могут быть выражены следующим образом:

$$k_{ij} = \frac{1}{R_{ij}C_i}; \quad k_{ij} = \frac{R_{o.c.i}}{R_{ij}}. \quad (9.19)$$

Для определения коэффициентов аналогии (масштабов моделирования) и конкретных параметров цепей решающих блоков необходимо составить *машинное уравнение* структурной модели и установить аналогию между ним и исследуемым дифференциальным уравнением. Для структурной модели, показанной на рис. 9.4, а, можно записать следующие машинные уравнения отдельных блоков (см. табл. 7.1):

$$\begin{aligned} 1) \quad & \Phi_1(p) = -[k_{1,1}\Phi_3(p) + k_{1,2}\Phi_5(p) - k_{1,3}\Phi_f(p) + k_{1,4}\Phi_6(p)]; \\ 2) \quad & \Phi_2(p) = -\frac{1}{p} k_{2,1}\Phi_1(p); \quad \Phi_1(p) = -\frac{1}{k_{2,1}} p\Phi_2(p) = -\frac{1}{k_{2,1}k_{3,1}k_{4,1}k_{5,1}} \times \\ & \times p^4\Phi_5(p); \\ 3) \quad & \Phi_3(p) = -\frac{1}{p} k_{3,1}\Phi_2(p); \quad \Phi_2(p) = -\frac{1}{k_{3,1}} p\Phi_3(p) = -\frac{1}{k_{3,1}k_{4,1}k_{5,1}} \times \\ & \times p^3\Phi_5(p); \\ 4) \quad & \Phi_4(p) = -\frac{1}{p} k_{4,1}\Phi_3(p); \quad \Phi_3(p) = -\frac{1}{k_{4,1}} p\Phi_4(p) = \frac{1}{k_{4,1}k_{5,1}} p^2\Phi_5(p); \\ 5) \quad & \Phi_5(p) = -\frac{1}{p} k_{5,1}\Phi_4(p); \quad \Phi_4(p) = -\frac{1}{k_{5,1}} p\Phi_5(p); \\ 6) \quad & \Phi_6(p) = -[k_{6,1}\Phi_2(p) + k_{6,2}\Phi_4(p)] = \frac{k_{6,1}}{k_{3,1}k_{4,1}k_{5,1}} p^3\Phi_5(p) + \frac{k_{6,2}}{k_{5,1}} \times \\ & \times p\Phi_5(p). \end{aligned} \quad (9.20)$$

После подстановки соотношений для блоков (2)—(6) в машинное уравнение суммирующего блока (1) и ряда дополнительных преобразований можно получить

$$p^4 \Phi_5(p) = k_{1,3} k_{2,1} k_{3,1} k_{4,1} k_{5,1} \Phi_f(p) - k_{1,2} k_{2,1} k_{3,1} k_{4,1} k_{5,1} \Phi_5(p) - \\ - k_{1,4} k_{6,2} k_{2,1} k_{3,1} k_{4,1} p \Phi_5(p) - k_{1,1} k_{2,1} k_{3,1} p^2 \Phi_5(p) - k_{1,4} k_{6,1} k_{2,1} p^3 \Phi_5(p). \quad (9.21)$$

При переходе к оригиналам (9.21) принимает вид

$$\frac{d^4 \varphi(\tau)}{d\tau^4} = k_{1,3} k_{2,1} k_{3,1} k_{4,1} k_{5,1} \varphi_f(\tau) - k_{1,2} k_{2,1} k_{3,1} k_{4,1} k_{5,1} \varphi(\tau) - \\ - k_{1,4} k_{6,2} k_{2,1} k_{3,1} k_{4,1} \frac{d\varphi(\tau)}{d\tau} - k_{1,1} k_{2,1} k_{3,1} \frac{d^2 \varphi(\tau)}{d\tau^2} - k_{1,4} k_{6,1} k_{2,1} \frac{d^3 \varphi(\tau)}{d\tau^3}. \quad (9.22)$$

Уравнения (9.17)—(9.22) совпадают по форме и включают в себя три пары сходственных величин:

$$x(t) \rightarrow \varphi(\tau); \quad f(t) \rightarrow \varphi_f(\tau); \quad t \rightarrow \tau. \quad (9.23)$$

В соответствии с этим вводится три коэффициента аналогии (масштаба моделирования):

$$M_x = \frac{x(t)}{\varphi(\tau)}; \quad M_f = \frac{f(t)}{\varphi_f(\tau)}; \quad M_t = \frac{t}{\tau}. \quad (9.24)$$

Заменяв с помощью (9.24) машинные переменные в (9.22) натурными, можно записать *машинное* уравнение в виде

$$\frac{d^4 x(t)}{dt^4} = k_{1,3} k_{2,1} k_{3,1} k_{4,1} k_{5,1} \frac{M_x}{M_f M_t^4} f(t) - k_{1,2} k_{2,1} k_{3,1} k_{4,1} k_{5,1} \frac{1}{M_t^4} x(t) - \\ - k_{1,4} k_{6,2} k_{2,1} k_{3,1} k_{4,1} \frac{1}{M_t^3} \frac{dx(t)}{dt} - k_{1,1} k_{2,1} k_{3,1} \frac{1}{M_t^2} \frac{d^2 x(t)}{dt^2} - \\ - k_{1,4} k_{6,1} k_{2,1} \frac{1}{M_t} \frac{d^3 x(t)}{dt^3}. \quad (9.25)$$

Уравнение (9.25) полностью совпадает с уравнением (9.17) при условиях

$$k_{1,3} k_{2,1} k_{3,1} k_{4,1} k_{5,1} \frac{M_x}{M_f M_t^4} \equiv \beta; \\ k_{1,4} k_{6,1} k_{2,1} \frac{1}{M_t} \equiv \alpha_3; \quad k_{1,1} k_{2,1} k_{3,1} \frac{1}{M_t^2} \equiv \alpha_2; \\ k_{1,4} k_{6,2} k_{2,1} k_{3,1} k_{4,1} \frac{1}{M_t^3} \equiv \alpha_1; \\ k_{1,2} k_{2,1} k_{3,1} k_{4,1} k_{5,1} \frac{1}{M_t^4} \equiv \alpha_0. \quad (9.26)$$

Соотношения (9.26) представляют собой индикаторы аналогии.

После подстановки в (9.26) соотношений (9.19) получаем

$$\begin{aligned} \frac{R_{0.c1}M_x}{R_{1,3}R_{2,1}R_{3,1}R_{4,1}R_{5,1}C_2C_3C_4C_5M_t^4M_f} &= \beta; \\ \frac{R_{0.c1}R_{0.c6}}{R_{1,4}R_{6,1}R_{2,1}C_2M_t} &= \alpha_3; \\ \frac{R_{0.c1}}{R_{1,1}R_{2,1}R_{3,1}C_2C_3M_t^2} &= \alpha_2; \\ \frac{R_{0.c1}R_{0.c6}}{R_{1,4}R_{6,2}R_{2,1}R_{3,1}R_{4,1}C_2C_3C_4M_t^3} &= \alpha_1, \\ \frac{R_{0.c1}}{R_{1,2}R_{2,1}R_{3,1}R_{4,1}R_{5,1}C_2C_3C_4C_5M_t^4} &= \alpha_0. \end{aligned} \quad (9.27)$$

Соотношения (9.27) позволяют выбрать коэффициенты аналогии (масштабы моделирования) и параметры цепей решающих блоков. При этом соотношений (9.27) всегда меньше, чем определяемых по ним величин (в данном случае 19 величин связаны пятью соотношениями), что обеспечивает известный произвол в выборе параметров. При этом следует учитывать некоторые дополнительные ограничения. Так, например, емкости конденсаторов, а также сопротивления цепей обратных связей обычно фиксированы или изменяются ступенчато.

Возможен другой подход к построению структурной схемы аналоговой модели. Одно дифференциальное уравнение высокого порядка можно заменить системой дифференциальных уравнений первого порядка. В частности, дифференциальное уравнение 4-го порядка (9.16) можно свести к такой системе с помощью вспомогательных переменных x_1, x_2, x_3 :

$$\begin{aligned} \frac{dx}{dt} &= x_1; \quad \frac{dx_1}{dt} = x_2; \quad \frac{dx_2}{dt} = x_3; \\ \frac{dx_3}{dt} &= \beta f(t) - \alpha_0 x - \alpha_1 x_1 - \alpha_2 x_2 - \alpha_3 x_3. \end{aligned} \quad (9.28)$$

Структурная модель для решения системы уравнений (9.28), состоящая из одного интегросуммирующего и трех интегрирующих блоков, показана на рис. 9.5. Легко видеть, что эта модель полностью совпадает с моделью для исследования уравнения (9.17), показанной на рис. 9.4, б. Полностью совпадут и машинные уравнения.

Случай построения структурной модели для изучения системы двух дифференциальных уравнений разных порядков будет рассмотрен в разд. 9.3.

Напряжение в структурных аналоговых моделях может меняться в пределах $-100 \text{ В} \leq \varphi \leq +100 \text{ В}$ (в некоторых АВМ $-10 \text{ В} \leq \varphi \leq +10 \text{ В}$). Выход машинных переменных за эти границы недопустим, так как избыточные напряжения просто срезаются, что приводит к неизбежным ошибкам решения. С другой стороны, точность решения возрастает с увеличением напряжения. Поэтому наилучшими являются такие масштабы, при которых максимальные значения машинных переменных будут близки к 100 В. Если мак-

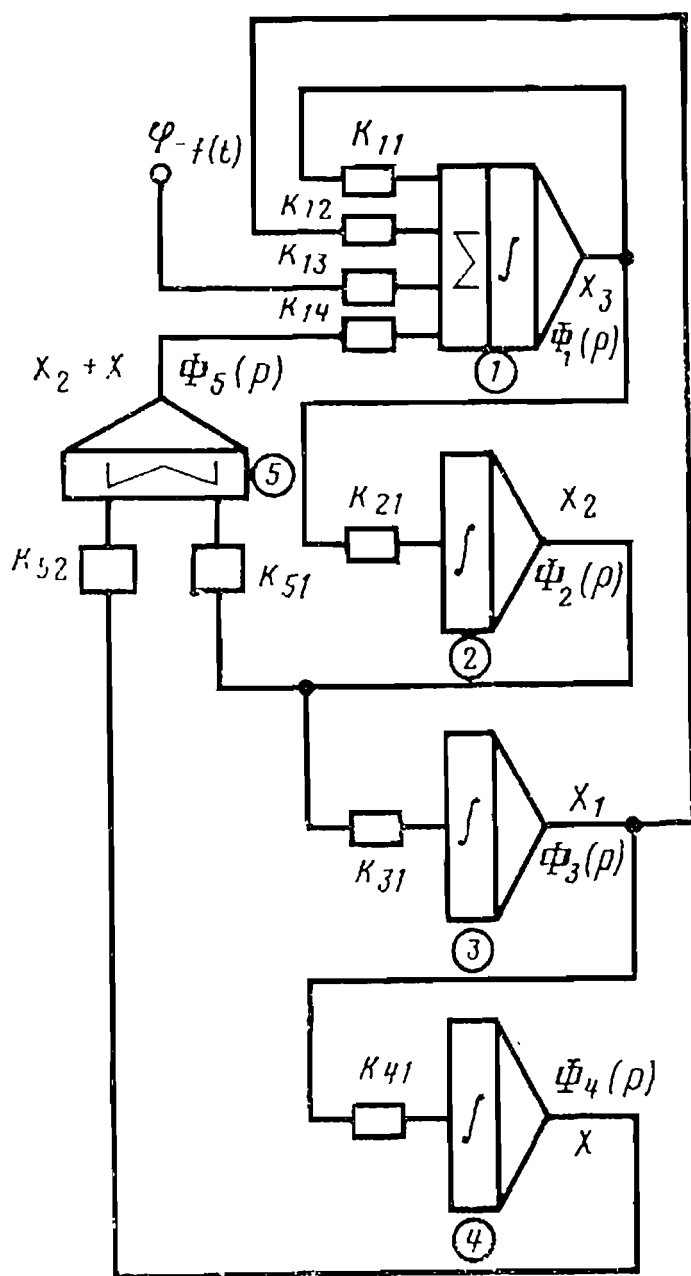
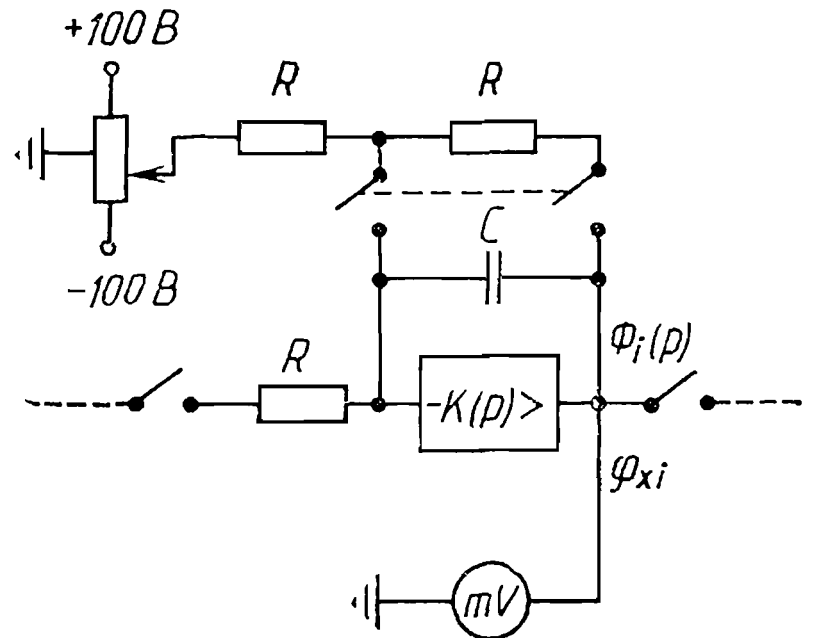


Рис. 9.5. Структурная модель для решения системы дифференциальных уравнений 1-го порядка

Рис. 9.6. Схема задания начальных условий для структурных моделей



симальные значения $x(t)$ и $f(t)$ известны, выбирают следующие масштабы моделирования:

$$M_x = \frac{x_{\max}(t)}{100}; \quad M_f = \frac{f_{\max}(t)}{100}. \quad (9.29)$$

Если значения $x_{\max}(t)$, $f_{\max}(t)$ неизвестны, подбор масштабов, удовлетворяющих (9.29), может быть осуществлен методом последовательных приближений.

Масштаб времени M_t позволяет «сжать» слишком медленные процессы ($M_t > 1$) или растянуть быстрые ($M_t < 1$). При $M_t = 1$ моделирование осуществляется в натуральном масштабе времени. Теоретически продолжительность модельного процесса может быть выбрана произвольно. На практике точность моделирования существенно зависит от токов утечек. Поэтому слишком большая длительность моделирования нежелательна. Обычно время моделирования не должно превышать 1—2 мин.

Для решения обыкновенных дифференциальных уравнений должны быть заданы начальные условия (4.12). В рассматриваемом случае должны быть заданы начальные значения функции и ее первых производных:

$$x(0) = C_0; \quad \frac{dx(0)}{dt} = C_1; \quad (9.30)$$

$$\frac{d^2x(0)}{dt^2} = C_2; \quad \frac{d^3x(0)}{dt^3} = C_3.$$

Соответствующие значения машинных переменных (с учетом знаков напряжений (см. рис. 9.4, *a*) определяются соотношениями

$$\begin{aligned}\varphi(0) &= \frac{C_0}{M_x}; \quad \frac{d\varphi(0)}{d\tau} = -\frac{M_t}{M_x} C_1; \\ \frac{d^2\varphi(0)}{d\tau^2} &= \frac{M_t^2}{M_x} C_2; \\ \frac{d^3\varphi(0)}{d\tau^3} &= -\frac{M_t^3}{M_x} C_3.\end{aligned}\tag{9.31}$$

Начальные значения машинных переменных подаются к соответствующим точкам структурной модели с помощью специальных цепей (рис. 9.6). При этом осуществляется предварительная (до начала работы) подзарядка конденсаторов интегрирующих блоков и разделение элементов структурной модели. При включении модели одновременно включаются основные цепи и отключаются цепи начальных условий.

При работе модели происходит изменение ее характеристик (потенциалов $\varphi_1, \varphi_2, \varphi_3$ и т. д.) в функции времени. Эти изменения можно зафиксировать с помощью электронного или шлейфного осциллографа. Интерпретация изменения потенциала как изменения исследуемой переменной $x(t)$ осуществляется с помощью коэффициентов аналогии M_x, M_t . Важно подчеркнуть, что осциллограф может быть подключен к любой точке модели. Поэтому можно изучить изменение как функции $x(t)$, так и ее производных ($\dot{x}(t), \ddot{x}(t) \dots$).

Работа модели может осуществляться или в режиме однократного включения или с периодизацией решения.

При наличии нелинейностей в модель могут быть включены стандартные («ограничения», «зона нечувствительности», «люфт») или универсальные нелинейности. В качестве примера стандартного нелинейного преобразователя может служить рассмотренный выше множительный блок (см. рис. 9.3). Универсальные нелинейные преобразователи, позволяющие задать произвольную функциональную зависимость, могут быть собраны в виде вентильных схем, включающих в себя наборы диодных ячеек; на основе функциональных устройств, выполненных на электроно-лучевых трубках, на основе коммутаторов с конденсаторами¹. Сложные функциональные зависимости могут быть получены также как результат решения тех или иных дифференциальных уравнений с помощью автономных вспомогательных аналоговых моделей.

¹ Тетельбаум И. М. Электрическое моделирование. М.: Физматгиз, 1959, С. 172—189.

1. Составьте структурные схемы для моделирования следующих дифференциальных уравнений:

$$a_5 \frac{d^5 x}{dt^5} + a_4 \frac{d^4 x}{dt^4} + a_3 \frac{d^3 x}{dt^3} + a_2 \frac{d^2 x}{dt^2} + a_1 \frac{dx}{dt} + a_0 x = f(t);$$

$$a_5 \frac{d^5 x}{dt^5} + a_3 \frac{d^3 x}{dt^3} + a_1 \frac{dx}{dt} + a_0 x = f(t);$$

$$a_3 \frac{d^3 x}{dt^3} + a_0 x = 0.$$

2. Составьте машинные уравнения для исследования уравнений, приведенных в п. 1.

9.3. ИССЛЕДОВАНИЕ СИСТЕМ АВТОМАТИЧЕСКОГО РЕГУЛИРОВАНИЯ

Наиболее широко структурные модели применяются при изучении работы систем автоматического регулирования (САР) ДЛА. Необходимость использования моделей при изучении САР диктуется сложностью анализа многоконтурных систем регулирования, наличием в них нелинейных элементов и переменных параметров, потребностью в сравнении многочисленных вариантов регулирования.

С помощью моделей могут быть решены следующие задачи исследования САР:

- определение устойчивости системы;
- нахождение частотных характеристик цепей регулирования;
- определение характеристик переходных процессов;
- изучение влияния малых случайных отклонений параметров на работу системы;
- оптимизация параметров системы.

Существует два подхода к исследованию САР на структурных аналоговых моделях. Первый подход связан с непосредственным моделированием полной системы дифференциальных уравнений САР, второй — с разбиением САР на блоки и моделированием этих блоков отдельными аналоговыми блоками-моделями.

Для иллюстрации возможностей первого подхода рассмотрим схему автоматического регулирования, предназначенную для поддержания стационарного состояния ДЛА при наличии возмущений (рис. 9.7, а). Предполагается, что возмущение, вызывающее отклонение состояния объекта от номинала, $f(t)$ известно. Система регулирования получает информацию об изменении состояния объекта по двум каналам. Для определенности можно считать, что в систему регулирования поступает некоторая координата $x(t)$, характеризующая состояние объекта, и ее производная $\dot{x}(t)$, характеризующая скорость изменения состояния объекта. Например, в качестве $x(t)$ может выступать частота вращения ротора ГТД, а в качестве $\dot{x}(t)$ — скорость изменения этой частоты. Система

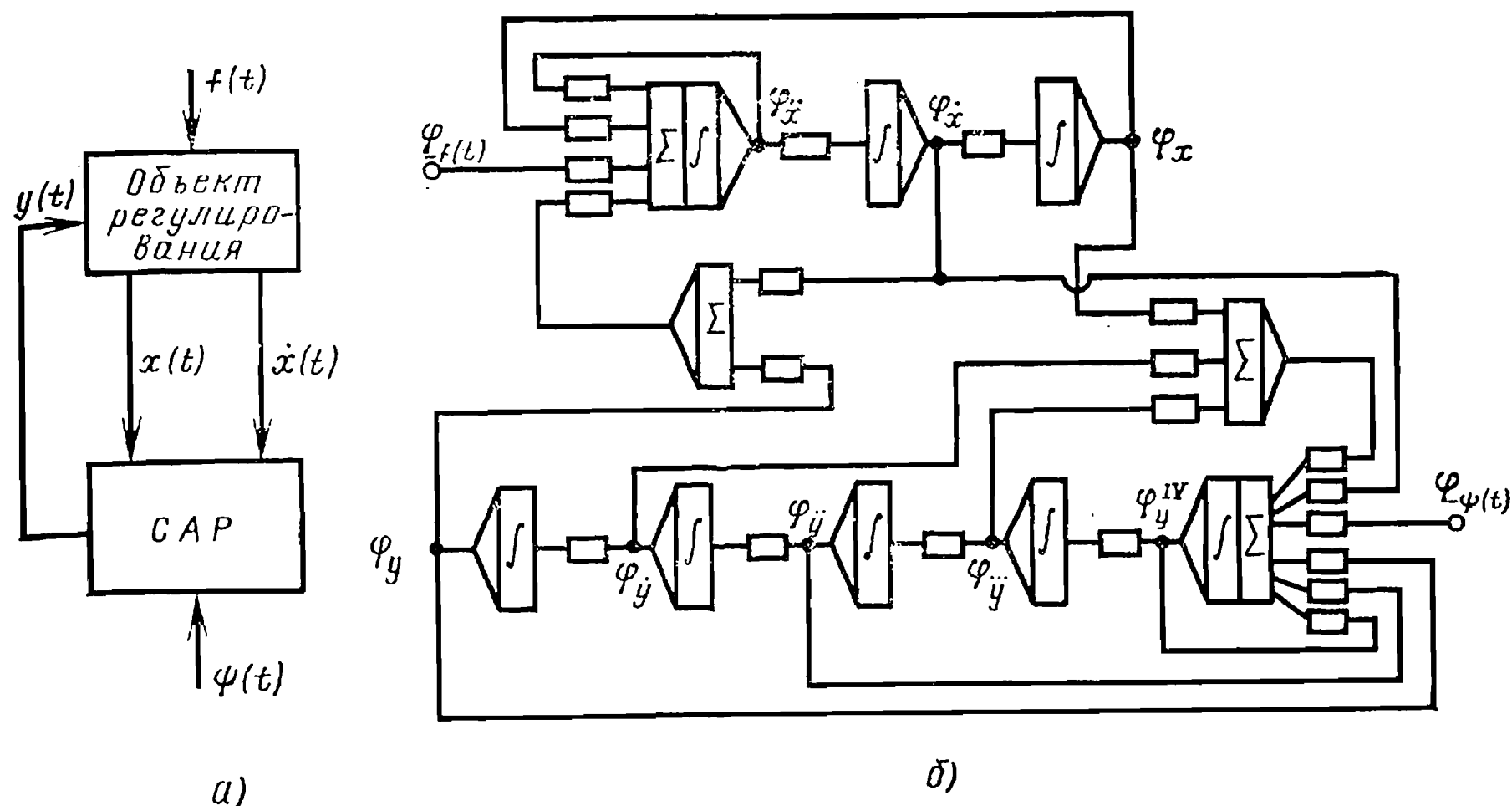


Рис. 9.7. Использование структурных моделей при изучении САР:
а — схема САР; б — схема модели

регулирования вырабатывает управляющий сигнал $y(t)$, возвращающий объект в стационарное состояние. В свою очередь, на систему регулирования действует возмущающий сигнал (шум) $\psi(t)$.

Пусть объект регулирования и система регулирования описываются обыкновенными дифференциальными уравнениями соответственно 3-го и 5-го порядка. При сделанных предположениях эти уравнения имеют вид

$$\begin{aligned}
 a_3 \frac{d^3 x(t)}{dt^3} + a_2 \frac{d^2 x(t)}{dt^2} + a_1 \frac{dx(t)}{dt} + a_0 x(t) &= c_1 f(t) + c_2 y(t); \\
 b_5 \frac{d^5 y(t)}{dt^5} + b_4 \frac{d^4 y(t)}{dt^4} + b_3 \frac{d^3 y(t)}{dt^3} + b_2 \frac{d^2 y(t)}{dt^2} + b_1 \frac{dy(t)}{dt} + b_0 y(t) &= \\
 &= d_0 \psi(t) + d_1 x(t) + d_2 \dot{x}(t). \quad (9.32)
 \end{aligned}$$

Начальное состояние системы характеризуется начальными условиями

$$\begin{aligned}
 x(0) &= C_0; \quad \dot{x}(0) = C_1; \quad \ddot{x}(0) = C_2; \\
 y(0) &= D_0; \quad \dot{y}(0) = D_1; \quad \ddot{y}(0) = D_2; \quad \dddot{y}(0) = D_3; \quad y^{IV}(0) = D_4. \quad (9.33)
 \end{aligned}$$

Для исследования рассмотренной системы может быть использована структурная модель, показанная на рис. 9.7, б. Модель включает в себя две цепочки интегрирующих и суммирующеинтегрирующих блоков для исследования каждого из уравнений (9.32). Обе эти цепочки объединены связями, учитывающими взаимное влияние объекта и регулятора. Для этой структурной модели могут быть записаны машинные уравнения, выбраны масштабы моделирования ($M_x, M_y, M_f, M_\psi, M_t$) и определены параметры решающих

элементов. В соответствующих масштабах должны быть заданы начальные условия. Рассмотренная модель позволяет изучить состояние системы в любой момент времени при любом характере возмущений и любых характеристиках регулятора.

Система автоматического регулирования может быть представлена в виде структурной схемы, отражающей ее физическую структуру. Такое представление САР позволяет создать аналоговую структурную модель в виде системы укрупненных блоков, каждый из которых соответствует определенному элементу схемы регулирования. Блочная модель позволяет не только изучить САР в целом, но и исследовать ее отдельные элементы. Кроме того, блочная модель открывает возможности сочтения модели с реальной аппаратурой.

На рис. 9.8, а представлена структурная схема рассмотренной выше САР для поддержания стационарного состояния ДЛА. Система регулирования включает в себя чувствительные элементы (датчики) для измерения отклонения объекта от заданного режима¹; управляющий элемент регулятора (УЭР), вырабатывающий на основании показаний датчиков управляющий сигнал; силовой привод, осуществляющий управление объектом (например сервопривод для регулирования сопла двигателя) и цепь обратной связи, которая для исключения перерегулирования начинает гасить управляющий сигнал при включении привода.

Каждый из элементов САР может быть представлен отдельной структурной аналоговой моделью. Такая модель может включать в себя один или несколько решающих элементов с набором входных цепей и цепей обратных связей (см., в частности, табл. 9.1). Так, в простейшем случае для рассмотренной структуры САР (рис. 9.8, б) датчик состояния объекта можно представить как колебательное звено; управляющий элемент регулятора — как сум-

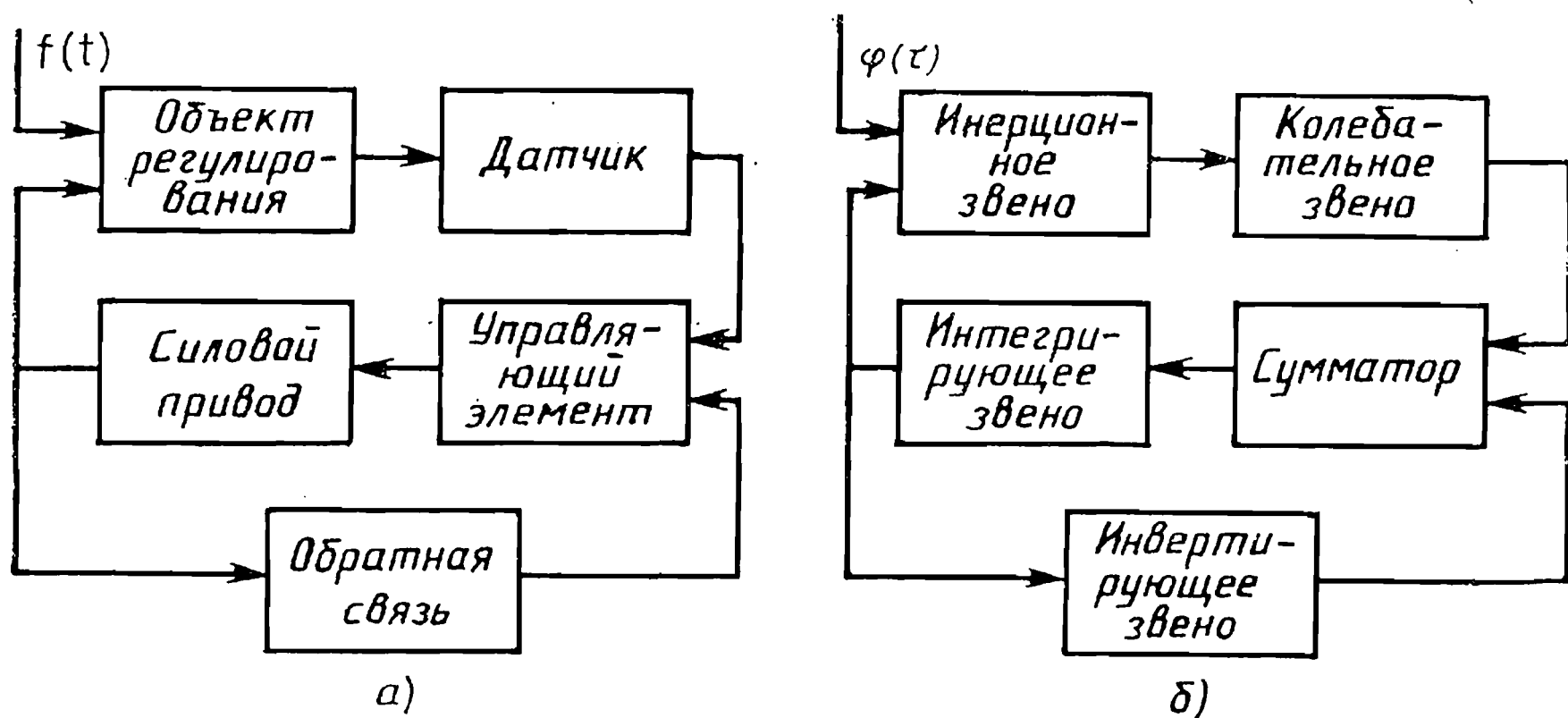


Рис. 9.8. Блочный принцип моделирования САР:

а — схема САР; б — схема модели

¹ При регулировании по двум каналам в состав САР входят два датчика.

матор; силовой привод — как интегрирующее звено; элемент цепи обратной связи — как инвертор; объект — как инерционное звено. Для реальных САР модели элементов будут более сложными. Возмущающий сигнал (шум) для каждого из элементов регулятора в данном случае должен задаваться самостоятельно.

Использование при моделировании блочной модели позволяет получать информацию о работе каждого элемента САР и изучать его влияние на работу системы в целом. Например, можно исследовать различные схемы выполнения чувствительного элемента и влияние каждого из вариантов выполнения чувствительного элемента на характеристики САР.

Блочный подход открывает возможности моделирования с *реальной аппаратурой*. В этом случае модель части системы сочетается с реальными блоками другой ее части. Например, модель ДЛА можно сочетать с реальной САР или модель ДЛА и части САР сочетать с реальным элементом САР, например, с реальным датчиком. Такой подход позволяет проводить натурные испытания регуляторов без затрат и риска, связанных с натурными испытаниями всей системы. В частности, в этом случае удастся осуществить испытания САР на *предельных* и *запредельных* режимах работы ДЛА.

По этому же принципу выполнены различные тренировочные установки для обучения шоферов, пилотов, операторов сложных устройств. Эти установки представляют собой модель объекта в сочетании с человеком-оператором и с целью обучения могут воспроизводить ту или иную ситуацию как в реальном, так и в замедленном темпе.

Сочетание модели с натурными блоками требует применения специальных преобразователей — переходных устройств для согласования сигналов реальных элементов и модели. В ряде случаев могут быть использованы стандартные преобразователи: напряжение — напряжение, код — напряжение, напряжение — код, угол поворота — напряжение, напряжение — угол поворота и т. д.

Вопросы для самоконтроля

1. Составьте структурные модели для исследования САР, описываемых следующими системами уравнений:

$$\left. \begin{aligned} a_4 \frac{d^4 x(t)}{dt^4} + a_2 \frac{d^2 x(t)}{dt^2} + a_0 x(t) &= C_0 f(t) + C_1 y(t) + C_2 \dot{y}(t); \\ b_3 \frac{d^3 y(t)}{dt^3} + b_2 \frac{d^2 y(t)}{dt^2} + b_0 y(t) &= d_0 \psi(t) + d_1 x(t) + d_2 \dot{x}(t); \\ a_5 \frac{d^5 x(t)}{dt^5} + a_1 \frac{dx(t)}{dt} + a_0 x(t) &= C_0 f(t) + C_2 y(t); \\ b_2 \frac{d^2 y(t)}{dt^2} + b_0 y(t) &= d_0 \psi(t) + d_1 x(t) + d_2 \dot{x}(t) + d_3 \ddot{x}(t). \end{aligned} \right\}$$

2. В чем заключается главный принцип блочного моделирования САР?
3. На чем основано моделирование САР с реальной аппаратурой?

МОДЕЛИ И МЕТОДЫ ИДЕНТИФИКАЦИИ ДВИГАТЕЛЕЙ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

При разработке систем математического моделирования часто возникает задача включения в модель экспериментальных данных, полученных при испытаниях ДЛА или их отдельных узлов. Результаты таких испытаний всегда представляют собой наборы величин, характеризующих работу ДЛА при различных сочетаниях управляющих параметров. Результаты численных исследований отдельных элементов и узлов ДЛА также представляют собой наборы величин и с точки зрения дальнейшего использования не отличаются от экспериментальных данных. Наиболее эффективным средством представления результатов экспериментов и численных расчетов в системах математического моделирования являются модели идентификации.

При построении модели идентификации обычно предполагается, что физическая теория работы объекта отсутствует или по тем или иным причинам не может быть использована. Объект идентификации представляет собой так называемый «черный ящик» с некоторым числом регулируемых (или, по крайней мере, измеряемых) входов x и одним или несколькими наблюдаемыми (измеряемыми) выходами. Задачей идентификации является построение модели объекта по результатам наблюдений его реакции на возмущения внешней среды. При этом необходимо учитывать ошибки, возникающие при измерении (вычислении) характеристик объекта (см. рис. 1.3, б).

ГЛАВА 10. ОБЩИЕ ПРИНЦИПЫ ПОСТРОЕНИЯ МОДЕЛЕЙ ИДЕНТИФИКАЦИИ

10.1. ВЫБОР ФОРМЫ МОДЕЛИ ИДЕНТИФИКАЦИИ

Математически задача идентификации формулируется следующим образом. Имеется n экспериментов ($r=1, 2, \dots, n$), для каждого из которых определен вектор входных параметров $x_r(x_{r1}x_{r2}, \dots, x_{rk})$ и известна (измерена или вычислена) выходная функция системы W_r .

В общем случае следует говорить о векторе выходных функций $W_r (W_{1r} \dots W_{mr})$, однако при идентификации модели для каждой из этих функций строятся *независимо* друг от друга, что позволяет ограничиться рассмотрением скалярного выхода. Требуется построить зависимость (модель)

$$W = f(x), \tag{10.1}$$

которая описывает характеристики изучаемой системы.

Обычно предполагается, что входные параметры (факторы) x_{ri} *взаимно независимы*. Однако это условие при построении моделей идентификации в общем случае не является строго обязательным. Уравнение (10.1) называется уравнением *регрессии* и описывает поверхность (гиперповерхность) отклика, характеризующую модель идентификации. Предполагается также, что имеющиеся экспериментальные (численные) данные дают достаточно информации для воссоздания математического описания объекта.

На рис. 10.1 в качестве примера показано (пока без каких-либо пояснений) решение задачи идентификации (прямая отклика) для некоторого набора экспериментальных данных ($n=7, k=1$), полученное с помощью линейной регрессионной зависимости:

$$W = a + bx. \tag{10.2}$$

Сформулированная задача в матричной форме может быть записана с помощью прямоугольной матрицы переменных (входов) $X(x_{ri})$ размерности $n \times k$ и матрицы столбца (в случае нескольких выходов — набора матриц столбцов) результатов $W(W_r)$ размерности $n \times 1$:

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}; \quad W = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{pmatrix}, \tag{10.3}$$

где каждому эксперименту соответствует строка матрицы X и соответствующий элемент матрицы $W(x_{r1}, x_{r2} \dots x_{rk}, W_r)$. Матрицы

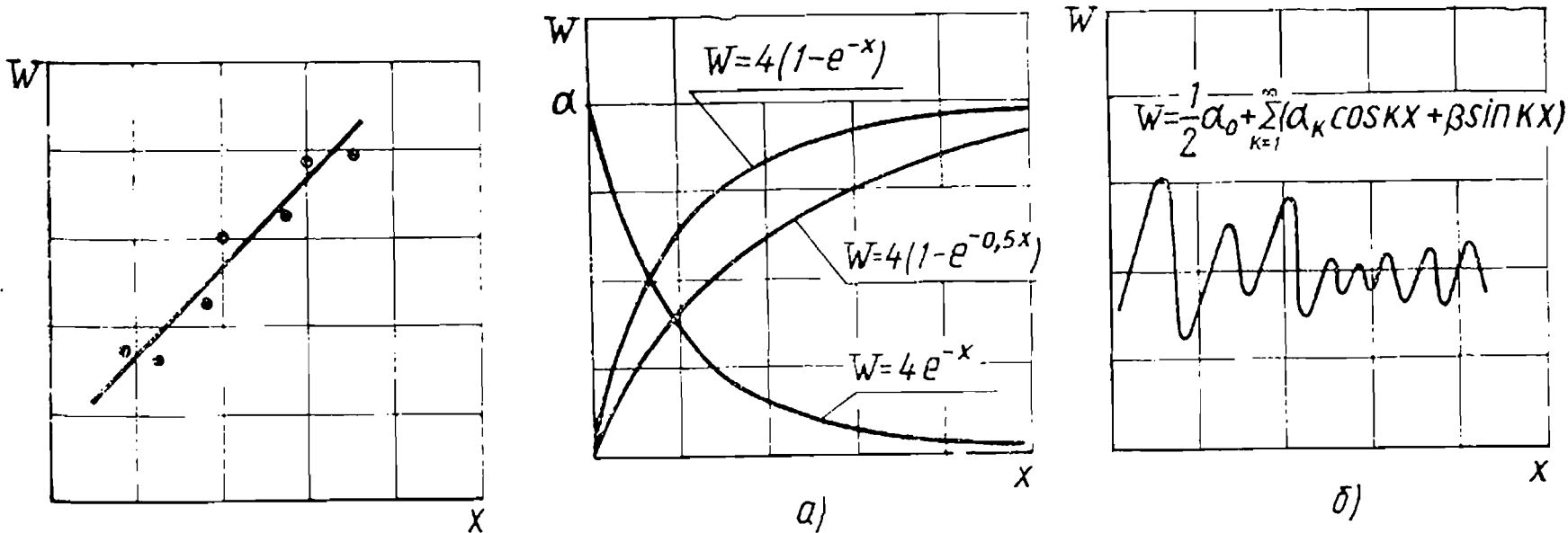


Рис. 10.1. Схема построения модели идентификации

Рис. 10.2. Выбор формы модели идентификации: а — форма модели задана заранее; б — использование отрезка гармонического ряда

(10.3) невырождены, если эксперименты и переменные не зависят друг от друга.

Построение модели идентификации начинается с выбора формы модели, т. е. вида зависимости (10.1). При этом на практике может встретиться два случая.

1. Форма математической модели *известна заранее*, и задача идентификации сводится к определению коэффициентов этой модели. Так, описание ряда затухающих или развивающихся процессов дается зависимостями экспоненциального типа (рис. 10.2, а)

$$\left. \begin{aligned} W &= \alpha \exp(\beta x); & W &= \alpha \exp(-\beta x); \\ W &= \alpha (1 - \exp(-\beta x)), \end{aligned} \right\} \quad (10.4)$$

и задачей исследования является определение коэффициентов α , β . При описании ряда физических процессов известна структура математических зависимостей. Так, например, при изучении термоэмиссионных преобразователей энергии для определения термоэмиссионного тока используется формула Ричардсона-Дешмана

$$J = AT^2 \exp\left(-\frac{e\varphi}{kT}\right), \quad (10.5)$$

где J — плотность термоэмиссионного тока; T — абсолютная температура; e — заряд электрона; k — постоянная Больцмана; φ — работа выхода металла; A — постоянная Ричардсона.

Формула (10.5) включает в себя одну величину, определяемую экспериментально, — работу выхода φ . Задачей идентификации в данном случае является подбор значений этой величины на основе экспериментальных данных.

Такая же ситуация имеет место при изучении теплообмена, течений жидкости, переходных процессов в системах автоматического регулирования. Иногда модель может быть сформулирована в виде критериальной зависимости и задача идентификации тогда заключается в определении степенных показателей критериев. В этих случаях количество подлежащих определению коэффициентов модели идентификации точно известно.

2. Форма математической модели *заранее неизвестна*. В этом случае для построения модели идентификации используются отрезки бесконечных рядов, а задача заключается в определении числа членов ряда и коэффициентов при этих членах.

Модель может быть представлена в виде

$$W = \sum_{i=1}^k \beta_{0i} f_0(x_i) + \sum_{i=1}^k \beta_{1i} f_1(x_i) + \dots + \sum_{i=1}^k \beta_{li} f_l(x_i), \quad (10.6)$$

где $f_q(x_i)$ — некоторые заданные функции; β_{qi} — коэффициенты регрессии ($q=0, 1, \dots, l$).

В одномерном случае ($k=1$) уравнение (10.6) принимает более простой вид:

$$W = \beta_0 f_0(x) + \beta_1 f_1(x) + \dots + \beta_l f_l(x). \quad (10.6a)$$

Конкретный вид модели зависит от выбора функций $f_q(x)$, по которым производится разложение W . Например, при описании колебательных процессов (рис. 10.2, б) удобно использовать ряд Фурье

$$W = \alpha_0 + \sum_{k=1}^{\infty} (\alpha_k \cos kx + \beta_k \sin kx). \quad (10.7)$$

При использовании в модели первых $k+1$ членов ряда необходимо определить $2k+1$ коэффициентов. Однако пока остается не вполне ясным, сколько же членов ряда обеспечивает наилучшее описание изучаемого процесса.

Если предварительная информация о характере поверхности отклика отсутствует, целесообразно использовать для построения модели наиболее общее разложение функции в ряд Тейлора (5.9). Уравнение регрессии в этом случае можно записать в виде

$$W = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{\substack{i=1 \\ j>i}}^k \beta_{ij} x_i x_j + \dots;$$

$$\beta_i = \frac{\partial W(x_0)}{\partial x_i}; \quad \beta_{ij} = \frac{\partial^2 W(x_0)}{\partial x_i \partial x_j}. \quad (10.8)$$

При $k=1$ (10.8) принимает вид

$$W = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_p x^p. \quad (10.8a)$$

Из сравнения (10.6a) и (10.8a) видно, что в качестве функций $f_0(x)$, $f_1(x)$, ..., $f_p(x)$ выступают степенные функции x^0 , x^1 , x^2 , ..., x^p .

Если ограничиться первыми членами разложения, то уравнения (10.8) и (10.8a) сведутся к рассмотренным выше линейным, квадратичным и другим полиномиальным моделям (см. разд. 5.2).

Для удобства дальнейшего изложения целесообразно изменить систему обозначения переменных. С этой целью вводятся новые переменные u :

$$u_0 = 1; \quad u_1 = x_1; \quad u_2 = x_2; \dots; u_k = x_k;$$

$$u_{k+1} = x_1^2; \dots, u_{2k} = x_k^2; \quad u_{2k+1} = x_1 x_2;$$

$$u_{2k+2} = x_1 x_3; \dots, u_l = x_{k-1} x_k; \quad u_{l+1} = x_1^3;$$

$$\dots u_{l+k} = x_k^3; \quad u_{l+k+1} = x_1 x_2 x_3; \dots; u_p,$$

$$(10.9)$$

и соответствующие обозначения коэффициентов регрессии β . В новой системе обозначений уравнение (10.8) принимает вид

$$W = \beta_0 u_0 + \beta_1 u_1 + \dots + \beta_p u_p = \sum_{i=0}^p \beta_i u_i, \quad (10.10)$$

где величина p и общее число коэффициентов регрессии m определяются размерностью задачи k и степенью полинома d :

$$m = C_{k+d}^d = \frac{(k+d)!}{k! d!}, \quad p = m - 1. \quad (10.11)$$

Таким образом, новая система обозначений позволяет представить модель в виде некоторой квазилинейной формы (10.10), что существенно упрощает ряд последующих выкладок.

Исходный экспериментальный материал (10.3) и коэффициенты регрессии могут быть представлены в виде

$$\mathbf{U}_{(n \times m)} = \begin{bmatrix} u_0, u_{11}, \dots, u_{1k}, u_{1k+1}, \dots, u_{1p} \\ u_0, u_{21}, \dots, u_{2k}, u_{2k+1}, \dots, u_{2p} \\ \vdots \\ u_0, u_{n1}, \dots, u_{nk}, u_{nk+1}, \dots, u_{np} \end{bmatrix};$$

где $\mathbf{U}$ — расширенная матрица входных параметров (добавлены первый столбец, состоящий из единицы, и столбцы с $k+1$ -го по p -й, содержащие произведения элементов первых k столбцов); $\mathbf{W}$ — матрица-столбец результатов; $\mathbf{B}$ — матрица-столбец коэффициентов регрессии.

Использование конечного отрезка ряда вместо бесконечного разложения и определение коэффициентов регрессии на основе ограниченного набора экспериментальных данных не позволяют *точно* восстановить зависимости (10.8), (10.10). Поэтому при идентификации ставится задача нахождения *приближенной* модели в виде:

$$\widehat{W} = b_0 u_0 + b_1 u_1 + \dots + b_p u_p = \sum_{i=0}^p b_i u_i, \quad (10.12)$$

где коэффициенты b_i являются оценкой истинных коэффициентов регрессии β_i ($b_i \approx \beta_i$).

Возможны три соотношения между количеством экспериментов (n) и размерностью массива коэффициентов регрессии (m).

1. Количество коэффициентов регрессии равно количеству экспериментальных точек ($m=n$). В этом случае соотношения (10.3) с учетом (10.9) приводят к системе m уравнений с m неизвестными:

[illegible]

Если матрица (10.3) не вырождена, система (10.13) имеет единственное решение и позволяет определить все коэффициенты регрессии (**b**). В данном случае можно найти поверхность (кривую) отклика, *точно* проходящую через *все* экспериментальные точки.

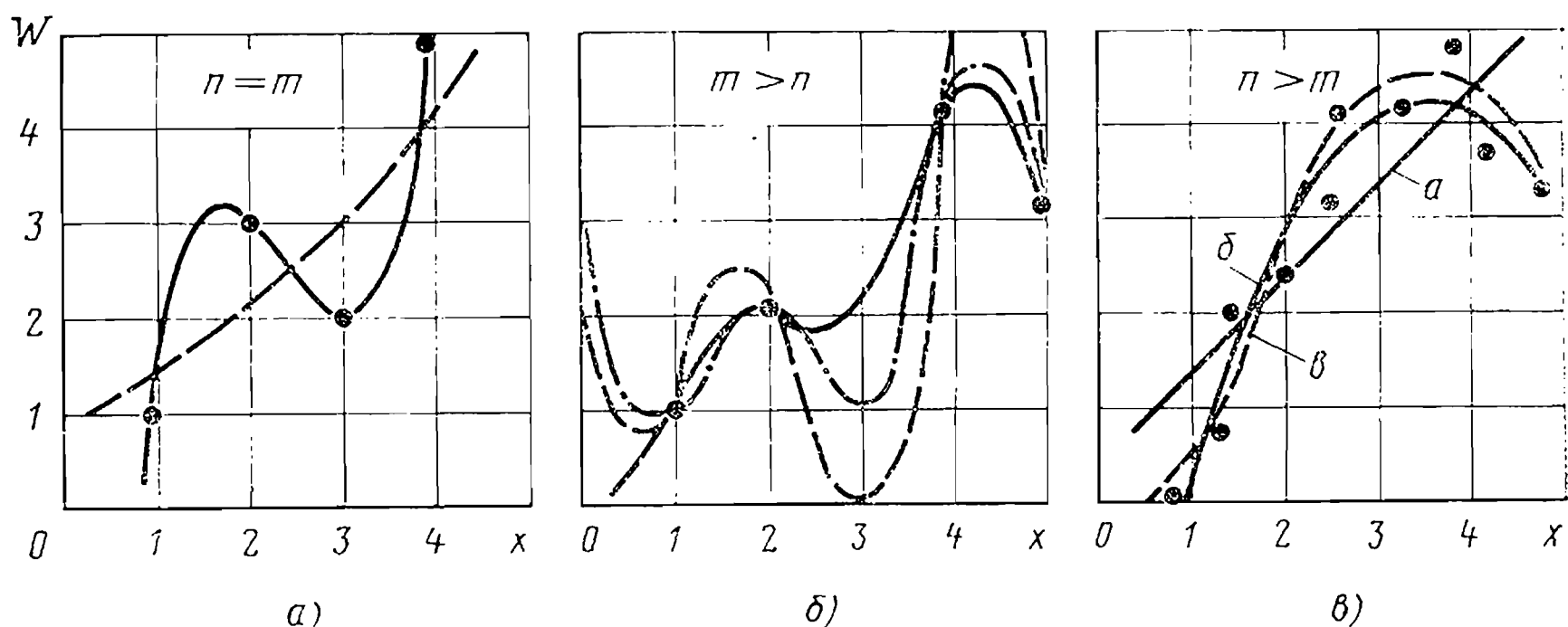


Рис. 10.3. Выбор порядка аппроксимирующего полинома

В качестве примера на рис. 10.3, а показана такая кривая отклика, построенная в виде полинома 3-й степени ($k=1$, $m=4$) по четырем экспериментальным точкам. Для *опорных* точек полученное решение является точным, т. е. обладает нулевой ошибкой. Однако точность описания процесса в промежуточных точках и достоверность самого характера модели при этом достаточно сомнительны. В частности, для экспериментальных данных, показанных на рис. 10.3, а, более точной является совсем другая модель (кривая отклика показана пунктиром). Дело в том, что в случае $n=m$ для определения коэффициентов регрессии используются все экспериментальные точки и не остается информации — степеней свободы (см. гл. 3) для проверки точности модели. Поэтому случай $n=m$ с точки зрения идентификации не является оптимальным.

2. Количество коэффициентов регрессии больше количества экспериментальных точек ($m > n$). В этом случае в системе (10.13) уравнений меньше, чем неизвестных. Система недоопределена и, следовательно, существует *неограниченное* число моделей, обладающих нулевой ошибкой (т. е. неограниченное число поверхностей отклика, проходящих через все экспериментальные точки). Например, на рис. 10.3, б показано несколько кривых шестого порядка, построенных по четырем экспериментальным точкам ($k=1$, $n=4$, $m=6$). Модели этого типа неустойчивы и не могут быть использованы на практике.

3. Количество коэффициентов регрессии меньше количества экспериментальных точек ($m < n$). В этом случае в системе (10.13) уравнений больше, чем неизвестных. Система *переопределена* и, следовательно, не существует ни одного решения, удовлетворяющего всем уравнениям (предполагается, что система не вырождена). Это значит, что нельзя построить поверхность отклика, проходящую через все экспериментальные точки. При этом, однако, можно построить приближенную модель, обеспечивающую в некотором смысле *наилучшее совпадение* с экспериментальными данными. Так, прямая, представленная на рис. 10.1, соответствует приближенной линейной модели, построенной по семи эксперимен-

тальным точкам методом наименьших квадратов. Легко видеть, что эта прямая достаточно хорошо соответствует исходному экспериментальному материалу, хотя и не проходит ни через одну из экспериментальных точек.

Именно условие $m < n$ представляет наибольший интерес с точки зрения идентификации, так как открывает возможности построения достаточно общих и точных моделей. Важным достоинством этих моделей является наличие $n - m$ степеней свободы, обеспечивающих статистическую достоверность получаемых результатов.

Вопросы для самоконтроля

1. В чем заключается задача идентификации?
2. От чего зависит форма модели идентификации?
3. Что такое коэффициенты регрессии и уравнения регрессии?
4. Как определяется число коэффициентов регрессии при использовании разложения в ряд Тейлора и в ряд Фурье?
5. Как может соотноситься между собой количество коэффициентов регрессии и количество экспериментальных точек? Что характеризует это соотношение?

10.2. КРИТЕРИИ ТОЧНОСТИ МОДЕЛЕЙ ИДЕНТИФИКАЦИИ

Для любой экспериментальной выборки могут быть предложены различные модели идентификации. Конкретная форма модели зависит от выбора функций $f_q(x)$ и количества членов ряда. Например, на рис. 10.3, в показано несколько кривых отклика, которые могут соответствовать представленной группе экспериментальных точек ($k=1, n=10, m=2, 3, 4$). Таким образом, сама постановка задачи идентификации включает в себя элемент *неопределенности*, возможность множественности решений. Важно выбрать лучшее или по крайней мере достаточно хорошее из этих решений.

Прежде всего можно просто положиться на здравый смысл и с помощью лекала или гибкой линейки провести на диаграмме кривую так, чтобы она проходила «посередине» выборки. Однако очевидно, что этот метод, во-первых, просто непригоден для построения многомерных моделей, во-вторых, необъективен, так как каждый из исследователей предложит свою кривую. Поэтому важно выработать объективные критерии соответствия модели рассматриваемой экспериментальной выборке.

Для оценки модели естественно использовать величины отклонений, полученных в эксперименте величин W_r и их оценок W_{rm} , предсказанных моделью $\{\epsilon_r\}$ (3.2). В качестве меры «значимости» отклонений $\{\epsilon_r\}$ можно использовать соотношения (3.8), (3.9), (3.10).

Доминирующее, более того, почти исключительное распространение при идентификации получил метод наименьших квадратов, требующий минимизации остаточной суммы квадратов (3.10) отклонений реальных значений оцениваемой величины от значений, предсказанных моделью.

Однако выбор метода построения модели не исчерпывает вопросов, связанных с ее точностью. С помощью критерия (3.10) можно подобрать для данной экспериментальной выборки наилучшую модель *данного вида*, т. е. принятой формы. Ответственность за выбор этой формы лежит на исследователе. К чему может привести неудачный выбор формы модели, видно из рис. 10.3, в, где, в частности, показаны результаты аппроксимации набора экспериментальных точек линейной моделью ($k=1$, $n=10$, $m=2$). Полученная модель является наилучшей линейной моделью (при построении использован метод наименьших квадратов), однако явно неудовлетворительно отражает характер экспериментальной выборки. Значительно лучшие результаты дает, например, квадратичная модель

$$W = b_0 + b_1x + b_2x^2. \quad (10.14)$$

Методы выбора оптимальной модели, основанные на прямом или многоступенчатом переборе вариантов, рассмотрены в разд. 10.3.

Непосредственное использование метода наименьших квадратов (3.10) приводит к оценке модели по экспериментальным данным, *уже использованным* при ее построении. Более объективная оценка может быть получена на основе дополнительных экспериментальных данных, образующих для модели некоторое *внешнее дополнение*. В этом случае весь экспериментальный массив делится на две последовательности: *обучающую* (n_1) и *проверочную* (n_2)¹.

Обучающая последовательность служит для определения коэффициентов регрессии модели, а проверочная — для оценки модели и сравнения моделей между собой. По существу, при этом вводится двухуровневая иерархическая система принятия решений, содержащая уровень (слой) выбора характеристик модели и уровень (слой) оценки моделей. При этом на вычислительную машину, осуществляющую перебор, возлагается часть функций, выполняемых обычно человеком.

В качестве критерия первого уровня для определения коэффициентов регрессии используется критерий наименьших квадратов. Критерии второго уровня, называемые критериями *регуляризации*, также обычно строятся на основе метода наименьших квадратов. Так, для оценки регулярности модели (точности краткосрочного прогноза) может быть использован критерий минимума средне-квадратичной ошибки, определенный на проверочной последовательности

$$K_p = \frac{1}{n_2} \sum_{r=1}^{n_2} (W_r - W_{r_m})^2 \Rightarrow \min, \quad (10.15)$$

¹ Можно выделить еще и третью «экзаменующую» последовательность для проверки оптимальности разбиения данных на обучающую и проверочную последовательности.

или относительный критерий

$$\bar{K}_p = \frac{\sum_{r=1}^{n_2} (W_r - W_{rM})^2}{\sum_{r=1}^{n_2} W_r^2} \Rightarrow \min \quad (10.15a)$$

Оптимальные по этим критериям модели мало чувствительны к небольшим изменениям исходных данных.

Число точек обучающей последовательности должно быть по крайней мере на единицу больше числа коэффициентов регрессии ($n_1 > m + 1$). Для повышения достоверности результатов этот запас должен быть существенно увеличен ($n_1 \geq (2 \dots 3)m$). Проверочная последовательность должна включать в себя хотя бы одну точку.

В ряде случаев в качестве критерия регуляризации удобно использовать критерий *несмещенности*, обеспечивающий наименьшее изменение модели при изменении состава обучающей последовательности. При этом весь экспериментальный массив разбивается на две одинаковые по величине последовательности ($n_1 = n_2$), каждая из которых поочередно используется в качестве обучающей. В результате их использования определяются две независимые, одинаковые по форме модели:

$$n_1 \Rightarrow W_M^* = f_1(u); \quad n_2 \Rightarrow W_M^{**} = f_2(u), \quad (10.16)$$

а критерий несмещенности рассчитывается по всем точкам выборки как средний квадрат разности *предсказанных* значений функции:

$$K_H = \frac{1}{n_1 + n_2} \sum_{r=1}^{n_1 + n_2} (W_{rM}^* - W_{rM}^{**})^2 \Rightarrow \min. \quad (10.17)$$

Лучшей из рассматриваемых является модель, обеспечивающая минимальное значение критерия несмещенности. В качестве критерия несмещенности можно использовать также отношение

$$\bar{K}_H = \frac{\sum_{r=1}^{n_1 + n_2} (W_{rM}^* - W_{rM}^{**})^2}{\sum_{r=1}^{n_1 + n_2} W_r^2}. \quad (10.17a)$$

Могут быть использованы и другие критерии регуляризации. В частности, в работе [7] рекомендуется использование в качестве критериев второго уровня комбинированных критериев, отражающих, например, требования регулярности и несмещенности модели:

$$K_{\text{комб}} = \sqrt{\lambda K_p^2 + (1 - \lambda) K_H^2}; \quad 0 < \lambda < 1. \quad (10.18)$$

Константа λ учитывает вес каждого из критериев.

1. В чем суть метода наименьших квадратов?
2. В чем заключается принцип двухуровневого выбора модели идентификации?
3. Что такое критерии регуляризации? Для чего они используются?
4. Запишите выражение критерия несмещенности и объясните его содержание.

ГЛАВА 11. МОДЕЛИ ИДЕНТИФИКАЦИИ ДЛЯ (НЕОРГАНИЗОВАННЫЙ ЭКСПЕРИМЕНТ)

11.1. РЕГРЕССИОННЫЙ АНАЛИЗ И МОДЕЛИ ИДЕНТИФИКАЦИИ

Специальные методы планирования эксперимента позволяют существенно повысить объем получаемой информации, улучшают характеристики моделей идентификации, а также упрощают процедуру обработки экспериментальных данных. Однако на практике очень часто приходится иметь дело с неорганизованным (пассивным) экспериментом. Связано это по крайней мере с тремя причинами.

Во-первых, в ряде случаев исследователь может только *наблюдать* входы системы (см. рис. 1.1), но не может их *регулировать*, что полностью исключает возможность планирования эксперимента (типичная ситуация: астроном — галактика).

Во-вторых, могут быть неизвестны *диапазоны* возможного изменения переменных (входов), что затрудняет планирование эксперимента и исключает возможность использования ряда эффективных методов планирования.

В-третьих, часто приходится строить модели идентификации на основе *уже полученных* ранее беспорядочных экспериментальных данных.

Основы методологии анализа неорганизованных данных были заложены в начале прошлого века Гауссом и Лапласом, сформулировавшими принцип наименьших квадратов. В настоящее время имеется ряд фундаментальных руководств по регрессионному анализу. Ниже рассмотрены лишь некоторые общие аспекты метода, необходимые для построения моделей идентификации.

Итак, имеется n результатов наблюдений величины W , зависящей от k независимых переменных, которые нужно представить степенным полиномом (10.10) степени d . При этом необходимо определить m коэффициентов модели.

При построении модели идентификации методами регрессионного анализа обычно предполагается, что:

- 1) результаты наблюдений $W_1, W_2, \dots, W_n$ представляют собой *независимые, нормально распределенные случайные* величины;
- 2) дисперсии этих случайных величин $\sigma^2(W_r)$ равны друг другу, т. е. разброс результатов эксперимента в рассматриваемой области факторного пространства *постоянен*;

$$U = \begin{vmatrix} u_0, & u_{11}, & \dots & u_{1p} \\ u_0, & u_{21}, & \dots, & u_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ u_0, & u_{n1}, & \dots, & u_{np} \end{vmatrix} \quad W = \begin{vmatrix} W_1 \\ W_2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ W_n \end{vmatrix}$$

$$U^*U = \begin{vmatrix} \sum_{r=1}^n u_0^2, & \sum_{r=1}^n u_{r1}u_0, & \dots, & \sum_{r=1}^n u_0u_{rp} \\ \sum_{r=1}^n u_{r1}u_0, & \sum_{r=1}^n u_{r1}^2, & \dots, & \sum_{r=1}^n u_{r1}u_{rp} \\ \cdot & \cdot & \cdot & \cdot \\ \sum_{r=1}^n u_{rp}u_0, & \sum_{r=1}^n u_{rp}u_{r1}, & \dots, & \sum_{r=1}^n u_{rp}^2 \end{vmatrix} \quad U^*W = \begin{vmatrix} \sum_{r=1}^n u_0W_r \\ \sum_{r=1}^n u_{r1}W_r \\ \cdot & \cdot & \cdot \\ \sum_{r=1}^n u_{rp}W_r \end{vmatrix}$$

$$U^*U^{-1} = \begin{vmatrix} c_{00}, & c_{01}, & \dots, & c_{0p} \\ c_{10}, & c_{11}, & \dots, & c_{1p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ c_{p0}, & c_{p1}, & \dots, & c_{pp} \end{vmatrix} \quad B = \begin{vmatrix} b_0 \\ b_1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ b_p \end{vmatrix}$$

Система нормальных уравнений в матричной форме может быть записана в виде

$$U^*UB = U^*W, \quad (11.5)$$

где U^* — транспонированная матрица U . Умножение слева обеих частей уравнения (11.5) на обратную матрицу $(U^*U)^{-1}$ дает

$$(U^*U)^{-1}(U^*U)B = (U^*U)^{-1}U^*W. \quad (11.6)$$

Но произведение матрицы на обратную всегда равно единичной матрице $(U^*U)^{-1}(U^*U) = E$. Поэтому из (11.6) следует

$$B = (U^*U)^{-1}U^*W. \quad (11.7)$$

Из (11.7) следует, что коэффициенты регрессии определяются выражениями (см. табл. 11.1)

$$b_i = \sum_{j=0}^p c_{ij} \sum_{r=1}^n u_{rj}W_r. \quad (11.8)$$

Матрица U^*U должна быть невырожденной. Это условие будет выполнено при линейной независимости переменных.

Таким образом, для определения коэффициентов регрессии необходимо произвести следующие операции:

составить матрицу независимых переменных $U(n \times m)$ и матрицу-столбец результатов $W(n \times 1)$;

найти транспонированную матрицу переменных $U^*(m \times n)$ и произведения матриц $U^*U(m \times m)$, $U^*W(m \times 1)$;

найти обратную матрицу $(U^*U)^{-1}(m \times m)$ и умножить ее на матрицу U^*W , т. е. определить матрицу-столбец $B(m \times 1)$.

Все вычисления производятся, как правило, с помощью ЭВМ. Математическое обеспечение большинства современных ЭВМ включает в себя готовые программы регрессионного анализа, предусматривающие также статистический анализ полученных результатов. Если же таких программ нет, то можно воспользоваться стандартными программами матричной алгебры для выполнения отдельных операций с матрицами.

Из формул (11.7), (11.8) следует, что коэффициенты регрессии могут быть определены только *совместно*. Действительно, сумма произведений, определяющая i -й коэффициент регрессии (11.8), состоит из m членов, соответствующих полному набору коэффициентов регрессии. Поэтому, если после вычислений и статистического анализа возникает необходимость изменения степени полинома или исключения отдельных членов модели, *все вычисления* нужно проводить заново. При этом меняются все коэффициенты регрессии.

Взаимосвязь коэффициентов исключает или по крайней мере сильно затрудняет их физическую интерпретацию. Поэтому модели идентификации, построенные методами регрессионного анализа, должны рассматриваться как *интерполяционные формулы*, коэффициенты которых лишены четкого физического смысла.

После вычисления коэффициентов регрессии необходимо провести статистический анализ полученной модели. Коэффициенты регрессии (11.8) по определению обеспечивают получение минимальной величины остаточной суммы квадратов, однако, очевидно, не могут свести ее к нулю. Величина остаточной суммы квадратов [квадратичной формы (11.1)] характеризует точность модели.

С помощью матричных преобразований для остаточной суммы квадратов может быть получено эквивалентное (11.1) выражение

$$S = \sum_{r=1}^n W_r^2 - \sum_{i=0}^p b_i \sum_{r=1}^n u_{ri} W_r. \quad (11.9)$$

Вычисление S по формуле (11.9) значительно проще, чем непосредственно по формуле (11.1), так как значения коэффициентов

b_i известны, а суммы $\sum_{r=1}^n u_{ri} W_r$ уже подсчитывались при вычислении

b_i (11.3). Однако уравнение (11.9) обладает повышенной чувствительностью к ошибкам округления, так как часто элементы суммы

$\sum_{i=0}^p b_i \sum_{r=1}^n u_{ri} W_r$ имеют один и тот же порядок величины и разные

знаки. Иногда рекомендуют с целью проверки всей системы вычислений определять S дважды как по формуле (11.1), так и по формуле (11.9) и сравнивать полученные значения.

Разделив остаточную сумму квадратов (11.1), (11.9) на число ее степеней свободы, можно найти средний квадрат этой суммы

$$s^2 = \frac{S}{f_s}; \quad f_s = n - m \quad (11.10)$$

и использовать для оценки адекватности модели действительности критерий Фишера (см. разд. 3.3).

Для оценки связи коэффициентов регрессии между собой используются методы корреляционного анализа (см. разд. 3.4). При этом строится так называемая *матрица ошибок*, представляющая собой ковариационную матрицу $\mathbf{D}(b_i)$:

$$\mathbf{D}(b_i) = \sigma^2(W) (\mathbf{U}^* \mathbf{U})^{-1}, \quad (11.11)$$

или в развернутом виде [14]

$$\begin{vmatrix} \sigma^2(b_0) & \text{cov}(b_0, b_1) & \dots & \text{cov}(b_0, b_p) \\ \text{cov}(b_1, b_0) & \sigma^2(b_1) & \dots & \text{cov}(b_1, b_p) \\ \dots & \dots & \dots & \dots \\ \text{cov}(b_p, b_0) & \text{cov}(b_p, b_1) & \dots & \sigma^2(b_p) \end{vmatrix} = \sigma^2(W) \begin{vmatrix} c_{00}c_{01} & \dots & c_{0p} \\ c_{10}c_{11} & \dots & c_{1p} \\ \dots & \dots & \dots \\ c_{p0}c_{p1} & \dots & c_{pp} \end{vmatrix}. \quad (11.12)$$

Из соотношений (11.11), (11.12) следует, что

$$\sigma^2(b_i) = c_{ii}\sigma^2(W); \quad \text{cov}(b_i, b_j) = c_{ij}\sigma^2(W); \quad (11.13)$$

$$r(b_i, b_j) = \frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}}.$$

Диагональные элементы матрицы (11.12) определяют дисперсии коэффициентов регрессии, а недиагональные — взаимосвязь этих коэффициентов. Ковариации характеризуют меру неопределенности, возникающей вследствие того, что коэффициенты регрессии определяются совместно. Далее будет показано, что при специальном *ортогональном* планировании экспериментов ковариационная матрица диагональна, а коэффициенты регрессии независимы, что обеспечивает модели идентификации ряд полезных свойств.

Для изучения роли отдельных членов можно использовать дополнительную сумму квадратов SS (3.20) и средний квадрат этой суммы s_{SS}^2 (3.21). Таким образом, можно, например, оценить необходимость включения в линейную модель членов второго порядка, а также роль каждого отдельного члена уравнения регрессии.

Вопросы для самоконтроля

1. Перечислите предпосылки регрессионного анализа и объясните их смысл.
2. Запишите нормальные уравнения для вычисления коэффициентов регрессии. Прodelайте выкладки, необходимые для их получения.

3. Запишите уравнения регрессионного анализа в матричной форме. Объясните смысл матричных преобразований.

4. Перечислите процедуры, необходимые для вычисления коэффициентов регрессии на ЭВМ.

5. С чем связаны трудности интерпретации коэффициентов регрессии?

6. Запишите матрицу ошибок для определения взаимосвязи коэффициентов регрессии.

11.2. ПРИМЕР ПОСТРОЕНИЯ МОДЕЛИ ИДЕНТИФИКАЦИИ

В качестве простого примера построения модели идентификации методами регрессионного анализа рассмотрим задачу восстановления математического описания известного процесса по результатам искусственно организованного математического эксперимента.

Предполагается, что процесс описывается одномерным уравнением второго порядка

$$W = 1 + 4x - 0,5x^2, \quad 0 \leq x \leq 6. \quad (11.14)$$

Истинный характер процесса показан на рис. 11.1 сплошной линией. Считается, что уравнение (11.14) неизвестно, но величина W может быть вычислена для любого x заданного диапазона. При этом величина x измеряется точно, а W — с ошибкой, имеющей нормальное распределение с нулевым математическим ожиданием и единичной дисперсией:

$$\tilde{W} = W + \varepsilon = 1 + 4x - 0,5x^2 + \varepsilon; \quad M(\varepsilon) = 0; \quad \sigma^2(\varepsilon) = 1. \quad (11.15)$$

Для получения выборки «экспериментальных» данных осуществляется математический эксперимент по следующей схеме:

1) с помощью датчика (таблицы) равномерно распределенных на отрезке $[0, 1]$ случайных чисел определяется случайное число ξ_p ($0 \leq \xi_p \leq 1$);

2) величина x определяется нормированием случайного числа на интервал $[0, 6]$ ($x = 6\xi_p$);

3) с помощью датчика (таблицы) нормально распределенных $(0, 1)$ случайных чисел определяется случайное число ξ_N , представляющее собой «ошибку» эксперимента ($\varepsilon = \xi_N$);

4) «наблюдаемая» величина $\tilde{W}$ определяется по (11.15) как сумма регулярной и случайной составляющих.

Результаты математического эксперимента (выборка 10 случайных пар

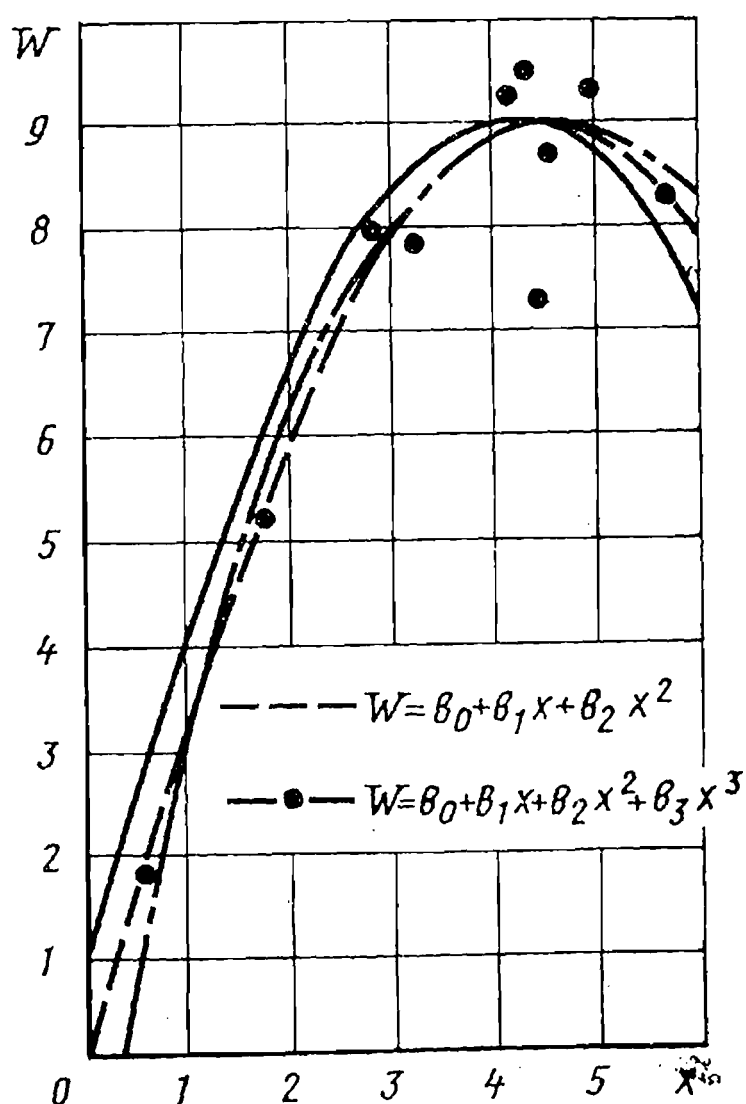


Рис. 11.1. Пример построения модели идентификации

$(x, \tilde{W})$) представлены в виде точек на рис. 11.1. Задача сводится к построению модели идентификации

$$W_m = b_0 u_0 + b_1 u_1 + b_2 u_2; \quad u_0 = 1; \quad u_1 = x; \quad u_2 = x^2 \tag{11.16}$$

при $n=10, k=1, p=2, m=3$. Матрицы, с помощью которых записывается решение, приведены в табл. 11.2. Полученная зависимость

$$W_m = -0,286 + 4,065x - 0,446x^2 \tag{11.17}$$

показана на рис. 11.1 пунктиром. Из рисунка видно, что полученная модель качественно правильно воспроизводит исходное описание процесса.

Таблица 11.2

$X =$	1	4,8606	23,629		$W =$	9,28	
	1	4,2396	17,974			9,40	
	1	2,7792	7,724			7,88	
	1	0,5988	0,359			1,86	
	1	3,2136	10,327			7,77	
	1	4,5156	20,391			8,73	
	1	5,9340	35,212			8,33	
	1	1,5852	2,516			5,16	
	1	4,4880	20,142			7,28	
	1	4,0932	16,754			9,22	
$X * X =$	10,02	36,31	155,02		$X * W =$	75,61	
	36,31	155,02	709,9			303,25	
	155,02	709,9	3396,1			1327,12	
$X * X^{-1} =$	1,596	−0,940	0,124		$B =$	−0,286	
	−0,940	0,704	−0,104			4,065	
	0,124	−0,104	0,016			−0,446	

Статистические характеристики модели имеют следующие значения:

$$S_2 = 4,681; \quad f_s = 7; \quad s_2^2 = 0,669; \quad F_{7,\infty} = 0,669. \tag{11.18}$$

По таблицам распределения критерия Фишера $F_{7,\infty}$, при 5%-м уровне значимости находим критическое значение $F_{7,\infty;0,05} = 2,01$ [16]. Так как полученное значение F меньше критического, гипотеза об адекватности модели (11.17) реальному процессу (11.14) принимается.

Ковариационная матрица и коэффициенты корреляции имеют вид

$$D(b) = \begin{vmatrix} 1,596; & -0,940; & 0,124 \\ -0,940; & 0,704; & -0,104 \\ 0,124; & -0,104; & 0,016 \end{vmatrix}; \quad \begin{matrix} r_{0,1} = -0,886 \\ r_{0,2} = 0,245 \\ r_{1,2} = -0,311 \end{matrix} \tag{11.19}$$

Таким образом, ошибка при определении коэффициентов регрессии соизмерима с ошибкой эксперимента $\sigma^2(b_i) \approx \sigma^2(W)$, а коэффициенты регрессии достаточно сильно связаны между собой.

Поскольку форма модели заранее неизвестна, целесообразно рассмотреть возможность использования кубической модели

$$\begin{aligned}
 W_m &= b_0u_0 + b_1u_1 + b_2u_2 + b_3u_3; \\
 u_0 &\equiv 1; \; u_1 = x; \; u_2 = x^2, \; u_3 = x^3.
 \end{aligned}
 \tag{11.20}$$

В результате расчетов можно найти коэффициенты такой модели (рис. 11.1):

$$W_m = -0,878 + 4,98x - 0,768x^2 + 0,032x^3. \tag{11.17a}$$

Для проверки целесообразности включения в модель члена третьей степени вычисляется остаточная сумма квадратов для уравнения (11.17а), дополнительная сумма квадратов, средний квадрат и величина критерия Фишера:

$$\begin{aligned}
 S_3 &= 4,537; \; SS = S_2 - S_3 = 0,144; \; f_{SS} = 1; \\
 s_{SS}^2 &= \frac{SS}{f_{SS}} = 0,144; \; F_{1,7} = \frac{s_{SS}^2}{s_2^2} = 0,215.
 \end{aligned}
 \tag{11.21}$$

По таблицам распределения определяется критическое значение критерия Фишера $F_{1; 7; 0,05} = 5,59$. В связи с тем, что полученное из расчета значение критерия Фишера (11.21) меньше критического, можно считать, что член третьего порядка не добавляет существенной информации и, следовательно, переход от (11.17) к (11.17а) нецелесообразен.

Вопросы для самоконтроля

Используя шесть пар случайных величин (x, W)

x	0,63,	0,01,	1,96,	0,31,	2,52,	3,59
W	1,95,	1,05,	1,85,	1,51,	3,01,	2,92

следует:

а) методами регрессионного анализа построить линейную модель идентификации в виде

$$W = b_0u_0 + b_1u_1; \; u_0 \equiv 1; \; u_1 = x;$$

б) провести статистический анализ модели;

в) проверить целесообразность включения в модель квадратичного члена, т. е. перехода к модели

$$W = b_0u_0 + b_1u_1 + b_2u_2; \; u_0 \equiv 1; \; u_1 = x; \; u_2 = x^2.$$

При расчетах целесообразно использовать клавишные вычислительные машины или малые ЭВМ.

11.3. ВЫБОР ОПТИМАЛЬНОЙ МОДЕЛИ ИДЕНТИФИКАЦИИ

Принцип наименьших квадратов позволяет найти наилучшую модель идентификации для описания исследуемой экспериментальной выборки заданным уравнением регрессии (10.6). Если имеются достаточно веские основания для выбора формы

этого уравнения, никаких проблем не возникает. Однако в большинстве случаев при исследовании ДЛА конкретная форма модели заранее неизвестна и может, вообще говоря, быть различной.

На первый взгляд может показаться, что более сложная модель (увеличение степени полинома) всегда обеспечивает получение большей точности. На самом деле это не так. При переходе к полиномам более высокой степени можно, конечно, получить лучшее *согласие* регрессионной кривой с экспериментальными данными. Для $m=n$ это согласие будет абсолютным (см. рис. 10.3, а), но при этом получится *худшее* согласие с истинным характером процесса $W(x)$. Дело в том, что экспериментальные данные представляют собой случайные величины и содержат лишь ограниченную информацию о характере $W(x)$. Увеличение степени полинома целесообразно лишь до тех пор, пока из экспериментальной выборки извлекается надежная информация. С другой стороны, увеличение степени полинома при заданном объеме экспериментальной выборки приводит к сокращению числа степеней свободы. Таким образом возникает проблема оптимального выбора формы модели.

В настоящее время наметилось два подхода к решению этой проблемы. Первый из них основан на статистическом исследовании уравнений регрессии [6] (см. разд. 11.2), а второй — на использовании критериев регуляризации (см. разд. 10.3) и селективном отборе моделей [7].

В первом случае вся имеющаяся информация используется при построении регрессионной зависимости (точнее, ряда зависимостей). Для выбора наилучшей из них может быть использован ряд методов.

Метод всех возможных регрессий основан на последовательном изучении всех возможных моделей ($m < n$). При этом все модели разбиваются на серии в зависимости от числа введенных переменных (в табл. 11.3 представлено несколько первых серий таких моделей). Для каждой из серий отбирается лучшая модель, причем в качестве критерия часто используется так называемый *множественный коэффициент детерминации*:

$$R = \frac{\sum_{r=1}^n (W_{rM} - \bar{W})^2}{\sum_{r=1}^n (W_r - \bar{W})^2}; \quad \bar{W} = \frac{1}{n} \sum_{r=1}^n W_r, \quad (11.22)$$

характеризующий полноту использования информации в регрессионной модели ($0 \leq R \leq 1$). Лучшей модели соответствует большее значение R . Лучшие модели серий сравниваются между собой.

Метод представляется мало пригодным для анализа сложных моделей ДУ ЛА, так как, во-первых, отличается высокой трудоемкостью (общее число возможных вариантов равно 2^m), а во-вторых, недостаточно надежен статистически: при $n > m$ увеличение числа переменных почти всегда приводит к росту R , что, вообще

Полиномы серий		
1-й	2-й	3-й
$W = b_0u_0$ $W = b_1u_1$ $W = b_mu_m$	$W = b_0u_0 + b_1u_1$ $W = b_0u_0 + b_2u_2$ $W = b_0u_0 + b_mu_m$ $W = b_1u_1 + b_2u_2$ $W = b_1u_1 + b_mu_m$ $W = b_{m-1}u_{m-1} + b_mu_m$ } 1	$W = b_0u_0 + b_1u_1 + b_2u_2$ $W = b_0u_0 + b_1u_1 + b_3u_3$ $W = b_0u_0 + b_1u_1 + b_mu_m$ $W = b_0u_0 + b_2u_2 + b_3u_3$ $W = b_0u_0 + b_2u_2 + b_mu_m$ $W = b_0u_0 + b_{m-1}u_{m-1} + b_mu_m$ $W = b_1u_1 + b_2u_2 + b_3u_3$ $W = b_1u_1 + b_2u_2 + b_mu_m$ $W = b_1u_1 + b_3u_3 + b_4u_4$ $W = b_1u_1 + b_3u_3 + b_mu_m$ $W = b_1u_1 + b_{m-1}u_{m-1} + b_mu_m$ $W = b_2u_2 + b_3u_3 + b_4u_4$ $W = b_{m-2}u_{m-2} + b_{m-1}u_{m-1} + b_mu_m$
$m + 1$	Количество полиномов $\sum_{i=1}^m i$	$\sum_{k=1}^{m-1} \left(\sum_{i=1}^k i \right)$

говоря, еще не является достаточным основанием для усложнения модели.

Метод исключения предполагает исследование наиболее полной (в пределах разумного) модели и последовательную проверку на значимость всех ее членов. При этом для каждого из членов модели вычисляется величина критерия F в предположении, что все остальные члены уже введены в модель. На основе полученного множества $\{F_i\}$ выбирается член уравнения регрессии, соответствующий минимальному значению критерия F_l . Если это минимальное значение меньше критического при выбранном уровне значимости ($F_l < F_{кр}$), то соответствующий член исключается из регрессионного уравнения как несущественный, после чего все коэффициенты регрессии пересчитываются заново и вновь осуществляется проверка их значимости.

Если $F_l > F_{кр}$, то все члены модели существенны и уравнение регрессии остается в первоначальном виде. Однако, если это произошло уже на первом шаге исследования, стоит рассмотреть целесообразность усложнения первоначальной модели.

Трудоемкость метода исключения меньше, чем метода всех возможных регрессий, а статистические основы надежнее.

Метод включения по существу противоположен методу исключения и предусматривает последовательное включение в модель новых членов с проверкой их статистической значимости. Конкретная процедура включения практически уже рассмотрена в разд. 11.2. Определенную роль играет порядок включения новых членов. В частности, если проведен корреляционный анализ, целесообразно в первую очередь включать в модель члены, зависящие от переменных, сильно коррелированных с результирующей функцией W . Трудоемкость метода включения существенно меньше трудоемкости рассмотренных выше методов.

Существуют и некоторые другие методы подбора оптимального уравнения регрессии [6].

Общим недостатком всех рассмотренных ранее методов является использование для оценки модели того же экспериментального материала, на основе которого эта модель построена. Иной подход к выбору оптимальной модели основан на использовании критериев регуляризации (см. разд. 10.3). При этом подходе все экспериментальные данные разбиваются на две части (последовательности): обучающую и проверочную. Первая из них используется для определения коэффициентов регрессии модели, вторая — для оценки модели в целом.

Критерии регуляризации могут быть использованы в сочетании с полным перебором моделей, что в известной мере соответствует методу всех возможных регрессий. При переборе происходит постепенное усложнение моделей (например, в соответствии с табл. 11.3). При этом критерий регуляризации сначала уменьшается (точность модели растет), а потом начинает возрастать. Оптимальной считается модель, соответствующая прекращению уменьшения (или снижению до определенной величины скорости уменьшения) критерия. Критерий регуляризации в отличие от критерия R (11.22) всегда имеет четко выраженный минимум [7], что обеспечивает объективное выделение модели оптимальной сложности. Процедура полного перебора всех моделей в данном случае так же трудоемка, как и при использовании метода всех регрессий и, следовательно, мало пригодна для исследования сложных моделей ДЛА, зависящих от большого числа переменных.

Для уменьшения объема перебора используется так называемый **метод многорядной селекции моделей**, основанный на принципах селекции растений или животных.

Процедура селекции состоит в следующем. Высеивается некоторое количество семян данного вида. В результате перекрестного опыления образуются сложные наследственные комбинации. Из полученных растений отбирается некоторая часть растений, лучше других соответствующих задачам селекции. Семена этих растений вновь высеиваются (второй ряд селекции) для образования новых, еще более сложных наследственных комбинаций. Операции продолжают в течение ряда поколений до вырождения процесса, т. е. до прекращения улучшения свойств вида.

Характерные особенности процедуры селекции заключаются в том, что, во-первых, в каждый новый ряд (цикл) селекции пропускается только *часть* растений предыдущего ряда (что ограничивает объем перебора); во-вторых, отбирается *некоторое количество* (а не одно единственное!) лучших растений, что обеспечивает свободу выбора в следующих поколениях и снижает риск потери наиболее перспективных решений на промежуточных этапах; в-третьих, для отбора лучших решений используется внешний по отношению и процедуре селекции критерий.

С точки зрения системотехники, многорядная селекция представляет собой многоуровневую иерархическую процедуру последовательного выбора решений. Очевидно, этим же путем осуществляется и естественный отбор в природе, причем в качестве критерия отбора выступает требование выживаемости вида.

Для построения модели идентификации методами многорядной селекции разработаны алгоритмы *многогруппового учета аргументов* (МГУА) [7]. Общая схема этих алгоритмов сводится к следующему.

Полная модель идентификации заменяется несколькими рядами частных описаний с постепенным усложнением функций. Например, к первому ряду селекции можно отнести функции парных взаимодействий:

$$y_{01} = f(x_0, x_1); y_{02} = f(x_0, x_2); \dots y_{k-1,k} = f(x_{k-1}, x_k). \quad (11.23)$$

Причем общее число уравнений, очевидно, равно $C_k^2 = \frac{1}{2} k(k-1)$.

Второй ряд селекции включает в себя функции от функций первого ряда:

$$z_1 = f(y_{01}, y_{02}), \dots, z_r = f(y_{k-2,k}, y_{k-1,k}). \quad (11.24)$$

К третьему ряду относятся функции от функций второго ряда:

$$v_1 = f(z_1, z_2), \dots, v_t = f(z_{r-1}, z_r) \quad (11.25)$$

и т. д. При этом каждая из функций второго ряда зависит от четырех (или трех, если одна из них входит дважды) первичных переменных. Функции третьего ряда в общем случае зависят от восьми переменных. Таким образом, сложность описания повышается с каждым рядом селекции. Вид функции f во всех рядах селекции сохраняется постоянным. Из каждого ряда для дальнейшего использования отбирается лишь некоторая часть L лучших функций (моделей).

Различные алгоритмы МГУА отличаются выбором вида опорных функций f . Если f — линейная функция, то первый ряд селекции включает в себя уравнения

$$\begin{aligned} y_1 &= b_0 + b_1 x_1; y_2 = b_0 + b_2 x_2; \dots; \\ y_r &= b_{k-1} x_{k-1} + b_k x_k. \end{aligned} \quad (11.23a)$$

Ко второму ряду относятся уравнения

$$\begin{aligned} z_1 &= a_1 y_1 + a_2 y_2; \dots; \\ z_r &= a_{r-1} y_{r-1} + a_r y_r. \end{aligned} \quad (11.24a)$$

При этом легко видеть, что

$$\begin{aligned} z_1 &= a_0(b_0 + b_1x_1) + a_1(b_0 + b_2x_2) = B_{01} + B_{11}x_1 + B_{21}x_2; \\ z_r &= a_{r-1}(b_{k-2}x_{k-2} + b_kx_k) + a_r(b_{k-1}x_{k-1} + b_kx_k) = \\ &= B_{k-2,k-1}x_{k-2}x_{k-1} + B_{k-2,k}x_{k-2}x_k + B_{k-1,k}x_{k-1}x_k + B_kx_k^2. \end{aligned} \quad (11.26)$$

Известны алгоритмы МГУА, в которых в качестве опорных функций используются квадратичные полиномы, функции Байеса и другие функции.

При отборе моделей для следующего ряда селекции часто используется принцип *постоянства числа переменных*. Это значит, что если в задаче рассматривается, например, десять первичных переменных (при этом можно составить 45 соотношений типа (11.23а)), во второй ряд селекции пропускается 10 лучших по критерию регуляризации моделей. При этом количество возможных моделей во всех рядах селекции, очевидно, одинаково. Могут быть предложены и другие принципы отбора.

Для оценки эффективности каждого очередного ряда селекции определяется *средний* критерий регуляризации для лучших моделей, пропускаемых в следующий ряд. Этот средний критерий сравнивается с соответствующим критерием предыдущего ряда. Достаточно интенсивное улучшение критерия от ряда к ряду подтверждает эффективность селекции и необходимость ее продолжения. Когда средний критерий регуляризации, определенный на группе лучших моделей, достигает минимума, селекцию следует прекратить, а лучшую из моделей последней группы признать оптимальной.

На практике селекцию прекращают несколько раньше, когда улучшение среднего критерия становится незначительным. Это приводит к более простым моделям и практически не сказывается на их точности.

Использование принципа многорядной селекции с полным перебором парных сочетаний и отбором лучших моделей в каждом из рядов существенно снижает трудоемкость по сравнению с процедурой перебора всех моделей. Однако при существенном увеличении числа первичных переменных ($k > 100$) этот выигрыш уже становится недостаточным. В этом случае приходится отказаться от полного перебора пар в каждом из рядов и перейти к *случайному выбору* парных сочетаний. При этом для каждого ряда селекции строится ограниченное число моделей, зависящих от случайно выбранных переменных (функций) предыдущего ряда. Этот механизм еще больше приближает рассматриваемый метод к естественному процессу селекции, так как при селекции опыление (скрещивание) в каждом из поколений осуществляется случайным образом.

Вопросы для самоконтроля

1. Почему необходимо искать оптимальную форму модели?
2. В чем смысл метода всех возможных регрессий, метода исключения, метода включения?

3. Каковы особенности и назначение алгоритма многорядной селекции моделей?

4. Приведите пример моделей многорядной селекции, построенных на основе учета парных взаимодействий.

5. Как оценивается эффективность каждого очередного ряда селекции? Когда нужно прекратить селекцию?

ГЛАВА 12. МОДЕЛИ ИДЕНТИФИКАЦИИ ДЛЯ (ПЛАНИРУЕМЫЙ ЭКСПЕРИМЕНТ)

12.1. ПОЛНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ

Планирование эксперимента позволяет повысить его информативность и, следовательно, увеличить точность построенных на его основе моделей идентификации или уменьшить объем экспериментальных данных, что особенно существенно в связи с высокой стоимостью экспериментального изучения ДЛА. Одновременно упрощаются методы построения модели. Планирование эксперимента подразумевает возможность активного вмешательства в эксперимент, т. е. возможность управления факторами x_i . Кроме того, они не должны зависеть друг от друга, т. е. допускать возможность задания каждого из факторов на любом желаемом уровне.

Планирование эксперимента возможно лишь в том случае, когда форма модели идентификации задана заранее и задача заключается в определении коэффициентов модели.

Задача планирования эксперимента состоит в выборе оптимальной совокупности точек k -мерного факторного пространства — векторов $\mathbf{x}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, эксперимент в которых обеспечивает получение максимума информации. Каждый фактор x_i меняется в определенном диапазоне $x_{i \min} \leq x_i \leq x_{i \max}$ (1.4), а фактически принимает ряд (r_i) значений из этого диапазона. Общее число возможных экспериментов дается соотношением:

$$n = \prod_{i=1}^k r_i, \quad (12.1)$$

а при одинаковом числе фиксированных значений (уровней) для всех факторов ($r_i = r = \text{const}$)

$$n = r^k. \quad (12.2)$$

Эксперимент, при котором реализуются все возможные сочетания факторов, принято называть *полным факторным экспериментом* (ПФЭ). Очевидно (12.2), что в многофакторных задачах при нескольких фиксированных уровнях факторов общее число экспериментов оказывается чрезвычайно большим. На практике обычно используются планы, предусматривающие фиксирование всех факторов на двух (планы 2^k , $n = 2^k$) или, существенно реже, на трех (планы 3^k , $n = 3^k$) уровнях.

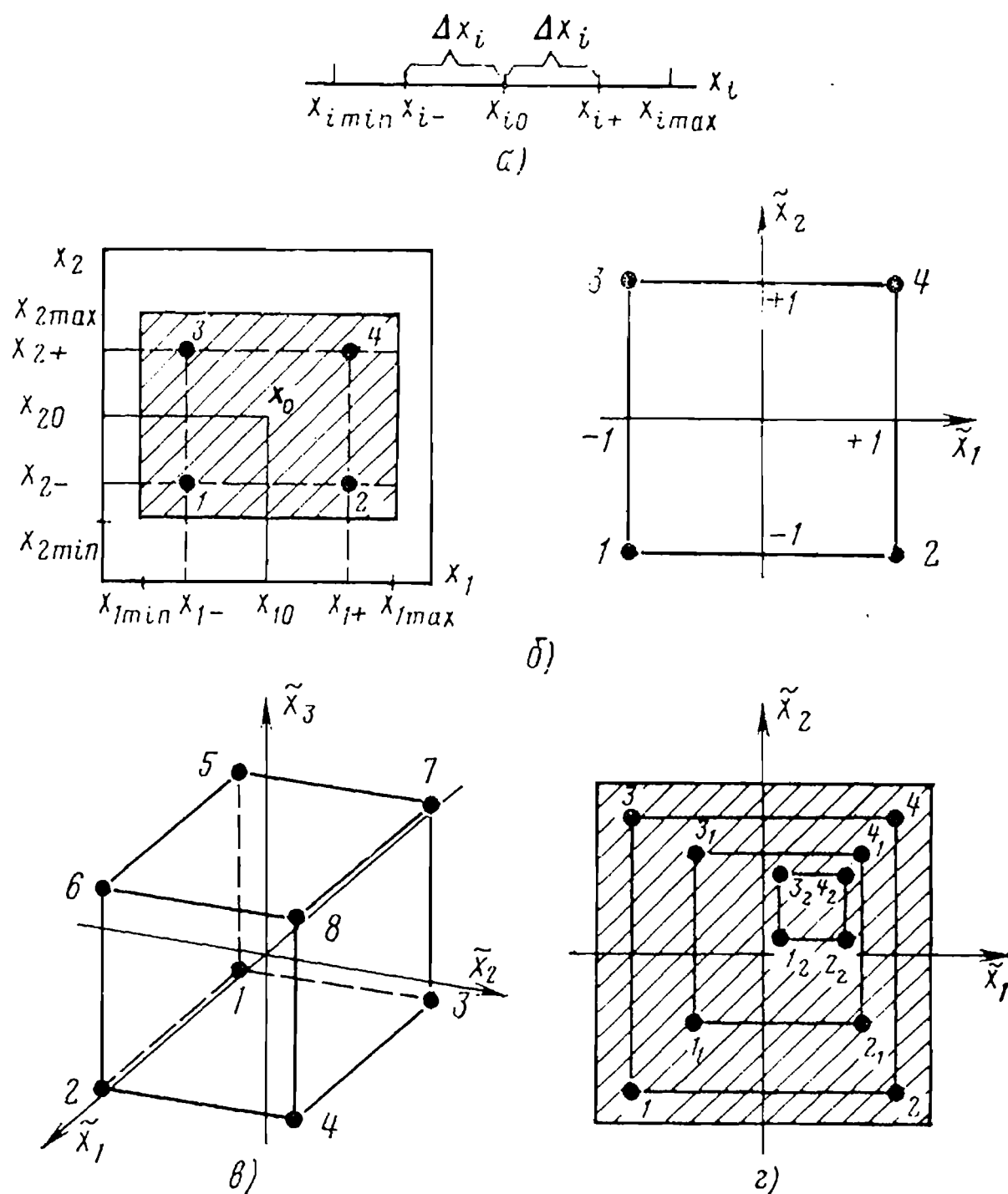


Рис. 12.1. Планирование эксперимента 2^k :
 а — выбор уровней факторов; б — план эксперимента 2^2 ; в — план эксперимента 2^3 ; г — последовательность сжимающихся планов

При составлении плана ПФЭ 2^k для каждого фактора выбирается нулевой (базовый) уровень x_{i0} , соответствующий *центру* эксперимента. Обычно, но не обязательно, он выбирается в центре диапазона варьирования (рис. 12.1, а):

$$x_{i0} = \frac{1}{2} (x_{i \max} - x_{i \min}). \quad (12.3)$$

Выбирается шаг варьирования Δx_i , характеризующий отклонение фактора от нулевого уровня. Эксперимент проводится при двух значениях фактора x_{i-} и x_{i+} :

$$x_{i-} = x_{i0} - \Delta x_i; \quad x_{i+} = x_{i0} + \Delta x_i. \quad (12.4)$$

Шаг варьирования выбирается на основе опыта и интуиции исследователя и обычно составляет

$$\Delta x = (0,2 \dots 0,3) (x_{i \max} - x_{i \min}). \quad (12.5)$$

Для облегчения расчетов удобно провести нормировку факторов с помощью преобразования

$$\tilde{x}_i = \frac{x_i - x_{i0}}{\Delta x_i}. \quad (12.6)$$

№ эксперимента	Матрица планирования 2 ²				Код	Резуль- таты W
	$\widetilde{x}_0$ $\widetilde{u}_0$	План		$\widetilde{x}_1\widetilde{x}_2$ $\widetilde{u}_3$		
		$\widetilde{x}_1$ $\widetilde{u}_1$	$\widetilde{x}_2$ $\widetilde{u}_2$			
1	+	—	—	+	(1)	W ₁
2	+	+	—	—	a	W ₂
3	+	—	+	—	b	W ₃
4	+	+	+	+	ab	W ₄

При этом нижнее и верхнее значения фактора $\tilde{x}_i$ (12.6) равны -1 и $+1$:

$$\tilde{x}_{i-} = -1; \tilde{x}_{i+} = +1. \tag{12.7}$$

(В таблицах планирования эксперимента обычно вместо -1 и $+1$ пишут просто $+u-$). Расположение экспериментов при планировании 2^k при $k=2$ и $k=3$ показано на рис. 12.1, б, в.

При разработке моделей идентификации для описания отдельных участков поверхности отклика часто используется последовательность «сжимающихся» планов. На первом этапе для описания области в целом параметры плана выбираются в соответствии с выражениями (12.3) — (12.5). Потом с помощью модели выделяется наиболее интересная подобласть (например, окрестность экстремума) и в ней ставится новый эксперимент. В качестве центра такого эксперимента можно выбрать, например, приближенно определенную точку экстремума. Процедура повторяется несколько раз (рис. 12.1, г).

Методику построения факторных планов легче всего рассмотреть на конкретных примерах. Пусть при исследовании процессов в камере сгорания ГТД варьируется температура $T(x_1)$ и давление $p(x_2)$ газа перед турбиной:

$$1200 \leq T \leq 1500 \text{ К}; 15 \cdot 10^5 \leq p \leq 30 \cdot 10^5 \text{ Н/м}^2.$$

С помощью соотношений (12.3) — (12.4) можно получить

$$x_{10} = 1350 \text{ К}; x_{20} = 22,5 \cdot 10^5 \text{ Н/м}^2;$$

$$x_{1-} = 1300\text{К}; x_{1+} = 1400 \text{ К}; x_{2-} = 18 \text{ Н/м}^2; x_{2+} = 27 \text{ Н/м}^2.$$

Все возможные комбинации для двух факторов, варьируемых на двух уровнях, будут исчерпаны в четырех экспериментах. План таких экспериментов после нормировки факторов с помощью соотношения (12.6) задается матрицей планирования 2^2 (табл. 12.1). В первом столбце матрицы указан номер эксперимента, во втором указаны значения фиктивной переменной $\tilde{x}_0=1$, в третьем и четвертом — значения переменных $\tilde{x}_1, \tilde{x}_2$ (нормированные температура и давление), образующие собственно планирование эксперимента, в

пятом — произведение факторов (двойное взаимодействие) $\tilde{x}_1\tilde{x}_2$, определяемое как скалярное произведение столбцов $\tilde{x}_1$ и $\tilde{x}_2$. Для удобства выкладок наряду с реальными переменными $\tilde{x}_i$ в столбцах 2—5 указаны условные переменные $\tilde{u}_i$. В шестом столбце приведено условное буквенное обозначение — *код эксперимента*, в седьмом — значение функции отклика в данном эксперименте. Первая строка соответствует первому эксперименту (рис. 12.1, б), когда оба фактора находятся на нижнем уровне, вторая строка — второму эксперименту (первый фактор на верхнем, второй — на нижнем уровне) и т. д.

Для повышения точности результатов целесообразно при каждом сочетании факторов проводить не один, а несколько экспериментов и в качестве выходного значения функции отклика принимать их среднее. План такого эксперимента для случая двух факторов при трех экспериментах для каждого сочетания уровней приведен в табл. 12.1а. Порядок проведения всех (в данном случае

Таблица 12.1а

№ экспе- римен- та	Матрица планирования 2 ²				Код	Результат			$\frac{W}{\overline{W}}$
	$\tilde{x}_0$ $\tilde{u}_0$	План		$\tilde{x}_1\tilde{x}_2$ $\tilde{u}_3$		1	2	3	
		$\tilde{x}_1$ $\tilde{u}_1$	$\tilde{x}_2$ $\tilde{u}_2$						
1	+	—	—	+	(1)	W_{11}	W_{12}	W_{13}	$\overline{W}_1$
2	+	+	—	—	a	W_{21}	W_{22}	W_{23}	$\overline{W}_2$
3	+	—	+	—	b	W_{31}	W_{32}	W_{33}	$\overline{W}_3$
4	+	+	+	+	ab	W_{41}	W_{42}	W_{43}	$\overline{W}_4$

12-ти) экспериментов должен быть *рандомизирован*, т. е. определен с помощью таблицы случайных чисел, что позволяет исключать влияние временного дрейфа характеристик исследуемой системы.

Планирование 2² позволяет определить четыре коэффициента регрессии. Модель может быть представлена в виде неполного квадратного уравнения, включающего свободный член, линейные эффекты и эффект двойного взаимодействия факторов:

$$W = b_0\tilde{x}_0 + b_1\tilde{x}_1 + b_2\tilde{x}_2 + b_{12}\tilde{x}_1\tilde{x}_2 \quad (W = b_0\tilde{u}_0 + b_1\tilde{u}_1 + b_2\tilde{u}_2 + b_3\tilde{u}_3) \quad (12.8)$$

В этом случае число оцениваемых коэффициентов равно числу экспериментов и не остается степеней свободы для проверки адекватности модели. Если есть основания считать, что изучаемый объект может быть описан линейной моделью

$$W = b_0\tilde{x}_0 + b_1\tilde{x}_1 + b_2\tilde{x}_2 \quad (W = b_0\tilde{u}_0 + b_1\tilde{u}_1 + b_2\tilde{u}_2) \quad (12.8a)$$

то после определения коэффициентов (b_0, b_1, b_2) остается одна степень свободы для проверки адекватности модели в целом.

№ экспе- римента	Матрица планирования 2 ³									Код	Резуль- тат W
	$\widetilde{x_0}$ $\widetilde{u_0}$	План			$\widetilde{x_1x_2}$ $\widetilde{u_4}$	$\widetilde{x_1x_3}$ $\widetilde{u_5}$	$\widetilde{x_2x_3}$ $\widetilde{u_6}$	$\widetilde{x_1x_2x_3}$ $\widetilde{u_7}$			
		$\widetilde{x_1}$ $\widetilde{u_1}$	$\widetilde{x_2}$ $\widetilde{u_2}$	$\widetilde{x_3}$ $\widetilde{u_3}$							
1	+	—	—	—	+	+	+	—	(1)	W ₁	
2	+	+	—	—	—	—	+	+	a	W ₂	
3	+	—	+	—	—	+	—	+	b	W ₃	
4	+	+	+	—	+	—	—	—	ab	W ₄	
5	+	—	—	+	+	—	—	+	c	W ₅	
6	+	+	—	+	—	+	—	—	ac	W ₆	
7	+	—	+	+	—	—	+	—	bc	W ₇	
8	+	+	+	+	+	+	+	+	abc	W ₈	

Переход к планированию 2³ требует двукратного повторения планирования 2², причем в первом случае $\tilde{x}_3$ поддерживается на нижнем, а во втором — на верхнем уровне (рис. 12.1, в). Матрица планирования приведена в табл. 12.2. Модель может быть представлена в виде

$$W = b_0\tilde{x}_0 + b_1\tilde{x}_1 + b_2\tilde{x}_2 + b_3\tilde{x}_3 + b_{12}\tilde{x}_1\tilde{x}_2 + b_{13}\tilde{x}_1\tilde{x}_3 + b_{23}\tilde{x}_2\tilde{x}_3 + b_{123}\tilde{x}_1\tilde{x}_2\tilde{x}_3 \tag{12.9}$$

$$(W = b_0\tilde{u}_0 + b_1\tilde{u}_1 + b_2\tilde{u}_2 + b_3\tilde{u}_3 + b_4\tilde{u}_4 + b_5\tilde{u}_5 + b_6\tilde{u}_6 + b_7\tilde{u}_7) ,$$

и включает в себя свободный член, линейные эффекты, эффекты двойных взаимодействий и тройного взаимодействия.

Модели (10.8), (10.9) не содержат квадратов факторов ($\tilde{x}_i^2$, $\tilde{x}_i^2\tilde{x}_j$) и линейны относительно каждой из переменных¹.

Аналогично можно построить матрицы планирования 2^k для задач большей размерности. При этом всегда дважды повторяется планирование меньшей размерности 2^{k-1}, причем первый раз новый фактор берется на нижнем, а второй раз — на верхнем уровне.

Для компактной записи плана эксперимента, а также отдельных экспериментов используется буквенное обозначение — код эксперимента. Каждому из варьируемых факторов сопоставляется буква латинского алфавита ($\tilde{x}_1 \rightarrow a$; $\tilde{x}_2 \rightarrow b$; $\tilde{x}_3 \rightarrow c...$), и в буквенное обозначение выносятся буквы, соответствующие факторам, находящимся на верхнем уровне.

Так, для факторного планирования 2³ код *ab* указывает на то, что на верхнем уровне заданы факторы $\tilde{x}_1$ и $\tilde{x}_2$, а фактор $\tilde{x}_3$ нахо-

¹ При фиксированных $\tilde{x}_2 = \text{const}$ и $\tilde{x}_3 = \text{const}$ выражение (12.9) линейно относительно $\tilde{x}_1$.

дится на нижнем уровне. Код bc означает, что фактор $\tilde{x}_1$ задан на нижнем, а факторы $\tilde{x}_2, \tilde{x}_3$ — на верхнем уровне. Для эксперимента, где все факторы заданы на нижнем уровне, используется обозначение (1).

Полные факторные планы 2^2 и 2^3 можно задать кодами

$$(1), a, b, ab; \quad (12.10)$$

$$(1), a, b, ab, c, ac, bc, abc.$$

Код ПФЭ 2^4 задается как дважды повторенный код ПФЭ 2^3 , причем во второй раз код домножается на символ фактора $\tilde{x}_4 \rightarrow d$:

$$(1), a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, abcd. \quad (12.11)$$

Кодовые обозначения удобны при построении и анализе ПФЭ большой размерности. Непосредственный анализ матриц планирования (табл. 12.1, 12.2) показывает, что полный факторный эксперимент обладает рядом важных свойств:

$$\begin{aligned} \sum_{r=1}^n \tilde{u}_{ri} &= 0; \quad i = 1, 2, \dots, p; \\ \sum_{r=1}^n \tilde{u}_{ri}^2 &= n; \quad i = 0, 1, 2, \dots, p; \end{aligned} \quad (12.12)$$

$$\sum_{r=1}^n \tilde{u}_{ri} \tilde{u}_{rj} = 0; \quad i, j = 0, 1, 2, \dots, p; \quad i \neq j,$$

где i, j — номер столбца в матрице планирования.

Первое из соотношений (12.12) (оно, очевидно, не может выполняться при $i=0$) характеризует симметричность расположения экспериментов относительно базовой точки и обеспечивает равномерность распределения информации в исследуемой области. Второе соотношение (12.12) является следствием нормировки и указывает на равенство суммы квадратов элементов для всех столбцов. Третье соотношение (12.12) указывает на ортогональность всех столбцов матрицы планирования (скалярное произведение любых двух столбцов равно нулю).

Из третьего соотношения (12.12) следует, что матрица коэффициентов нормальных уравнений $\mathbf{U}^* \mathbf{U}$ (см. табл. 11.1), а также обратная ей матрица $(\mathbf{U}^* \mathbf{U})^{-1}$ диагональны. В силу второго соотношения (12.12) все диагональные элементы матрицы $\mathbf{U}^* \mathbf{U}$ равны n , а диагональные элементы матрицы $(\mathbf{U}^* \mathbf{U})^{-1}$ равны $1/n$. При этом формулы (11.8) для определения коэффициентов регрессии существенно упрощаются и принимают вид

$$b_i = \frac{\sum_{r=1}^n \tilde{u}_{ri} W_r}{n}; \quad i = 0, 1, 2, \dots, p. \quad (12.13)$$

Таким образом, для определения коэффициента регрессии b_i достаточно просто просуммировать все значения функции отклика с учетом знаков элементов соответствующего столбца матрицы планирования и разделить полученный результат на общее число экспериментов.

Зависимости для статистического анализа коэффициентов регрессии также упрощаются:

$$\sigma^2(b_i) = \frac{\sigma^2(W)}{n}; \quad i=0, 1, 2, \dots, p;$$

$$\text{cov}(b_i, b_j) = 0; \quad i \neq j; \quad i, j = 0, 1, 2, \dots, p; \quad (12.14)$$

$$r(b_i, b_j) = 0; \quad i \neq j; \quad i, j = 0, 1, 2, \dots, p.$$

Все коэффициенты регрессии при ПФЭ определяются *независимо* друг от друга. Поэтому в отличие от неорганизованного эксперимента (см. гл. 11) коэффициентам регрессии можно дать обоснованное физическое толкование, а в случае необходимости отказаться от учета тех или иных членов модели без пересчета коэффициентов регрессии при остальных членах.

Из сказанного ясно, почему модели, построенные методами факторного планирования эксперимента, не включают в себя членов с квадратами факторов ($\tilde{x}_i^2$, $\tilde{x}_i^2 \tilde{x}_j$ и т. д.). Дело в том, что столбцы матрицы планирования для этих членов полностью совпали бы с уже имеющимися столбцами для свободного члена ($\tilde{x}_0$) и линейных эффектов ($\tilde{x}_j$). (Ведь все элементы столбца планирования для $\tilde{x}_i^2$ равны ± 1). Эффекты с одинаковыми столбцами не различимы и, следовательно, введение квадратов не имеет практического смысла. Коэффициенты регрессии b_0 и b_i по существу задают *совместные* оценки коэффициентов регрессии:

$$b_0 \Rightarrow \beta_0 + \sum_{i=1}^k \beta_{ii}; \quad b_i \Rightarrow \beta_i + \sum_{j=1}^k \beta_{ijj}. \quad (12.15)$$

При построении полной модели идентификации в виде (12.8), (12.9) все эксперименты, предусмотренные полным факторным планированием, используются для определения коэффициентов регрессии. Поэтому для проверки адекватности модели целесообразно *дополнительно* провести один или несколько (n_0) экспериментов в центральной точке ($\tilde{x}_1=0$, $\tilde{x}_2=0$, ..., $\tilde{x}_k=0$) (рис. 12.1, б, в).

Проверка адекватности модели осуществляется с помощью критерия Фишера (3.16). Оценка дисперсии модели s_M^2 определяется по остаточной сумме квадратов всех экспериментов¹:

$$s_M^2 = \frac{\sum_{r=1}^{n+n_0} (W_r - W_{rM})^2}{f_s}; \quad f_s = n_0, \quad (12.16)$$

а дисперсия эксперимента $\sigma^2(W)$ или задана из условий эксперимента, или определяется оценкой s_σ^2 по экспериментам в центральной точке:

$$s_\sigma^2 = \frac{\sum_{i=1}^{n_0} (W_{0i} - \bar{W}_0)^2}{n_0 - 1}, \quad (12.17)$$

где

$$\bar{W}_0 = \frac{1}{n_0} \sum_{i=1}^{n_0} W_{0i}. \quad (12.18)$$

Если критерий Фишера оказывается меньше порогового табличного значения при принятом уровне значимости α

$$F_{f_s, n_0-1} = \frac{s_M^2}{s_\sigma^2} \leq F_{f_s, [n_0-1, \alpha]}, \quad (12.19)$$

модель может быть признана адекватной действительности.

Кроме проверки адекватности модели в целом, необходимо проверить значимость отдельных членов модели. Эта оценка также осуществляется с помощью критерия Фишера, причем подсчитывается дополнительная сумма квадратов (3.20) и квадрат этой суммы (3.21), связанные с проверяемым членом (группой членов) модели (см. разд. 3.3). Для оценки роли отдельных членов модели можно использовать также методы корреляционного анализа (см. разд. 3.4).

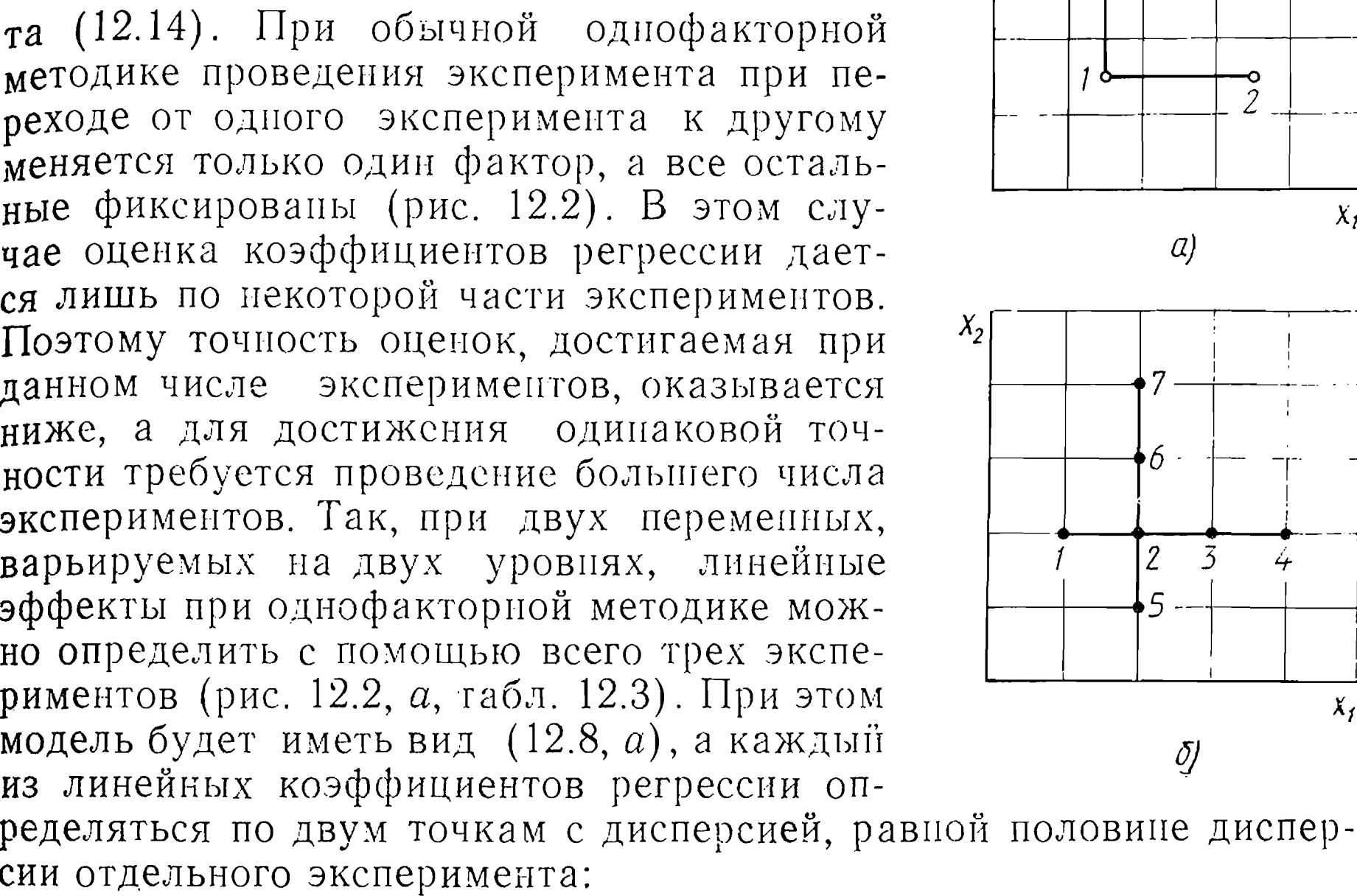
Факторное планирование является оптимальным планированием в широком смысле этого слова. В частности, показано, что при выполнении условий (12.12) коэффициенты регрессии оцениваются с наименьшей возможной дисперсией. Дело в том, что при факторном планировании одновременно варьируются все переменные, каждый коэффициент регрессии определяется по результатам *всех* экспериментов и, следовательно, дисперсия коэффициентов регрессии оказывается в n раз меньше дисперсии отдельного эксперимен-

¹ Если при каждом сочетании факторов проводится m экспериментов (табл. 12.1 а), оценка s_M^2 дается формулой

$$s_M^2 = \frac{\sum_{r=1}^{mn} (W_r - W_{rM})^2 + \sum_{r=1}^{n_0} (W_{0r} - W_{0rM})^2}{f_s};$$

$$f_s = (m-1)n + n_0.$$

Рис. 12.2. Схема расположения экспериментальных точек при однофакторном планировании:
а — три эксперимента; *б* — семь экспериментов



$$\sigma^2(b_i) = \frac{\sigma^2(W)}{2}; \quad i = 1, 2. \tag{12.20}$$

При ПФЭ 2^2 (см. рис. 12.1, б, табл. 12.1) при четырех экспериментах дисперсия всех, в том числе и линейных коэффициентов регрессии, будет в четыре раза меньше дисперсии отдельного эксперимента (12.14). Для достижения такой же точности оценки линейных коэффициентов регрессии при однофакторной методике потребовалось бы семь экспериментов (рис. 12.2, б).

Выигрыш от применения ПФЭ возрастает при увеличении размерности задачи. Так, при $k=3$ полное факторное планирование позволяет с помощью восьми экспериментов оценить коэффициенты регрессии с точностью $\sigma^2(W)/8$. Для достижения такой же точности

Т а б л и ц а 12.3

№ экспери- мента	Матрица планирования			Код	Результат W
	$\tilde{x}_0$ u_0	План			
		$\tilde{x}_1$ u_1	$\tilde{x}_2$ u_2		
1	+	—	—	(1)	W_1
2	+	+	—	a	W_2
3	+	—	+	b	W_3

линейных коэффициентов регрессии при однофакторном планировании потребовалось бы 22 эксперимента. При $k=5$ точность оценки коэффициентов регрессии $\sigma^2(W)/32$ достигается при полном факторном планировании при 32 экспериментах, а при однофакторном — при 156 экспериментах и т. д.

Другим важным свойством ПФЭ 2^k является его *ротатабельность*. Это значит, что информация, содержащаяся в модели, равномерно «размазана» по сфере (гиперсфере), т. е. модель является одинаково точной для эквидистантных сфер, описанных в области исследования из центральной точки эксперимента. При постановке эксперимента заранее неизвестно, какие точки области представляют наибольший интерес. Поэтому естественно стремиться к ротатабельному планированию. Важно подчеркнуть, что ПФЭ 2^k одновременно является и ортогональным и ротатабельным.

Таким образом, при ПФЭ 2^k :

— все коэффициенты регрессии определяются с одинаковой и минимальной дисперсией, т. е. извлекается максимальная информация при минимуме экспериментальных точек;

— все коэффициенты регрессии определяются независимо друг от друга (планирование ортогонально);

— информация, содержащаяся в модели, равномерно распределена вокруг центра эксперимента (планирование ротатабельно);

— вычисление коэффициентов регрессии осуществляется по очень простым формулам.

В некоторых случаях, особенно при больших размерностях факторного пространства, такая точность коэффициентов регрессии (12.14) является избыточно высокой. В этих случаях целесообразно использовать упрощенные факторные планы, так называемые дробные факторные планы (см. разд. 12.2).

Вопросы для самоконтроля

1. Что называется полным факторным экспериментом?
2. Запишите матрицы ПФЭ 2^2 , 2^3 , 2^4 .
3. Что такое код эксперимента? Для чего он используется?
4. Запишите код эксперимента 2^5 .
5. В чем заключается свойство ортогональности ПФЭ?
6. Запишите и объясните статистические характеристики ПФЭ, формулы для расчета коэффициентов регрессии при ПФЭ.
7. В чем заключается свойство ротатабельности ПФЭ?

12.2 ДРОБНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ

Полное факторное планирование 2^k может быть подвергнуто критике с двух прямо противоположных позиций.

Во-первых, при больших размерностях задачи число необходимых экспериментов оказывается слишком большим (при $k=7$ $n=128$, при $k=15$ $n=32768$), а полученная модель идентификации содержит много избыточных членов (например априори можно считать несущественными шестерные, семерные взаимодействия факторов, а также взаимодействия более высоких порядков).

Во-вторых, наоборот, особенно при малой размерности факторного пространства или в отдельных его областях точность модели, полученной с помощью ПФЭ 2^k , можно оказаться недостаточной.

Для упрощения модели и уменьшения числа экспериментов переходят к так называемым *дробным* факторным планам, представляющим собой отдельные блоки ПФЭ 2^k . Уточнение модели, связанное с увеличением числа входящих в нее членов, может быть осуществлено при переходе к ПФЭ 3^k или при использовании *композиционных* планов, построенных на основе ПФЭ 2^k и включающих в себя дополнительные точки (см. разд. 12.3).

При построении дробных факторных планов можно сохранить свойства ПФЭ 2^k (12.12) — (12.14). Метод построения дробных факторных планов легче всего продемонстрировать на конкретных примерах. Допустим, что требуется построить приближенную линейную модель идентификации поверхности отклика трех переменных (температура, давление, расход) в виде

$$W=b_0\tilde{x}_0+b_1\tilde{x}_1+b_2\tilde{x}_2+b_3\tilde{x}_3. \tag{12.21}$$

Для решения этой задачи достаточно четырех экспериментов, планируемых в соответствии с матрицей ПФЭ 2^2 (табл. 12.1), содержащей как раз четыре эксперимента, обладающих свойствами (12.12) — (12.14).

При построении плана дополнительная по отношению к плану 2^2 переменная $\tilde{x}_3$ приравнивается к двойному взаимодействию $\tilde{x}_1\tilde{x}_2$ (единственному нелинейному члену модели (12.8), причем возможно два равноценных *генерирующих соотношения*:

$$\tilde{x}_3=\tilde{x}_1\tilde{x}_2; \tilde{x}_3=-\tilde{x}_1\tilde{x}_2 \tag{12.22}$$

В результате будут получены две матрицы планирования для трех факторов, содержащие по четыре эксперимента (табл. 12.4, 12.5). Планы, записанные в этих таблицах, принято называть *полурепликами* полного факторного эксперимента и обозначать 2^{3-1} . Расположение экспериментов для этих полуреплик показано на рис. 12.3, б.

Легко видеть, что построенные планы 2^{3-1} :
— удовлетворяют уравнениям (12.12), а значит, и статистическим оценкам (12.14);

Т а б л и ц а 12.4

№ эксперимента	Матрица 2^{3-1}				Код	Результаты W
	$\widetilde{x_0}$ $\widetilde{u_0}$	План				
		$\widetilde{x_1}$ $\widetilde{u_1}$	$\widetilde{x_2}$ $\widetilde{u_2}$	$\widetilde{x_3}$ $\widetilde{u_3}$		
1	+	—	—	+	c	W_1
2	+	+	—	—	a	W_2
3	+	—	+	—	b	W_3
4	+	+	+	+	abc	W_4

№ экспери- мента	Матрица 2^{3-1}				Код	Результаты W
	$\tilde{x}_0$ $\tilde{u}_0$	План				
		$\tilde{x}_1$ $\tilde{u}_1$	$\tilde{x}_2$ $\tilde{u}_2$	$\tilde{x}_3$ $\tilde{u}_3$		
1	+	—	—	—	(1)	W_1
2	+	+	—	+	ac	W_2
3	+	—	+	+	bc	W_3
4	+	+	+	—	ab	W_4

— совместно образуют полное факторное планирование 2^3 ;
— строятся по специальному принципу, в соответствии с кото-
рым в одну полуреплику отобраны эксперименты из 2^3 с *нечетным*
количеством букв в коде (c, a, b, abc), а в другую — с *четным* ко-
личеством букв в коде ($(1), ab, ac, bc$).

Аналогично могут быть построены полуреплики при более вы-
сокой размерности факторного простран-
ства. Так, для оценки линейной модели четы-
рех переменных, содержащей пять членов,
можно использовать полуреплики 2^{4-1} ПФЭ
 2^4 . Генерирующие соотношения могут быть
заданы в виде

$$\tilde{x}_4 = \tilde{x}_1 \tilde{x}_2 \tilde{x}_3; \quad \tilde{x}_4 = -\tilde{x}_1 \tilde{x}_2 \tilde{x}_3 \quad (12.23)$$

а дробные факторные планы характеризу-
ются кодами¹

$$(1), ad, bd, ab, cd, ac, bc, abcd; \quad (12.24)$$

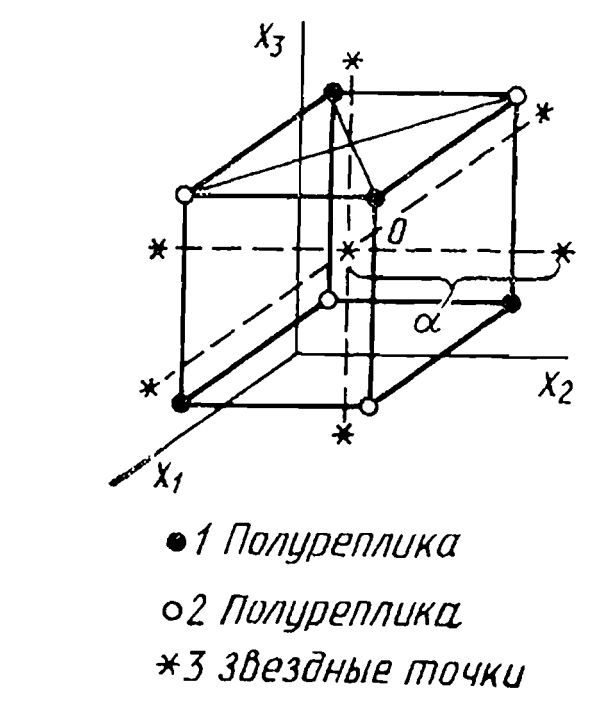
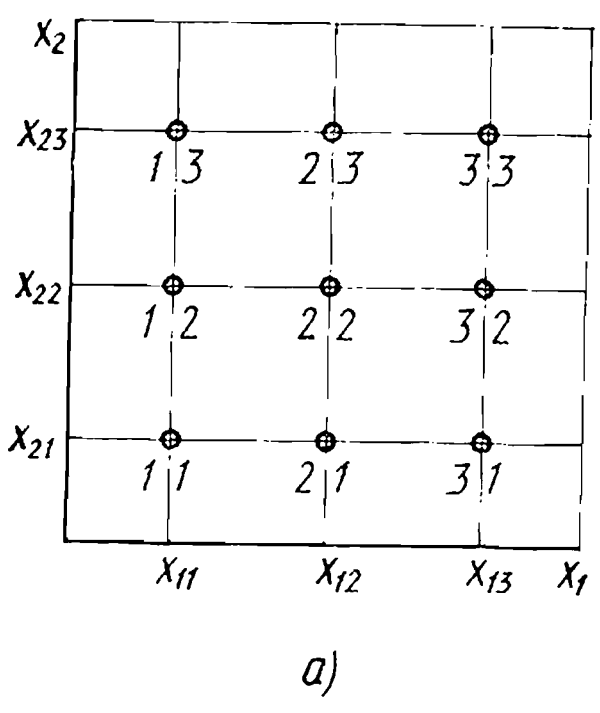
$$d, a, b, abd, c, acd, bcd, abc.$$

Как и в предыдущем случае, планы
(12.24) совместно образуют ПФЭ 2^4 . Каж-
дый план 2^{4-1} содержит 8 экспериментов.
Поэтому при оценке коэффициентов линей-
ной модели

$$W = b_0 \tilde{x}_0 + b_1 \tilde{x}_1 + b_2 \tilde{x}_2 + b_3 \tilde{x}_3 + b_4 \tilde{x}_4 \quad (12.25)$$

остается три степени свободы для оценки
адекватности модели. В принципе планиро-
вание 2^{4-1} позволяет наряду с линейными

Рис. 12.3. Схемы планирования эксперимента:
а — планирование 3^2 ; б — композиционное планирование



¹ Читателю рекомендуется построить необходимые матрицы и проделать вы-
кладки.

эффектами оценить и эффекты трех двойных взаимодействий ($\tilde{x}_1\tilde{x}_2$; $\tilde{x}_1\tilde{x}_3$; $\tilde{x}_2\tilde{x}_3$). Однако вводить соответствующие члены в модель целесообразно лишь в том случае, когда есть основания полагать, что их значимость выше значимости отброшенных эффектов ($\tilde{x}_1\tilde{x}_4$; $\tilde{x}_2\tilde{x}_4$; $\tilde{x}_3\tilde{x}_4$).

В соотношениях (12.23) дополнительные факторы $\tilde{x}_4$ приравнены наиболее сложным (тройным) взаимодействиям, так как их влияние на модель в целом наименее существенно. Однако возможны другие схемы построения полуреплик. Так, вместо (12.23) можно было бы воспользоваться *любой* из следующих пар генерирующих соотношений:

$$\begin{aligned} 1. \quad & \tilde{x}_4 = \tilde{x}_1\tilde{x}_2, \quad \tilde{x}_4 = -\tilde{x}_1\tilde{x}_2; \\ 2. \quad & \tilde{x}_4 = \tilde{x}_1\tilde{x}_3, \quad \tilde{x}_4 = -\tilde{x}_1\tilde{x}_3; \\ 3. \quad & \tilde{x}_4 = \tilde{x}_2\tilde{x}_3, \quad \tilde{x}_4 = -\tilde{x}_2\tilde{x}_3, \end{aligned} \quad (12.26)$$

что привело бы к другим планам 2^{4-1} . Так, при использовании первой пары генерирующих соотношений (12.26) планы эксперимента имеют вид

$$\begin{aligned} & d, a, b, abd, cd, ac, bc, abcd; \\ & (1), ad, bd, ab, c, acd, bcd, abc. \end{aligned} \quad (12.24a)$$

Так же как планы (12.24), планы (12.24a) совместно образуют полный факторный эксперимент 2^4 . Однако они включают в себя эксперименты как с четными, так и с нечетными кодами и образованы как бы смешением планов (12.24).

Если предполагается сохранить в модели кроме линейных членов какие-либо из взаимодействий, полезно провести предварительное статистическое исследование для выявления наиболее значимых из них. В этом случае выбор генерирующих соотношений осуществляется с учетом результатов такого исследования.

Для задач высокой размерности можно построить факторные планы более высокой степени дробности. Так, при построении линейной модели пяти переменных (модель содержит шесть членов) можно воспользоваться матрицей планирования ПФЭ 2^3 при одной из комбинаций генерирующих соотношений¹:

$$\begin{aligned} 1. \quad & \tilde{x}_5 = \tilde{x}_1\tilde{x}_2\tilde{x}_3; \quad \tilde{x}_4 = \tilde{x}_1\tilde{x}_2; \\ 2. \quad & \tilde{x}_5 = -\tilde{x}_1\tilde{x}_2\tilde{x}_3; \quad \tilde{x}_4 = -\tilde{x}_1\tilde{x}_2; \\ 3. \quad & \tilde{x}_5 = \tilde{x}_1\tilde{x}_2\tilde{x}_3; \quad \tilde{x}_4 = -\tilde{x}_1\tilde{x}_2; \\ 4. \quad & \tilde{x}_5 = -\tilde{x}_1\tilde{x}_2\tilde{x}_3; \quad \tilde{x}_4 = \tilde{x}_1\tilde{x}_2, \end{aligned} \quad (12.27)$$

приводящих к планам 2^{5-2} в виде

$$\begin{aligned} & d, ae, be, abd, cde, ac, bc, abcde; \\ & e, ad, bd, db e, c, acde, bcde, abc; \\ & (1), ade, bde, ab, ce, acd, bcd, abce; \\ & de, a, b, abde, cd, ace, bce, abcd; \end{aligned} \quad (12.28)$$

¹ Возможны также другие комбинации генерирующих соотношений, например $\tilde{x}_5 = \tilde{x}_1\tilde{x}_2\tilde{x}_3$; $\tilde{x}_4 = \tilde{x}_1\tilde{x}_3$ и т. д.

Планы (12.28) называются *четвертьрепликами*, не содержат одинаковых экспериментов и совместно образуют ПФЭ 2^5 .

Наконец, при оценке линейной модели семи переменных также можно воспользоваться планом ПФЭ 2^3 при полной замене всех взаимодействий новыми факторами. При этом возможны различные комбинации планов 2^{7-4} , один из которых задается генерирующими соотношениями

$$\tilde{x}_4 = \tilde{x}_1 \tilde{x}_2 \tilde{x}_3; \quad \tilde{x}_5 = \tilde{x}_1 \tilde{x}_2; \quad \tilde{x}_6 = \tilde{x}_1 \tilde{x}_3; \quad \tilde{x}_7 = \tilde{x}_2 \tilde{x}_3. \quad (12.29)$$

Важно подчеркнуть, что все рассмотренные дробные факторные планы несмотря на огромное многообразие в силу построения удовлетворяют соотношениям (12.12), (12.13), (12.14), т. е. обладают замечательными свойствами ПФЭ. Вместе с тем переход к дробным факторным планам не дается даром. Если в полном факторном эксперименте все коэффициенты регрессии оцениваются независимо друг от друга, то при дробном факторном планировании необходимо дополнительно исследовать связи между коэффициентами.

Например, планирование 2^{3-1} (см. табл. 12.4) позволяет независимо друг от друга оценить свободный член и линейные эффекты модели (12.9) при условии, что эффекты двойных взаимодействий и тройного взаимодействия равны нулю ($\beta_{i,j} \equiv 0, \beta_{i,j,z} \equiv 0$). Если же это не так, то найденные коэффициенты регрессии будут совместными оценками линейных эффектов и двойных взаимодействий, свободного члена и тройного взаимодействия:

$$\begin{aligned} b'_0 &\rightarrow \beta_0 + \beta_{1,2,3}; & b'_1 &\rightarrow \beta_1 + \beta_{23}; \\ b'_2 &\rightarrow \beta_2 + \beta_{13}; & b'_3 &\rightarrow \beta_3 + \beta_{12}. \end{aligned} \quad (12.30)$$

Объясняется это тем, что столбцы матрицы для этих эффектов *совпадают* между собой и, следовательно, не могут быть разделены оценки этих эффектов.

Легко показать, что вторая полуреблика (табл. 12.5) приводит к совместным оценкам вида

$$\begin{aligned} b''_0 &\rightarrow \beta_0 - \beta_{123}; & b''_1 &\rightarrow \beta_1 - \beta_{23}; \\ b''_2 &\rightarrow \beta_2 - \beta_{13}; & b''_3 &\rightarrow \beta_3 - \beta_{12}. \end{aligned} \quad (12.30a)$$

Поэтому, если после проведения экспериментов одной из полуреблик исследователь усомнится в правильности предположения о незначительности эффектов взаимодействий, он должен провести эксперименты второй полуреблики и разделить оценки по схеме

$$b_1 = \frac{b'_1 + b''_1}{2}; \quad b_{23} = \frac{b'_1 - b''_1}{2}. \quad (12.31)$$

Возможность такого разделения очевидна, так как две полуреблики образуют полное факторное планирование 2^3 .

Для исследования разрешающей способности дробных реплик вводится специальный аппарат *определяющих контрастов*. Опреде-

ляющим контрастом I называется выражение, задающее элементы столбца матрицы планирования, соответствующего фиктивной переменной x_0 , т. е. свободному члену. Все элементы этого столбца равны $+1$ (см. табл. 12.1, 12.2, 12.3 и т. д.). Очевидно, определяющий контраст может быть получен возведением в квадрат любого из столбцов матрицы планирования:

$$I = \tilde{u}_1^2 = \tilde{u}_2^2 = \dots = \tilde{u}_i^2 = \dots \quad (12.32)$$

Определяющие контрасты конкретных дробных факторных планов (реплик) могут быть получены из генерирующих соотношений умножением их на соответствующую переменную. Так, определяющие контрасты полуреplik 2^{3-1} получаются из (12.22) в виде

$$I_1 = \tilde{x}_3^2 = \tilde{x}_1 \tilde{x}_2 \tilde{x}_3, \quad I_2 = \tilde{x}_3^2 = -\tilde{x}_1 \tilde{x}_2 \tilde{x}_3. \quad (12.33)$$

С помощью определяющих контрастов легко найти соотношения, задающие все совместные оценки коэффициентов регрессии. Для этого достаточно последовательно умножить все переменные на определяющий контраст, помня, что $\tilde{x}_i^2 = 1$. В рассматриваемом случае это приведет к соотношениям

$$\begin{aligned} \tilde{x}_1 &= \tilde{x}_2 \tilde{x}_3, \quad \tilde{x}_2 = \tilde{x}_1 \tilde{x}_3, \quad \tilde{x}_3 = \tilde{x}_1 \tilde{x}_2; \\ \tilde{x}_1 &= -\tilde{x}_2 \tilde{x}_3, \quad \tilde{x}_2 = -\tilde{x}_1 \tilde{x}_3, \quad \tilde{x}_3 = -\tilde{x}_1 \tilde{x}_2. \end{aligned} \quad (12.34)$$

Кроме того, так как сам определяющий контраст равен элементам столбца $\tilde{x}_0$, то из (12.33) находим

$$\tilde{x}_0 = \tilde{x}_1 \tilde{x}_2 \tilde{x}_3; \quad \tilde{x}_0 = -\tilde{x}_1 \tilde{x}_2 \tilde{x}_3. \quad (12.35)$$

Из (12.34) и (12.35) следует, что коэффициенты регрессии, полученные с помощью рассмотренных полуреplik, являются оценками

$$\begin{aligned} b_0 &\Rightarrow \beta_0 + \beta_{123}; & b_0 &\Rightarrow \beta_0 - \beta_{123}; \\ b_1 &\Rightarrow \beta_1 + \beta_{23}; & b_1 &\Rightarrow \beta_1 - \beta_{23}; \\ b_2 &\Rightarrow \beta_2 + \beta_{13}; & b_2 &\Rightarrow \beta_2 - \beta_{13}; \\ b_3 &\Rightarrow \beta_3 + \beta_{12}; & b_3 &\Rightarrow \beta_3 - \beta_{12}, \end{aligned} \quad (12.36)$$

что совпадает с непосредственно полученным ранее результатом.

Для полуреplik 2^{4-1} (12.23) определяющие контрасты записываются в виде

$$I_1 = \tilde{x}_1 \tilde{x}_2 \tilde{x}_3 \tilde{x}_4; \quad I_2 = -\tilde{x}_1 \tilde{x}_2 \tilde{x}_3 \tilde{x}_4, \quad (12.37)$$

и приводят к совместным оценкам

$$\begin{aligned} b_0 &\Rightarrow \beta_0 + \beta_{1234}; & b_0 &\Rightarrow \beta_0 - \beta_{1234}; \\ b_1 &\Rightarrow \beta_1 + \beta_{234}; & b_1 &\Rightarrow \beta_1 - \beta_{234}; \\ b_2 &\Rightarrow \beta_2 + \beta_{134}; & b_2 &\Rightarrow \beta_2 - \beta_{134}; \\ b_3 &\Rightarrow \beta_3 + \beta_{124}; & b_3 &\Rightarrow \beta_3 - \beta_{124}; \\ b_4 &\Rightarrow \beta_4 + \beta_{123}; & b_4 &\Rightarrow \beta_4 - \beta_{123}. \end{aligned} \quad (12.38)$$

В данном случае линейные эффекты оцениваются совместно с эффектами тройных взаимодействий. Однако при построении полуреplik 2^{4-1} с помощью других генерирующих соотношений смещение оценок будет иным. Так, для первой пары генерирующих соотношений (12.26) (планирование (12.24a)) определяющие контрасты имеют вид

$$I_1 = \tilde{x}_1 \tilde{x}_2 \tilde{x}_4; \quad I_2 = -\tilde{x}_1 \tilde{x}_2 \tilde{x}_4, \quad (12.37a)$$

и приводят к оценкам вида

$$\begin{aligned} b_0 &\Rightarrow \beta_0 \pm \beta_{124}; \quad b_1 \Rightarrow \beta_1 \pm \beta_{24}; \\ b_2 &\Rightarrow \beta_2 \pm \beta_{14}; \quad b_3 \Rightarrow \beta_3 \pm \beta_{1234}; \\ b_4 &\Rightarrow \beta_4 \pm \beta_{12}, \end{aligned} \quad (12.38a)$$

т. е. в данном случае некоторые линейные эффекты оцениваются совместно с эффектами двойных взаимодействий. Кстати, именно поэтому планирование (12.23) в принципе предпочтительнее планирования (12.26).

Механизм определяющих контрастов работает и для дробных факторных планов большей дробности. При этом, однако, нужно помнить, что произведение определяющих контрастов в силу построения тоже является определяющим контрастом. Так, для четвертьреплики 2^{5-2} , заданной первой парой генерирующих соотношений (12.27), определяющий контраст имеет вид

$$I = \tilde{x}_1 \tilde{x}_2 \tilde{x}_3 \tilde{x}_5 = \tilde{x}_1 \tilde{x}_2 \tilde{x}_4 = \tilde{x}_3 \tilde{x}_4 \tilde{x}_5 \quad (12.39)$$

и задает следующие совместные оценки коэффициентов регрессии:

$$\begin{aligned} b_0 &\rightarrow \beta_0 + \beta_{1235} + \beta_{124} + \beta_{345}; \\ b_1 &\rightarrow \beta_1 + \beta_{235} + \beta_{24} + \beta_{1345}; \\ b_2 &\rightarrow \beta_2 + \beta_{135} + \beta_{14} + \beta_{2345}; \\ b_3 &\rightarrow \beta_3 + \beta_{125} + \beta_{1234} + \beta_{45}; \\ b_4 &\rightarrow \beta_4 + \beta_{12345} + \beta_{12} + \beta_{35}; \\ b_5 &\rightarrow \beta_5 + \beta_{123} + \beta_{1245} + \beta_{34}. \end{aligned} \quad (12.40)$$

Разрешающая способность этой четвертьреплики невелика, так как все линейные эффекты оцениваются совместно с эффектами двойных взаимодействий.

Вопросы для самоконтроля

1. Объясните правило построения дробных реплик.
2. Постройте дробные реплики 2^{4-1} , 2^{5-1} .
3. Что такое генерирующие соотношения?
4. Что такое определяющие контрасты? Для чего они служат?
5. Запишите определяющие контрасты и выявите связь оценок для дробных реплик

$$2^{4-1} (\tilde{x}_4 = \tilde{x}_1 \tilde{x}_2 \tilde{x}_3); \quad 2^{4-1} (\tilde{x}_4 = -\tilde{x}_2 \tilde{x}_3);$$

$$2^{5-2} (\tilde{x}_4 = \tilde{x}_1 \tilde{x}_2 \tilde{x}_3; \quad \tilde{x}_5 = \tilde{x}_2 \tilde{x}_3);$$

$$2^{7-4} (\tilde{x}_4 = \tilde{x}_1 \tilde{x}_2 \tilde{x}_3; \quad \tilde{x}_5 = \tilde{x}_2 \tilde{x}_3; \quad \tilde{x}_6 = \tilde{x}_1 \tilde{x}_3; \quad \tilde{x}_7 = \tilde{x}_1 \tilde{x}_2).$$

12.3. ПОСЛЕДОВАТЕЛЬНОЕ УСЛОЖНЕНИЕ МОДЕЛЕЙ ИДЕНТИФИКАЦИИ ПРИ ПЛАНИРУЕМОМ ЭКСПЕРИМЕНТЕ

Полный факторный эксперимент 2^k содержит информацию, достаточную для построения модели идентификации, учитывающей линейные эффекты и эффекты взаимодействий факторов. В отдельных случаях точность такой модели оказывается недостаточной. В частности, такая ситуация часто складывается в окрестности экстремума, где поверхность отклика имеет незначительную кривизну.

Увеличение точности модели за счет включения в модель членов типа $\tilde{x}_i^2$, $\tilde{x}_i^2\tilde{x}_j$ требует изменения плана эксперимента и увеличения числа опытов. Существует ряд путей усложнения плана эксперимента. К сожалению, все они приводят к существенному увеличению трудоемкости построения модели. Кроме того, недостаточно четкой является система критериев оптимальности таких планов.

Наиболее естественный способ усложнения плана эксперимента связан с переходом от ПФЭ 2^k к ПФЭ 3^k , при котором каждый из факторов фиксируется на трех уровнях. Схема такого эксперимента для двух факторов ($n=3 \times 3=9$) представлена на рис. 12.3, а. Такой план позволяет построить модель в виде

$$W = b_0\tilde{x}_0 + b_1\tilde{x}_1 + b_2\tilde{x}_2 + b_{11}\tilde{x}_1^2 + b_{22}\tilde{x}_2^2 + b_{12}\tilde{x}_1\tilde{x}_2 + \\ + b_{112}\tilde{x}_1^2\tilde{x}_2 + b_{122}\tilde{x}_1\tilde{x}_2^2 + b_{1122}\tilde{x}_1^2\tilde{x}_2^2. \quad (12.41)$$

Разработаны методы статистического анализа такого эксперимента. Однако уже для трех факторов при планировании 3^3 необходимо проведение 27 экспериментов, а полученная при этом модель (учитывающая члены вплоть до $\tilde{x}_1^2\tilde{x}_2^2\tilde{x}_3^2$) оказывается избыточно подробной. Поэтому на практике планирование 3^k используется редко.

Большее распространение получили так называемые *композиционные* планы, предусматривающие усложнение ПФЭ 2^k . В качестве примера такого плана на рис. 12.3, б показано центральное композиционное планирование для случая трех переменных. Планирование предусматривает проведение экспериментов в восьми точках, расположенных в вершинах куба в соответствии с ПФЭ 2^3 , одного или нескольких экспериментов — в центральной точке O и шести экспериментов — в «звездных» точках, расположенных за центрами граней куба на расстоянии α от центра эксперимента. Такое планирование требует проведения значительно меньшего по сравнению с ПФЭ 3^3 количества экспериментов (при проведении в центральной точке одного эксперимента планирование включает в себя 15 экспериментов вместо 27) и позволяет учесть в модели члены вида $\tilde{x}_i^2$:

$$W = b_0\tilde{x}_0 + \sum_{i=1}^3 b_i\tilde{x}_i + \sum_{\substack{i=1 \\ j>i}}^3 b_{ij}\tilde{x}_i\tilde{x}_j + b_{123}\tilde{x}_1\tilde{x}_2\tilde{x}_3. \quad (12.42).$$

При этом остается четыре степени свободы для проверки адекватности модели действительности. В случае четырех переменных композиционное планирование позволяет ограничиться проведением 25 экспериментов вместо 81 при ПФЭ 3^4 .

Для конкретизации композиционного плана необходимо выбрать величину плеча звездных точек α и задать количество экспериментов n_0 , проводимых в центральной точке. В общем виде композиционное планирование не является ни ортогональным, ни ротатабельным. Так, для матрицы центрального композиционного планирования скалярные произведения ряда столбцов не равны нулю:

$$\sum_{r=1}^n \tilde{x}_0 \tilde{x}_{ir} \neq 0; \sum_{r=1}^n \tilde{x}_{ir}^2 x_{jr}^2 \neq 0, i \neq j. \quad (12.43)$$

Все элементы столбца $\tilde{x}_0$ равны $+1$, а неотрицательные элементы столбцов $\tilde{x}_{ir}^2$ не могут все быть равны нулю. Для перехода к ортогональному планированию необходимо провести специальное преобразование квадратичных переменных и выбрать величину α . В случае трех переменных $\alpha=1,215$. Величины для некоторых задач другой размерности приведены в табл. 12.6 [14]. Однако ортогональное композиционное планирование не является ротатабельным. Для обеспечения ротатабельности плана α выбирается иначе (свойство ортогональности при этом будет утрачено) и увеличивается число экспериментов, проводимых в центральной точке. В случае трех переменных величина $\alpha=1,682$. В центральной точке плана нужно провести шесть экспериментов (общее количество экспериментов, предусмотренных таким планом, равно 20). Характеристики ротатабельных композиционных планов для некоторых задач другой размерности приведены в табл. 12.6. К сожалению, и ротатабельные планы нельзя считать оптимальными композиционными планами.

В ряде случаев при испытаниях ДЛА желательно наиболее полное заполнение исследуемой области экспериментами. При этом отказываются от требований ротатабельности и ортогональности и принимают $\alpha=1$ (план Хартли).

Главным достоинством композиционных планов является возможность их постепенного усложнения по мере изучения задачи. Действительно, реализация плана, представленного на рис. 12.3, б, осуществляется в три этапа. На первом этапе проводятся эксперименты в соответствии с одной из полуреплик факторного планирования 2^{3-1} (на рисунке эти точки показаны черными кружочками) и один эксперимент в центральной точке для оценки адекватности модели действительности. Если после проверки (например по критерию Фишера, см. разд. 12.1) модель оказывается недостаточно точной, проводят эксперименты, соответствующие второй полуреплике, т. е. *достраивают* план до ПФЭ 2^3 . После повторной проверки точности, если это окажется необходимым, проводят эксперименты в шести «звездных» точках и дополнительные эксперименты — в центральной точке. Таким образом, переход к более слож-

Т а б л и ц а 12.6

Планирование		Количество факторов k			
Тип	Характеристика	2	3	4	5
ПФЭ 3^k	Количество эксперимен- тальных точек N	9	27	81	243
Ортогональ- ное компози- ционное пла- нирование	Ядро планирования	2^2	2^3	2^4	2^{5-1} $I = x_1x_2x_3x_4x_5$
	Число вершин куба	4	8	16	16
	Число «звездных» точек	4	6	8	10
	Число точек в центре экс- перимента	1	1	1	1
	Общее число эксперимен- тов	9	15	25	27
	Плечо «звездных» точек α	1,000	1,215	1,414	1,547
Ротатабель- ное компози- ционное пла- нирование	Ядро планирования	2^2	2^3	2^4	2^{5-1} $I = x_1x_2x_3x_4x_5$
	Число вершин куба	4	8	16	16
	Число «звездных» точек	4	6	8	10
	Число точек в центре экс- перимента	5	6	7	6
	Общее число эксперимен- тов	13	20	31	32
	Плечо «звездных» точек α	1,414	1,682	2,000	2,000

ному плану каждый раз предусматривает использование уже про-
веденных экспериментов.

Легко видеть, что с увеличением числа переменных, а также тре-
бований к точности модели сложность плана быстро растет. На
практике реализация слишком сложных планов затруднительна.
Поэтому большое значение приобретает использование априорных
сведений о модели, позволяющих иногда отказаться от полиноми-
ального представления функции отклика. В этих случаях модель

может быть представлена в виде аналитического выражения, а задача исследования сводится к определению параметров этого выражения.

Вопросы для самоконтроля

1. Чем вызвана необходимость перехода от ПФЭ 2^k к более сложным планам?
2. В чем суть ПФЭ 3^k ?
3. Почему целесообразно переходить к композиционным планам?
4. Нарисуйте схему композиционного плана для трех переменных.
5. В чем заключается принцип последовательного композиционного планирования?

12.4. ПРИМЕРЫ МОДЕЛЕЙ ИДЕНТИФИКАЦИИ, ПОСТРОЕННЫХ ПО РЕЗУЛЬТАТАМ ПЛАНИРУЕМОГО ЭКСПЕРИМЕНТА

Модель идентификации для системы форинжекции ионов.

Рассмотрим щелевую ионно-оптическую систему (ИОС) для формирования и фокусировки ионного пучка. Важнейшими параметрами, характеризующими конструкцию ИОС (рис. 12.4), являются расстояние между антикатодом газоразрядного источника ионов и ускоряющим электродом l_1 , толщина ускоряющего электрода l_2 и разность потенциалов между антикатодом и ускоряющим электродом U_B . При исследовании ИОС ставилась задача формирования жестко сфокусированного пучка, в связи с чем в качестве характеристической функции было выбрано отношение максимальной ширины пучка в пределах d_n щели ускоряющего электрода к ширине этой щели $d_{щ}$. При этом плотность ионного тока и энергия ионов на выходе считались заданными и равными соответственно 75 А/м^2 и 1400 эВ .

По конструктивным и технологическим соображениям диапазоны изменения управляющих переменных были выбраны следующим образом:

$$0,003 \leq l_1 \leq 0,0042 \text{ м}; \quad 0,002 \leq l_2 \leq 0,004 \text{ м}; \quad 3500 \leq U_B \leq 5000 \text{ В}.$$

(12.44)

Исследование ионно-оптической системы осуществлялось методами аналогового моделирования, а для сокращения объема модельных экспериментов были использованы методы планирования.

Последовательно рассматривались регрессионные модели, полученные на основе полуреплики полного факторного эксперимента 2^{3-1} и полного факторного эксперимента 2^3 .

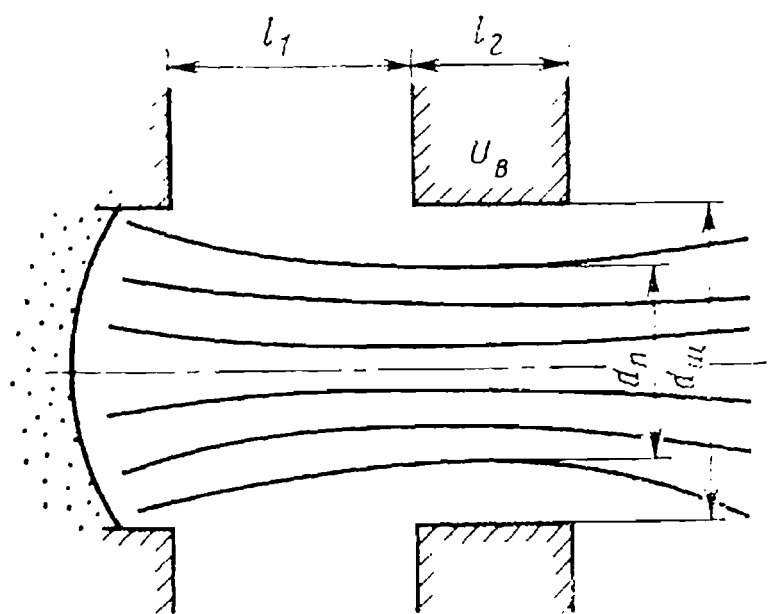


Рис. 12.4. Схема ионно-оптической системы форинжекции ионов

В первом случае оценивались коэффициенты линейной модели

$$W = b_0 + b_1 \tilde{l}_1 + b_2 \tilde{l}_2 + b_3 \tilde{U}_B, \tag{12.45}$$

во втором — коэффициенты неполного квадратного полинома

$$W = b_0 + b_1 \tilde{l}_1 + b_2 \tilde{l}_2 + b_3 \tilde{U}_B + b_{12} \tilde{l}_1 \tilde{l}_2 + b_{13} \tilde{l}_1 \tilde{U}_B + b_{23} \tilde{l}_2 \tilde{U}_B + b_{123} \tilde{l}_1 \tilde{l}_2 \tilde{U}_B. \tag{12.46}$$

Полученные результаты представлены в табл. 12.7 и 12.8.

Контрольные эксперименты, проведенные в центре плана ($\tilde{l}_1 = \tilde{l}_2 = \tilde{U}_B = 0$), показали, что точность регрессионной модели (12.45), полученной на основе полуреплики, недостаточна. Точность полной факторной модели выше (удовлетворяет критерию χ^2 при 5%-ном уровне риска).

Модель идентификации для процесса теплопередачи от твердого тела к жидкости. Методы использования планового эксперимента при построении модели идентификации в заданной форме проще всего показать на конкретном, хотя и несколько искусственном примере определения коэффициента теплоотдачи к жидкости, те-

Т а б л и ц а 12.7

№ эксперимента	$\tilde{x}_0$	$\tilde{l}_1$	$\tilde{l}_2$	$\tilde{U}_B$	Код	W
1	+	—	—	—	(1)	0,9
2	+	+	+	—	<i>ab</i>	1,2
3	+	+	—	+	<i>ac</i>	0,64
4	+	—	+	+	<i>bc</i>	0,6
b_i	0,835	0,085	0,065	—0,215	Для W	

Т а б л и ц а 12.8

№ эксперимента	$\tilde{x}_0$	$\tilde{l}_1$	$\tilde{l}_2$	$\tilde{U}_B$	$\tilde{l}_1 \tilde{l}_2$	$\tilde{l}_1 \tilde{U}_B$	$\tilde{l}_2 \tilde{U}_B$	$\tilde{l}_1 \tilde{l}_2 \tilde{U}_B$	Код	W
1	+	—	—	—	+	+	+	—	(1)	0,9
2	+	+	—	—	—	—	+	+	<i>a</i>	1
3	+	—	+	—	—	+	—	+	<i>b</i>	0,07
4	+	+	+	—	+	—	—	—	<i>ab</i>	1,2
5	+	—	—	—	+	—	—	+	<i>c</i>	0,46
6	+	+	—	+	—	+	—	—	<i>ac</i>	0,64
7	+	—	+	+	—	—	+	—	<i>bc</i>	0,6
8	+	+	+	+	+	+	+	+	<i>abc</i>	0,96
b_i	0,73	0,221	—0,021	—0,064	0,151	—0,086	0,136	—0,106	Для W	

кущей в круглой трубе. В случае турбулентного течения коэффициент теплоотдачи α зависит от семи факторов: скорости жидкости w , кинетической вязкости ν , удельной теплоемкости c , коэффициента теплопроводности λ , плотности ρ , диаметра D и длины трубы L . Таким образом, модель задается зависимостью вида

$$\alpha = f(w, \nu, c, \lambda, \rho, D, L). \quad (12.47)$$

Если искать зависимость (12.47) по схеме ПФЭ, необходимо провести в каждой серии $n = 2^7 = 128$ экспериментов, причем полученное приближение может дать хорошее описание функции в узком диапазоне изменения параметров. Первым шагом по сокращению объема эксперимента без ущерба точности может стать умелый выбор факторов как комплексов определяющих параметров. Так, использование правил теории подобия позволяет составить модель в виде следующей критериальной зависимости:

$$Nu = f(Pr, Re, L/D), \quad (12.48)$$

где $Nu = \frac{\alpha D}{\lambda}$ — критерий Нуссельта; $Pr = \frac{\nu}{\alpha}$ — критерий Прандтля; $Re = wD/\nu$ — критерий Рейнольдса.

Зависимость (12.48) можно исследовать по схеме ПФЭ вида 2^3 , причем необходимый объем эксперимента сокращается до восьми опытов в одной серии. Однако зависимость в виде неполного квадратного многочлена пригодна лишь для оценок в окрестностях базовой точки. Из теории теплообмена известно, что значительно более точные результаты дает критериальная зависимость вида

$$Nu = A Pr^a Re^b. \quad (12.49)$$

Обработка большого массива, полученного в пассивном эксперименте в диапазоне параметров $0,6 \leq Pr \leq 3000$ и $7 \cdot 10^3 \leq Re \leq 2 \cdot 10^6$ при $\frac{L}{D} \geq 50$, привела к зависимости

$$Nu = 0,023 Pr^{0,4} Re^{0,8}. \quad (12.50)$$

Интересно получить оценку констант A , a , b методами планируемого эксперимента и сравнить результат с зависимостью (12.50), которую можно считать точной, а также провести проверку гипотезы об автомодельности процесса относительно критерия L/D , для чего рассмотреть модель вида

$$Nu = A Pr^a Re^b \left(\frac{L}{D} \right)^c. \quad (12.51)$$

Логарифмирование (12.51) позволяет перейти к линейной модели

$$\lg Nu = \lg A + a \lg Pr + b \lg Re + c \lg \frac{L}{D}. \quad (12.52)$$

В качестве факторов выступают теперь следующие переменные:

$$x_0 = 1; x_1 = \lg Pr; x_2 = \lg Re; x_3 = \lg \frac{L}{D}. \quad (12.53)$$

удобно также обозначать

$W = \lg Nu; \lg A = b_0; a = b_1; b = b_2; c = b_3.$ (12.54)

Для определения коэффициентов b_i используется ПФЭ вида 2^3 .

Матрица планирования эксперимента, уровни натуральных параметров, а также результаты расчета коэффициентов регрессии представлены в табл. 12.9. Уравнение функции отклика имеет вид

$W = 3,39 + 0,19\tilde{x}_1 + 0,805\tilde{x}_2 + 0,01\tilde{x}_3 + 0,004\tilde{x}_1\tilde{x}_2 + 0,001\tilde{x}_1\tilde{x}_3 +$
 $+ 0,006\tilde{x}_2\tilde{x}_3 + 0,02\tilde{x}_1\tilde{x}_2\tilde{x}_3.$ (12.55)

Видно, что коэффициенты $b_{12}, b_{13}, b_{23}, b_{123}$, определяющие взаимодействия, почти на два порядка меньше коэффициентов при линейных членах, а коэффициент b_3 много меньше коэффициентов b_1 и b_2 .

Т а б л и ц а 12.9

№ экс- пери- мента	$\tilde{x}_0$	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_1\tilde{x}_2$	$\tilde{x}_1\tilde{x}_3$	$\tilde{x}_2\tilde{x}_3$	$\tilde{x}_1\tilde{x}_2\tilde{x}_3$	Код	W
1	+	—	—	—	+	+	+	—	(1)	2,42
2	+	+	—	—	—	—	+	+	a	2,74
3	+	—	+	—	—	+	—	+	b	3,96
4	+	+	+	—	+	—	—	—	ab	4,4
5	+	—	—	+	+	—	—	+	c	2,38
6	+	+	—	+	—	+	—	—	ac	2,77
7	+	—	+	+	+	—	+	—	bc	4,01
8	+	+	+	+	+	+	+	+	abc	4,39
	3,39	0,19	0,805	0,01	0,004	0,001	0,006	—0,02	W	
	x_1 3 2	x_2 6 —4	x_3 2,7 —1,7		$\hat{W} = 2,38\tilde{x}_0 + 2,76 \tilde{x}_1 + 3,99\tilde{x}_2 + 4,37\tilde{x}_3$					
+										
—										

Оценка значимости членов модели с помощью критерия Фишера показывает, что можно ограничиться рассмотрением модели

$W = -1,6 + 0,38\tilde{x}_1 + 0,805\tilde{x}_2.$ (12.56)

Используя критериальные переменные (12.51), окончательно можно получить

$Nu = 0,025Pr^{0,38}Re^{0,805},$ (12.57)

что хорошо совпадает с «точной» зависимостью (12.50).

Приведенный пример наглядно демонстрирует возможность сокращения объема эксперимента при максимальном использовании априорных сведений об объекте. Достигнуто сокращение эксперимента в 16 раз!

МЕТОДЫ ОПТИМАЛЬНОГО ПРОЕКТИРОВАНИЯ ДВИГАТЕЛЕЙ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

Центральной проблемой проектирования ДУ ЛА, в частности, при создании и использовании систем автоматизированного проектирования, является выбор оптимальных параметров конструкции и режимов работы ДУ. Задача оптимизации вообще характерна для деятельности как отдельного человека, так и любых человеческих коллективов. Выбор специальности и способа провести свободный вечер, схемы разрабатываемой машины и сроков посевной кампании, структуры капитальных вложений в народное хозяйство и внешнеполитических акций государства всегда предполагает предварительный перебор возможных вариантов. При выборе оптимальных технических решений принято говорить об *оптимальном проектировании* технических устройств. Задача оптимального проектирования несомненно важна для всех областей техники. Лишь в очень редких случаях необходимый для оптимизации перебор вариантов может быть выполнен экспериментально. Поэтому решение задач оптимального проектирования, как правило, основано на изучении математических моделей ДУ и их элементов, а также других подсистем ЛА.

При постановке задачи оптимального проектирования ДУ ЛА прежде всего необходимо определить условия и цели оптимизации, т. е. выбрать переменные, изменяемые при оптимизации, диапазоны возможного изменения этих переменных, ограничения, которым должны удовлетворять оптимальные решения, и, наконец, целевую функцию (или функции) для оценки совершенства ДУ. Если для оптимизации используется математическая модель ДУ, то нужно выбрать и обосновать форму и состав этой модели и оценить точность результатов моделирования.

ГЛАВА 13. ПОСТАНОВКА ЗАДАЧ ОПТИМАЛЬНОГО ПРОЕКТИРОВАНИЯ

13.1. ФОРМУЛИРОВКА ЗАДАЧИ ОПТИМАЛЬНОГО ПРОЕКТИРОВАНИЯ

Математически задача оптимизации системы (1.1) в наиболее общем виде формулируется как задача определения вектора $\hat{x}$, такого, что

$$W(\hat{x}, \omega) \leq W(x, \omega); \quad \forall x, \hat{x} \in X^* \subset X; \quad \forall \omega \in \Omega, \quad (13.1)$$

где $x = (x_1, x_2 \dots x_k)$ — вектор *варьируемых переменных*; $\omega = (\omega_1,$

рис. 13.1. Схема задачи оптимизации

$\omega_2, \dots, \omega_m$) — вектор *неопределенностей* (шумов); W — оценочная функция системы, называемая *критерием оптимальности* (а также функцией качества, или функцией цели).

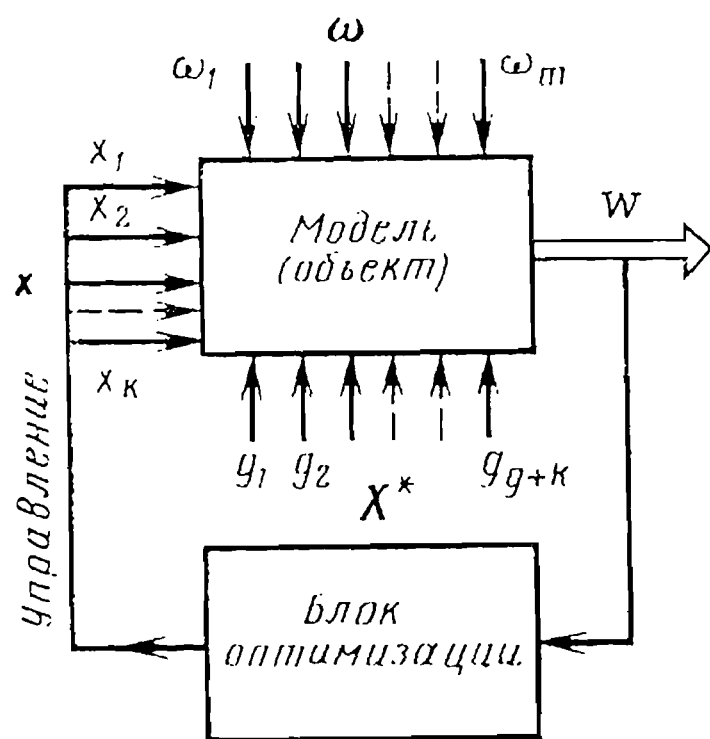
Сформулированная задача (13.1) представляет собой задачу *минимизации*, т. е. наилучшему решению соответствует минимальное значение функции цели (например, наилучшим является решение, обеспечивающее минимум массы или стоимости ДУ). На практике могут встретиться задачи, требующие *максимизации* тех или иных характеристик (например полезной нагрузки или дальности полета ЛА). Переход к случаю максимизации¹ осуществляется заменой функции качества W обратной ей функцией W^{-1} или $-W$.

Графическая схема задачи оптимизации представлена на рис. 13.1 и отличается от схемы рис. 1.1 наличием блока оптимизации, получающего информацию о величине функции цели и изменяющего вектор варьируемых переменных x .

Варьируемые переменные представляют собой набор параметров, изменяя которые оптимизатор может добиваться улучшения характеристик ДУ. Варьируемые параметры могут характеризовать или схему ДУ (количество параллельно работающих двигателей или двигательных блоков, баков топлива, источников или преобразователей энергии), или конструкцию двигателя и его элементов (количество ступеней компрессора, камер сгорания, степень двухконтурности ГТД; количество форсунок, степень расширения сопла, схема охлаждения ЖРД; размеры и количество газоразрядных камер, количество и конфигурация электродов ускоряюще-формирующей ионно-оптической системы, конфигурация и тип магнитной системы в ЭРД), или режимы работы двигателя (степень сжатия, расход рабочего тела для основной и форсажной камер сгорания, частота вращения ротора, температура газа перед турбиной в ГТД, давление и температура в камере сгорания, расход рабочего тела в ЖРД, напряжение разряда и напряженность магнитного поля, расход рабочего тела, ускоряющее напряжение в ЭРД), или, наконец, используемые материалы и рабочие тела.

Общее число переменных определяет *размерность* задачи оптимизации (размерность факторного пространства).

Варьируемые переменные не должны зависеть друг от друга, по крайней мере, в некоторой области факторного пространства (X^*). Наличие условий или связей типа равенств (1.3) фактически приводит к снижению размерности задачи оптимального проек-



¹ Далее, если это не оговорено особо, всегда рассматриваются задачи минимизации.

тирования, так как одна из переменных может быть исключена из рассмотрения или выражена через другие переменные.

Не все перечисленные ранее переменные отвечают этому требованию (например, в ЖРД выбор конфигурации камеры сгорания, рода рабочего тела и его расхода однозначно определяет температуру и давление продуктов сгорания). На практике необходимо тщательно проверять независимость варьируемых переменных. Совокупность варьируемых переменных должна быть *полной*, т. е. обеспечивать полное определение характеристик ДУ в рамках принятой модели.

Варьируемые переменные должны быть *значимы*, т. е. должны действительно влиять на характеристики ДУ. При недостаточной значимости переменных поверхность отклика слабоконтрастна (см. разд. 1.3). Как правило, желательно ранжировать переменные по их значимости для системы в целом.

Изменение варьируемых переменных может осуществляться как *непрерывно* (расход топлива, температура газа, частота вращения ротора ГТД, напряжение разряда, ускоряющие потенциалы), так и *дискретно* (количество двигательных блоков, входящих в ДУ, количество ступеней компрессора, количество электродов системы ускорения и формирования пучка ионов).

К сожалению, при оптимизации ДУ, как правило, приходится иметь дело с наиболее трудными для решения задачами, включающими в себя как непрерывно, так и дискретно меняющиеся переменные.

Область существования решения X^* задается ограничениями (1.3), (1.4). При этом, как уже отмечалось, ограничения типа равенств приводят к снижению размерности факторного пространства и фактически автоматически входят в математическую модель. Ограничения типа неравенств (1.4) или должны систематически проверяться в процессе оптимизации, или с помощью тех или иных искусственных приемов также должны быть включены в модель.

Наиболее эффективный метод такого включения связан с введением так называемых *функций штрафа*.

Метод функций штрафа предусматривает переход от реальной целевой функции W к модифицированной целевой функции W_m с одновременным переходом от области существования X^* к полному факторному пространству X , т. е. переход к задаче оптимизации *без ограничений*. Могут быть предложены разные формы функции W_m . Наибольшее распространение получили функции вида

$$W_m = W + \sum_j K_j g_j^2 h(g_j); \quad (13.2)$$

$$W_m = W + \sum_j K_j |g_j| h(g_j), \quad (13.3)$$

где g_j — ограничения (1.4), записанные в виде

$$g_i = (x_i - x_{i \max})(x_i - x_{i \min}) \leq 0; \quad g_j = f_j(x) - C_j \leq 0, \quad (13.4)$$

K_j — константы штрафа; $h(g_j)$ — пороговая функция (функция Хевисайда):

$$h(g_j) = \begin{cases} 0, & g_j \leq 0; \\ 1, & g_j > 0. \end{cases} \quad (13.5)$$

Легко убедиться, что запись (13.4) эквивалентна записи (1.4), так как при любом выходе переменной за ограничения (например $x_i > x_{i \max}$) соответствующие функции g_j становятся положительными.

В модифицированную функцию качества можно включить также и ограничения типа равенств. В этом случае выражения (13.2), (13.3) могут быть переписаны в виде

$$W_m = W + \sum_i K_i g_i^2 + \sum_j K_j g_j^2 h(g_j); \quad (13.2a)$$

$$W_m = W + \sum_i K_i |g_i| + \sum_j K_j |g_j| h(g_j). \quad (13.3a)$$

В области X^* , когда ограничения не нарушены, $W_m = W$, так как второй член правой части (13.2) равен нулю в силу равенства нулю $h(g_j)$ при $g_j \leq 0$. При нарушении же тех или иных ограничений во втором члене правой части (13.2) появляются положительные компоненты и $W_m > W$ (при максимизации знаки дополнительных членов в (13.2), (13.3) должны быть заменены на обратные и $W_m < W$). Таким образом, нарушение ограничений автоматически делает получаемое решение *невыгодным* и заставляет блок оптимизации пересматривать стратегию поиска, возвращая его в область X^* . Механизм действия функции штрафа в одномерном случае показан на рис. 13.2, а.

Величина «штрафа» и, следовательно, ухудшение целевой функции зависит от степени нарушения ограничения g_j^2 и величины константы штрафа K_j , характеризующей важность данного ограничения. При использовании функций штрафа допускается некоторый

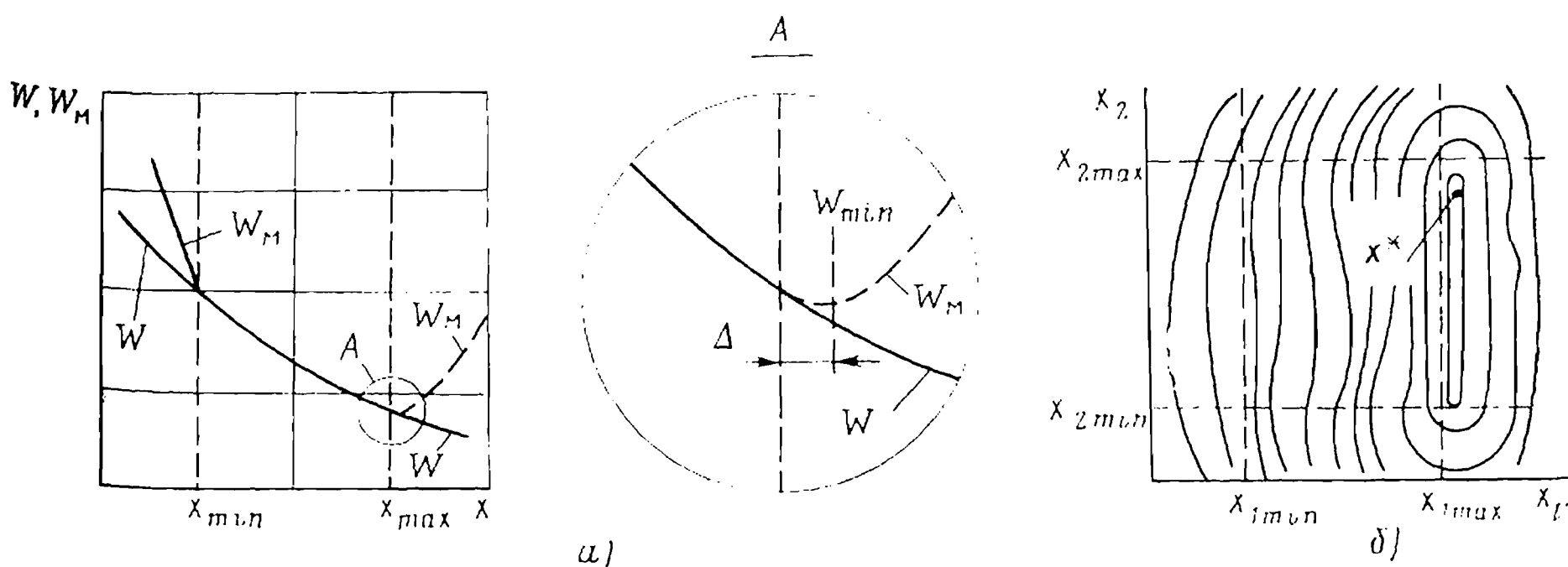


Рис. 13.2. Схема учета ограничений при использовании функции штрафа:
 а — исходный и модифицированный критерии оптимальности; б — возникновение «оврага» вдоль границы области при использовании функции штрафа

(правда, достаточно малый) выход за пределы разрешенной области X^* (рис. 13.2, а), что, вообще говоря, соответствует реальной ситуации, возникающей при решении большинства технических задач. Более существенно, что использование штрафов, особенно в многомерных задачах, может приводить к ухудшению *обусловленности* функции отклика в окрестности границы, т. е. к возникновению «оврагов», расположенных вдоль границы (рис. 13.2, б), затрудняющих поиск оптимума. Увеличение констант штрафа уменьшает возможный выход поиска за пределы X^* , но, с другой стороны, ухудшает обусловленность функции в окрестности границы. На практике величины K_j , K_i подбираются экспериментально и в ряде случаев уточняются в процессе поиска.

Вектор неопределенностей ω характеризует факторы, не учтенные в модели. При оптимизации ДЛА к таким факторам относятся упрощения, принятые при разработке модели, а также отклонения характеристик материалов, комплектующих узлов и внешних условий функционирования ДЛА от принятых значений. Наличие неопределенностей, естественно, приводит к огрублению результатов оптимизации.

Вопросы для самоконтроля

1. Дайте математическую формулировку задачи оптимального проектирования.
2. Определите понятия «варьируемая переменная», «шум», «критерий оптимальности».
3. Что такое функция штрафа?

13.2. КРИТЕРИИ ОПТИМАЛЬНОСТИ

Выбор критерия оптимальности W обычно или очевиден, или чаще всего весьма сложен. Дело в том, что довольно трудно формулировать цели и задачи функционирования действительно сложных систем. Кроме того, при проектировании часто приходится иметь дело с рядом противоречивых требований, предъявляемых к системе. Так, от современного самолета одновременно требуется высокая скороподъемность, большая скорость, дальность и предельная высота полета и т. д. Все это затрудняет выбор эффективных критериев оптимальности сложных систем.

Общим принципом построения критериев оптимальности является оценка эффективности исследуемой системы с точки зрения ее *полезности для системы старшего уровня* (надсистемы). Поэтому в качестве критерия оптимальности не должны использоваться внутренние характеристики системы. Так, например, при оценке ДУ как подсистемы ЛА нельзя использовать коэффициент полезного действия, удельную тягу или экономичность двигателя. Эти показатели очень важны, но характеризуют отдельные процессы в двигателе, а не ДУ в целом.

Вызывает возражения и использование в качестве критерия оптимальности ДУ характеристик ее надежности, в частности, веро-

ятности безотказной работы, которая всегда может быть повышена за счет снижения удельных характеристик, использования более стойких (а значит, и более дорогих) материалов и, наконец, за счет резервирования. При этом, однако, резко возрастает масса и стоимость ДУ. С другой стороны, вероятность безотказной работы ЛА в целом определяется вероятностью безотказной работы ряда подсистем и одностороннее ее увеличение для ДУ может привести лишь к незначительному улучшению характеристик аппарата в целом. Поэтому вероятность безотказной работы правильнее рассматривать как ограничение, заданное на определенном уровне.

Для комплексной оценки сложной системы могут быть использованы критерии двух типов: *стоимостные* и *технические* общесистемные критерии оптимальности.

Стоимостные критерии так или иначе отражают стоимость создания и функционирования системы. Иногда в качестве критерия оптимальности используется непосредственно величина стоимости конкретной полетной программы, выполняемой с помощью ДУ. Для оценки входящих в состав ДУ узлов или режимов их работы можно использовать величину перенесенной стоимости ДУ. Эта стоимость C_Σ складывается из стоимости научно-исследовательских и опытно-конструкторских работ $C_{н.о}$, стоимости изготовления $C_{из}$, стоимости хранения и эксплуатации $C_{э}$ и т. д.:

$$C_\Sigma = C_{н.о} + C_{из} + C_{э} + \dots \quad (13.6)$$

Каждая из компонент (13.6), в свою очередь, может быть выражена как функция технических параметров ДЛА. Широкое распространение получили линейные зависимости, с помощью которых компоненты (13.6) выражаются через массу M_i , мощность W_i , время работы τ_i и другие характеристики ДУ или ее узлов:

$$C_i = k_{M_i} M_i + k_{W_i} W_i + k_{\tau_i} \tau_i + \dots \quad (13.7)$$

Коэффициенты k_{ji} представляют собой удельные затраты и определяются на основе статистических данных. Использование функций (13.7) приводит к *линейным* моделям стоимости (см. разд. 5.1).

Часто для описания зависимости стоимости от технических характеристик системы используются функции Кобба — Дугласа

$$C = k \prod_{j=1}^m x_j^{\alpha_j}, \quad (13.8)$$

где k — константа; x_j — j -ная составляющая вектора технических характеристик ($j = 1, 2, \dots, m$); α_j — константа, характеризующая затраты на достижение некоторого значения характеристики x_j . Константы k, α_j определяются на основании статистических данных.

Использование функций (13.8) приводит к *логарифмическим* моделям стоимости. В отдельных случаях для описания стоимостных характеристик используются экспоненциальные зависимости.

При всей кажущейся очевидности использование критерия стоимости вызывает серьезные возражения, так как не учитывается

в явном виде *эффективность* исследуемой системы. Поэтому любое решение, даже обеспечивающее существенный рост эффективности ценой незначительного увеличения затрат, будет отвергнуто как невыгодное. В связи с этим критерий стоимости может быть использован лишь для сравнения различных вариантов решения одной и той же строго заданной задачи (например выбор типа самолета для доставки заданного груза на заданное расстояние).

Более гибкими оказываются критерии, отражающие *стоимость полезного эффекта*:

$$W_{\text{эф}} = \frac{C_{\Sigma}}{\text{полезный эффект}}, \quad (13.9)$$

например, стоимость тонно- или пассажирокилометра для гражданской авиации, стоимость единицы информации для систем наблюдения, стоимость маневра и т. д. Однако в ряде случаев трудно дать *количественную* оценку полезного эффекта (например, трудно оценить в рублях фактор времени при доставке грузов, ценность метеорологических наблюдений и т. д.).

Определение ряда компонент C_i представляет собой серьезные трудности. Корректный расчет стоимости требует учета таких «отдаленных» затрат, как затраты на развитие городов и подготовку специалистов, затраты, связанные с отвлечением сил и средств из других областей народного хозяйства, а также возможностей использования полученных результатов в других областях техники. Полная оценка таких затрат возможна только в рамках изучения экономики в целом. Кроме того, в ряде случаев стоимостные критерии мало контрастны, т. е. слабо изменяются при изменении конкретных технических параметров.

При оптимизации ДУ, ДЛА и их основных узлов широкое распространение получили так называемые *массовые* критерии оптимальности, требующие минимизации массы ДУ при выполнении комплекса заданных условий (заданная тяга, суммарный импульс, количество включений, вероятность безотказной работы в течение заданного ресурса и т. д.). Использование массовых критериев основано на том, что при выполнении заданных требований наилучшей с точки зрения ЛА является самая легкая ДУ. Вместе с тем нужно иметь в виду, что использование массовых критериев может привести к росту стоимости ДУ и полетных программ, например, за счет перехода к более дорогим материалам.

В качестве массового критерия может использоваться суммарная масса ДУ (M_{Σ}), включающая в себя собственно массу двигателя $M_{\text{д}}$, массу топлива $M_{\text{т}}$, массу системы подачи и хранения топлива $M_{\text{с.п.х}}$, массу источников $M_{\text{и}}$ и преобразователей $M_{\text{п}}$ энергии, массу системы управления $M_{\text{с.у}}$ и т. д.:

$$M_{\Sigma} = M_{\text{д}} + M_{\text{т}} + M_{\text{и}} + M_{\text{п}} + M_{\text{с.у}} + \dots \quad (13.10)$$

Каждая из составляющих (13.10) может быть связана с характеристиками конструкции и рабочего процесса ДУ, образующими вектор варьируемых параметров x . Однако критерию (13.10) свой-

ственными те же недостатки и ограничения, что и абсолютной стоимости.

Более эффективным, особенно для транспортных систем, представляется использование *относительной начальной массы* $\bar{M}_n$, представляющей собой отношение начальной массы ЛА к массе полезной нагрузки:

$$\bar{M}_n = M_0 / M_n. \quad (13.11)$$

Для управляющих систем (например систем коррекции, ориентации и стабилизации космических ЛА) удобно использовать так называемую *массовую цену* $M_{уд}$ суммарного импульса I_Σ :

$$M_{уд} = M_\Sigma / I_\Sigma. \quad (13.12)$$

Использование массовых критериев оптимальности позволяет разработать относительно простые и объективные модели ДУ и ее узлов. Кроме того, массовые критерии оптимальности обладают большей контрастностью по сравнению с более общими, а поэтому и менее чувствительными к изменению частных факторов x_i стоимостными критериями. Могут быть предложены и другие технические критерии оптимальности сложных систем.

Вопросы для самоконтроля

1. В чем заключается требование системности критерия оптимальности?
2. Сравните между собой стоимостные и массовые критерии оптимальности.

13.3. МНОГОКРИТЕРИАЛЬНЫЕ ЗАДАЧИ ОПТИМИЗАЦИИ

Часто к ДУ ЛА предъявляется ряд различных требований, приводящих к введению нескольких критериев оптимальности $W_1, W_2, \dots, W_p$. Например, желательно, чтобы ДУ имела минимальную стоимость и массу, максимальную тягу и живучесть, широкий диапазон регулирования и т. д. Наиболее простым и надежным способом оптимизации таких систем является выделение одной из этих характеристик и использование ее в качестве критерия. При этом остальные требования задаются в виде ограничений. Если же идет речь о действительной оптимизации нескольких функций, необходимо перейти к *векторным* критериям оптимальности $W(W_1, W_2, \dots, W_p)$.

Решение задачи векторной оптимизации тривиально, если существует вектор $\hat{x} \in X^*$, обеспечивающий экстремизацию всех компонент критерия оптимальности одновременно:

$$W_1(\hat{x}, \omega) \leq W_1(x, \omega); W_2(\hat{x}, \omega) \leq W_2(x, \omega); \quad (13.13)$$

$$W_p(\hat{x}, \omega) \leq W_p(x, \omega); \quad \forall \hat{x}, x \in X^*; \quad \forall \omega \in \Omega.$$

Правда, в этом случае оптимизация по W эквивалентна скалярной оптимизации по любой из компонент W_i .

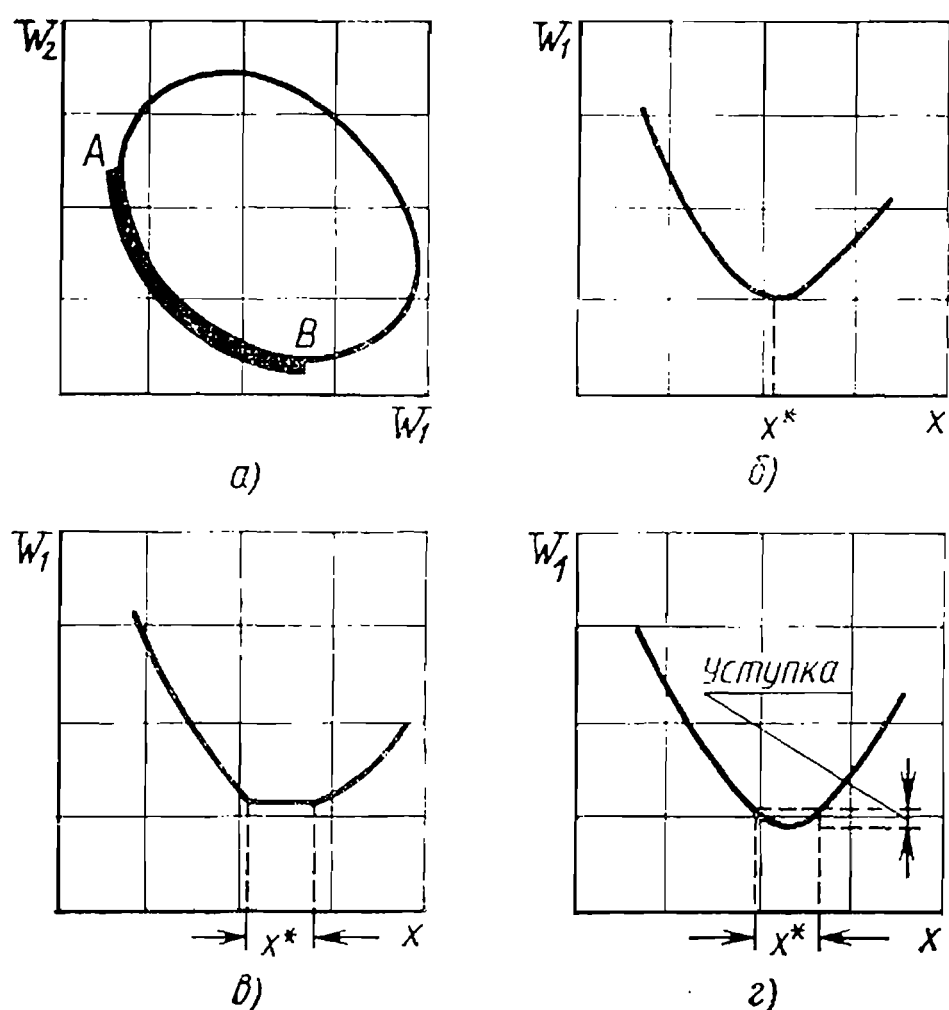


Рис. 13.3. Многокритериальная оптимизация:

α — множество Парето; β — единственная оптимальная по критерию W_1 точка; γ — множество равнозначных по критерию W_1 точек; δ — метод «уступок»

На практике такая ситуация встречается редко. Требования, предъявляемые к системе, обычно противоречивы. Например, минимизация массы потребует использования в ДУ более высококачественных и дорогих материалов и, следовательно, приведет к увеличению ее стоимости, расширение диапазона регулирования потребует увеличения и массы, и стоимости ДУ и т. д. Поэтому оптимальные векторы варьируемых переменных ($\hat{x}_j$) для разных компонент $\mathbf{W}$ не совпадают между собой, и необходимо искать компромиссное решение.

При векторной оптимизации рекомендуется первоначально выделить множество точек, не допускающих дальнейшего *одновременного* улучшения всех компонент $\mathbf{W}$ (множество Парето)¹, и в дальнейшем рассматривать только эти точки. Для случая двухкомпонентного вектора оптимальности $\mathbf{W}(W_1, W_2)$ это множество показано на рис. 13.3, α жирной линией.

Существует несколько подходов к оптимизации ДУ по векторным критериям оптимальности.

1. В ряде случаев задается отношение *строгого предпочтения* целевых функций

$$W_1 \succ W_2 \succ \dots \succ W_p, \quad (13.14)$$

означающее, что любое улучшение компоненты W_1 всегда безусловно важнее изменения всех других компонент критерия. Компонента W_2 играет второстепенную роль по сравнению с W_1 , но, в свою очередь, безусловно важнее W_3 и т. д. В этом случае оптимизация проводится последовательно начиная с важнейших целевых функций.

В результате оптимизации по W_1 может быть получена или оптимальная точка (рис. 13.3, β), или некоторая область *равнозначных* оптимальных решений (рис. 13.3, γ). В первом случае дальнейшая оптимизация невозможна, во втором для выбора одного из равнозначных по критерию W_1 решений может быть использован критерий W_2 и т. д. Нужно отметить, что области равнозначных

¹ По имени итальянского математика-экономиста Парето.

оптимальных решений характерны для задач с дискретно меняющимися переменными.

2. Может быть задан искусственный *комплексный* критерий, т. е. *весовая функция качества*:

$$W_{\Sigma} = a_1 W_1 + a_2 W_2 + \dots + a_p W_p, \quad (13.15)$$

где коэффициенты a_i характеризуют «вес», т. е. значимость соответствующих компонент W_i .

В этом случае существенное изменение даже второстепенного критерия может оказаться более важным по сравнению с малым изменением первоочередных критериев. Для того чтобы суммарное улучшение нескольких критериев не привело к недопустимому ухудшению одного из них, обычно вводятся дополнительные ограничения

$$W_1 \leq A_1; W_2 \leq A_2; \dots W_p \leq A_p. \quad (13.16)$$

Выбор весовых коэффициентов осуществляется с помощью экспертных оценок и уточняется по мере получения результатов.

3. Наконец, если отсутствуют данные как для задания отношения строгого предпочтения (13.14), так и для выбора весовых коэффициентов функции (13.15), может быть использован *метод уступок*. При этом предполагается, что небольшое отступление (например на 5 или 10%) от оптимума, даже по важнейшим критериям, не приводит к существенному ухудшению системы. Такое допущение может быть обосновано общими соображениями о приближенном характере моделирования или измерений, если для оптимизации используются экспериментальные данные.

Введение указанных отклонений, «уступок», позволяет получить *искусственные* множества равнозначных решений (рис. 13.3, г) и использовать методы последовательной оптимизации по ряду критериев. Выбор величин допустимых уступок также требует экспертных оценок.

Из сказанного ясно, что *математически* задача оптимизации всегда сводится к процедуре оптимизации по *скалярному* критерию (та или иная компонента W_i при последовательной оптимизации, комплексный критерий W_{Σ}). Поэтому далее при описании методов оптимизации всегда рассматривается случай скалярного критерия качества. С другой стороны, выбор этого критерия в той или иной форме должен учитывать *множественность* реальных технических требований, предъявляемых к ДУ.

Вопросы для самоконтроля

1. Каковы причины, требующие использования векторных критериев оптимизации?
2. Сформулируйте методы исследования задач многокритериальной оптимизации.
3. Что такое множество Парето?

Особое место среди задач оптимизации занимает задача оптимального проектирования *многоцелевых ДУ*, т. е. ДУ, предназначенных для выполнения ряда различных задач.

Физическая постановка задачи многоцелевой оптимизации очевидна. Например, оптимальные условия выполнения полетной задачи могут быть обеспечены, если для каждой из таких задач будет спроектирована специальная ДУ, учитывающая особенности каждой из них. Однако такой подход совершенно нереален, и на практике разрабатывается несколько типов ДУ для решения определенных классов задач.

Аналогичная ситуация возникает при проектировании основных узлов ДУ, летательных аппаратов и иных технических устройств (электродвигатели, генераторы, радиоаппаратура и т. д.).

Система, предназначенная для решения некоторой группы различных задач, как правило, не может быть оптимальной для каждой из них. Поэтому оптимизация многоцелевых систем всегда связана с некоторым *компромиссом*, с минимизацией проигрыша, неизбежного при использовании многоцелевой системы вместо системы, разработанной специально для данного случая. Причем при выборе параметров многоцелевой системы нельзя ориентироваться на среднюю расчетную задачу из заданного класса, так как такой подход не гарантирует выполнения всех задач класса [17]. Например, если при проектировании самолета для полетов в диапазоне дальностей 5—10 тыс. км исходить из средней дальности 7500 км, то спроектированная машина может вообще оказаться непригодной для беспосадочных полетов на 10 тыс. км. Точно также неверно определять оптимальные параметры самолета исходя из самых дальних перелетов.

При оптимизации многоцелевых объектов вводится непрерывное или дискретное *множество возможных задач* Λ , включающее в себя отдельные задачи $\lambda \in \Lambda$ как элементы, и множество систем (*модулей* или различных типов ДУ) $\{q\}$ ($q=1, 2, \dots, m$), решающих эти задачи, причем каждому модулю q соответствует свой оптимальный вектор варьируемых параметров x_q . Предполагается, что каждая из задач множества Λ решается по крайней мере одним из m модулей (одной из ДУ), а среди нескольких модулей, решающих одну и ту же задачу, всегда можно выделить один, обеспечивающий наилучшее решение. Тогда все множество Λ можно разбить в соответствии с номером оптимального модуля на Λ_i подмножеств таких, что

$$\bigcup_{i=1}^m \Lambda_i = \Lambda; \quad \Lambda_i \cap \Lambda_j = \emptyset; \quad i, j = 1, 2, \dots, m. \quad (13.17)$$

Для оценки принадлежности данной задачи λ к конкретному подмножеству Λ_i вводится *функция распределения* $B(\lambda)$, принимающая значение, равное номеру соответствующего модуля:

$$B(\lambda) = i \quad \text{при } \lambda \in \Lambda_i. \quad (13.18)$$

Возможны различные постановки задачи многоцелевой оптимизации:

1. Может быть задано множество задач Λ и требуется найти оптимальное число модулей m , оптимальные параметры этих модулей $\hat{x}_q$ и распределение задач между модулями $B(\lambda)$:

$$\Lambda \Rightarrow m, \hat{x}_q, B(\lambda). \quad (13.19)$$

2. Может быть задано множество задач Λ и число модулей m , а требуется найти параметры модулей и функцию распределения:

$$\Lambda, m \Rightarrow \hat{x}_q, B(\lambda). \quad (13.20)$$

2. Наконец, могут быть заданы и множество задач Λ , и число модулей m , и их параметры $\hat{x}_q$. В этом случае задачей исследования является определение функции распределения

$$\Lambda, m, \hat{x}_q \Rightarrow B(\lambda). \quad (13.21)$$

Каждая следующая из сформулированных задач является частным случаем предыдущей.

Решение задачи многоцелевой оптимизации в наиболее общем виде (13.19) требует использования критериев оптимальности, в той или иной форме учитывающих стоимость ДУ и ДЛА. Для технических критериев оптимальности (например при минимизации массы ДУ) наилучшим, очевидно, окажется решение, требующее разработки для каждой из задач собственной ДУ. При введении же стоимости в задачу появляется *обратная связь*, требующая уменьшения числа модулей. При заданном числе модулей (13.20) могут быть использованы любые из перечисленных выше критериев.

Меняется форма используемых критериев. В данном случае речь идет о минимизации *проигрыша*, возникающего из-за применения многоцелевой системы вместо системы, истинно оптимальной для данных условий. Поэтому в качестве критерия целесообразно использовать величину этого проигрыша, полученную при сравнении соответствующих значений целевых функций:

$$W_{\Pi}(\lambda) = W(\lambda) - W_{\lambda}; \quad W_{\Pi}(\lambda) = \frac{W(\lambda)}{W_{\lambda}}; \quad (13.22)$$

$$W_{\Pi}(\lambda) = \frac{W(\lambda) - W_{\lambda}}{W_{\lambda}},$$

где $W(\lambda)$ — величина целевой функции для многоцелевой системы; W_{λ} — величина целевой функции, специально спроектированной для задачи λ (истинно оптимальной) системы.

Для каждой из задач $\lambda \in \Lambda$ будет получено собственное значение проигрыша (13.22). Оптимизация системы модулей в целом может быть проведена или с целью минимизации *среднего* проигрыша для всех задач множества или с целью минимизации *наи-*

большого из проигрышей. В первом случае для системы вводится осредненная (интегральная) оценка

$$W = \sum_{i=1}^m \sum_{\lambda \in \Lambda_i} a_{\lambda} W_{\pi}(\lambda), \quad (13.23)$$

где a_{λ} — коэффициенты, характеризующие важность или частоту появления задач.

Во втором случае вводится *гарантированная* оценка

$$W = \max_m \max_{\lambda \in \Lambda_i} \{W_{\pi_i}(\lambda)\} = \max_{\lambda} W_{\pi}(\lambda). \quad (13.24)$$

Функции (13.23), (13.24) и являются в конечном счете критериями, характеризующими оптимальность многоцелевой системы.

Вопросы для самоконтроля

1. Сформулируйте задачу многоцелевой оптимизации.
2. Сформулируйте критерии оптимальности для задач многоцелевой оптимизации.

13.5. КЛАССИФИКАЦИЯ МЕТОДОВ ОПТИМИЗАЦИИ

Каждый вычислительный алгоритм, в частности, алгоритм определения оптимума моделей ДУ и ДЛА должен обладать следующими важнейшими свойствами.

Во-первых, алгоритм должен быть *полным*, т. е. правила, образующие алгоритм, должны быть сформулированы четко и однозначно, охватывать все возможные в ходе исследования ситуации. При использовании алгоритма человек или вычислительная машина должны иметь возможность, получив исходные данные, довести вычисления до конца, не прибегая к каким-либо дополнительным рассуждениям или дополнительной информации. Ряд рассмотренных ниже алгоритмов предусматривает использование «случайных» решений (случайный выбор начального приближения, случайный выбор направления движения в факторном пространстве). Однако случайным оказывается лишь результат той или иной процедуры. Сам же алгоритм, т. е. порядок использования случайности всегда *детерминирован*, т. е. задан однозначно.

Во-вторых, алгоритм должен быть *сходящимся*, т. е. действительно сходится к искомому решению (например к точке, соответствующей минимуму модели). При этом всегда важно знать, может ли быть найдено решение за конечное число итераций. Если решение является приближенным, то каким является интервал неопределенности, соответствующий последней итерации.

В-третьих, для каждого алгоритма должна быть четко определена *область применимости*, т. е. те классы моделей (например линейные модели, унимодальные модели и т. д.), для которых он предназначен и обладает указанными выше свойствами.

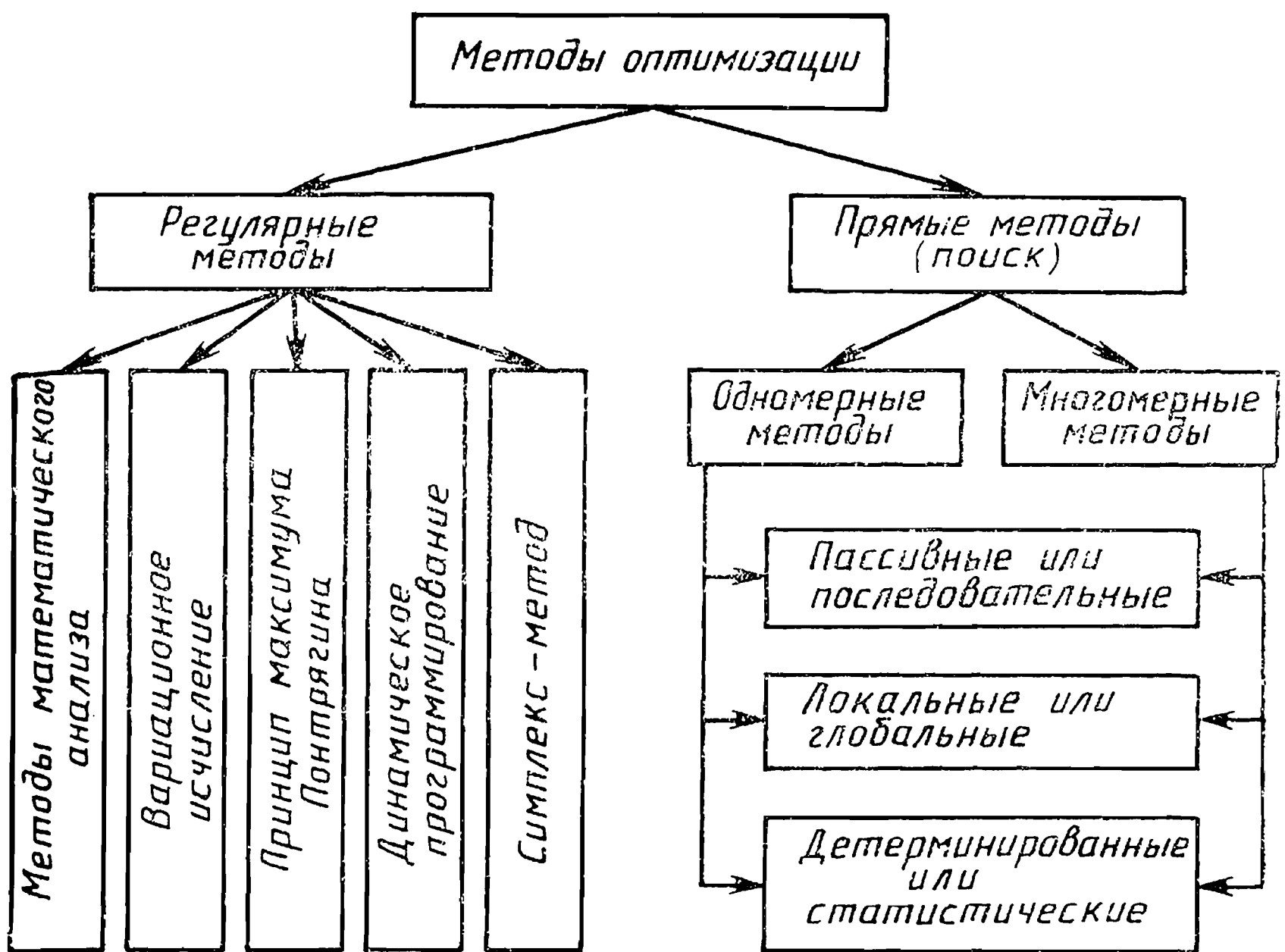


Рис. 13.4. Классификация методов оптимизации

Существуют различные схемы классификации методов поиска оптимальных (экстремальных) решений. Одна из таких схем представлена на рис. 13.4. Как и любая классификация, эта классификация условна и может быть подвергнута обоснованной критике. Однако она позволяет внести определенный порядок в изучение методов оптимизации.

Все методы оптимизации можно разделить на две большие группы: *регулярные* и *прямые*.

Регулярные методы предусматривают решение задачи оптимизации с помощью дифференциального исчисления или рекуррентных соотношений и опираются на точную формулировку необходимых и достаточных условий экстремума. Несмотря на то, что при практическом решении задач и в этом случае широко используются численные методы, всегда подразумевается возможность хотя бы частичного аналитического исследования модели. К регулярным методам оптимизации относятся методы поиска безусловного и условного экстремума с помощью исследования производных, вариационное исчисление, методы, основанные на принципе максимума Понтрягина, динамическое программирование, симплекс-метод.

Краткая характеристика этих методов дана в следующей главе. Общей особенностью регулярных методов является принципиальная возможность точного определения оптимальной точки факторного пространства. Так как регулярные методы опираются на аналитическое исследование задачи, они могут быть использованы

исключительно при исследовании математических моделей. Возможность проведения аналитического исследования делает регулярные методы очень эффективными. Однако именно из-за необходимости аналитического исследования предъявляются очень высокие требования к математической модели. Например, может потребоваться существование аналитических выражений для производных модели по всем переменным. Поэтому регулярные методы могут быть использованы при изучении достаточно узких классов задач.

Прямые методы оптимизации не предполагают проведения каких-либо аналитических исследований, использования аналитических выражений производных, дифференциальных или рекуррентных соотношений. Единственная процедура, используемая при прямых методах, состоит в определении величины критерия оптимальности при заданной комбинации варьируемых переменных. Эта особенность, во-первых, приводит к минимальным требованиям к математической модели, что придает прямым методам наиболее общий характер; во-вторых, методически объединяет математические и экспериментальные методы поиска оптимума, так как определение величины функции качества может осуществляться как математически, так и экспериментально, например, при стендовых испытаниях ДЛА.

Характерной особенностью прямых методов оптимизации является принципиальная невозможность *точного* определения оптимальной точки факторного пространства. Действительно, поскольку с помощью прямых методов оптимизации можно определить лишь величину критерия оптимальности в некоторой точке факторного пространства, любая стратегия поиска, задающая последовательность точек, обеспечивающих улучшение критерия, даже при очень большой длине не может привести к самой лучшей (*истинно оптимальной*) точке. Большинство практических задач оптимального проектирования ДУ и ее узлов решается именно прямыми методами.

Прямые методы, в свою очередь, могут быть разделены на ряд групп. Так, по размерности задачи методы оптимизации принято делить на одномерные и многомерные, по характеру стратегии — на пассивные и последовательные, по глубине использования накопленной информации — на локальные и глобальные, по характеру выбора очередного шага — на детерминированные и статистические.

В настоящее время разработаны сотни различных методов оптимизации и непрерывно появляются все новые и новые методы. Существует ряд руководств [18, 20, 21], посвященных подробному исследованию этих методов и содержащих в ряде случаев конкретные программы оптимизации для современных ЭВМ. Вместе с тем не существует и, видимо, не может существовать метод, обеспечивающий наилучшие результаты для всех встречающихся задач.

В настоящем пособии нет ни возможности, ни необходимости в подробном изучении каждого из методов оптимизации. Далее рас-

смотрены типичные классы методов, даны характерные для этих классов алгоритмы и сделана попытка классификации ситуаций, для которых целесообразно использование методов каждого из классов.

Вопросы для самоконтроля

1. В чем заключается разница между прямыми и регулярными методами оптимизации?
2. Нарисуйте схему классификации методов оптимизации.

ГЛАВА 14. РЕГУЛЯРНЫЕ МЕТОДЫ ОПТИМИЗАЦИИ

14.1 ОПРЕДЕЛЕНИЕ ЭКСТРЕМУМА ДИФФЕРЕНЦИРУЕМОЙ ФУНКЦИИ

Если исследуемая модель описывается непрерывной и дифференцируемой в рассматриваемой области X^* функцией $W(x)$ всех переменных x , для определения оптимума могут быть использованы методы математического анализа. В экстремальной точке x должны выполняться соотношения

$$\frac{\partial W(\hat{x})}{\partial x_i} = 0; \quad i = 1, 2, \dots, k. \quad (14.1)$$

Если якобиан системы (14.1) отличен от нуля, эта система может быть разрешена относительно $\hat{x}$. Однако соотношения (14.1) представляют собой лишь *необходимые* условия экстремума, и стационарная точка $\hat{x}$ может быть как экстремальной точкой, так и точкой перегиба. Если функция $W(x)$ дважды дифференцируема, могут быть сформулированы и *достаточные* условия экстремума. Так, решение системы (14.1) $\hat{x}$ будет точкой минимума, если второй дифференциал функции $W(x)$ в ней положителен:

$$d^2W(\hat{x}) = \sum_{i=1}^k \sum_{j=1}^k \frac{\partial^2 W(\hat{x})}{\partial x_i \partial x_j} dx_i dx_j > 0. \quad (14.2)$$

Это значит, что для всех соседних с $\hat{x}$ точек значения критерия $W(\hat{x})$ больше (хуже!), чем в точке $\hat{x}$. Условие достижения в точке x максимума имеет вид $d^2W(\hat{x}) < 0$.

Если обозначить

$$a_{ij} = \frac{\partial^2 W(\hat{x})}{\partial x_i \partial x_j}; \quad dx_i = \xi_i; \quad dx_j = \xi_j, \quad (14.3)$$

то условие (14.2) сведется к требованию положительной определенности некоторой квадратичной формы:

$$\sum_{i=1}^k \sum_{j=1}^k a_{ij} \xi_i \xi_j > 0. \quad (14.4)$$

Согласно теореме Сильвестра квадратичная форма (14.4) будет положительно определена, если положительно определены все угловые миноры и определитель матрицы коэффициентов (a_{ij}) :

$$\Delta_1 = a_{11} > 0; \quad \Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0; \dots; \\ \Delta_k = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{vmatrix} > 0. \quad (14.5)$$

Соотношения (14.5) позволяют проверить, является ли найденная стационарная точка $\hat{\mathbf{x}}$ минимумом. В точке максимума квадратичная форма (14.4) должна быть отрицательно определена, для чего необходимо чередование знаков угловых миноров, причем первый из них должен быть отрицательным:

$$\Delta_1 = a_{11} < 0; \quad \Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0; \dots \\ \Delta_k > 0, \text{ если } (-1)^k > 0, \\ \Delta_k < 0, \text{ если } (-1)^k < 0. \quad (14.6)$$

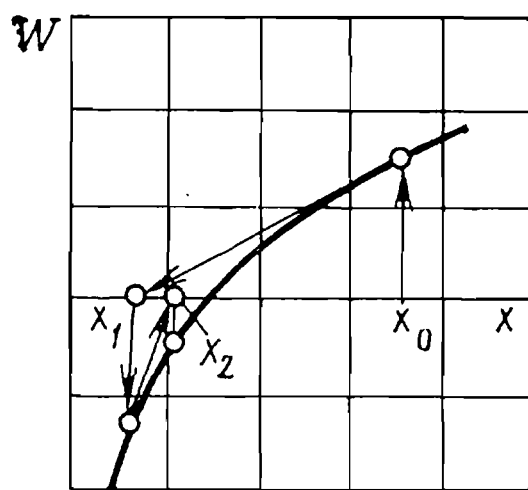
Если функция полимодальна, соотношения (14.1) — (14.6) позволяют выявить все локальные экстремумы, из которых необходимо выбрать оптимальную точку, сообщающую функции $W(\mathbf{x})$ минимальное (максимальное) значение (*глобальный экстремум*). При исследовании функции в ограниченной области эту точку необходимо дополнительно сравнить с граничными точками области, где может достигаться экстремум, не отвечающий условиям (14.1), (14.2).

Аналитическое решение системы уравнений (14.1) возможно лишь в простейших случаях. Для численного решения можно использовать метод Ньютона, позволяющий построить сходящуюся последовательность решений $\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^n$. Геометрическая иллюстрация (для случая одной переменной) и структурная схема программы алгоритма, реализующего метод Ньютона, представлены на рис. 14.1.

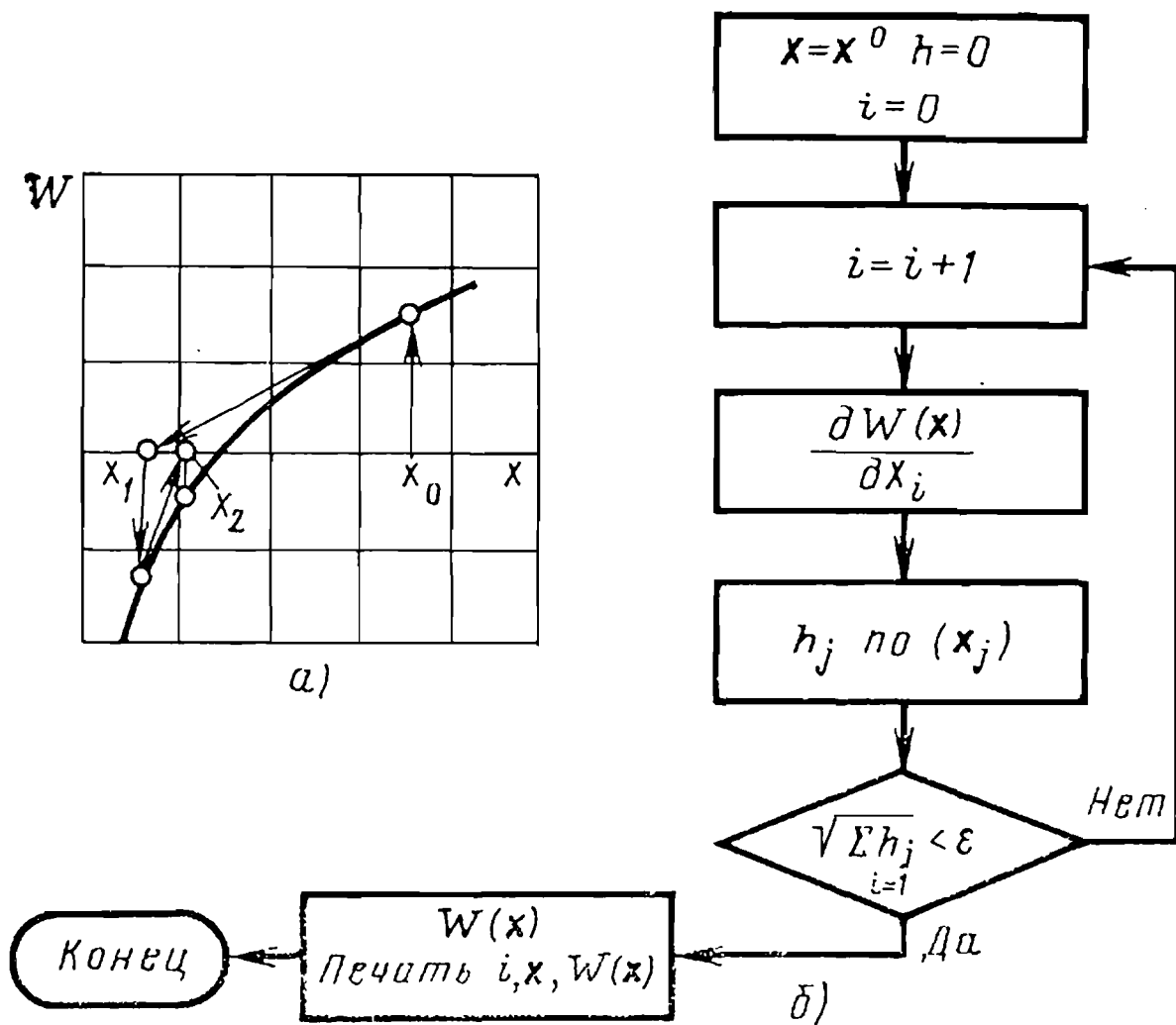
Метод предусматривает произвольное задание начальной точки $\mathbf{x}^0(x_1^0, x_2^0, \dots, x_k^0)$, вычисление производных в этой точке (поскольку решается система (14.1), записанная для производных функции W , речь идет о вторых производных этой функции) и опреде-

Рис. 14.1. Метод Ньютона:

a — схема определения экстремума; *б* — структурная схема программы для реализации метода



a)



б)

ление новой точки $\mathbf{x}^1(x_1^1, x_2^1, \dots, x_k^1)$ с помощью следующей системы уравнений:

$$x_j^1 = x_j^0 + h_j; \quad \sum_{j=1}^k b_{ij} h_j = -b_i; \quad (14.7)$$

$$b_i = \frac{\partial W(\mathbf{x})}{\partial x_i}; \quad b_{ij} = \frac{\partial^2 W(\mathbf{x})}{\partial x_i \partial x_j}; \quad j, i = 1, 2, \dots, k.$$

Процесс итераций $(\mathbf{x}^0 \rightarrow \mathbf{x}^1 \rightarrow \dots \rightarrow \mathbf{x}^n)$ продолжается до тех пор, пока сумма поправок не станет достаточно малой:

$$\sqrt{\sum_{j=1}^k h_j^2} < \varepsilon. \quad (14.8)$$

Величина ε характеризует точность вычислений.

К сожалению, при исследовании реальных моделей сложных систем, использовать необходимые условия экстремума (14.1) обычно не удастся.

Вопросы для самоконтроля

1. Сформулируйте необходимые и достаточные условия экстремума.
2. Сформулируйте критерий Сильвестра.
3. В чем заключается метод Ньютона?

14.2. ВАРИАЦИОННОЕ ИСЧИСЛЕНИЕ. ПРИНЦИП МАКСИМУМА ПОНТЯГИНА

Решение ряда задач, сводящихся к минимизации (максимизации) некоторых специальных функционалов, может быть получено методами, основанными на *вариационном исчислении* и

принципе максимума Понтрягина. Каждый из этих методов представляет собой самостоятельную математическую дисциплину [10, 11].

Многие законы механики и физики сводятся к утверждению, что некий функционал, определенный на заданном классе функций (см. разд. 4.1), должен достигать экстремального значения (принцип наименьшего действия, закон сохранения энергии, принцип возрастания энтропии, принцип Ферма в оптике и т. д.). Применение вариационных принципов при исследовании ДУ приводит к ряду вариационных задач, например, при определении движения заряженных частиц в ускорителе ионов, при определении оптимальной конфигурации сопла или при распределении топлива по ступеням многоступенчатой ракеты. К вариационным сводятся также многочисленные транспортные задачи, связанные с выбором оптимальной траектории ракеты или космического аппарата, оптимальных трасс самолетов и т. д.

Методы классического вариационного исчисления пригодны для оптимизации функционалов, определенных на классе гладких функций, у которых в рассматриваемой области непрерывна первая производная (рис. 14.2, а), или на классе кусочно-гладких функций, у которых первая производная имеет конечное число разрывов первого рода (рис. 14.2, б). Основное соотношение вариационного исчисления — знаменитое уравнение Эйлера, выводится из анализа изменений *вариаций* функционала $I(W(x))$, играющих роль производных функции $W(x)$.

Вывод уравнения Эйлера может быть продемонстрирован на примере простейшего функционала вида ¹

$$I(W(x)) = \int_a^b F(x, W, \dot{W}) dx, \quad (14.9)$$

определенного на классе гладких дважды дифференцируемых функций $W(x)$, проходящих через фиксированные начальную и конечную точки a, b (рис. 14.2, а), причем функция F трижды диффе-

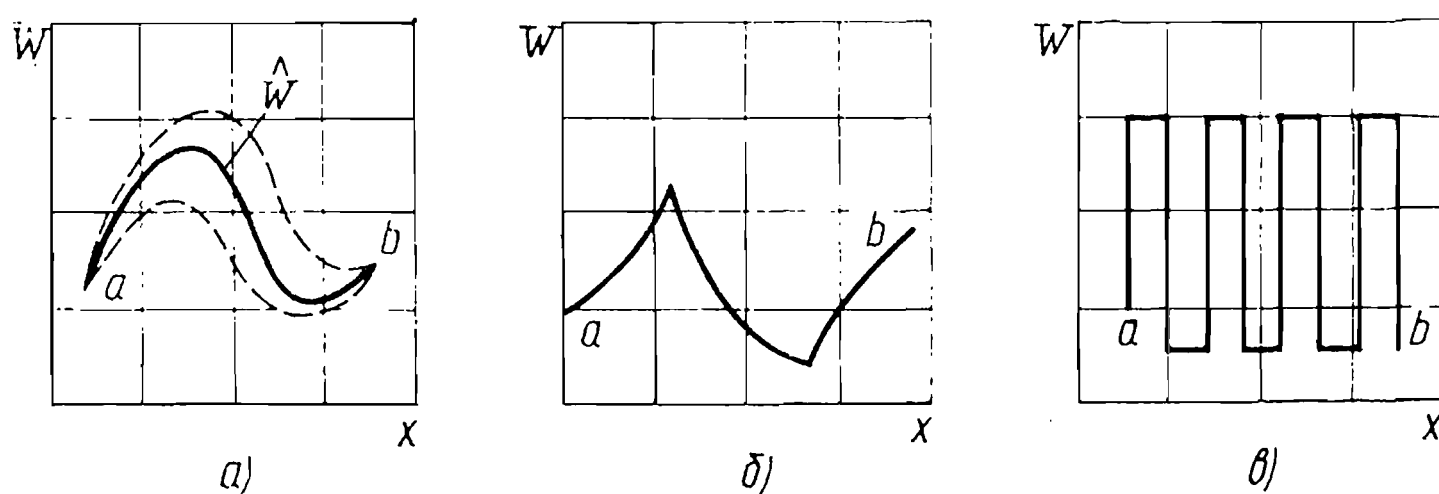


Рис. 14.2. Различные виды функционалов:

а — непрерывные функции с непрерывной производной; **б** — непрерывные функции с разрывной производной; **в** — ступенчатые функции с производными, обращающимися в бесконечность

¹ Вывод уравнений Эйлера дается в соответствии с работой [10].

ренцируема. Предполагается, что оптимальное (минимальное) значение функционала достигается на кривой $\hat{W}(x)$:

$$I_{\min} = I(\hat{W}(x)) = \int_a^b F(x, \hat{W}, \dot{\hat{W}}) dx. \quad (14.9a)$$

Близкие к $\hat{W}(x)$ функции, также проходящие через точки a и b (см. рис. 14.2), можно записать в виде

$$W(x) = \hat{W}(x) + \alpha \eta(x), \quad \eta(a) = \eta(b) = 0, \quad (14.10)$$

где α — произвольное число; $\eta(x)$ — произвольная гладкая функция.

Для каждой из функций $W(x)$ функционал (14.9) имеет вид

$$I = \int_a^b F(x, \hat{W}(x) + \alpha \eta(x), \dot{\hat{W}}(x) + \alpha \dot{\eta}(x)) dx. \quad (14.11)$$

Разность функций $\hat{W}(x)$ и $W(x)$ называется вариацией функции

$$\delta W(x) = W(x) - \hat{W}(x) = \alpha \eta(x), \quad (14.12)$$

а приращение функционала при переходе от $\hat{W}(x)$ к $W(x)$ записывается в виде

$$\Delta I = \int_a^b F(x, \hat{W} + \alpha \eta, \dot{\hat{W}} + \alpha \dot{\eta}) dx - \int_a^b F(x, \hat{W}, \dot{\hat{W}}) dx. \quad (14.13)$$

Это приращение, определенное на классе функций (14.10), зависит только от значения параметра α . При этом минимум функционала ($\Delta I = 0$) достигается на экстремали ($\alpha = 0$).

Разложив приращение ΔI в окрестности экстремали $\hat{W}(x)$ в ряд Тейлора (производные берутся на экстремали, т. е. при $\alpha = 0$), можно получить

$$\Delta I = \alpha \frac{dI}{d\alpha} + \frac{\alpha^2}{2} \frac{d^2 I}{d\alpha^2} + \dots \quad (14.14)$$

Первый и второй члены разложения (14.14) называются первой и второй вариациями функционала

$$\delta I = \alpha \frac{dI}{d\alpha}; \quad \delta^2 I = \frac{\alpha^2}{2} \frac{d^2 I}{d\alpha^2} \quad (14.15)$$

и играют роль первого и второго дифференциалов функции W (см. разд. 14.1). При малых α можно практически считать

$$\Delta I \simeq \delta I = \alpha \frac{dI}{d\alpha}. \quad (14.16)$$

Если функция $\hat{W}(x)$ сообщает функционалу I минимум, на всех соседних функциях $W(x)$ функционал имеет большие значения и, следовательно,

$$\Delta I \simeq \delta I = \alpha \frac{dI}{d\alpha} \geq 0. \quad (14.17)$$

Так как α может быть произвольным числом, из (14.17) следует

$$\frac{dI}{d\alpha} = 0. \quad (14.18)$$

В случае максимума условие (14.17) имеет вид $\alpha \frac{dI}{d\alpha} \leq 0$ и также приводит к соотношению (14.18).

С другой стороны, из (14.11) следует

$$\begin{aligned} \frac{dI}{d\alpha} &= \int_a^b \frac{d}{d\alpha} F(x, \hat{W} + \alpha\eta, \hat{\dot{W}} + \alpha\dot{\eta}) dx = \\ &= \int_a^b \left[\frac{\partial F}{\partial W} \frac{d(\hat{W} + \alpha\eta)}{d\alpha} + \frac{\partial F}{\partial \dot{W}} \frac{d(\hat{\dot{W}} + \alpha\dot{\eta})}{d\alpha} \right] dx = \\ &= \int_a^b \left(\frac{\partial F}{\partial W} \eta + \frac{\partial F}{\partial \dot{W}} \dot{\eta} \right) dx. \end{aligned} \quad (14.19)$$

Интегрируя второй член (14.19) по частям, с учетом условия $\eta(a) = \eta(b) = 0$ можно получить

$$\int_a^b \frac{\partial F}{\partial \dot{W}} \dot{\eta} dx = F_{\dot{W}} \eta \Big|_a^b - \int_a^b \eta \frac{d}{dx} F_{\dot{W}} dx = - \int_a^b \eta \frac{d}{dx} F_{\dot{W}} dx, \quad (14.20)$$

и, следовательно,

$$\frac{\partial I}{\partial \alpha} = \int_a^b \left(F_W - \frac{d}{dx} F_{\dot{W}} \right) \eta(x) dx. \quad (14.21)$$

$\eta(x)$ — произвольная функция, и поэтому в силу основной леммы вариационного исчисления (лемма Лагранжа) из (14.21) следует

$$F_{\hat{W}} - \frac{d}{dx} F_{\hat{\dot{W}}} = 0, \quad \forall x \in [a, b]. \quad (14.22)$$

Соотношение (14.22) и представляет собой уравнение Эйлера, выражающее необходимое условие экстремума и в той или иной форме лежащее в основе всех задач вариационного исчисления. Общее решение уравнения Эйлера содержит две постоянные, для определения которых, как правило, задаются значения функционала в начале и конце исследуемого интервала $W(a)$, $W(b)$. В общем случае функция F нелинейна относительно W и $\dot{W}$, поэтому уравнение Эйлера нелинейно и не имеет аналитического решения. Для численного (прямого) решения задач вариационного исчисления разработан ряд методов, в частности, методы Ритца и Эйлера.

Так как на экстремали первая вариация обращается в нуль, знак приращения определяется второй вариацией. В случае мини-

или в векторной форме

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(t, \mathbf{x}, \mathbf{u}); \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}; \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix}; \quad \mathbf{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_k \end{pmatrix} \quad (14.28a)$$

с краевыми условиями

$$\mathbf{x}(t_0) \in S_0; \quad \mathbf{x}(t_1) \in S_1; \quad t_0 \leq t \leq t_1, \quad (14.29)$$

где S_0, S_1 — заданные многообразия; $\mathbf{u}(u_1, u_2, u_m)$ — вектор управления, принадлежащий классу кусочно-непрерывных функций $\mathbf{u} \in U$, включающему, в частности, функции с бесконечной производной. Требуется найти вектор $\mathbf{u}$, удовлетворяющий условиям (14.28), (14.29) и минимизирующий некоторый функционал

$$I(\mathbf{x}) = \int_{t_0}^{t_1} F(\mathbf{x}, \mathbf{u}, t) dt. \quad (14.30)$$

В отличие от формулировки задачи классического вариационного исчисления при использовании принципа Понтрягина:

1) в качестве независимой переменной используется время t , что четко определяет ориентацию метода на решение динамических задач;

2) для обозначения фазовых координат вместо $W(t)$ в (14.24) вводятся переменные $\mathbf{x}(t)$, а вместо производных фазовых координат $\dot{W}(t)$ — управления $\mathbf{u}(u_1, u_2 \dots u_m)$:

3) роль ограничений играют дифференциальные уравнения (14.28), описывающие поведение системы.

Для решения задачи вводится специальная функция Гамильтона:

$$H = -F(\mathbf{x}, \mathbf{u}, t) + \sum_{i=1}^k \Psi_i(t) f_i(\mathbf{x}, \mathbf{u}, t). \quad (14.31)$$

С помощью ряда преобразований ценой увеличения числа переменных и связей из рассмотрения исключается время t и задача сводится к системе дифференциальных уравнений, включающих в себя вспомогательные функции Ψ_i :

$$\frac{\partial \Psi_i}{\partial t} = - \sum_{j=0}^{n+1} \frac{\partial f_i(\mathbf{x}(t), \mathbf{u})}{\partial x_j} \Psi_j; \quad i = 0, 1, \dots, n+1. \quad (14.32)$$

Для решения системы (14.32) используются итерационные численные методы.

Методы классического вариационного исчисления и методы, основанные на принципе максимума Понтрягина, широко используются при оптимизации систем автоматического регулирования, при решении навигационных задач, например, при определении оптимальных трасс для перелетов и перевозки грузов, а также при изучении переходных процессов включения (выключения) двигателя или его элементов.

Вопросы для самоконтроля

1. Сформулируйте задачу вариационного исчисления.
2. Какова роль первой и второй вариаций при решении задач вариационного исчисления?
3. Сравните вариационное исчисление и методы, основанные на принципе максимума Понтрягина.

14.3. ДИНАМИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

Методом, позволяющим решать вариационные задачи при сложном виде возможных управлений, является метод *динамического программирования*, разработанный Р. Беллманом [2].

В основе динамического программирования лежит *принцип оптимальности Беллмана*. Согласно этому принципу оптимальное управление в любой момент времени не зависит от предыстории системы и определяется только целью управления и состоянием системы в рассматриваемый момент. Это значит, что если предпоследнее состояние системы известно, то независимо от того, как система попала в это состояние, управление на последнем участке должно быть наилучшим из возможных управлений. Таким образом, существенной предпосылкой метода является предположение об отсутствии последствия предыдущих управлений: их результат полностью характеризуется состоянием системы.

Нужно подчеркнуть, что принцип оптимальности не может быть распространен на любой участок траектории движения. Действительно, при беге на длинную дистанцию неправильно требовать от бегуна максимальной скорости на каждом круге (хотя эта скорость и оптимальна для каждого круга в отдельности). Однако на *последнем* круге уже нет смысла экономить силы, и оптимальная для этого круга максимальная скорость будет одновременно оптимальной и для дистанции в целом.

Пусть критерий оптимальности выражается функционалом (14.30). Оптимальное решение, обеспечивающее минимум функционала, зависит от начальных условий и может быть записано в виде

$$S(x_0) = \min_x \int_{t_0}^{t_1} F(x, u, t) dt. \quad (14.33)$$

В силу аддитивности операции интегрирования

$$\int_{t_0}^{t_1} = \int_{t_0}^{t_1 - \Delta t} + \int_{t_1 - \Delta t}^{t_1}. \quad (14.34)$$

Тогда в соответствии с принципом оптимальности

$$S(x_0) = \min_{x(t_0, t_1 - \Delta t)} \left[\int_{t_0}^{t_1 - \Delta t} F(x, u, t) dt + S(x_{t_1 - \Delta t}) \right]. \quad (14.35)$$

Соотношение (14.35) называется *функциональным уравнением Беллмана* и определяет рекуррентную процедуру для решения задач оптимизации на ЭВМ. Действительно, заменяя интеграл в выражении для функционала (14.33) суммой, а дифференциальные Уравнения (14.28) разностными, можно получить

$$I(x) = \sum_{i=0}^{N-1} F(x_i, u_i, i\Delta t) \Delta t; \quad x_{i+1} = x_i + f(x_i, t_i, i\Delta t); \quad (14.36)$$

$$t_0 = 0; \quad t_1 = N\Delta t.$$

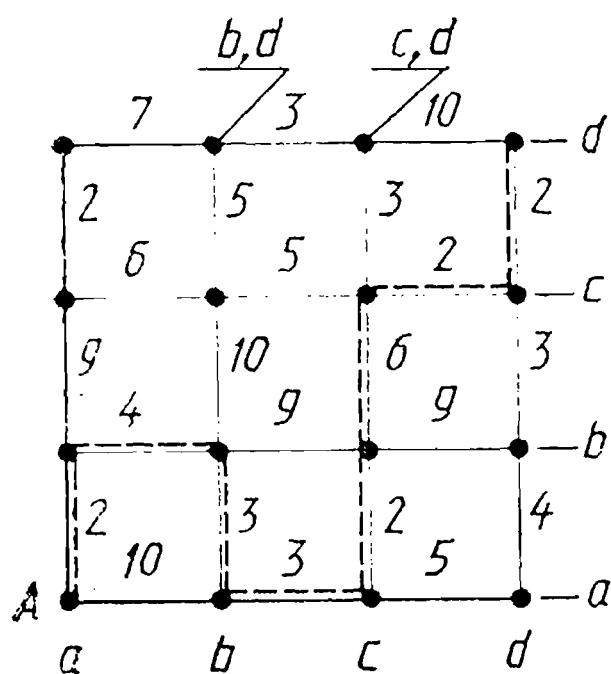


Рис. 14.3. Пример решения задачи методом динамического программирования

Из (14.35) следует

$$S(x_0) = \min_{x_k} \{ F(x_k, u_k, k\Delta t) \Delta t + S_{k+1} \times [x_k + f(x_k, u_k, k\Delta t) \Delta t] \}. \quad (14.37)$$

Соотношение (14.37) позволяет на каждом участке траектории выбирать оптимальное управление начиная с $N-1$ -го участка. После того как, начиная с конца, на каждом участке

$$[(N-1)\Delta t, N\Delta t]; [(N-2)\Delta t, N\Delta t]; \dots; [0, N\Delta t] \quad (14.38)$$

оптимальное управление получено, необходимо двигаться от начала в конец и получить оптимальную траекторию в фазовом пространстве, состоящую из оптимальных отрезков и соединяющую начальную и конечную точки.

С помощью перехода к дифференциальным соотношениям можно показать [10], что функциональное уравнение Беллмана по существу эквивалентно уравнениям, вытекающим из принципа максимума Понтрягина.

Рекуррентные соотношения динамического программирования (14.37) основаны на дискретном представлении исследуемого функционала и потому в равной мере пригодны для задач с непрерывно и дискретно меняющимися переменными.

Для иллюстрации метода динамического программирования можно рассмотреть в качестве примера решение простейшей транспортной задачи.

Пусть минимизируемый функционал представляет собой время, затрачиваемое на движение автомобиля из пункта А в пункт В. Предполагается, что эти пункты соединены прямоугольной сетью дорог (рис. 14.3), причем для каждого участка известно время движения в минутах, указанное на рисунке цифрами. Для простоты будем считать, что движение осуществляется в три этапа, на каждом из которых автомобиль перемещается на следующую вертикальную полосу.

Рассмотрим третий этап движения. Автомобиль из любого возможного положения на вертикальной линии с должен попасть в точку В. Если перед последним этапом автомобиль находился в точке (c, d) (первая буква характеризует вертикаль, вторая — горизонталь), возможны четыре варианта пути, приводящие к соответствующим затратам времени:

$$(c, d) \Rightarrow B, t_3 = 10 \text{ мин};$$

$$(c, d) \Rightarrow (c, c) \Rightarrow (d, c) \Rightarrow B, t_3 = 7 \text{ мин}; \quad (14.39)$$

$$(c, d) \Rightarrow (c, c) \Rightarrow (c, b) \Rightarrow (d, b) \Rightarrow (d, c) \Rightarrow \mathbf{B}, t_3 = 23 \text{ мин};$$

$$(c, d) \Rightarrow (c, c) \Rightarrow (c, b) \Rightarrow (c, a) \Rightarrow (d, a) \Rightarrow (d, b) \Rightarrow (d, c) \Rightarrow \mathbf{B},$$

$$t_3 = 25 \text{ мин.}$$

Очевидно, оптимален второй вариант пути:

$$(c, d) : (c, d) \Rightarrow (c, c) \Rightarrow (d, c) \Rightarrow \mathbf{B}, t_3 = 7 \text{ мин.}$$

Аналогично рассчитываются оптимальные траектории из других точек вертикальной линии c . Эти траектории будут

$$(c, c) : (c, c) \Rightarrow (d, c) \Rightarrow \mathbf{B}, t_3 = 4;$$

$$(c, b) : (c, b) \Rightarrow (c, c) \Rightarrow (d, c) \Rightarrow \mathbf{B}, t_3 = 10; \quad (14.40)$$

$$(c, a) : (c, a) \Rightarrow (c, b) \Rightarrow (c, c) \Rightarrow (d, c) \Rightarrow \mathbf{B}, t_3 = 12.$$

На предпоследнем этапе (этапе 2) осуществляется переход с вертикальной линии b на линию c . Перед этим этапом автомобиль может находиться в любой точке вертикали b . Если автомобиль находится в точке (b, d) , то попасть в точку вертикали c он может четырьмя путями:

$$(b, d) \rightarrow (c, d) \rightarrow \dots, t_2 = 3, t_2 + t_3 = 10;$$

$$(b, d) \rightarrow (b, c) \rightarrow (c, c) \rightarrow \dots, t_2 = 10, t_2 + t_3 = 14;$$

$$(b, d) \rightarrow (b, c) \rightarrow (b, b) \rightarrow (c, b) \rightarrow \dots, t_2 = 24, t_2 + t_3 = 34; \quad (14.41)$$

$$(b, d) \rightarrow (b, c) \rightarrow (b, b) \rightarrow (b, a) \rightarrow (c, a) \rightarrow \dots, t_2 = 21, t_2 + t_3 = 33.$$

Очевидно, оптимален по суммарному времени $t_2 + t_3$ первый путь:

$$(b, d) : (b, d) \rightarrow (c, d) \rightarrow \dots, t_2 + t_3 = 10 \text{ мин.}$$

Важно подчеркнуть, что расчет ведется лишь для *текущего* этапа, так как оптимальные траектории для последнего этапа *уже определены*.

Оптимальные траектории из других точек второй вертикали определяются соотношениями

$$(b, c) : (b, c) \rightarrow (c, c) \rightarrow \dots, t_2 = 5, t_2 + t_3 = 9;$$

$$(b, b) : (b, b) \rightarrow (b, a) \rightarrow (c, a) \rightarrow \dots, t_2 = 6, t_2 + t_3 = 18; \quad (14.42)$$

$$(b, a) : (b, a) \rightarrow (c, a) \rightarrow \dots, t_2 = 3, t_2 + t_3 = 15.$$

Наконец, на первом (третьем с конца!) этапе автомобиль перемещается из точки $\mathbf{A}$ в одну из точек второй вертикали. Это можно сделать четырьмя способами:

$$\mathbf{A} \rightarrow (a, b) \rightarrow (a, c) \rightarrow (a, d) \rightarrow (b, d) \rightarrow \dots, t_1 = 20, t_2 = 30;$$

$$\mathbf{A} \rightarrow (a, b) \rightarrow (a, c) \rightarrow (b, c) \rightarrow \dots, t_1 = 17, t_2 = 26; \quad (14.43)$$

$$\mathbf{A} \rightarrow (a, b) \rightarrow (b, b) \rightarrow \dots, t_1 = 6, t_2 = 24;$$

$$\mathbf{A} \rightarrow (b, a) \rightarrow \dots, t_1 = 10, t_2 = 25.$$

Легко видеть, что оптимальная траектория движения (на рис. 14.3 показана пунктиром)

$$A \rightarrow (a, b) \rightarrow (b, b) \rightarrow (b, a) \rightarrow (c, a) \rightarrow (c, b) \rightarrow (c, c) \rightarrow (d, c) \Rightarrow B. \quad (14.44)$$

Достоинством метода динамического программирования по сравнению с прямым перебором вариантов является усечение дерева перебора за счет предварительного выбора оптимальных траекторий на последующих этапах. В частности, в рассмотренной задаче необходимо рассчитать 30 траекторий, в то время как при прямом переборе — 16384 траектории, причем при увеличении членения задачи выигрыш растет.

Широкое применение метод динамического программирования получил при решении транспортных задач, задач распределения ресурсов, замены оборудования и т. д. При исследовании элементов ДУ метод динамического программирования может быть использован для оптимизации конфигураций деталей, если их поверхность представлена как некоторая траектория в пространстве.

Вопросы для самоконтроля

1. В чем заключается принцип оптимальности Беллмана?
2. Для решения каких задач используется метод динамического программирования?

14.4. ПОИСК ОПТИМУМА ЛИНЕЙНЫХ МОДЕЛЕЙ (СИМПЛЕКС-МЕТОД)

Для линейных моделей (см. разд. 5.1) может быть предложен детерминированный метод перебора возможных решений, позволяющий за конечное число итераций найти точное оптимальное решение. По существу при этом происходит последовательное исследование вершин некоторого полиэдрального множества, в связи с чем этот метод известен под названием *симплекс-метода* [3].

Предварительный анализ задачи линейного программирования показывает, что для поиска оптимального решения необходимо включить ограничения в функцию цели, а затем найти оптимальные значения переменных. В общем случае эта задача не является тривиальной. Симплекс-метод предусматривает построение некоторого возможного *базисного* решения (предполагается, что множество допустимых решений не пусто и, значит, какие-то решения возможны) и его последовательное улучшение. Изложение метода целесообразно начать с рассмотрения простого примера (5.3). Из рис. 5.1, а ясно, что задача имеет бесконечное множество оптимальных решений, расположенных на прямой $2x_1 + 3x_2 = 18$, причем $W = -18$.

Для формального поиска оптимума вводятся дополнительные свободные переменные (5.4) по числу ограничений x_3, x_4, x_5 , позволяющие свести задачу к каноническому виду:

$$\begin{aligned} W + 2x_1 + 3x_2 &= 0; \\ 2x_1 + 3x_2 + x_3 &= 18; \\ x_1 - 4x_2 + x_4 &= 4; \\ -2x_1 + x_2 + x_5 &= 2. \end{aligned} \quad (14.45)$$

Дополнительные свободные переменные удовлетворяют соотношениям

$$x_3 \geq 0; x_4 \geq 0; x_5 \geq 0. \quad (14.46)$$

В качестве базисного может быть выбрано любое решение из области допустимых решений (рис. 5.1, а). Удобно выбрать это решение в виде

$$x_1=0, x_2=0, x_3=18, x_4=4, x_5=2, W=0. \quad (14.47)$$

Очевидно, это решение (на рис. 5.1, а оно отмечено точкой с цифрой 1 в кружочке) не относится к числу лучших, но может быть улучшено при оптимизации. Отличные от нуля переменные (x_3, x_4, x_5) называют *базисными* переменными, переменные, равные нулю (x_1, x_2), *внебазисными*.

Улучшение базисного плана может быть достигнуто за счет изменения (увеличения) внебазисных переменных. При этом увеличение на единицу переменной x_1 приводит к уменьшению W на 2 единицы, увеличение на единицу переменной x_2 — к уменьшению W на 3 единицы.

Формальным признаком возможности улучшения W (*критерий улучшения плана — симплекс-критерий I*) является наличие в первой строке системы (14.45) переменных с положительными коэффициентами. Действительно, увеличение таких переменных должно приводить к уменьшению W . Если таких переменных нет, решение признается *неулучшаемым*, т. е. оптимальным, и вычисления прекращаются. В рассматриваемом случае обе переменные (x_1, x_2) входят в выражение для W с положительными коэффициентами и значит, улучшение возможно. Однако увеличение x_1, x_2 в силу ограничений приводит к уменьшению других переменных (x_3, x_4, x_5) а они, в свою очередь, не должны быть меньше нуля (14.46). Порядок улучшения решения определяется *симплекс-критерием II*, который требует выбора наиболее сильно влияющей на W переменной (в данном случае x_2), определения ее максимально возможного значения и включения ее в базис вместо одной из базисных переменных.

В силу первого ограничения (2-я строка системы (14.45)) x_2 не может быть больше 6, в силу третьего ограничения (4-я строка системы (14.45)) x_2 не может быть больше 2. Второе ограничение (3-я строка системы (14.45)), куда x_2 входит с отрицательным

коэффициентом, при увеличении x_2 не действует. Это, кстати, очевидно из рис. 5.1, а. Таким образом, необходимо изменить значения x_2 и x_5 :

$$x_2=2, \; x_5=0, \tag{14.48}$$

и включить x_2 в базис вместо x_5 . Результаты первой итерации формализованы в табл. 14.1.

Процедура замены базиса требует исключения новой базисной переменной (x_2) из выражения для W и преобразования ограниче-

Таблица 14.1

Базисное решение	Пробное решение	Допустимое значение x_2	Минимальное значение x_2	Следующее пробное решение
W_0	0	—		
x_3	18	6		
x_4	4	8		
x_5	2	2	2	$x_2=2, \; x_5=0$

Таблица 14.2

Базисное решение	Пробное решение	Допустимое значение x_1	Минимальное значение x_1	Следующее пробное решение
W_0	—6	—		
x_2	2	3/2		
x_3	12	8		
x_4	12	8	3/2	$x_1=3/2, \; x_3=0$

ний. Для этого четвертую строку системы (14.45) — действующее ограничение на x_2 — нужно умножить на 3 и вычесть из первой и второй строк системы, а потом умножить на 4 и прибавить к третьей строке системы:

$$\begin{array}{rcll} W & +8x_1 & & -3x_5 = -6; \\ & 8x_1 & +x_3 & -3x_5 = 12; \\ & -7x_1 & & +x_4 +4x_5 = 12; \\ & -2x_1 & +x_2 & +x_5 = 2. \end{array} \tag{14.49}$$

После преобразования новая базисная переменная (x_2) входит только в одну строку системы (14.49), эквивалентной исходной системе (14.45). Новое базисное решение (на рис. 5.1, а показано точкой с цифрой 2 в кружочке)

$$x_1=0, \; x_2=2, \; x_3=12, \; x_4=12, \; x_5=0, \; W=-6 \tag{14.50}$$

очевидно, лучше исходного решения, но оно по-прежнему (в силу симплекс-критерия I) допускает улучшение за счет переменной x_1 . Переменная x_1 в силу второй строки системы (14.49) не может быть больше 3/2. Третья и четвертая строка системы не накладывают ограничений на увеличение x_1 . Поэтому необходимо включать в базис переменную x_1 вместо x_3 :

$$x_1=3/2, \; x_3=0. \tag{14.51}$$

Результаты второй итерации формализованы в табл. 14.2.

Преобразование системы (14.49) осуществляется так же, как и в предыдущем случае, однако необходимо дополнительно преоб-

разовать вторую строку системы (действующее ограничение на x_1) с тем, чтобы новая базисная переменная x_1 входила в нее с коэффициентом 1:

$$\begin{aligned} W & -x_3 & & = -18; \\ x_1 & + 1/8x_3 & - 3/8x_5 & = 12/8; \\ & 7/8x_3 & + x_4 & + 11/8x_5 = 180/8; \\ x_2 & + 2/8x_3 & & + 2/8x_5 = 40/8. \end{aligned} \quad (14.52)$$

Новое базисное решение (точка 3 в кружочке на рис. 5.1, а)

$$x_1 = 3/2; x_2 = 5; x_3 = 0; x_4 = 45/2; x_5 = 0; W = -18. \quad (14.53)$$

в силу симплекс-критерия I не допускает дальнейшего улучшения (попытка увеличивать x_3 приводит лишь к увеличению W) и, следовательно, является оптимальным. Оптимальная точка, разумеется, лежит на прямой $2x_1 + 3x_2 = 18$. Остается убедиться, что значения переменных (14.53) действительно удовлетворяют (5.3). Из рис. 5.1, а видно, что поиск оптимума осуществляется путем перебора вершин области возможных решений.

Формально изложенный метод сводится к выполнению следующих процедур.

1. В соответствии с числом ограничений вводятся дополнительные свободные переменные и задается первоначальное базисное решение, включающее в себя нулевые действительные переменные и ненулевые свободные переменные.

2. Проверяется возможность улучшения плана (симплекс-критерий I). Если улучшение возможно, осуществляется переход к процедуре 3. Если улучшение невозможно, решение считается окончательным (оптимальным) и вычисления прекращаются.

3. В соответствии с симплекс-критерием II выбирается путь улучшения плана, т. е. новая базисная переменная, и определяется ее максимально допустимое значение. Одновременно определяется переменная, которую нужно исключить из базиса.

4. Изменяется базисное решение. Преобразуется система уравнений задачи, после чего осуществляется переход к процедуре 2.

Схема алгоритма симплекс-метода представлена на рис. 14.4.

Как уже отмечалось, задачи линейного программирования могут иметь множество равнозначных оптимальных решений. Симплекс-метод гарантирует получение *одного* из этих решений. Сама процедура симплекс-метода неоднозначна. Определенный произвол заключен как в выборе первоначального базисного решения, так и пути совершенствования этого решения. В частности, в рассмотренном примере при первоначальном улучшении по переменной x_1 будет получено¹ равнозначное (14.53) оптимальное решение (см. точку 4 в кружочке на рис. 5.1, а):

$$x_1 = 84/11; x_2 = 10/11; x_3 = 0; x_4 = 0; x_5 = 180/11; W = -18, \quad (14.54)$$

также удовлетворяющее системе (5.3) и лежащее на прямой $2x_1 + 3x_2 = 18$. В тех случаях, когда линейная модель имеет более одного оптимального решения, она имеет бесконечное число таких решений. При этом можно показать, что любое

¹ Читателю рекомендуется проделать соответствующие выкладки.

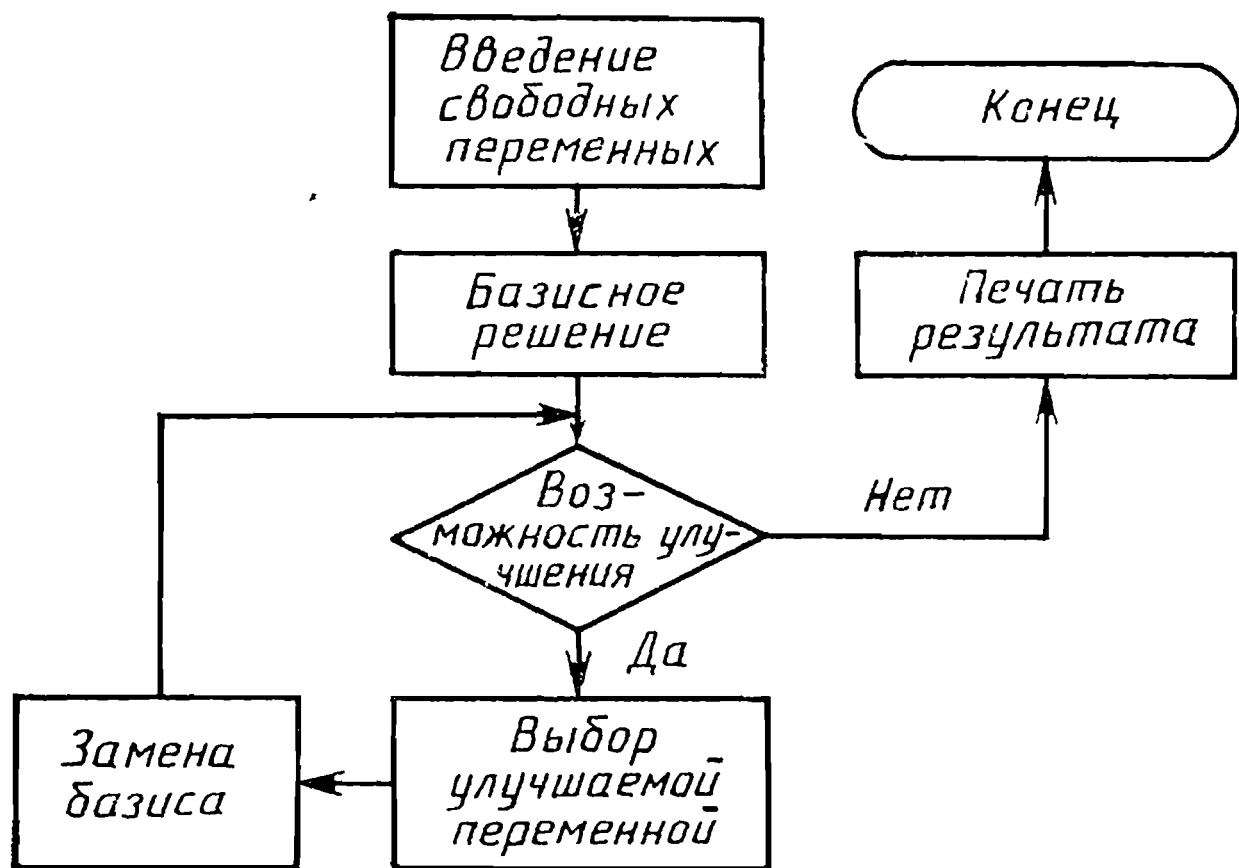


Рис. 14.4. Структурная схема программы для решения задачи линейного программирования симплекс-методом

положительно взвешенное среднее двух оптимальных решений тоже является эквивалентным оптимальным решением. Для рассмотренного примера будут оптимальны все значения x_1 и x_2 , отвечающие условиям

$$x_1 = \frac{3}{2} w + \frac{84}{11} (1 - w); \quad x_2 = 5w + \frac{10}{11} (1 - w);$$

$$0 < w < 1. \quad (14.55)$$

Рассмотренный пример сопровождался рядом громоздких записей, преследовавших цель пояснения алгоритма. На практике объем этих записей можно существенно сократить, если использовать табличное представление данных (табл. 14.3 или 14.4 [3]). Первая таблица содержит базис каждой итерации, пробные значе-

Таблица 14.3

Итерация	Базис	Значения переменных	Коэффициенты при переменных					Строка
			x_1	x_2	x_3	x_4	x_5	
Исходное решение	W	0	2	3	—	—	—	1
	x_3	18	2	3	1	—	—	2
	x_4	4	1	—4	—	1	—	3
	x_5	2	—2	1	—	—	1	4
1	W	—6	8	—	—	—	3	1
	x_2	2	8	—	1	—	3	2
	x_3	12	—7	—	—	1	4	3
	x_4	12	—2	1	—	—	1	4
2	W	—18	—	—	1	—	—	1
	x_1	8/2	1	—	1/8	—	3/8	2
	x_2	5	—	—	7/8	1	11/8	3
	x_4	45/2	—	1	2/8	—	2/8	4

Итерация	Базис	Пробное решение	x_1	x_2	Строка
Исходное решение	W	0	2	3	1
	x_3	18	2	3	2
	x_4	4	1	-4	3
	x_5	2	-2	1	4
			x_1	x_5	
1	W	-6	8	3	1
	x_2	2	8	3	2
	x_3	12	-7	4	3
	x_4	12	2	1	4
			x_3	x_5	
2	W	-18	1	-	1
	x_1	3/2	1/8	3/8	2
	x_2	5	7/8	11/8	3
	x_4	45/2	2/8	2/8	4

ния переменных базиса и W , а также коэффициенты при переменных в каждой из строк системы уравнений. Во второй таблице отсутствуют столбцы базисных переменных.

Вопросы для самоконтроля

- 1. Для каких задач используется симплекс-метод?
- 2. Изложите последовательность действий при использовании симплекс-метода.
- 3. В чем состоит смысл симплекс-критерия I и симплекс-критерия II?
- 4. С помощью симплекс-метода найдите оптимум следующей линейной модели:

$W = 2x_1 - 5x_2$

при ограничениях

$x_1 + x_2 = 10; x_1 - 2x_2 = 6; -2x_1 + x_2 = 5;$
 $x_1 \geq 0, x_2 \geq 0.$

ГЛАВА 15. ПРЯМЫЕ МЕТОДЫ ОПТИМИЗАЦИИ ДЛЯ (ОДНОМЕРНЫЙ ПОИСК)

Прямые методы оптимизации предполагают использование единственной процедуры — процедуры определения величины критерия оптимальности при заданной комбинации варьируемых параметров.

Прямые методы предъявляют минимальные требования к исследуемой математической модели. Эта модель должна представлять собой замкнутую систему зависимостей, позволяющую перейти от значений варьируемых переменных к однозначно определяемому критерию. При этом не требуется использования, например, аналитических выражений для производных. Можно обойтись и вообще без математической модели: варьируемые переменные могут быть заданы, а критерий измерен непосредственно в эксперименте. Возможны и комбинированные модели, в которых сочетаются эксперимент и частичная математическая обработка данных. Таким образом, прямые методы обеспечивают единый методический подход к задачам численной и экспериментальной оптимизации. Поэтому выражения «вычисление функции цели» в данной точке факторного пространства и «эксперимент» в этой точке употребляются как равнозначные.

Прямые методы оптимизации принципиально носят *приближенный* характер. При отсутствии аналитических соотношений результаты поиска представляют собой последовательность более или менее случайных точек с вычисленными в них значениями критерия. Вероятность того, что среди них окажется собственно экстремальная точка, очевидно, равна нулю. В лучшем случае можно рассчитывать на определение той или иной *окрестности* экстремума и величины критерия, характерной для этой окрестности.

Область X^{**} , в которой лежит экстремум x^* , называется *интервалом неопределенности*. До проведения исследования интервал неопределенности совпадает с областью определения экстремума $X^{**} = X^*$ (при использовании функции штрафа с полным факторным пространством X). Лучшей будет та стратегия поиска, которая обеспечивает наибольшее сокращение интервала неопределенности ($X^{**}/X^* \rightarrow \min$) при фиксированных затратах на поиск или наименьшие затраты на поиск при заданном сокращении интервала неопределенности.

При оптимизации основная доля усилий и основные затраты машинного времени связаны с многократными вычислениями критерия оптимальности. Поэтому затраты на поиск принято определять числом исследованных точек факторного пространства n .

В зависимости от характера стратегии методы поиска принято делить на *пассивные* и *последовательные*. К пассивным относятся методы, при которых все точки факторного пространства для определения величины критерия задаются заранее до начала поиска. При этом дополнительная информация, полученная в процессе поиска, для выбора этих точек не используется. Последовательные методы, наоборот, предусматривают выбор следующей точки факторного пространства на основе анализа результатов в уже исследованных точках. Последовательные методы приводят к более сложным алгоритмам, но обеспечивают более высокую эффективность поиска.

Наконец, в зависимости от числа варьируемых переменных методы поиска делят на *одномерные* ($k=1$) и *многомерные* ($k \geq 2$).

Подавляющее большинство задач оптимального проектирования ДУ ЛА и ее узлов многомерно. Более того, попытка свести такие задачи к одномерным чревата опасностью существенных ошибок. Тем не менее методы одномерной оптимизации представляют значительный интерес: во-первых, в некоторых случаях при известной осторожности все-таки удастся свести рассматриваемые задачи к одномерным; во-вторых, методы одномерной оптимизации являются составной частью большинства многомерных методов, и в-третьих, анализ одномерных методов позволяет наглядно продемонстрировать основные идеи поиска вообще.

15.1. ПРИНЦИП МИНИМАКСА. ОПТИМАЛЬНЫЙ ВЫБОР ДВУХ ОПОРНЫХ ТОЧЕК НА ИНТЕРВАЛЕ НЕОПРЕДЕЛЕННОСТИ

Для исследования одномерных унимодальных моделей широко используются стратегии, основанные на принципе *минимакса*. Идея принципа минимакса может быть пояснена на примере оптимального выбора двух опорных точек для определения значения функции отклика (двух экспериментов) на интервале неопределенности.

Очевидно, исследование функции отклика в единственной точке исходного интервала неопределенности $L_0^{(1)}$ хотя и позволяет судить об общем уровне функции отклика, но не дает оснований для сокращения длины интервала. Единственное значение функции отклика не позволяет судить о характере этой функции (рис. 15.1, а). Поэтому $L_1 = L_0$. Для сокращения интервала неопределенности необходимо исследовать по крайней мере две точки этого интервала. Пусть для определенности эти точки заданы соотношениями

$$\begin{aligned} x_1 &= x_0 + 0,2L_0; \quad x_2 = x_0 + 0,7L_0; \\ x_1, x_2 &\in L_0; \quad x_1 < x_2. \end{aligned} \quad (15.1)$$

В результате исследования для этих точек будут получены два значения функции отклика:

$$W_1 = W(x_1), \quad W_2 = W(x_2), \quad (15.2)$$

которые при условии, что функция *унимодальна*, позволяют указать новый интервал неопределенности L_2 , содержащий оптимум (минимум), причем $L_2 < L_0$.

Действительно, в результате исследования может быть получен один из трех результатов:

$$W_1 < W_2; \quad W_1 > W_2; \quad W_1 = W_2. \quad (15.3)$$

Если $W_1 < W_2$ (рис. 15.1, б), минимум функции не может лежать правее x_2 (в противном случае функция отклика не будет унимодальной!) и интервал неопределенности будет

$$L_2 = x_2 - x_0 = 0,7L_0. \quad (15.4)$$

¹ Индекс при L характеризует количество исследованных точек интервала.

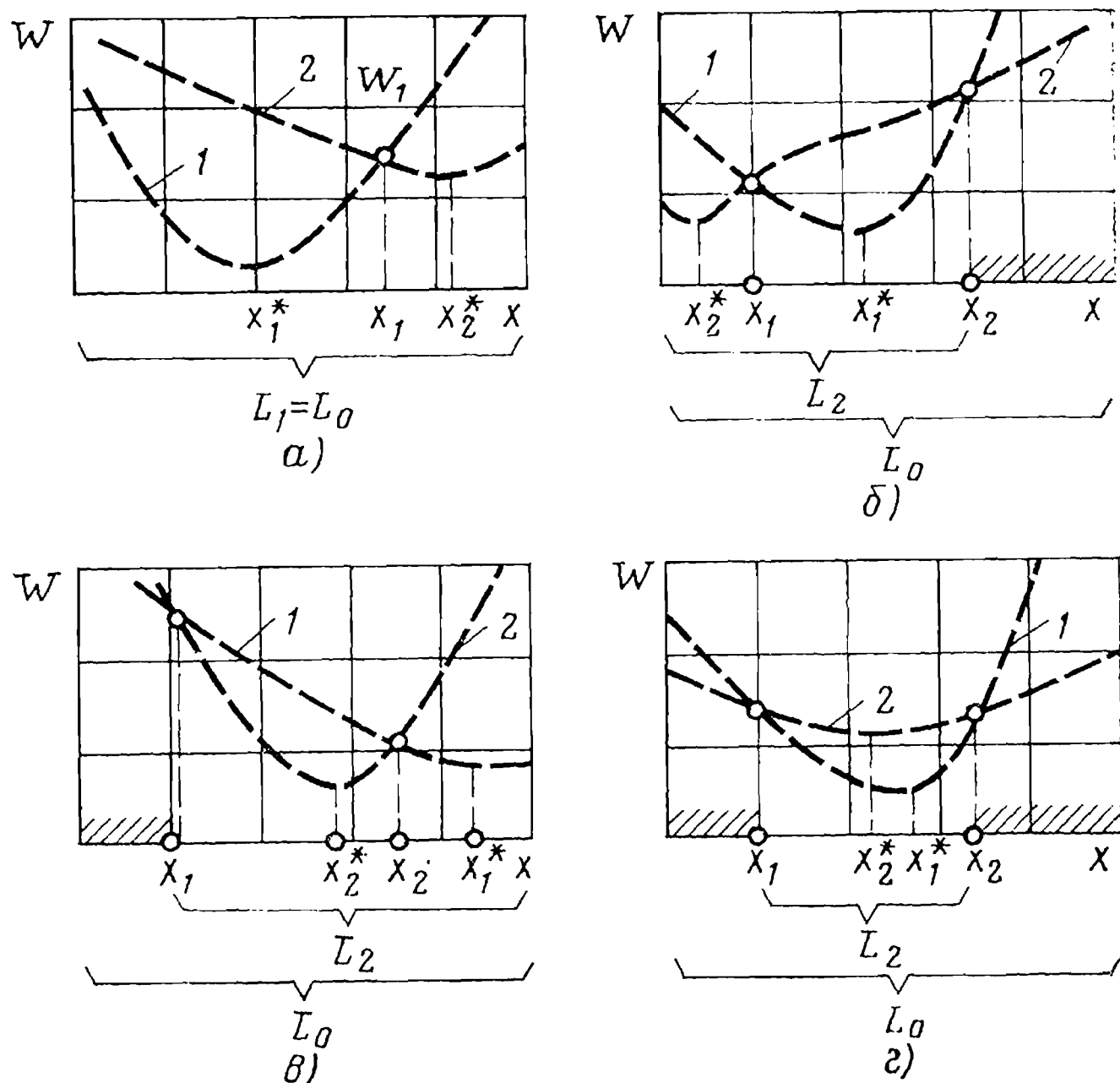


Рис. 15.1. Схема размещения экспериментов при изучении одномерной модели: а — один эксперимент ($L_1 = L_0$); б, в, г — различные исходы в случае двух экспериментов ($L_2 < L_0$)

При $W_1 > W_2$ (рис. 15.1, в) минимум не может находиться левее x_1 :

$$L_2 = x_0 + L_0 - x_1 = 0,8L_0. \quad (15.5)$$

Наконец, в третьем, вообще говоря, маловероятном, случае минимум может находиться только между x_1 и x_2 (рис. 15.1, г):

$$L_2 = x_2 - x_1 = 0,5L_0. \quad (15.6)$$

Легко видеть, что в различных ситуациях величина L_2 оказывается различной, а точное значение L_2 определяется только после эксперимента. Однако для выбора оптимальной стратегии поиска необходимо иметь такой критерий эффективности, который позволял бы судить о конечном интервале неопределенности до проведения исследования и не зависел бы от удачи исследователя.

Из рис. 15.1, б, в видно, что интервал L_2 включает в себя отрезки, лежащие по обе стороны от точки, давшей лучшее значение отклика $W_{\min}$, и ограничен второй (худшей) исследованной точкой и границей интервала L_0 . Аналогично при исследовании n точек интервал L_n ограничен двумя точками, соседними с точкой x_i , где получено лучшее значение отклика (одна из этих точек может быть границей интервала L_0):

$$L_n = x_{i+1} - x_{i-1}, \quad W_{\min} = W(x_i). \quad (15.7)$$

До проведения исследования неизвестно, в какой именно точке будет достигнуто лучшее значение отклика, поэтому необходимо рассмотреть все возможные интервалы:

$$L_2 = \{x_2 - x_0, x_0 + L_0 - x_1\}; L_n = \{x_2 - x_0, x_{i+1} - x_{i-1}, x_0 + L_0 - x_{n-1}\}. \\ i = 1, 2, \dots, n - 1. \quad (15.8)$$

При этом величина всех возможных интервалов L_2, L_n определяется выбором точек $x_1, x_2, \dots, x_n$. При предварительных оценках естественно ориентироваться на худший случай и предполагать, что значения L_2, L_n окажутся

$$L_2 = \max \{L_2\} = \max \{x_2 - x_0, x_0 + L_0 - x_1\}; \quad (15.9) \\ L_n = \max \{L_n\}.$$

(Достаточно искусственный случай $W_1 = W_2$ (см. рис. 15.1, з) или аналогичный ему случай $W_k = W_{k+1}$ приводят к заведомо меньшему интервалу неопределенности и интереса не представляют).

Реальные результаты, полученные при исследовании, могут оказаться только лучше (но не хуже!) этой оценки. Целесообразно так расположить точки $x_1, x_2, \dots, x_n$ на интервале, чтобы эта, вообще говоря, довольно пессимистическая оценка (15.9) все-таки была наилучшей:

$$L_{n \text{ опт}} = \min L_n = \min \max \{L_n\}; L_{2 \text{ опт}} = \min \max \{L_2\}. \quad (15.10)$$

Соотношение (15.10) и представляет собой формулировку принципа минимакса.

Таким образом, принцип минимакса требует ориентации на самый плохой из возможных случаев (в рассмотренном примере это $L_2 = 0,8 L_0$) и выбора стратегии, минимизирующей этот наихудший из возможных интервалов неопределенности. Отсюда и название принципа: минимакс (минимум максимума). Принцип минимакса имеет игровую природу и заимствован из теории антагонистических игр, когда приходится иметь дело с сознательным противником. Стратегии, основанные на принципе минимакса, получили название *минимаксных стратегий*.

Величину интервала неопределенности L_2 можно выразить в функции положения точек x_1, x_2

$$L_2 = \max \{x_2 - x_0, x_0 + L_0 - x_1\} \quad (15.11)$$

или в долях исходного интервала L_0

$$\bar{L}_2 = \max \{\bar{x}_2, 1 - \bar{x}_1\}; \bar{L}_2 = \frac{L_2}{L_0}; \bar{x}_2 = \frac{x_2 - x_0}{L_0}; \quad (15.12)$$

$$\bar{x}_1 = \frac{x_1 - x_0}{L_0}.$$

Несколько необычные линии уровня функции $\bar{L}_2$ показаны на рис. 15.2, а. Например, величина $\bar{L}_2 = 0,8$ может быть получена или при $\bar{x}_2 = 0,8, 1 - \bar{x}_1 \leq 0,8$ (линия $\bar{x}_2 = 0,8; 0,2 \leq \bar{x}_1 \leq 1$), или при

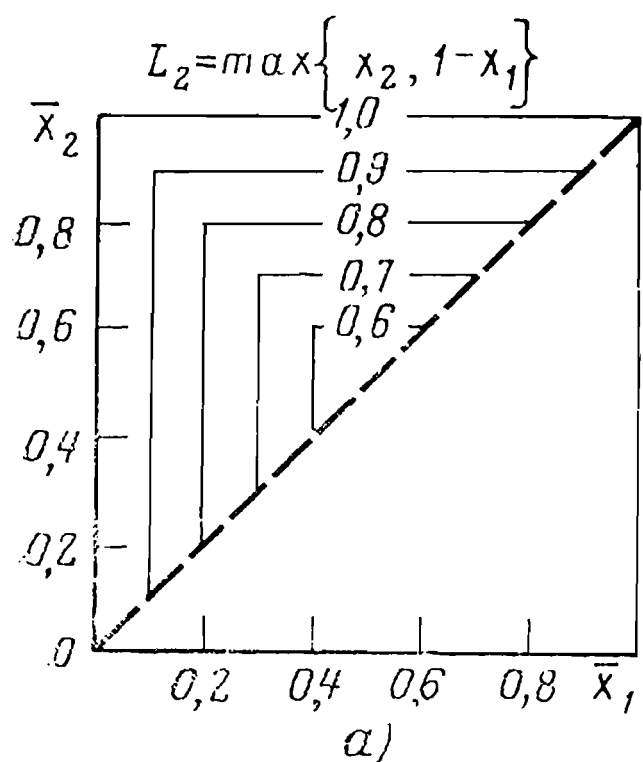
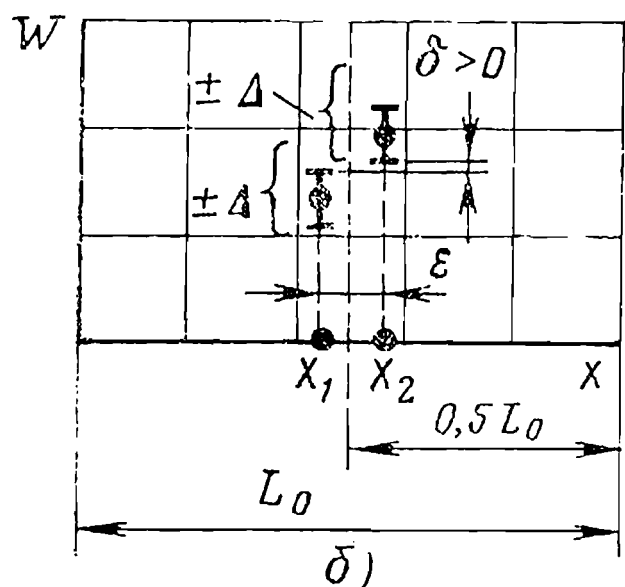


Рис. 15.2. Оптимальное размещение двух экспериментов на интервале неопределенности:
а — линии уровня; *б* — оптимальное размещение двух экспериментов

$1 - \bar{x}_1 = 0,8 \quad \bar{x}_2 \leq 0,8$ (линия $\bar{x}_1 = 0,2$; $0 \leq \bar{x}_2 \leq 0,8$). Из рисунка следует, что наилучшему (наименьшему) минимаксному значению $\bar{L}_2 = 0,5$ соответствовало бы размещение обеих точек $\bar{x}_1, \bar{x}_2$ в центре интервала L_0 . Однако при этом было бы нарушено исходное условие $\bar{x}_1 < \bar{x}_2$, т. е. условие исследования *двух* различных точек интервала.



Точки $\bar{x}_1$ и $\bar{x}_2$ должны отстоять друг от друга на некоторое расстояние ε , которое позволяет трактовать их как две различные точки. Это расстояние ε зависит как от темпа изменения, так и от точности измерения критерия оптимальности W . Дело в том, что и при вычислениях, и тем более при эксперименте величина W определяется с некоторой погрешностью Δ , т. е. на самом деле приходится иметь дело с величиной $W \pm \Delta$. Для того чтобы два значения $W(\bar{x}_1)$ и $W(\bar{x}_2)$ были действитель-

но различны (т. е. для уверенного выделения лучшей и худшей точек), должно соблюдаться условие

$$|W(\bar{x}_1) - W(\bar{x}_2)| > 2\Delta. \quad (15.13)$$

Именно условие (15.13) и определяет величину ε .

Таким образом, оптимальным будет расположение двух точек симметрично относительно центра интервала L_0 на расстоянии ε друг от друга (см. рис. 15.2, б). При этом интервал неопределенности L_2 составит

$$L_2 = 0,5L_0 + \varepsilon/2 \quad (\bar{L}_2 = 0,5 + \varepsilon/2L_0). \quad (15.14)$$

Добавление третьего эксперимента (третьей точки) не позволяет существенно сократить интервал неопределенности. Можно показать [20], что минимаксная оценка L_3 составляет $L_3 = 0,5L_0$, т. е. улучшение, достигаемое за счет третьей точки, равно $\varepsilon/2$.

Вопросы для самоконтроля

1. В чем заключается принцип минимакса?
2. Как оптимальным образом расположить на интервале неопределенности две исследуемые точки?
3. Какова физическая природа величины ε ?

Стратегии поиска, использующие минимаксные оценки интервала неопределенности, называются минимаксными или с учетом величины ε — ε -минимаксными стратегиями.

Возможны различные минимаксные стратегии пассивного поиска. Если исследователь располагает четным числом $n=2p$ точек (экспериментов), то наилучшее размещение соответствует разделению точек на пары, расположенные около равноотстоящих друг от друга центров. При этом в каждой паре точки отстоят друг от друга на ε (подразумевается, что расстояние между парами существенно больше ε), но необязательно симметричны относительно своего центра (рис. 15.3, а). Как легко видеть из рисунка, величина интервала неопределенности после n экспериментов определяется выражением

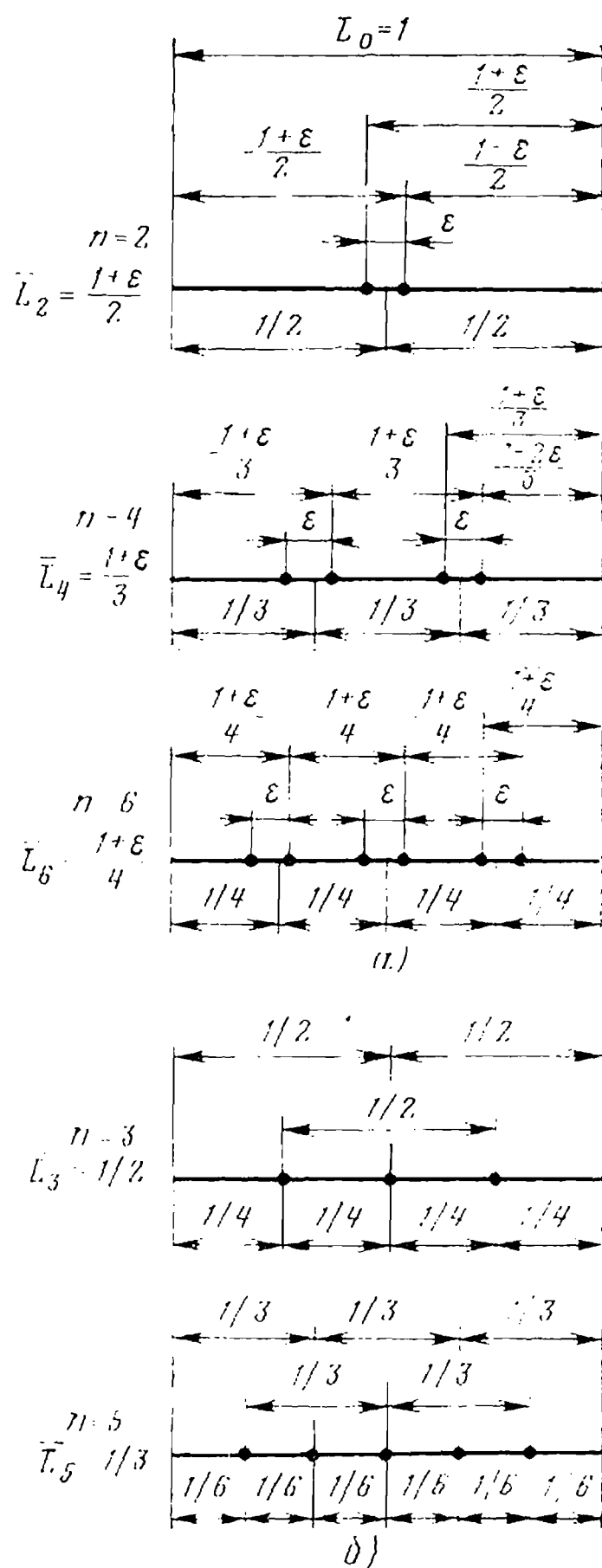
$$L_n = \frac{L_0 + \varepsilon}{\frac{n}{2} + 1}; \quad n = 2p. \quad (15.15)$$

Рассмотренная стратегия является ε -минимаксной. Легко убедиться, что попытка сдвинуть любую из зафиксированных на рис. 15.3, а точек приводит лишь к увеличению минимаксной оценки L_n . Этот метод получил название *метода оптимизации однородными парами*. Из (15.15) следует, что при достаточно малой величине ε для сокращения интервала L_0 , например в 100 раз (точнее, в $100/(1 + \frac{\varepsilon}{L_0})$ раза), необходимо исследовать 198 точек этого интервала.

При нечетном числе исследуемых точек $n=2p+1$ существует множество равноценных минимаксных стратегий. Одна из них предусматривает равномерное расположение точек на интервале L_0 (рис. 15.3, б). Интервал неопределенности при этом составляет

$$L_n = \frac{L_0}{\frac{n-1}{2} + 1}; \quad n = 2p + 1. \quad (15.16)$$

Рис. 15.3. Пассивные ε -минимаксные стратегии: а — метод однородных пар (четное число экспериментов); б — нечетное число экспериментов



Так же, как и при сравнении двух и трех точек, введение дополнительной нечетной точки позволяет сократить интервал неопределенности на некоторую долю ε :

$$L_{2p} - L_{2p+1} = \frac{L_0 + \varepsilon}{p+1} - \frac{L_0}{p+1} = \frac{\varepsilon}{p+1}. \quad (15.17)$$

Таким образом, целесообразность постановки дополнительного нечетного эксперимента существенно зависит от величины ε .

Эффективность поиска может быть существенно увеличена, если при выборе очередной точки использовать информацию, полученную при исследовании предыдущих точек, т. е. при переходе к так называемым *последовательным* стратегиям поиска.

Пусть первая пара точек (предварительной информации пока еще нет!) выбрана в соответствии с рассмотренной выше ε -минимаксной стратегией (см. рис. 15.2, б). После исследования этих точек из дальнейшего рассмотрения в силу унимодальности функции W может быть исключена почти половина интервала L_0 (рис. 15.4, а) и, следовательно, отпадет необходимость в исследовании дополнительных точек в этой области. Для полученного интервала неопределенности L_2 (15.14) может быть спланирована новая оптимально расположенная пара точек, в результате исследования которых будет получен интервал неопределенности L_4 :

$$L_4 = \frac{1}{2} L_2 + \frac{\varepsilon}{2} = \frac{1}{2} \left(\frac{1}{2} L_0 + \frac{\varepsilon}{2} \right) + \frac{\varepsilon}{2} = \frac{1}{4} L_0 + \frac{3}{4} \varepsilon. \quad (15.18)$$

Дальнейшая последовательность действий ясна из рис. 15.4, а. После проведения $n=2p$ экспериментов интервал неопределенности составит

$$L_n = 2^{-\frac{n}{2}} L_0 + \left(1 - 2^{-\frac{n}{2}} \right) \varepsilon. \quad (15.19)$$

Структурная схема алгоритма, реализующего метод, представлена на рис. 15.4, б. Рассмотренный метод получил название метода *дихотомии* (деления пополам).

Легко сравнить эффективность методов дихотомии и однородных пар. Из (15.19) следует, что при использовании метода дихотомии уже после 14 экспериментов интервал неопределенности сократится почти в 128 раз $\left(L_{14} = \frac{1}{128} L_0 + \frac{127}{128} \varepsilon \right)$, в то время как при методе однородных пар для сокращения интервала неопределенности в 100 раз потребуется 198 экспериментов. Этот выигрыш, возрастающий с ростом заданной точности определения экстремума, достигается за счет систематического использования накопленной информации.

Несмотря на высокую эффективность, метод дихотомии не является оптимальным ε -минимальным методом. На каждом шаге дихотомии требуется исследование *двух новых* точек интервала неопределенности. Между тем после проведения первого шага в интервале L_2 (а также во всех последующих интервалах L_{2p}) уже

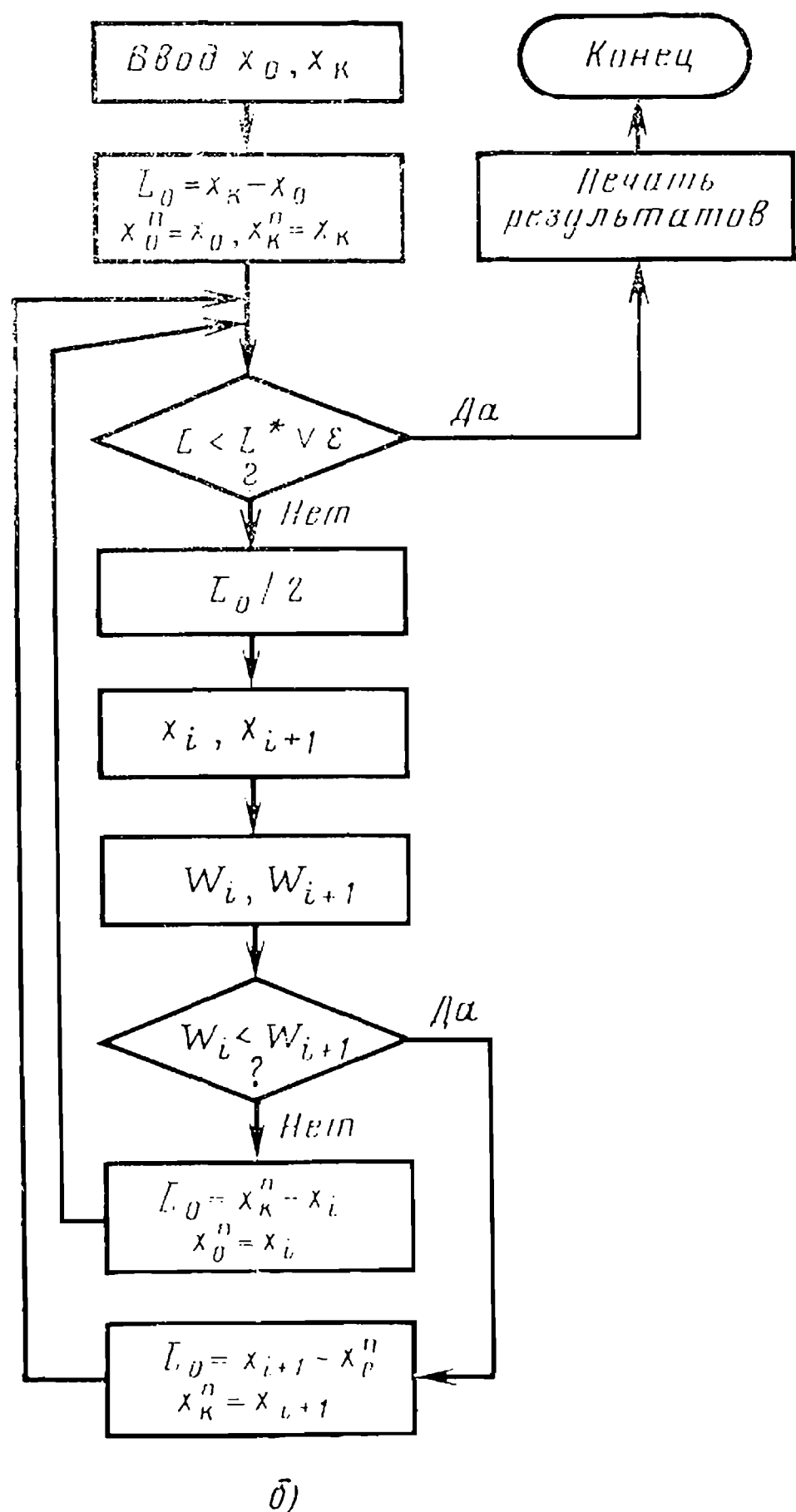
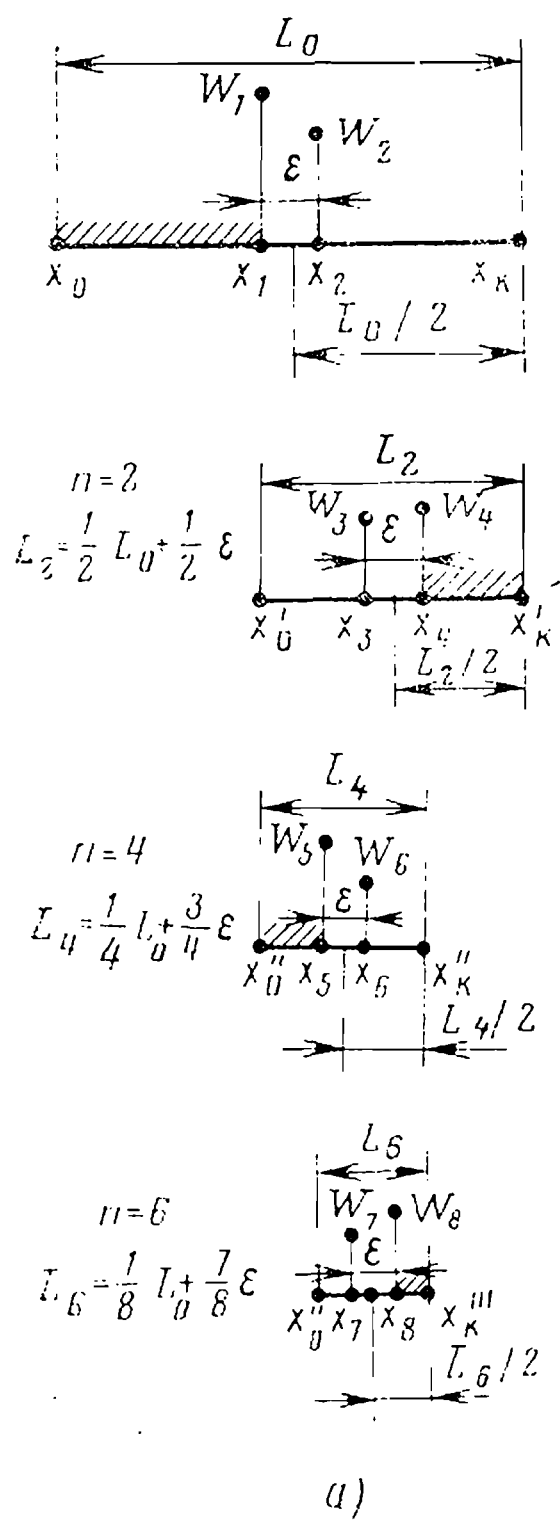


Рис. 15.4. Метод дихотомии:

а — схема поиска оптимума; б — структурная схема программы

имеется одна исследованная точка (рис. 15.4, а), оставшаяся от предыдущего шага. При дальнейшей оптимизации эта точка не используется, так как она лежит очень близко (на расстоянии ϵ) от края интервала. Таким образом, погоня за оптимальностью каждого отдельного шага оптимизации приводит к необходимости исследования дополнительных точек.

При достаточно большом количестве шагов оказывается целесообразным отказаться от оптимальности каждого отдельного шага в пользу оптимальности стратегии в целом. Эта идея реализована в методах, предложенных в 1953 г. американским ученым Кифером и основанных на последовательности чисел Фибоначчи и золотом сечении Эвклида [20].

Пусть для сокращения исходного интервала неопределенности L_0 предполагается исследовать n точек интервала¹. Перед прове-

¹ Изложение метода Кифера ведется по схеме, заимствованной из работы [20].

дением исследования в последней n -й точке исследователь располагает интервалом неопределенности L_{n-1} , внутри которого есть одна исследованная точка x_{n-1} , где получено пока лучшее (минимальное) значение отклика $W_{n-1} = W(x_{n-1})$ (рис. 15.5, а). Расположение точки x_{n-1} внутри L_{n-1} полностью определяется принятой исследователем стратегией поиска. Разумеется, исследователь может также выбрать положение последней точки x_n . Поскольку в данном случае речь идет о выборе положения двух точек, а заботиться о дальнейшем уже не нужно (наблюдается аналогия с вы-

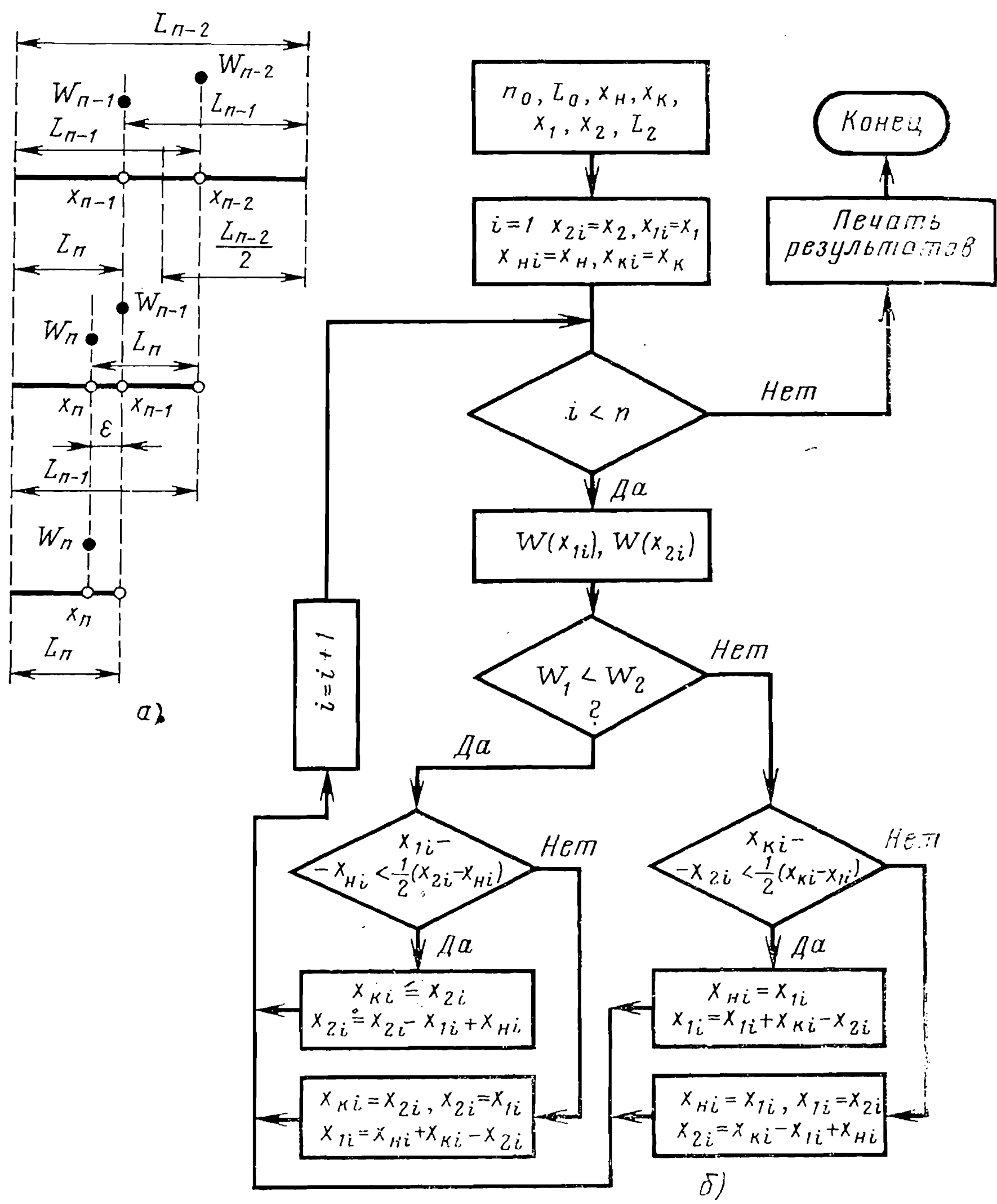


Рис. 15.5. Метод Кифера (чисел Фибоначчи):
а — схема поиска оптимума; б — структурная схема программы; x_{Hi} и x_{Ki} — начальная и конечная точки интервала неопределенности L_i

бором управления на последнем этапе при динамическом программировании, см. разд. 14.3), естественно расположить эти точки оптимальным образом симметрично середине интервала на расстоянии ε друг от друга.

Вне зависимости от того, какая из точек (x_n, x_{n-1}) окажется лучше, длины интервалов L_n и L_{n-1} связаны соотношением (рис. 15.5, а)

$$L_{n-1} = 2L_n - \varepsilon. \quad (15.20)$$

Аналогично в интервале неопределенности L_{n-2} , полученном после исследования $n-2$ точек, имеется лучшая из этих точек x_{n-2} ($W_{n-2} = W(x_{n-2})$) (рис. 15.5, а). В этом интервале необходимо разместить точку x_{n-1} , после исследования которой будет осуществлен переход $L_{n-2} \rightarrow L_{n-1}$. На рис. 15.5, а значение отклика в точке x_{n-2} больше (хуже!), чем в точке x_{n-1} , в результате чего точка x_{n-2} становится граничной точкой интервала L_{n-1} . Но ведь это *один из возможных* исходов и заранее неизвестно, в какой из точек x_{n-1} , x_{n-2} будет получен лучший результат. Величина же интервала L_{n-1} не должна зависеть от конкретного результата испытания. Поэтому расстояние от левого конца L_{n-2} до правой исследуемой точки (x_{n-2}) должно быть равно расстоянию от левой исследуемой точки (x_{n-1}) до правого конца L_{n-1} . Оба эти расстояния равны L_{n-1} . В то же время расстояние каждой из точек (x_{n-1}, x_{n-2}) до ближайшей границы интервала L_{n-2} равно L_n . Следовательно,

$$L_{n-2} = L_{n-1} + L_n = 3L_n - \varepsilon. \quad (15.21)$$

Обобщая (15.21) не любые три последовательные интервала неопределенности можно получить

$$L_{j-1} = L_j + L_{j+1}; \quad 1 < j < n. \quad (15.22)$$

Длины нескольких последних интервалов неопределенности даются соотношениями

$$\begin{aligned} L_{n-1} &= 2L_n - \varepsilon; \quad L_{n-2} = L_{n-1} + L_n = 3L_n - \varepsilon; \\ L_{n-3} &= L_{n-2} + L_{n-1} = 5L_n - 2\varepsilon; \quad L_{n-4} = L_{n-3} + L_{n-2} = 8L_n - 3\varepsilon; \\ L_{n-5} &= L_{n-4} + L_{n-3} = 13L_n - 5\varepsilon; \dots \end{aligned} \quad (15.23)$$

Коэффициенты при L_n и ε в правой части (15.23) образуют последовательность знаменитых чисел Фибоначчи F , в которой, как известно, первые два числа равны 1, а каждое следующее число является суммой двух предыдущих:

$$\begin{aligned} F_0 &= F_1 = 1; \quad F_j = F_{j-1} + F_{j-2}; \quad j = 2, 3, \dots \\ 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \dots \end{aligned} \quad (15.24)$$

С помощью чисел Фибоначчи соотношения (15.23) можно записать в общем виде:

$$L_{n-k} = F_{k+1}L_n - F_{k-1}\varepsilon. \quad (15.25)$$

При $k=n-1$ в левой части (15.25) стоит исходный интервал неопределенности $L_1=L_0$. Поэтому из (15.25) следует

$$L_0 = F_n L_n - F_{n-2} \varepsilon, \quad (15.26)$$

откуда можно получить выражение для конечного интервала неопределенности после исследования n точек:

$$L_n = \frac{L_0}{F_n} + \frac{F_{n-2}}{F_n} \varepsilon. \quad (15.27)$$

Для определения положения первой точки на интервале неопределенности (x_1) можно воспользоваться соотношением (15.25) для L_2 :

$$L_2 = F_{n-1} L_n - F_{n-3} \varepsilon. \quad (15.28)$$

Подставляя (15.27) в (15.28), можно получить

$$L_2 = \frac{F_{n-1}}{F_n} L_0 + \frac{F_{n-1} F_{n-2} - F_n F_{n-3}}{F_n} \varepsilon. \quad (15.29)$$

Доказано, что

$$F_{n-1} F_{n-2} - F_n F_{n-3} = (-1)^n, \quad (15.30)$$

и, следовательно, интервал L_2 , задающий расстояние первых двух исследуемых точек от концов исходного интервала неопределенности

$$L_2 = \frac{F_{n-1}}{F_n} L_0 + (-1)^n \frac{\varepsilon}{F_n}. \quad (15.31)$$

Структурная схема алгоритма метода с учетом переобозначения точек на каждом шаге представлена на рис. 15.5, б.

Метод Кифера, основанный на использовании чисел Фибоначчи, является оптимальным ε -минимаксным методом и обеспечивает максимально возможное сокращение интервала неопределенности. Так, сокращение исходного интервала неопределенности почти в 144 раза (при оценке пренебрегается величиной $\sim 0,4\varepsilon$) требует изучения функции отклика всего в 11 точках интервала. Однако для использования метода чисел Фибоначчи в силу (15.31) нужно *заранее* задать общее число исследуемых точек n . Это удобно, если целью исследования является сокращение интервала неопределенности, но в ряде случаев оптимизацию ДУ ЛА нужно вести до тех пор, пока какая-либо ее характеристика не достигнет заданного уровня и общее число необходимых для этого исследования точек неизвестно. В этой ситуации вместо чисел Фибоначчи может быть использовано золотое сечение¹. В этом случае, так же как и при использовании чисел Фибоначчи, предусматривается исследование на каждом (кроме первого) шаге одной точки интервала, расположенной симметрично точке, оставшейся от предыдущего шага.

¹ Напомним, что золотым сечением называется такое деление отрезка на две части, при котором отношения длин всего отрезка к большей части и большей части к меньшей равны между собой.

Однако расположение первой точки определяется существенно проще:

$$L_2 = \frac{L_0}{\tau}; \quad \tau = \frac{1 + \sqrt{5}}{2} \simeq 1,618..., \tag{15.32}$$

что как раз и соответствует делению отрезка методом золотого сечения. После n экспериментов интервал неопределенности равен

$$L_n = \frac{L_0}{\tau^{n-1}}. \tag{15.33}$$

Можно показать, что при достаточно больших n имеет место соотношение

$$\frac{F_{n-1}}{F_n} \approx \frac{1}{\tau}. \tag{15.34}$$

в силу чего при достаточно больших n конечный интервал неопределенности при использовании золотого сечения примерно равен соответствующему интервалу при использовании чисел Фибоначчи:

$$L_{n_{з.с}}/L_{n_{\phi}} \simeq 1,17. \tag{15.35}$$

Некоторые сравнительные характеристики методов поиска экстремума унимодальных моделей приведены в табл. 15.1 [20].

Т а б л и ц а 15.1

Методы	L_0/L_n					
	Число экспериментов					
	4	6	10	15	20	25
пассивного поиска	3	4	6	8	11	19
дихотомии	4	8	32	128	1024	4096
чисел Фибоначчи	5	13	89	987	10946	121393
золотого сечения	4,24	11,1	76	849	9349	103680

Легко показать, что при возрастании n отношение двух последовательных чисел Фибоначчи быстро стремится к τ (15.32):

$$\frac{F_4}{F_3} = 1,666; \quad \frac{F_5}{F_4} = 1,600; \quad \frac{F_6}{F_5} = 1,625; \quad \frac{F_7}{F_6} = 1,615...$$

Поэтому при достаточно больших n соотношение (15.32) можно использовать для задания начальных точек при поиске методом, основанным на числах Фибоначчи.

При практическом проведении поиска необходимо помнить, что точность оптимизации ограничена условием разрешимости, запрещающим сближать рассматриваемые точки ближе чем на ϵ (см. разд. 15.1). Так, если ϵ составляет 1% исходного интервала неопределенности, при использовании чисел Фибоначчи бессмысленно проводить исследование более чем в 10 точках (см. табл. 15.1).

1. Приведите описание минимаксных стратегий пассивного эксперимента.
2. Дайте развернутое описание методов дихотомии и алгоритмов Кифера, основанных на использовании чисел Фибоначчи и золотого сечения.
3. Запишите основные математические зависимости для метода дихотомии и метода чисел Фибоначчи.
4. За счет чего достигается повышение эффективности при использовании чисел Фибоначчи по сравнению с методом дихотомии?
5. Чем ограничено целесообразное число исследуемых точек при последовательных методах оптимизации?

15.3 МЕТОДЫ ОПТИМИЗАЦИИ ПОЛИМОДАЛЬНЫХ МОДЕЛЕЙ

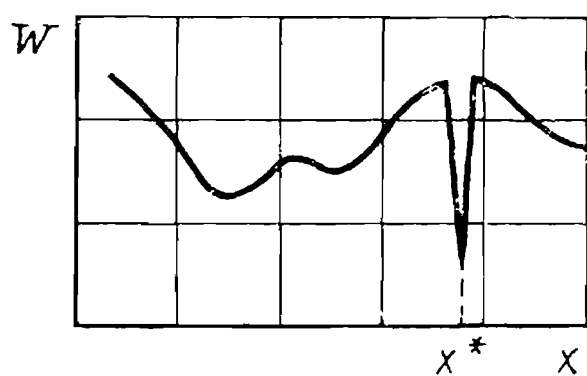
В предыдущих разделах рассмотрены методы поиска экстремума для *унимодальных* моделей сложных систем. Однако на практике при изучении ДЛА иногда встречаются *полимодальные* модели, содержащие несколько экстремумов. Появление нескольких экстремумов на поверхности отклика в ряде случаев связано с дискретным характером изменения варьируемых переменных. Изучение полимодальных моделей значительно сложнее, чем унимодальных. Рассмотренные в предыдущем разделе методы позволяют обнаружить один из локальных экстремумов модели, но совершенно не гарантируют выделение самого лучшего, *глобального* экстремума. Более того, для общего случая полимодальной модели вообще не существует методов, гарантирующих выделение глобального экстремума, так как даже при самом подробном изучении поверхности отклика всегда существует риск пропуска глобального экстремума с малой областью притяжения (рис. 15.6, а). К счастью, такие патологические функции качества на практике не встречаются.

Методы исследования полимодальных моделей всегда в той или иной форме предусматривают исследование всей области существования L_0 . В первую очередь, это методы *сканирования* или перебора, иногда называемые методами слепого поиска. В наиболее простом варианте на интервал неопределенности накладывается некоторая (достаточно частая) сетка (рис. 15.6, б), в узлах которой определяются значения критерия оптимальности. Лучшая из этих точек принимается за точку оптимума. При шаге Δ перебор задается правилом

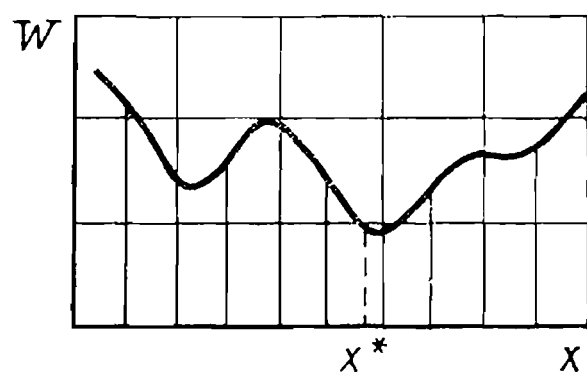
$$x_{i+1} = x_i + \Delta \Rightarrow x_0, x_0 + \Delta, x_0 + 2\Delta, \dots, x_0 + n\Delta \quad (15.36)$$

и заканчивается при $x_{i+1} = x_0 + L_0$. При этом в памяти ЭВМ сохраняется лучшее значение координаты x^* и критерия оптимальности W^* в соответствии с правилом

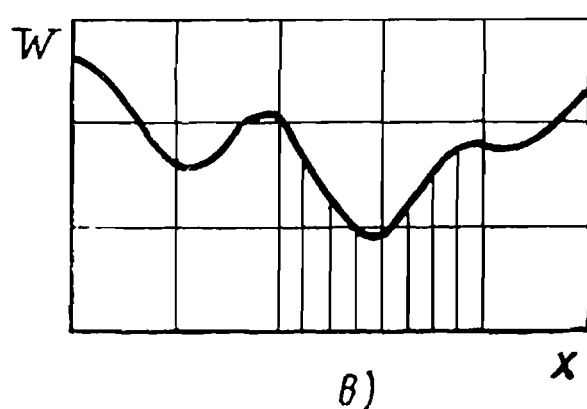
$$x_{i+1}^* = \begin{cases} x_i^*, & W(x_{i+1}) > W(x_i^*) \\ x_{i+1}, & W(x_{i+1}) < W(x_i^*) \end{cases};$$



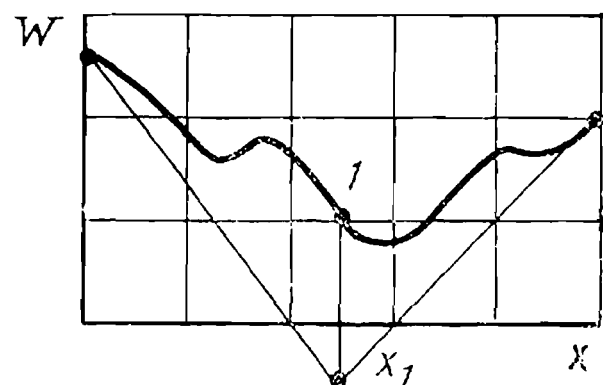
а)



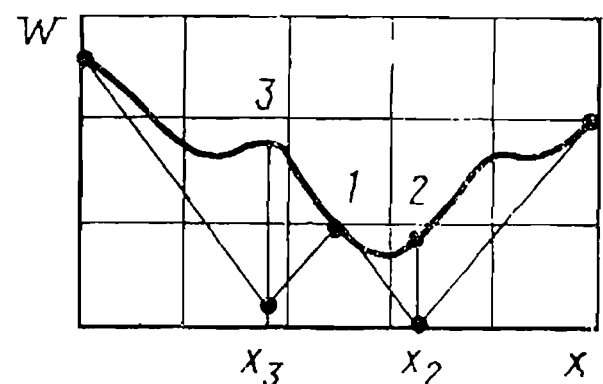
б)



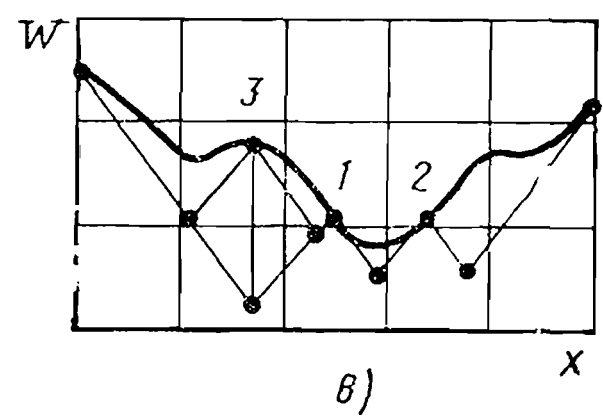
в)



а)



б)



в)

Рис. 15.6. Полимодальные поверхности отклика:

а — глобальный минимум с малой зоной притяжения; б — поиск глобального минимума методом сканирования; в — последовательное сканирование с уменьшением шага

Рис. 15.7. Метод ломаных:

а — первый шаг; б — второй шаг; в — третий шаг

$$W_{i+1}^* = \begin{cases} W_i^*, & W(x_{i+1}) > W_i^* \\ W(x_{i+1}), & W(x_{i+1}) < W_i^* \end{cases} \quad (15.37)$$

На первом шаге полагают $x_{i+1}^* = x_0$, $W_0^* = W(x_0)$.

Может быть организовано и случайное сканирование, при котором очередное значение x_i определяется с помощью датчика равномерно распределенных на отрезке $[0, 1]$ случайных чисел ξ_{pi} :

$$x_i = x_0 + \xi_{pi} L_0. \quad (15.38)$$

Количество точек случайного сканирования n может быть произвольным.

Возможны различные модификации метода сканирования, опирающиеся на дополнительную информацию о функции отклика. Например, если известно, что зоны притяжения экстремумов (по крайней мере глобального экстремума) достаточно велики, можно использовать двухступенчатое (многоступенчатое) сканирование, при котором первоначально рассматривается грубая сетка с большим шагом Δ , затем в окрестности лучшей точки накладывается более частая сетка и т. д. (рис. 15.6, в).

Случайный (статистический) аналог ступенчатого сканирования предусматривает n_1 случайных попыток на интервале L_0 ; n_2 случайных попыток на суженном (например, в пять или десять раз) интервале в окрестности лучшей из n_1 точек и т. д. Могут быть предложены также различные комбинации сканирования для выделения окрестности глобального экстремума и любого из рассмотренных в разд. 5.2 методов для нахождения собственно экстремальной точки.

Эффективный и очень изящный метод исследования предложен¹ для частного класса полимодальных моделей, отвечающих дополнительному условию ограниченности темпа изменения W .

Математически это условие записывается как условие Липшица:

$$|W(x_i) - W(x_j)| < K |x_i - x_j|; \quad \forall x_i, x_j \in L_0; \quad (15.39)$$

т. е. условие ограниченности производной (K — некоторая постоянная). Суть метода, иногда называемого *методом ломаных*, состоит в построении ломаной линии, *мажорирующей* поверхность отклика снизу и непрерывно приближающейся к ней.

Предусматривается проведение следующих операций (рис. 15.7):

1. Определяются значения функции качества в концах отрезка L_0

$$W(x_0); \quad W(x_0 + L_0).$$

2. Из точек $(x_0, W(x_0))$ и $(x_0 + L_0, W(x_0 + L_0))$ строятся лучи с наклоном, характеризуемым величиной K :

$$g_1(x) = W(x_0) - K(x - x_0); \quad g_2(x) = W(x_0 + L_0) - K(x_0 + L_0 - x). \quad (15.40)$$

В силу условий (15.39) оба эти луча целиком лежат *под* поверхностью отклика (так как имеют наклон, больший наклона поверхности отклика в любой точке изучаемого интервала).

3. Определяется точка пересечения лучей x_1

$$x_1 = x_0 + \frac{1}{2} L_0 + \frac{1}{2K} [W(x_0) - W(x_0 + L_0)], \quad (15.41)$$

и величина $W(x_1)$.

4. Из точки $(x_1, W(x_1))$ вновь строятся лучи с наклоном K , также в силу (15.39) целиком лежащие под поверхностью отклика:

$$g_3 = W(x_1) - K(x - x_1); \quad g_4 = W(x_1) - K(x_1 - x), \quad (15.42)$$

и определяются точки пересечения x_2, x_3 . При этом двухзвенная ломаная заменяется четырехзвенной. Для более «глубокой» из точек x_2, x_3 процедура повторяется. Дальнейший ход рассуждений ясен из рис. 15.7. На каждом шаге число звеньев и вершин ломаной увеличивается, для дальнейшего разбиения выбирается самая

¹ Пиавский С. А. Один алгоритм отыскания абсолютного экстремума функции. — Журнал вычислительной математики и математической физики. 1972, 12 № 4, 8 с.

«глубокая» вершина ломаной, а сама ломаная непрерывно приближается к поверхности отклика, мажорируя ее снизу. После достаточного числа шагов минимум ломаной достаточно хорошо совпадает с глобальным минимумом изучаемой модели.

На практике все реальные модели ДУ ЛА (кроме некоторых моделей процессов управления) отвечают условиям (15.39), однако величина константы K , как правило, неизвестна. Конечно, можно застраховать себя от ошибки, выбрав константу K достаточно большой. При этом, однако, резко возрастает количество необходимых шагов и трудоемкость метода. При уменьшении K возрастает риск нарушения условия (15.39) и потери глобального экстремума. Обычно величина K подбирается экспериментально методом проб и ошибок.

Вопросы для самоконтроля

1. Дайте характеристику метода сканирования.
2. Дайте характеристику метода ломаных.

ГЛАВА 16. ПРЯМЫЕ МЕТОДЫ ОПТИМИЗАЦИИ ДЛЯ (МНОГОМЕРНЫЙ ПОИСК)

16.1. «ПРОКЛЯТИЕ» РАЗМЕРНОСТИ

Реальные модели оптимального проектирования ДУ, ДЛА и их элементов почти всегда многомерны. При их исследовании возникает задача изучения многомерного факторного пространства, многомерной поверхности отклика. Изучение многомерных задач сопряжено с рядом трудностей, не встречающихся при одномерном поиске. Эти трудности резко нарастают с увеличением числа факторов и столь велики, что американский математик и оптимизатор Р. Беллман называл их «проклятием» размерности. Можно выделить три проблемы, порождаемые многомерностью.

Во-первых, возрастание числа переменных делает менее вероятной унимодальность поверхности отклика и часто приводит к необходимости изучения полимодальных функций.

Во-вторых, для многомерного случая не удастся найти универсальную, не зависящую от удачи исследователя меру эффективности поиска, аналогичную принципу минимакса.

В-третьих, при переходе к многомерному пространству уменьшается относительная эффективность поиска.

Последнее положение может быть проиллюстрировано рис. 16.1, а. Для одномерного случая (верхнее переднее ребро куба) десятикратное сокращение интервала неопределенности приводит к такому же уточнению координаты экстремума x_1 . В двумерном случае (передняя грань куба) аналогичное сокращение интервала неопределенности (площади грани) приводит к чуть более чем трехкратному сокращению интервала неопределенности по каждой

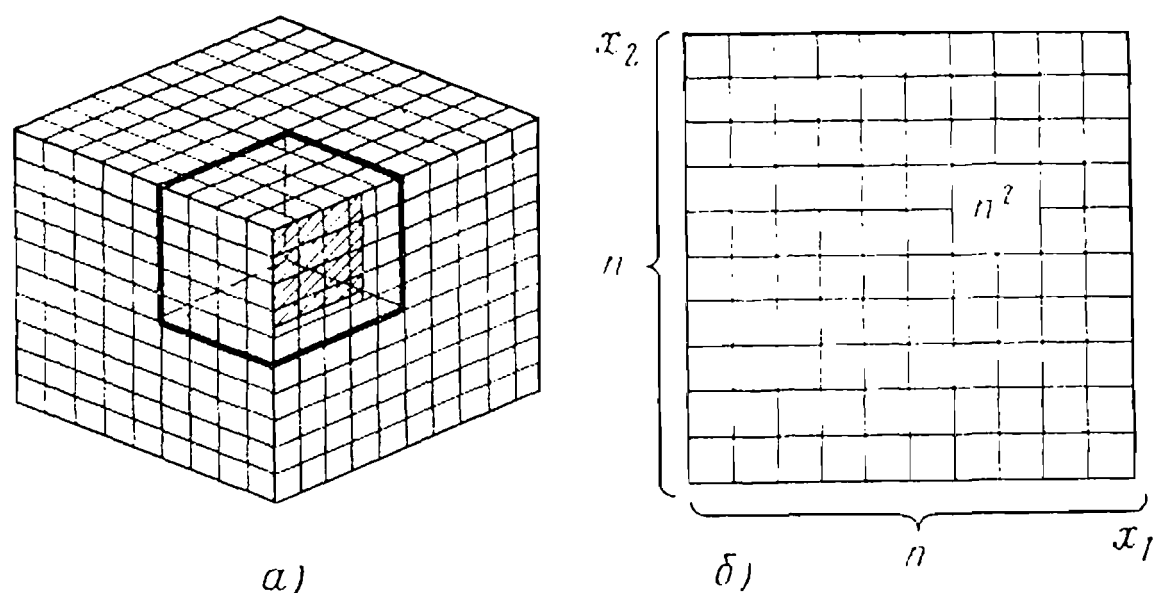


Рис. 16.1. Зависимость затрат на поиск от размерности задачи:

а — уменьшение интервала неопределенности при одинаковом количестве расчетов; *б* — зависимость числа расчетов от размерности задачи при постоянном уменьшении интервала неопределенности по каждой переменной

из переменных x_1, x_2 . Наконец, для трехмерной задачи (куб) интервал неопределенности для переменных x_1, x_2, x_3 в этом случае уменьшится чуть более чем в два раза. В случае 50-ти переменных интервалы неопределенности по каждой из них составят при 10-кратном сокращении объема 0,93 величины исходных интервалов.

Можно подойти к этой проблеме с другой стороны. При использовании пассивной стратегии поиска для 10-кратного сокращения одномерного интервала неопределенности необходимо 19 расчетов функции качества (см. разд. 15.2). Для двумерной задачи аналогичный результат для каждой из переменных потребует проведения $19^2 = 361$ расчетов (рис. 16.1, б). В случае же десяти переменных требуется определить $19^{10} \approx 6 \cdot 10^{12}$ значений функции качества. Даже если предположить, что для однократного исследования модели необходима только 1 секунда работы ЭВМ (а для реальных моделей ДЛА это весьма оптимистическая оценка!), решение задачи потребует около двухсот тысяч лет. А ведь существуют задачи с несколькими десятками и сотнями переменных.

Из приведенных рассуждений, в частности, следует бесперспективность использования пассивных методов при решении многомерных ($k \geq 4$) задач. Далее будут рассмотрены только последовательные методы оптимизации.

Существует множество различных методов многомерного прямого поиска. Многие из них воплощены в стандартных программах, вошедших в справочники и библиотеки стандартных программ современных ЭВМ. В настоящей главе рассмотрены лишь некоторые наиболее типичные из этих методов, разбитые на группы в соответствии с классификацией, приведенной на рис. 13.4.

При разработке процедуры многомерного поиска всегда возникает три вопроса:

1. Откуда (из какой точки $x^0 \in X^*$) нужно начинать поиск?
2. В каком направлении необходимо двигаться в факторном пространстве $x^n \rightarrow x^{n+1}$?
3. Когда необходимо прекратить поиск?

Выбор начальной точки более или менее универсален для всех рассмотренных ниже методов оптимизации. Эта точка обычно задается в центре области X^* . Наряду с детерминированно заданными в ряде алгоритмов используются случайные начальные точки.

Выбор направления движения в факторном пространстве существенно зависит от конкретного алгоритма оптимизации. Этот выбор может опираться на информацию о локальном (в точке или ее окрестности) изменении функции качества (*локальные методы*) или учитывать предысторию поиска (*глобальные методы*), быть точным (*детерминированные методы*) или случайным (*статистические методы*).

16.2. ДЕТЕРМИНИРОВАННЫЕ ЛОКАЛЬНЫЕ МЕТОДЫ ПОИСКА ЭКСТРЕМУМА С ЛИНЕЙНОЙ ТАКТИКОЙ

Данная группа методов оптимизации предназначена для изучения унимодальных многофакторных моделей и включает в себя ряд алгоритмов, предусматривающих последовательный детерминированный перебор точек факторного пространства на основе информации, полученной в последней изученной точке при линейной аппроксимации поверхности отклика. Детерминированные локальные методы поиска с линейной тактикой при определении направления движения обычно используют информацию о градиенте поверхности отклика, но так как аналитические выражения производных неизвестны, составляющие градиента определяются численно с помощью пробных шагов по каждой из координат (рис. 16.2, а):

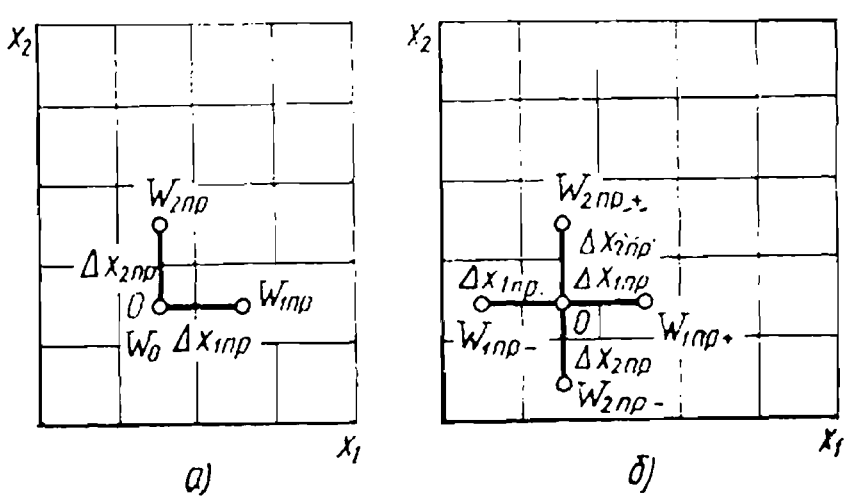
$$\text{grad } W = \sum_{i=1}^k \frac{\partial W}{\partial x_i} k_i \simeq \sum_{i=1}^k \frac{W_{i\text{np}} - W_0}{\Delta x_{i\text{np}}} k_i. \quad (16.1)$$

Определение направления градиента по формуле (16.1) для k -мерной поверхности отклика требует $k+1$ вычисления функции качества. Более точно определяется градиент при использовании *парных* проб, расположенных по обе стороны от исследуемой точки (рис. 16.2, б). В этом случае

$$\text{grad } W = \sum_{i=1}^k \frac{\partial W}{\partial x_i} k_i \simeq \sum_{i=1}^k \frac{W_{i\text{np}+} - W_{i\text{np}-}}{2\Delta x_{i\text{np}}} k_i. \quad (16.1a)$$

Для определения градиента этим способом необходимо провести $2k$ вычислений функции качества. При минимизации W движение должно осуществляться против градиента, т. е. в *антиградиентном* направлении.

Рис. 16.2. Определение направления градиента поверхности отклика:
а — при однократной пробе; б — при парной пробе



Правило окончания поиска (правило остановки) также зависит от конкретного алгоритма оптимизации. Часто используется одно из двух условий:

— уменьшение до определенной величины темпа изменения критерия оптимальности (изменения критерия оптимальности за один шаг):

$$|W^{n+1} - W^n| < \varepsilon_1 \quad (16.2)$$

— уменьшение до наперед заданной величины шага поиска

$$|\mathbf{x}^{n+1} - \mathbf{x}^n| < \varepsilon_2 \Rightarrow |x_i^{n+1} - x_i^n| < \varepsilon_2. \quad (16.3)$$

Иногда требуется одновременное выполнение условий (16.2), (16.3):

$$|W^{n+1} - W^n| < \varepsilon_1 \wedge |\mathbf{x}^{n+1} - \mathbf{x}^n| < \varepsilon_2. \quad (16.4)$$

Однако нужно иметь в виду, что условия (16.2) — (16.4) могут выполняться не только в окрестности экстремума, но и при некоторых других условиях: на пологом склоне, в овражной ситуации и т. д.

Поэтому условия (16.2) — (16.4) нельзя считать исчерпывающими, и в ряде алгоритмов оптимизации предусматривается дополнительное исследование точек факторного пространства, отвечающих этим условиям.

Методы отсечений предусматривают исследование поверхности отклика в одной из точек изучаемой области $\mathbf{X}^*$ и *исключение* из дальнейшего рассмотрения части этой области, признанной бесперспективной с точки зрения поиска экстремума. Структурная схема и схема работы простейшего из методов отсечений представлены на рис. 16.3, а, б. Метод предусматривает:

1) численное определение направления градиента (и антиградиента) в некоторой (начальной) точке факторного пространства $\mathbf{x}^1 \in \mathbf{X}_0^*$;

2) построение прямой (гиперплоскости), ортогональной направлению градиента. В исследованной точке эта прямая (плоскость) касательна к линии (поверхности) уровня (рис. 16.3, б);

3) исключение из дальнейшего рассмотрения области, где функция качества возрастает (в случае минимизации), как бесперспективной с точки зрения поиска экстремума.

Для выделенной после отсечения области $\mathbf{X}_1^*$ все вычисления повторяются. После ряда отсечений остается достаточно малая область (интервал неопределенности) $\mathbf{X}_n^*$, содержащая экстремум.

Метод отсечений внешне очень напоминает одномерные методы оптимизации, основанные на принципе минимакса (например, метод дихотомии), однако в данном случае нельзя заранее определить размеры конечной области неопределенности. Высока эффективность метода. Если допустить, что при каждой операции отсечения исключается в среднем $1/2$ рассматриваемой области, то

$$\mathbf{X}_n^* \simeq \frac{1}{2^n} \mathbf{X}_0^*; \quad N = n(k + 1), \quad (16.5)$$

где N — общее число вычислений функции качества.

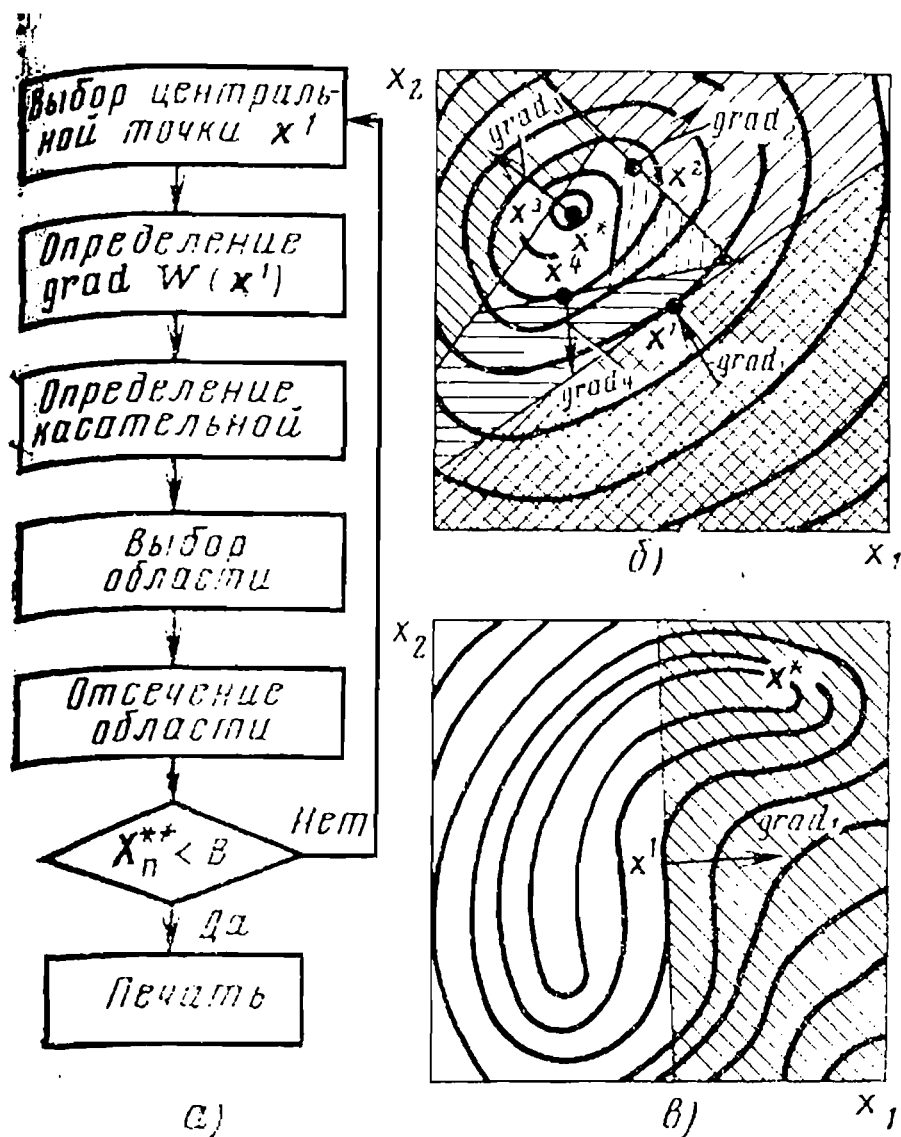


Рис. 16.3. Метод отсечений:

а — структурная схема программы поиска;
б — схема поиска оптимума; *в* — схема поиска оптимума при наличии на поверхности отклика криволинейного оврага

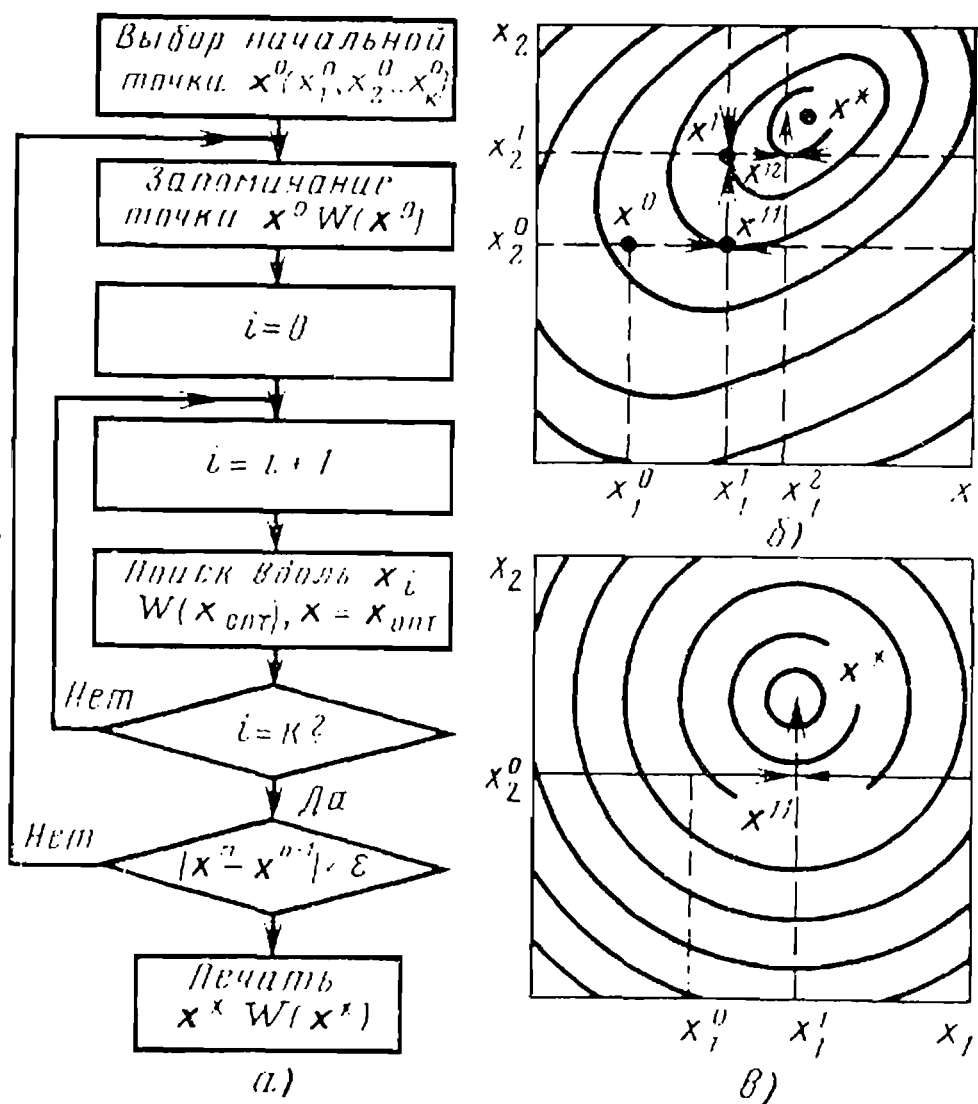


Рис. 16.4. Метод покоординатного спуска:

а — структурная схема программы поиска;
б — схема поиска оптимума; *в* — схема поиска оптимума для сферической поверхности отклика

Возможны различные модификации метода отсечений.

Методы отсечений, однако, пригодны только для изучения строго унимодальных моделей (см. разд. 1.4). В более общем случае при отсечении может быть выброшена область, содержащая экстремум (рис. 16.3, *в*). Этот эффект является результатом слишком *сильных* решений (отбрасывание целой области), принимаемых на основе *локальной* информации о характере поверхности отклика в одной точке. Поэтому на практике методы отсечений можно использовать лишь в том случае, когда имеются достаточные основания считать модель строго унимодальной.

Методы покоординатного спуска (иногда их называют релаксационными методами) являются естественным приложением методов одномерного поиска к исследованию многомерных моделей. Эти методы предусматривают последовательную циклическую оптимизацию по каждой из варьируемых переменных x_i .

Структурная схема программы и схема поиска для одного из таких методов показаны на рис. 16.4, *а*, *б*. Для заданной начальной точки $x^0(x_1^0, x_2^0, \dots, x_k^0)$ фиксируются все переменные, кроме первой, и осуществляется одномерный поиск оптимального решения вдоль координаты x_1 . В результате поиска определяется точка $x^{11}(x_1^1, x_2^0, \dots, x_k^0)$. Затем фиксируются все переменные, кроме второй, причем x_1 фиксируется на полученном оптимальном уровне и определяется точка $x^{12}(x_1^1, x_2^1, x_3^0, \dots, x_k^0)$. После последователь-

ной оптимизации по всем переменным определяется точка $x^1(x_1^1, x_2^1 \dots x_k^1)$ и осуществляется второй цикл оптимизации

$$(x_1^2, x_2^1 \dots x_k^1) \rightarrow (x_1^2, x_2^2 \dots x_k^1) \rightarrow (x_1^2, x_2^2 \dots x_k^2). \tag{16.6}$$

Поиск завершается при условии, что очередной цикл оптимизации не дает существенного улучшения результатов (16.2); (16.3); (16.4).

Могут быть предложены различные модификации метода, например, предусматривающие случайный выбор очередной исследуемой координаты.

Методы покоординатного спуска более гибки, чем методы отсечений, так как не требуется отбрасывание части области. Однако эффективность этих методов существенно зависит от характера поверхности отклика. В идеальном случае сферической (гиперсферической) поверхности отклика (рис. 16.4, в) покоординатный спуск приводит к окрестности экстремума за один цикл. Хуже обстоит дело при сложных поверхностях отклика. На рис. 16.5 представлены ситуации, при которых даже для двумерных унимодальных моделей покоординатный спуск не позволяет приблизиться к оптимуму. В первом случае (рис. 16.5, а) рассмотрена овражная ситуация с оврагом, вытянутым вдоль биссектрисы угла $x_1 \wedge x_2$. Легко видеть, что уже для первой точки, лежащей в окрестности дна оврага, возникает «неулучшаемая» ситуация, воспринимаемая алгоритмом как окончательный экстремум, хотя истинная точка оптимума x^* может лежать существенно ниже по оврагу. Во втором случае (рис. 16.5, б) трудности возникают в результате введения ограничений. Видно, что в зависимости от выбора очередности шагов будут получены две *различные* оптимальные точки, причем обе они достаточно далеки от истинного оптимума.

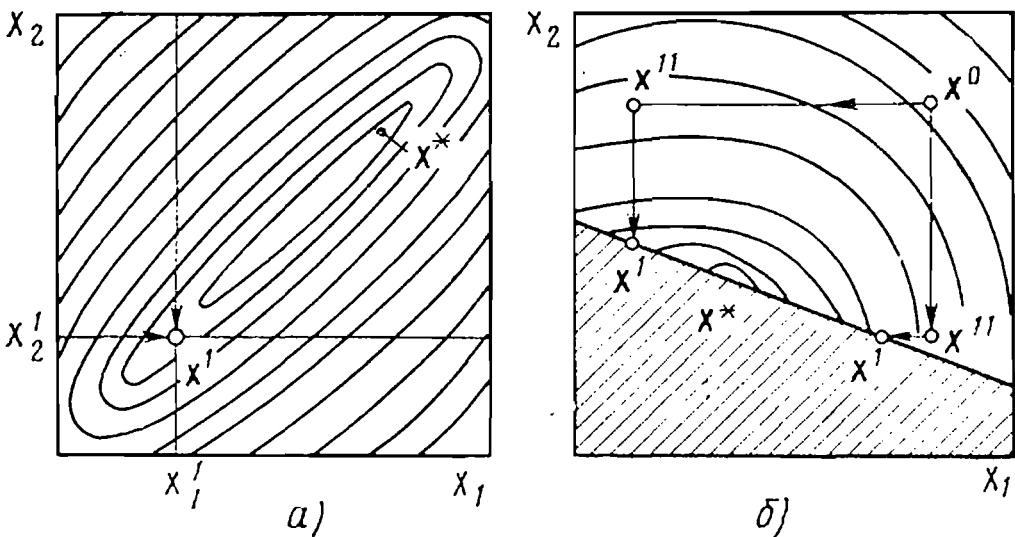
Главным недостатком метода покоординатного спуска является жесткая детерминированность направлений движения, совпадающих с координатами факторного пространства. Уменьшить влияние этого недостатка можно с помощью перехода к новым искусственно введенным координатам. Действительно, за счет выбора масштабов переменных можно добиться ситуации, при которой линии уровня модели по крайней мере в некоторой области поверхности отклика, будут близки к сферическим. Однако такие преобразования сложны и требуют значительного количества информации.

Кроме того, эти преобразования, строго говоря, нужно повторять при каждом очередном шаге поиска.

Градиентные методы.
К градиентным относится

Рис. 16.5. Метод покоординатного спуска:

а — в овражной ситуации; б — при расположении оптимума на границе области



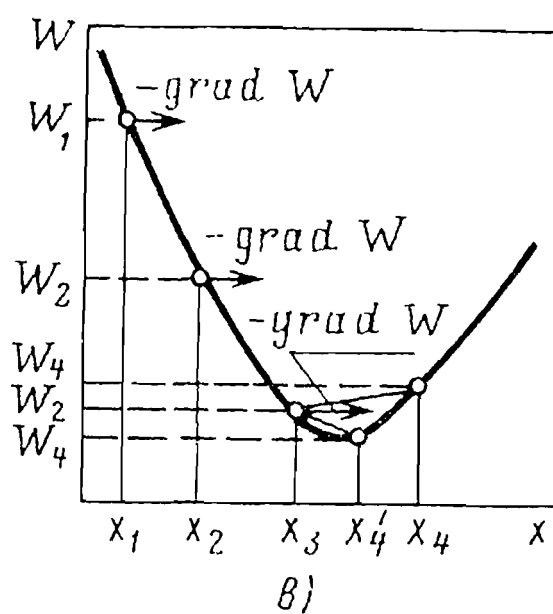
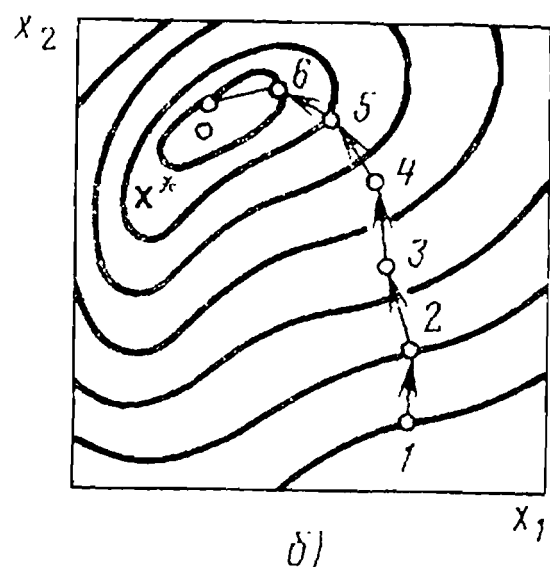
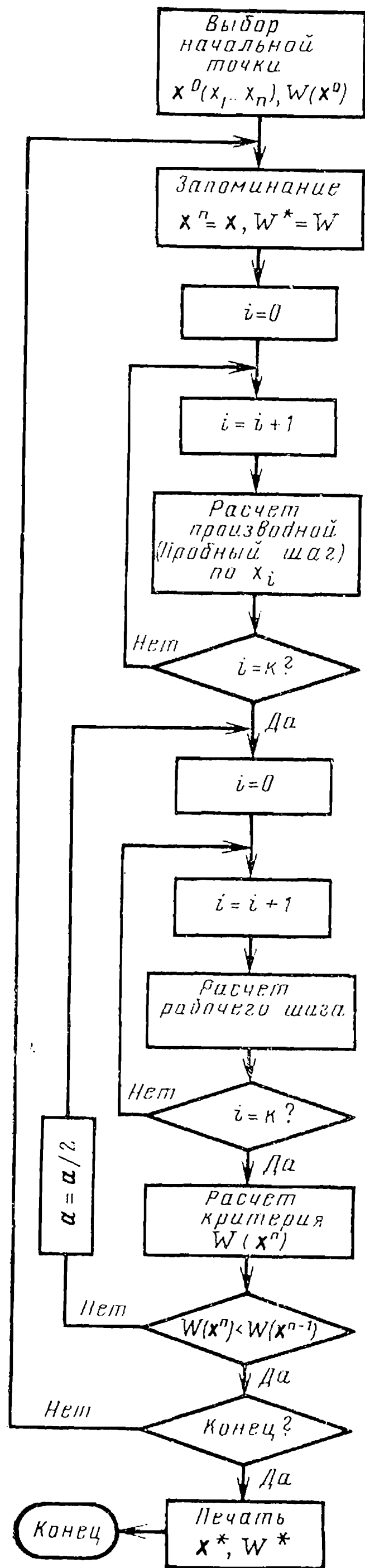


Рис. 16.6. Градиентный метод поиска:
 а — структурная схема программы поиска;
 б — схема поиска оптимума; в — уменьшение шага поиска в окрестности экстремума

большая группа методов. Структурная схема программы и схема поиска оптимума простейшим градиентным методом показаны на рис. 16.6. Метод предусматривает:

- 1) определение в исследуемой точке $\mathbf{x}^n \in X^*$ направления антиградиента;
- 2) выполнение по этому направлению рабочего шага $\mathbf{x}^n \rightarrow \mathbf{x}^{n+1}$;
- 3) определение в новой точке функции качества $W(\mathbf{x}^{n+1})$; проверку условия $W(\mathbf{x}^{n+1}) < W(\mathbf{x}^n)$;
- 4) повторение операций 1, 2 для новой точки $\mathbf{x}^{n+1}$.

Рабочий шаг определяется соотношением

$$x_i^{n+1} = x_i^n - a_i \frac{\partial W^n}{\partial x_i}, \quad (16.7)$$

где a_i — модуль шага по i -й координате. Первоначальная величина a_i выбирается как доля интервала неопределенности по каждой переменной, например,

$$a_i = \frac{x_{i\max} - x_{i\min}}{10}. \quad (16.8)$$

Движение в антиградиентном направлении *по определению* должно приводить к улучшению функции качества. Если это не так $W(\mathbf{x}^{n+1}) > W(\mathbf{x}^n)$, можно предположить, что поиск просто «проскочил» оптимальную точку. В этом случае следует уменьшить модуль шага (обычно в два раза $a_i \Rightarrow \frac{1}{2} a_i$) и повторить вычисления (см. рис. 16.6, в). Уменьшение модуля шага до заданной величины (например, в 16 раз) может служить основанием для прекращения поиска. На каждом шаге градиентного спуска необходимо провести $k+1$ (в случае уменьшения модуля шага $k+2$) вычислений функции качества.

Существует много улучшенных модификаций метода градиентного поиска. Наиболее популярной из них является так называемый **метод наискорейшего спуска**, предложенный Л. В. Канторовичем. Структурная схема программы и схема поиска оптимума для этого метода показаны на рис. 16.7.

Идея метода заключается в отказе от фиксированной величины рабочего шага. После определения направления антиградиента осуществляется одномерный поиск по этому направлению и определяется лучшая точка $\mathbf{x}$ на заданной прямой. В этой точке повторяется операция определения направления антиградиента и осуществляется новый спуск. Очевидна аналогия методов наискорейшего и покоординатного спусков. Существенно, что при наискорейшем спуске направления одномерного поиска не связаны с фиксированными координатами факторного пространства, что сообщает поиску большую гибкость.

Эффективность градиентных методов зависит от удачи выбора начальной точки поиска (рис. 16.7, в). Спуск из точки 1 сразу приводит к оптимуму. Спуск из точки 2 состоит из ряда последователь-

Выбор
начальной
точки
 $x^0(x_1, x_2, \dots, x_n), W(x^0)$

Запоминание
 $x^* = x \quad W^* = W$

$i = 0$

$i = i + 1$

Расчет
производной
(Пробный шаг
по x_i)

Нет
 $i = k?$

Да

Построение
прямой спуска
в фазовом
пространстве

Поиск миниму-
ма вдоль прямой
одномерным ме-
тодом x^n

Расчет
критерия
 $W(x^n)$

Нет
Конец?

Да

Печать
 $x^* \quad W^*$

Конец

$a,$

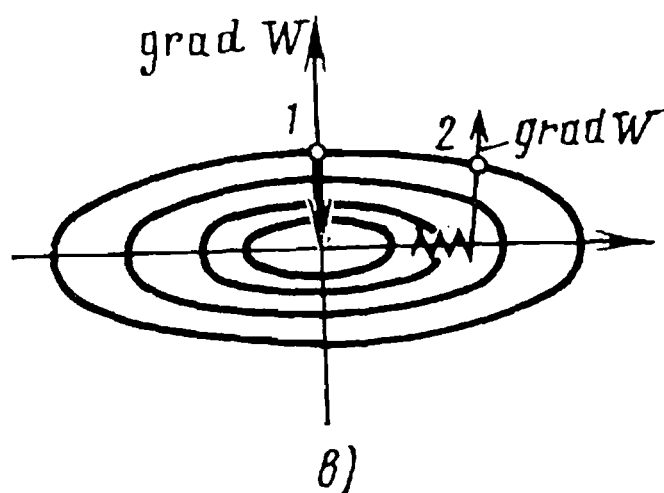
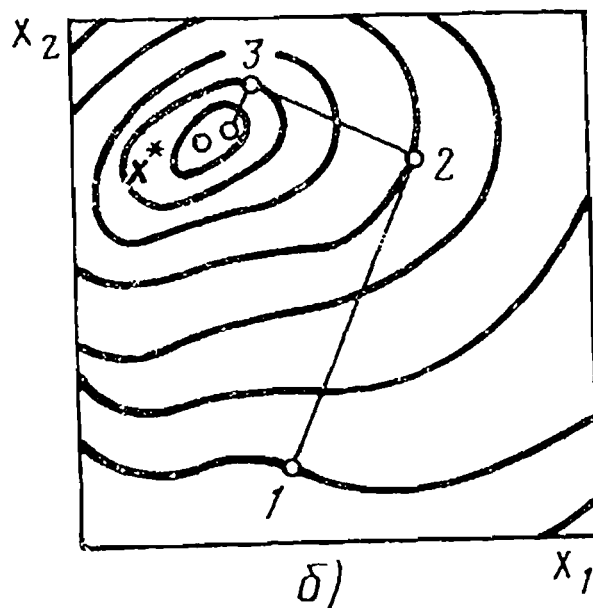


Рис. 16.7. Наискорейший спуск:
а — структурная схема программы
поиска; б — схема поиска миниму-
ма; в — схема поиска оптимума в
случае эллиптической поверхности
отклика

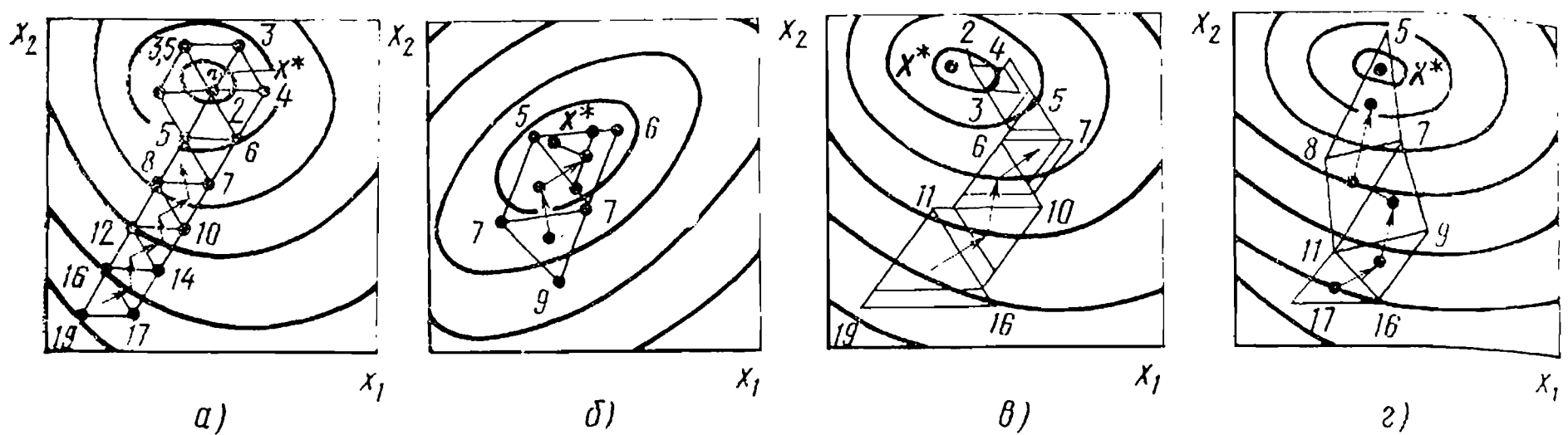


Рис. 16.8. Поиск по многограннику:

а — схема поиска оптимума; *б* — уменьшение шага поиска в окрестности экстремума; *в* — поиск при переменном ребре многогранника; *г* — поиск по деформируемому многограннику

ных шагов и, как будет показано в разд. 16.4, скорее всего закончится достаточно далеко от оптимума.

К классу локальных методов с линейной тактикой относится также получивший широкое распространение в последние годы **метод поиска по равностороннему многограннику**. Схема поиска оптимума для этого метода показана на рис. 16.8. Метод предусматривает последовательное изучение функции качества в вершинах $k+1$ -мерного многогранника ($\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^{k+1} \in X^*$), рис. 16.8, *а*. Поиск, как всегда, начинается в центральной части области X^* . На первом этапе определяются и сравниваются между собой значения функции качества в вершинах многогранника (на плоскости это треугольник) $W(\mathbf{x}^1), W(\mathbf{x}^2), W(\mathbf{x}^3)$. Точка, соответствующая худшему значению функции качества $\mathbf{x}^3$, отбрасывается и заменяется отраженной через противоположную грань точкой $\mathbf{x}^4$. Для нового многогранника ($\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^4$) операция повторяется, причем на всех этапах, кроме первого, необходимо определять значение функции качества только в одной точке.

Если отраженная вершина в новом многограннике оказывается наихудшей, т. е. должно произойти обратное отражение, нужно вернуться к исходному многограннику и произвести отражение вершины со вторым, наихудшим значением функции качества. Вращение многогранника вокруг какой-либо точки может служить основанием для прекращения поиска, так как указывает на достижение окрестности экстремума. Для уточнения положения экстремума в этом случае целесообразно уменьшить величину многогранника (рис. 16.8, *б*), сохраняя лучшую из имеющихся вершин. Величина последнего многогранника характеризует область неопределенности, содержащую экстремальную точку.

Достоинствами метода являются простота алгоритма и хорошая адаптация к поверхности отклика.

Существует ряд модификаций поиска по многограннику. Для ускорения сходимости используется поиск с переменным шагом (рис. 16.8, *в*), предусматривающий сокращение размеров многогранника по мере приближения к экстремуму. Метод обеспечивает быстрое перемещение к экстремуму на начальном этапе и достаточно точное определение точки экстремума на завершающем этапе

поиска. При сокращении размеров сохраняется лучшая из исследованных вершин. Для сокращения длины ребра используют одну из следующих зависимостей:

$$L_n = \gamma_n L_0; \quad \gamma_n = \frac{1}{1 + dn}; \quad \gamma_n = e^{-dn}, \quad (16.9)$$

где d — положительное число.

В зависимости от соотношения значений функции качества в вершинах может изменяться форма многогранника (рис. 16.8, з). При этом он может вытягиваться или сокращаться, что обеспечивает увеличение гибкости поиска.

Вопросы для самоконтроля

1. Дайте развернутое описание метода отсечений, методов покоординатного спуска, градиентных методов и поиска по многограннику.
2. Что общего в методах покоординатного и наискорейшего спуска?

16.3. СТАТИСТИЧЕСКИЕ ЛОКАЛЬНЫЕ МЕТОДЫ ПОИСКА ЭКСТРЕМУМА

Наряду с детерминированными широко используются так называемые *статистические* методы поиска экстремума, опирающиеся на случайный выбор изучаемых точек факторного пространства. Все методы этого типа можно разбить на две большие группы: методы *случайного наброса* и методы *случайного спуска*. Методы случайного наброса всегда предусматривают параллельное исследование поверхности отклика в некотором множестве случайных точек факторного пространства. Одна из первых стратегий этого типа, разработанная С. Бруксом, предусматривает последовательное проведение нескольких серий расчетов (или экспериментов). При этом на первом этапе исследуется функция отклика в m случайных точках первоначальной области неопределенности. Координаты этих точек определяются с помощью датчика случайных чисел ξ_r (в k -мерном пространстве необходимо $k \times m$ случайных чисел):

$$x_i^j = x_{i \min} + \xi_{ri}^j (x_{i \max} - x_{i \min}); \quad (16.10)$$

$$i = 1, 2, \dots, k, \quad j = 1, 2, \dots, m.$$

Среди этих m точек выбирается лучшая ($W(\mathbf{x}^r) < W(\mathbf{x}^j)$, $r, j = 1, 2, \dots, m, j \neq r$) и в ее окрестности (например, можно потребовать сокращения интервала неопределенности в два раза) ставится новая серия расчетов (экспериментов), вновь предусматривающая изучение m случайных точек. Процедура повторяется несколько раз вплоть до желательного сокращения области неопределенности. Структурная схема алгоритма и схема поиска оптимума методом случайного наброса показана на рис. 16.9.

Методы случайного наброса просты и наглядны. Однако их эффективность при исследовании многомерных моделей невелика.

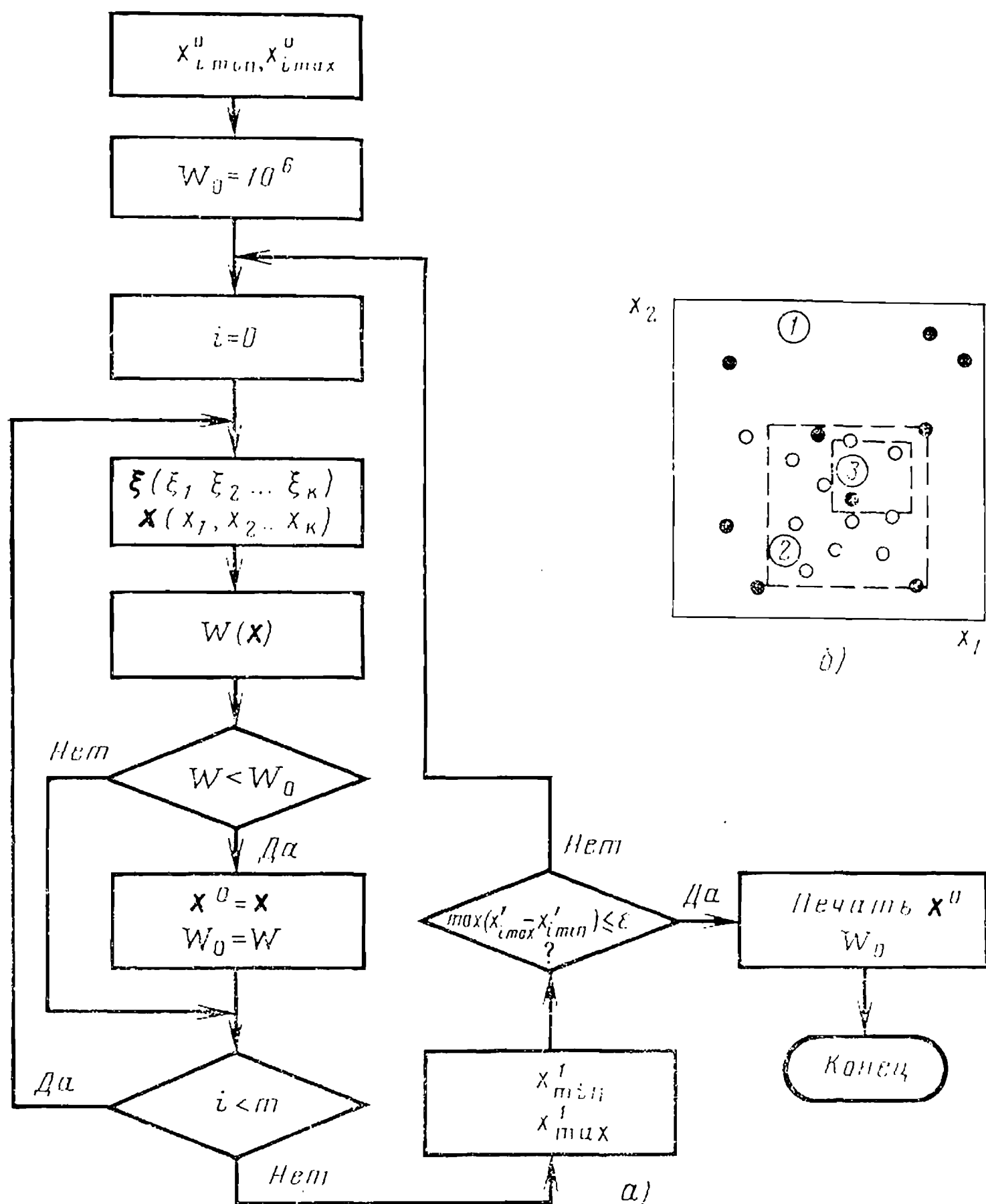


Рис. 16.9. Случайный «набросовый» поиск:
 а — структурная схема программы поиска; б — схема поиска оптимума

Дело в том, что значительная часть исследуемых точек (при алгоритме Брукса все точки каждой из серий) факторного пространства выбирается на основе пассивной стратегии, что в случае многофакторных моделей приводит к быстрому росту затрат на поиск.

Более эффективны методы случайного спуска [18], являющиеся статистическими аналогами детерминированных методов спуска, рассмотренных в предыдущем разделе. В основе всех методов этого типа лежит идея случайного выбора направления движения и величины шага в факторном пространстве. При этом всегда используется не самое лучшее из возможных (антиградиентное) направление движения, а просто какое-либо из удачных направлений, приводящих к уменьшению критерия оптимальности, и не самый хороший, детерминированно заданный, а случайный шаг. Однако при этом существенно снижаются затраты на поиск этого шага.

Рис. 16.10. Статистический (случайный) спуск:

а — структурная схема программы простого статистического спуска; б — схема поиска оптимума при простом статистическом спуске; в — схема поиска оптимума при статистическом спуске с пересчетом

Для определения направления и величины шага используется датчик случайных чисел (см. разд. 6.1).

Методы статистического (случайного) спуска представляют собой развитие универсального метода проб и ошибок и, вообще говоря, отражают процесс эволюции в природе. Как отметил Л. А. Растригин, именно разнообразие и совершенство живой природы представляет собой лучшее доказательство эффективности этого метода. Элемент случайности (случайный шаг) проявляется в виде мутаций, которые либо улучшают, либо ухудшают шансы мутанта в борьбе за существование. В первом случае возникает возможность закрепления мутационных изменений в потомстве, что в конечном счете ведет к совершенствованию вида.

К простейшим алгоритмам статистического поиска относится **статистический поиск с возвратом**, иногда называемый простым случайным поиском. Структурная схема программы и схема поиска оптимума показаны на рис. 16.10. Суть метода заключается в следующем: для определения очередного случайного шага (например, из начальной точки 1 или любой промежуточной точки 2, 4, 5 и т. д. рис. 16.10, б) определяется $2k$ равномерно распределенных случайных чисел ξ_p $[0, 1]$, образующих два случайных вектора,

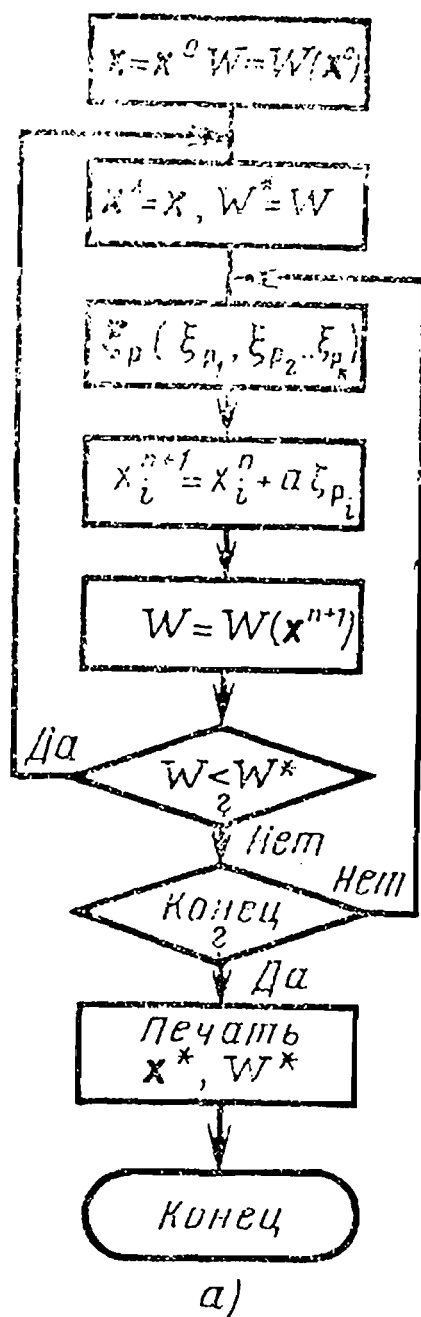
$$\xi_p^1(\xi_{p1}^1, \xi_{p2}^1, \dots, \xi_{pk}^1), \xi_p^2(\xi_{p1}^2, \xi_{p2}^2, \dots, \xi_{pk}^2), \quad (16.11)$$

и случайный шаг задается соотношениями

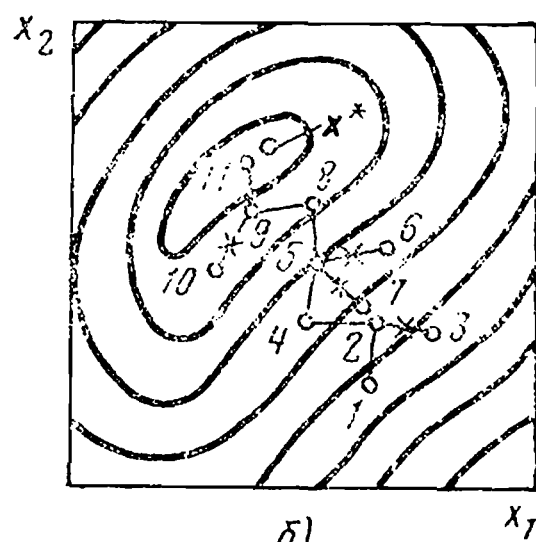
$$x_i^{n+1} = \begin{cases} x_i^n + a_i \xi_{pi}^1 & \text{при } \xi_{pi}^2 < 0,5, \\ x_i^n - a_i \xi_{pi}^1 & \text{при } \xi_{pi}^2 \geq 0,5. \end{cases} \quad (16.12)$$

При этом случайные числа ξ_{pi}^1 задают величину, а ξ_{pi}^2 — направление соответствующих компонент шага.

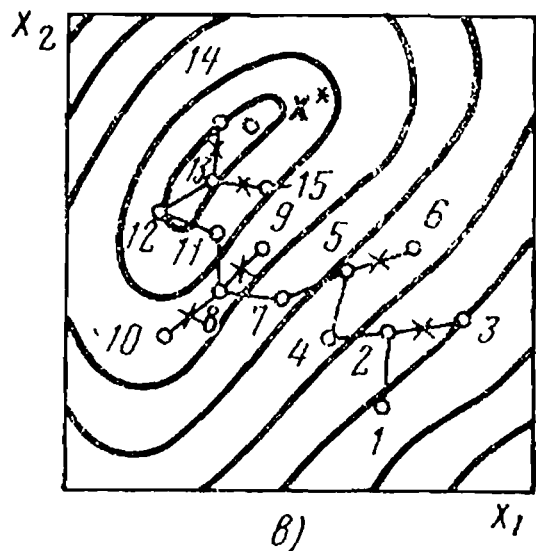
В новой точке определяется значение критерия оптимальности $W(x^{n+1})$. Это значение сравнивается с величиной $W(x^n)$ в исходной точке и, если $W(x^{n+1}) < W(x^n)$, шаг признается «удачным», а поиск смещается в точку x^{n+1} . В противном случае шаг признается не-



а)



б)



в)

удачным, поиск остается в точке $\mathbf{x}^n$; откуда и предпринимается новая случайная попытка. Таким образом,

$$\mathbf{x}^{n+1} = \begin{cases} \mathbf{x}^{n+1} & \text{при } W(\mathbf{x}^{n+1}) < W(\mathbf{x}^n); \\ \mathbf{x}^n & \text{при } W(\mathbf{x}^{n+1}) \geq W(\mathbf{x}^n). \end{cases} \quad (16.13)$$

В среднем удачным оказывается каждый второй пробный шаг, однако в окрестности экстремума вероятность удачного случайного шага снижается. Величина a_i представляет собой средний модуль шага по i координате. При приближении к точке экстремума этот модуль может оказаться слишком большим, что, так же как и в градиентном методе (см. разд. 16.2), приводит к «перескакиванию» через экстремум. Правда, в данном случае речь должна идти о «вероятности перескока», так как величина шага случайна.

По мере приближения к экстремуму модуль числа a_i должен уменьшаться. Сигналом для очередного уменьшения шага (например, в два раза) служит определенное количество (например, 10 или 20) следующих *подряд* неудачных попыток. Уменьшение модуля до определенной величины служит основанием для прекращения поиска, а конечная величина модуля шага характеризует «радиус» конечной области неопределенности, содержащей экстремум $\mathbf{x}^*$:

$$x_i^n - a_i^n \leq x_i^* \leq x_i^n + a_i^n. \quad (16.14)$$

Эффективность статистического поиска может быть существенно повышена, если наряду со случайностью ввести в алгоритм поиска некоторые детерминированные правила. Одним из таких методов является **статистический поиск с пересчетом**.

Этот метод (схема поиска оптимума показана на рис. 16.10, в) предназначен для поверхностей отклика с относительно стабильным характером изменения. В основе метода лежит предположение о том, что если случайный шаг оказался неудачным, то противоположный ему шаг должен привести к улучшению функции качества. Поэтому вслед за неудачным шагом делается не случайный, а детерминированный шаг в *противоположном* направлении:

$$\mathbf{x}^{n+1} = \begin{cases} \mathbf{x}^n + \mathbf{a}\xi_p^1 & \text{при } W(\mathbf{x}^n + \mathbf{a}\xi_p^1) < W(\mathbf{x}^n); \\ \mathbf{x}^n - \mathbf{a}\xi_p^1 & \text{при } W(\mathbf{x}^n + \mathbf{a}\xi_p^1) > W(\mathbf{x}^n); \end{cases} \quad (16.15)$$

$$\mathbf{a}(a_1, a_2, \dots, a_k); \quad \xi_p(\xi_{p1}, \xi_{p2}, \dots, \xi_{p3}).$$

Изменение знака вектора ξ_p^1 в (16.15) предполагает замену на противоположные знаков всех его компонент. В том случае, когда и пересчитанный шаг не дает улучшения функции качества $W(\mathbf{x}^n - \mathbf{a}\xi_p^1) \geq W(\mathbf{x}^n)$ (точка 8 на рис. 16.10, в), нужно сделать из точки $\mathbf{x}^n$ новый случайный шаг. Обычно в этом случае рекомендуется также уменьшить модуль шага $\mathbf{a}$. Введение пересчета позволяет повысить быстродействие поиска.

Если достаточно оснований считать модель линейной, возможно дальнейшее совершенствование метода. В этом случае можно не только изменить на проти-

воположное направление неудачного случайного шага (16.15), но и предположить, что сохраняется абсолютная величина приращения функции качества:

$$W(x^{n+1}) = \begin{cases} W(x^{n+1}), & W(x^{n+1}) < W(x^n); \\ W^*(x^{n+1}) = 2W(x^n) - W(x^{n+1}), & W(x^{n+1}) > W(x^n), \end{cases} \quad (16.16)$$

где функция $W^*(x^{n+1})$ не рассчитывается (определяется) заново, а пересчитывается по уже известному результату $W(x^{n+1})$ неудачного шага. При этом вдвое сокращаются затраты на поиск. Этот метод получил название **статистического поиска с линейным пересчетом функции качества**. Разумеется, достоверность такого пересчета невелика, поэтому данный метод может быть использован только на начальных этапах исследования ДЛА и лучше в комбинации (например, через раз) с обычным методом статистического поиска с пересчетом.

Одним из наиболее популярных статистических методов поиска является **статистический поиск с наказанием случайностью**, основанный на идее возможности повторного (или многократного) использования однажды найденного удачного направления.

Действительно, если какое-либо направление движения в факторном пространстве оказалось удачным, можно предполагать, что оно останется удачным на протяжении ряда шагов. Поэтому после определения случайного удачного направления поиск сохраняет это направление до тех пор, пока идет улучшение функции качества. При первом неудачном шаге (направление исчерпало свои возможности) из последней удачной точки совершается случайный шаг (серия шагов) для определения нового удачного направления. Алгоритм поиска с наказанием случайностью может быть записан в виде (рис. 16.11)

$$x^{n+1} = \begin{cases} x^n + \Delta x^n & \text{при } W(x^{n+1}) < W(x^n); \\ x + a\xi_p & \text{при } W(x^{n+1}) \geq W(x^n). \end{cases} \quad (16.17)$$

Легко видеть, что поиск с наказанием случайностью представляет собой *статистический аналог* метода наискорейшего спуска, от-

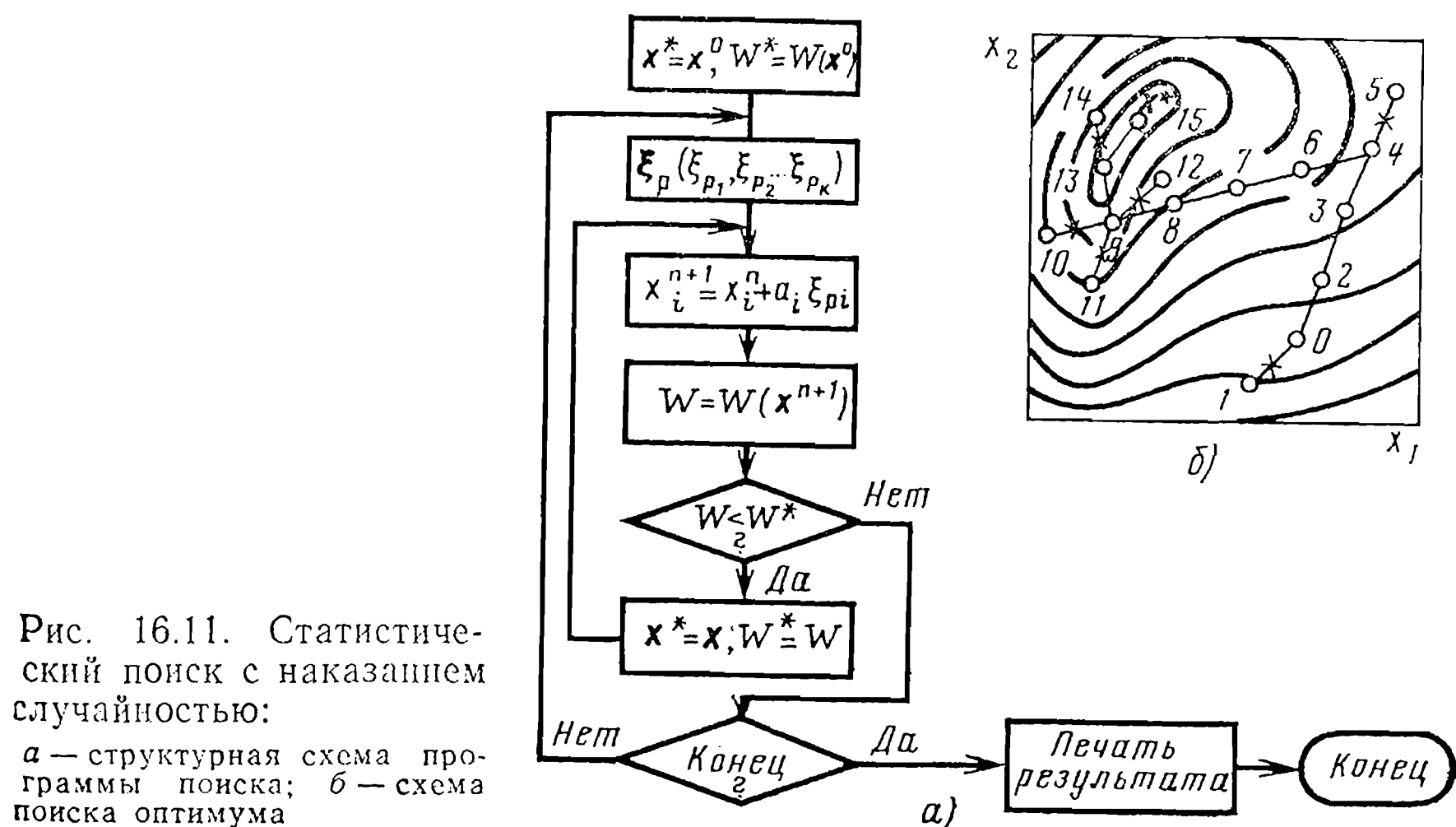


Рис. 16.11. Статистический поиск с наказанием случайностью:

a — структурная схема программы поиска; b — схема поиска оптимума

личающийся тем, что направления спуска выбираются случайно, а не заданы детерминированно.

Выбор направления во всех рассмотренных выше методах основан на единичной случайной пробе. Поэтому это направление, хотя и является по определению «удачным», вероятно, будет существенно отличаться от оптимального (антиградиентного) направления. Направление спуска может быть существенно улучшено при использовании нескольких случайных проб, приводящих к методу **статистического градиента**.

Суть метода заключается в следующем. Из исходной точки x^n делается m случайных пробных шагов:

$$x_j^{n+1} = x^n + a_j \xi_{pj}; \quad j = 1, 2, \dots, m, \quad (16.18)$$

где a_j ($a_{1j}, a_{2j}, \dots, a_{kj}$) — модуль пробного шага, и определяются соответствующие приращения критерия оптимальности

$$\Delta W_j = W(x^n + a_j \xi_{pj}) - W(x^n). \quad (16.19)$$

Векторная сумма

$$Y_m = \sum_{j=1}^m \xi_{pj} \Delta W_j \quad (16.20)$$

задает статистическую оценку градиента

$$\lim_{m \rightarrow \infty} \frac{Y_m}{|Y_m|} \rightarrow \frac{\text{grad } W(x)}{|\text{grad } W(x)|} \quad (16.21)$$

Рабочий шаг совершается в антиградиентном направлении

$$x^{n+1} = x^n - a \frac{Y_m}{|Y_m|}. \quad (16.22)$$

Разумеется, статистический антиградиент не может (при конечной выборке m) точно совпасть с действительным антиградиентом поверхности отклика. Однако направление статистического антиградиента будет существенно удачнее просто случайного удачного направления. При этом увеличиваются и затраты на поиск, так как для выбора каждого рабочего шага необходимо определить $m+1$ значения функции качества.

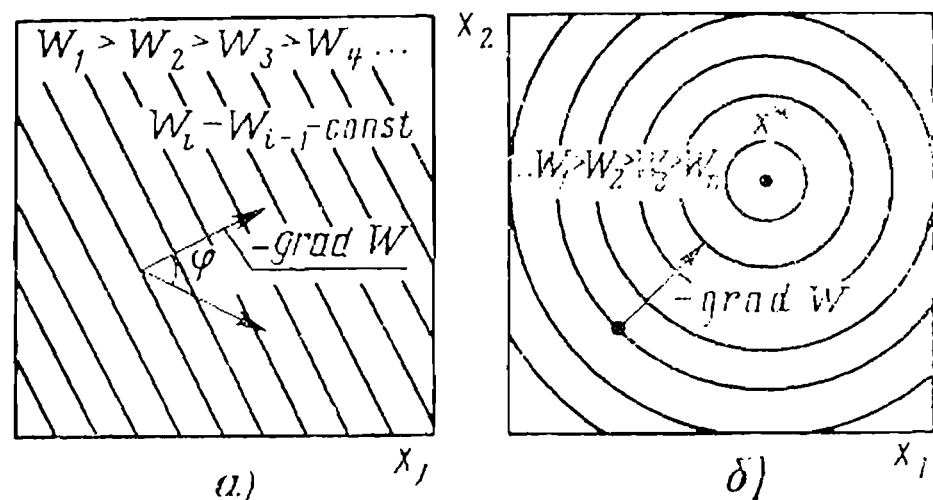
Использовать метод статистического градиента целесообразно при $m < k$ (при $m \approx k$ лучше рассчитать истинный градиент!) Поэтому этот метод обычно используется при оптимизации моделей большой размерности $k \geq (30 \dots 50)$. Если движение в направлении статистического антиградиента не приведет к успеху, необходимо так же, как и при градиентном методе, повторить вычисления при уменьшенном модуле шага a .

При сравнении статистических и детерминированных методов поиска нужно иметь в виду, что эффективность статистических методов является случайной величиной, существенно зависит от удачи и меняется от случая к случаю. Поэтому можно говорить лишь о статистической оценке эффективности этих методов. При этом

Рис. 16.12. Модели для оценки эффективности поиска:

а — линейная модель; б — центральная модель

«цена» (под ценой подразумевается необходимое количество вычислений функции качества) случайного удачного шага в многофакторном пространстве всегда меньше «цены» детерминированного шага. Однако эффективность случайного шага тоже меньше.



Целесообразность использования тех или иных методов поиска существенно зависит от характера поверхности отклика, однако при исследовании реальных моделей характер поверхности отклика заранее неизвестен. Для оценок обычно используются искусственные модели, в частности, линейная модель (рис. 16.12, а), характеризующая условия поиска вдали от экстремума, где кривизна линий уровня незначительна, и центральная модель (рис. 16.12, б), характеризующая поверхность отклика в непосредственной окрестности экстремума.

Для линейной модели вероятность удачного случайного шага из исходной точки равна 1/2. Поэтому для совершения рабочего шага в зависимости от принятого алгоритма требуются в среднем два (простой случайный поиск, поиск с пересчетом) или одно (поиски с линейным пересчетом функции качества) определение величины функции качества.

Эффективность рабочего шага оценивается смещением в сторону экстремума и зависит от модуля шага a и угла φ между случайным и антиградиентным направлениями (рис. 16.12, а):

$$\Delta W = -a \cos \varphi, \quad M(\Delta W) = a \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos \varphi p(\varphi) d\varphi, \quad (16.23)$$

где $M(\Delta W)$ — математическое ожидание среднего смещения за один шаг; $p(\varphi)$ — плотность распределения вероятности угла φ .

При равной вероятности всех направлений величина $M(\Delta W)$ уменьшается с увеличением размерности пространства k [18]. Однако темп этого уменьшения не превышает $1/\sqrt{k}$. Затраты на поиск определяются количеством расчетов функции качества, необходимых для смещения в сторону экстремума, и составляют

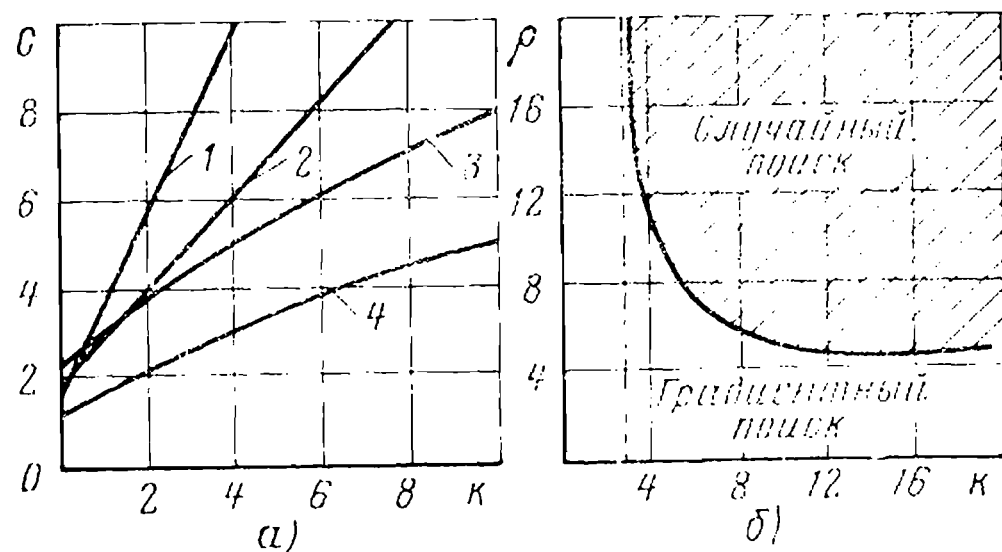
$$C = \frac{2}{M(\Delta W)}; \quad C = \frac{1}{M(\Delta W)} \quad (16.24)$$

в случае простого случайного поиска (поиска с пересчетом) и в случае поиска с линейным пересчетом функции качества соответственно.

При градиентном поиске смещение к экстремуму на расстояние a достигается за $n+1$ вычислений функции качества. Сравнение затрат на различном числе

Рис. 16.13. Сравнительная оценка эффективности статистического и градиентного поиска:

а — вдали от экстремума (линейная модель); б — в окрестности экстремума (центральная модель); 1 — градиентный поиск с парной пробой; 2 — градиентный поиск; 3 — статистический поиск с возвратом; 4 — статистический поиск с пересчетом



варьируемых переменных показано на рис. 16.13, а. Легко видеть, что простой статистический поиск и поиск с пересчетом оказываются эффективнее градиентного поиска при $k > 3$, а поиск с линейным пересчетом функции качества — при всех k .

Преимущество статистических методов растет с возрастанием размерности задачи. Поэтому статистические методы поиска могут быть рекомендованы для исследования многомерных моделей по крайней мере вдали от экстремума на первых этапах исследования.

Для центральной модели (см. рис. 16.12, б) вероятность удачного шага падает, что, конечно, приводит к снижению эффективности статистических методов поиска. При этом вероятность удачного шага зависит как от числа варьируемых переменных (k), так и от относительного расстояния от экстремальной точки ($\rho = r/a$). Области целесообразного использования статистического и градиентного поиска, полученные в рамках центральной модели, показаны на рис. 16.13, б [18]. Легко видеть, что в окрестности экстремума эффективность статистических методов уступает эффективности градиентного метода, а значит, применение их нецелесообразно.

Вопросы для самоконтроля

1. В чем заключается идея статистических методов поиска?
2. Как вводится при статистическом поиске элемент случайности?
3. В чем заключается разница методов случайного наброса и случайного спуска?
4. Дайте развернутую характеристику алгоритмов статистического поиска:
— простого статистического поиска,
— статистического поиска с пересчетом,
— статистического поиска с наказанием случайностью.
5. При каких условиях статистические методы поиска экстремума эффективнее детерминированных?

16.4. ПОИСК ОПТИМУМА В «ОВРАЖНОЙ» СИТУАЦИИ

В предыдущих разделах рассмотрены методы, предназначенные для изучения моделей, зависящих от примерно равнозначных переменных. Однако реальные модели ДЛА часто включают в себя переменные, различно влияющие на функцию качества. Поверхность отклика в этом случае содержит так называемые «овраги» (при поиске максимума — «гребни»), образующие достаточно протяженную область слабого изменения функции качества.

Математически овражная ситуация характеризуется наличием двух групп существенно различных по величине производных:

$$\left| \frac{\partial W(\mathbf{x})}{\partial x_i} \right| > \delta, \quad i = 1, 2, \dots, m < k; \quad \left| \frac{\partial W(\mathbf{x})}{\partial x_i} \right| < \delta, \quad i = m + 1, \dots, k \quad (16.25)$$

Величина δ называется *пороговым* значением. Параметры x_i при $i = 1, 2, \dots, m$, производные которых имеют большие значения, принято, в какой-то мере вопреки естественному смыслу, называть *несущественными* переменными (их влияние на величину W ярко выражено, и поэтому их легко оптимизировать!); переменные x_i при $i = m + 1, \dots, k$, влияние которых более слабо, а потому труднее поддается учету, принято называть *существенными*.

В реальных задачах оптимизации овражная ситуация возникает как непосредственно из-за учета факторов, оказывающих различ-

ное влияние на оптимизируемые характеристики, так и по формальным причинам, в частности, из-за взаимосвязи варьируемых переменных. Кроме того, на границах области исследования X^* возникают овраги, связанные с введением функции штрафа (см. разд. 13.1).

По конфигурации овраги принято делить на прямолинейные (дно оврага ориентировано вдоль некоторой прямой) и криволинейные. Для изучения процесса поиска оптимума в овражной ситуации вводят модельные овражные функции. Наиболее простой из них является прямолинейный эллиптический овраг

$$W(x) = \sum_{i=1}^m a_i x_i^2 + \sum_{i=m+1}^k x_i^2, \quad (16.26)$$

где a_i — достаточно большие числа ($a_i \gg 1$). Дно оврага задается гиперплоскостью $x_1 = x_2 = \dots = x_m = 0$, а минимум — точкой начала координат $x_1 = x_2 = \dots = x_k = 0$.

На рис. 16.14, а показан двумерный эллиптический овраг, задаваемый моделью

$$W(x_1, x_2) = x_1^2 + 10x_2^2. \quad (16.27)$$

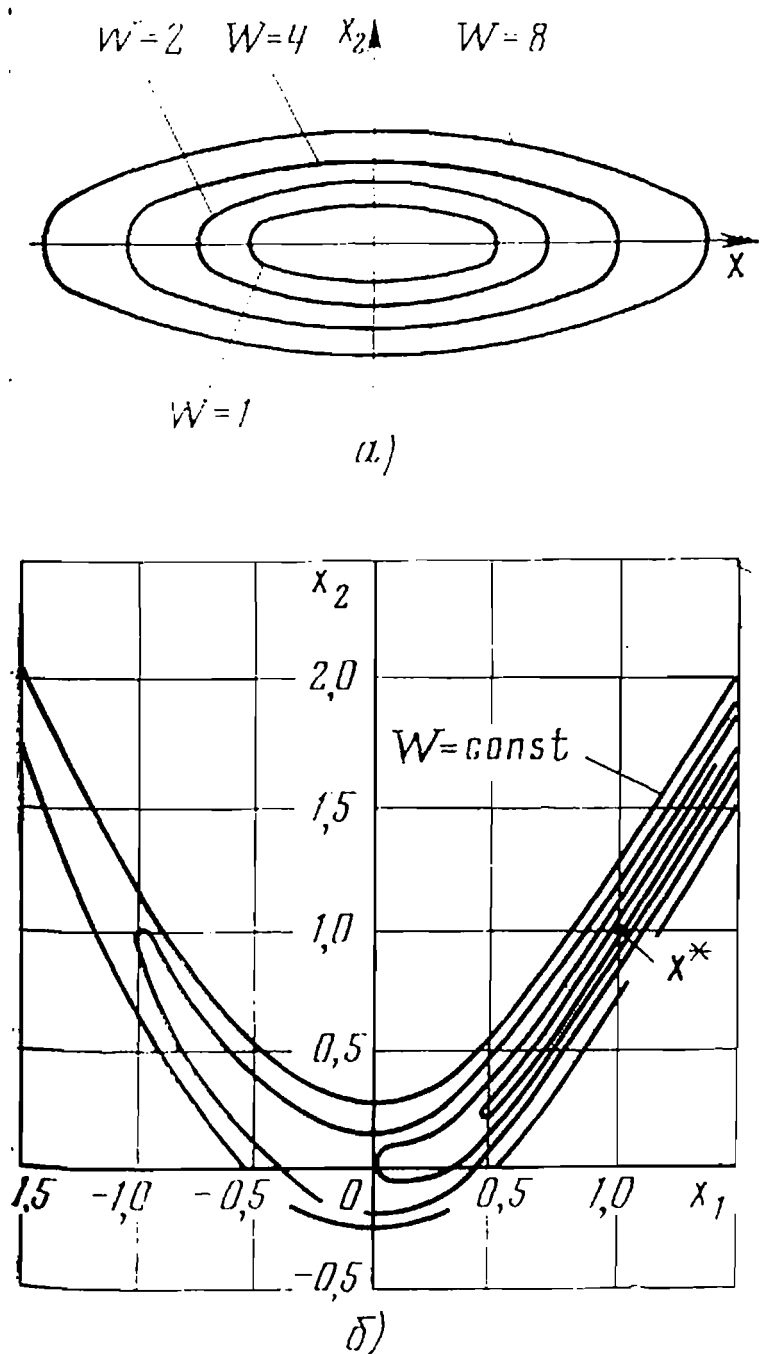


Рис. 16.14. Модели овражных поверхностей отклика:

а — эллиптический овраг; б — функция Розенброка

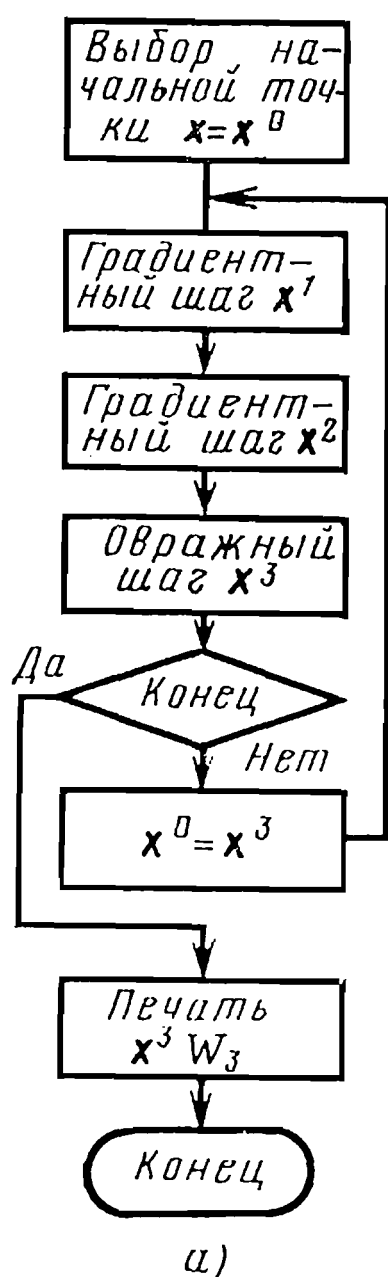
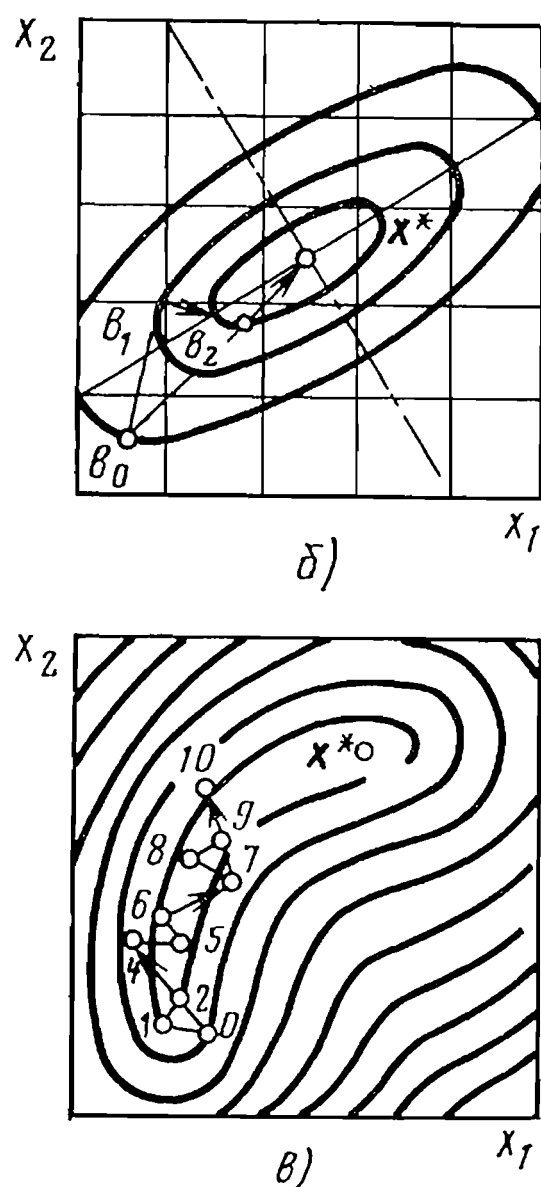


Рис. 16.15. Нелокальный метод наискорейшего спуска:

а — структурная схема программы поиска; б — схема поиска оптимума на эллиптическом овраге; в — схема поиска оптимума на криволинейном овраге



Наиболее распространенной моделью криволинейного оврага, на которой принято проверять различные методы оптимизации, является функция Розенброка:

$$W = 100(x_2 - x_1^2)^2 + (1 - x_1)^2, \quad (16.28)$$

задающая неглубокую криволинейную долину с минимумом в точке $x_1 = x_2 = 1$ (рис. 16.14, б)¹. Можно построить и многомерный аналог функции Розенброка.

Все рассмотренные локальные методы поиска малоэффективны в овражной ситуации. На рис. 16.5 показана плохая сходимость в овражной ситуации метода покоординатного спуска. Не эффективны в овражной ситуации и другие локальные методы, в частности, градиентный метод, наискорейший спуск, статистические методы поиска.

Действительно, например, при применении градиентного поиска очень быстро достигается дно оврага (оптимум несущественной переменной), а затем начинается колебательное движение между *откосами* оврага. При этом непрерывно уменьшаются шаг a_i и темп изменения функции качества (ΔW), что приводит к прекращению поиска. Благоприятная ситуация, при которой направление движения совпадает с направлением дна оврага и быстро приводит к экстремальной точке, возможна только в случае прямолинейного оврага, но и в этом случае крайне маловероятна.

Хуже всего то, что при использовании локальных методов оптимизации результаты исследования в овражной ситуации полностью совпадают с результатами исследования в окрестности экстремума. Поэтому, если изучаемая модель включает в себя овраги, никогда нельзя быть уверенным в том, что найденная точка факторного пространства действительно оптимальна.

Неэффективность в овражной ситуации локальных методов оптимизации заставляет переходить к более сложным, так называемым овражным методам поиска, всегда в той или иной мере предполагающим учет предыстории поиска, т. е. имеющим нелокальный характер.

Все методы овражного поиска включают в себя три основных этапа:

- изучение поверхности отклика в некоторой произвольной окрестности для определения направления движения вдоль оврага;
- движение вдоль оврага до выявления окрестности экстремума;
- изучение окрестности экстремума с целью уточнения его положения.

Нелокальный метод наискорейшего спуска. Метод разработан в первую очередь для исследования эллиптических оврагов и обеспечивает в этом случае быструю сходимость к экстремальной точке.

¹ Здесь в качестве несущественной переменной, определяющей дно оврага, выступает комплекс $(x_2 - x_1^2)^2$, в качестве существенной $(1 - x_1)^2$.

Структурная схема и схема поиска оптимума этим методом показаны на рис. 16.15.

Если поверхность отклика представляет собой эллипсоид вращения (рис. 16.7, в) и начальная точка поиска лежит на одной из его главных осей (точка 1 на рис. 16.7, в), метод наискорейшего спуска приводит к экстремуму в результате одного шага поиска. Однако очевидно, что вероятность такого удачного выбора начальной точки чрезвычайно мала. Во всех других случаях (точка 2 на рис. 16.7, в) траектория движения представляет собой ломаную линию с непрерывно уменьшающимся шагом, лежащую внутри некоторого конуса (на рисунке — в области, ограниченной двумя прямыми), вершина которого находится в экстремальной точке. При этом очевидно, что точка экстремума не может быть достигнута за конечное число шагов, так как при приближении к экстремуму длина шага непрерывно сокращается.

Характер траектории наводит на мысль о возможности ускорения поиска за счет движения *вдоль образующей* конуса. Метод предусматривает (рис. 16.15, б):

- 1) проведение двух шагов поиска методом наискорейшего спуска (в принципе можно использовать любой локальный поиск) от b_0 к b_1 и от b_1 к b_2 ;

- 2) построение прямой через точки b_0 и b_2 ;

- 3) одномерный поиск оптимума (любым из методов, рассмотренных в гл. 15) вдоль прямой b_0b_2 .

Для идеализированной эллиптической модели метод непосредственно приводит к окрестности экстремальной точки и не требует дополнительных исследований. Этот метод может быть распространен на реальные не эллиптические модели, в том числе на криволинейные овраги (рис. 16.15, в). Однако в этом случае рассмотренная процедура представляет собой лишь один шаг поиска и должна быть повторена из полученной в результате этого шага точки b_3 . Одномерный поиск вдоль прямой b_0b_2 не производится, а просто совершается достаточно длинный рабочий шаг. Точка b_3 может оказаться отнюдь не лучше предыдущих точек (например, b_2) так как возможен *выход* поиска на «откос» оврага, однако в целом траектория $b_0, b_3, b_6...$ позволяет *отследить* направление оврага.

В окрестности оптимальной точки поиск начинает совершать колебания и для исследования этой окрестности целесообразно использовать один из локальных методов оптимизации (см. разд. 16.2) или методы нелинейной аппроксимации (см. разд. 16.5).

Овражный поиск Гельфанда—Цетлина. Одним из наиболее простых и в то же время эффективных алгоритмов овражного поиска является поиск Гельфанда — Цетлина, предложенный ими для исследования задач рентгеноструктурного анализа. Структурная схема и схема поиска показаны на рис. 16.16.

Метод представляет собой комбинацию локального метода спуска на дно оврага и движения вдоль оврага. Сначала из про-

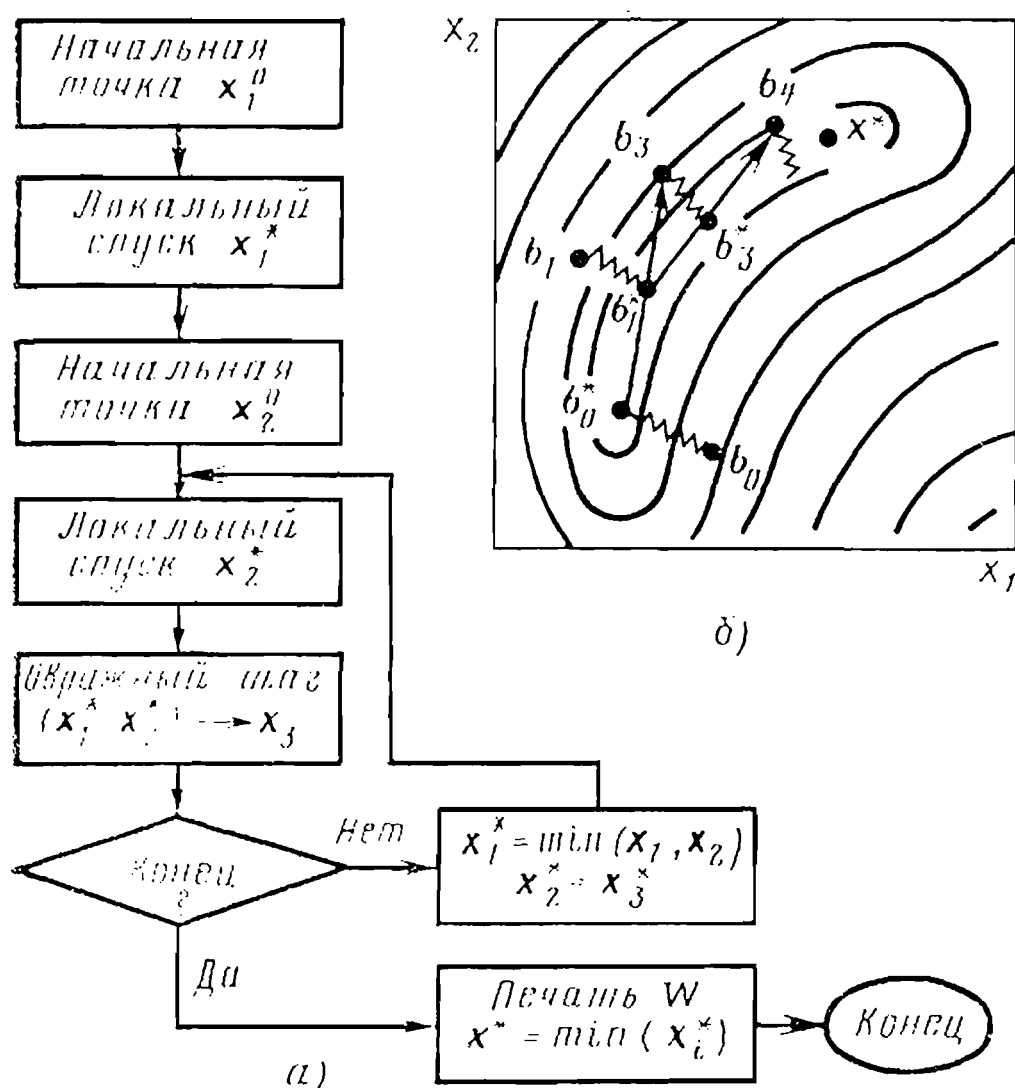


Рис. 16.16. Метод Гельфанда — Цетлина:

a — структурная схема программы поиска; *б* — схема поиска оптимума

извольной начальной точки b_0 (рис. 16.16, б) с помощью любого из локальных методов производится спуск на дно оврага в точку b_0^* . Далее из другой произвольной точки b_1 осуществляется такой же спуск в точку b_1^* . Точки b_0^* и b_1^* лежат на дне оврага, точнее в окрестности дна оврага, и, следовательно, соединяющая их прямая определяет его направление. Вдоль этой прямой делается рабочий овражный шаг в

сторону лучшей из точек b_0^* , b_1^* к точке b_2 . Точка b_2 не обязательно приводит к улучшению критерия оптимальности, так как может лежать на склоне оврага. Из этой точки вновь осуществляется локальный спуск в точку b_2^* . Новый овражный шаг строится для точки b_2^* и лучшей из точек b_0^* , b_1^* . По мере приближения к экстремуму величина овражного шага должна уменьшаться. Сигналом для уменьшения овражного шага может служить возникновение колебательного процесса со сменой направлений движения после каждого спуска.

В окрестности экстремума целесообразно провести дополнительное изучение поверхности отклика.

Могут быть предложены различные модификации и усовершенствования овражного поиска Гельфанда — Цетлина.

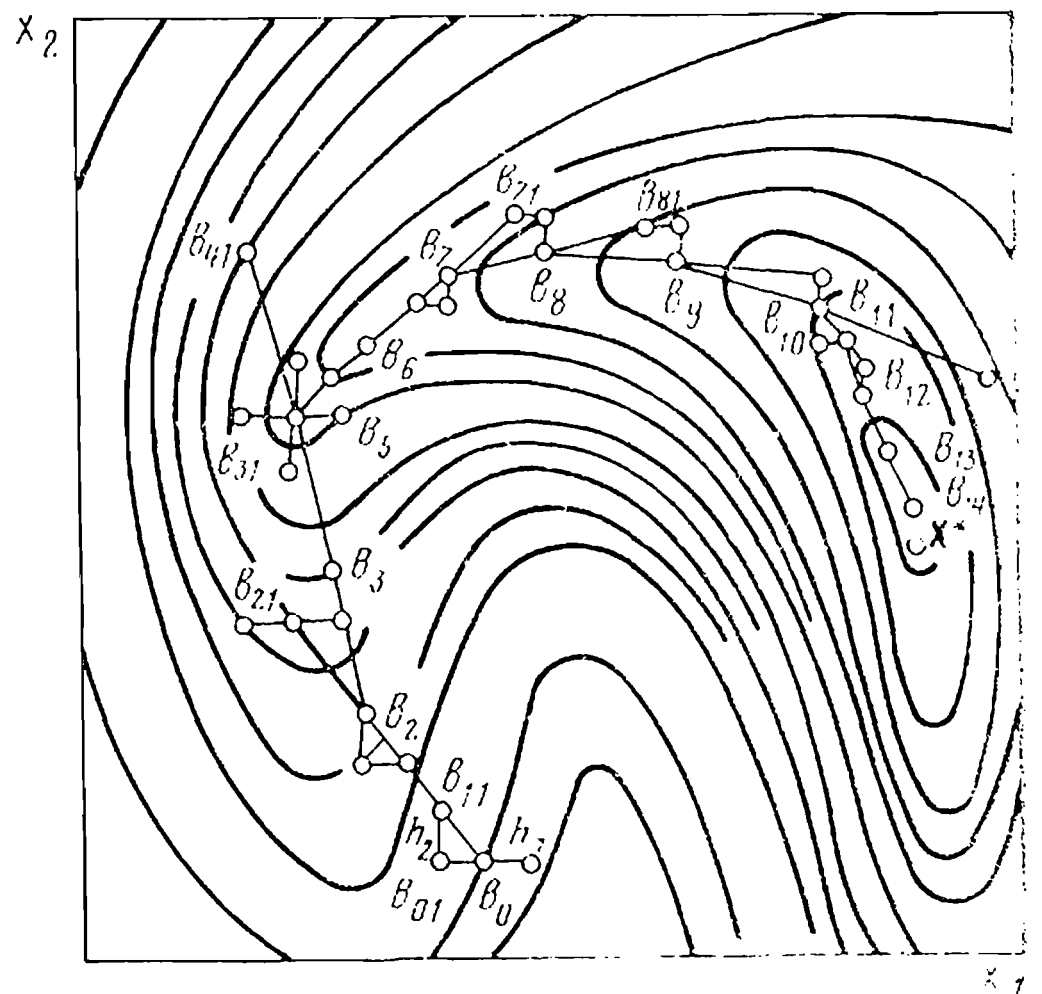
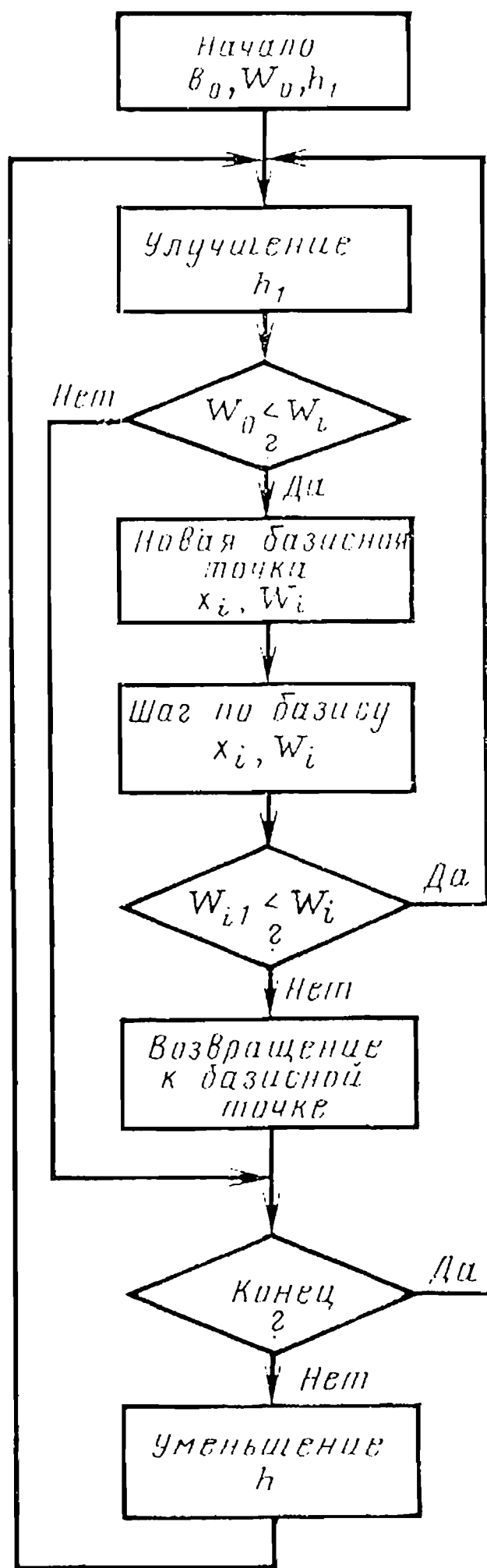
Метод конфигураций. Широкое распространение при исследовании овражных моделей получил также метод конфигураций. Структурная схема программы и схема поиска оптимума этим методом показаны на рис. 16.17.

Метод предусматривает локальное изучение поверхности отклика с помощью небольших пробных шагов и ускоренное движение вдоль дна оврага.

Поиск начинается в произвольно выбранной базовой точке b_0 и предусматривает на первом этапе проведение пробных шагов h_i по каждой из координат x_i . При этом сначала из точки b_0 делают пробный шаг h_1 по координате x_1 и, если он приводит к улучшению функции качества ($W(b_0 + h_1) < W(b_0)$), принимают точку $b_0 + h_1$ за «временную» вершину b_{01} . Если шаг h_1 не приводит к улучшению функции качества, делают противоположный ему шаг — h_1 в точку

Рис. 16.17. Метод конфигураций (Хука — Дживса):

а — структурная схема программы поиска;
б — схема поиска оптимума



б)

$b_0 \rightarrow h_1$. Если и этот шаг не дает улучшения $W(b_0 + h_1) > W(b_0)$, временную вершину сохраняют в точке b_0 . Таким образом,

$$b_{01} = \begin{cases} b_0 + h_1 & \text{при } W(b_0 + h_1) < W(b_0); \\ b_0 - h_1 & \text{при } W(b_0 - h_1) < W(b_0) \leq W(b_0 + h_1); \\ b_0 & \text{при } W(b_0) \leq \min \{W(b_0 + h_1), W(b_0 - h_1)\}. \end{cases} \quad (16.29)$$

Из точки b_{01} точно так же делается шаг h_2 по второй координате x_2 и т. д. В результате полного исследования всех переменных находится вторая базовая точка b_1 . Таким образом исключается смещение поиска со дна оврага (если базовая точка уже находится на дне оврага!), т. е. исключаются из рассмотрения несущественные переменные. В этом случае точка b_1 тоже лежит на дне оврага, причем $W(b_1) < W(b_0)$. Если базовая точка b_0 лежит на склоне

оврага (именно эта ситуация показана на рис. 16.17), процедура (16.29) обеспечивает спуск на дно оврага.

Пара базовых точек b_0 и b_1 определяет *первую конфигурацию*. Опираясь на эту конфигурацию, делается рабочий шаг, равный двойному расстоянию b_0b_1 в точку b_{11} , представляющую собой новую «временную» вершину. Вокруг точки b_{11} совершается цикл пробных шагов с целью коррекции направления движения. Этот цикл завершается определением вершины b_2 , которая в частном случае может совпадать с вершиной b_{11} .

Новый рабочий шаг делается на основе конфигурации b_1b_2 . В результате этого шага определяется временная вершина b_{21} . Аналогично определяются точки b_3b_4 . В процессе поиска систематически проверяется условие улучшения функции качества

$$W(b_4) < W(b_3) < W(b_2) < W(b_1) < W(b_0). \quad (16.30)$$

Интересно отметить, что величина коррекции направления движения влияет на величину рабочего шага, причем правильно выбранное направление приводит к увеличению длины шага. Так, в точке b_{11} (рис. 16.17) изменения направления движения не потребовалось, что привело к увеличению длины шага вдвое. В точке b_{21} потребовалась незначительная коррекция, причем длина шага тоже увеличилась. В точке b_{31} цикл пробных шагов не привел к улучшению временной вершины ($b_4 = b_{31}$). При этом направление движения и длина шага остались неизменными.

Если очередной рабочий шаг не приводит к улучшению функции качества ($W(b_{41}) > W(b_4)$), конфигурация разрушается. В этом случае необходимо продолжить поиск из последней удачной точки (b_4) при уменьшенной (обычно в два раза) величине пробных шагов. Таким образом, удастся построить новую конфигурацию и продолжить движение вдоль оврага. При этом в зависимости от характера коррекции направления движения длина шага может возрастать, оставаться неизменной или уменьшаться. Если после очередного разрушения конфигурации (в точке b_{14}) и уменьшения пробных шагов h_i до некоторой заранее заданной величины построить новую конфигурацию не удастся, поиск завершается, а последняя точка поиска задает окрестность экстремума.

В овражной ситуации могут быть использованы и некоторые статистические методы, в частности статистические методы с *самообучением* [18].

Общей особенностью всех овражных методов является зависимость каждого рабочего шага от предыдущих шагов, от предыстории поиска. Это коренным образом отличает овражные методы от локальных методов, рассмотренных в разделах 16.2, 16.3, сообщает им глобальный характер. При этом одновременно увеличивается трудоемкость методов, затраты на поиск. Вместе с тем все рассмотренные методы предполагают унимодальность поверхности отклика и, вообще говоря, непригодны для изучения полимодальных моделей.

1. Дайте общую характеристику овражной ситуации.
2. Дайте подробную характеристику нелокального метода наискорейшего спуска.
3. Дайте подробную характеристику овражного поиска Гельфанда — Цетлина.
4. Дайте подробную характеристику метода конфигураций.

16.5. НЕЛИНЕЙНАЯ АППРОКСИМАЦИЯ ПОВЕРХНОСТИ ОТКЛИКА

Методы многопараметрической оптимизации, рассмотренные в разд. 16.2—16.4, так или иначе основаны на линейной аппроксимации поверхности отклика, т. е. на замене реальной модели ее линейным образом — гиперплоскостью. Возможность такой замены по крайней мере для некоторого участка поверхности отклика следует из возможности разложения функции качества в ряд Тейлора (5.9) при сохранении в разложении только линейных членов:

$$W(x) \approx \tilde{W}(x) = W(x^0) + \sum_{i=1}^k b_i (x_i - x_i^0);$$

$$b_i = \frac{\partial W(x^0)}{\partial x_i}. \quad (16.31)$$

Именно на линейной аппроксимации основано использование для выбора направления поиска численного антиградиента, случайных проб, рабочих шагов вдоль оврага в овражной ситуации и т. д. Поэтому все рассмотренные выше методы поиска принято называть *линейными* методами или методами поиска с *линейной тактикой*.

Линейные методы приводят к относительно небольшим затратам на поиск и эффективны для достаточно контрастных поверхностей отклика. Однако в непосредственной окрестности экстремума, а также в окрестностях точек перегиба, когда функция качества меняется слабо, линейная аппроксимация может оказаться неэффективной. Движение по антиградиенту (равно, как и случайные пробы) не приводит к существенному улучшению функции качества, в результате чего поиск останавливается достаточно далеко от экстремальной точки, а при наличии помех (см. разд. 16.7) возможны ошибки в выборе направления движения.

В этих случаях для описания поверхности отклика целесообразно использовать *нелинейные* многообразия, в частности, *квадратичные* формы, полученные при сохранении в разложении Тейлора (5.9) не только линейных, но и квадратичных членов:

$$W(x) \approx \tilde{W}(x) = W(x^0) + \sum_{i=1}^k b_i (x_i - x_i^0) + \sum_{\substack{i=1 \\ j \neq i}}^k b_{ij} (x_i - x_i^0) (x_j - x_j^0);$$

$$b_i = \frac{\partial W(x^0)}{\partial x_i}; \quad b_{ij} = \frac{\partial^2 W(x^0)}{\partial x_i \partial x_j}. \quad (16.32)$$

В векторной форме (16.32) можно записать в виде

$$\tilde{W}(x) = W(x^0) + B^*X + X^*HX, \quad (16.33)$$

где

$$B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix}; \quad X = \begin{pmatrix} x_1 - x_1^0 \\ x_2 - x_2^0 \\ \dots \\ x_k - x_k^0 \end{pmatrix};$$

$$H = \begin{pmatrix} b_{11}, & \frac{1}{2} b_{12}, \dots, & \frac{1}{2} b_{1k} \\ \frac{1}{2} b_{21}, & b_{22}, \dots, & \frac{1}{2} b_{2k} \\ \dots & \dots & \dots \\ \frac{1}{2} b_{k1}, & \frac{1}{2} b_{k2}, \dots, & b_{kk} \end{pmatrix}; \quad (16.34)$$

знак $(*)$ обозначает транспонирование матрицы, а матрица H для (16.32) играет роль матрицы Гессе.

Если базовая точка x^0 расположена не слишком далеко от истинного экстремума модели x^* , то оценка положения этого истинного экстремума дается экстремумом квадратичной формы (16.32) x^{**} . Для уточнения этой оценки можно в суженной окрестности x^{**} построить новую квадратичную форму и определить новое значение x_1^{**} и т. д. Последовательность возникающих при этом приближений (в одномерной интерпретации) ясна из рис. 16.18.

Задача поиска экстремума при этом распадается на две задачи: построение аппроксимирующей квадратичной формы; определение экстремума этой квадратичной формы x^{**} .

Затраты на квадратичную аппроксимацию поверхности отклика существенно превосходят затраты, связанные с линейной аппроксимацией. Квадратичная форма (16.32) включает в себя C_{k+2}^2 коэффициентов и, следовательно, требует определения функции качества как минимум в C_{k+2}^2 точках факторного пространства¹. Так, при квадратичной аппроксимации функции 10 переменных необходимо знать ее значения в 66 точках факторного пространства вместо 11 при линейной аппроксимации, при 25 переменных — в 351 точке вместо 26 и т. д.

Для определения экстремума квадратичной формы (16.33) (и вообще нели-

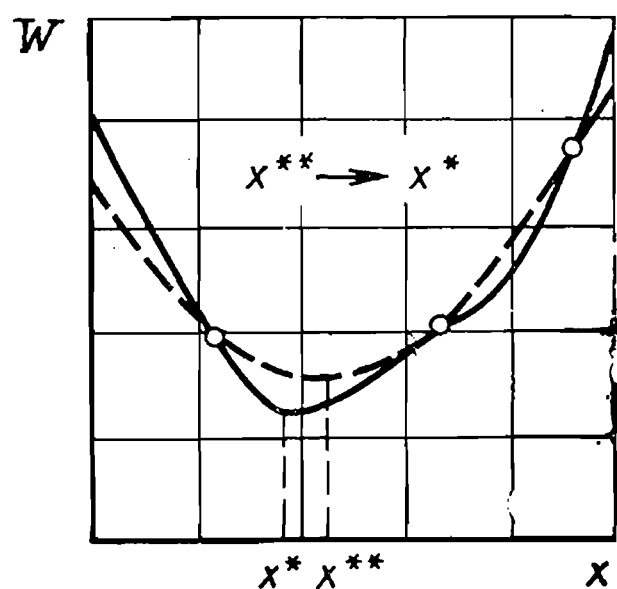


Рис. 16.18. Нелинейная аппроксимация поверхности отклика

¹ При наличии ошибок (помех) это число должно быть увеличено в 2—3 раза (см. гл. 10—12).

нейных аппроксимационных моделей поверхности отклика) разработан ряд эффективных методов, основанных на использовании вторых производных (так называемые методы 2-го порядка). В отличие от реальной поверхности отклика (истинной модели) квадратичная аппроксимационная модель имеет аналитические вторые производные, образующие матрицу $\mathbf{H}$ (16.34), которая может быть использована для улучшения поиска экстремума $\mathbf{x}^{**}$.

Наиболее распространенным из таких методов является метод Ньютона — Рафсона, представляющий собой развитие метода Ньютона (см. разд. 14.1) и предусматривающий использование вторых производных для построения средневзвешенных значений первых производных, которые, в свою очередь, определяют направление движения в факторном пространстве. Таким образом, метод Ньютона — Рафсона представляет собой *модифицированный градиентный метод*.

В соответствии с методом Ньютона — Рафсона следующая за $\mathbf{x}^n$ точка $\mathbf{x}^{n+1}$ определяется соотношением (сравнить с (16.7))

$$\mathbf{x}^{n+1} = \mathbf{x}^n - \mathbf{H}^{-1}(\mathbf{x}^n) \dot{\mathbf{W}}(\mathbf{x}^n), \quad (16.35)$$

где $\dot{\mathbf{W}}(\mathbf{x}^n)$ — матрица-столбец первых производных.

Для определения экстремума $\mathbf{x}^{**}$ квадратичной формы (16.32) достаточно одного шага (16.35). При этом точка $\mathbf{x}^{**}$ определяется точно. В общем случае нелинейной аппроксимационной модели необходимо осуществить некоторую последовательность шагов (16.35). Критерием сходимости метода Ньютона—Рафсона в рамках нелинейной аппроксимационной модели при условии, что $\mathbf{W}(\mathbf{x})$ дважды дифференцируема, служит положительная определенность матрицы $\mathbf{H}^{-1}$, обратной матрице $\mathbf{H}$.

Метод Ньютона — Рафсона требует проведения большого объема вычислений. В частности, количество умножений при обращении матрицы ($k \times k$) приблизительно пропорционально k^3 , т. е. представляет собой большую величину даже для задач умеренной размерности.

При использовании квадратичных моделей для последовательной аппроксимации реальной поверхности отклика остается открытым вопрос о сходимости последовательности значений $\mathbf{x}^{**}$ к истинному значению экстремума $\mathbf{x}^*$. Можно утверждать, что сходимость метода Ньютона—Рафсона будет обеспечена, если начальная точка $\mathbf{x}^0$ достаточно близка к точке экстремума $\mathbf{x}^*$. Это обстоятельство, равно как и высокая трудоемкость, позволяют рекомендовать использование метода Ньютона—Рафсона на последних этапах поиска экстремума, когда окрестность экстремума уже определена каким-либо из линейных методов.

Вопросы для самоконтроля

1. На чем основаны и когда применяются методы нелинейной аппроксимации?
2. Сравните между собой линейные и нелинейные методы поиска экстремума.
3. В чем заключается процедура последовательного приближения поверхности отклика квадратичными моделями?

16.6. ИССЛЕДОВАНИЕ ПОЛИМОДАЛЬНЫХ МОДЕЛЕЙ. ПОИСК ГЛОБАЛЬНОГО ЭКСТРЕМУМА

Методы многопараметрической оптимизации, рассмотренные в разд. 16.2—16.5, предназначены для исследования унимодальных моделей. Для полимодальных моделей эти методы непригодны, так как в общем случае приводят к *ближайшему* локальному экстремуму (рис. 16.19, а) и не гарантируют определения самого главного — *глобального* экстремума. Между тем исследователя, конечно, интересует самое хорошее из возможных решений.

Исследование полимодальных моделей представляет собой очень сложную задачу. Более того, не существует абсолютно надежных методов поиска глобального экстремума. В то же время при оптимальном проектировании полимодальные модели встречаются достаточно часто. Одной из причин полимодальности моделей ДУ и ДЛА являются дискретно изменяющиеся переменные. Такими переменными являются варианты структуры или компоновки ДУ, количество тех или иных элементов конструкции, используемые материалы и рабочие тела и т. д.

Например, при оптимизации параметров воздушно-реактивных двигателей может быть выбрана или одноконтурная или двухконтурная схема двигателя. Оба варианта схемы могут стать предметом локальной оптимизации. Однако общая задача оптимизации двигателя требует исследования *двухэкстремальной задачи*. В рассмотренном простейшем случае, конечно, достаточно исследовать две локальные задачи, а потом сравнить между собой полученные результаты. В общем случае число дискретно изменяющихся переменных может быть большим (дискретно меняется число ступеней компрессора и турбины, количество камер сгорания и т. д.), что приводит к очень большому числу локальных задач. Полимодальность модели может быть следствием также тех или иных особенностей физических процессов. Поэтому возникает необходимость разработки специальных глобальных методов поиска экстремума.

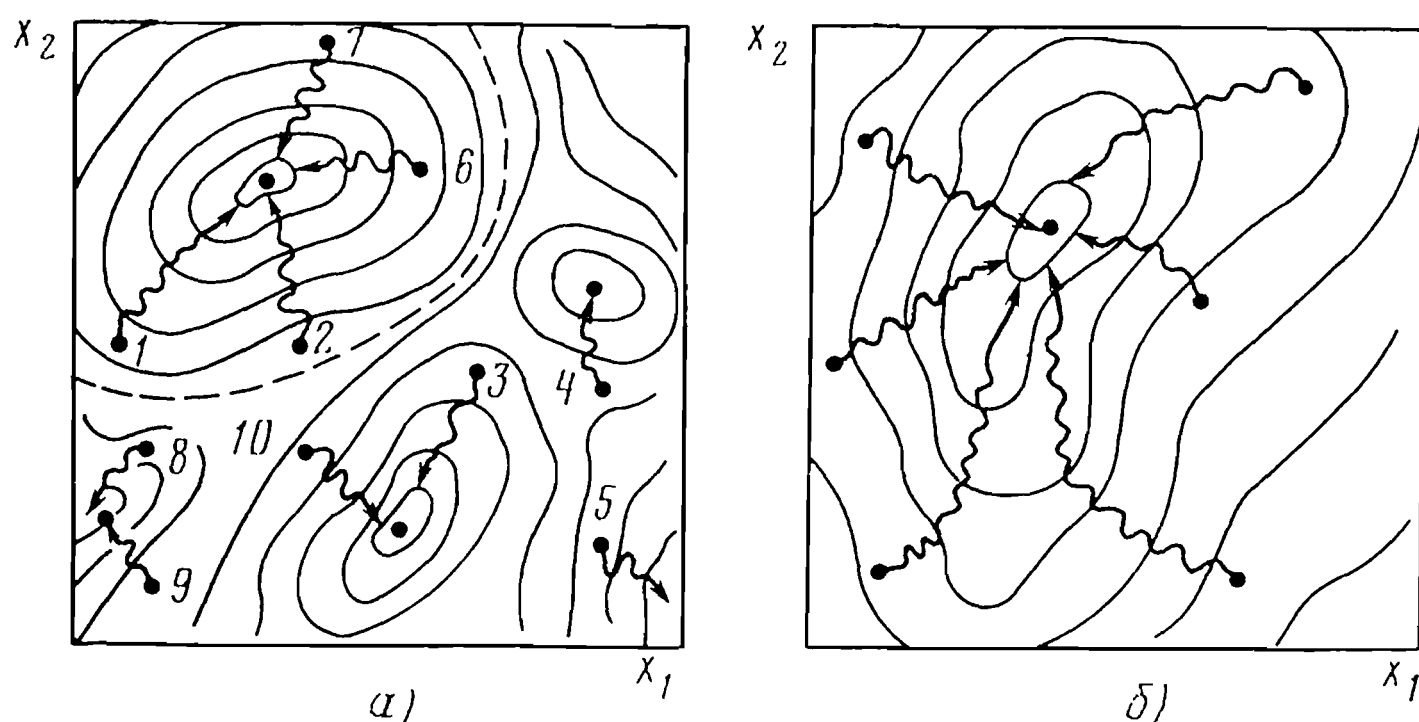


Рис. 16.19. Схема поиска глобального оптимума:

а — поиск методом многократного спуска; б — доказательство унимодальности поверхности отклика при сходимости многократных спусков к одной точке

Для отработки *глобальных* методов разработан ряд искусственных математических моделей. В качестве одной из них может быть использована функция с m экстремумами [18]

$$W(x) = W_0 - \sum_{j=1}^m \frac{a_j}{t_j (x - x_j^*)^2 + 1}, \quad (16.36)$$

где x_j^* — положение j -го экстремума; t_j — коэффициент «заостренности» j -го экстремума.

Графически полимодальная модель может быть представлена как система «долин» («вершин»), разделенных хребтами с перевалами (рис. 16.19, а, $k=2$, $m=5$). Каждый из экстремумов имеет свою «зону притяжения», характеризующую область сходимости локальных методов поиска.

Среди глобальных методов поиска экстремума в первую очередь нужно назвать пассивные **методы детерминированного или случайного сканирования**, требующие параллельного исследования поверхности отклика в некотором множестве точек факторного пространства. При достаточном количестве таких точек и не слишком узких «зонах притяжения» сканирование позволяет выявить все локальные экстремумы модели.

Для уточнения положения локальных экстремумов можно использовать несколько последовательных серий сканирующего поиска, стягивая изучаемые точки факторного пространства к предполагаемым экстремумам. Однако метод сканирования, как уже указывалось (разд. 16.1), очень трудоемок и поэтому может быть использован лишь при малом числе варьируемых переменных ($k \leq 3$).

Непосредственным развитием метода сканирования является метод, предусматривающий ряд параллельных спусков из различных точек факторного пространства (см. рис. 16.19, а). Начальные точки или задаются случайно, или равномерно распределены в изучаемой области. Метод, таким образом, сочетает элементы сканирования (выбор начальных точек) и локальные методы поиска экстремума. Шаг сканирования в этом случае, разумеется, существенно больше, чем при чистом сканировании. Если спусков достаточно много, а каждый из локальных экстремумов обладает достаточной зоной притяжения, можно надеяться, что будут обнаружены все локальные экстремумы. Если все спуски сходятся в одну окрестность (см. рис. 16.19, б), имеются достаточные основания считать модель унимодальной.

В ряде случаев, в частности, при большом числе дискретно заданных элементов (число ступеней компрессора или камер сгорания воздушно-реактивного двигателя) на поверхности отклика возникают локальные экстремумы с *малой зоной притяжения* (рис. 16.20). Для преодоления такой «зарифленности» поверхности отклика может быть предложен ряд методов с *инерцией*, позволяющей продолжить поиск при временных неудачах.

Метод тяжелого шарика основан на аналогии смещения поиска в факторном пространстве и движения тяжелого шарика по криво-

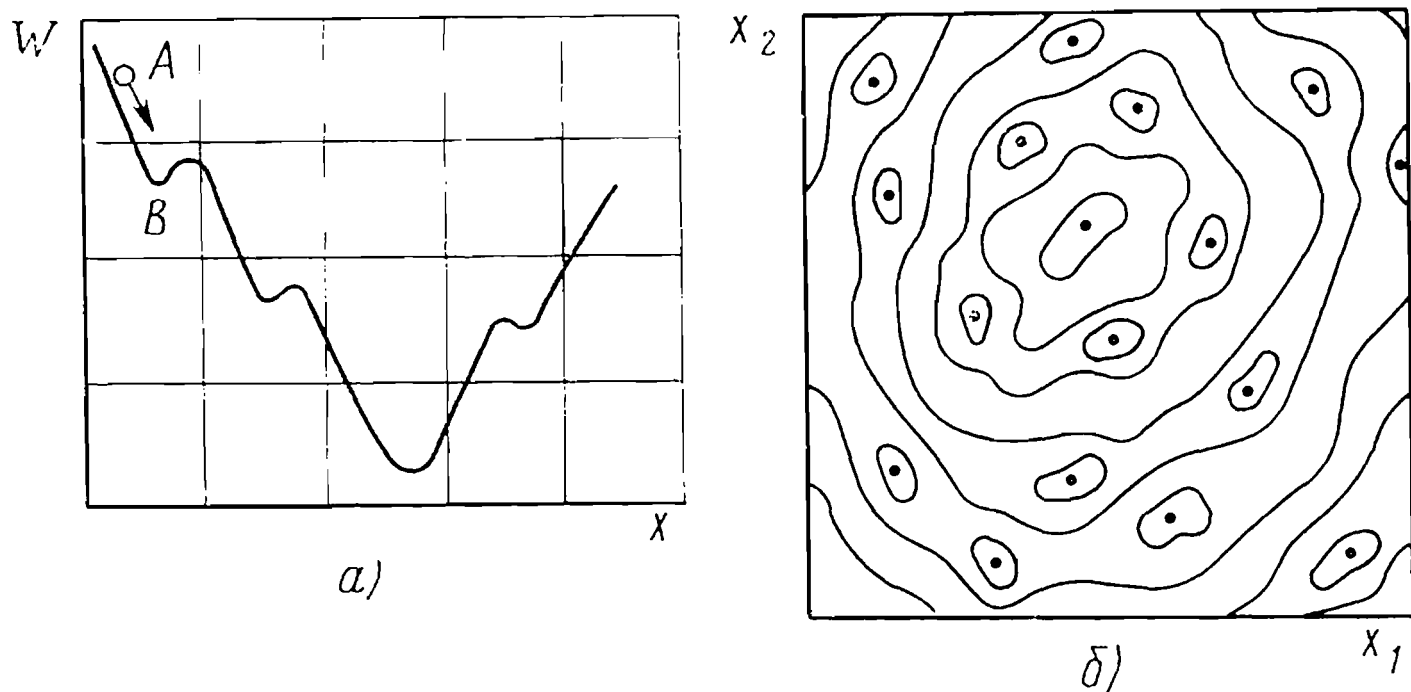


Рис. 16.20. Схема поиска глобального оптимума при наличии локальных экстремумов с малой зоной притяжения (зарифленность поверхности отклика):
a — одномерный случай; *б* — двумерный случай

линейной поверхности. Если шарик с некоторой массой начинает движение из точки *A* (рис. 16.20, *a*), то под действием силы тяжести он будет набирать скорость и получит возможность за счет инерции преодолеть локальный экстремум (точка *B*), если высота барьера окажется ниже начальной точки *A*. В случае удачи (а вероятность удачи растет с числом попыток) среди пройденных шариком экстремумов должен оказаться и глобальный экстремум.

При оптимизации роль инерции выполняет специальная функция, накапливающая информацию о результатах предыдущих шагов поиска. Например, алгоритм определения очередного шага поиска на основе метода градиента можно записать следующим образом:

$$x_i^n = x_i^{n-1} - \beta (x_i^{n-1} - x_i^{n-2}) - a_i \frac{\partial W(x^{n-1})}{\partial x_i}, \quad (16.37)$$

где β — коэффициент «инерции» поиска, накопленной на предыдущем шаге ($0 \leq \beta \leq 1$).

Можно предусмотреть накопление информации за ряд предшествующих шагов поиска:

$$x_i^n = x_i^{n-1} - \sum_{j=1}^{n-1} \beta_j (x_i^j - x_i^{j-1}) - a_i \frac{\partial W(x^{n-1})}{\partial x_i}, \quad (16.38)$$

причем естественно предположить, что вес более давних шагов меньше ($\beta_{n-1} > \beta_{n-2} > \beta_{n-3} \dots$). При $\beta \approx 1$ инерция велика и поиск совершает колебательные движения, при малых β колебания быстро затухают, при отсутствии инерции ($\beta = 0$) соотношения (16.37), (16.38) совпадают с соотношением (16.7) для обычного градиентного метода. Наличие инерции позволяет поиску сохранять направление движения, приносящее успех при предыдущих шагах, и преодолевать на этом пути временные препятствия.

Определенными глобальными свойствами обладают также **статистические методы поиска экстремума с самообучением**.

Локальные статистические методы (см. разд. 16.3) при каждом очередном шаге или начинают поиск заново (делают случайные пробы, полностью игнорируя предыдущий опыт), или, наоборот, «слепо» сохраняют прежнее удачное направление движения. Оптимальная стратегия лежит где-то посередине и предусматривает сочетание случайного выбора с анализом накопленного опыта, т. е. с самообучением в процессе поиска. Механизм самообучения и сообщает методу определенную инерцию.

Существует ряд вариантов статистического поиска с самообучением [18]. В качестве примера можно рассмотреть самообучение при простом случайном поиске, который предусматривает одинаковую вероятность движения как в положительном, так и в отрицательном направлении по каждой из координат (16.12). Для организации самообучения можно ввести в поиск дополнительный *вектор памяти* $\mathbf{w}$ ($w_1, w_2, \dots, w_k$), изменяющий вероятностные характеристики поиска. При этом случайный шаг задается соотношениями

$$x_i^{n+1} = \begin{cases} x_i^n + a_i \xi_{pi}^1 & \text{при } \xi_{pi}^2 < 1 - 0,5 \exp(-a_i w_i^n); \\ x_i^n - a_i \xi_{pi}^1 & \text{при } \xi_{pi}^2 \geq 1 - 0,5 \exp(-a_i w_i^n); \end{cases} \quad (16.39)$$

$$w_i^n = w_i^{n-1} - \delta \operatorname{sign}((x_i^n - x_i^{n-1})(W(\mathbf{x}^n) - W(\mathbf{x}^{n-1}))); \quad w_i^0 = 0.$$

При $w_i = 0$ соотношения (16.39) совпадают с соотношением (16.12), поэтому поиск начинается в равновероятной ситуации. По мере накопления информации вырабатывается предпочтительное направление поиска. Поиск «детерминируется» в некотором направлении. Темп перестройки вероятностных характеристик поиска определяется «шагом памяти» δ ($0 < \delta < 1$), а введение в память величины рабочего шага a учитывает базу поиска, на которой получена информация.

Например, если для i -й компоненты в течение нескольких (например, десяти) шагов оказывается выгодным положительное направление ($x_i^n - x_i^{n-1} > 0$, $W(\mathbf{x}^n) - W(\mathbf{x}^{n-1}) < 0$) и $\operatorname{sign}((x_i^n - x_i^{n-1})(W(\mathbf{x}^n) - W(\mathbf{x}^{n-1}))) = -1$, то величина соответствующей компоненты памяти становится достаточно большой ($w_i^n = 10\delta$), а величина $\exp(-a w_i)$ — достаточно малой, а это значит, что вероятность выбора следующего шага в положительном направлении резко возрастает.

Рассмотренный вариант статистического поиска с самообучением не позволяет непосредственно двигаться в неудачном направлении. Однако концентрация случайных проб в определенном направлении в сочетании со случайной величиной шага ($a_i \xi_{pi}^1$) повышает вероятность выхода поиска из локального экстремума.

Наиболее ярко глобальные свойства выражены у **статистического поиска с направляющим конусом**. Суть метода состоит в ограничении возможных случайных направлений движения некоторым гиперконусом (рис. 16.21). В факторном пространстве вектором $\mathbf{a}$ определен k -мерный гиперконус с вершиной в точке $\mathbf{x}^n$ (точка 1 на

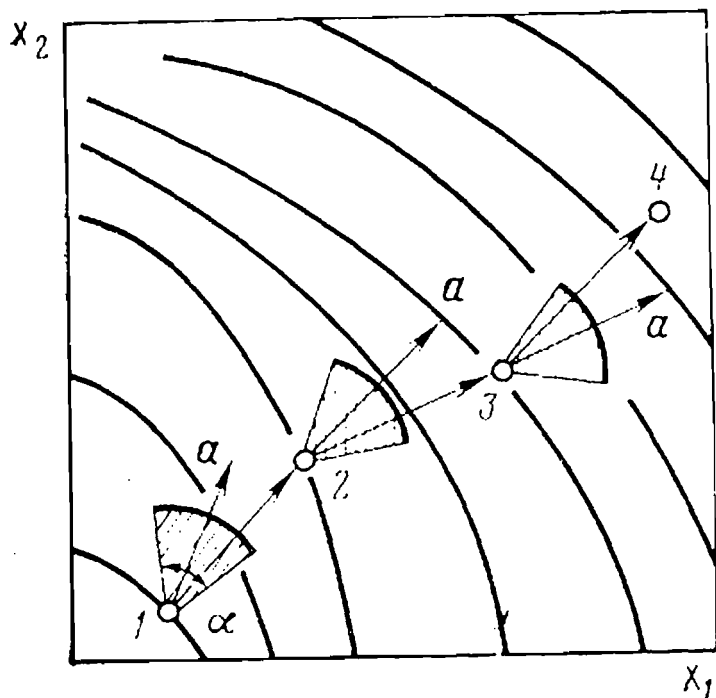


Рис. 16.21. Метод случайного поиска с направляющим конусом

рис. 16.21) и постоянным углом раскрытия α . В пределах этого конуса делается ряд случайных пробных шагов $b\xi^1, b\xi^2, \dots, b\xi^m$:

$$\xi^j(\xi_1^j, \xi_2^j \dots \xi_k^j); \xi_i^j = \alpha(\xi_p[0,1] - 0,5)$$

и среди них выбирается шаг, соответствующий наилучшему значению функции качества. В направлении этого наилучшего шага делается рабочий шаг

$$x^{n+1} = x^n + c\xi^*; W(x^n + c\xi^*) = \min_{i=1,2,\dots,m} W(x^n + c\xi^i), \quad (16.40)$$

и ось гиперконуса совмещается с направлением рабочего шага. В новой точке x^{n+1} (точка 3 на рис. 16.21) процедура повторяется.

Особенностью рассмотренного алгоритма является взаимная *связанность* двух следующих друг за другом шагов: их направления не могут отличаться друг от друга более, чем на $\alpha/2$. При этом не допускается крутых разворотов поиска, что заставляет его следовать вдоль оврага. Рассматриваемый метод обладает известными глобальными свойствами, так как позволяет поиску не только спускаться, но и подниматься вдоль оврага (среди наилучших проб могут оказаться пробы, не приводящие к уменьшению функции качества!), проходить «возвышенности» рельефа функции качества по «перевалам», за которыми может лежать глобальный минимум.

Наконец, для многомерных полимодальных моделей, удовлетворяющих условию Липшица (15.39), т. е. имеющих во всей рассматриваемой области *ограниченную* производную, может быть построен многомерный аналог *метода ломаных* (см. разд. 15.3), приводящий к построению системы гиперконусов, мажорирующих поверхность отклика снизу и непрерывно приближающихся к ней.

Вопросы для самоконтроля

1. В чем заключаются особенности поиска глобального экстремума для полимодальных моделей?
2. В чем заключается процедура многократного спуска при поиске глобального экстремума?
3. Дайте характеристику метода «тяжелого шарика».
4. В чем суть самообучения при статистическом поиске?
5. Дайте характеристику метода статистического поиска с направляющим конусом.

16.7. ПОИСК ЭКСТРЕМУМА В ОБСТАНОВКЕ ПОМЕХ

В предыдущих разделах почти полностью игнорировались ошибки измерений, искажающие величину критерия качества в изучаемой точке факторного пространства. Такой подход оправдан при поиске оптимума математических моделей, так как ошибки счета малы и ими можно пренебречь. Однако при использовании прямых методов для оптимизации экспериментальных моделей с этими ошибками необходимо считаться. Можно выделить следующие источники ошибок (помех).

1. Шумы объекта, связанные как с влиянием неучтенных в модели факторов, так и с возможным временным дрейфом параметров объекта, в частности, за счет его старения.

2. Шумы устройства измерения, включающие в себя собственно ошибки измерения и ошибки передачи результатов измерений, в частности, в измерительно-вычислительных системах.

3. Ошибки обработки информации (ошибки округления и представления данных).

При измерении функции качества могут возникать как систематические, так и случайные ошибки. Однако систематические ошибки фактически эквивалентны смещению поверхности отклика в факторном пространстве и, хотя и искажают величину критерия качества, вообще говоря, не вызывают погрешности при определении экстремальной точки. Более существенны случайные ошибки, результатом которых может быть ошибочное определение экстремальной точки.

Формально введение случайной помехи эквивалентно замене истинного значения функции качества $W(x)$ приближенным случайным значением $\tilde{W}(x)$:

$$\tilde{W}(x) = W(x) + \varepsilon, \quad (16.41)$$

включающим в себя случайную составляющую ε . Соотношение (16.41) позволяет строить простейшие математические модели для изучения эффективности методов поиска в обстановке помех. Так, например, можно построить случайный аналог квадратичной формы (16.32), функции Розенброка (16.28) и т. д.

При поиске оптимума при наличии помех обычно предполагается, что $\tilde{W}(x)$ является несмещенной оценкой $W(x)$ (см. гл. 3):

$$M(\tilde{W}(x)) = W(x), \quad (16.42)$$

а разброс случайных наблюдений ограничен:

$$\sigma^2(\tilde{W}(x)) = M[\tilde{W}(x) - W(x)]^2 < C. \quad (16.43)$$

Возможно несколько путей организации поиска оптимума в обстановке помех.

Во-первых, можно добиться подавления помех за счет *многократных* измерений функции качества в каждой из исследуемых

точек. В силу (3.5) среднее n измерений $\bar{W}(\mathbf{x})$ в точке $\mathbf{x}$ является хорошей оценкой истинной величины $W(\mathbf{x})$:

$$\lim_{n \rightarrow \infty} \bar{W}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \tilde{W}_i(\mathbf{x}) \Rightarrow W(\mathbf{x}). \quad (16.44)$$

Однако при этом существенно увеличиваются затраты на эксперимент.

Во-вторых, можно ввести *фильтрацию* результатов измерения для подавления случайной составляющей. Простейшей является так называемая пороговая фильтрация, основанная на том, что случайные помехи забивают слабые, но не могут существенно исказить сильные сигналы. Алгоритм поиска оптимума (16.7) при пороговой фильтрации можно записать в виде

$$x_i^{n+1} = \begin{cases} x_i^n - \alpha_i \frac{\partial W(\mathbf{x}^n)}{\partial x_i} & \text{при } |W^{n+1} - W^n| \geq \delta; \\ x_i^n & \text{при } |W^{n+1} - W^n| < \delta. \end{cases} \quad (16.45)$$

Таким образом, рабочий шаг происходит только в том случае, когда выигрыш от этого шага превосходит некоторую минимальную величину. Аналогично можно ввести фильтрацию и в другие методы оптимизации. Пороговая фильтрация хорошо сочетается со статистическими алгоритмами поиска со случайной длиной шага (см. разд. 16.3): при длинном шаге вероятность существенного улучшения критерия качества выше, чем при коротком шаге. Однако введение пороговой фильтрации уменьшает быстродействие поиска за счет неиспользованных шагов, а в многомерном случае может привести к неправильному выбору направления движения. В частности, очевидно, что введение пороговой фильтрации исключает смещение в направлении существенных переменных в овражной ситуации (см. разд. 16.4). Непригодна пороговая фильтрация также в окрестности экстремума, где приращение функции качества невелико.

Разработан ряд схем поиска оптимума при наличии помех, не требующих повторных измерений функции качества в каждой точке факторного пространства или пропуска части шагов. В основе всех этих методов лежит предположение о возможности *компенсации* случайных ошибок, допущенных на различных шагах поиска. Однако для того чтобы такая компенсация была возможна, поиск, с одной стороны, не должен *слишком быстро* сходиться к оптимуму, а, с другой стороны, новая (возможно ошибочная!) информация не должна слишком сильно влиять на поиск.

Необходимо отметить, что некоторые, уже рассмотренные алгоритмы поиска оптимума обладают определенными помехозащитными свойствами. К ним, в частности, относятся методы с «инерцией» (метод тяжелого шарика, статистические методы с самообучением), а также методы с «осторожным» выбором направления

движения в факторном пространстве (определение градиента на основе парной пробы, метод статистического градиента).

Ряд методов определения положения экстремума при наличии помех получил название методов **стохастической аппроксимации**. Здесь термин *аппроксимация* указывает на приближенный характер определения экстремума, а термин *стохастический* — на случайный характер ошибок эксперимента. Процедуры стохастической аппроксимации в целом подобны обычным процедурам поиска экстремума, но оснащены механизмом подавления помех.

Одной из наиболее эффективных процедур стохастической аппроксимации является **процедура Кифера—Вольфовица**, предусматривающая движение в антиградиентном направлении при специальном законе изменения величины шага.

Очередная рабочая точка поиска определяется соотношением

$$x_i^{n+1} = x_i^n - a_{in} \frac{W^i(x^n) - W^{-i}(x^n)}{2h_{in}};$$

$$W_n^i = W(x_1^n, x_2^n, \dots, x_{i-1}^n, x_i^n + h_{in}, x_{i+1}^n, \dots, x_k^n);$$

$$W_n^{-i} = W(x_1^n, x_2^n, \dots, x_{i-1}^n, x_i^n - h_{in}, x_{i+1}^n, \dots, x_k^n),$$
(16.46)

где a_n — n -й член последовательности положительных чисел; h_n — n -й пробный шаг при оценке градиента.

В пределе длины рабочего и пробного шагов должны стремиться к нулю:

$$\lim_{n \rightarrow \infty} a_n = 0; \quad \lim_{n \rightarrow \infty} h_n = 0,$$
(16.47)

что обеспечивает асимптотическую сходимость процесса.

Чтобы величина суммарной коррекции начального положения (начальной точки) поиска была достаточна для достижения экстремума и осуществлялось подавление помех, должны выполняться условия

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} \left(\frac{a_n}{h_n} \right)^2 < \infty.$$
(16.48)

В качестве последовательности a_n можно использовать гармоническую последовательность

$$a_n = \frac{1}{n}; \quad n = 1, 2, \dots$$
(16.49)

Метод стохастической аппроксимации обеспечивает сходимость поиска к экстремальной точке, если поверхность отклика в окрестности экстремума близка к выпуклой, а скорость изменения целевой функции в этой окрестности ограничена. На практике эти условия, как правило, выполняются.

1. В чем заключается влияние помех при оценке оптимума функции качества?
2. В чем суть пороговой фильтрации при оптимизации «шумящих» объектов?
3. В чем суть метода Кифера — Вольфовица? Чем отличается метод Кифера — Вольфовица от обычного градиентного метода?

16.8 КОМБИНИРОВАННЫЕ СТРАТЕГИИ ПОИСКА ЭКСТРЕМУМА

Каждый из рассмотренных методов оптимизации обладает своими достоинствами и недостатками, хорошо «работает» в одних ситуациях и плохо в других. Поэтому при оптимальном проектировании сложных систем, в частности ДУ ЛА, редко применяются «чистые» оптимизационные стратегии. Среди методов, которые все-таки могут быть непосредственно использованы при решении задач оптимизации, нужно отметить метод наискорейшего спуска, статистический поиск с наказанием случайностью, симплексный поиск, овражный метод Гельфанда—Цетлина. Правда, последний из названных методов по существу представляет собой *комбинированную* стратегию. Чистые стратегии поиска оказываются эффективными при исследовании относительно простых моделей для задач самых верхних или самых нижних уровней иерархического описания ДУ ЛА.

Для решения большинства задач оптимального проектирования сложных систем используются *комбинированные* стратегии, предусматривающие совместное использование двух или нескольких методов оптимизации. При этом выгодно сочетать методы, отличающиеся своими свойствами. Например, весьма эффективны комбинации статистических методов поиска, позволяющих при большой размерности задачи быстро достигнуть окрестности экстремума, и детерминированных методов, более эффективных в этой окрестности. Другая типичная комбинация предусматривает сочетание линейных методов на первых этапах исследования и методов нелинейной аппроксимации окрестности экстремума. Возможна более сложная комбинированная стратегия, предусматривающая последовательное использование статистических, линейных детерминированных методов и нелинейной аппроксимации. Каждая комбинированная стратегия предполагает существование правил перехода от одного метода оптимизации к другому.

Это правило всегда связано с характером изменения от шага к шагу величины функции качества, а также с изменением величины самого шага. Например, чрезмерное уменьшение шага может служить сигналом о вырождении метода поиска. Для анализа ситуации можно использовать вектор-столбец частных производных функции качества. В частности, деление этих производных по величине на две группы указывает на наличие оврага.

Разработаны и более сложные комбинированные методы. Так, например, успешно используется комбинированный метод, предусматривающий:

1) выбор некоторой последовательности случайных начальных точек

$$x_1^0, x_2^0, \dots, x_n^0 \in X^*; \quad (16.50)$$

2) спуск из каждой начальной точки x_i^0 к точке минимума x_i^* с помощью последовательно используемых статистического поиска с наказанием случайностью и наискорейшего спуска;

3) исследование каждой из точек x_i^* с помощью ряда контрольных *больших* шагов в случайных направлениях с целью преодоления локального экстремума;

4) последовательное исследование получаемых точек x_i^* , включающее в себя проверку тождественности точек $|x_i^* - x_{i+1}^*| < \varepsilon$, соединение нетождественных точек прямой и одномерный поиск с целью определения наилучшей для этой прямой точки x_{i+1}^{**} , заменяющей в дальнейших исследованиях точки x_i^* и x_{i+1}^* .

Рассмотренный метод, конечно, отличается высокой трудоемкостью. Однако он обеспечивает надежное определение экстремума даже при очень сложном характере поверхности отклика.

Следующим шагом по отношению к сложным комбинированным стратегиям оптимизации является разработка *многоуровневых* систем оптимизации, включающих в себя набор алгоритмов оптимизации, систему оценок, обеспечивающих выбор алгоритма в зависимости от возникающей в процессе поиска ситуации, и систему изменения этих оценок по мере накопления опыта. Можно показать, что сложные многоуровневые комбинированные стратегии поиска весьма эффективны при изучении сложных моделей и при высоких требованиях к точности результата. В то же время для более простых случаев применение таких стратегий неоправдано, так как приводит к изменению затрат на поиск.

Вопросы для самоконтроля

1. В чем смысл использования комбинированных стратегий оптимизации?
2. Приведите примеры комбинированных стратегий оптимизации.

16.9. ЧУВСТВИТЕЛЬНОСТЬ ОПТИМАЛЬНЫХ РЕШЕНИЙ

Как ни важно найти оптимальное решение задачи, знание этого решения само по себе обычно не может удовлетворить исследователя. Как правило, необходимо определить *изменения* этого оптимального решения, возникающие при отклонении всех или некоторых переменных от их оптимальных значений.

Во-первых, часто точное решение, полученное при оптимизации, по тем или иным причинам просто не может быть реализовано и поневоле приходится удовлетворяться решением приближенным. Такая ситуация, в частности, всегда возникает при замене в модели дискретных переменных (число ступеней компрессора, возможные конструктивные материалы с их свойствами) фиктивными непрерывными переменными.

0,01251	0,38489	0,77362	0,33847	0,04503	0,45406	0,97184	0,84302	0,49830	0,13283
0,14233	0,94557	0,12333	0,64203	0,04821	0,66178	0,08113	0,18590	0,32573	0,71368
0,71028	0,24312	0,54895	0,40125	0,11756	0,00385	0,05251	0,92255	0,40014	0,14853
0,66259	0,67058	0,76323	0,09174	0,93838	0,44475	0,44756	0,07823	0,05413	0,66946
0,63541	0,29415	0,27768	0,43806	0,68132	0,06118	0,17577	0,96298	0,44882	0,26861
0,43918	0,36452	0,47792	0,19881	0,45377	0,03247	0,75243	0,04683	0,32783	0,57336
0,11105	0,32913	0,95743	0,64329	0,40940	0,04954	0,63010	0,53691	0,70221	0,73577
0,45475	0,16364	0,33133	0,04703	0,99861	0,62628	0,08177	0,28474	0,28191	0,77806
0,98647	0,21440	0,26455	0,28066	0,82639	0,72727	0,38108	0,94781	0,78069	0,41219
0,80126	0,08051	0,38302	0,98591	0,51085	0,39761	0,55420	0,42083	0,63524	0,10049
0,23232	0,91017	0,97236	0,10145	0,41097	0,52985	0,70297	0,64579	0,84929	0,06810
0,62321	0,72467	0,07284	0,22896	0,39797	0,20041	0,02495	0,19095	0,12664	0,81495

Продолжение

0,2005	0,4043	1,0828	—0,8888	—0,1557	0,0723	1,4399	—1,2982	1,2249	—1,5188	—0,2401
1,1609	0,4686	—1,3566	0,8115	—0,2033	—0,8472	0,6264	—0,3035	0,4177	0,0283	0,2634
0,5864	1,0423	—0,5446	0,5405	1,2237	—1,1630	0,3172	1,1409	2,8545	—1,2804	0,2446
0,1425	1,8818	—0,0831	—1,1929	1,3876	—0,5332	—0,5412	—0,3315	1,7019	—0,2011	1,7287
0,9516	—1,1147	—0,5863	—1,3596	—1,4647	—0,4254	—0,4525	1,1407	—0,8211	—0,0327	—0,6078
1,1922	0,6379	1,1572	0,4167	—1,2384	—2,1401	—0,1673	—0,2092	0,2244	0,1818	2,1936
—0,6690	1,4664	—0,4428	—0,2676	—0,1316	0,6057	0,7770	—0,8648	—1,3384	1,3351	0,8285
—0,9245	—1,8149	—0,3924	—0,6022	—0,7003	1,8800	1,3063	0,1408	0,1878	—0,3766	0,5849
—0,2863	0,7390	0,8319	—0,0572	0,9780	0,6769	0,3513	—0,4552	—0,0541	—2,6891	—1,2694
—0,0077	—0,4428	0,9990	—0,8513	0,2506	—0,5111	1,9888	1,4988	1,3989	—0,3886	—1,4345
0,0904	1,1803	1,7981	0,0093	2,1961	0,5955	1,7748	1,1105	—0,0843	0,1176	—1,2114
1,2809	—0,2736	0,4270	—0,5061	—0,7679	0,8878	0,0959	0,5407	0,7196	0,7349	—0,8690
0,0348	—2,3006	—0,1032	1,1054	0,1451	0,3961	1,3912	—0,8433	0,2175	—1,1073	0,8420
1,5068	0,0033	0,6141	0,2119	0,8373	—0,9567	—0,7831	1,0572	0,7661	—0,7246	1,9777
—0,6200	0,1009	2,3965	—0,7944	—1,9278	—0,0307	0,0564	—0,8212	—0,1925	2,6179	—0,4280
—1,0324	0,5652	—1,3493	0,5456	—0,1932	1,9745	—1,6344	—1,3690	1,3532	—1,5846	0,7267
—2,1421	1,3361	0,5343	0,1161	—0,3836	—0,6722	1,1758	—0,8117	0,2758	—0,9516	—2,1545

Во-вторых, иногда уже имеются разработанные агрегаты, узлы и элементы конструкции или технологические процессы, близкие к оптимальным, и возникает естественное желание использовать эти готовые элементы. Так, при создании новой ДУ естественно использовать наиболее подходящие из уже имеющихся двигателей, агрегатов питания, преобразователей энергии.

В-третьих, приближенный характер математического моделирования, а следовательно, и полученных оптимальных решений, сам по себе вызывает сомнения в необходимости абсолютно точного соблюдения полученных рекомендаций.

Поэтому результаты оптимизации всегда должны быть дополнены анализом *чувствительности* полученных решений, т. е. должна быть определена та «цена», которую придется платить за отклонение варьируемых параметров от их оптимальных значений:

$$\Delta W = W(x^*) - W(x^* + \Delta x). \quad (16.51)$$

Для определения чувствительности оптимального решения строят так называемые *разрезы* поверхности отклика по отдельным переменным.

Наиболее просто построить такие разрезы, зафиксировав все переменные, кроме одной, на оптимальном уровне и изменяя эту последнюю переменную в некотором диапазоне в окрестности оптимума

$$x_{i \text{ опт}} - \Delta x_i \leq x_i \leq x_{i \text{ опт}} + \Delta x_i; \quad x_j = x_{j \text{ опт}}, \quad i \neq j. \quad (16.52)$$

При этом будет получена одномерная кривая $W = W(x_i)$, характеризующая зависимость целевой функции от исследуемой переменной. Процедура построения разреза в этом случае потребует лишь нескольких вычислений функции качества.

Однако очевидно, что такой подход вполне правомочен лишь для сепарабельных моделей, когда можно выделить влияние каждой из переменных. В общем случае, когда все факторы взаимосвязаны, а именно этот случай характерен для ДУ ЛА, отклонение от оптимума *одной* из переменных приводит к отклонению от оптимума и *остальных* факторов. Поэтому при изучении чувствительности сложных моделей целесообразно при каждом заданном значении исследуемого фактора проводить *частную* оптимизацию модели в $(k-1)$ -мерном пространстве, полученном путем исключения фактора x_i , и определять для каждой точки собственное оптимальное значение. Правда, в этом случае существенно возрастают объемы необходимых расчетов.

Очевидно, чувствительность модели, определенная с учетом частной оптимизации, всегда оказывается ниже чувствительности, определенной при фиксированных оптимальных параметрах. Поэтому оценка чувствительности при фиксировании параметров (16.52) может рассматриваться как предельная *гарантированная* оценка возможных потерь.

Вопросы для самоконтроля

1. Почему необходимо исследование чувствительности оптимальных решений?
2. Какие методы построения разрезов поверхности отклика Вам известны?

ПРИЛОЖЕНИЯ

1. Равномерно распределенные на отрезке [0,1] случайные числа ξ_p

0,86515	0,89342	0,18867	0,38132	0,70492	0,03564	0,98088	0,81305	0,41899	0,11641
0,69186	0,78416	0,00607	0,83935	0,07824	0,16325	0,66186	0,29835	0,05498	0,70358
0,41686	0,36538	0,90316	0,66447	0,89571	0,33071	0,39453	0,32795	0,88898	0,82041
0,86522	0,88955	0,50961	0,75120	0,77757	0,53163	0,44326	0,76154	0,39976	0,44292
0,72587	0,33181	0,52452	0,24520	0,66131	0,26623	0,29142	0,32764	0,22148	0,23277
0,90795	0,67248	0,76773	0,64294	0,10274	0,21302	0,75162	0,58846	0,69207	0,83373
0,3393	0,27589	0,04825	0,79130	0,76044	0,62766	0,59157	0,35801	0,51512	0,95873
0,42163	0,12332	0,87113	0,25731	0,40903	0,96929	0,67981	0,54313	0,23775	0,06923
0,93000	0,72664	0,42499	0,19976	0,73244	0,91730	0,65877	0,91090	0,60102	0,99497
0,66155	0,09082	0,97526	0,15218	0,23974	0,17381	0,97496	0,69558	0,66785	0,14726
0,85181	0,94377	0,84778	0,37525	0,59598	0,91682	0,49687	0,39072	0,30649	0,01795
0,88059	0,91641	0,57616	0,75601	0,51435	0,87653	0,40837	0,07876	0,36694	0,38547
0,66434	0,12311	0,27256	0,49286	0,14783	0,85444	0,59097	0,41675	0,87225	0,23541
0,99224	0,14480	0,80317	0,45420	0,05312	0,85739	0,16021	0,22700	0,52141	0,83991
0,60580	0,23963	0,77898	0,40607	0,04510	0,32882	0,76018	0,35244	0,44904	0,12202
0,46105	0,64737	0,33322	0,53855	0,98006	0,86678	0,17221	0,13280	0,67279	0,25016

—1,4527	1,2464	—1,4222	—1,2432	—0,9137	0,8308	—1,5597	0,1402	—0,3996	0,3107	0,8222
—0,7401	—1,7729	1,1764	—1,5747	—0,5137	—0,9590	—1,3960	0,1267	—2,2602	—0,0893	—1,2937
0,8522	0,4673	—0,1719	—0,1824	1,3322	0,2432	0,2810	—0,3456	—1,5751	1,3521	—0,5502
—0,5972	—0,2171	1,2580	—0,2568	—0,3199	—0,9768	0,5451	2,0035	0,5975	0,0904	1,0333
0,3184	—0,3627	0,9895	0,5588	0,5768	—0,8153	—0,2544	—0,1016	0,0464	0,8055	—0,1767
—1,8206	—0,9108	0,5028	—0,3690	0,8286	2,5666	0,1089	0,4794	—1,1670	—1,6222	1,7626
—2,1394	—0,7298	0,9064	0,2136	—0,8159	—0,0971	—1,9540	1,3062	—1,7540	0,9620	—0,3662
—0,5367	0,4394	1,5037	—0,0006	—1,2330	—0,6119	—0,4258	0,8132	0,0814	—0,8897	0,4426
1,7609	—1,7181	—0,1026	—0,1572	0,8314	0,3077	0,3663	0,8073	1,0686	—0,9519	—0,3837
—0,1418	—1,4312	—1,6218	—0,5767	1,6976	0,4413	0,1807	0,7560	1,4139	0,2610	38,79
—1,1800	0,6535	—0,2647	0,6989	1,1220	—2,3175	—0,5102	0,1999	—0,6202	0,6500	—0,5912

3. Критические значения критерия χ^2

f	$\alpha=0,1$	$\alpha=0,05$	$\alpha=0,01$	$\alpha=0,005$	f	$\alpha=0,1$	$\alpha=0,05$	$\alpha=0,01$	$\alpha=0,005$
1	2,706	3,841	6,635	7,879	13	19,812	22,362	27,688	29,819
2	4,605	5,991	9,210	10,597	14	21,064	23,685	29,141	31,319
3	6,251	7,815	11,345	12,838	15	22,307	24,996	30,578	32,801
4	7,779	9,488	13,277	14,860	16	23,542	26,296	32,000	34,267
5	9,236	11,071	15,086	16,750	17	24,769	27,587	33,409	35,718
6	10,645	12,592	16,812	18,548	18	25,989	28,869	34,805	37,156
7	12,017	14,067	18,475	20,278	19	27,204	30,144	36,191	38,582
8	13,362	15,507	20,090	21,955	20	28,412	31,410	37,566	39,997
9	14,684	16,919	21,666	23,589	21	29,615	32,671	38,932	41,401
10	15,987	18,307	23,209	25,188	22	30,813	33,924	40,289	42,796

f	$\alpha=0,1$	$\alpha=0,05$	$\alpha=0,01$	$\alpha=0,005$	f	$\alpha=0,1$	$\alpha=0,05$	$\alpha=0,01$	$\alpha=0,005$
25	34,382	37,652	44,314	46,928	55	68,796	73,311	82,292	85,749
26	35,563	38,885	45,642	48,290	60	74,397	79,082	88,379	91,952
27	36,741	40,113	46,963	49,645	70	85,527	90,531	100,425	104,215
28	37,916	41,337	48,278	50,993	80	96,578	101,879	112,329	116,321
29	39,087	42,557	49,588	52,336	85	102,079	107,522	118,236	122,325
30	40,256	43,773	50,892	53,672	90	107,565	113,145	124,116	128,299
35	46,059	49,802	57,342	60,275	100	118,498	124,342	135,807	140,169
40	51,805	55,758	63,691	66,766	120	140,233	146,567	158,950	163,648
45	57,505	61,656	69,957	73,166	140	161,827	168,613	181,840	186,847
50	53,167	67,505	76,154	79,490					

4. Критические значения статистики Колмагорова — Смирнова λ

	$\alpha=0,1$	$\alpha=0,05$	$\alpha=0,01$	n	$\alpha=0,1$	$\alpha=0,05$	$\alpha=0,01$	n	$\alpha=0,1$	$\alpha=0,05$	$\alpha=0,01$
10	0,36866	0,40925	0,48893	22	0,25283	0,28087	0,33666	50	0,16959	0,18841	0,22604
11	0,35242	0,39122	0,46770	23	0,24746	0,27490	0,32954	55	0,16186	0,17981	0,21574
12	0,33815	0,37543	0,44905	24	0,24242	0,26931	0,32286	60	0,15511	0,17231	0,20673
13	0,33815	0,36143	0,43247	25	0,23768	0,26404	0,31657	65	0,14913	0,16567	0,19877
14	0,31417	0,34890	0,41762	26	0,23320	0,25907	0,31064	70	0,14381	0,15975	0,19167
15	0,30397	0,33760	0,40420	27	0,22898	0,25438	0,30502	75	0,13901	0,15442	0,18528
17	0,28627	0,31796	0,38086	29	0,22117	0,24571	0,29466	85	0,13072	0,14520	0,17421
19	0,27136	0,30143	0,36117	35	0,20185	0,22425	0,26897	95	0,12375	0,13746	0,16493
20	0,26473	0,29408	0,35241	40	0,18913	0,21012	0,25205	100	0,12067	0,13403	0,16081

при $\alpha=0,1$

Число степеней свободы знаменателя f_s	Степени свободы числителя f_s									
	1	2	3	4	5	6	7	10	20	∞
1	39,864	49,500	53,593	55,833	57,241	58,204	58,906	60,195	61,740	63,328
2	8,5263	9,0000	9,1618	9,2434	9,2926	9,3255	9,3491	9,3916	9,4413	9,4913
3	5,5383	5,4624	5,3908	5,3427	5,3092	5,2847	5,2662	5,2304	5,1845	5,1337
4	4,5443	4,3246	4,1908	4,1013	4,0506	4,0098	3,9790	3,9199	3,8443	3,7607
5	4,0604	3,7797	3,6195	3,5202	3,4530	3,4045	3,3679	3,2974	3,2067	3,1050
6	3,7760	3,4633	3,2888	3,1808	3,1075	3,0546	3,0145	2,9369	2,8363	2,7222
7	3,5894	3,2574	3,0741	2,9605	2,8833	2,8274	2,7849	2,7025	2,5947	2,4708
10	3,2850	2,9245	2,7277	2,6053	2,5216	2,4606	2,4140	2,3226	2,2007	2,0554
20	2,9747	2,5893	2,3801	2,2489	2,1582	2,0913	2,0397	1,9367	1,7938	1,6074
∞	2,7055	2,3026	2,0838	1,9449	1,8473	1,7741	1,7167	1,5987	1,4206	1,0000

Число степеней свободы знаменателя f_s	Степени свободы числителя f_s									
	1	2	3	4	5	6	7	10	20	∞
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	241,88	248,01	254,32
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,396	19,446	19,496
3	10,128	9,5521	9,2766	9,1172	9,0135	8,9406	8,8868	8,7855	8,6602	8,5265
4	7,7086	6,9443	6,5914	6,3883	6,2560	6,1631	6,0942	5,9644	5,8025	5,6281
5	6,6079	5,7861	5,4095	5,1922	5,0503	4,9503	4,8759	4,7351	4,5581	4,3650
6	5,9874	5,1433	4,7571	4,5337	4,3874	4,2839	4,2066	4,0600	3,8747	3,6688
7	5,5914	4,7374	4,3468	4,1203	3,9715	3,8660	3,7870	3,6365	3,4445	3,2298
10	4,9646	4,1028	3,7083	3,4780	3,3258	3,2172	3,1355	2,9782	2,7740	2,5379
20	4,3513	3,4928	3,0984	2,8661	2,7109	2,5990	2,5140	2,3479	2,1242	1,8432
∞	3,8415	2,9957	2,6049	2,3719	2,2141	2,0986	2,0096	1,8307	1,5705	1,0000

При $\alpha=0,01$

Продолжение

1	4052,2	4999,5	5403,3	5624,6	5763,7	5859,0	5928,3	6055,8	6208,7	6366,0
2	98,503	99,000	99,166	99,249	99,299	99,332	99,356	99,399	99,449	99,499
3	34,116	30,817	29,457	28,710	28,237	27,911	27,672	27,229	26,690	26,125
4	21,198	18,000	16,694	15,977	15,522	15,207	14,976	14,546	14,020	13,463
5	16,258	13,274	12,060	11,392	10,967	10,672	10,456	10,051	9,5527	9,0204
6	13,745	10,925	9,7795	9,1483	8,7459	8,4661	8,2600	7,8741	7,3958	6,8801
7	12,246	9,5466	8,4513	7,8467	7,4604	7,1914	6,9928	6,6201	6,1554	5,6495
10	10,044	7,5594	6,5523	5,9943	5,6363	5,3858	5,2001	4,8492	4,4054	3,9090
20	8,0960	5,8489	4,9382	4,4307	4,1027	3,8714	3,6987	3,3682	2,9377	2,4212
∞	6,6349	4,6052	3,7816	3,3192	3,0173	2,8020	2,6393	2,3209	1,8783	1,0000

1	16211	20000	21615	22500	23056	23437	23715	24224	24836	25465
2	198,50	199,00	199,17	199,25	199,30	199,33	199,36	199,40	199,45	199,51
3	55,552	49,799	47,467	46,195	45,392	44,838	44,434	43,686	42,778	41,829
4	31,333	26,284	24,259	23,155	22,456	21,975	21,622	20,967	20,167	19,325
5	22,785	18,314	16,530	15,556	14,940	14,513	14,200	13,618	12,903	12,144
6	18,635	14,544	12,917	12,028	11,464	11,073	10,786	10,250	9,5888	8,8793
7	16,236	12,404	10,882	10,050	9,5221	9,1554	8,8854	8,3803	7,7540	7,0760
10	12,826	9,4270	8,0807	7,3428	6,8723	6,5446	6,3025	5,8467	5,2740	4,6385
20	9,9439	6,9865	5,8177	5,1743	4,7616	4,4721	4,2569	3,8470	3,3178	2,6904
∞	7,8794	5,2983	4,2794	3,7151	3,3499	3,0913	2,8968	2,5188	1,9998	1,0000

1. Алемасов В. Е., Дергалин А. Ф., Тишин А. П. Теория ракетных двигателей. М.: Машиностроение, 1980. 531 с.
2. Беллман Р., Дрейфус С. Прикладные задачи динамического программирования. М.: Наука, 1965. 458 с.
3. Вагнер Г. Основы исследования операций. М.: Мир, 1972. Т. I. 335 с.
4. Гришин С. Д., Лесков Л. В., Козлов Н. П. Электрические ракетные двигатели. М.: Машиностроение, 1975. 270 с.
5. Гродзовский Г. Л., Иванов Ю. Н., Токарев В. В. Механика космического полета. Проблемы оптимизации. М.: Наука, 1975. 701 с.
6. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. М.: Статистика, 1973. 388 с.
7. Ивахненко А. Г. Индуктивный метод самоорганизации моделей сложных систем. Киев: Наукова думка, 1982. 295 с.
8. Карплюс У. Моделирующие устройства для решения задач теории поля. М.: ИЛ, 1962. 486 с.
9. Крамер Г. Математические методы статистики. М.: Мир, 1975. 648 с.
10. Кузин Л. Т. Основы кибернетики, Т. I. М., Энергия, 1973. 502 с.
11. Математическая теория оптимальных процессов / Л. С. Понтрягин, В. Г. Болтянский, Р. В. Гамкрелидзе, Е. Ф. Мищенко. М.: Физматгиз, 1961. 382 с.
12. Месарович М., Мако Д., Такахара И. Теория иерархических многоуровневых систем. М.: Мир, 1973. 342 с.
13. Морозов А. И. Физические основы космических электроракетных двигателей. Т. I. М.: Атомиздат, 1978. 325 с.
14. Налимов В. В., Чернова Н. А. Статистические методы планирования экстремальных экспериментов. М.: Наука, 1965. 338 с.
15. Основы автоматизированного проектирования двигателей летательных аппаратов / Д. В. Хронин, В. И. Баулин, Ю. П. Кирпичкин, М. К. Леонтьев. М.: Машиностроение, 1984. 182 с.
16. Оуэн Д. Б. Сборник статистических таблиц. М.: Изд. АН СССР. Мат. ин-т им. В. А. Стеклова, ВЦ АН СССР. 1966. 586 с.
17. Пиавский С. А., Брусов В. С., Хвилон Е. А. Оптимизация параметров многоцелевых летательных аппаратов. М.: Машиностроение, 1974. 167 с.
18. Растринин Л. А. Системы экстремального управления. М.: Наука, 1974. 630 с.
19. Теория и расчет энергосиловых установок космических летательных аппаратов / Л. А. Латышев, Д. Д. Севрук, В. Б. Тихонов, Л. А. Квасников. М.: Машиностроение, 1984. 325 с.
20. Уайлд Д. Дж. Методы поиска экстремума. М.: Наука, 1967. 267 с.
21. Химмельблау Д. Прикладное нелинейное программирование. М.: Мир, 1973. 534 с.
22. Худсон Д. Статистика для физиков. М.: Мир, 1967. 242 с.
23. Шеннон Р. Имитационное моделирование систем. Искусство и наука. М.: Мир, 1978. 418 с.
24. Шевяков А. А. Автоматика авиационных и ракетных силовых установок. М.: Машиностроение, 1970, с. 338—339, 579—583.
25. Шредер Ю. А. Метод статистических испытаний (метод Монте-Карло). М.: Физматгиз, 1962. 331 с.

А

Адекватность модели действительности 14, 29, 146, 154, 169
 Аналогии индикатор 92, 93, 104, 107, 108, 110, 112, 132
 Аналогии коэффициенты 90, 93, 104, 108, 132, 134
 Аналогии принцип 85
 Аналоговая вычислительная машина 85, 87, 121
 Аналоговые модели 85
 — — систем автоматического регулирования 136
 — — физических полей 90, 113, 117
 — — на сетках сопротивлений 103, 115
 — — на сплошных средах 97, 113
 — — нестационарных полей 106, 117
 — — стационарных полей 97
 — — физических систем 120
 Аналоговый решающий блок 121, 122, 123
 — — — дифференцирующий 126
 — — — инерционный 126
 — — — интегрирующий 126
 — — — интегросуммирующий 126
 — — — множительный 128
 — — — с отрицательной обратной связью 124
 — — — с пассивной цепью на входе 123
 — — — с положительной обратной связью 124
 — — — суммирующий 125

В

Вариационное исчисление 203
 Вариация 205
 Варьируемые переменные 186
 Взаимосвязи моделей 18, 20, 21
 Внешнего дополнения принцип 147
 Выбор оптимальной модели идентификации 157
 — — — методом включения 160
 — — — методом всех возможных регрессий 158

— — — методом исключения 159
 — — — методом многорядной селекции моделей 160
 Выборочные статистики 30

Г

Гарантированная оценка 198
 Генерирующие соотношения 173, 175
 Гидро-тепловая аналогия 118
 Гипотеза об адекватности модели действительности 29
 Градиентный метод оптимизации 238

Д

Датчик случайных чисел 68
 Декомпозиция сложной системы на подсистемы 20
 Детерминированные методы оптимизации 223, 235
 Динамическое программирование 209
 Дихатомии метод оптимизации 224
 Дополнительная сумма квадратов 36
 Дробный факторный эксперимент 172
 Дуальности свойство 95
 — — аналоговых моделей 95
 — — физических полей 96

З

Золотого сечения метод оптимизации 228

И

Иерархический метод описания 18
 — — — на основе концепции слоев 20
 — — — на основе концепции страт 19
 — — — на основе концепции эшелонов 20
 Иерархическая многоуровневая система моделей 19, 20
 Интервал неопределенности 200, 219, 234

К

Классификация математических моделей 12

Классификация методов оптимизации 198

Ковариация 38, 154

Ковариационная матрица 154

Код планируемого эксперимента 167

— — — для дробных реплик 174

— — — для полного факторного эксперимента 167

Комбинированные методы оптимизации 268

Конечно-разностная аппроксимация уравнений 54

Координация подсистем 19, 28

— — методом оценки взаимодействия 28

— — методом прогнозирования взаимодействий 28

— — методом согласования взаимодействий 28

Корреляции коэффициент 37

Красные условия 49, 94

— — граничные 49

— — начальные 45, 50

Критерий 29, 31, 187, 190

— адекватности модели действительности 29

— Колмогорова — Смирнова 33

— Пирсона 33

— Стьюдента 32

— Фишера 35

— оптимальности 187, 190

— — векторный 193

— — массовый 192

— — многоцелевых систем 196

— — стоимостной 191

М

Массовых блужданий метод 83

Математическое моделирование 5, 8

Матрицы планирования эксперимента 165

Машинная переменная 121

Машинное уравнение 131

Метод ломаных 232

Методы оптимизации 198

— глобальные 260

— детерминированные 235

— локальные 235

— многомерные 233

— наискорейшего спуска 240

— — — нелокальный 252

— одномерные 217

— пассивные 218, 223

— последовательные 218, 224

— прямые 200, 217

— регулярные 199, 201

— статистические 232, 243

Минимакса принцип 219

Минимаксные стратегии оптимизации 223

Минимизация 187

Множественный коэффициент детерминации 158

Многогруппового учета аргументов метод 161

Модели 8

— геометрические 10

— математические 12

— — аналитические 13, 42

— — аналоговые 85

— структурные 87, 120

— идентификации 12

— квадратичные 13, 60

— конкурирующие 15

— линейные 13, 56

— нелинейные 13, 60

— полиномиальные 60

— позиционные 61

— регулярные 13, 42

— сепарабельные 62

— численные 56

— — детерминированные 56

— — статистические 62, 72

— физические 10

Н

Нелинейная аппроксимация 257

Неорганизованный эксперимент 149

Нормальное уравнение 150

Ньютона метод 202

Ньютона — Рафсона метод 259

О

Область определения модели 15

Овражный метод поиска 250

— — Гельфанда — Цетлина 253

— — конфигураций 254

Ограничения 15

Однородных пар метод оптимизации 223

Определяющий контраст 176

Оптимальный принцип Белмана 209

Ортогональное планирование эксперимента 168, 181

Остаточная сумма квадратов 32, 147, 150

Отсеченный метод оптимизации 236

Ошибок матрица 154

П

Паретовское множество оптимальных решений 194

Планирование эксперимента 163

— — композиционное 170, 181

План Хартли 180

Поверхность отклика 15

— — полимодальная 17

— — унимодальная 16

Поиск оптимума по многограннику 242

Поиск оптимума при наличии помех 265

Покоординатный спуск 237
Полный факторный эксперимент 163
Полуреплика полного факторного эксперимента 173
Принцип максимума Понтрягина 203, 207

Р

Распределение случайных величин 63
— — — биномиальное 66
— — — гамма 66
— — — нормальное 65
— — — Пуассона 66
— — — равномерное 65
— — — экспоненциальное 66
Регрессия 141
— линейная 141
— нелинейная 142
Регуляризации критерии 147
Рототабельное планирование 172

С

Сильвестра критерий 202
Симплекс метод оптимизации 212
Синтез сложных систем 26
Системный подход 8, 18, 26
Система автоматизированного проектирования 5
Сканирование 218, 230, 261
— детермированное 218, 230, 233
— случайное 243, 261
Сложные системы 8
— — иерархические 18
Случайных блужданий метод 81
Статистический метод оптимизации 243, 256
— — — по статистическому градиенту 248
— — — с возвратом 245
— — — с наказанием случайностью 247

— — — с направляющим конусом 263
— — — с пересчетом 246
— — — с самообучением 256, 262
Статистических испытаний метод 62
Стохастическая аппроксимация 267
Структурные схемы для АВМ 130, 134
Сходственные величины 90, 129

Т

Тяжелого шарика метод оптимизации 261

У

Управляющие параметры 9, 140, 187

Ф

Фибоначчи числа 227
Физические поля 48, 90
— — стационарные 48, 97
— — нестационарные 49, 106
Функциональное уравнение Беллмана 209

Ц

Цикл проектирования сложной системы 26

Ч

Чувствительность оптимального решения 269

Ш

Штрафа константа 188
Штрафа функции 188
Шумы в задачах оптимизации 187, 190

Э

Электрическое моделирование 87
Электролитическая ванна 97, 101
Электропроводная бумага 97, 101

Введение	3
<i>Часть I.</i>	
ОБЩИЕ ПРИНЦИПЫ ПОСТРОЕНИЯ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ДВИГАТЕЛЕЙ ЛЕТАТЕЛЬНЫХ АППАРАТОВ	5
Глава 1. Математические модели в системах автоматизированного про- ектирования	5
1.1. Система автоматизированного проектирования.	5
1.2. Математические модели двигательной установки и ее элементов	8
1.3. Классификация математических моделей	12
1.4. Геометрическое представление математических моделей	15
Глава 2. Иерархические принципы построения математических моделей	18
2.1. Иерархический подход к описанию ДЛА	18
2.2. Декомпозиция математического описания	20
2.3. Взаимосвязи моделей. Цикл проектирования	26
Глава 3. Статистические методы проверки адекватности математиче- ских моделей	29
3.1. Основные положения теории проверки моделей	29
3.2. Критерии согласия (χ^2 и Смирнова — Колмогорова)	33
3.3. Анализ дисперсий. Критерий Фишера	35
3.4. Оценка роли переменных. Коэффициент корреляции	37
3.5. Оценка адекватности модели сложной системы по данным об адек- ватности подсистем	39
<i>Часть II.</i>	
РЕГУЛЯРНЫЕ МАТЕМАТИЧЕСКИЕ МОДЕЛИ ДВИГАТЕЛЕЙ ЛЕТА- ТЕЛЬНЫХ АППАРАТОВ	42
Глава 4. Математические модели аналитического типа	42
4.1. Функция и функционал	42
4.2. Математические модели в виде обыкновенных дифференциальных уравнений	45
4.3. Модели, заданные в виде уравнений в частных производных	48
Глава 5. Численные детерминированные модели	56
5.1. Линейные детерминированные модели	56
5.2. Нелинейные детерминированные модели	60

Глава 6. Статистические модели. Метод статистических испытаний . . .	62
6.1. Общие принципы построения и исследования статистических моде- лей	62
6.2. Статистические модели элементов ДЛА	72
<i>Часть III.</i>	
АНАЛОГОВЫЕ МОДЕЛИ И АНАЛОГОВОЕ МОДЕЛИРОВАНИЕ ЭЛЕ- МЕНТОВ ДВИГАТЕЛЕЙ ЛЕТАТЕЛЬНЫХ АППАРАТОВ	85
Глава 7. Общие принципы аналогового моделирования	85
Глава 8. Аналоговое моделирование физических полей в элементах дви- гателей летательных аппаратов	90
8.1. Общие принципы моделирования полей	90
8.2. Модели стационарных полей	97
8.3. Модели нестационарных полей	106
8.4. Примеры моделирования физических полей в элементах ДЛА	113
8.5. Неэлектрические аналоговые модели физических полей	117
Глава 9. Аналоговое моделирование физических систем	120
9.1. Решающие элементы	120
9.2. Решение обыкновенных дифференциальных уравнений на аналого- вых вычислительных машинах	129
9.3. Исследование систем автоматического регулирования	136
<i>Часть IV.</i>	
МОДЕЛИ И МЕТОДЫ ИДЕНТИФИКАЦИИ ДВИГАТЕЛЕЙ ЛЕТА- ТЕЛЬНЫХ АППАРАТОВ	140
Глава 10. Общие принципы построения моделей идентификации	140
10.1. Выбор формы модели идентификации	140
10.2. Критерии точности моделей идентификации	146
Глава 11. Модели идентификации ДЛА (неорганизованный эксперимент)	149
11.1. Регрессионный анализ и модели идентификации	149
11.2. Пример построения модели идентификации	155
11.3. Выбор оптимальной модели идентификации	157
Глава 12. Модели идентификации ДЛА (планируемый эксперимент) . .	163
12.1. Полный факторный эксперимент	163
12.2. Дробный факторный эксперимент	172
12.3. Последовательное усложнение моделей идентификации при пла- нируемом эксперименте	179
12.4. Примеры моделей идентификации, построенных по результатам пла- нируемого эксперимента	182
<i>Часть V.</i>	
МЕТОДЫ ОПТИМАЛЬНОГО ПРОЕКТИРОВАНИЯ ДВИГАТЕЛЕЙ ЛЕ- ТАТЕЛЬНЫХ АППАРАТОВ	186
Глава 13. Постановка задач оптимального проектирования	186
13.1. Формулировка задачи оптимального проектирования	186
13.2. Критерии оптимальности	190
13.3. Многокритериальные задачи оптимизации	193
13.4. Оптимизация многоцелевых двигателей ЛА	196
13.5. Классификация методов оптимизации	198
	287

Глава 14. Регулярные методы оптимизации	201
14.1. Определение экстремума дифференцируемой функции	201
14.2. Вариационное исчисление. Принцип максимума Понтрягина	203
14.3. Динамическое программирование	209
14.4. Поиск оптимума линейных моделей (симплекс-метод)	212
Глава 15. Прямые методы оптимизации ДЛА (одномерный поиск) . . .	217
15.1. Принцип минимакса. Оптимальный выбор двух опорных точек на интервале неопределенности	219
15.2. Минимаксные стратегии поиска оптимума	223
15.3. Методы оптимизации полимодальных моделей	230
Глава 16. Прямые методы оптимизации ДЛА (многомерный поиск) . . .	233
16.1. «Проклятие» размерности	233
16.2. Детерминированные локальные методы поиска экстремума с линейной тактикой	235
16.3. Статистические локальные методы поиска экстремума	243
16.4. Поиск оптимума в «овражной» ситуации	250
16.5. Нелинейная аппроксимация поверхности отклика	257
16.6. Исследование полимодальных моделей. Поиск глобального экстремума	260
16.7. Поиск экстремума в обстановке помех	265
16.8. Комбинированные стратегии поиска экстремума	268
16.9. Чувствительность оптимальных решений	269
Приложения	271
1. Равномерно распределенные на отрезке $[0, 1]$ случайные числа ξ_p . .	271
2. Нормально распределенные случайные числа $\xi_N (M(\xi_N)=0; \sigma(\xi_N)=1)$.	273
3. Критические значения критерия χ^2	275
4. Критические значения статистики Колмогорова — Смирнова λ_α . .	277
5. Критические значения для распределения Фишера (F) при $\alpha=0,1; \alpha=$ $=0,05; \alpha=0,01; \alpha=0,005$	278
Список литературы	282
Предметный указатель	283

УЧЕБНОЕ ИЗДАНИЕ

ЧУЯН Ростислав Константинович

МЕТОДЫ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ДВИГАТЕЛЕЙ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

Редактор *Суворова И. А.*

Художественный редактор *В. В. Лебедев*

Технический редактор *Т. И. Андреева*

Корректоры *Н. Г. Богомолова, Л. Л. Георгиевская*

ИБ № 5679

Сдано в набор 16. 06. 87. Подписано в печать 30.11.87. Т-21557. Формат 60×90¹/₁₆. Бумага кн.-журн. № 2. Гарнитура литературная. Печать высокая. Усл. печ. л. 18,0. Усл. кр.-отт. 18,0. Уч.-изд. л. 18,74. Тираж 4000 экз. Заказ 1532. Цена 1 р. 10 к.

Ордена Трудового Красного Знамени издательство «Машиностроение», 107076, Москва, Стромынский пер., 4.

Московская типография № 8 Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли, 101898. Москва, Центр, Хохловский пер., 7.