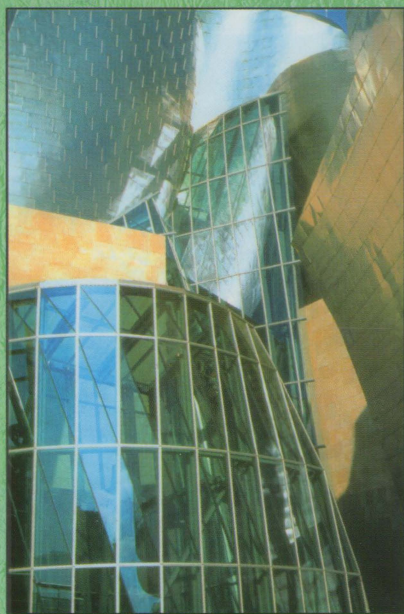


**НОВАЯ
УНИВЕРСИТЕТСКАЯ
БИБЛИОТЕКА**

ВВЕДЕНИЕ В МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ



УЧЕБНОЕ ПОСОБИЕ

**Новая
Университетская
Библиотека**

Каменистые тропы науки – это горы литературы,
уступы книг, которые нужно прочесть, усвоить.
Но книги – это путеводитель, по которому можно
ориентироваться на дорогах науки.

А.Я. Яншин, академик

ВВЕДЕНИЕ В МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Учебное пособие

*Допущено Министерством образования
Российской Федерации в качестве учебного пособия
для студентов высших учебных заведений,
обучающихся по направлению
511200 – «Математика. Прикладная математика»*



Москва
«Логос»
2007

УДК 510.6:683.3:531

ББК 22 в.6

В24

Рецензенты

Доктор физико-математических наук, профессор *А.Р. Абдуллаев*

Доктор технических наук, профессор,

академик РАН *В.П. Матвеевко*

Авторы

В.Н. Ашихмин, М.Б. Гитман, И.Э. Келлер, О.Б. Наймарк,

В.Ю. Столбов, П.В. Трусов, П.Г. Фрик

В24 Введение в математическое моделирование: Учеб. пособие / Под ред. П.В. Трусова. — М.: Университетская книга, Логос, 2007. — 440 с.

ISBN 978-5-98704-037-X

Рассмотрены основные понятия, определения, положения и подходы математического моделирования, представлена классификация математических моделей. Описаны основные этапы, технология построения математических моделей, приведены простые примеры ее применения. Анализируются особенности математического моделирования в условиях различных типов неопределенности, разработки моделей с применением структурного и имитационного подходов. Особое внимание уделено анализу линейных и нелинейных моделей, выявлению их качественных различий. Приведены сведения о современных разделах математики (вейвлеты, фракталы, клеточные автоматы), эффективно используемых при решении различных проблем нелинейной физики. Каждый из разделов снабжен перечнем заданий для самостоятельной работы.

Для студентов высших учебных заведений, обучающихся по направлению 510000 — «Естественные науки и математика» и специальности 010200 — «Прикладная математика». Представляет интерес для специалистов в области математического моделирования физико-механических процессов и явлений.

ББК 22в.6

ISBN 978-5-98704-037-X

© Авторы, указанные на обороте титульного листа, 2004

© «Университетская книга», 2006

© «Логос», 2007

ПРЕДИСЛОВИЕ

...Становится ясно, что в следующем веке понадобятся не только эксперты по некоторым аспектам отдельных стадий избранных процессов. Понадобятся специалисты по решению проблем... По-видимому, междисциплинарность будет в цене. А в институтах будут стараться учить не «предметам», а стилям мышления.

Г.Г.Малинецкий

Понятие «математическое моделирование» в последние два–три десятилетия является едва ли не самым распространенным в научной литературе, по крайней мере в естественно-научной и технической. Сегодня трудно представить себе проектную или конструкторскую организацию, не использующую в своей практике в той или иной мере математические модели. Все более распространенным и эффективным становится применение математического моделирования в научных исследованиях. Интенсивно разрабатываются математические модели в экономике, управлении, истории, биологии и многих других областях знаний. Подавляющее большинство известных авторам диссертационных работ по специальностям естественно-научных и инженерно-технических направлений связано с разработкой и использованием соответствующих математических моделей. В последние 10–15 лет эта тенденция получает все более широкое распространение при подготовке дипломных работ выпускников вузов. Математическое моделирование используется в учебно-исследовательской работе учащихся физико-математических школ и лицеев.

Можно констатировать, что математическое моделирование в последние десятилетия оформилось в отдельную междисциплинарную область знаний с присущими ей объектами, подходами и методами исследования. В связи с этим все более актуальной становится задача целенаправленной подготовки специалистов-«матмодельеров» в вузах

различного профиля, в рамках различных направлений и специальностей. Кроме того, по мнению авторов, эту подготовку, формирование соответствующего «стиля мышления» можно (а возможно, и необходимо) начинать со старших классов общеобразовательной школы. Определенный положительный опыт подобной работы с учащимися специализированных физико-математических школ и классов позволяет по крайней мере не отвергать данную гипотезу.

Реализация образовательного процесса по подготовке специалистов-«матмодельеров», естественно, требует наличия соответствующего методического обеспечения. За последние годы издано немало прекрасных монографий, статей, научно-популярных брошюр, пособий, часть из которых цитируется в тексте. К их числу в первую очередь следует отнести работы отечественных ученых: А.А.Самарского, Н.Н.Моисеева, С.П.Курдюмова, Г.Г.Малинецкого и многих других, трудами которых математическое моделирование и превратилось в самостоятельную область знаний. В этих работах, большей частью написанных математиками, достаточно подробно и прозрачно освещены такие вопросы, как предмет, подходы, методы математического моделирования, приведено огромное количество ярких примеров математических моделей. Как правило, в работах этого направления основное внимание уделяется методам исследования собственно математических моделей, качественному анализу решений, новым эффектам в исследуемых процессах и явлениях.

Следует отметить, что в настоящее время в значительной части учебников и учебных пособий по различным дисциплинам включаются некоторые понятия, методы и примеры применения математического моделирования. Здесь обычно используется некоторый набор базовых моделей данной дисциплины или смежных с ней: из этих «кубиков» в дальнейшем строится модель анализируемого процесса. При этом, как правило, базовые модели принимаются в качестве данности, не обсуждается правомерность их использования, область применимости, степень адекватности описания. «За кадром» в большинстве случаев остается и собственно процесс создания математической модели, процесс перехода от «языка природы» к «языку математики».

Понятно, что указанные обстоятельства обусловлены сложившимися в различных областях традициями, спецификой дисциплин, личными склонностями авторов. В то же время нам представляется полезным появление пособия, в котором более детально рас-

крывалась бы «кухня» разработчиков математических моделей. Несмотря на то, что создание любой новой модели – процесс творческий, близкий к искусству, существуют достаточно общие подходы, методы, «инструменты», пригодные для различных предметных областей. Именно этому кругу вопросов – технологии создания математических моделей – будет уделено наибольшее внимание.

Предлагаемое пособие основано на материале специальных семинаров и курсов лекций по математическому моделированию, которые авторы читают студентам специальности «Прикладная математика и информатика» (специализация «Математическое моделирование»), курсов лекций по дисциплине «Концепции современного естествознания», читаемых для ряда инженерно-технических и гуманитарных специальностей Пермского государственного технического университета, а также на собственном опыте авторов в области разработки математических моделей (главным образом физико-механических процессов, что сказалось и на содержании работы). Естественно, широко использовались монографии, статьи, пособия, посвященные данной тематике.

Пособие ориентировано в первую очередь на студентов младших курсов математических и естественно-научных специальностей (главным образом – математиков-прикладников, физиков, механиков), на студентов технических специальностей (инженеров-механиков), а также на учителей и учащихся старших классов физико-математических школ. Книга может представлять интерес для аспирантов и научных сотрудников, специализирующихся в области математического моделирования физико-механических процессов и явлений. Для работы с большинством разделов пособия достаточно знания вузовского курса математики (математический анализ, линейная алгебра и аналитическая геометрия, элементы теории вероятности, теории обыкновенных дифференциальных уравнений и уравнений математической физики), изучаемого обычно на младших курсах. При этом мы стремились сделать пособие «замкнутым», т.е. содержащим по возможности все необходимые понятия, определения и другие сведения.

Глава 1 является вводной, содержащей описание предмета, основные понятия и определения, связанные с моделированием в целом и в частности с математическим моделированием, классификацию моделей. Для самой юной части читателей этот материал может показаться излишним; тем не менее, для дальнейшего чтения необходимо познакомиться хотя бы с основными понятиями и

определениями. Глава 2 посвящена технологии построения математических моделей и представляется нам ключевой для специалистов, занимающихся построением конкретных моделей. В главе 3 мы попытались на простейших примерах, доступных всем отмеченным выше категориям читателей, показать эту технологию в действии (аналогичную цель в той или иной степени преследуют и примеры всех остальных глав). Главы 4 и 5 содержат необходимые сведения из отдельных разделов математики, широко используемых при построении различных математических моделей (системного анализа, теории вероятностей, теории нечетких множеств), а также качественные примеры построения моделей с применением этих знаний. В главе 6, требующей для прочтения достаточной математической подготовки, авторы попытались отразить современные тенденции моделирования, связанные в первую очередь с созданием нелинейных моделей и анализом возникающих при этом качественно новых эффектов. Здесь же отражены некоторые сведения о современных методах (вейвлетах, фракталах), используемых при разработке математических моделей физико-механических процессов. Глава 7 посвящена одному из широко используемых в настоящее время при построении различных моделей «инструментов» — имитационному подходу.

Каждая глава снабжена набором вопросов для самопроверки и/или заданий для самостоятельной работы, выполнение которых мы считаем обязательным элементом работы с предлагаемым пособием. Возможно, некоторые задания покажутся читателю тривиальными; в случае правильности этого предположения выполнение задания не займет много времени. В то же время нельзя исключать возможности, что посылка была не верна. Кроме того, большинство заданий можно сделать сложными настолько, насколько это приемлемо для конкретного читателя. В любом случае, выполнение заданий не приведет к отрицательному результату в усвоении содержания.

Главы 1 и 2 предназначены в основном для преподавателей, читающих курсы «Введение в математическое моделирование» и «Концепции современного естествознания», и студентов специальности «Прикладная математика и информатика». Остальные главы, по мнению авторов, полезны для всех категорий читателей, указанных выше. Студенты и учащиеся старших классов найдут в них сведения о современных методах и подходах, широко применяемых при исследовании нелинейных проблем, примеры построения ма-

тематических моделей и задания для самостоятельной разработки моделей широкого спектра явлений и процессов, различного уровня сложности и глубины. Учителя старших классов физико-математических школ могут использовать материал глав 3–7 при подготовке заданий по естественно-научным (физика, биология) и математическим дисциплинам, а также для внеклассной (кружковой) работы. Преподаватели вузов могут воспользоваться материалом указанных глав при подготовке заданий по курсовым работам по указанным выше дисциплинам, а также в рамках курсов «Уравнения математической физики», «Дифференциальные уравнения в частных производных», «Численные методы» и др.

Следует особо подчеркнуть, что приведенные примеры и задания соотносятся с математическими моделями сложных реальных процессов примерно так же, как холмы с Эверестом. Но пособие и не рассчитано на «покорителей вершин». Нам представляется целесообразным начинать работу в математическом моделировании именно с простейших моделей, не отягченных математическими сложностями. Будущий специалист, избравший этот не самый легкий путь, должен обладать широкими и глубокими знаниями не только во многих разделах «чистой» и «прикладной» математики и информатики, но и аналогичными познаниями в относящихся к объекту моделирования естественно-научных (физике, механике, химии и др.) и/или гуманитарных дисциплинах. Только доскональное знание объекта моделирования, соответствующей предметной области, а также возможность говорить на одном языке со специалистами-«заказчиками» модели позволяет надеяться на успешную реализацию того или иного проекта по созданию математической модели процесса или явления. По крайней мере, такими качествами должен обладать «постановщик» задачи построения той или иной модели.

Надо сказать, что модели сложных процессов и явлений, как правило, разрабатываются коллективами научных сотрудников различных специальностей, «постановщик» обычно является руководителем коллектива. Именно в подобных специалистах ощущается острый (и резко возрастающий) дефицит, а следовательно, подготовке специалистов данного профиля необходимо уделять повышенное внимание. Возможно, хотя бы в малой степени, предлагаемое пособие будет способствовать решению этой задачи.

Авторы выражают искреннюю признательность своим студентам и аспирантам, вопросы и замечания которых в немалой степе-

ни способствовали становлению курса, а в итоге – появлению предлагаемого пособия. Не меньшую признательность и благодарность мы выражаем рецензентам – доктору физико-математических наук, профессору А.Р.Абдуллаеву и академику РАН, доктору технических наук, профессору В.П.Матвеевко, обсуждение с которыми замечаний и предложений по содержанию пособия позволило, как нам представляется, сделать последнее более подходящим «для употребления». Авторы благодарят доцентов кафедры «Математическое моделирование систем и процессов» Н.Д.Няшину и И.Ю.Зубко, написавших соответственно параграфы 6.17 и 7.4.

Глава 1

ОПРЕДЕЛЕНИЕ И НАЗНАЧЕНИЕ МОДЕЛИРОВАНИЯ

С процессом моделирования и различными моделями человек начинает сталкиваться с самого раннего детства. Так, еще не научившись уверенно ходить, малыш начинает играть с кубиками, сооружая из них (на первых порах с помощью старших) различные конструкции (точнее, модели последних). Его окружают разнообразные игрушки: плюшевые, резиновые, металлические, различающиеся размерами, формой, цветом, предназначением и т.д. При этом большинство игрушек в большей или меньшей степени воспроизводят (моделируют) отдельные свойства и форму реально существующих предметов и объектов. В этом смысле такие игрушки также можно рассматривать в качестве моделей соответствующих объектов.

В школе практически все обучение построено на использовании моделей в той или иной форме. Действительно, для знакомства с основными конструкциями и правилами родного языка используются различные структурные схемы и таблицы, которые можно считать моделями, отражающими свойства языка. Процесс написания сочинения следует рассматривать как моделирование некоторого события или явления средствами родного языка. На уроках биологии, физики, химии и анатомии к плакатам и схемам (т.е. моделям) добавляются макеты (тоже модели) изучаемых реальных объектов. На уроках рисования или черчения на листе бумаги либо ватмана создаются модели различных объектов, выраженные изобразительным языком либо более формализованным языком чертежа.

Даже такую трудно формализуемую область знания, как история, также можно считать непрерывной эволюционирующей совокупностью моделей прошлого какого-либо народа, государства и т.д. Устанавливая закономерности в наступлении разных исторических

событий (революций, войн, ускорений либо застоев исторического развития), можно не только выяснить причины, приведшие к данным событиям, но и прогнозировать и даже управлять их появлением и развитием в будущем.

Так, моделями можно считать картину, написанную художником, художественное произведение и скульптуру. Даже жизненный опыт человека, его представления о мире является примером модели. Причем поведение человека определяется моделью, сформировавшейся в его сознании. Психолог или учитель, изменяя параметры такой внутренней модели, способен в отдельных случаях существенно влиять на поведение человека.

Без преувеличения можно утверждать, что в своей осознанной жизни человек имеет дело исключительно с моделями тех или иных реальных объектов, процессов, явлений. При этом один и тот же объект воспринимается различными людьми по-разному, иногда с точностью «до наоборот» (как говорится в известной поговорке, «на вкус и цвет ...»). Это восприятие, мысленный образ объекта также является разновидностью модели последнего (так называемой *когнитивной моделью*) и существенным образом зависит от множества факторов: качества и объема знаний, особенностей мышления, эмоционального состояния конкретного человека «здесь и сейчас» и от множества других, зачастую не доступных рациональному осознанию. Особенно велика роль моделей и моделирования в современной науке и технике.

Можно ли обойтись в технике без применения тех или иных видов моделей? Очевидный ответ — нет! Безусловно, что новый самолет можно построить «из головы» (без предварительных расчетов, чертежей, экспериментальных образцов, т.е. используя только единственную идеальную модель, существующую в мыслях конструктора), но едва ли это будет достаточно эффективная и надежная конструкция. Единственное ее достоинство — уникальность. Ведь даже автор не сможет повторно изготовить точно такой же самолет, так как в процессе изготовления первого экземпляра будет получен некоторый опыт, который обязательно изменит идеальную модель в голове самого конструктора.

Чем более сложным и надежным должно быть техническое изделие, тем большее число видов моделей потребуется на этапе его проектирования.

Как правило, сложные изделия создаются целыми коллективами разработчиков. Вся совокупность применяемых ими разнообразных моделей позволяет сформировать общую для всего коллектива

идеальную модель разрабатываемого изделия. Реальное техническое изделие можно рассматривать как материальную модель (аналог) созданной авторами идеальной модели.

1.1. ЧТО ТАКОЕ МОДЕЛЬ?

*Что значит знать? Вот, друг мой, в чем вопрос.
На этот счет у нас не все в порядке.*

Гёте

Место моделирования среди методов познания

Метод есть душа знания, его жизнь, им порождаются отдельные научные системы, им же они и низвергаются как недостаточно разрешающие задачу научного построения.

С.И. Гессен

Безусловно, моделирование является далеко не единственным методом изучения окружающего мира. Существует целая область знания, которая специально занимается изучением методов познания и которую принято именовать *методологией*. Методология дословно означает «учение о методах» (ибо происходит этот термин от двух греческих слов: *metodos* – метод, путь к чему-либо и *logos* – учение). Изучая закономерности человеческой познавательной деятельности, методология вырабатывает на этой основе методы ее осуществления. Важнейшей задачей методологии является изучение происхождения, сущности, эффективности и других характеристик методов познания.

Понятие «*метод*» означает совокупность приемов и операций практического и теоретического освоения действительности. Метод вооружает человека системой принципов, требований, правил, руководствуясь которыми он может достичь намеченной цели. Владеть методом – это значит знать, каким образом, в какой последовательности нужно совершать те или иные действия для решения различных задач, и уметь реализовать эти знания на практике. Умению грамотно применять тот или иной метод на практике возможно научиться только при решении различных практических задач. Как отмечал С.И. Гессен, «*Овладеть методом науки можно,*

только применяя этот метод к решению конкретных проблем опытного знания».

Учение о методе появилось в науке в XVI в. Ее представители считали правильный метод ориентиром в движении к надежному, истинному знанию. Так, видный философ XVII в. Ф.Бэкон сравнивал метод познания с фонарем, освещающим дорогу путнику, идущему в темноте. Другой известный ученый и философ того же периода Р.Декарт изложил свое понимание метода: «Под методом я разумею точные и простые правила, строгое соблюдение которых ... без лишней траты умственных сил, но постепенно и непрерывно увеличивая знания, способствует тому, что ум достигает истинного познания всего, что ему доступно».

Методы научного познания принято подразделять по степени их общности, т.е. по широте применимости в процессе научного исследования (рис. 1.1), на всеобщие, общенаучные и частнонаучные.

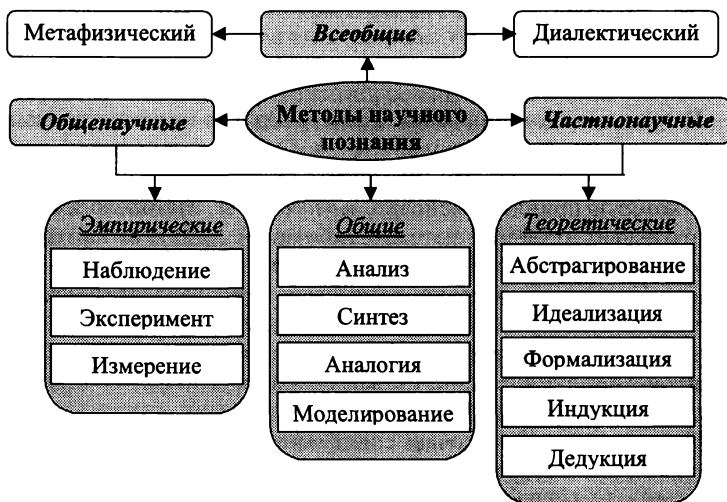


Рис. 1.1. Методы научного познания

Из всеобщих методов в истории познания известны два: диалектический и метафизический. Это общеподобные методы. При метафизическом подходе объекты и явления окружающего мира рассматриваются изолированно друг от друга, без учета их взаим-

ных связей и как бы в застывшем, фиксированном, неизменном состоянии. Диалектический подход, наоборот, предполагает изучение объектов, явлений со всем богатством их взаимосвязей, с учетом реальных процессов их изменения, развития. С середины XIX в., в период третьей научной революции, метафизический метод начал все больше и больше вытесняться из естествознания диалектическим методом.

Общенаучные методы используются в самых различных областях науки, т.е. имеют весьма широкий междисциплинарный спектр применения. Классификация этих методов тесно связана с понятием уровней научного познания. Различают два уровня научного познания: *эмпирический* и *теоретический*. Одни общенаучные методы применяются только на эмпирическом уровне (наблюдение, эксперимент, измерение), другие – только на теоретическом (идеализация, формализация), но есть и такие (например, моделирование), которые используются как на эмпирическом, так и на теоретическом уровне.

Эмпирический уровень научного познания характеризуется непосредственным исследованием реально существующих, чувственно воспринимаемых объектов. На этом уровне путем проведения наблюдений, выполнения разнообразных измерений, постановки экспериментов осуществляется процесс накопления информации об исследуемых объектах, явлениях, производится первичная систематизация получаемых фактических данных в виде таблиц, схем, графиков и т.п. Кроме того, на эмпирическом уровне научного познания – как следствие обобщения научных фактов – возможно формулирование некоторых эмпирических закономерностей.

Теоретический уровень научного исследования присущ рациональной (логической) ступени познания. На данном уровне происходит раскрытие наиболее глубоких, существенных сторон, связей, закономерностей, относящихся к изучаемым объектам, явлениям. Теоретический уровень – более высокая ступень в научном познании. Результатами теоретического познания становятся гипотезы, теории, законы.

Выделяя в научном исследовании два различных уровня – эмпирический и теоретический, не следует, однако, отрывать их один от другого и противопоставлять, поскольку они тесно взаимосвязаны. Эмпирический уровень выступает в качестве основы, фундамента теоретического осмысления научных фактов и получаемых статистических данных. В то же время теоретическое мышление неизбежно опирается на чувственно-наглядные образы (в том числе

схемы, графики и т.п.), с которыми имеет дело эмпирический уровень исследования.

В свою очередь эмпирический уровень научного познания не может существовать без достижений теоретического уровня. Эмпирическое исследование обычно опирается на определенную теоретическую конструкцию, которая определяет направление этого исследования, обуславливает и обосновывает применяемые при этом методы.

К группе *частнонаучных* методов научного познания относятся методы, используемые только в рамках исследований какой-либо конкретной науки или какого-либо конкретного явления. Каждая частная наука (биология, химия, геология и т.д.) имеет свои специфические методы исследования.

Как правило, частнонаучные методы содержат в различных сочетаниях те или иные общенаучные методы познания, базируются на них и могут включать наблюдения, измерения, индуктивные или дедуктивные умозаключения и т.д. Характер сочетания различных методов и его использования зависит от условий исследования, природы изучаемых объектов. Таким образом, частнонаучные методы не оторваны от общенаучных, напротив, тесно связаны с ними, а также со всеобщим диалектическим методом, который как бы преломляется через них. Например, всеобщий диалектический принцип развития проявился в биологии в виде открытого Ч.Дарвином естественно-исторического закона эволюции животных и растительных видов.

К сказанному остается добавить, что любой метод сам по себе еще не предопределяет успеха в познании тех или иных сторон материальной действительности. Важно еще уметь правильно применить его в процессе познания.

Итак, *моделирование* — метод познания окружающего мира, который можно отнести к общенаучным методам, применяемым как на эмпирическом, так и на теоретическом уровне познания. При построении и исследовании модели (см. ниже) могут применяться практически все остальные методы познания.

Определение модели

Научное познание сосредоточено на изучении предметов, явлений и процессов, существующих вне нашего сознания и называемых *объектами исследования* (от лат. *objectum* — предмет).

Понятия модели и моделирования наиболее распространены в сфере обучения, научных исследованиях, проектно-конструкторских работах, в серийном техническом производстве. В каждой из этих областей моделирование имеет свои особенности. Далее моделирование будет рассматриваться главным образом применительно к научным исследованиям. Чаще всего термин «модель» используют для обозначения:

- устройства, воспроизводящего строение или действие какого-либо другого устройства (уменьшенное, увеличенное или в натуральную величину);
- аналога (чертежа, графика, плана, схемы, описания и т.д.) какого-либо явления, процесса или предмета.

К недостаткам термина «модель» следует отнести его многозначность. В словарях можно найти до восьми различных значений данного термина, из которых в научной литературе наиболее распространены два:

- модель как аналог реального объекта;
- модель как образец будущего изделия.

Важную роль при разработке моделей играют *гипотезы* (от греч. hypothesis – основание, предположение), т.е. определенные предсказания, предположительные суждения о причинно-следственных связях явлений, основанные на некотором количестве опытных данных, наблюдений, догадок. Формулирование и проверка правильности гипотез основывается, как правило, на аналогиях.

Аналогия (от греч. analogia – соответствие, соразмерность) – это представление о каком-либо частном сходстве двух объектов, причем такое сходство может быть как существенным, так и несущественным. Существенность сходства или различия двух объектов условна и зависит от уровня абстрагирования (от лат. abstrahere – отвлекать), определяемого конечной целью исследования. Уровень абстрагирования зависит от набора учитываемых параметров объекта исследования. Например, при изучении механических свойств в качестве объектов исследования могут быть выделены материалы из дерева, металла, пластмассы и т.д. В свою очередь материалы из дерева можно подразделить по видам древесины на лиственные и хвойные, лиственные – на «березу», «тополь», «ясень» и т.д.

В данном примере степень абстрагирования снижается при добавлении учитываемых параметров. Следует заметить, что уровень абстрагирования данного объекта всегда устанавливается по отношению к другим объектам.

Гипотезы и аналогии, в определенной мере отражающие реальный, объективно существующий мир, должны обладать наглядностью или сводиться к удобным для исследования логическим схемам. Именно подобные логические схемы, упрощающие рассуждения и логические построения, а также позволяющие проводить эксперименты, приводящие к пониманию явлений природы, называют моделями. Другими словами, *модель* — это объект-заменитель объекта-оригинала, обеспечивающий изучение некоторых интересующих исследователя свойств оригинала.

Под моделью (от лат. modulus — мера, образец, норма) понимают такой материальный или мысленно представляемый объект, который в процессе познания (изучения) замещает объект-оригинал, сохраняя некоторые важные для данного исследования типичные его черты. Процесс построения и использования модели называется моделированием.

Представленное определение является достаточно общим и может трактоваться по-разному. В частности, любое знание можно рассматривать как некоторую идеальную модель природного объекта или явления. В свою очередь, любой искусственный (т.е. созданный человеком) объект или процесс есть материальная модель, построенная на основе соответствующих знаний (идеальных моделей). В этом смысле можно говорить о трех реальностях или трех сферах (что близко к мыслям, высказанным В.И.Вернадским еще в 1922 г.), в которых живет человек.

Первой реальностью является живая и неживая природа, законы развития которой не зависят от человека. Поэтому природные объекты и явления нельзя рассматривать как модели по отношению к человеку. Однако познание и использование человеком природных объектов возможно только через их модели, которые в результате изучения самих объектов также изменяются.

Объекты природы находят свое отражение во второй реальности, или ноосфере, включающей знания, накопленные всем человечеством и практически мало зависящие от конкретного человека. Данная реальность, состоящая из идеальных моделей, зависит от эволюции человечества и изменяется в процессе познания, пополняясь новыми и изменяя старые модели. Можно сказать, что процесс познания в любой области знаний представляет собой непрерывное совершенствование существующих и построение новых моделей исследуемых объектов. Этот ряд моделей (с оптимистической точки зрения) бесконечен.

Наконец, третья реальность, или *техносфера*, которая может рассматриваться как отражение второй реальности, включает все материальные модели, созданные человеком. К составляющим техносферы

ры следует отнести также искусственное разведение животных и растений, их селекцию и (в последнее время) их клонирование. Хотя сами живые организмы являются представителями живой природы и моделями не являются, но процесс их появления управляем человеком на основании некоторых модельных представлений о данных объектах.

В контексте данных рассуждений и приведенного выше определения модели можно сделать вывод о том, что человек в своей жизни в основном занимается знакомством с уже разработанными ранее моделями и созданием на их основе новых идеальных или материальных моделей. Поэтому понятие «человек моделирующий» можно считать тождественным понятию «человек разумный».

Свойства моделей

В настоящее время нет предпосылок к выделению «самых элементарных» и «неделимых» кирпичиков мироздания. Поэтому можно утверждать, что любой объект исследования является бесконечно сложным и характеризуется бесконечным числом параметров. При построении модели исследователь всегда исходит из поставленных *целей*, учитывает только наиболее существенные для их достижения факторы. Поэтому любая модель нетождественна объекту-оригиналу и, следовательно, *неполна*, поскольку при ее построении исследователь учитывал лишь важнейшие с его точки зрения факторы. Другие факторы, несмотря на свое относительно малое влияние на поведение объекта по сравнению с выбранными факторами, в совокупности все же могут приводить к значительным различиям между объектом и его моделью. «Полная» модель, очевидно, будет полностью тождественна оригиналу. Эту мысль хорошо выразили Артуро Розенблют и Норберт Винер [90]: «*наилучшей моделью кота является другой кот, а еще лучше – тот же самый кот*». В то же время, как отметил М. Вартофский [17], при моделировании должно «*исключаться какое то бы ни было самоотнесение, ничто не может быть моделью самого себя*».

Если результаты моделирования удовлетворяют исследователя и могут служить основой для прогнозирования поведения или свойств исследуемого объекта, то говорят, что модель *адекватна* (от лат. *adaequatus* – приравненный) объекту. При этом адекватность модели зависит от целей моделирования и принятых критериев. Учитывая заложенную при создании неполноту модели, можно утверждать, что идеально адекватная модель принципиально невозможна.

В качестве одной из характеристик модели может выступать *простота* (или *сложность*) *модели*. Очевидно, что из двух моделей, позволяющих достичь желаемой цели и получить требуемые результаты с заданной точностью, предпочтение должно быть отдано более простой. При этом адекватность и простота модели далеко не всегда являются противоречивыми требованиями. Учитывая бесконечную сложность любого объекта исследования, можно предположить существование бесконечной последовательности его моделей, различающихся по степени полноты, адекватности и простоты.

В качестве еще одного свойства модели можно рассматривать *потенциальность модели* (от лат. *potentia* — мощь, сила), или предсказательность с позиций возможности получения новых знаний об исследуемом объекте. Данное свойство модели подчеркивается в определении Н.Н. Моисеева [74]: *«Под моделью мы будем понимать упрощенное, если угодно, упакованное знание, несущее вполне определенную, ограниченную информацию о предмете (явлении), отражающее те или иные его свойства. Модель можно рассматривать как специальную форму кодирования информации. В отличие от обычного кодирования, когда известна вся исходная информация, и мы лишь переводим ее на другой язык, модель, какой бы язык она не использовала, кодирует и ту информацию, которую люди еще не знали. Можно сказать, что модель содержит в себе потенциальное знание, которое человек, исследуя ее, может приобрести, сделать наглядным и использовать в своих практических жизненных нуждах»*. По этому же поводу высказываются Т.Тоффоли и Н.Марголус [107]: *«В науке мало пользы от моделей, которые рабски подчиняются нашим желаниям. Мы хотим иметь модели, которые дерзят нам; модели, которые имеют свой собственный ум. Мы хотим получать от моделей больше, чем в них вложили»*. Эти «дерзость», «собственный ум» моделей — есть проявление множества внутренних связей, осознать совместное действие (синергетические эффекты) которых их создатели зачастую не в состоянии (по крайней мере — на стадии разработки). Именно свойство потенциальности (иногда называемое *богатством модели*) позволяет модели выступать в качестве самостоятельного объекта исследования.

В научных исследованиях модели, не обладающие определенной «предсказательностью», едва ли могут считаться удовлетворительными.

Известно немало случаев, когда изучение или использование моделей позволило сделать открытия. В качестве примера можно привести открытие планеты Нептун, положение которой было

предсказано французским астрономом Лаверье на основании расчетов, выполненных с использованием закона всемирного тяготения (т.е. модели) и данных о движении планеты Уран. В наше время только на основании результатов теоретического моделирования открыты «черные дыры» в астрофизике и кварки в физике элементарных частиц (и те, и другие экспериментально подтверждены пока косвенно), высокотемпературный Т-слой в плазме (подтвержден экспериментально), использование которого позволяет значительно повысить коэффициент полезного действия магнито-гидродинамических генераторов, которые в настоящее время рассматриваются как перспективные устройства для получения электрической энергии.

Цели моделирования

Хорошо построенная модель, как правило, доступнее, информативнее и удобнее для исследователя, нежели реальный объект.

Рассмотрим основные цели, преследуемые при моделировании в научной сфере. Самым важным и наиболее распространенным предназначением моделей является их применение при изучении и прогнозировании поведения сложных процессов и явлений. Следует учитывать, что некоторые объекты и явления вообще не могут быть изучены непосредственным образом. Недопустимы, например, широкомасштабные «натурные» эксперименты с экономикой страны или со здоровьем ее населения (хотя и те, и другие с определенной периодичностью ставятся и реализуются). Принципиально неосуществимы эксперименты с прошлым какого-либо государства или народа («*История не терпит сослагательного наклонения*»). Невозможно (по крайней мере, в настоящее время) провести эксперимент по прямому исследованию структуры звезд. Многие эксперименты неосуществимы в силу своей дороговизны или рискованности для человека и/или среды его обитания.

Как правило, в настоящее время всесторонние предварительные исследования различных моделей явления предшествуют проведению любых сложных экспериментов. Более того, эксперименты на моделях с применением ЭВМ позволяют разработать план натурных экспериментов, выяснить требуемые характеристики измерительной аппаратуры, наметить сроки проведения наблюдений, а также оценить стоимость такого эксперимента.

Другое, не менее важное, предназначение моделей состоит в том, что с их помощью выявляются наиболее существенные фак-

торы, формирующие те или иные свойства объекта, поскольку сама модель отражает лишь некоторые основные характеристики исходного объекта, учет которых необходим при исследовании того или иного процесса или явления. Например, исследуя движение массивного тела в атмосфере вблизи поверхности Земли, на основании известных экспериментальных данных и предварительного физического анализа можно выяснить, что ускорение существенно зависит от массы и геометрической формы этого тела (в частности, от величины поперечного к направлению движения сечения объекта), в определенной степени — от шероховатости поверхности, но не зависит от цвета поверхности. При рассмотрении движения того же тела в верхних слоях атмосферы, где сопротивлением воздуха можно пренебречь, несущественными становятся и форма, и шероховатость поверхности.

Конечно, модель любого реального процесса или явления «беднее» его самого как объективно существующего (процесса, явления). В то же время хорошая модель «богаче» того, что понимается под реальностью, поскольку в сложных системах понять всю совокупность связей «разом» человек (или группа людей), как правило, не в состоянии. Модель же позволяет «играть» с ней: включать или отключать те или иные связи, менять их для того, чтобы понять важность для поведения системы в целом.

Модель позволяет научиться правильно управлять объектом путем апробирования различных вариантов управления. Использовать для этого реальный объект часто бывает рискованно или просто невозможно. Например, получить первые навыки в управлении современным самолетом безопаснее, быстрее и дешевле на тренажере (т.е. модели), чем подвергать себя и дорогую машину риску.

Если свойства объекта с течением времени меняются, то особое значение приобретает задача прогнозирования состояний такого объекта под действием различных факторов. Например, при проектировании и эксплуатации любого сложного технического устройства желательно уметь прогнозировать изменение надежности функционирования как отдельных подсистем, так и всего устройства в целом.

Итак, модель нужна для того, чтобы:

1) понять, как устроен конкретный объект: какова его структура, внутренние связи, основные свойства, законы развития, саморазвития и взаимодействия с окружающей средой;

2) научиться управлять объектом или процессом, определять наилучшие способы управления при заданных целях и критериях;

3) прогнозировать прямые и косвенные последствия реализации заданных способов и форм воздействия на объект.

1.2. КЛАССИФИКАЦИЯ МОДЕЛЕЙ

Очевидно, не существует классификации мира, которая бы не была произвольной и проблематичной. Причина весьма проста: мы не знаем, что такое мир... Невозможность постигнуть божественную схему мира не может, однако, отбить у нас охоту создавать наши, человеческие схемы, хотя мы понимаем, что они — временны.

Хорхе Луис Борхес

В действительности вообще нет никаких строго проведенных межей и граней к великой горести всех систематиков.

А.И. Герцен

Отмеченная выше неоднозначность термина «модель», огромное число типов моделирования и их быстрое развитие затрудняют в настоящее время построение логически законченной, удовлетворяющей всех классификации моделей. Любая подобная классификация условна в силу того, что она отражает, с одной стороны, пристрастия авторов, а с другой — ограниченность их знаний в конечном числе областей научного познания. Данную классификацию следует рассматривать как попытку построения некоторого инструмента или модели для исследования свойств и характеристик самого процесса моделирования.

Как было отмечено ранее, моделирование относится к общенаучным методам познания. Использование моделирования на эмпирическом и теоретическом уровнях исследования приводит к делению (условному) моделей на материальные и идеальные.

Материальное моделирование — это моделирование, при котором исследование объекта выполняется с использованием его материального аналога, воспроизводящего основные физические, геометрические, динамические и функциональные характеристики данного объекта. К таким моделям, например, можно отнести использование макетов в архитектуре, моделей и экспериментальных образцов при создании различных транспортных средств.

Идеальное моделирование отличается от материального тем, что оно основано не на материализованной аналогии объекта и моде-

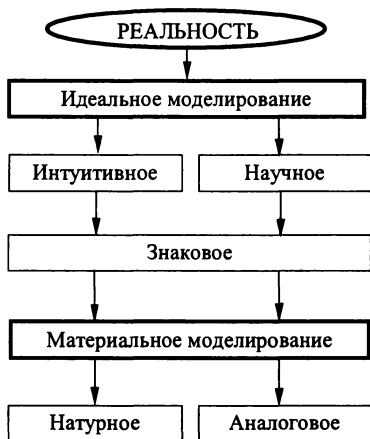


Рис. 1.2. Виды моделирования

ли, а на аналогии идеальной, мыслимой и всегда носит теоретический характер.

Учитывая, что идеальное моделирование является первичным по отношению к материальному (вначале в сознании человека формируется идеальная модель, а затем на ее основе строится материальная), существующие типы моделирования можно представить так, как показано на рис. 1.2. Знакомство с видами моделирования начнем с материального, которое хотя и является вторичным, но более наглядно и просто для понимания.

Материальное моделирование

Основными разновидностями материального моделирования являются натурное и аналоговое. При этом оба вида моделирования основаны на свойствах геометрического или физического подобия. Две геометрические фигуры подобны, если отношение всех соответственных длин и углов одинаковы. Если известен коэффициент подобия – масштаб, то простым умножением размеров одной фигуры на величину масштаба определяются размеры другой, ей подобной геометрической фигуры. Два явления физически подобны, если по заданным характеристикам одного можно получить характеристики другого простым пересчетом, который аналогичен переходу от одной системы единиц измерения к другой. Изучением условий подобия явлений занимается теория подобия.

Натурное моделирование – это такое моделирование, при котором реальному объекту ставится в соответствие его увеличенный или уменьшенный материальный аналог, допускающий исследование (как правило, в лабораторных условиях) с помощью последующего перенесения свойств изучаемых процессов и явлений с модели на объект на основе теории подобия.

К примерам натуральных моделей можно отнести макеты в архитектуре, модели судов в судостроении. Следует отметить, что имен-

но с натуральных моделей судов в середине XIX века моделирование стало развиваться как научная дисциплина, а сами модели — активно использоваться при проектировании новых технических устройств. Середина XIX века связана в судостроении с окончанием эпохи парусных судов и началом эпохи парового флота. Оказалось, что использование паровых машин требует принципиального изменения конструкции судов. В первую очередь это осознали строители военных кораблей. Как известно, в условиях морского сражения время жизни судна зависит главным образом от его маневренности и скорости. Для парусных судов в результате многовекового опыта были выработаны оптимальные сочетания формы корпуса и парусов. Для кораблей с паровой машиной скорость определяется в значительной мере мощностью последней. В тот период тепло для машин получали от сжигания угля в топках котлов. Поэтому чем выше требуемая мощность машины, тем большее количество котлов необходимо использовать и иметь на судне большой запас угля. Все это утяжеляло судно и снижало его скорость, сводя к нулевому эффекту увеличение мощности машины.

Учитывая, что строительство одного крейсера занимало несколько лет, а его стоимость была весьма значительной, можно понять стремление судостроителей найти более быстрый и дешевый (по сравнению с традиционным методом проб и ошибок) способ поиска оптимальных параметров судна. Выход был найден в моделировании. Протягивая в бассейнах небольшие модели будущих судов и измеряя силу сопротивления, конструкторы нашли рациональные решения, как по форме корпуса судна, так и по мощности силовой установки.

В настоящее время методы натурального моделирования находят самое широкое применение в судостроении, авиастроении, автомобилестроении, ракетостроении и других областях. Например, при разработке нового самолета большое значение имеют эксперименты с натурными моделями, испытываемыми в аэродинамической трубе. Проведенные исследования позволяют изучить особенности обтекания фюзеляжа воздушными потоками, найти наиболее рациональную форму корпуса и отдельных узлов. Натурные модели используют и при исследовании причин крупных аварий и катастроф. Активно применяются натурные модели в сочетании с другими методами моделирования (например, компьютерного) при съемке кинофильмов. Так, на съемках американского фильма «Титаник» для сцен гибели корабля было использовано более десяти моделей судна.

Аналоговое моделирование – это моделирование, основанное на аналогии процессов и явлений, имеющих различную физическую природу, но одинаково описываемых формально (одними и теми же математическими соотношениями, логическими и структурными схемами).

В основу аналогового моделирования положено совпадение математических описаний различных объектов. Примерами аналоговых моделей могут служить электрические и механические колебания, которые с точки зрения математики описываются одинаковыми соотношениями, но относятся к качественно отличающимся физическим процессам. Поэтому изучение механических колебаний можно вести с помощью электрической схемы, и наоборот. При некоторых допущениях аналогичными можно считать процессы распространения тепла в теле, диффузии примесей и просачивания жидкости.

К числу интересных примеров можно отнести известную в теории упругости аналогию Прандтля, который показал, что уравнения для функции напряжений (по которой простым дифференцированием по координатам определяются компоненты тензора напряжений) в задаче о кручении стержня произвольного поперечного сечения идентичны уравнениям, определяющим прогиб нерастяжимой мембраны, натянутой на упругий контур той же формы, под действием равномерного давления. Это позволяет заменить отнюдь не простые эксперименты по определению компонент тензора напряжений в скручиваемом стержне простыми измерениями прогибов мембраны.

Модели физического и аналогового типов являются материальным отражением реального объекта и тесно связаны с ним своими геометрическими, физическими и прочими характеристиками. Фактически процесс исследования моделей данного типа сводится к проведению ряда натуральных экспериментов, где вместо реального объекта используется его физическая или аналоговая модель.

Идеальное моделирование

Идеальное моделирование разделяют на два основных типа: интуитивное и научное.

В качестве наиболее яркого примера интуитивной модели окружающего мира можно считать жизненный опыт любого человека. К данному типу моделирования можно отнести умения и знания, накопленные многовековым опытом и передающиеся от по-

коления к поколению (например, умение лечить болезни с использованием трав и приемов народной медицины). Любое эмпирическое (т.е. полученное из эксперимента или в процессе наблюдения) знание без объяснения причин и механизмов наблюдаемого явления также следует считать интуитивным.

Интуитивное моделирование – это моделирование, основанное на интуитивном (не обоснованном с позиций формальной логики) представлении об объекте исследования, не поддающимся формализации или не нуждающимся в ней.

Необходимо подчеркнуть чрезвычайно важную роль интуиции, интуитивных моделей в науке, без них не обходится не одно сколь-нибудь новое знание. Последнее недостижимо только методами формальной логики. В связи с этим интересным представляется высказывание А. де Сент-Экзюпери: *«Теоретик верит в логику. Ему кажется, будто он презирает мечту, интуицию, поэзию. Он не замечает, что они, эти три феи, просто переоделись, чтобы обольстить его, как влюбчивого мальчишку... Они являются ему под именем «рабочих гипотез», «произвольных гипотез», «аналогий», и может ли теоретик подозревать, что, слушая их, он изменяет суровой логике и внемлет напевам муз».*

Научное моделирование – это всегда логически обоснованное моделирование, использующее минимальное число предположений, принятых в качестве гипотез на основании наблюдений за объектом моделирования.

Главное отличие *научного* моделирования от интуитивного заключается не только в умении выполнять необходимые операции и действия по собственно моделированию, но и в знании «внутренних» механизмов, которые используются при этом. Можно сказать, что научное моделирование знает не только, как необходимо моделировать, но и почему так нужно делать.

Необходимо заметить, что в основе любого логического рассуждения лежат гипотезы или аксиомы, принимаемые на веру и не противоречащие имеющемуся опыту или эксперименту. Поэтому можно говорить об интуитивной первооснове любого научного знания. В то же время следует подчеркнуть, что во многих областях знаний, особенно в естественных науках, аксиомы являются, как правило, обобщением огромного количества накопленных эмпирических данных.

В результате познания человек разбирается в причинах тех или иных явлений, отодвигая границы своих интуитивных представлений об окружающем мире. Учитывая бесконечность этого процесса, можно предположить, что интуитивная первооснова у любого научного знания будет присутствовать всегда. О мере интуитивности знания мож-

но судить по числу использованных гипотез и аксиом. В указанном смысле деление моделирования на интуитивное и научное следует признать относительным.

Знаковым называют моделирование, использующее в качестве моделей знаковые изображения какого-либо вида: схемы, графики, чертежи, иероглифы, руны, наборы символов, включающее также совокупность законов и правил, по которым можно оперировать с выбранными знаковыми образованиями и элементами. В качестве примеров таких моделей можно назвать любой язык, например: устного и письменного человеческого общения, алгоритмический, химических формул, живописи, нот для записи музыкальных произведений и т.д. Знаковая форма используется для передачи как научного, так и интуитивного знания. Моделирование с помощью математических соотношений также является примером знакового моделирования.

Интуитивное и научное (теоретическое) моделирование ни в коей мере нельзя противопоставлять одно другому. Они хорошо дополняют друг друга, разделяя области своего применения. Привычная буржуазно-цифровая (знаковая) форма представления научного знания исторически сложилась как технология «передела» готового знания. Создание же качественно нового знания, рождение принципиально новых научных идей не может быть сведено к процессу чистой дедукции, к процессу формально-логического вывода следствий из множества уже открытых, готовых фактов, гипотез и теорий, составляющих информационную базу данных сегодняшней науки. *«Подлинной ценностью, — говорил А. Эйнштейн [116], — является, в сущности, только интуиция. Для меня не подлежит сомнению, что наше мышление протекает, в основном минуя символы (слова) и к тому же бессознательно».*

Не менее определенно в пользу внелогического характера научного творчества высказывался А. Пуанкаре [88]: *«Чистая логика всегда привела бы нас только к тавтологии; она не могла бы создать ничего нового; сама по себе она не может дать начало никакой науке... Для того, чтобы создать арифметику, как и для того, чтобы создать геометрию или какую бы то ни было науку, нужно нечто другое, чем чистая логика. Для обозначения этого другого у нас нет иного слова, кроме слова “интуиция”».*

Даже в самой абстрактной области фундаментальной науки — математике — интуиция играет определяющую роль: *«Вы должны догадаться о математической теореме прежде, чем вы ее докажете; вы должны догадаться об идее доказательства, прежде чем вы его проведете в деталях ...; доказательство открывается ... с помощью догадки» [86].*

Сходные мысли высказывает Ф.Клейн: *«Извечный секрет необычайной продуктивности гения — в его умении находить новые постанов-*

ки задач, интуитивно предугадывать теоремы, приводящие к новым значительным результатам и к установлению важных зависимостей. Не будь новых концепций, новых идей, математика с присущей ей строгостью логических выводов вскоре исчерпала бы себя и пришла в упадок, ибо весь материал оказался бы израсходованным. В этом смысле можно сказать, что математику движут вперед в основном те, кто отмечен даром интуиции, а не строгого доказательства».

Итак, интуитивное знание является генератором нового знания. Однако далеко не все догадки и идеи выдерживают последующую проверку экспериментом и методами формальной логики, собственными научному подходу, выступающему в виде своеобразного фильтра для выделения наиболее ценных знаний.

Говоря о научном моделировании, следует пояснить смысл таких терминов, как «модель» и «теория». В современной научной литературе данные понятия трактуются неоднозначно, а граница между ними размыта. В методологии науки признаны в настоящее время следующие трактовки:

Модель – инструмент, ориентированный в первую очередь на исследование поведения и свойств конкретного объекта в целях управления им или предсказания его свойств.

Теория – более абстрактное, чем модель, средство, основной целью которого является объяснение поведения или свойств не конкретного объекта, а некоторого класса объектов. Можно сказать, что теория содержит конечную или даже бесконечную совокупность конкретных моделей.

Например, для описания течения ньютоновской или нелинейно-вязкой жидкости в канале заданной формы разрабатывается соответствующая модель. При создании конкретной модели используются законы и уравнения соответствующей теории, в данном случае – механики жидкостей. Можно сказать, что модель дает ответы на вопросы: «Каким образом?» и «Почему?» для конкретного объекта, а теория – для целого семейства объектов, обладающих схожими свойствами. Следует, однако, отметить, что при разработке моделей сложных процессов и явлений зачастую приходится использовать понятия и соотношения нескольких теорий, относящихся к различным разделам, дисциплинам и даже областям знаний.

Когнитивные, концептуальные и формальные модели

Как уже отмечалось выше, при наблюдении за объектом-оригиналом в голове исследователя формируется некий мысленный образ объекта, его идеальная модель, которую в научной литерату-

ре принято называть *когнитивной* (мысленной, способствующей познанию) [84]. Формируя такую модель, исследователь, как правило, стремится ответить на конкретные вопросы, поэтому от бесконечно сложного устройства объекта отсекается все ненужное с целью получения его более компактного и лаконичного описания.

Представление когнитивной модели на естественном языке называется *содержательной* моделью.

Когнитивные модели субъективны, так как формируются умозрительно («в голове» исследователя) на основе всех его предыдущих знаний и опыта. Получить представление о когнитивной модели можно только описав ее в знаковой форме. Нельзя утверждать, что когнитивные и содержательные модели эквивалентны, поскольку первые могут содержать элементы, которые исследователь не сможет или не хочет сформулировать. В то же время, если содержательная модель сформулирована кем-то другим или является продуктом коллективного творчества, то ее интерпретация, уровень понимания, степень доверия могут существенно изменяться в зависимости от того или иного интерпретатора. В естественно-научных дисциплинах и в технике содержательную модель часто называют технической постановкой проблемы.

По функциональному признаку и целям содержательные модели подразделяются на описательные, объяснительные и прогностические. *Описательной* моделью можно назвать любое описание объекта. *Объяснительная* модель позволяет ответить на вопрос, почему что-либо происходит. Наконец, *прогностическая* модель должна описывать будущее поведение объекта. Можно заметить, что прогностическая модель не обязана включать в себя объяснительную.

Концептуальной моделью принято называть содержательную модель, при формулировке которой используются понятия и представления предметных областей знания, занимающихся изучением объекта моделирования.

В более широком смысле под концептуальной моделью понимают содержательную модель, базирующуюся на определенной концепции или точке зрения. Выделяют три вида концептуальных моделей: логико-семантические, структурно-функциональные и причинно-следственные.

Логико-семантическая модель является описанием объекта в терминах и определениях соответствующих предметных областей знаний, включающим все известные логически непротиворечивые утверждения и факты. Анализ таких моделей осуществляется сред-

ствами логики с привлечением знаний, накопленных в соответствующих предметных областях.

При построении *структурно-функциональной* модели объект обычно рассматривается как целостная система, которую расчленяют на отдельные элементы или подсистемы. Части системы связываются структурными отношениями, описывающими подчиненность, логическую и временную последовательность решения отдельных задач. Для представления подобных моделей удобны различного рода схемы, карты и диаграммы.

Причинно-следственная модель часто используется для объяснения и прогнозирования поведения объекта. Данные модели ориентированы в основном на следующее: 1) выявление главных взаимосвязей между составными элементами изучаемого объекта; 2) определение того, как изменение одних факторов влияет на состояние компонентов модели; 3) понимание того, как в целом будет функционировать модель и будет ли она адекватно описывать динамику интересующих исследователя параметров.

Формальная модель является представлением концептуальной модели с помощью одного или нескольких формальных языков (например, языков математических теорий, универсального языка моделирования (UML) или алгоритмических языков).

В гуманитарных науках процесс моделирования во многих случаях заканчивается созданием концептуальной модели объекта.

В естественно-научных дисциплинах, как правило, удается построить формальную модель. Таким образом, когнитивные, содержательные и формальные модели составляют три взаимосвязанных уровня моделирования. Перечисленные выше разновидности моделей нельзя рассматривать изолированно одну от другой. «Взаимоотношения» моделей между собой показаны на рис. 1.3.

Взаимовлияние уровней моделирования связано со свойством потенциальности моделей. Создание любой модели сопряжено с появлением новых знаний об исследуемом объекте, что ведет к переоценке и уточнению концепций и взглядов на объект моделирования. Данное обстоятельство приводит, в свою очередь, к пересмотру соответствующих содержательных и когнитивных моделей, обеспечивая спиральное развитие всех уровней моделирования исследуемого объекта.

Следует обратить внимание на такие моменты: если значение содержательных и формальных моделей для процесса познания более или менее осознается исследователями, то роль когнитивных моделей часто недооценивается. Это связано с субъективностью



Рис. 1.3. Взаимосвязь моделей

таких моделей и скрытостью процесса мышления. Однако существуют объекты и процессы, для которых роль когнитивных моделей особенно велика. Например, оператор или лицо, принимающее решения, осуществляет управление объектом или процессом главным образом на основании собственных когнитивных моделей. Велика роль данного типа моделей и в социальных науках. В настоящее время изучением свойств и особенностей когнитивных моделей занимается новая, быстро развивающаяся дисциплина – *когнитология*.

Как уже отмечалось выше, одним из видов знакового моделирования является математическое моделирование.

Математическое моделирование – это идеальное научное знаковое формальное моделирование, при котором описание объекта осуществляется на языке математики, а исследование модели проводится с использованием тех или иных математических методов.

В качестве примера математического моделирования приведем классическую механику точки И. Ньютона, с помощью которой

можно описать движение любого материального объекта, размеры которого малы по сравнению с характерными расстояниями, проходимыми телом. Фактически все современные разделы физики посвящены построению и исследованию математических моделей различных физических объектов и явлений. Так, физики-«ядерщики» до проведения экспериментов выполняют серьезные исследования с применением математических моделей. При этом на основании результатов теоретического моделирования разрабатывается и уточняется методика натуральных экспериментов, выясняется, какие эффекты, где и когда следует ожидать, когда и что регистрировать. Такой подход позволяет значительно снизить затраты на проведение эксперимента, повысить его эффективность. Аналогичные замечания можно сделать по другим современным дисциплинам.

Как правило, значительные успехи в биологии и химии в последнее время были связаны с разработкой и исследованием математических моделей для биологических систем и химических процессов. В настоящее время широким фронтом идут работы по созданию математических моделей в экологии, экономике и социологии. Нельзя переоценить использование математических моделей в медицине и промышленности. Появилась возможность на научной (т.е. логически обоснованной) основе подходить ко многим экологическим и медицинским проблемам: имплантации и замене различных органов, прогнозированию развития эпидемий, обоснованной разработке планов ликвидации последствий крупных аварий и катастроф. Очень часто методы математического моделирования являются единственно возможными. Например, всестороннее математическое моделирование и «проигрывание» различных вариантов на ЭВМ позволило в кратчайшие сроки (1–2 недели) обоснованно спланировать и приступить к реализации плана ликвидации последствий Чернобыльской катастрофы. Уникальные результаты были получены по проекту «Гея», связанному с математическим моделированием последствий ядерной войны. Было выяснено, что в результате сильного запыления атмосферы возможно значительное глобальное похолодание («ядерная зима») и связанное с этим вымирание практически всего живого. Подобных примеров эффективного использования математических моделей можно привести очень много. В настоящее время это один из самых результативных и наиболее часто применяемых методов научного исследования.

Следует отметить, в сравнении с натурным экспериментом определенные преимущества математического моделирования:

- экономичность (в частности, сбережение ресурсов реальной системы);
- возможность моделирования гипотетических, т.е. не реализованных в природе объектов (прежде всего на разных этапах проектирования);
- возможность реализации режимов, опасных или труднопроизводимых в натуре (критический режим ядерного реактора, работа системы противоракетной обороны);
- возможность изменения масштаба времени;
- простота многоаспектного анализа;
- большая прогностическая сила вследствие возможности выявления общих закономерностей;
- универсальность технического и программного обеспечения проводимой работы (ЭВМ, системы программирования и пакеты прикладных программ широкого назначения).

Приведенное выше определение математического моделирования нельзя считать единственно возможным. Учитывая то, что сама математика подразделяется (конечно, с большей степенью условности) на «чистую» и «прикладную» [13], можно привести несколько вариантов подобного определения.

Пример определения математической модели, характерный для «чистой» математики, можно найти в известном справочнике Т.Корн и Г.Корн [55], где под математической моделью понимается «класс абстрактных и символьных математических объектов, таких, как числа или векторы, и отношения между ними». Под математическим отношением понимается «гипотетическое правило, связывающее два или более символических объекта». Вводятся абстрактное и конструктивное определения математической модели. При абстрактном определении новая модель задается «непротиворечивым набором правил (определяющих аксиом), вводящих операции, которыми можно пользоваться, и устанавливающих общие отношения между их результатами. Конструктивное определение вводит новую математическую модель, пользуясь уже известными математическими понятиями (например, определение сложения и умножения матриц в терминах сложения и умножения чисел)».

Излишний формализм, оперирование абстрактными понятиями можно считать, с одной стороны, недостатком приведенного определения «чистых» математиков. С другой стороны, абстрактность понятий повышает общность определения модели, делает его применимым к моделям самых разных по природе, не похожих объектов и явлений. Кроме того, несомненными достоинствами этого определения является его строгость и логичность. Для «при-

кладных» математиков характерна меньшая оторванность от реальной жизни, поскольку математические соотношения связывают не просто абстрактные математические объекты, а вполне конкретные параметры реальных физических, химических, биологических или социальных явлений или процессов. Желательно иметь определение математической модели, сохраняющее строгость и логичность последнего в «чистой» математике и, в то же время, пригодное для классификации существующих и создаваемых моделей, сравнения их между собой. Вариант подобного определения можно сформулировать в следующих выражениях.

Любая математическая модель, предназначенная для научных исследований, позволяет по заданным исходным данным найти значения интересующих исследователя параметров моделируемого объекта или явления. Поэтому можно предположить, что суть любой подобной модели заключается в отображении некоторого заданного множества Ω_X значений входных параметров X на множество значений Ω_Y выходных параметров Y . Данное обстоятельство позволяет рассматривать математическую модель как некоторый математический оператор A и сформулировать следующее определение.

Под *математической моделью* будем понимать любой оператор A , позволяющий по соответствующим значениям входных параметров X установить выходные значения параметров Y объекта моделирования:

$$A: X \rightarrow Y, \quad X \in \Omega_X, \quad Y \in \Omega_Y,$$

где Ω_X и Ω_Y — множества допустимых значений входных и выходных параметров для моделируемого объекта. В зависимости от природы моделируемого объекта элементами множеств Ω_X и Ω_Y могут являться любые математические объекты (числа, векторы, тензоры, функции, множества и т.п.).

Понятие оператора в приведенном определении может трактоваться достаточно широко. Это может быть как некоторая функция, связывающая входные и выходные значения, так и отображение, представляющее символическую запись системы алгебраических, дифференциальных, интегродифференциальных или интегральных уравнений. Наконец, это может быть некоторый алгоритм, совокупность правил или таблиц, обеспечивающих нахождение (или установление) выходных параметров по заданным исходным значениям.

Определение математической модели через понятие оператора является более конструктивным с точки зрения построения клас-

сификации таких моделей, поскольку включает в себя все многообразие имеющихся в настоящее время математических моделей.

С развитием вычислительной техники большое распространение получили *информационные модели*, представляющие, по существу, автоматизированные справочники, реализованные с помощью систем управления базами данных. Получая на входе некоторый запрос на поиск требуемой информации, подобные модели позволяют найти всю имеющуюся в базе данных информацию по интересующему вопросу. Однако такие модели не могут генерировать новое знание, отсутствующее в базе данных. Можно сказать, что это модели с нулевым потенциалом. В то же время в сочетании даже с весьма простыми математическими моделями (например, с применением регрессионного анализа) информационные модели могут привести к открытию новых закономерностей, позволить прогнозировать развитие исследуемых процессов.

Так как в качестве входных или выходных параметров моделей могут выступать математические объекты, а саму процедуру поиска данных можно представить в виде некоторого оператора, то информационные модели можно считать специфической разновидностью математических моделей. В данном пособии информационные модели и особенности их построения рассматриваться не будут.

1.3. Классификация математических моделей

Черная королева покачала головой: «Вы, конечно, можете назвать это чушью, но я-то встречала чушь такую, что в сравнении с ней эта кажется толковым словарем».

Л. Кэрролл

Классификационные признаки

Бурное развитие методов математического моделирования и многообразие областей их использования привело к появлению огромного количества моделей самого разного типа. В связи с этим возникает необходимость в определенном упорядочивании, классификации существующих и появляющихся математических моделей. Учитывая большое число возможных классификационных признаков и субъективность их выбора, появление все новых классов моделей, следует отметить условность и незавершенность рассматриваемой ниже классификации.

Представляется возможным подразделить математические модели на различные классы в зависимости от:

- сложности объекта моделирования;
- оператора модели (подмодели);
- входных и выходных параметров;
- способа исследования модели;
- цели моделирования.

Классификация математических моделей в зависимости от сложности объекта моделирования

В качестве объекта моделирования может выступать как некоторое материальное тело или конструкция, так и природный, технологический или социальный процесс либо явление. Все объекты моделирования можно разделить на две группы: простые и объекты-системы (рис. 1.4). В первом случае при моделировании не рассматривается внутреннее строение объекта, не выделяются составляющие его элементы или подпроцессы. В качестве примера подобного объекта можно привести материальную точку в классической механике.



Рис. 1.4. Классификация объектов моделирования

Система есть совокупность взаимосвязанных элементов, в определенном смысле обособленная от окружающей среды и взаимодействующая с ней как целое.

Для сложных систем характерно наличие большого числа взаимно связанных, взаимодействующих между собой элементов. При этом связь между элементами *A* и *B* системы может отличаться от связи между элементами *B* и *A*.

Если система имеет N элементов и каждый элемент связан с каждым, то общее число связей равно $N(N-1)$. Если все N элементов имеют по M состояний, то для такой системы общее число

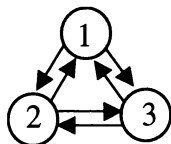


Рис. 1.5. Схема электронной системы

состояний S равно M^N . Например, пусть некоторая электронная система (рис. 1.5) состоит из трех блоков ($N = 3$) и каждый блок может находиться в двух состояниях ($M = 2$, например, включен и выключен). Для подобной системы имеем $S = 2^3 = 8$ состояний. Максимальное число связей в подобной системе равно 6. Если поведение системы описывается процессом перехода блока из одного состояния в другое, то общее число возможных переходов равно S^2 . Для рассматриваемого

примера число сценариев возможного поведения системы равно $S = 8^2 = 64$.

Поведение системы быстро усложняется с ростом числа ее элементов системы. Так, для системы из 10 элементов при $M = 2$ число состояний $S = 1024$, а число сценариев равно 1 048 576. Данное обстоятельство, с одной стороны, говорит о сложности систем и многовариантности их поведения. С другой стороны, следует ожидать наличия больших трудностей, возникающих при изучении и моделировании систем.

Конечно, деление объектов исследования на «простые» и «сложные» условно. Поскольку для любых известных процессов, явлений, материальных тел невозможно выделить их «элементарные кирпичики», «атомы», то любой объект исследования можно считать бесконечно сложным. Упрощение его строения при разработке модели выполняется в результате отбрасывания малозначимых, несущественных для достижения поставленных на данный момент целей исследования связей между составляющими объект элементами. При изменении целей исследования или повышении требований к точности и глубине модели приходится, как правило, пересматривать уровень детализации объекта.

Модели объектов-систем, учитывающие свойства и поведение отдельных элементов, а также взаимосвязи между ними, называются *структурными*. Более подробно модели подобного типа будут рассмотрены в гл. 4.

Среди структурных динамических систем выделяют в отдельный подкласс *имитационные* системы, состоящие из конечного числа элементов, каждый из которых имеет конечное число состояний. Число связей между элементами также предполагается конечным. Моделирование взаимодействий элементов внутри системы осуществляется с помощью некоторого алгоритма, реализуемого обычно с использованием ЭВМ. Для моделирования на ЭВМ ре-

ального времени вводится понятие системного времени. В качестве моделей отдельных элементов могут быть использованы модели любого типа. Модели имитационного типа более подробно рассмотрены в гл. 7.

Как правило, взаимодействие внешней среды со сложной системой полностью проследить не удастся, что приводит к неопределенности внешних воздействий и, как следствие, неоднозначности в поведении самой системы. Наличие подобной неопределенности является характерной особенностью сложных систем.

Классификация математических моделей в зависимости от оператора модели

Выше отмечалось, что любая математическая модель может рассматриваться как некоторый оператор A , который является алгоритмом или определяется совокупностью уравнений – алгебраических, обыкновенных дифференциальных уравнений (ОДУ), систем ОДУ (СОДУ), дифференциальных уравнений в частных производных (ДУЧП), интегродифференциальных уравнений (ИДУ) и др. (рис. 1.6).

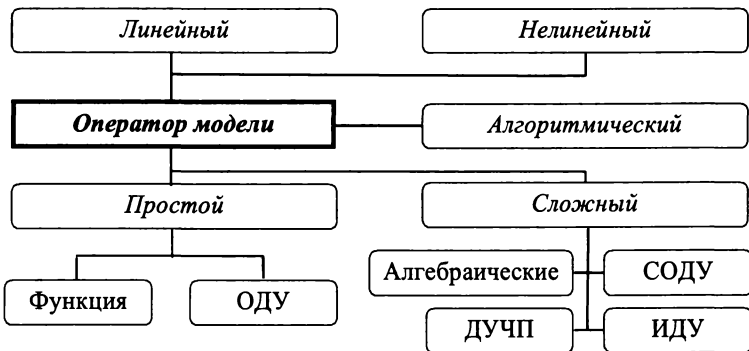


Рис. 1.6. Классификация в зависимости от оператора модели

Если оператор обеспечивает линейную зависимость выходных параметров Y от значений входных параметров X , то математическая модель называется *линейной* (рис. 1.7). Линейные модели более просты для анализа. Например, из свойства линейности следует свойство суперпозиции решений, т.е. если известны решения Y_1 при

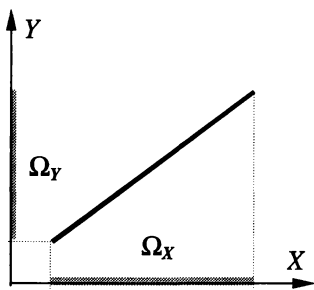


Рис. 1.7. Иллюстрация линейной модели

X_1 и Y_2 при X_2 , то решение для выходных параметров при $X = X_1 + X_2$ есть $Y = Y_1 + Y_2$. Предельные значения Y для линейных моделей достигаются, как правило, на границах областей Ω_X допустимых значений входных параметров.

Исторически первыми стали разрабатываться и исследоваться именно линейные математические модели. Область применения подобных моделей очень широка. Она охватывает классическую механику, электродинамику, аналитическую химию и биологию. Методы их построения, разрабатывавшиеся в течение столетий, обладают большой общностью и эффективностью.

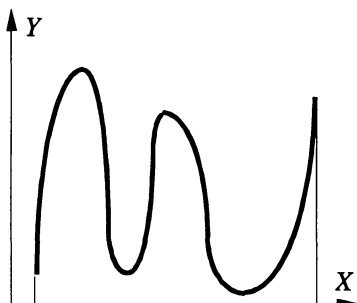


Рис. 1.8. Схематичное изображение нелинейной модели

Линейное поведение свойственно относительно простым объектам. Системам, как правило, присуще нелинейное многовариантное поведение (рис. 1.8).

В настоящее время все чаще возникает потребность не только в повышении точности моделирования, но и в создании качественно новых моделей, учитывающих нелинейность поведения реальных объектов исследования. Анализ подобных моделей намного сложнее, чем линейных, причем разработка методики и общих подходов

к исследованию в настоящее время далека от завершения. Являясь более богатым и сложным, мир нелинейных моделей представляется для современной науки более перспективным в плане открытия новых закономерностей и описания сложных явлений. Например, такие явления, как солитоны и хаос (см. гл. 6), нельзя с достаточной степенью адекватности описать в рамках традиционных линейных моделей. Методы исследования нелинейных моделей в настоящее время быстро прогрессируют, складываясь в новые научные направления. К таким относительно новым направлениям

можно отнести, например, синергетику – науку о сложных самоорганизующихся системах.

В зависимости от вида оператора математические модели можно разделить на *простые* и *сложные*.

В случае, когда оператор модели является алгебраическим выражением, отражающим функциональную зависимость выходных параметров Y от входных X , модель будем называть *простой*.

Простые модели чаще всего являются результатом обобщения и анализа экспериментальных данных, полученных в результате наблюдений за исследуемым объектом или явлением. На основании анализа таких данных выдвигается гипотеза о возможной функциональной связи входных и выходных параметров. После этого гипотеза проверяется на имеющемся экспериментальном материале, уточняется степень ее адекватности, т.е. степень соответствия результатов моделирования, полученных с применением данной гипотезы, имеющимся знаниям об исследуемом объекте. Если результаты проверки неудовлетворительны, то принятая гипотеза отвергается и заменяется новой. Процесс повторяется до получения желаемой степени соответствия результатов эксперимента и модели.

В качестве примеров простых моделей можно привести многие законы физики (закон всемирного тяготения, закон Ома, закон Гука, закон трения Амонтона–Кулона), а также все эмпирические, т.е. полученные из опыта, алгебраические зависимости между входными и выходными параметрами. Так, в теории резания металлов очень часто используются соотношения, ставящие время и стоимость обработки детали на станке в зависимость от скоростей ее вращения (скорости резания) и осевого перемещения (скорости подачи) резца.

Модель, включающая системы дифференциальных и интегральных соотношений, уже не может быть отнесена к простым, так как для своего исследования требует применения довольно сложных математических методов. Однако в двух случаях она может быть сведена к простым:

1) если полученная для подобной модели система математических соотношений может быть разрешена аналитически;

2) если результаты вычислительных экспериментов со сложной моделью аппроксимированы некоторой алгебраической зависимостью. В настоящее время известно достаточно большое число подходов и методов аппроксимации (например, метод наименьших квадратов или метод планирования экспериментов).

На практике довольно часто возникают ситуации, когда удовлетворительное описание свойств и поведения объекта моделиро-

вания (как правило, сложной системы) не удастся выполнить с помощью математических соотношений. Однако в большинстве случаев удастся построить некоторый имитатор поведения и свойств такого объекта с помощью алгоритма, который можно считать оператором модели.

Например, если в результате наблюдения за объектом получена таблица соответствия между входными X и выходными Y значениями параметров, то определить оператор A , позволяющий получить «выход» по заданному «входу», зачастую бывает проще с помощью алгоритма.

Классификация математических моделей в зависимости от параметров модели (рис. 1.9)

В общем случае параметры, описывающие состояние и поведение объекта моделирования, разбиваются на ряд непересекающихся подмножеств:



Рис. 1.9. Классификация математических моделей в зависимости от параметров

- совокупность входных (управляемых) воздействий на объект (Ω_X);
- совокупность воздействий внешней среды (неуправляемых) (Ω_E);
- совокупность внутренних (собственных) параметров объекта (Ω_P);
- совокупность выходных характеристик (Ω_Y).

Например, при моделировании движения твердого тела в атмосфере в поле сил тяжести входными параметрами могут быть начальное положение и начальная скорость точки, принятой за полюс, а также угловая скорость в момент времени $t = 0$. Сила сопротивления и сила тяжести характеризуют воздействие внешней среды. Масса тела и его форма являются собственными параметрами тела. Координаты и скорости точек тела (при $t > 0$) относятся к выходным параметрам.

В то же время отнесение параметров к входным или выходным зависит от постановки конкретной задачи. Например, в приведенном примере можно переформулировать задачу, сделав ее обратной к исходной: определить начальное положение и скорости (линейную скорость полюса и угловую скорость тела) по заданному положению и скоростям в момент времени $t_1 > 0$. Понятно, что в данном случае входные и выходные параметры меняются местами.

Входные параметры X , параметры, описывающие воздействие внешней среды E , и внутренние (собственные) характеристики I объекта относят обычно к *независимым (экзогенным)* величинам. Выходные параметры Y — *зависимые (эндогенные)* величины.

В общем случае оператор модели A преобразует экзогенные параметры в эндогенные $A: \{X, E, I\} \rightarrow Y$.

Следует отметить, что введенные здесь внутренние характеристики, являющиеся независимыми (от внешних воздействий) величинами, не следует смешивать с так называемыми внутренними переменными, широко используемыми в механике сплошных сред, термодинамике и других естественно-научных дисциплинах. К таким внутренним переменным относится, например, плотность дислокаций в моделях физики твердого тела и мезомеханики, которая, безусловно, зависит от внешних воздействий.

Количество параметров всех типов в математических моделях, как правило, конечно (хотя параметры или функции могут принадлежать любому бесконечномерному функциональному пространству). При этом каждый из параметров может иметь различную «математическую природу»: быть постоянной величиной или функци-

ей, скаляром или вектором (или тензором второго, третьего и выше рангов), четким или нечетким множеством и т.д. В математических моделях, рассматриваемых в естественных науках, наиболее распространенными являются параметры, являющиеся тензор-значными функциями (напомним, что скаляр – тензор нулевого ранга, а вектор – тензор первого ранга). В качестве независимых переменных (аргументов) при этом обычно выступают координаты точек трехмерного пространства и/или время (или некоторый неубывающий параметр – аналог времени).

По своей природе характеристики объекта могут быть как качественными, так и количественными. Введение тех или иных количественных характеристик объекта моделирования возможно при наличии эталона сравнения. Например, для характеристики размеров тела используется эталонный образец – метр. Для количественной характеристики вводятся числа, выражающие отношения между данным параметром и эталоном. Кроме того, количественные значения параметра могут выражаться дискретными или непрерывными величинами. Качественные характеристики находятся, например методом экспертных оценок. В зависимости от вида используемых множеств параметров модели могут подразделяться на *качественные и количественные, дискретные и непрерывные, а также смешанные*.

При построении моделей реальных объектов и явлений очень часто приходится сталкиваться с недостатком информации. Как правило, для любого исследуемого объекта распределение свойств, параметры воздействия и начальное состояние известны с той или иной степенью неопределенности. Это связано с множеством трудно учитываемых факторов, ограниченностью числа используемых параметров модели, конечной точностью экспериментальных измерений. При построении модели возможны следующие варианты описания неопределенности параметров:

1) *детерминированное* – значения всех параметров модели определяются детерминированными величинами (т.е. каждому параметру соответствует конкретное целое, вещественное или комплексное число либо соответствующая функция). Данный способ соответствует полной определенности параметров;

2) *стохастическое* – значения всех или отдельных параметров модели определяются случайными величинами, заданными плотностями вероятности. В литературе наиболее полно исследованы случаи нормального (гауссова) и показательного распределения случайных величин;

3) *случайное* – значения всех или отдельных параметров модели устанавливаются случайными величинами, заданными оценками плотностей вероятности, полученными в результате обработки ограниченной экспериментальной выборки данных параметров. Эта форма описания тесно связана с предыдущей. Однако в рассматриваемом случае получаемые результаты моделирования будут существенным образом зависеть от точности оценок моментов и плотностей вероятности случайных параметров, от постулируемых законов распределения и объема выборок;

4) *интервальное* – значения всех или отдельных параметров модели описываются интервальными величинами, заданными интервалом, образованным минимальным и максимально возможными значениями параметра;

5) *нечеткое* – значения всех или отдельных параметров модели описываются функциями принадлежности соответствующему нечеткому множеству. Такая форма используется, когда информация о параметрах модели задается экспертом на естественном языке, а следовательно, в «нечетких» (с позиции математики) терминах типа «много больше пяти», «около нуля».

В теории игр [34, 40, 77] встречается еще один вид неопределенности параметров модели, называемый *игровой неопределенностью*. В данном пособии указанный вид неопределенности не рассматривается.

Разделение моделей на *одномерные, двухмерные и трехмерные* применимо для таких моделей, в число параметров которых входят координаты пространства, и связано с особенностями реализации этих моделей, равно как и с резким увеличением их сложности при возрастании размерности (с «проклятием размерности» по образному выражению Р. Беллмана). Как правило, увеличение размерности модели приводит к росту числа используемых математических соотношений. Особенно сложны в реализации трехмерные модели, требующие высокопроизводительной вычислительной техники с большим объемом оперативной и дисковой памяти. Реализация таких моделей стала возможной лишь с появлением вычислительных машин третьего поколения и потребовала создания специальных вычислительных методов и приемов.

Как правило, эффективная реализация трехмерных моделей возможна лишь на многопроцессорных вычислительных комплексах с использованием языков параллельных вычислений. Среди характерных вычислительных трудностей, с которыми сталкиваются при создании моделей в трехмерной постановке, можно отметить

необходимость хранения и решения систем уравнений большой размерности (10 тысяч уравнений и более), проблемы подготовки исходной информации и ее проверки, наглядное отображение полученных результатов. При разработке модели стараются (если это возможно) понизить размерность. Однако необоснованное понижение размерности модели может существенно исказить результаты моделирования. Например, если для исследования движения брошенного мяча в вертикальной плоскости использование двухмерной модели может быть оправдано, то для исследования движения бумеранга такую модель строить бесполезно.

Из всей совокупности параметров при разработке различных моделей отдельно следует рассмотреть учет времени. Как и координаты, время относится к независимым переменным, от которых могут зависеть остальные параметры модели. В различных ситуациях объект исследования может по-разному испытывать влияние времени. Обычно чем меньше масштаб объекта, тем существеннее зависимость его параметров от времени. Если сравнивать скорости изменения различных объектов по отношению к масштабам, характерным для Земли, то можно отметить, что для галактик время заметных изменений измеряется миллионами лет, а для элементарных частиц — миллионными долями секунды. Учитывая, что весь окружающий нас материальный мир состоит из постоянно изменяющихся и взаимодействующих элементарных частиц и полей, то все без исключения объекты исследования следует считать изменяющимися во времени.

Любой объект стремится перейти в некоторое равновесное состояние как с окружающей его средой, так и между отдельными элементами самого объекта. Нарушение этого равновесия приводит к изменениям различных параметров объекта и его переходу в новое равновесное состояние.

При построении модели важным является сравнение времени существенных изменений внешних воздействий и характерных времен перехода объекта в новое равновесное состояние с окружающей средой, а также времени релаксации, определяющего установление равновесия между отдельными элементами внутри объекта. Если скорости изменения внешних воздействий на объект моделирования существенно меньше скорости релаксации, то явной зависимостью от времени в модели можно пренебречь. В этом случае говорят о *квазистатическом процессе*.

Например, если скорость появления микротрещин в элементах конструкции моста, связанная с сезонными колебаниями темпера-

туры и переменностью нагрузок, невелика, то расчет его максимальной несущей способности можно проводить в рамках статической модели. Срок службы моста в этом случае можно определить с помощью квазистатической модели, использующей зависимость прочностных свойств материала моста от суммарного числа циклов нагружения до разрушения.

Совокупность значений параметров модели в некоторый момент времени или на данной стадии называется *состоянием объекта*.

Если скорости изменения внешних воздействий и параметров состояния изучаемого объекта достаточно велики (по сравнению со скоростями релаксации), то учет времени необходим. В этом случае объект исследования рассматривают в рамках *динамического процесса*.

Условие движения отдельных элементов исследуемого объекта не является обязательным условием включения времени в число параметров модели. Для примера рассмотрим течение жидкости в длинной трубе постоянного сечения. Эксперименты показывают, что на достаточно большом удалении от входа в трубу частицы жидкости движутся параллельно оси трубы (рис. 1.10).

При этом если условия на входе не изменяются и скорость течения невелика (ламинарный режим течения), то профиль скоростей частиц в данном сечении трубы с течением времени остается неизменным. В этом случае в каждой фиксированной точке исследуемого пространства значения параметров модели не зависят от времени.

Подобные процессы называют *стационарными*. Как правило, стационарные модели применяются для описания различных потоков (жидкости, газа, тепла) в случае постоянства условий на входе и выходе потока. Для таких процессов время может быть исключено из числа независимых переменных.

Если в качестве одной из существенных независимых переменных модели необходимо использовать время (или его аналог), то модель называется *нестационарной*. Примером нестационарной модели является модель движения жидкости в трубе, но вытекающей

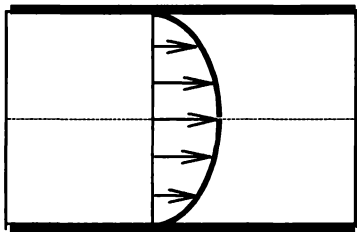


Рис. 1.10. Поле скоростей при ламинарном течении

из некоторого сосуда. По мере понижения уровня жидкости в сосуде давление на входе в трубу будет уменьшаться, что приведет к изменению параметров течения жидкости в любой точке трубы.

Заметим, что для значительной части реальных процессов стационарные режимы являются наиболее предпочтительными. После их определения (с применением той или иной математической модели) проверяется устойчивость стационарного режима (решения), что во многих случаях требует постановки и решения нестационарной задачи для возмущений стационарного решения. В ряде случаев, когда определение стационарных режимов из аналитического решения или некоторых эвристических соображений затруднено, их поиск осуществляется методом установления соответствующей нестационарной задачи (т.е. ищется решение нестационарной задачи, асимптотически стремящееся к стационарному). Следует отметить, что этим методом довольно часто пользуются при решении стационарных задач численными методами, поскольку методы решения нестационарных задач часто оказываются существенно эффективнее, чем стационарных.

Классификация математических моделей в зависимости от целей моделирования (рис. 1.11)

Целью *дескриптивных моделей* (от лат. *descriptio* – описание) является установление законов изменения параметров модели. В качестве примера такой модели можно привести модель движения материальной точки под действием приложенных сил, использующая второй закон Ньютона. Задавая положение и скорость точки в начальный момент времени (входные параметры), массу (собственный параметр) и закон изменения прикладываемых сил (внешние воздействия), можно определить скорость и координаты материальной точки в любой момент времени (выходные параметры). Полученная модель описывает зависимость выходных параметров от входных. Поэтому дескриптивные модели являются реализацией

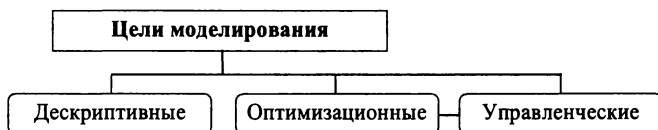


Рис. 1.11. Классификация в зависимости от целей моделирования

описательных и объяснительных содержательных моделей на формальном уровне моделирования.

Другой пример дескриптивной модели — модель движения ракеты после старта с поверхности Земли. В качестве параметров модели в данном случае могут выступать начальное положение и начальная скорость ракеты (входные), ее начальная масса, импульс двигателя, режим его работы (собственные параметры), закон изменения сил притяжения и сил сопротивления атмосферы (внешние воздействия). Выходными параметрами будут положение и скорость центра масс ракеты и ее ориентация в пространстве в произвольный момент времени.

Оптимизационные модели предназначены для определения оптимальных (наилучших) с точки зрения некоторого критерия параметров моделируемого объекта или же для поиска оптимального (наилучшего) режима управления некоторым процессом. Часть параметров модели относят к параметрам управления, изменяя которые можно получать различные варианты наборов значений выходных параметров. Как правило, данные модели строятся с использованием одной или нескольких дескриптивных моделей и включают некоторый критерий, позволяющий сравнивать различные варианты наборов значений выходных параметров между собой с целью выбора наилучшего. На область значений входных параметров могут быть наложены ограничения в виде равенств и неравенств, связанные с особенностями рассматриваемого объекта или процесса. Целью оптимизационных моделей является поиск таких допустимых параметров управления, при которых критерий выбора достигает своего «наилучшего значения».

Примером оптимизационной модели может служить моделирование процесса запуска ракеты с поверхности Земли с целью подъема ее на заданную высоту за минимальное время при ограничениях на величину импульса двигателя, время его работы, начальную и конечную массу ракеты. Математические соотношения дескриптивной модели движения ракеты выступают в данном случае в виде ограничений типа равенств.

Отметим, что для большинства реальных процессов, конструкций требуется определение оптимальных параметров сразу по нескольким критериям, т.е. мы имеем дело с так называемыми многокритериальными задачами оптимизации. При этом нередкими являются ситуации противоречивости критериев; например, при оптимизации конструкции рамы грузового автомобиля можно потребовать максимальной жесткости, минимальной массы и мини-

мальной стоимости. Для решения подобных задач используются специальные методы и алгоритмы.

Управленческие модели применяются для принятия эффективных управленческих решений в различных областях целенаправленной деятельности человека. В общем случае принятие решений является процессом, по своей сложности сравнимым с процессом мышления в целом [20]. Однако на практике под принятием решений обычно понимается выбор некоторых альтернатив из заданного их множества, а общий процесс принятия решений представляется как последовательность таких выборов альтернатив [20]. Например, на предприятии освободилась должность главного инженера, и задача директора состоит в выборе из имеющегося множества кандидатов на эту должность одного, отвечающего заданным требованиям. Сложность задачи заключается в наличии неопределенности как по исходной информации (неполные данные о кандидатах) и характеру воздействия внешних условий (случайное: выбранный кандидат заболел или отказался; игровое: министерство против выбранной кандидатуры), так и по целям (противоречивые требования к выбираемой кандидатуре: должен быть хорошим специалистом и администратором, опытен, энергичен, молод и пр.). Поэтому в отличие от оптимизационных моделей, где критерий выбора считается определенным и искомое решение устанавливается из условий его экстремальности, в управленческих моделях необходимо введение специфических критериев оптимальности, которые позволяют сравнивать альтернативы при различных неопределенностях задачи.

Поскольку оптимальность принятого решения даже в одной и той же ситуации может пониматься по-разному, вид критерия оптимальности в управленческих моделях заранее не фиксируется. Именно в этом состоит основная особенность данных моделей.

Методы формирования критериев оптимальности в зависимости от вида неопределенности рассматриваются в теории выбора и принятия решений [20, 105], которая базируется на теории игр и исследовании операций [34, 77].

Классификация математических моделей в зависимости от методов реализации (рис. 1.12)

Метод реализации модели относят к *аналитическим*, если он позволяет получить выходные параметры в виде *аналитических выра-*

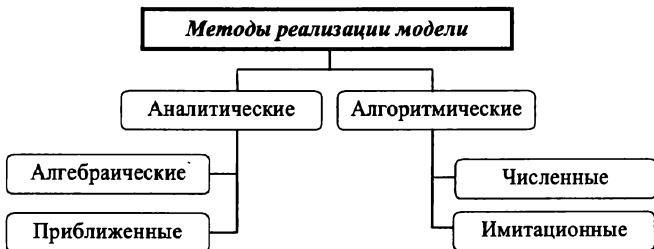


Рис. 1.12. Классификация в зависимости от методов реализации

жений, т.е. выражений, в которых используется не более чем счетная совокупность арифметических операций и переходов к пределу.

Примеры аналитических выражений:

$$\sum_{k=0}^{\infty} \frac{a_k x^k}{x^k + 1}, \quad \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n.$$

Частным случаем аналитических выражений являются *алгебраические выражения*, в которых используется конечное или счетное число арифметических операций, операций возведения в целочисленную степень и извлечения корня. Примеры алгебраических выражений:

$$ax^2 + bx + c, \quad a + b\sqrt{x^3 + 4ac}.$$

Очень часто аналитическое решение для модели представляют в элементарных или специальных функциях: показательных, логарифмических, тригонометрических, гиперболических и т.п. Для получения значений этих функций при конкретных значениях входных параметров используют их разложение в ряды (например, Тейлора). Так, показательная функция может быть представлена следующим рядом:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

Учитывая различное число членов ряда, можно вычислять значения функции с различной степенью точности. Например, учет первых шести членов ряда в разложении показательной функции обес-

печивает точность в 10^{-4} , а первых десяти — 10^{-8} . Таким образом, значение функции при каждом значении аргумента в этом случае определяется приближенно. Модели, использующие подобный прием, называются *приближенными*.

Аналитические методы реализации модели являются более ценными в том плане, что позволяют с меньшими вычислительными затратами изучить свойства объекта моделирования, применяя традиционные хорошо развитые математические методы анализа аналитических функций. Существенно, что применение аналитических методов возможно без использования ЭВМ (за исключением случаев, когда аналитическое решение определяется в рядах и для его доведения до числа требуются трудоемкие вычисления с применением ЭВМ). Кроме того, знание аналитического выражения для искомых параметров позволяет исследовать фундаментальные свойства объекта, его качественное поведение, строить новые гипотезы о его внутренней структуре. Следует отметить, что возможности аналитических методов существенно зависят от уровня развития соответствующих разделов математики.

В настоящее время мощный всплеск интереса к аналитическим методам при реализации моделей связан с появлением пакетов математических вычислений (*Derive, MatLab, Mathcad, Maple, Mathematica, Scientific WorkPlace* и др.). Спектр решаемых данными пакетами задач очень велик и постоянно расширяется (элементарная математика, символьные операции с полиномами, производными и интегралами, с векторами и матрицами, задачи теории поля и векторного анализа, метод конечных элементов и т.п.). Применение подобных программных средств не только упрощает процедуру получения аналитического решения, но и облегчает последующий анализ полученного решения с применением различного рода визуализаторов. Достаточно подробное описание математических пакетов, их сравнение и анализ возможностей можно найти в [92].

К сожалению, существующие в настоящее время математические методы позволяют получить аналитические решения только для относительно несложных математических моделей в узком диапазоне значений параметров. В большинстве случаев при исследовании моделей приходится использовать алгоритмические подходы, позволяющие получить лишь приближенные значения искомых параметров.

При *численном подходе* совокупность математических соотношений модели заменяется конечномерным аналогом. Это чаще всего достигается дискретизацией исходных соотношений, т.е. пе-

реходом от функций непрерывного аргумента к функциям дискретного аргумента. После дискретизации исходной задачи выполняется построение вычислительного алгоритма, т.е. последовательности арифметических и логических действий, выполняемых на ЭВМ и позволяющих за конечное число шагов получить решение дискретной задачи. Найденное решение дискретной задачи принимается за приближенное решение исходной математической задачи.

Степень приближения определяемых с помощью численного метода искомым параметром модели зависит как от *погрешностей самого метода*, связанных с заменой исходной модели ее дискретным аналогом, так и от *ошибок округления*, возникающих при выполнении любых расчетов на ЭВМ в связи с конечной точностью представления чисел в ее памяти. Основным требованием к вычислительному алгоритму является необходимость получения решения исходной задачи с заданной *точностью* за конечное число шагов.

К настоящему времени круг вопросов, связанных с разработкой и использованием численных методов, а также с построением на их основе вычислительных алгоритмов, выделился в самостоятельный быстро развивающийся и обширный раздел — *вычислительную математику*.

Если при численном подходе дискретизации подвергалась полученная система математических соотношений, то при *имитационном подходе* на отдельные элементы разбивается сам объект исследования [66]. В этом случае система математических соотношений для объекта-системы в целом не записывается, а заменяется некоторым алгоритмом, моделирующим ее поведение и учитывающим взаимодействие друг с другом моделей отдельных элементов системы. Модели отдельных элементов могут быть как аналитическими, так и алгоритмическими.

Алгоритмические модели, использующие как численный, так и имитационный подход, не позволяют получить решения задач в аналитической форме, что затрудняет и усложняет процесс анализа результатов моделирования. Так как применение моделей данного типа возможно лишь при наличии вычислительной техники, то их эффективность зависит от мощности и быстродействия ЭВМ. Несомненным достоинством алгоритмических моделей является отсутствие принципиальных ограничений на сложность модели, что позволяет применять их для исследования систем произвольной сложности.

Использование математической модели, построенной алгоритмическими методами, аналогично проведению экспериментов с ре-

альным объектом, только вместо реального эксперимента с объектом проводится *вычислительный эксперимент* с его моделью. Задавая конкретным набором значений исходных параметров модели, в результате вычислительного эксперимента находим конкретный набор приближенных значений искомым параметрам. Для исследования поведения объекта при новом наборе исходных данных необходимо проведение нового вычислительного эксперимента.



Вопросы для самопроверки

1. Что такое модель и моделирование? Цели моделирования?
2. В каких областях человеческой деятельности применяются модели?
3. Можно ли отнести мифологию к моделированию? Почему?
4. Какие типы моделей используются в изучаемых вами дисциплинах (включая дисциплины вузовского и/или школьного курса)?
5. Какие существуют типы моделирования?
6. В чем отличие моделирования натурального от мысленного?
7. Назовите характерные особенности аналоговых моделей.
8. Что такое когнитивная модель?
9. Какие модели называют содержательными?
10. Назовите разновидности содержательных моделей.
11. Чем концептуальная модель отличается от содержательной?
12. Что такое формальная модель?
13. Какое моделирование называется математическим?
14. Какие примеры математических моделей вам известны?
15. Сформулируйте достоинства математических моделей.
16. Приведите и проанализируйте различные примеры определений математических моделей.
17. Что может выступать в качестве оператора при математическом моделировании?
18. Почему информационные модели нельзя считать разновидностью математических?
19. По каким классификационным признакам можно разделять математические модели?
20. Чем простые модели отличаются от сложных?
21. В чем заключается сложность моделирования систем?
22. Какие типы моделей можно выделить по виду оператора моделирования?
23. Чем отличаются линейные и нелинейные модели?
24. Какие типы моделей выделяются по виду параметров моделирования?

25. Чем характерна дескриптивная модель?
26. Для каких целей служит оптимизационная модель?
27. Чем отличаются стационарные и нестационарные модели?
28. Как влияет размерность на сложность модели? Почему?
29. Перечислите способы описания неопределенности параметров модели.
30. Назовите основные методы реализации моделей, перечислите их достоинства и недостатки.



Задания для самостоятельного выполнения

1. Разработайте когнитивную, содержательную и концептуальную (структурно-функциональную и причинно-следственную) модели оптимального (с вашей точки зрения) расписания движения общественного транспорта. Попытайтесь оценить различие когнитивной и содержательной моделей. С использованием доступных вам математических методов разработайте вариант математической модели.

2. Проведите классификацию математических моделей классической физики.

3. Для задания №1 разработайте простую аналитическую модель, проанализируйте ее с позиций сложности объекта, операторов, типа входных и выходных параметров, цели моделирования. Проведите численные эксперименты для различных значений входных переменных.

4. Проведите анализ и классификацию нескольких математических моделей в интересующей вас области знаний. Установите аналоги рассматриваемых математических моделей в других областях.

5. Аналогично предыдущему пункту – для процесса развития эпидемии в регионе. Для математической модели найдите аналоги в известных моделях механики сплошной среды.

6. Сформулируйте несколько вариантов содержательных постановок задач моделирования работы:

- продовольственного магазина;
- стационара больницы или поликлиники;
- ремонтной мастерской;
- регулируемого перекрестка дорог;
- автодорожной сети микрорайона;
- участка цеха;
- факультета вуза или средней школы.

Глава 2

ЭТАПЫ ПОСТРОЕНИЯ МАТЕМАТИЧЕСКОЙ МОДЕЛИ

Процесс создания математических моделей трудоемок, длителен и связан с использованием труда различных специалистов достаточно высокого уровня, обладающих хорошей подготовкой как в предметной области, связанной с объектом моделирования, так и в области прикладной математики, современных численных методов, программирования, знающих возможности и особенности современной вычислительной техники. Отличительной особенностью математических моделей, создаваемых в настоящее время, является их комплексность, связанная со сложностью моделируемых объектов. Например, при моделировании процессов деформирования различных конструкций под действием приложенной нагрузки приходится учитывать не только происходящие при деформировании процессы массопереноса, но и теплоперенос, а также связанные с этими процессами изменения структуры и свойств материала.

В некоторых случаях необходимо учитывать влияние различных видов излучения, воздействия гравитационных и электромагнитных полей, предьстории деформирования. Кроме того, для современных моделей характерно представление объекта моделирования в виде более или менее сложной системы взаимодействующих элементов.

Все отмеченные выше особенности приводят к усложнению модели и необходимости совместного использования нескольких теорий (нередко — из разных областей знания), применения современных вычислительных методов и вычислительной техники для получения и анализа результатов моделирования.

Внедрение вычислительной техники во все сферы человеческой деятельности привело к повсеместному использованию математических моделей. Заметим, что ЭВМ — это только «железо», а «умным» и полезным его делают программы, которые в большинстве случаев являются реализациями алгоритмов соответствующих математических моделей. Поэтому необходимо создание большого количества разнообразных математических моделей с широкими возможностями, отвечающих различным, зачастую противоречивым, требованиям. В случае сложных объектов удовлетворить всем предъявляемым требованиям в одной модели обычно невозможно. Приходится создавать целый спектр моделей одного и того же объекта (в некоторых случаях — иерархическую совокупность «вложенных» одна в другую моделей), каждая из которых наиболее эффективно решает возложенные на нее задачи. Например, в конструкторской и технологической практике, как правило, применяется широкий спектр моделей — от простых расчетных формул (часть из которых представляет собой аппроксимацию экспериментальных данных) на первоначальной стадии до весьма сложных моделей, приближающихся к исследовательским, — на завершающей стадии разработки конструкции или технологического процесса.

Модели, ориентированные на исследовательские цели, способны представлять объект в широком диапазоне исходных параметров с удовлетворительной точностью. При этом практически нет ограничений по сложности подобной модели, а также времени, затрачиваемом на получение результатов.

Исследовательские модели могут быть ориентированы как на количественные, так и на качественные результаты. К моделям, используемым в автоматизированных системах управления (АСУ), в отличие от исследовательских, предъявляются достаточно жесткие ограничения относительно времени, затрачиваемого на получение результатов, а также точности самих результатов.

Необходимость массового построения моделей требует разработки некоторой совокупности правил и подходов, которые позволили бы снизить затраты на разработку моделей и уменьшить вероятность появления трудно устранимых впоследствии ошибок. Подобную совокупность правил можно было бы назвать *технологией создания математических моделей*.

Процесс построения любой математической модели можно представить последовательностью этапов, представленных на рис. 2.1.



Рис. 2.1. Этапы построения математической модели

2.1. ОБСЛЕДОВАНИЕ ОБЪЕКТА МОДЕЛИРОВАНИЯ

Первоначальная формулировка проблемы является самой трудной частью, так как здесь необходимо все время использовать свои мысли, позднее взамен их может использоваться математика.

А. Эддингтон

Математические модели, особенно использующие численные методы и вычислительную технику, требуют для своего построения значительных интеллектуальных, финансовых и временных затрат. Поэтому решение о разработке новой модели принимается лишь в случае отсутствия иных, более простых путей решения возникших проблем (например, модификации одной из существующих моделей). Необходимость в новой модели может появиться в связи с про-

ведением научных исследований (особенно — на стыке различных областей знания), выполнением проектных и конструкторских работ на производстве, созданием систем автоматического управления, планирования и контроля. Человека или организацию, заинтересованных в разработке новой математической модели, для краткости будем называть *заказчиком*. После принятия решения о необходимости построения новой математической модели заказчик ищет *исполнителя* своего заказа. В качестве исполнителя, как правило, может выступать рабочая группа, включающая специалистов разного профиля: прикладных математиков, специалистов, хорошо знающих особенности объекта моделирования, программистов.

Итак, если решение о создании модели принято и рабочая группа сформирована, то можно приступать к этапу обследования объекта моделирования. Основной целью данного этапа является подготовка содержательной постановки задачи моделирования.

Перечень сформулированных в содержательной (словесной) форме основных вопросов об объекте моделирования, интересующих заказчика, составляет содержательную постановку задачи моделирования.

Подготовка списка вопросов, на которые должна ответить новая модель, зачастую является самостоятельной проблемой, требующей для своего решения специалистов со специфическими знаниями и способностями. Они должны не только хорошо разбираться в предметной области моделирования, знать возможности современной вычислительной математики и техники, но и быть достаточно коммуникабельными, т.е. уметь общаться с людьми, «разговорить» практиков, хорошо «чувствующих» объект моделирования, нюансы его поведения. Подобных специалистов в настоящее время называют *постановщиками задач*.

На основании анализа всей собранной информации постановщик задачи должен сформулировать такие требования к будущей модели, которые, с одной стороны, удовлетворяли бы заказчика, а с другой — позволяли бы реализовать модель в заданные сроки и в рамках выделенных материальных средств.

Специалисты-постановщики должны обладать способностью из большого объема слабо формализованной разнообразной информации об объекте моделирования, из различных нечетко высказанных и сформулированных пожеланий и требований заказчика к будущей модели выделить то главное, что может быть действительно реализовано. Из перечисленных требований, предъявляемых к постановщикам задач, видно, насколько велика возложенная на них от-

ответственность и насколько могут быть тяжелы допущенные ими ошибки и просчеты. Неправильная оценка срока и стоимости реализации требуемой модели может привести к неудаче всего проекта, к напрасной потере времени и средств.

Специалисты, предрасположенные к работе в качестве постановщиков задач, особенно ценятся и являются, без преувеличения, золотым фондом научных коллективов. По этому поводу Г. Биркгоф отмечает, что прикладники-математики, *«способные к глубокому общению с другими учеными и инженерами и знакомые с мощью и ограничениями цифровых машин, ... призваны стать вождями завтрашнего математического мира, но их будет крайне трудно найти и развить»* [12]. С учетом данного высказывания, а также имея в виду конечную цель деятельности рабочей группы – построение математической модели, – представляется целесообразным рекомендовать в качестве руководителя группы именно прикладника-математика.

Этап обследования проводится членами рабочей группы под руководством постановщиков задач и включает следующие работы:

- тщательное обследование собственно объекта моделирования с целью выявления основных факторов, механизмов, влияющих на его поведение, определения соответствующих параметров, позволяющих описывать моделируемый объект;
- сбор и проверка имеющихся экспериментальных данных об объектах-аналогах, проведение при необходимости дополнительных экспериментов;
- аналитический обзор литературных источников, анализ и сравнение между собой построенных ранее моделей данного объекта (или подобных рассматриваемому объекту);
- анализ и обобщение всего накопленного материала, разработка общего плана создания математической модели.

На основе собранной информации об объекте моделирования постановщики совместно с заказчиком формулируют содержательную постановку задачи моделирования, которая, как правило, не бывает окончательной и может уточняться и конкретизироваться в процессе разработки модели. Однако, с учетом изложенного выше, все последующие уточнения и изменения содержательной постановки должны носить частный, не принципиальный характер. Если объектом моделирования является технологический процесс, машина, конструкция или деталь, то содержательную постановку задачи моделирования очень часто называют технической постановкой задачи.

Весь собранный в результате обследования материал о накопленных к данному моменту знаниях об объекте, содержательная постановка задачи моделирования, дополнительные требования к реализации модели и представлению результатов оформляются в виде *технического задания на проектирование и разработку модели*.

Техническое задание является итоговым документом, заканчивающим этап обследования, который, как уже отмечалось, является очень важным и ответственным. Чем более полную информацию удастся собрать об объекте на этапе обследования, тем более четко можно выполнить содержательную постановку задачи, более полно учесть накопленный опыт и знания, избежать многих сложностей на последующих этапах разработки модели. Особенно строго необходимо формулировать требования к будущей модели. Неконкретные и нечеткие требования могут серьезно затруднить процесс сдачи модели заказчику, вызвать бесконечные доработки и улучшения. В целом этап проработки технического задания может составлять до 30% времени, отпущенного на создание всей модели, и даже более – с учетом возможного уточнения и переформулировки.

Понимая огромную важность рассматриваемого этапа, техническое задание следует подвергать внутренней (внутри организации) и внешней экспертизе независимыми экспертами, не участвующими в его разработке. Обязательным условием на этапе разработки технического задания является участие в его обсуждении всех членов рабочей группы. Ниже приведен пример содержательной постановки задачи о баскетболистке.

Пример. *Содержательная постановка задачи о баскетболистке.*

Разработать математическую модель, позволяющую описать полет баскетбольного мяча, брошенного игроком в баскетбольную корзину.

Модель должна позволять:

- вычислять положение мяча в любой момент времени;
- определять точность попадания мяча в корзину после броска при различных начальных параметрах.

Исходные данные:

- масса и радиус мяча;
- начальные координаты, начальная скорость и угол броска мяча;
- координаты центра и радиус корзины.

2.2. КОНЦЕПТУАЛЬНАЯ ПОСТАНОВКА ЗАДАЧИ МОДЕЛИРОВАНИЯ

Наилучшие гипотезы — это простые гипотезы, которые легко подтвердить экспериментально, если они верны, и легко опровергнуть с помощью надлежащим образом подобранных решающих экспериментов или наблюдений, если они неверны.

Н.Бейли

В отличие от содержательной концептуальная постановка задачи моделирования, как правило, формулируется членами рабочей группы без привлечения представителей заказчика, на основании разработанного на предыдущем этапе технического задания, с использованием имеющихся знаний об объекте моделирования и требований к будущей модели.

Анализ и совместное обсуждение членами рабочей группы всей имеющейся информации об объекте моделирования позволяет сформировать содержательную модель объекта, являющуюся синтезом когнитивных моделей, сложившихся у каждого из членов рабочей группы. На основании содержательной модели разрабатывается *концептуальная*, или «естественно-научная» (физическая, химическая, биологическая и т.д.), постановка задачи моделирования, служащая основой для *концептуальной модели объекта*.

Концептуальная постановка задачи моделирования — это сформулированный в терминах конкретных дисциплин (физики, химии, биологии и т.д.) перечень основных вопросов, интересующих заказчика, а также совокупность гипотез относительно свойств и поведения объекта моделирования.

Наибольшие трудности при формулировке концептуальной постановки приходится преодолевать в моделях, находящихся на «стыке» различных дисциплин. Различия традиций, понятий и языков, используемых для описания одних и тех же объектов, являются очень серьезными препятствиями, возникающими при создании «междисциплинарных» моделей. Например, такие понятия как «прибыль» и «баланс» вызывают совершенно разные ассоциации у экономиста и математика-прикладника.

Можно сказать, что когнитивные модели, стоящие за этими понятиями, у этих двух специалистов совершенно различны. Если экономист, говоря о прибыли и балансе, связывает с этими понятиями конкретное производство, цену и себестоимость продукции, то для математика данные понятия выглядят более формально — как

результаты решения некоторых математических уравнений. При этом практически невозможно научить математика мыслить как экономиста, а экономиста – как математика. И тот, и другой способ восприятия имеет свои достоинства и недостатки. Экономист никогда не сделает ошибок, которые может допустить математик, обращаясь с параметрами модели формально, без должных знаний в рассматриваемой предметной области. В то же время, используя формальные преобразования математических соотношений, математик может получить решения, которые очень сложно получить экономисту, пользуясь своими подходами и методами (обычно более простыми с точки зрения математики). Поэтому эффективность деятельности рабочей группы в большой степени зависит от способности ее членов поставить себя на место специалиста другого профиля, изучить его точку зрения (т.е. особенности его когнитивной модели) и найти некоторый компромисс, учитывающий все ценное.

Выше отмечалось, что концептуальная модель строится как некоторая идеализированная модель объекта, записанная в терминах конкретных (например, естественно-научных) дисциплин. Для этого формулируется совокупность гипотез о поведении объекта, его взаимодействии с окружающей средой, изменении внутренних параметров. Как правило, эти гипотезы правдоподобны в том смысле, что для их обоснования могут быть приведены некоторые теоретические доводы и использованы экспериментальные данные, основанные на собранной ранее информации об объекте. В выборе и обосновании принимаемых гипотез в значительной степени проявляется искусство, опыт и знания, накопленные членами рабочей группы. Согласно принятым гипотезам определяется множество параметров, описывающих состояние объекта, а также перечень законов, управляющих изменением и взаимосвязью этих параметров между собой.

Пример. *Концептуальная постановка задачи о баскетболисте.*

Движение баскетбольного мяча может быть описано в соответствии с законами классической механики Ньютона (рис. 2.2).

Примем следующие гипотезы:

- объектом моделирования является баскетбольный мяч радиуса R ;
- мяч будем считать материальной точкой массой m , положение которой совпадает с центром масс мяча;
- движение происходит в поле сил тяжести с постоянным ускорением свободного падения g и описывается уравнениями классической механики Ньютона;

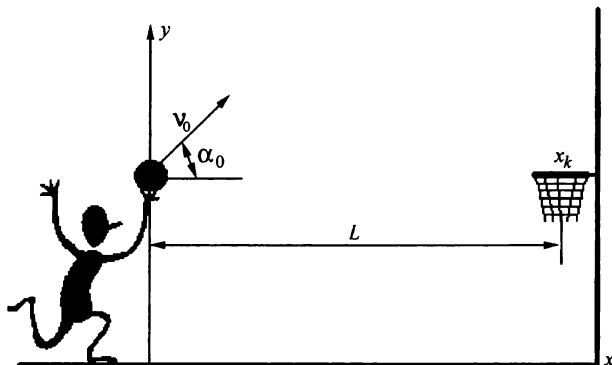


Рис. 2.2. Схема к постановке задачи о баскетболисте

- движение мяча происходит в одной плоскости, перпендикулярной поверхности Земли и проходящей через точку броска и центр корзины;
- пренебрегаем сопротивлением воздуха и возмущениями, вызванными собственным вращением мяча вокруг центра масс.

В соответствии с изложенными гипотезами в качестве параметров движения мяча можно использовать координаты (x и y) и скорость (ее проекции v_x и v_y) центра масс мяча. Тогда для определения положения мяча в любой момент времени достаточно найти закон движения центра масс мяча, т.е. зависимость координат x , y и проекций вектора скорости v_x и v_y центра мяча от времени. В качестве оценки точности броска Δ можно рассматривать величину расстояния по горизонтали (вдоль оси x) от центра корзины до центра мяча в момент, когда последний пересекает горизонтальную плоскость, проходящую через плоскость кольца корзины.

С учетом вышеизложенного можно сформулировать концептуальную постановку задачи о баскетболисте в следующем виде: определить закон движения материальной точки массой m под действием силы тяжести, если известны начальные координаты точки x_0 и y_0 , ее начальная скорость v_0 и угол бросания α_0 . Центр корзины имеет координаты x_k и y_k . Вычислить точность броска $\Delta = x(t_k) - x_k$, где t_k определяется из условий: $t_k > 0$, $v_y < 0$, $y(t_k) = y_k$.

Рассмотрим особенности приведенной в примере концептуальной постановки задачи о баскетболисте. Первая из перечисленных гипотез особенно важна, так как она выделяет объект моделирования. В данном случае объект можно считать простым. Однако в качестве объекта моделирования можно рассматривать систему «игрок – мяч – кольцо». Требуемая для описания подобной системы модель будет уже

намного сложнее, так как игрок в свою очередь представляет сложную биомеханическую систему и его моделирование является далеко не тривиальной задачей. В данной ситуации выбор в качестве объекта моделирования только мяча обоснован, поскольку именно его движение требуется исследовать, а влияние игрока можно учесть достаточно просто через начальные параметры броска. Для сложных систем выбор объекта моделирования – далеко не простая и неоднозначная задача.

Гипотеза о том, что мяч можно считать материальной точкой, широко применяется для исследования движений тел в механике. В рассматриваемом случае она оправдана в силу симметрии формы мяча и малости его радиуса по сравнению с характерными расстояниями перемещения мяча. Предполагается, что последний является шаром с одинаковой толщиной стенки.

Гипотезу о применимости в данном случае законов классической механики можно обосновать огромным экспериментальным материалом, связанным с изучением движения тел вблизи поверхности Земли со скоростями много меньше скорости света. Учитывая, что высота полета мяча лежит в пределах 5–10 м, а дальность – 5–20 м, предположение о постоянстве ускорения свободного падения также представляется обоснованным. Если бы моделировалось движение баллистической ракеты при дальности и высоте полета более 100 км, то пришлось бы учитывать изменение ускорения свободного падения в зависимости от высоты и широты места.

Гипотеза о движении мяча в плоскости, перпендикулярной поверхности Земли, ограничивает класс рассматриваемых траекторий и значительно упрощает модель. Траектория мяча может не лежать в одной плоскости, если при броске он сильно подкручивается вокруг вертикальной оси. В этом случае скорости точек поверхности мяча относительно воздуха на различных сторонах мяча будут различны. Для точек, движущихся навстречу потоку, относительная скорость выше, а для точек противоположной стороны, движущихся по потоку, – ниже скорости центра масс мяча. В соответствии с законом Бернулли, давление газа на поверхность больше там, где его относительная скорость меньше. Поэтому для ситуации, изображенной на рис. 2.3, на мяч будет действовать дополнительная сила, направленная (для данной схемы) сверху вниз. Этот эффект будет проявляться тем больше, чем больше скорость центра масс мяча и скорость его вращения. Для баскетбола характерны относительно низкие скорости полета мяча (до 10 м/с). При этом довольно редко используется подкрутка мяча рукой. Поэтому гипотеза о движении мяча в одной плоскости кажется оправданной. Ее

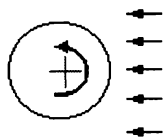


Рис. 2.3. Обтекание мяча потоком воздуха

использование позволяет отказаться от построения значительно более сложной трехмерной модели движения мяча.

Гипотеза об отсутствии влияния сопротивления воздуха наименее обоснована. При движении тела в газе или жидкости сила сопротивления увеличивается с ростом скорости движения. Учитывая невысокие скорости движения мяча, его правильную обтекаемую форму и малые дальности бросков, указанная гипотеза может быть принята в качестве первого приближения.

Следует отметить, что концептуальная постановка задачи моделирования в отличие от содержательной постановки использует терминологию конкретной дисциплины (в рассматриваемом случае – механики). При этом моделируемый реальный объект (мяч) заменяется его механической моделью (материальной точкой). Фактически в приведенном примере концептуальная постановка свелась к постановке классической задачи механики о движении материальной точки в поле сил тяжести. Концептуальная постановка более абстрактна по отношению к содержательной, так как материальной точке можно сопоставить произвольный материальный объект, брошенный под углом к горизонту: футбольный мяч, ядро, камень или артиллерийский снаряд.

2.3. МАТЕМАТИЧЕСКАЯ ПОСТАНОВКА ЗАДАЧИ МОДЕЛИРОВАНИЯ

Задача любого вида сводится к математической задаче.

Р.Декарт

Законченная концептуальная постановка позволяет сформулировать математическую постановку задачи моделирования, включающую совокупность различных математических соотношений, описывающих поведение и свойства объекта моделирования.

Математическая постановка задачи моделирования – это совокупность математических соотношений, описывающих поведение и свойства объекта моделирования.

Как было отмечено в гл. 1, совокупность математических соотношений определяет вид оператора модели. Наиболее простым будет оператор модели в случае, если он представлен системой алгебраических уравнений. Подобные модели можно назвать моделями *аппроксимационного типа*, так как для их получения часто используют различные методы аппроксимации имеющихся экспериментальных данных о поведении выходных параметров объекта

моделирования в зависимости от входных параметров и воздействий внешней среды, а также от значений внутренних параметров объекта.

Однако область применения моделей подобного типа ограничена. Для создания математических моделей сложных систем и процессов, применимых для широкого класса реальных задач требуется, как уже отмечалось выше, привлечение большого объема знаний, накопленных в рассматриваемой дисциплине (а в некоторых случаях и в смежных областях). В большинстве дисциплин (особенно естественно-научных) эти знания сконцентрированы в аксиомах, законах, теоремах, имеющих четкую математическую формулировку.

Следует отметить, что во многих областях знаний (механике, физике, биологии и т.д.) принято выделять законы, справедливые для всех объектов исследования данной области знаний, и соотношения, описывающие поведение отдельных объектов или их совокупностей. К числу первых в физике и механике относятся, например, уравнения баланса массы, количества движения, энергии и т.д., справедливые при определенных условиях для любых материальных тел, независимо от их конкретного строения, структуры, состояния, химического состава. Уравнения этого класса подтверждены огромным количеством экспериментов, хорошо изучены и в силу этого применяются в соответствующих математических моделях как данность. Соотношения второго класса в физике и механике называют *определяющими*, или *физическими* уравнениями, или *уравнениями состояния*. Они устанавливают особенности поведения материальных объектов или их совокупностей (например, жидкостей, газов, упругих или пластических сред и т.д.) при воздействиях различных внешних факторов.

В качестве классических примеров определяющих соотношений можно привести закон Гука в теории упругости или уравнение Клапейрона для идеальных газов. Очевидно, определяющие соотношения должны отражать реальное атомно-молекулярное строение исследуемых материальных объектов.

Соотношения второго класса гораздо менее изучены, а в ряде случаев их приходится устанавливать самому исследователю (особенно при анализе объектов, состоящих из новых материалов). Необходимо отметить, что определяющие соотношения – это основной элемент, «сердцевина» любой математической модели физико-механических процессов. Именно ошибки в выборе или установлении определяющих соотношений приводят к количествен-

но (а в некоторых случаях и качественно) неверным результатам моделирования.

Совокупность математических соотношений указанных двух классов определяет оператор модели. В большинстве случаев оператор модели включает в себя систему обыкновенных дифференциальных уравнений (ОДУ), дифференциальных уравнений в частных производных (ДУЧП) и/или интегродифференциальных уравнений (ИДУ). Для обеспечения корректности постановки задачи к системе ОДУ или ДУЧП добавляются начальные и/или граничные условия, которые, в свою очередь, могут быть алгебраическими или дифференциальными соотношениями различного порядка.

Можно выделить несколько наиболее распространенных типов задач для систем ОДУ или ДУЧП:

- *задача Коши, или задача с начальными условиями*, в которой по заданным в начальный момент времени переменным (начальным условиям) определяются значения этих искомым переменных для любого момента времени;
- *начально-граничная, или краевая, задача*, когда условия на искомую функцию выходного параметра задаются в начальный момент времени для всей пространственной области и на границе последней в каждый момент времени (на исследуемом интервале);
- *задачи на собственные значения*, в формулировку которых входят неопределенные параметры, определяемые из условия качественного изменения поведения системы (например, потеря устойчивости состояния равновесия или стационарного движения, появление периодического режима, резонанс и т.д.).

Для контроля правильности полученной системы математических соотношений требуется проведение ряда обязательных проверок [13]:

- *Контроль размерностей*, включающий правило, согласно которому приравниваться и складываться могут только величины одинаковой размерности. При переходе к вычислениям данная проверка сочетается с контролем использования одной и той же системы единиц для значений всех параметров.
- *Контроль порядков*, состоящий из грубой оценки сравнительных порядков складываемых величин и исключением малозначимых параметров. Например, если для выражения $x + y + z = 0$ в результате оценки установлено, что в рассмат-

риваемой области значений параметров модели $|z| \ll |x|$ и $|z| \ll |y|$, то третьим слагаемым в исходном выражении можно пренебречь.

- *Контроль характера зависимостей* заключается в проверке того, что направление и скорость изменения выходных параметров модели, вытекающие из выписанных математических соотношений, такие, как это следует непосредственно из «физического» смысла изучаемой модели.
- *Контроль экстремальных ситуаций* — проверка того, какой вид принимают математические соотношения, а также результаты моделирования, если параметры модели или их комбинации приближаются к предельно допустимым для них значениям, чаще всего к нулю или бесконечности. В подобных экстремальных ситуациях модель часто упрощается, математические соотношения приобретают более наглядный смысл, упрощается их проверка. Например, в задачах механики деформируемого твердого тела деформация материала в исследуемой области в изотермических условиях возможна лишь при приложении нагрузок, отсутствие же нагрузок должно приводить к отсутствию деформаций.
- *Контроль граничных условий*, включающий проверку того, что граничные условия действительно наложены, что они использованы в процессе построения искомого решения и что значения выходных параметров модели на самом деле удовлетворяют данным условиям.
- *Контроль физического смысла* — проверка физического или иного, в зависимости от характера задачи, смысла исходных и промежуточных соотношений, появляющихся по мере конструирования модели.
- *Контроль математической замкнутости*, состоящий в проверке того, что выписанная система математических соотношений дает возможность, притом однозначно, решить поставленную математическую задачу. Например, если задача свелась к отысканию n неизвестных из некоторой системы алгебраических или трансцендентных уравнений, то контроль замкнутости состоит в проверке того факта, что число независимых уравнений должно быть n . Если их меньше n , то надо установить недостающие уравнения, а если их больше n , то либо уравнения зависимы, либо при их составлении допущена ошибка. Однако если уравнения получаются из эксперимента или в результате наблюдений, то возможна

постановка задачи, при которой число уравнений превышает n , но сами уравнения удовлетворяются лишь приближенно, а решение ищется, например, по методу наименьших квадратов. Неравенств среди условий также может быть любое число, как это бывает, например, в задачах линейного программирования.

Свойство математической замкнутости системы математических соотношений тесно связано с введенным Ж. Адамаром понятием корректно поставленной математической задачи [93], т.е. задачи, для которой решение существует, оно *единственно* и *непрерывно* зависит от исходных данных. В данном случае решение считается непрерывным, если малому изменению исходных данных соответствует достаточно малое изменение решения.

Понятие корректности задачи имеет большое значение в прикладной математике. Например, численные методы решения оправдано применять лишь к корректно поставленным задачам. При этом далеко не все задачи, возникающие на практике, можно считать корректными (например, так называемые обратные задачи). Доказательство корректности конкретной математической задачи — достаточно сложная проблема, она решена только для некоторого класса математически поставленных задач. Проверка математической замкнутости является менее сложной по сравнению с проверкой корректности математической постановки. В настоящее время активно исследуются свойства некорректных задач, разрабатываются методы их решения. Аналогично понятию «корректно поставленная задача» можно ввести понятие «корректная математическая модель».

Математическая модель является *корректной*, если для нее осуществлен и получен положительный результат всех контрольных проверок: размерности, порядков, характера зависимостей, экстремальных ситуаций, граничных условий, физического смысла и математической замкнутости.

Пример. Математическая постановка задачи о баскетболисте.

Математическую постановку задачи о баскетболисте можно представить как в векторной, так и в координатной форме (рис. 2.4).

1. *Векторная форма.*

Найти зависимости векторных параметров от времени — $r(t)$ и $v(t)$ — из решения системы обыкновенных дифференциальных уравнений

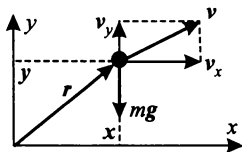


Рис. 2.4. Расчетная схема

$$m \frac{dv}{dt} = mg, \quad v = \frac{dr}{dt} \quad (2.1)$$

при начальных условиях

$$r(0) = 0, \quad v(0) = v_0. \quad (2.2)$$

Вычислить параметр Δ по формуле

$$\Delta = r_x(t_k) - r_{xk}, \quad (2.3)$$

где t_k определить из следующих условий:

$$t_k > 0, \quad v_y(t_k) < 0, \quad y(t_k) = y_k. \quad (2.4)$$

Проецируя векторные соотношения (2.1)–(2.4) на оси координат, получим математическую постановку задачи о баскетболисте в координатной форме.

2. Координатная форма.

Найти зависимости $x(t)$, $y(t)$ и $v_x(t)$, $v_y(t)$ из решения системы дифференциальных уравнений:

$$\begin{aligned} m \frac{dv_x}{dt} &= 0, & v_x &= \frac{dx}{dt}, \\ m \frac{dv_y}{dt} &= -mg, & v_y &= \frac{dy}{dt} \end{aligned} \quad (2.5)$$

при следующих начальных условиях:

$$\begin{aligned} x(0) &= x_0, & y(0) &= y_0, \\ v_x(0) &= v_0 \cos \alpha_0, & v_y(0) &= v_0 \sin \alpha_0. \end{aligned} \quad (2.6)$$

Вычислить параметр Δ по формуле

$$\Delta = x(t_k) - x_k, \quad (2.7)$$

где t_k определить из условий

$$t_k > 0, \quad v_y(t_k) < 0, \quad y(t_k) = y_k. \quad (2.8)$$

Как можно видеть, с математической точки зрения задача о баскетболисте свелась к задаче Коши для системы ОДУ первого порядка с заданными начальными условиями. Полученная система уравнений является замкнутой, так как число независимых уравнений (четыре дифференциальных и два алгебраических) равно числу искомых параметров задачи (x , y , v_x , v_y , Δ , t_k). Выполним контроль размерностей задачи:

уравнение динамики

$$m \frac{dv}{dt} = \sum F \Rightarrow [\text{кг}] \frac{[\text{м/с}]}{[\text{с}]} = [H] \Rightarrow \left[\frac{\text{кг} \cdot \text{м}}{\text{с}^2} \right] = \left[\frac{\text{кг} \cdot \text{м}}{\text{с}^2} \right],$$

связь скорости и перемещения

$$\frac{dr}{dt} = v \Rightarrow \frac{[\text{м}]}{[\text{с}]} = \left[\frac{\text{м}}{\text{с}} \right].$$

Существование и единственность решения задачи Коши доказана математиками. Поэтому данную математическую модель можно считать корректной.

Математическая постановка задачи еще более абстрактна, чем концептуальная, так как сводит исходную задачу к чисто математической (например, к задаче Коши), методы решения которой достаточно хорошо разработаны. Умение свести исходную проблему к известному классу математических задач и обосновать правомочность такого сведения требует высокой квалификации математика-прикладника и особенно высоко ценится в исследовательских коллективах.

2.4. ВЫБОР И ОБОСНОВАНИЕ ВЫБОРА МЕТОДА РЕШЕНИЯ ЗАДАЧИ

Метод решения хорош, если с самого начала мы можем предвидеть — и далее подтвердить это, — что следуя этому методу, мы достигнем цели.

Г. Лейбниц

При использовании разработанных математических моделей, как правило, требуется найти зависимость некоторых неизвестных заранее параметров объекта моделирования (например, координат и скорости центра масс тела, точности броска), удовлетворяющих определенной системе уравнений. Таким образом, поиск решения задачи сводится к отысканию некоторых зависимостей искомых величин от исходных параметров модели. Как уже было отмечено в гл. 1, все методы решения задач, составляющих «ядро» математических моделей, можно подразделить на аналитические и алгоритмические.

Следует отметить, что при использовании аналитических решений для получения результатов «в числах» также часто требуется

разработка соответствующих алгоритмов, реализуемых на ЭВМ. Однако исходное решение при этом представляет собой аналитическое выражение (или их совокупность). Решения же, основанные на алгоритмических методах, принципиально не сводимы к точным аналитическим решениям рассматриваемой задачи.

Выбор того или иного метода исследования в значительной степени зависит от квалификации и опыта членов рабочей группы. Как уже было отмечено, аналитические методы более удобны для последующего анализа результатов, но применимы лишь для относительно простых моделей. В случае, если математическая задача (хотя бы и в упрощенной постановке) допускает аналитическое решение, последнее, без сомнения, предпочтительнее численного (см. гл. 1). Как справедливо говорит по этому поводу А.Б. Мигдал [72], «...раньше чем пользоваться счетными машинами, задачу необходимо всесторонне исследовать аналитическими методами. Аналитические методы — «старое, но грозное оружие» — не теряют своего значения».

Алгоритмические методы сводятся к некоторому алгоритму, реализующему вычислительный эксперимент с использованием ЭВМ. Точность моделирования в подобном эксперименте существенно зависит от выбранного метода и его параметров (например, шага интегрирования). Алгоритмические методы, как правило, более трудоемки в реализации, требуют от членов рабочей группы хорошего знания методов вычислительной математики, обширной библиотеки специального программного обеспечения и мощной вычислительной техники. Современные модели на базе алгоритмических методов разрабатываются в исследовательских организациях, которые рекомендовали себя как авторитетные научные школы в соответствующей области знания.

Как отмечалось ранее, приближенные и численные методы исследования поставленных математических задач относятся к обширному разделу — современной вычислительной математике [8, 10, 36, 51, 58, 69, 94, 95, 111, 112]. Численные методы применимы лишь для корректных математических задач, что существенно ограничивает использование их в математическом моделировании.

Общим для всех численных методов является сведение математической задачи к конечномерной [94]. Это чаще всего достигается дискретизацией исходной задачи, т.е. переходом от функции непрерывного аргумента к функциям дискретного аргумента. Например, траектория центра тяжести баскетбольного мяча определяется не как непрерывная функция времени, а как табличная (дискретная) функция координат от времени, т.е. определяющая зна-

чения координат лишь для конечного числа моментов времени. Полученное решение дискретной задачи принимается за приближенное решение исходной математической задачи.

Применение любого численного метода неминуемо приводит к погрешности результатов решения задачи. Выделяют три основных составляющих возникающей погрешности при численном решении исходной задачи:

- *неустраняемая погрешность*, связанная с неточным заданием исходных данных (начальные и граничные условия, коэффициенты и правые части уравнений);
- *погрешность метода*, связанная с переходом к дискретному аналогу исходной задачи (например, заменяя производную $y'(x)$ разностным аналогом $(y(x + \Delta x) - y(x))/\Delta x$, получаем погрешность дискретизации, имеющую при $\Delta x \rightarrow 0$ порядок Δx);
- *ошибка округления*, связанная с конечной разрядностью чисел, представляемых в ЭВМ.

Естественным требованием для конкретного вычислительного алгоритма является согласованность в порядках величин перечисленных трех видов погрешностей.

Численный, или приближенный, метод реализуется всегда в виде вычислительного алгоритма. Поэтому все требования, предъявляемые к алгоритму, применимы и к вычислительному алгоритму. Прежде всего, алгоритм должен быть реализуем — обеспечивать решение задачи за допустимое машинное время. Важной характеристикой алгоритма является его точность, т.е. возможность получения решения исходной задачи с заданной точностью $\epsilon > 0$ за конечное число $Q(\epsilon)$ действий. Очевидно, чем меньше ϵ , тем больше затрачиваемое машинное время. Для очень малых значений ϵ время вычислений может быть недопустимо большим. Поэтому на практике добиваются некоторого компромисса между точностью и затрачиваемым машинным временем. Очевидно, что для каждой задачи, алгоритма и типа ЭВМ имеется свое характерное значение достигаемой точности.

Время работы алгоритма зависит от числа действий $Q(\epsilon)$, необходимых для достижения заданной точности. Для любой математической задачи, как правило, можно предложить несколько алгоритмов, позволяющих получить решение с заданной точностью, но за разное число действий $Q(\epsilon)$. Алгоритмы, включающие меньшее число действий для достижения одинаковой точности, будем называть более *экономичными*, или более *эффективными*.

В процессе работы вычислительного алгоритма на каждом акте вычислений возникает некоторая погрешность. При этом от действия к действию она может возрастать или не возрастать (а в некоторых случаях даже уменьшаться). Если погрешность в процессе вычислений неограниченно возрастает, то такой алгоритм называется *неустойчивым*, или *расходящимся*. В противном случае алгоритм называется *устойчивым*, или *сходящимся*.

Выше уже отмечалось, что вычислительная математика объединяет огромный пласт разнообразных, быстро развивающихся численных и приближенных методов, поэтому практически невозможно привести их законченную классификацию. Стремление получить более точные, эффективные и устойчивые вычислительные алгоритмы приводит к появлению многочисленных модификаций, учитывающих специфические особенности конкретной математической задачи или даже особенности моделируемых объектов.

Можно выделить следующие группы численных методов [8] по объектам, к которым они применяются:

- интерполяция и численное дифференцирование;
- численное интегрирование;
- определение корней линейных и нелинейных уравнений;
- решение систем линейных уравнений (подразделяют на прямые и итерационные методы);
- решение систем нелинейных уравнений;
- решение задачи Коши для обыкновенных дифференциальных уравнений;
- решение краевых задач для обыкновенных дифференциальных уравнений;
- решение уравнений в частных производных;
- решение интегральных уравнений.

В данном пособии нет возможности остановиться более подробно на достоинствах и недостатках тех или иных численных методах даже одной группы. В приложении 2 приведены примеры некоторых простейших вариантов вычислительных алгоритмов. Более подробные сведения по численным методам, оценкам их точности, устойчивости и эффективности можно найти в [8, 10, 94, 95, 111, 112].

Огромное разнообразие численных методов в значительной степени затрудняет выбор того или иного метода в каждом конкретном случае. Поскольку для реализации одной и той же модели можно использовать несколько альтернативных алгоритмических методов, то выбор конкретного метода производится с учетом того,

какой из них больше подходит для данной модели с точки зрения обеспечения эффективности, устойчивости и точности результатов, а также более освоен и знаком членам рабочей группы. Освоение нового метода, как правило, очень трудоемко и связано с большими временными и финансовыми затратами. При этом основные затраты связаны с разработкой и отладкой необходимого программного обеспечения для соответствующего класса ЭВМ, обеспечивающего реализацию данного метода.

Следует отметить, что вычислительная математика в определенном смысле являет собой более искусство, нежели науку (в понимании последней как области культуры, базирующейся на формальной логике). Очень часто эффективность применяемых методов, разработанных программ определяется нарабатываемыми годами и десятками лет интуитивными приемами, не обоснованными с математических позиций. В связи с этим эффективность одного и того же метода может весьма существенно отличаться при его применении различными исследователями.

Пример. Аналитическое решение задачи о баскетболисте.

Для получения решения рассмотренной выше задачи о баскетболисте в постановке (2.5)–(2.8) можно использовать как аналитические, так и численные методы. Проинтегрировав соотношения (2.5) по времени, получим

$$\begin{aligned} x(t) &= C_2 + C_1 t, & y(t) &= C_4 + C_3 t - gt^2/2, \\ v_x(t) &= C_1, & v_y(t) &= C_3 - gt. \end{aligned} \quad (2.9)$$

Константы интегрирования найдем из начальных условий (2.6). Тогда решение задачи можно записать следующим образом:

$$\begin{aligned} x(t) &= x_0 + v_0 t \cos \alpha_0, & y(t) &= y_0 + v_0 t \sin \alpha_0 - gt^2/2, \\ v_x(t) &= v_0 \cos \alpha_0, & v_y(t) &= v_0 \sin \alpha_0 - gt. \end{aligned} \quad (2.10)$$

Примем для простоты, что в момент броска мяч находится в начале координат и на одном уровне с корзиной (т.е. $x_0 = y_0 = y_k = 0$). Под дальностью L броска будем понимать расстояние вдоль оси Ox , которое пролетит мяч от точки броска до пересечения с горизонтальной плоскостью, проходящей через кольцо корзины. Из соотношений (2.10) дальность броска выразится следующим образом:

$$L = \frac{v_0^2}{g} \sin 2\alpha_0. \quad (2.11)$$

С учетом (2.7) точность броска

$$\Delta = L - x_k. \quad (2.12)$$

Например, при броске мяча со штрафной линии можно принять следующие исходные данные: $x_0 = y_0 = y_k = 0$; $x_k = 4,225$ м; $v_0 = 6,44$ м/с; $\alpha_0 = 45^\circ$. Тогда из (2.11) и (2.12) имеем $L = 4,225$ м; $\Delta = 0$ м.

Пример. Алгоритмическое решение задачи о баскетболисте.

В простейшем случае можно использовать метод Эйлера [см. приложение 2]. Алгоритм решения данной задачи на псевдокоде приведен ниже.

Алгоритм 2.1

program Задача о баскетболисте.

Данные: m, R – масса и радиус мяча;
 x_0, y_0 – начальные координаты мяча;
 v_0, α_0 – начальная скорость и угол броска мяча;
 x_k, y_k – координаты центра корзины;
 t – текущее время;
 dt – шаг по времени;
 f_x, f_y – силы, действующие на мяч;
 x, y, v_x, v_y – текущие координаты и проекции скорости мяча.

Результаты: L и Δ – дальность и точность броска.

start

```
g := 9,81
m := 0,6           R := 0,12
v0 := 6,44         alpha0 := 45
x0 := 0            y0 := 0
xk := 4,225        yk := 0
vx := v0 cos alpha
vy := v0 sin alpha
dt := 0,1
t := 0
x := x0; y := y0
while ((vy >= 0) or ((vy < 0) and (y <= yk)))
  t := t + dt
  fx := 0           } – силы, действующие на мяч
  fy := -mg
  vx := vx + fx dt/m } – компоненты скорости
  vy := vy + fy dt/m
  x := x + dt vx
  y := y + dt vy   } – координаты мяча
end while
L := x - x0
Delta := x - xk
```

stop

2.5. РЕАЛИЗАЦИЯ МАТЕМАТИЧЕСКОЙ МОДЕЛИ В ВИДЕ ПРОГРАММЫ ДЛЯ ЭВМ

Компьютеры бесподобны: за несколько минут они могут совершить такую ошибку, которую не в состоянии сделать множество людей за многие месяцы.

Лоуренс Дж. Питер

При создании различных программных комплексов, используемых для решения разнообразных исследовательских, проектно-конструкторских и управленческих задач, в настоящее время, основой, как правило, служат математические модели. В связи с этим возникает необходимость реализации модели в виде программы для ЭВМ. Процесс разработки надежного и эффективного программного обеспечения является не менее сложным, чем все предыдущие этапы создания математической модели. Успешное решение данной задачи возможно лишь при уверенном владении современными алгоритмическими языками и технологиями программирования, знании возможностей вычислительной техники, имеющегося программного обеспечения, особенностей реализации на ЭВМ методов вычислительной математики, наличии опыта решения подобных задач.

Процесс создания программного обеспечения можно разбить на ряд этапов [68]:

- составление технического задания на разработку пакета программ программного обеспечения;
- проектирование структуры программного комплекса;
- кодирование алгоритма;
- тестирование и отладка;
- сопровождение и эксплуатация.

Техническое задание на разработку программного обеспечения оформляют в виде *спецификации*. Примерная форма спецификации включает следующие семь разделов:

1. *Название задачи* – дается краткое определение решаемой задачи, название программного комплекса, указывается система программирования для его реализации и требования к аппаратному обеспечению (компьютеру, внешним устройствам и т.д.).

2. *Описание* – подробно излагается математическая постановка задачи, описываются применяемая математическая модель для за-

дач вычислительного характера, метод обработки входных данных для задач не вычислительного (логического) характера и т.д.

3. *Управление режимами работы программы* – формируются основные требования к способу взаимодействия пользователя с программой (интерфейс «пользователь-компьютер»).

4. *Входные данные* – описываются входные данные, указываются пределы, в которых они могут изменяться, значения, которые они не могут принимать, и т.д.

5. *Выходные данные* – описываются выходные данные, указывается, в каком виде они должны быть представлены (в числовом, графическом или текстовом), приводятся сведения о точности и объеме выходных данных, способах их сохранения и т.д.

6. *Ошибки* – перечисляются возможные ошибки пользователя при работе с программой (например, ошибки при вводе входных данных), указываются способы диагностики (в данном случае под диагностикой понимается выявление, обнаружение ошибок при работе программного комплекса) и защиты от этих ошибок на этапе проектирования, а также возможная реакция пользователя при совершении им ошибочных действий и реакция программного комплекса (компьютера) на эти действия.

7. *Тестовые задачи* – приводятся один или несколько тестовых примеров, на которых в простейших случаях проводится отладка и тестирование программного комплекса.

Пример. Спецификация задачи о баскетболисте.

1. *Название задачи*

Название программы Basketball

Система программирования Delphi

Компьютер IBM PC Pentium

Операционная система Windows-9x, NT

2. *Описание*

Приводится математическая постановка задачи и описание метода ее решения. (см. разд. 2.3 и 2.4).

3. *Управление режимами работы программы*

Для управления режимами работы программы необходимо использовать интерфейс Windows с использованием меню, диалоговых окон, полей ввода данных, кнопок.

4. *Входные данные*

Радиус и масса мяча, его начальные координаты и скорость, угол бросания, координаты корзины.

5. *Выходные данные*

Траектория центра мяча, расчетная величина дальности и точность броска. Выходные данные представляются в табличном и графическом виде.

6. Ошибки

При вводе исходных данных предусмотреть контроль:

- все вводимые значения должны быть положительными;
- угол бросания должен быть в пределах от 5° до 85° ;
- начальная скорость мяча должен быть в пределах от 0 до 30 м/с;
- горизонтальная координата центра корзины должна быть больше начальной горизонтальной координаты мяча.

При диагностировании перечисленных ошибок программа должна выдавать соответствующие сообщения, которые могут сопровождаться звуковым сигналом, и предлагать повторить ввод.

7. Тестовые примеры

При $x_0 = y_0 = y_k = 0$; $x_k = 4,225$; $v_0 = 6,44$; $\alpha_0 = 45^\circ$ получаем: $L = 4,225$; $\Delta = 0$.

На этапе проектирования формируется общая структура программного комплекса. Вся программа разбивается на программные модули. Для каждого программного модуля формулируются требования по реализуемым функциям и разрабатывается алгоритм, выполняющий эти функции. Определяется схема взаимодействия программных модулей, называемая схемой потоков данных программного комплекса. Разрабатывается план и задаются исходные данные для тестирования отдельных модулей и программного комплекса в целом.

Большинство программ, реализующих математические модели, состоят из трех основных частей:

- препроцессора (подготовка и проверка исходных данных модели);
- процессора (решение задачи, реализация вычислительного эксперимента);
- постпроцессора (отображение полученных результатов).

Лишь для относительно простых случаев эти три составные части могут быть оформлены в виде одной программы. При решении современных задач по моделированию поведения жидкостей, газов и твердых тел каждая из частей может включать в себя целый комплекс программ. Например, постпроцессор должен уметь представлять информацию не только в табличном, но и графическом виде (диаграммы, графики зависимости от различных параметров, отображение скалярных, векторных (тензорных) полей и т.п.). Возможности пре- и постпроцессора наиболее широко реализуются в современных системах автоматизированного проектирования (САПР), где они в значительной степени могут сократить время на получение данных и оценку результатов моделирования.

Как правило, создание современных математических моделей в какой-либо области и доведение их до программных комплексов требует значительных временных затрат (минимум 3–5 лет). Требуется время не только на освоение методик и подходов к моделированию в исследуемой области, но и на наработку библиотек программ по решению возникающих математических задач, по подготовке исходных данных и отображению получаемых результатов. Качественные, надежные, обладающие дружественным интерфейсом, легко модифицируемые и хорошо сопровождаемые программные комплексы можно создавать лишь при наличии хорошо продуманной стратегии развития программного обеспечения, обеспечивающей его модульность и совместимость по входным и выходным параметрам.

Большое значение следует придавать освоению современных технологий программирования: структурной, абстрактной, объектно-ориентированной и визуальной. Назначение любой технологии – это в первую очередь повышение надежности программного обеспечения и увеличение производительности труда программиста. Причем чем серьезней и объемней программный проект, тем большее значение приобретают вопросы использования современных технологий программирования. Пренебрежение данными вопросами может привести к значительным временным издержкам и снижению надежности программного комплекса.

Важнейшим фактором, определяющим надежность и малые сроки создания программного комплекса для решения конкретного класса задач, является наличие развитой библиотеки совместимых между собой программных модулей. Программа получается более надежной и создается за меньшие сроки при максимальном использовании стандартных программных элементов. Для эффективной разработки программного обеспечения в области математического моделирования необходимо обратить внимание на создание следующих стандартных библиотек программ:

- приближенные и численные методы (процессоры);
- средства подготовки исходных данных (препроцессоры);
- средства визуализации и представления результатов (постпроцессоры).

Разработка таких общих библиотек программ возможна лишь при стандартизации потоков передачи данных между препроцессором, процессором и постпроцессором. В простейшем случае речь может идти об унификации форматов передаваемых файлов.

2.6. ПРОВЕРКА АДЕКВАТНОСТИ МОДЕЛИ

Хороший теоретик может объяснить почти любые полученные результаты, верные или неверные, и, по крайней мере, может потерять массу времени на выяснение того, верны они или нет.

Р. Хемминг

Под *адекватностью* математической модели будет пониматься степень соответствия результатов, полученных по разработанной модели, данным эксперимента или тестовой задачи. Прежде чем переходить к проверке адекватности модели, необходимо убедиться в правильном комплексном функционировании всех алгоритмов и программ модели, выполнить независимое тестирование и отладку всех отдельных алгоритмов (например, используемых программных модулей, реализующих используемый численный метод).

Проверка адекватности модели преследует две цели:

1) убедиться в справедливости совокупности гипотез, сформулированных на этапах концептуальной и математической постановок. Переходить к проверке гипотез следует лишь после проверки использованных методов решения, комплексной отладки и устранения всех ошибок и конфликтов, связанных с программным обеспечением;

2) установить, что точность полученных результатов соответствует точности, оговоренной в техническом задании.

Проверка разработанной математической модели выполняется путем сравнения с имеющимися экспериментальными данными о реальном объекте или с результатами других, созданных ранее и хорошо себя зарекомендовавших моделей. В первом случае говорят о проверке путем сравнения с экспериментом, во втором — о сравнении с результатами решения тестовой задачи.

Решение вопроса о точности моделирования зависит от требований, предъявляемых к модели, и ее назначения. При этом должна учитываться точность получения экспериментальных результатов или особенности постановок тестовых задач. В моделях, предназначенных для выполнения оценочных и прикидочных расчетов, удовлетворительной считается точность 10–15%. В моделях, используемых в управляющих и контролирующих системах, требуемая точность может быть 1–2% и даже более.

Как правило, различают качественное и количественное совпадение результатов сравнения. При качественном сравнении требу-

ется лишь совпадение некоторых характерных особенностей в распределении исследуемых параметров (например, наличие экстремальных точек, положительное или отрицательное значение параметра, его возрастание или убывание и т.д.). Фактически при качественном сравнении оценивается совпадение лишь вида функции распределения параметров (убывающая или возрастающая, с одним экстремумом или с несколькими). Вопрос о количественном сравнении можно ставить лишь после удовлетворительного ответа на вопрос о качественном соответствии результатов. При количественном сравнении большое значение следует придавать точности исходных данных для моделирования и соответствующих им значений сравниваемых параметров.

Неадекватность результатов моделирования возможна, по крайней мере, по трем причинам:

а) значения задаваемых параметров модели не соответствуют допустимой области этих параметров, определяемой принятой системой гипотез. Например, в задаче о баскетболисте гипотезу об отсутствии сопротивления воздуха можно использовать лишь при относительно малых (менее 5 м/с) скоростях движения тела. При больших значениях начальной скорости мяча влияние силы сопротивления будет существенным;

б) принятая система гипотез верна, но константы и параметры в использованных определяющих соотношениях установлены не точно. Например, в случае задачи о баскетболисте значение ускорения свободного падения g может быть уточнено в зависимости от широты местности, где находится баскетболист;

в) неверна исходная совокупность гипотез.

Все три случая требуют дополнительного исследования как моделируемого объекта (с целью накопления новой дополнительной информации о его поведении), так и исследования самой модели (с целью уточнения границ ее применимости).

Замечание. В данном случае не анализируется влияние выбранного численного метода на точность получаемого решения, а значит, и на адекватность модели. Вопрос о сходимости алгоритма и устойчивости получаемого выбранным численным методом решения, а также накопление погрешностей, связанных с ошибками округления при использовании ЭВМ, здесь не рассматривается.

При возникновении проблем, связанных с адекватностью модели, ее корректировку требуется начинать с последовательного анализа всех возможных причин, приведших к расхождению результатов моделирования и результатов эксперимента. В первую очередь

требуется исследовать модель и оценить степень ее адекватности при различных значениях варьируемых параметров (начальных и граничных условиях, параметров, характеризующих свойства объектов моделирования). Если модель неадекватна в интересующей исследователя области параметров, то можно попытаться уточнить значения констант и исходных параметров модели. Если же и в этом случае нет положительных результатов, то единственной возможностью улучшения модели остается изменение принятой системы гипотез. Данное решение фактически означает возвращение ко второму этапу процесса разработки модели и может повлечь не только серьезное изменение математической постановки задачи, но и методов ее решения (например, переход от аналитических к численным), полной переработки программного обеспечения и нового цикла проверки модели на адекватность. Поэтому решение об изменении принятой системы гипотез должно быть всесторонне взвешено и приниматься только в том случае, если исчерпаны все прочие возможности по улучшению адекватности модели.

Примечание. Надлежит предостеречь начинающих исследователей от попыток «перепрыгнуть» рассмотренный этап моделирования, от желания быстрее перейти к решению «настоящей задачи». Как показывает собственный опыт авторов, подобный образ действий приводит к огромным временным издержкам (не говоря уже о психологических). Особенно опасной является ситуация, в которой при решении реальной задачи с использованием не проверенной должным образом модели получают правдоподобные результаты. Для других условий модель может дать качественно неверные результаты, но истоки ошибок разработчики будут искать уже не в модели...

Пример. Проверка адекватности решения задачи о баскетболисте.

Соотношения (2.10) являются аналитическим решением задачи о баскетболисте и позволяют определить значения координат и скоростей центра масс мяча в любой момент времени. Для координат x и y соотношения (2.10) есть уравнения параболы в параметрической форме. Мяч при броске движется по траектории, близкой к параболе. Поэтому в данном случае можно говорить о качественном совпадении результатов моделирования и экспериментальных данных. Вопрос о количественном совпадении результатов моделирования и эксперимента скорее всего будет решен отрицательно, так как отказ от учета силы сопротивления воздуха является грубым предположением. Для удовлетворительной оценки точности попадания мяча в корзину расхождение результатов моделирования и эксперимента не должно превышать 1–2 см. Поэтому гипотезу об отсутствии силы сопротивления воздуха в концептуальной постановке задачи заменим

новой: сила сопротивления воздуха прямо пропорциональна скорости мяча:

$$F_{\text{сопр}} = -k_{\text{сопр}}v,$$

где $k_{\text{сопр}}$ – коэффициент сопротивления, зависящий от свойств среды и формы тела.

Для тела в форме шара коэффициент сопротивления определяется по формуле Стокса:

$$k_{\text{сопр}} = 6\pi\mu R,$$

где μ – динамическая вязкость среды (для воздуха при температуре 20°C и давлении 1 атм $\mu = 0,0182 \text{ Н}\cdot\text{с}/\text{м}^2$).

В соответствии с принятыми в баскетболе правилами длина окружности мяча может изменяться от 0,749 до 0,780 м, что соответствует изменению радиуса мяча от 0,119 до 0,124 м. Масса мяча должна быть в пределах от 0,567 до 0,650 кг. Примем для определенности радиус мяча $R = 0,12 \text{ м}$, а массу мяча $m = 0,6 \text{ кг}$. Тогда коэффициент сопротивления среды составляет $k_{\text{сопр}} = 0,0412 \text{ Н}\cdot\text{с}/\text{м}$.

Соотношения (2.5) в математической постановке задачи следует заменить на новые:

$$\begin{aligned} m\frac{dv_x}{dt} &= -k_{\text{сопр}}v_x, & v_x &= \frac{dx}{dt}, \\ m\frac{dv_y}{dt} &= -mg - k_{\text{сопр}}v_y, & v_y &= \frac{dy}{dt}. \end{aligned}$$

Разделяя переменные и интегрируя с учетом начальных условий (2.6), можно получить решение в следующем виде:

$$\begin{aligned} x &= m_k v_{x0} (1 - \exp(-t/m_k)), \\ y &= m_k (m_k g + v_{y0}) (1 - \exp(-t/m_k)) - m_k g t, \\ v_x &= v_{x0} \exp(-t/m_k), \\ v_y &= (m_k g + v_{y0}) \exp(-t/m_k) - m_k g, \end{aligned} \tag{2.13}$$

где $m_k = m/k_{\text{сопр}}$.

Для определения дальности броска следует второе соотношение в (2.13) записать в виде

$$y_k = m_k (m_k g + v_0 \sin \alpha_0) (1 - \exp(-t/m_k)) - m_k g t,$$

где y_k – координата центра корзины. При $y_k = 0$ получим нелинейное уравнение относительно времени

$$(m_k g + v_0 \sin \alpha_0)(1 - \exp(-t/m_k)) = gt. \quad (2.14)$$

Определив время из решения (2.14) и подставив его в первое из соотношений (2.13), можно найти дальность броска. Однако в данном случае нет возможности определить аналитическое решение соотношения (2.14), что не позволяет построить соотношение для дальности L броска, аналогичное (2.11). В этом случае можно только сравнить результаты решения (2.10) и (2.13) для некоторых фиксированных значений v_0 и α_0 .

Сила сопротивления воздуха зависит не только от $k_{\text{сопр}}$, но и от скорости мяча. Ниже приведены данные сравнения силы сопротивления по отношению к силе тяжести.

$v, \text{ м/с}$	5	10	15	20
$F_{\text{сопр}}, \text{ Н}\cdot\text{с/м}^2$	0,206	0,412	0,618	0,824
$F_{\text{сопр}}/(mg), \%$	3,5	7,0	10,5	14,0

Как можно видеть, сила сопротивления воздуха при скорости движения мяча 20 м/с не превышает 14% величины силы тяжести. Однако даже такое незначительное воздействие на движение мяча может существенно сказаться на точности попадания. Так, при броске мяча под углом 45° с начальной скоростью 6,44 м/с дальность броска с учетом и без учета сил сопротивления будет отличаться на 17 см. При радиусе корзины $R_k = 23,3$ см разница составляет более половины R_k .

Учет сил сопротивления приводит к изменению результатов в задаче о попадании мяча в корзину. Так, при моделировании без учета сопротивления среды броска мяча со штрафной линии при начальных данных $x_0 = y_0 = y_k = 0$, $x_k = 4,225$ м, $\alpha_0 = 45^\circ$ полученный в разд. 2.4 результат для начальной скорости v_0 составил 6,44 м/с. С учетом сопротивления среды начальная скорость мяча должна быть равной 6,575 м/с, т.е. увеличится на 2,1%.

2.7. ПРАКТИЧЕСКОЕ ИСПОЛЬЗОВАНИЕ ПОСТРОЕННОЙ МОДЕЛИ И АНАЛИЗ РЕЗУЛЬТАТОВ МОДЕЛИРОВАНИЯ

Цель расчетов — не числа, а понимание.
Р. Хемминг

Дескриптивные модели, рассмотренные выше, предназначены для описания исследуемых параметров некоторого явления или про-

цесса, а также для изучения закономерностей изменения этих параметров. Эти модели могут использоваться:

- для изучения свойств и особенностей поведения исследуемого объекта при различных сочетаниях исходных данных и разных режимах;
- как моделирующие блоки в различных САПР и автоматизированных системах управления (АСУ);
- при построении оптимизационных моделей и моделей-имитаторов сложных систем и комплексов.

Модели, разрабатываемые для исследовательских целей, как правило, не доводятся до уровня программных комплексов, предназначенных для передачи сторонним пользователям. Время их существования чаще всего ограничено временем выполнения исследовательских работ по соответствующему направлению. Эти модели отличает поисковый характер, применение новых вычислительных процедур и алгоритмов, неразвитый программный интерфейс.

Модели и построенные на их основе программные комплексы, предназначенные для последующей передачи сторонним пользователям или коммерческого распространения, имеют развитый дружественный интерфейс, мощные пре- и постпроцессоры. Данные модели обычно строятся на апробированных и хорошо себя зарекомендовавших постановках и вычислительных процедурах. Программные комплексы имеют подробные и качественно составленные описания и руководства для пользователя; по всем неясным вопросам фирма-производитель проводит консультации. Однако следует помнить, что такие коммерческие модели предназначены только для решения четко оговоренного класса задач. Как правило, они не могут быть модернизированы и усовершенствованы. Так, пользователь не имеет возможности самостоятельно расширять библиотеку используемых численных методов или изменять систему исходных гипотез.

Независимо от области применения созданной модели группа разработчиков обязана провести качественный и количественный анализ результатов моделирования.

Работая с моделью, разработчики становятся специалистами в области, связанной с объектом моделирования. Они достаточно хорошо представляют свойства объекта, могут предсказать и объяснить его поведение. Поэтому всесторонний анализ результатов моделирования позволяет:

- выполнить модификацию рассматриваемого объекта, найти его оптимальные характеристики или, по крайней мере, лучшим образом учесть его поведение и свойства;

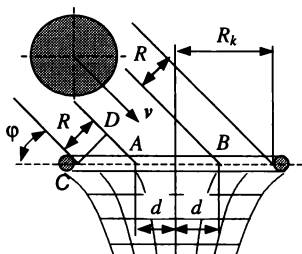
- обозначить область применения модели, что особенно важно в случае использования моделей для систем автоматического управления;
- проверить обоснованность гипотез, принятых на этапе математической постановки, оценить возможность упрощения модели с целью повышения ее эффективности при сохранении требуемой точности;
- показать, в каком направлении следует развивать модель в дальнейшем.

Пример. Анализ результатов решения задачи о баскетболисте.

Соотношения (2.10)–(2.12) представляют аналитическое решение задачи о баскетболисте без учета сил сопротивления среды, а соотношения (2.13)–(2.14) – с учетом этих сил. Достоинством первого решения является его простота, а к недостаткам можно отнести меньшую по сравнению с (2.13), (2.14) точность. Невозможность получения аналитических оценок для дальности броска следует считать недостатками решения (2.13) и (2.14). Это обстоятельство затрудняет аналитический анализ данного решения. Вместе с тем, как следует из приведенных в предыдущем разделе оценок начальной скорости броска, ее изменение для решения (2.13) и (2.14) составляет несколько процентов, что позволяет не отбрасывать решение (2.10)–(2.12) и выполнить его анализ.

Из соотношения (2.11) можно заключить, что заданную величину дальности броска можно определить при двух значениях угла бросания, обеспечивающих *настильную* (при $\alpha_0 < 45^\circ$) и *навесную* (при $\alpha_0 > 45^\circ$) траектории движения мяча. При $\alpha_0 = 45^\circ$ указанные траектории совпадают. Для обеспечения одинаковой точности (при отсутствии сопротивления) для навесной и настильной траекторий начальные скорости мяча должны быть одинаковы.

Для оценки точности попадания мяча в кольцо рассмотрим ситуации, возникающие при подлете мяча к корзине со скоростью v под углом φ к плоскости корзины (рис. 2.5).



Отрезок AB длиной $2d$ определяет ширину коридора так называемого «чистого» попадания мяча в корзину.

Задачу определения d можно свести к чисто геометрической. Для этого достаточно определить длину гипотенузы прямоугольного треугольника ACD и вычесть ее из величины внутреннего радиуса корзины:

Рис. 2.5. Схема к оценке точности броска

$$d = R_k - R/\sin \varphi. \quad (2.15)$$

Величина d получилась зависящей от угла падения φ мяча. При $d = 0$ можно найти минимальное значение φ_{\min} , при котором еще возможно «чистое» попадание мяча:

$$\sin \varphi_{\min} = R/R_k. \quad (2.16)$$

Принимая внутренний радиус кольца корзины R_k равным 0,225 м, получаем значение минимального угла $\varphi_{\min} = 32,2^\circ$.

Если рассматривать броски при условии $y_k = y_0$, то в отсутствии силы сопротивления воздуха угол падения φ мяча равен углу его бросания α_0 . В этом случае для обеспечения чистого попадания мяча в корзину угол α_0 должен быть больше $\varphi_{\min} = 32,2^\circ$.

Проведенный анализ позволяет ввести ограничение на точность броска: $-d \leq \Delta \leq d$. В предельном случае, учитывая (2.12) и выражая дальность, получим $L = x_k \pm d$. Подставив выражение для дальности (2.11) и определив из полученного соотношения начальную скорость мяча, найдем

$$v_0 = \sqrt{\frac{g}{\sin 2\alpha_0}} (x_k \pm d). \quad (2.17)$$

Соотношение (2.17) определяет интервал начальных скоростей, обеспечивающих чистое попадание мяча в корзину при заданном угле бросания.

На рис. 2.6,а приведена графическая иллюстрация соотношения (2.17). Допустимые начальные скорости мяча образуют С-образную полосу. Черная полоска соответствует чистому попаданию по модели (2.10)–(2.12). Серая полоска соответствует попаданию по модели (2.13). Изменение ширины полосок в зависимости от величины угла бросания приведено на рис. 2.6,б. Учет силы сопротивления приво-

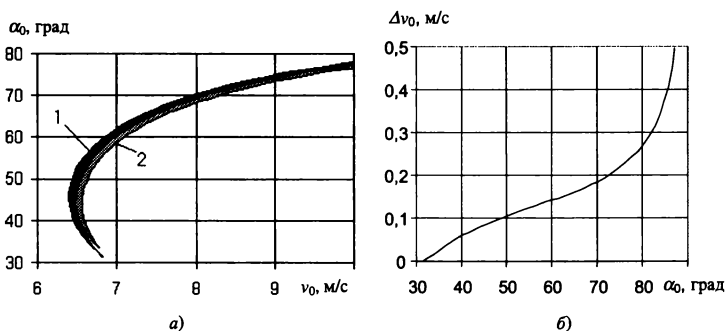


Рис. 2.6. Соотношения начальной скорости и угла бросания со штрафной линии, обеспечивающие чистое попадание мяча в корзину (1 – без учета сопротивления; 2 – с учетом сопротивления)

дит к незначительному увеличению начальной скорости мяча. При этом ширина полоски при учете сопротивления воздуха практически не изменяется.

Особенно быстро ширина полоски допустимых скоростей увеличивается при $\alpha_0 > 80^\circ$. Однако в этом случае сильно возрастает также начальная скорость мяча (при $\alpha_0 = 89^\circ$ она должна быть более 34 м/с). Можно предположить, что чем выше начальная скорость броска, тем с большей погрешностью баскетболист может ее сообщить мячу. Поэтому штрафные броски под углом $\alpha_0 > 80^\circ$ надо исключить из рассмотрения.

Приведенный здесь анализ попадания мяча в корзину при броске со штрафной линии следует считать приближенным, так как при определении допустимых скоростей и углов бросания мяча не учитывалась возможность попадания за счет отскоков от щита и кольца корзины. Более подробно эта задача будет рассмотрена в гл. 4.



Вопросы для самопроверки

1. Кто участвует в разработке содержательной постановки задачи?
2. На основании какой информации выполняется формулировка концептуальной постановки задачи моделирования?
3. Какие функции выполняет постановщик задач?
4. Какая из постановок задачи (содержательная, концептуальная или математическая) является самой абстрактной?
5. Что включает понятие корректности математической задачи?
6. Каким условиям должна удовлетворять корректная модель?
7. К каким математическим задачам можно применять численные методы?
8. Назовите три составляющие погрешности численных методов.
9. Какие цели преследует проверка адекватности модели?
10. Перечислите причины возможной неадекватности модели.
11. Для решения каких задач может быть использована математическая модель?



Задания для самостоятельного выполнения

1. Используя приведенную в данной главе «технология», разработайте математическую модель движения железнодорожного состава. В первом приближении вагоны можно считать абсолютно жесткими телами, связями между ними – линейными или нелинейно упругими.

2. Выполните содержательную, концептуальную и математическую постановки для математической модели, описывающей взлет космического аппарата с Луны.

3. Выполните содержательную, концептуальную и математическую постановки для математической модели, описывающей посадку спутника в атмосфере Земли.

4. Выполните содержательную, концептуальную и математическую постановки для математической модели, описывающей движение шарика в сферической ямке. Выберите и реализуйте метод решения полученной математической задачи. Исследуйте траектории шарика в зависимости от начальных условий.

5. Выполните содержательную, концептуальную и математическую постановки для математической модели, описывающей движение срубленного дерева.

6. Выполните содержательную, концептуальную и математическую постановки для математической модели, описывающей движение заряженной частицы в магнитном поле.

7. Космический аппарат совершает движение по орбите вокруг Земли и может быть виден некоторым наблюдателем, находящимся на ее поверхности в точке с заданными координатами. Выполните содержательную, концептуальную и математическую постановки для математической модели, описывающей движение космического аппарата по небосводу Земли с точки зрения наблюдателя.

8. Выполните содержательную, концептуальную и математическую постановки для математической модели, описывающей движение лыжника, выполняющего прыжок с трамплина.

9. Выполните содержательную, концептуальную и математическую постановки для математической модели, описывающей процесс нагревания и закипания чайника.

10. Разработайте математическую модель какого-либо процесса в интересующей вас области знаний, опишите особенности каждого из этапов моделирования, сравните их с этапами построения математической модели предыдущего задания.

Глава 3

ПРИМЕРЫ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

— Должна же я стерпеть двух-трех гусениц, — сказала Роза, — если хочу познакомиться с бабочками. Они, должно быть, прелестны.

А. де Сент-Экзюпери

Настоящая глава посвящена иллюстрации рассмотренных выше этапов разработки математических моделей на простых примерах, часть которых известна из различных предметов школьного курса. Подобный выбор примеров обусловлен стремлением с достаточной (даже для школьников старших классов и студентов младших курсов) прозрачностью показать именно «технология» создания математических моделей и в то же время обратить внимание читателей на некоторые нюансы, присущие большинству из них. Практически все приведенные здесь результаты могут быть получены самостоятельно теми, на кого ориентировано предлагаемое пособие. При этом следует отметить, что разработка математических моделей реальных процессов и явлений существенно сложнее, чем в примерах данной главы, и требует весьма глубоких знаний, интуиции, догадки.

Как уже отмечалось, аналитические модели позволяют получить искомое решение в виде конечной (или счетной) комбинации элементарных функций, что дает возможность более полно изучить свойства объекта моделирования, его *качественное* поведение, применяя хорошо развитые методы исследования аналитических функций.

В тех случаях, когда по тем или иным причинам приходится ограничиваться некоторым конечным «набором» аналитических выражений (например, несколькими членами ряда, если решения

представляют собой бесконечные ряды), не являющимся точным решением задачи, модель получается приближенной, допускающей определенные погрешности в результатах. Однако и в этом случае сохраняется возможность качественного анализа ее поведения.

Следует отметить, что в настоящее время аналитические и приближенные решения имеются для весьма небольшого числа задач, представляющих практический интерес. Как правило, прикладные научные и технические проблемы требуют учета такого громадного числа параметров и связей, в том числе нелинейных, что получение для них аналитических или приближенных решений на современном уровне развития математики не представляется возможным.

При численном подходе математические соотношения модели заменяются некоторым приближенным аналогом, что достигается *дискретизацией* принятых уравнений и в итоге заменой непрерывного решения конечным числом его значений для ограниченного числа значений аргументов.

Благодаря развитию средств вычислительной техники численные методы решения сложных прикладных задач получили широкое распространение. Задачи теплопроводности, механики жидкостей и газов, механики деформируемого твердого тела и многие другие были решены в основном благодаря широкому использованию *сеточных** методов. Вместе с тем следует иметь в виду, что некачественное применение разностных схем к решению уравнений, описывающих эволюционные процессы, приводит к получению решений, далеких от истинных. Поэтому понятен интерес к теории численных методов, позволяющей еще на стадии разработки алгоритма решения сложной инженерной задачи выяснить условия успешной реализации вычислительной процедуры модели. Основная идея решения задач численными методами на ЭВМ заключается в сведении исходной задачи к решению системы алгебраических (линейных или нелинейных) уравнений. При этом, естественно, возникает вопрос о *разрешимости* этой системы. Чис-

* В более широком смысле следует говорить о проекционных методах отыскания решения операторного уравнения в заданном подпространстве, основанных на *проектировании* искомого решения уравнения на некоторое (вообще говоря, другое, как правило, – конечномерное) подпространство. *Проекционные* методы являются основой построения различных вычислительных схем решения краевых задач. К их числу относятся такие широко известные методы, как методы Галеркина, Рунге, конечных элементов, наименьших квадратов, конечно-разностные, коллокаций и ряд других.

ленное решение *сходится* к точному, если при неограниченном увеличении числа алгебраических уравнений решение дискретизированной задачи стремится к решению исходной задачи. Поскольку решить систему с бесконечным числом алгебраических уравнений невозможно, весьма актуальным представляется вопрос об оценке *погрешности* получаемых численных решений исходной задачи.

3.1. СТАТИЧЕСКИЙ АНАЛИЗ КОНСТРУКЦИЙ

В качестве первого примера рассмотрим порядок построения простейшей математической модели, описывающей состояние конструкции в условиях равновесия. Необходимость в построении подобных моделей возникает как при анализе прочности и устойчивости конструкции в целом, так и при анализе усилий (или напряжений), возникающих в отдельных ее элементах. Если моделирование исследуемого объекта ведется в условиях равновесия, то время в качестве отдельного параметра явно в модели не учитывается. Задачи подобного типа относят к задачам статики, а анализ состояния объекта, проведенный с использованием моделей данного типа, называют статическим. Рассмотрим статический анализ однопролетного арочного моста (рис. 3.1).

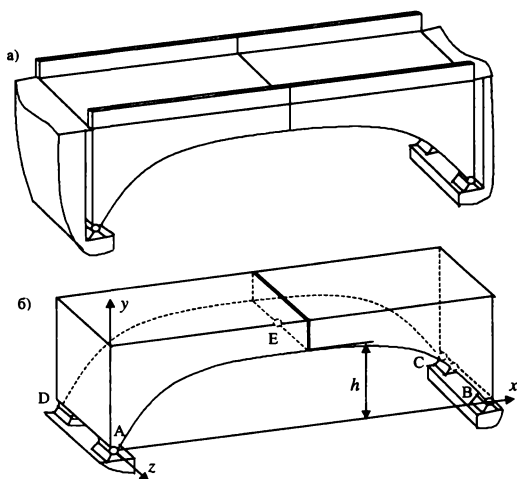


Рис. 3.1. Конструкция однопролетного арочного моста

Содержательная постановка задачи

Арка моста состоит из двух одинаковых половинок, соединенных между собой с помощью одного шарового шарнира E , расположенного на средней линии моста (рис. 3.1,б). К опорам арка крепится с помощью четырех цилиндрических шарниров A , B , C и D . Подобный способ крепления позволяет половинкам моста поворачиваться друг относительно друга, компенсируя изменение своих размеров при сезонных колебаниях температуры.

Известны геометрические размеры элементов конструкции арки моста: длина моста (расстояние между опорами A и B) $L = 14$ м; ширина моста $w = 4$ м; высота арки моста $h = 6$ м. Обе половинки моста имеют одинаковый вес $G = G_1 = G_2 = 120$ кН. Центр тяжести половинок моста расположен в продольной вертикальной плоскости симметрии моста. Линия действия силы тяжести G левой половинки арки моста пересекает плоскость XZ (рис. 3.1,б) на расстоянии $a = 1,5$ м от оси Z .

Модель должна позволять определять усилия в узлах крепления арки моста к опорам.

Концептуальная постановка задачи

Статический анализ конструкции проведем при следующих допущениях:

- объектом исследования является арка моста, состоящая из двух одинаковых половинок;
- арка крепится к опорам моста с помощью четырех цилиндрических шарниров, оси которых параллельны оси Z и лежат в плоскости XZ ;
- половинки арки моста соединены между собой шаровым шарниром, расположенным в точке E вертикальной плоскости продольной симметрии моста;
- все элементы конструкции считаются абсолютно твердыми телами, т.е. пренебрегаем их деформациями при нагружении внешними усилиями;
- пренебрегаем наличием люфта и трения в шарнирах A , B , C , D и E .

С учетом сформулированных допущений требуется определить величины реакции опор R_A , R_B , R_C , R_D арки моста.

Тела, препятствующие свободному перемещению исследуемой конструкции, называются *связями*. В данном случае в качестве связей выступают опоры арки моста. При построении модели стати-

ческого равновесия сами связи исключаются из модели, а их воздействие на исследуемый объект заменяется соответствующими усилиями, называемыми *реакциями связи*. Для цилиндрического шарнира без трения в качестве реакции связи принимается сила, лежащая в плоскости, перпендикулярной оси шарнира. Учитывая данное обстоятельство и раскладывая реакции связи шарниров по осям координат получаем, что для каждого из узлов А, В, С и D следует найти по два параметра, т.е. всего получаем восемь иско- мых величин. Полученная система сил приведена на рис. 3.2.

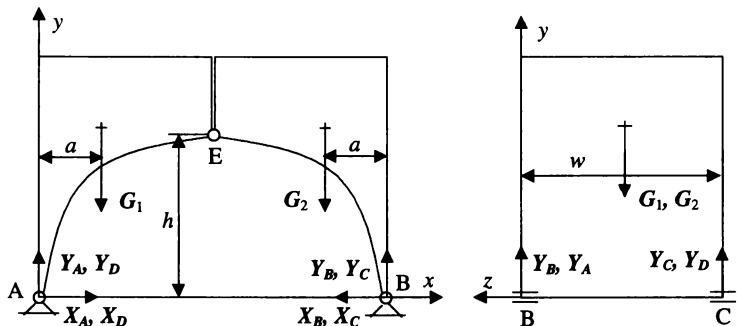


Рис. 3.2. Система сил, приложенных к аркам моста

Математическая постановка задачи

В общем случае движение твердого тела с массой m можно разложить на поступательное движение вместе с некоторой точкой P — полюсом (например, центром тяжести тела) и на вращательное движение относительно этого полюса:

$$M \frac{dV_P}{dt} = F, \quad J_P \frac{d\omega_P}{dt} = M, \quad (3.1)$$

где J_P — момент инерции тела относительно полюса; F — главный вектор приложенной к телу системы сил; M — главный момент этой системы сил.

В условиях равновесия движение тела отсутствует. Поэтому левые части соотношений (3.1) равны нулю, что позволяет сформулировать условия равновесия тела:

$$F = 0, \quad M = 0. \quad (3.2)$$

В проекциях на три оси координат система уравнений (3.2) превращается в систему шести алгебраических уравнений:

$$\begin{aligned}
 F_X &= X_A + X_D - X_B - X_C = 0, \\
 F_Y &= Y_A + Y_D + Y_B + Y_C - G_1 - G_2 = 0, \\
 F_Z &= 0, \\
 M_X &= Y_C w + Y_D w - G_1 w/2 - G_2 w/2 = 0, \\
 M_Y &= X_C w - X_D w = 0, \\
 M_Z &= Y_C L + Y_B L - G_1 a - G_2 (L - a) = 0.
 \end{aligned}
 \tag{3.3}$$

Система (3.3) состоит из шести уравнений, содержащих восемь неизвестных величин. В разд. 2.3 отмечалось, что если число неизвестных больше числа уравнений, то задача является математически незамкнутой и не может быть решена однозначно. В статике, как разделе теоретической механики, задачи подобного типа относят к статически неопределимыми.

В данном случае конструктивные особенности моста позволяют привести задачу вычисления реакций опор к статически определимой и математически корректной задаче. Половины моста связаны друг с другом только в одном узле E , являющемся шаровым шарниром. Выберем в качестве дополнительного объекта исследования, например, только левую половину арки моста. Тогда правую половину арки следует рассматривать в качестве связи, а ее воздействие на левую часть заменить соответствующей реакцией связи. Для сферического шарнира без трения реакцией связи является сила, линия действия которой проходит через центр шарнира. Данную реакцию связи можно разложить на три компоненты — X_E , Y_E и Z_E , параллельные осям координат (рис. 3.3).

Запишем систему уравнений равновесия для левой половины арки моста:

$$\begin{aligned}
 F_X &= X_A + X_D + X_E = 0, \\
 F_Y &= Y_A + Y_D + Y_E - G_1 = 0, \\
 F_Z &= Z_E = 0, \\
 M_X &= Y_D w + Y_E w/2 - G_1 w/2 = 0, \\
 M_Y &= -X_E w/2 - Z_E w/2 - X_D w = 0, \\
 M_Z &= Y_E L/2 - X_E h - G_1 a = 0.
 \end{aligned}
 \tag{3.4}$$

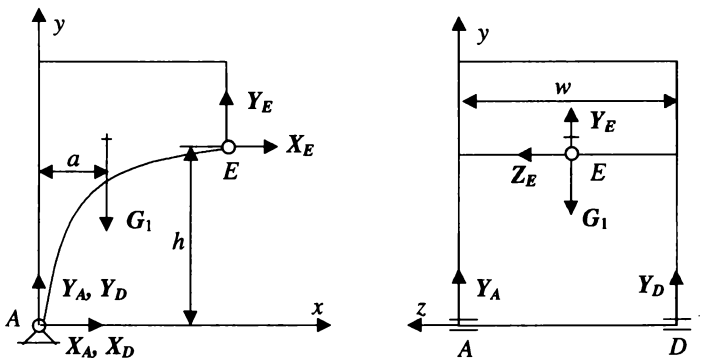


Рис. 3.3. Система сил, приложенных к левой арке моста

Вместе системы (3.3) и (3.4) имеют 12 уравнений и 11 неизвестных. Учитывая, что третье уравнение системы (3.3) является тождеством и может быть исключено, получаем систему из 11 уравнений с 11 неизвестными. Таким образом, имеем статически определенную математически корректную задачу.

Решение задачи

Перепишем объединенную систему уравнений (3.3) и (3.4) в следующем виде:

$$\begin{aligned}
 X_A + X_D - X_B - X_C &= 0, & X_A + X_D + X_E &= 0, \\
 Y_A + Y_D + Y_B + Y_C &= 2G, & Y_A + Y_D + Y_E &= G, \\
 Y_C + Y_D &= G, & Y_D + Y_E/2 - G/2 &= 0, \\
 X_C &= X_D, & X_E/2 + X_D &= 0, \\
 Y_C + Y_B &= G, & Y_E L/2 - X_E h - Ga &= 0. \\
 Z_E &= 0, & &
 \end{aligned} \tag{3.5}$$

Решение системы (3.5) можно получить непосредственно, выражая неизвестные из одних уравнений и подставляя в другие. Выполнив подобные действия, находим решение в следующем виде:

$$\begin{aligned}
 X_A = X_B = X_C = X_D &= (G/2)(a/h), \\
 Y_A = Y_C = Y_B = Y_D &= G/2, \\
 X_E = -Ga/h, & Y_E = Z_E = 0.
 \end{aligned} \tag{3.6}$$

Знак минус для компоненты X_E указывает, что направление реакции на рис. 3.3 следует изменить на обратное.

Решение системы (3.5) можно получить и с использованием численных процедур. В приложении 2 рассматривается алгоритм метода Гаусса как одного из прямых методов решения систем алгебраических линейных уравнений. В общем случае систему алгебраических уравнений в матричном виде можно представить так:

$$[A]\{X\} = \{B\}, \quad (3.7)$$

где $[A]$ – матрица коэффициентов; $\{B\}$ – вектор-столбец правой части системы уравнений; $\{X\}$ – вектор неизвестных. Необходимым условием реализации обычного метода Гаусса (без выбора главного элемента) является неравенство нулю элементов на главной диагонали матрицы $[A]$ коэффициентов системы уравнений. Выполнить данное условие достаточно просто, если расставить уравнения в такой последовательности, чтобы в i -м по порядку уравнении присутствовало слагаемое с i -й неизвестной. Например, в первом уравнении присутствовал член с X_A , во втором – с X_B и т.д. Ниже приведен пример матрицы $[A]$ и правой части $\{B\}$ для системы уравнений (3.5). Пустые клетки соответствуют нулевым значениям коэффициентов.

$[A] =$

№	X_A	X_B	X_C	X_D	Y_A	Y_B	Y_C	Y_D	X_E	Y_E	Z_E
1	1			1					1		
2	1	-1	-1	1							
3			1	-1							
4				2					1		
5					1	1	1	1			
6						1	1				
7							1	1			
8					1			1		1	
9									$2h$	L	
10								2		1	
11											1

$\{B\}^T =$

№	1	2	3	4	5	6	7	8	9	10	11
b_i	0	0	0	0	$2G$	G	G	G	$-2Ga$	G	0

Анализ результатов

Аналитическое решение (3.6) позволяет заключить, что все четыре опоры являются равнонагруженными. Данный результат можно считать ожидаемым, учитывая симметричность конструкции моста, что может служить подтверждением правильности полученных результатов. Вертикальные составляющие реакции связи каждой опоры равны $1/4$ веса всего моста. Горизонтальные составляющие реакции связи зависят от отношения a/h , т.е. реакция тем больше, чем ближе центр тяжести половинки арки расположен к шарниру E , и тем меньше, чем выше пролет моста.

Для приведенных выше исходных данных получаем $X_A = X_B = X_C = X_D = 15$ кН; $Y_A = Y_C = Y_B = Y_D = 60$ кН; $X_E = 30$ кН; $Y_E = Z_E = 0$.

3.2. МОДЕЛЬ СПРОСА–ПРЕДЛОЖЕНИЯ*

Содержательная постановка задачи

Некий фермер каждый год выращивает на своем поле пшеницу на продажу. Запасов, которые хранились бы больше года, он не делает. Решение о том, сколько пшеницы сеять, принимается с учетом цен предыдущего года. Причем если цены были высокие, то в предстоящем году надо сеять пшеницы больше, а если низкие – меньше. Спрос на пшеницу в течение года зависит от ее цены в момент продажи. Когда цена растет, спрос падает.

Необходимо описать поведение цен в ближайшие годы как функцию от первоначальной цены.

Концептуальная постановка

В качестве параметров модели используем следующие:

- p_n – цена за единицу массы пшеницы в n -й год ($p_n \geq 0$);
 - s_n – предложение (объем поставок) пшеницы в n -й год (в принятых единицах массы, $s_n \geq 0$);
 - d_n – спрос на пшеницу в n -й год (в единицах массы, $d_n \geq 0$).
- Построение модели будем вести при некоторых предпосылках:
- Объектом исследования является зависимость цены p_n на пшеницу от ее первоначальной цены p_0 .

* Условие и решение задачи заимствовано из [99].

- Предложение s_{n+1} будущего года зависит линейно от цены p_n в этом году, причем чем выше p_n , тем больше s_{n+1} :

$$s_{n+1} = ap_n - b,$$

где a и b – положительные константы, неизменные на протяжении всего анализируемого периода времени. Очевидно, что цена на пшеницу не должна быть меньше некоторой минимальной величины, покрывающей затраты на ее производство, так как лишь в этом случае величина предложения s_{n+1} будет больше нуля.

- Спрос будущего года d_{n+1} зависит линейно от цены p_{n+1} в том же году, причем чем выше цена p_{n+1} , тем меньше спрос d_{n+1} :

$$d_{n+1} = -cp_{n+1} + g,$$

где c и g – положительные константы, неизменные на протяжении всего анализируемого периода времени. Очевидно, что самый большой спрос на пшеницу должен существовать при $p_{n+1} = 0$.

- Рыночная цена p_{n+1} определяется равновесием между спросом d_{n+1} и предложением s_{n+1} .

Требуется описать поведение цен p_1, p_2, p_3, \dots в зависимости от значения цены p_0 .

Математическая постановка задачи

Считая значение p_0 заданным, найти последовательность значений p_1, p_2, p_3, \dots , удовлетворяющих следующим уравнениям:

$$s_{n+1} = ap_n - b, \quad (3.8)$$

$$d_{n+1} = -cp_{n+1} + g, \quad (3.9)$$

$$s_{n+1} = d_{n+1}, \quad (3.10)$$

где a, b, c, g – положительные вещественные числа, причем отношения b/a и g/c характеризуют, соответственно, минимально и максимально допустимые цены (при выполнении условий $s_{n+1} \geq 0$ и $d_{n+1} \geq 0$), а величина g – максимально возможный спрос (при $p_{n+1} = 0$). В результате имеем систему трех уравнений с тремя неизвестными, т.е. модель математически корректна.

Методика решения задачи

Геометрически данную задачу можно проиллюстрировать следующим образом (рис. 3.4). Зависимости спроса от цены $d(p)$ и предложения от цены $s(p)$ являются линейными и представляются на рис. 3.4 прямыми линиями. Соотношение (3.10) определяет точку пересечения P данных прямых, которую можно было бы назвать точкой сбалансированности рынка, или его равновесия.

В данном случае соотношения (3.8) и (3.9) рассматривались как непрерывные в зависимости от цены p . Однако процесс изменения цены на протяжении ряда лет образует дискретную последовательность величин p_0, p_1, p_2, \dots , которую, собственно, и необходимо найти при решении данной задачи. Определять искомую последовательность будем следующим образом.

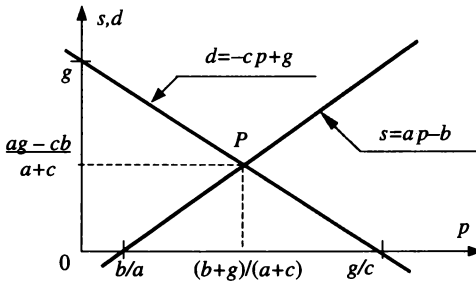


Рис. 3.4. Схема к постановке задачи

Подставляя (3.8) и (3.9) в (3.10), получим

$$ap_n - b = -cp_{n+1} + g.$$

Проведя преобразования и введя новые положительные константы

$$A = a/c > 0, \quad B = (b/c + g/c) > 0,$$

приходим к соотношению

$$p_{n+1} = -Ap_n + B. \quad (3.11)$$

Уравнение (3.11) представляет собой линейное рекуррентное соотношение, которое позволяет построить последовательность интересующих нас решений p_1, p_2, p_3, \dots . Перепишем (3.11) в виде

$$p_{n+1} + Ap_n = B. \quad (3.12)$$

Так как нас интересует функциональная зависимость $p_n(p_0)$, рассмотрим следующую схему решения. Будем искать n -е решение в виде суммы решения однородного уравнения

$$p_{n+1} + Ap_n = 0 \quad (3.13)$$

и частного решения уравнения (3.12).

Решение однородного уравнения: $p_{n+1} = -Ap_n$.

Предположим, что $p_0 = C$. Тогда

$$p_1 = C(-A); \quad p_2 = C(-A)^2; \quad p_3 = C(-A)^3; \dots$$

или в общем случае $p_n = C(-A)^n$.

Частное решение неоднородного уравнения: исходя из вида правой части (3.12) решение будем искать в виде константы $p_n = D$ для всех n . Подставим в (3.12) и получим $D + AD = B$ или $D = B/(A+1)$.

Следовательно, общее решение (3.12) имеет вид

$$p_n = C(-A)^n + B/(A+1). \quad (3.14)$$

Из (3.8) при $n = 0$ получим

$$C = p_0 - B/(A+1). \quad (3.15)$$

Подставив (3.15) в (3.14), найдем окончательное решение задачи в виде

$$p_n = p_0 (-A)^n + [B/(A+1)][1 - (-A)^n]. \quad (3.16)$$

Анализ результатов

При рассмотрении соотношения (3.16) можно выделить три характерные области значений A :

1) $0 < A < 1$. Модуль $(-A)^n$ с увеличением n стремится к нулю, следовательно, $p_n \rightarrow B/(A+1)$. Этот результат можно изобразить графически, построив зависимости функций s_{n+1} и d_{n+1} от цены p_n (рис. 3.5).

Проанализируем полученные результаты. Начнем с цены p_0 . Точка на графике $s(p)$, соответствующая p_0 , дает значение s_1 . Двигаясь горизонтально, находим значение d_1 , т.е. выполняем условие $d_1 = s_1$. Цена, соответствующая d_1 , равна p_1 . Этой цене в свою оче-

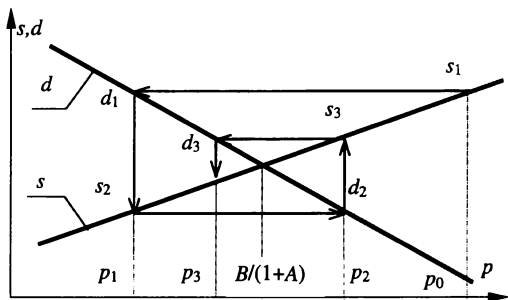


Рис. 3.5. Иллюстрация к решению задачи ($0 < A < 1$)

речь соответствует предложение s_2 . Выполняя условие $d_2 = s_2$, движемся горизонтально и находим d_2 . Процесс продолжается до стягивания полученной «паутины» к цене $B/(A + 1)$, соответствующей пересечению графиков $s(p)$ и $d(p)$.

2) $A = 1$. В этом случае уравнение (3.16) примет вид

$$p_n = p_0 (-1)^n + B[1 - (-1)^n]/2.$$

Следовательно:

$$p_n = -p_0 + B, \text{ если } n - \text{нечетное};$$

$$p_n = p_0, \text{ если } n - \text{четное}.$$

Геометрически эта ситуация показана на рис. 3.6.

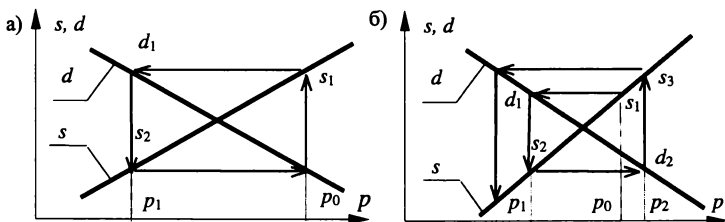


Рис. 3.6. Геометрическая иллюстрация решения задачи при $A = 1$ (а) и $A > 1$ (б)

3) $A > 1$. Из уравнения (3.16) видно, что с возрастанием n растет амплитуда колебаний p_n . Графически это показано на рис. 3.6, б.

Из рассмотренных трех случаев первый соответствует *устойчивому равновесию*, т.е. сбалансированному рынку, когда спрос соответствует предложению. Во втором случае получили *неустойчивое*

равновесие, так как идет периодическое снижение и повышение цены. В третьем случае наступает так называемый *коллапс*, когда рынок полностью разбалансирован. По аналогии с принятой в качественной теории дифференциальных уравнений терминологией, точки пересечения графиков $s(p)$ и $d(p)$ (особые точки) для трех рассмотренных случаев можно назвать соответственно устойчивым фокусом ($0 < A < 1$), центром ($A = 1$) и неустойчивым фокусом ($A > 1$).

Величина A в соотношении (3.16) представляет собой отношение коэффициентов a и c , т.е. тангенсов углов наклона прямых спроса и предложения. При устойчивом состоянии рынка ($0 < A < 1$) прямая спроса круче прямой предложения, т.е. в этом состоянии при увеличении цены на единицу спрос падает быстрее, чем растет предложение.

Необходимо подчеркнуть, что мы сделали предположение о линейной зависимости спроса и предложения от цены. Если это предположение заменить более сложной зависимостью, то задача становится намного сложнее и может вообще не иметь аналитического решения. Нелинейную зависимость спроса d и предложения s от цены p можно ввести множеством способов. Например, учитывая те факты, что любой производитель не может бесконечно наращивать объемы продукции и что спрос при низких ценах растет быстрее, чем при высоких, зависимости спроса и предложения от цены можно представить следующими выражениями:

$$s_{n+1} = ap_n^m - b, \quad d_{n+1} = g \exp(-cp_{n+1}), \quad (3.17)$$

где $a, b, c, g > 0$; $0 < m < 1$. Графически данная ситуация представлена на рис. 3.7. (Решение задачи и анализ результатов для таких условий предлагается произвести читателю.)

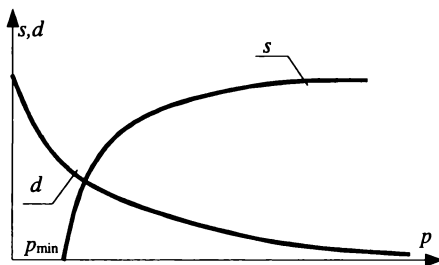


Рис. 3.7. Нелинейные зависимости спроса и предложения

3.3. ДИНАМИКА ПОПУЛЯЦИЙ

Одним из важнейших разделов современной биологии и экологии является биология популяций. Популяцией в биологии называют сообщество особей одного вида, занимающих некоторую область пространства на нашей планете. Вопросы, которые приходится решать биологии популяций, разнообразны. Например, что получится, если поместить тысячу карасей в пруд с ограниченными пищевыми ресурсами? Что изменится, если выпустить туда еще пятьдесят щук, поедающих в среднем по два карася в день? Какая судьба постигнет вирус, вызывающий гибель определенного вида животных и распространяющийся с известной скоростью, зависящей от плотности популяции? Какими темпами будет увеличиваться численность людей на Земле? Список подобных вопросов можно продолжать и продолжать.

Характерной особенностью биологии как науки является то, что одним из основных методов исследования данной дисциплины все еще остается наблюдение. Объяснить это можно множеством причин, одной из которых является сложность формализации параметров, описывающих различные виды живых существ. В то же время отличительной чертой биологии популяций является то, что для исследования динамики популяций достаточно интенсивно используется математическое моделирование.

В природе встречаются популяции организмов как с дискретным, так и с непрерывным периодом размножения. К первым можно отнести многие виды насекомых (бабочки, майские жуки, саранча), рыб, птиц и млекопитающих с привязкой периода размножения к определенным временам года. Ко вторым относятся живые организмы, процесс размножения которых не привязан ко временам года (в первую очередь это домашние животные и человек). Модели для динамики популяций с дискретным и непрерывным периодом размножения существенно отличаются. В частности, для популяций с непрерывным периодом размножения применим аппарат дифференциального исчисления. Далее будут рассматриваться только популяции с непрерывным периодом размножения.

Первая модель динамики популяций была предложена священником Томасом Мальтусом еще в 1778 г. в опубликованной им работе «Трактат о народонаселении». Хотя модель, предложенная Мальтусом, касалась народонаселения Земли, ее можно распространить на любую популяцию живых организмов. Рассмотрим эту модель более подробно.

Содержательная постановка задачи

Как будет изменяться численность популяции, если сдерживающие факторы (болезни, хищники, конкурирующие виды, ограниченность питания и т.п.) отсутствуют?

Концептуальная постановка задачи

Исследование популяции провести при следующих допущениях:

- объектом исследования является некоторая популяция организмов;
- сдерживающие факторы роста популяции отсутствуют;
- скорость прироста численности популяции прямо пропорциональна величине численности популяции.

Последние два предположения являются относительно грубыми. Их применение оправдано на довольно коротком начальном этапе развития популяции (например, при начальном развитии колонии бактерий в автоклаве при достаточно интенсивном перемешивании биомассы).

Математическая постановка задачи для модели Мальтуса

Пусть $x(t)$ — численность популяции в момент времени t . Функцией прироста $R(t)$ называют относительное изменение численности за время Δt :

$$R(t) = \frac{x(t + \Delta t) - x(t)}{x(t)\Delta t}.$$

Если эта величина — константа r , то закон, управляющий динамикой в модели Мальтуса, имеет вид

$$\frac{x(t + \Delta t) - x(t)}{\Delta t} = rx(t).$$

Переходя к пределу по Δt , получим следующее обыкновенное дифференциальное уравнение:

$$\frac{dx}{dt} = rx. \quad (3.18)$$

Итак, для решения поставленной задачи необходимо найти решение уравнения (3.18) при начальном условии

$$x(0) = x_0. \quad (3.19)$$

Решение задачи

Для решения уравнения (3.18) можно воспользоваться методом разделения переменных:

$$\int_{x_0}^x \frac{dx}{x} = r \int_0^t dt \quad \text{или} \quad \ln(x/x_0) = rt.$$

Окончательно имеем:

$$x = x_0 e^{rt}. \quad (3.20)$$

Анализ результатов

Полученное решение (3.20) по модели Мальтуса предсказывает неограниченный рост численности популяции по экспоненциальному закону. В действительности неограниченный рост невозможен, так как сдерживающие факторы присутствуют всегда. Численность популяции, как правило, испытывает небольшие колебания относительно некоторой величины.

Одним из первых обратил на это внимание П.Ф. Ферхюльст, сформулировав в 1845 г. закон, содержащий ограничение на рост популяции. Он объяснил это тем, что любая экологическая ниша может обеспечить существование популяции только определенного максимального размера x_{\max} и что коэффициент прироста должен снижаться, когда размеры популяции приближаются к x_{\max} . Будем измерять численность популяции в относительных единицах:

$$X = x/x_{\max}.$$

Тогда функцию прироста по Ферхюльсту можно записать следующим образом:

$$R(X) = r(1 - X). \quad (3.21)$$

С учетом (3.21) запишем математическую постановку задачи для модели Ферхюльста.

Математическая постановка задачи для модели Ферхюльста

Найти решение задачи Коши

$$\frac{dX}{dt} = r(1-X)X \quad (3.22)$$

при начальных условиях $X(0) = X_0$.

Решение задачи

Уравнение (3.22) можно проинтегрировать методом разделения переменных:

$$\frac{dX}{X(X-1)} = \frac{dX}{X-1} - \frac{dX}{X} = -rdt,$$

в результате чего получим решение

$$X = \frac{X_0 e^{rt}}{1 - X_0(1 - e^{rt})}. \quad (3.23)$$

Анализ результатов

На рис. 3.8 показано изменение относительной численности популяции во времени при различных начальных значениях X_0 и r (условие $X_0 > 1$ возможно, если в период времени до начала рассмотрения окружающие условия для популяции были более благоприятны).

Как можно видеть, с течением времени величина $X \rightarrow 1$. Проанализируем данное обстоятельство подробнее.

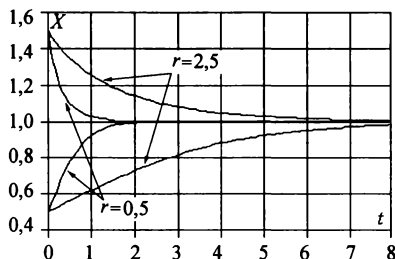


Рис. 3.8. Изменение относительной численности популяции

Рассмотрим исходное дифференциальное уравнение (3.22):

$$\dot{X} = r(1 - X)X = f(X). \quad (3.24)$$

В теории дифференциальных уравнений точки X^* , в которых правая часть уравнения $f(X)$ обращается в нуль, т.е. $f(X^*) = 0$, называются *положениями равновесия*, или *стационарными* (иногда *особыми*) точками. Положение равновесия может быть как устойчивым, так и неустойчивым. Понятия устойчивости и неустойчивости имеют большое прикладное значение и являются одними из наиболее важных понятий при анализе различных математических моделей. В частности, характер эволюции некоторой системы из состояния равновесия существенно зависит от того, каким это равновесие является: устойчивым или неустойчивым. В случае неустойчивого равновесия в результате даже очень малых начальных отклонений система может быть отброшена от стационарного состояния (механическая иллюстрация приведена на рис. 3.9), и движение станет либо очень сложным, либо система перейдет в другое стационарное состояние, весьма далекое от исходного.

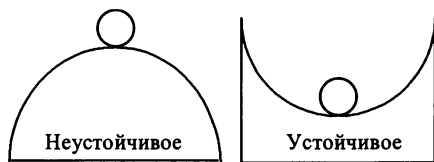


Рис. 3.9. Механическая иллюстрация положения равновесия

Говорят, что система находится в положении устойчивого равновесия X^* , если при малых отклонениях от него система останется вблизи X^* при любых t . На математическом языке это утверждение формулируется следующим образом.

Стационарная точка дифференциального уравнения называется *устойчивой*, или, точнее, *устойчивой по Ляпунову*, если для любого $\epsilon > 0$ существует $\delta(\epsilon) > 0$ такое, что для всякого решения $X(t)$ того же уравнения, начальные значения которого удовлетворяют неравенству

$$\rho(X(t_0), X^*) < \delta(\epsilon)$$

для всех $t \geq t_0$ справедливо

$$\rho(X(t), X^*) < \varepsilon.$$

Здесь ρ – расстояние (по введенной метрике) между точками X и X^* .

Если же положение равновесия X^* не только устойчиво, но, кроме того, удовлетворяет условию $\lim_{t \rightarrow \infty} \rho(X(t), X^*) = 0$, то стационарная точка X^* в этом случае называется *асимптотически устойчивой*.

Предположим, что исследуемая система в некоторый момент времени t^* находилась в состоянии $X(t^*)$, близком к состоянию равновесия X^* :

$$X(t^*) = X^* + \delta\xi(t^*), \quad (3.25)$$

где δ – малый параметр ($\delta \ll 1$), $\xi(t)$ – функция из того же класса, что и $X(t)$.

При каких условиях функция $X(t)$ и для $t > t^*$ останется близкой к X^* ? Ответ на этот вопрос дает *теорема Ляпунова об устойчивости по первому приближению*. Суть ее состоит в следующем. Рассмотрим уравнение (3.24) при условии (3.25). Тогда, переходя в (3.24) к переменной $\delta\xi(t)$ и разлагая правую часть в ряд, получим

$$\delta\dot{\xi} = f(X^* + \delta\xi) = f(X^*) + \frac{\partial f}{\partial X}(X^*)\delta\xi + \frac{1}{2} \frac{\partial^2 f}{\partial X^2}(X^*)(\delta\xi)^2 + \dots$$

Учитывая $f(X^*) = 0$, а также малость $\delta\xi(t)$ и отбрасывая нелинейные слагаемые, получим

$$\delta\dot{\xi} = \frac{\partial f}{\partial X}(X^*)\delta\xi = \lambda\delta\xi. \quad (3.26)$$

Теорема Ляпунова утверждает, что если $\lambda < 0$, то положение равновесия X^* асимптотически устойчиво по линейризованной схеме, при $\lambda > 0$ – неустойчиво, и при $\lambda = 0$ об устойчивости или неустойчивости данного положения равновесия ничего сказать нельзя.

Исследуем исходное уравнение (3.24) и найдем положения равновесия, приравняв правую часть уравнения к нулю. Получаем две точки: $X_1^* = 0$ и $X_2^* = 1$. Найдем зависимость $\delta\xi$ по соотношению (3.26):

$$\delta\dot{\xi} = r(1 - 2X^*)\delta\xi, \quad (3.27)$$

т.е. $\lambda = r(1 - 2X^*)$.

Для $X_1^* = 0$ величина $\lambda_1 = r$, т.е. если величина коэффициента прироста r популяции положительна, то данное положение равновесия является неустойчивым. Если же $r < 0$ (т.е. популяция вымирающая), то положение равновесия $X_1^* = 0$ является устойчивым. Для $X_1^* = 1$ величина $\lambda_2 = -r$, т.е. получаем обратную картину: при $r > 0$ данное положение равновесия устойчиво, а при $r < 0$ — неустойчиво.

Проведенный качественный анализ уравнения (3.24) позволяет не только оценить тенденции эволюции популяции при различных исходных данных, но и определить значимость различных параметров модели. Так, в данном случае проведенный анализ показал определяющую роль на судьбу популяции такого параметра, как коэффициент прироста g . Коэффициент прироста зависит от множества факторов, действующих как разово, так и в течение многих лет. Например, стихийные бедствия и эпидемии можно отнести к разовым воздействиям, а климатические условия — к действующим в течение длительного времени. Для популяций большой численности отрицательная величина коэффициента прироста, связанная с разовыми воздействиями, не так страшна. Этого нельзя сказать о малых популяциях. Для них велика вероятность вымирания в результате неблагоприятных погодных условий, стихийных бедствий, эпидемий и т.п.

Не менее важным обоснованием необходимости анализа устойчивости является то, что во многих случаях исследователя в большей степени интересует вопрос о самих точках равновесия, а не о том, каким образом исследуемая система перейдет в эти точки. Как правило, если внешние условия для системы остаются неизменными, то сама система достаточно быстро перейдет в одно из устойчивых состояний равновесия и будет находиться там до изменения внешних условий. Определение точек равновесия и исследование их на устойчивость для моделей, сводящихся к системе обыкновенных дифференциальных уравнений, — более простая задача, чем получение аналитического или численного решения. Рассмотрим особенности численного исследования модели Ферхюльста.

Численное исследование модели Ферхюльста

Динамику популяции, описываемую моделью Ферхюльста, в зависимости от начальных условий и коэффициента прироста можно исследовать с использованием численной процедуры. Исполь-

зую метод Эйлера, для исходного дифференциального уравнения получим следующий аналог уравнения (3.22) в конечных разностях

$$X_{n+1} = X_n + r\Delta t(1 - X_n)X_n, \quad (3.28)$$

где Δt — величина шага по времени; X_n — относительная численность популяции на начало шага по времени; X_{n+1} — численность популяции на конец шага по времени. Задаваясь начальной численностью X_0 и величиной шага по времени Δt , по соотношению (3.28) можно получить таблицу значений численности в различные моменты времени.

На рис. 3.10 показано изменение численности популяции в зависимости от величины произведения шага по времени на коэффициент прироста. При малых значениях данного произведения изменение численности (рис. 3.10, а), полученное численной процедурой метода Эйлера, стремится к положению равновесия ($X^* = 1$) и практически совпадает с аналитическим решением (3.23). Рассмотрим поведение малых отклонений $\Delta_n = X_n - X^*$ от положения равновесия. Подставляя в (3.28) и линеаризуя, получим

$$\Delta_{n+1} = \Delta_n + r\Delta t(1 - 2X^*)\Delta_n.$$

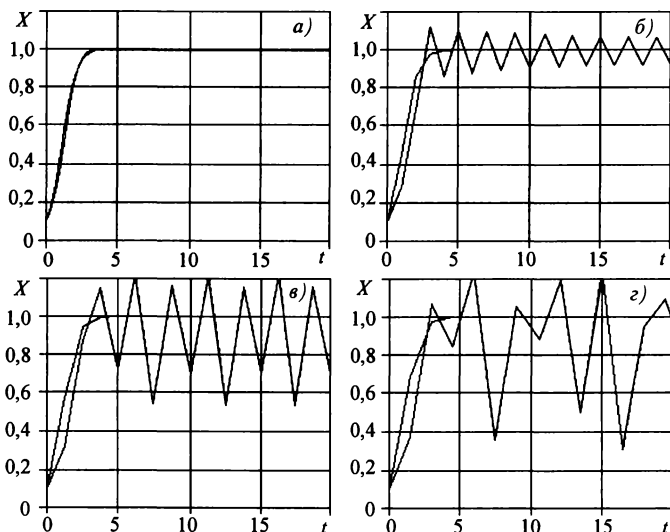


Рис. 3.10. Изменение численности популяции при различных значениях произведения $r\Delta t$: а — $r\Delta t = 0,5$; б — $r\Delta t = 2$; в — $r\Delta t = 2,5$; г — $r\Delta t = 3$

Подставляя значение X^* и деля на Δ_n , найдем

$$\Delta_{n+1}/\Delta_n = 1 - r\Delta t.$$

Из последнего соотношения можно заключить, что последовательность Δ_{n+1}/Δ_n будет сходящейся, если $0 < r\Delta t < 2$, причем при $0 < r\Delta t < 1$ сходимость будет монотонной, при $1 < r\Delta t < 2$ происходят колебания вблизи нуля. При $r\Delta t > 2$ численное решение начинает колебаться около положения равновесия (рис. 3.10, б). При $r\Delta t = 2,5$ (рис. 3.10, в) происходят устойчивые колебания с периодом 4. Дальнейший рост величины $r\Delta t$ приводит к последовательному удвоению периода колебаний при все большем сближении значений $r\Delta t$. Наконец, при $r\Delta t = 2,57$ процесс вообще перестает быть периодическим (рис. 3.10, г). Теперь X все время меняется около бесконечного числа значений, так что поведение процесса, несмотря на его полную изначальную детерминированность, практически невозможно прогнозировать на большие периоды времени. Подобное поведение обычно называют *хаотическим*.

Более полное представление о поведении численного решения можно получить с помощью так называемой бифуркационной диаграммы (рис. 3.11). На диаграмме по горизонтальной оси откладывают значения $r\Delta t$, а по вертикальной — численность популяции X . Для каждого значения $r\Delta t$ определяют значения ряда X_0, X_1, X_2, \dots соотношения (3.28), причем первые 4000 членов X_n на диаграмме не отображаются. Это делается для того, чтобы процесс успел выйти к положению равновесия. Следующие 200 членов ряда нанесены на диаграмму. Как можно видеть на диаграмме, для метода Эйлера (рис. 3.11, а) при $r\Delta t < 2$ все 200 членов ряда отображаются в одну точку, что соответствует процессу сходимости, приведенному на рис. 3.10, а. При $2 < r\Delta t < \sqrt{6}$ на диаграмме получаем две точки (процесс сходимости представлен на рис. 3.10, б), затем 4 (рис. 3.10, в), 8, 16 и т.д. точек вплоть до области хаоса, где точки могут заполнять целые полосы (рис. 3.10, г).

Если использовать метод Рунге-Кутты, имеющий четвертый порядок точности, то получим другой вид бифуркационной диаграммы (рис. 3.11, б). До значения $r\Delta t = 2,75$ решение сходится к 1. Затем точка сходимости начинает изменяться, уменьшаясь по сравнению с 1, и при значениях $r\Delta t > 3,475$ решение начинает осциллировать.

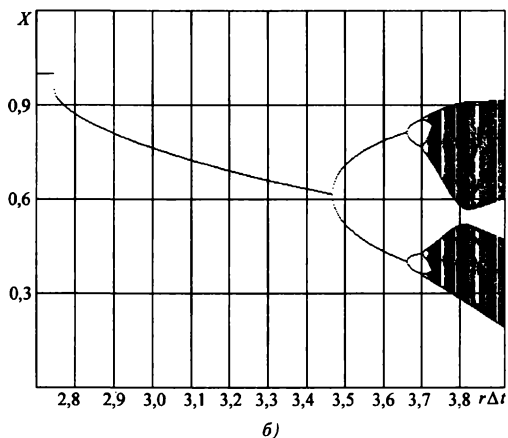
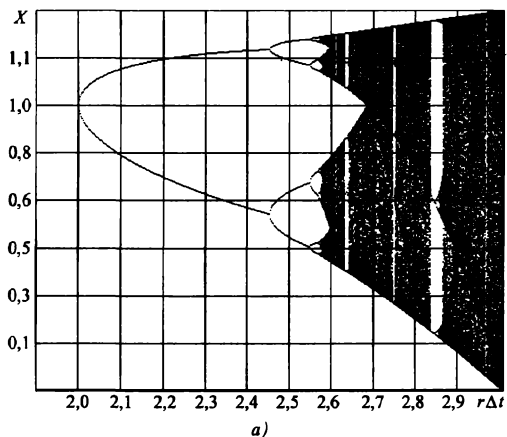


Рис. 3.11. Бифуркационная диаграмма для модели Ферхюльста:
 а – схема Эйлера; б – схема Рунге-Кутты

Рассмотренный пример показывает, что при выборе численного метода для получения приближенного решения задачи следует очень внимательно относиться к выбору шага интегрирования для соответствующих значений параметров модели. Для численного метода можно ввести понятие вычислительной устойчивости или неустойчивости. При этом нельзя отождествлять неустойчивость состояния самого объекта моделирования с неустойчивостью вы-

числительной процедуры. Надлежит таким образом выбирать параметры численной процедуры, чтобы исключить появление вычислительной неустойчивости при исследовании модели.

3.4. МОДЕЛЬ КОНКУРЕНЦИИ ДВУХ ПОПУЛЯЦИЙ

Рассмотрим ситуацию, когда в одной и той же местности проживают две популяции, имеющие схожий рацион питания. В этом случае популяции начинают конкурировать между собой из-за источника питания. Обозначим численность одной популяции через x , а другой через y . Пусть нам известны предельные численности популяций x^* и y^* , которых они могли бы достигать в данной области обитания при отсутствии конкуренции, причем обозначение популяций введем так, чтобы выполнялось ограничение $x^* \leq y^*$. Тогда функции прироста с учетом конкуренции можно записать следующим образом:

$$R_x(x, y) = k_x(x^* - x) - k_{xy}y, \quad (3.29a)$$

$$R_y(x, y) = k_y(y^* - y) - k_{yx}x, \quad (3.29б)$$

где k_x, k_y, k_{xy}, k_{yx} — положительные коэффициенты пропорциональности, учитывающие особенности каждой популяции и их взаимовлияние.

С учетом (3.29) запишем математическую постановку данной задачи.

Математическая постановка задачи

Найти решение задачи Коши

$$\begin{aligned} \frac{dx}{dt} &= k_x(x^* - x)x - k_{xy}xy, \\ \frac{dy}{dt} &= k_y(y^* - y)y - k_{yx}xy \end{aligned} \quad (3.30)$$

при начальных условиях $x(0) = x_0, \quad y(0) = y_0$.

Как и для модели Ферхюльста, уравнения (3.30) можно привести к безразмерному виду. Рассмотрим популяции схожих организ-

мов. Например, это могут быть травоядные животные двух разных видов или животные-хищники двух видов. В этом случае можно ввести относительные численности популяции:

$$X = x/x^*; \quad Y = y/x^*. \quad (3.31)$$

Тогда математическую постановку можно переписать в следующем виде: найти решение задачи Коши

$$\begin{aligned} \frac{dX}{dt} &= r_x(1 - X - \mu_x Y)X, \\ \frac{dY}{dt} &= r_y(Y^* - Y - \mu_y X)Y \end{aligned} \quad (3.32)$$

при начальных условиях $X(0) = X_0$, $Y(0) = Y_0$. Здесь $r_x = k_x x^*$, $r_y = k_y x^*$, $\mu_x = k_{xy}/k_x$, $\mu_y = k_{yx}/k_y$, причем $Y^* = y^*/x^* \geq 1$.

Качественный анализ задачи

Проведем качественный анализ данной задачи. Найдем точки равновесия для системы уравнений (3.32), приравняв правые части нулю:

$$\begin{aligned} r_x(1 - X - \mu_x Y)X &= 0, \\ r_y(Y^* - Y - \mu_y X)Y &= 0. \end{aligned}$$

Имеем четыре точки равновесия:

$$\begin{aligned} 1) X^{(1)} = Y^{(1)} = 0; \quad 2) X^{(2)} = 0, Y^{(2)} = Y^*; \\ 3) X^{(3)} = 1, Y^{(3)} = 0; \quad 4) X^{(4)} = \frac{1 - \mu_x Y^*}{1 - \mu_x \mu_y}, Y^{(4)} = \frac{Y^* - \mu_y}{1 - \mu_x \mu_y}. \end{aligned} \quad (3.33)$$

Исходя из ограничений на численность (невымирание популяций)

$$X^{(4)} > 0 \quad \text{и} \quad Y^{(4)} > 0 \quad (3.34)$$

получаем ограничения на μ_x и μ_y :

$$a) \mu_x \mu_y < 1, \quad 0 < \mu_x < 1/Y^*, \quad 0 < \mu_y < Y^*; \quad (3.35a)$$

$$\text{б) } \mu_x \mu_y > 1, \mu_x > 1/Y^*, \mu_y > Y^*. \quad (3.356)$$

Если эти условия не выполняются, то одна из популяций вымирает. Из принятых соглашений следуют ограничения на численность популяций в точке равновесия:

$$0 < X^{(4)} < 1, \quad 0 < Y^{(4)} < Y^*. \quad (3.36)$$

Из условия (3.33) следует, что $X^{(4)} < Y^{(4)}$, если $1 + \mu_y < (1 + \mu_x)Y^*$.

Для данной модели имеем два параметра X и Y , образующие двумерное фазовое пространство. Параметры X и Y называются *фазовыми переменными*. Любое решение $X(t)$, $Y(t)$ системы (3.32) можно интерпретировать геометрически как кривую в трехмерном пространстве переменных t , X , Y , которая называется *интегральной кривой*. Проекция интегральной кривой на фазовое пространство называется *фазовой траекторией*.

Уравнения (3.32) в левой части содержат производные по времени, которые можно интерпретировать как скорости изменения переменных X и Y . Поэтому можно считать, что уравнения (3.32) в каждой точке фазового пространства определяют некоторую скорость, задаваемую вектором $\tau = (f_1, f_2)$, компоненты которого задаются величиной правой части соответствующего уравнения в данной точке фазового пространства. Такая область фазового пространства с заданным в каждой точке направлением называется *полем скоростей*, или *полем направлений*. Пример такого поля для системы (3.32) приведен на рис. 3.12.

Как можно видеть, при приближении к точкам равновесия длина вектора скорости уменьшается и в самой точке равновесия становится равной нулю. Геометрически интегрирование системы уравнений (3.32) сводится к нахождению кривых, у которых направление касательной в каждой точке совпадает с направлением вектора τ в данной точке поля направлений.

Проверка устойчивости точек равновесия для систем дифференциальных уравнений имеет свои особенности. Согласно теореме Ляпунова скорости изменения малого параметра при разложении в ряд в точке равновесия имеют вид

$$\begin{aligned} \delta \xi_1 &= \frac{\partial f_1}{\partial X} \delta \xi_1 + \frac{\partial f_1}{\partial Y} \delta \xi_2, \\ \delta \xi_2 &= \frac{\partial f_2}{\partial X} \delta \xi_1 + \frac{\partial f_2}{\partial Y} \delta \xi_2. \end{aligned}$$

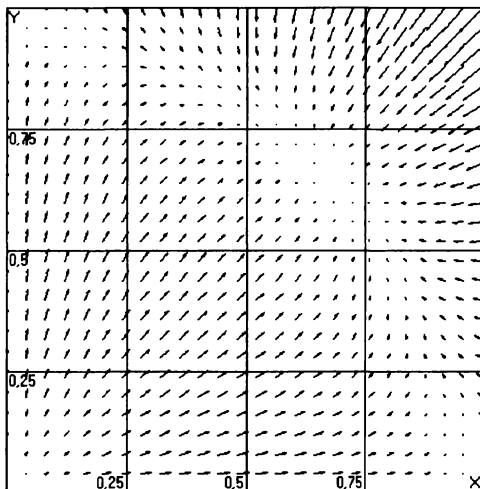


Рис. 3.12. Поле скоростей для системы (3.32)

$$(\mu_x = \mu_y = 0,5; r_x = r_y = 0,1; X^{(4)} = Y^{(4)} = 2/3; Y^* = 1)$$

Коэффициенты при $\delta\xi_1$ и $\delta\xi_2$ образуют матрицу линеаризации

$$A = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial X} & \frac{\partial f_1}{\partial Y} \\ \frac{\partial f_2}{\partial X} & \frac{\partial f_2}{\partial Y} \end{bmatrix}.$$

Собственные значения λ_k этой матрицы найдем из следующего соотношения:

$$\det \begin{bmatrix} a_{11} - \lambda & a_{21} \\ a_{12} & a_{22} - \lambda \end{bmatrix} = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} = 0. \quad (3.37)$$

Если все собственные значения λ_k матрицы A удовлетворяют неравенству $\text{Re } \lambda_k < 0$, $k = 1, 2$, то данное положение равновесия асимптотически устойчиво. Если же хотя бы для одного собственного значения вещественная часть больше нуля, то положение неустойчиво.

В зависимости от значений λ_k особые точки системы из двух обыкновенных дифференциальных уравнений могут быть четырех типов:

1. *Узел* – корни λ_1 и λ_2 действительные и одного знака (рис. 3.13, а, б).

2. *Седло* – корни λ_1 и λ_2 действительные, но разных знаков (рис. 3.13, в). Положение равновесия неустойчиво. Траектории I и II называются *сепаратрисами*.

3. *Фокус* – корни λ_1 и λ_2 комплексно-сопряженные (но не чисто мнимые) (рис. 3.13, з, д).

4. *Центр* – корни λ_1 и λ_2 чисто мнимые (рис. 3.13, е).

Запишем (3.37) в виде $\lambda^2 - B\lambda + C = 0$, где $B = a_{11} + a_{22}$, $C = a_{11}a_{22} - a_{12}a_{21}$. Соответственно дискриминант D уравнения имеет вид

$$D = B^2 - 4C = (a_{11} - a_{22})^2 + 4a_{12}a_{21}.$$

Для собственных значений получим

$$\lambda_{1,2} = \frac{1}{2}(a_{11} + a_{22} \pm \sqrt{D}). \quad (3.38)$$

Несложный анализ решения (3.38) позволяет сформулировать следующие условия:

1. Если дискриминант равен нулю или один из коэффициентов a_{12} или a_{21} равен нулю, то приходим к состоянию устойчивого или неустойчивого узла.

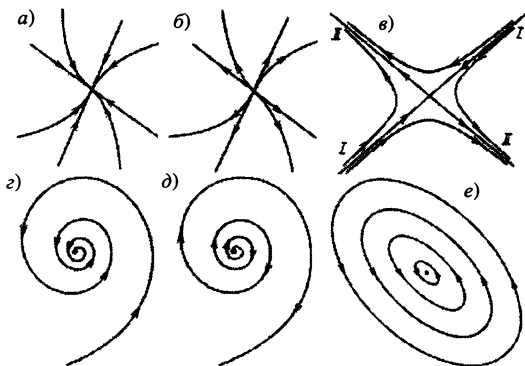


Рис. 3.13. Типы особых точек системы двух дифференциальных уравнений: а – устойчивый узел; б – неустойчивый узел; в – седловая точка; з – устойчивый фокус; д – неустойчивый фокус; е – центр [64]

2. Если коэффициенты a_{12} и a_{21} одного знака и не равны нулю, то корни уравнения (3.37) действительные, и возможно состояние как узловой, так и седловой точки.

3. Если коэффициенты a_{12} и a_{21} имеют разные знаки, не равны нулю и выполняется условие $(a_{11} - a_{22})^2 < |4a_{12}a_{21}|$, то при $a_{11} = -a_{22}$ получаем чисто мнимые корни уравнения (3.37) и, соответственно, условие центральной точки. Иначе неминуемо приходим к состоянию устойчивого или неустойчивого фокуса.

Выполним анализ положений равновесия для системы уравнений (3.32). Для первой точки равновесия ($X^{(1)} = Y^{(1)} = 0$) имеем

$$A = \begin{bmatrix} r_x & 0 \\ 0 & r_y \end{bmatrix}, \text{ т.е. } \lambda_1 = r_x, \lambda_2 = r_y.$$

Так как r_x и r_y — положительные вещественные числа, то первая точка равновесия является неустойчивым узлом. Для второй точки равновесия ($X^{(2)} = 0, Y^{(2)} = Y^*$) имеем

$$A = \begin{bmatrix} r_x(1 - \mu_x Y^*) & 0 \\ -r_y \mu_y Y^* & -r_y Y^* \end{bmatrix}, \text{ т.е. } \lambda_1 = r_x(1 - \mu_x Y^*), \lambda_2 = -r_y Y^*.$$

При $\mu_x Y^* < 1$ значение λ_1 положительно, т.е. это положение равновесия соответствует седловой точке. При $\mu_x Y^* > 1$ значение λ_1 отрицательно, т.е. имеем устойчивый узел, следовательно, при выполнении ограничения (3.35б) популяция X вымирает.

Для третьей точки равновесия ($X^{(3)} = 1, Y^{(3)} = 0$) имеем

$$A = \begin{bmatrix} -r_x & -r_x \mu_x \\ 0 & r_y(1 - \mu_y) \end{bmatrix}, \text{ т.е. } \lambda_1 = -r_x, \lambda_2 = r_y(1 - \mu_y).$$

При $\mu_y < 1$ значение λ_2 положительно, т.е. данное положение равновесия также является седловой точкой. При $\mu_y > 1$ значение λ_2 отрицательно и опять получаем устойчивый узел. Таким образом, при выполнении ограничения (3.35б) популяция Y вымирает.

Для четвертой точки равновесия

$$\left(X^{(4)} = \frac{1 - \mu_x Y^*}{1 - \mu_x \mu_y}, Y^{(4)} = \frac{Y^* - \mu_y}{1 - \mu_x \mu_y} \right)$$

получим
$$A = \begin{bmatrix} -r_x X^{(4)} & -r_x \mu_x X^{(4)} \\ -r_y \mu_y Y^{(4)} & -r_y Y^{(4)} \end{bmatrix}.$$

Если коэффициенты μ_x и μ_y положительны, то имеем коэффициенты a_{12} и a_{21} одного знака и не равные нулю, т.е. собственные значения λ_1 и λ_2 действительные, и возможно состояние как узловой, так и седловой точки. Так как $X^{(4)}$ и $Y^{(4)}$ положительные, то сумма элементов a_{11} и a_{22} матрицы линеаризации имеет отрицательное значение. Следовательно, чтобы получить отрицательные собственные значения λ_1 и λ_2 , необходимо выполнить условие $(a_{11} + a_{22})^2 > D$ или $a_{11}a_{22} > a_{12}a_{21}$.

Подставляя значения коэффициентов матрицы линеаризации, находим, что

$$\mu_x \mu_y < 1. \quad (3.39)$$

Полученное ограничение справедливо при выполнении ограничения (3.35а).

Таким образом, проведенный анализ позволяет заключить, что совместное существование популяций возможно при выполнении ограничения (3.35а). На рис. 3.14 показана допустимая область значений для устойчивого совместного существования популяций. Область представляет собой прямоугольник, один из узлов которого находится на кривой $\mu_x \mu_y = 1$.

Для построения области необходимо:

- 1) выбрать значение Y^* и отложить его на оси ординат;
- 2) провести через полученную точку горизонтальную линию до пересечения с кривой $\mu_x \mu_y = 1$;
- 3) из точки пересечения провести вертикальную линию до пересечения с осью абсцисс.

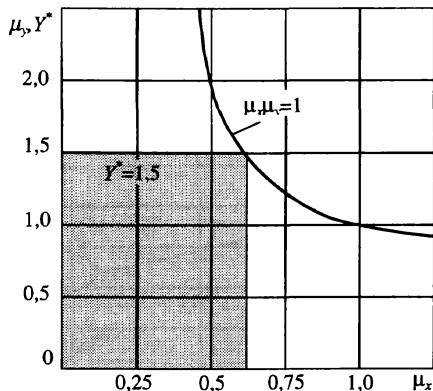


Рис. 3.14. Область устойчивого положения равновесия при $Y^* = 1,5$ ($r_x = r_y = 1$)

Численное исследование модели конкуренции популяций

Рассмотрим особенности численного исследования модели конкуренции популяций. Применим метод Эйлера для решения системы (3.32):

$$\left. \begin{aligned} X_{n+1} &= X_n + r_x \Delta t (1 - X_n - \mu_x Y_n) X_n, \\ Y_{n+1} &= Y_n + r_y \Delta t (Y^* - Y_n - \mu_y X_n) Y_n, \end{aligned} \right\} \quad (3.40)$$

где Δt — величина шага по времени; X_n и Y_n — относительная численность популяций на начало шага по времени; X_{n+1} и Y_{n+1} — численность популяций на конец шага по времени. Задаваясь начальной численностью X_0 и Y_0 и величиной шага по времени Δt , получаем таблицу значений численности в различные моменты времени.

На рис. 3.15 приведены бифуркационные диаграммы для соотношений (3.40) при фиксированных значениях исходных данных. Как можно видеть, сходимость процесса существенно зависит от шага интегрирования. При значении шага более 1,94 процесс перестает сходиться и переходит к колебаниям, а при шаге более 2,5 становится хаотическим.

На рис. 3.16 приведены диаграммы сходимости (первые 50 шагов) в фазовом пространстве и в зависимости от времени при различных величинах шага интегрирования. Анализ приведенных ди-

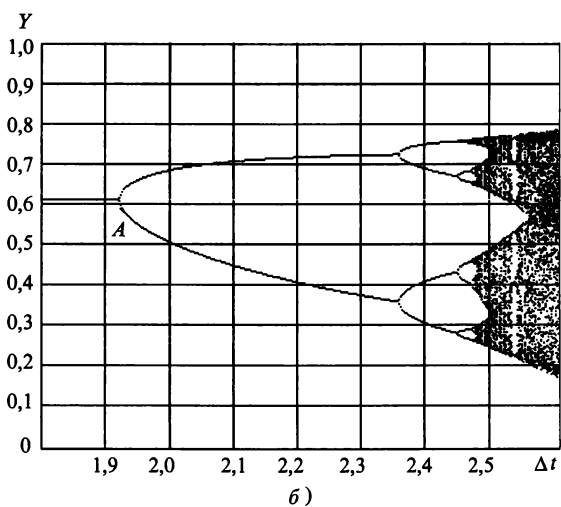
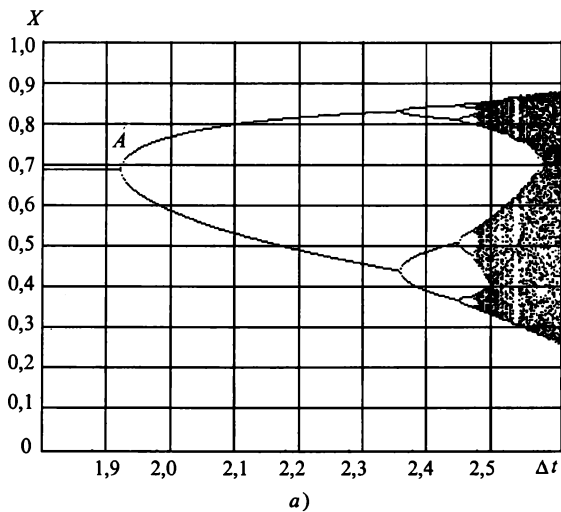


Рис. 3.15. Бифуркационные диаграммы для соотношений (3.40)
 $(X_0 = 0,2; Y_0 = 0,5; r_x = r_y = 1; \mu_x = 0,5; \mu_y = 0,7; Y^* = 1,1)$

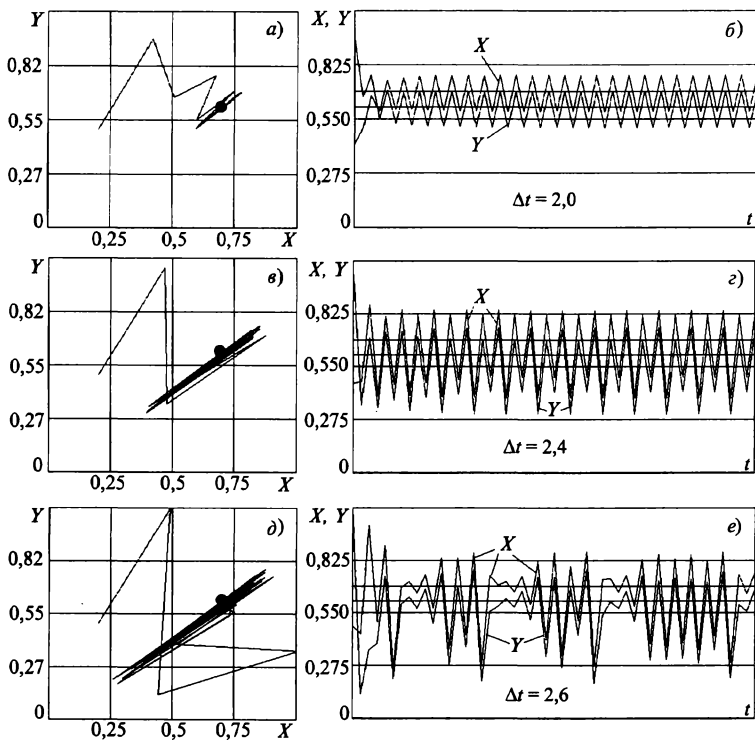


Рис 3.16. Диаграммы сходимости для различных величин шага интегрирования: а, в, д – фазовые диаграммы; б, г, е – временные диаграммы; а, б – $\Delta t = 2,0$; в, г – $\Delta t = 2,4$; д, е – $\Delta t = 2,6$

аграмм позволяет заключить, что данный процесс интегрирования быстро выходит в окрестность точки равновесия и далее совершаются колебательные движения вблизи этой точки. Следует отметить синхронное изменение численности популяций.

С увеличением шага интегрирования амплитуда и сложность колебаний нарастают. При хаотическом режиме возможны ситуации, когда численность одной из популяций становится меньше нуля.

На рис. 3.17 приведены кривые изменения величины шага интегрирования Δt_A , при котором процесс сходимости переходит от монотонного к колебательному режиму (точка А на рис. 3.15). При-

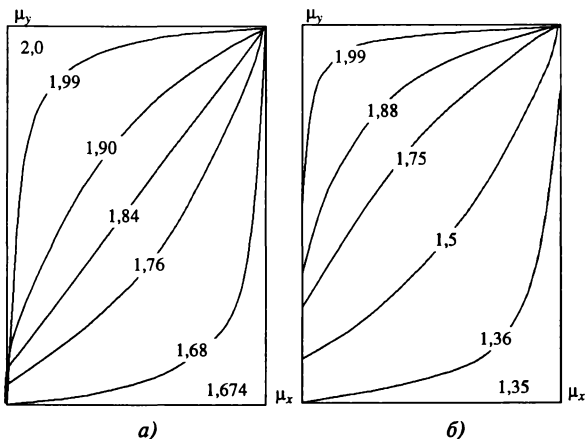


Рис. 3.17. Величина шага интегрирования в зависимости от μ_x и μ_y :
 а - $Y^* = 1,2$; б - $Y^* = 1,5$ ($X_0 = 0,2$; $Y_0 = 0,5$; $r_x = r_y = 1$)

веденные диаграммы построены в зависимости от значений μ_x и μ_y , которые удовлетворяли ограничению (3.35а). Построение выполнено по результатам вычислительного эксперимента, в котором значение Δt_A искалось для 2500 точек допустимой области (разбиение 50×50 точек).

Следует отметить, что величина Δt_A для допустимой области не превышает 2. При этом она уменьшается с ростом μ_x и уменьшением μ_y . С увеличением Y^* наименьшее значение шага монотонной сходимости уменьшается ($\Delta t_A = 1,674$ при $Y^* = 1,2$ и $\Delta t_A = 1,35$ при $Y^* = 1,5$). Оценку для Δt_A можно получить из анализа устойчивости соотношений (3.40).

Положим:

$$\begin{aligned} X_n &= X_p + \Delta_{x(n)}, & Y_n &= Y_p + \Delta_{y(n)}, \\ X_{n+1} &= X_p + \Delta_{x(n+1)}, & Y_{n+1} &= Y_p + \Delta_{y(n+1)}, \end{aligned} \quad (3.41)$$

где X_p, Y_p – точка равновесия; Δ_x, Δ_y – малые отклонения от равновесия.

Подставляя в (3.40) и преобразуя, получим:

$$\begin{aligned} \Delta_{x(n+1)} &= \Delta_{x(n)} - r_x \Delta t X_p (\Delta_{x(n)} + \mu_x \Delta_{y(n)}), \\ \Delta_{y(n+1)} &= \Delta_{y(n)} - r_y \Delta t Y_p (\Delta_{y(n)} + \mu_y \Delta_{x(n)}) \end{aligned}$$

или, поделив на $\Delta_{x(n)}$ и $\Delta_{y(n)}$,

$$\begin{aligned}\Delta_{x(n+1)} / \Delta_{x(n)} &= 1 - r_x \Delta t X_p (1 + \mu_x \Delta_{y(n)} / \Delta_{x(n)}), \\ \Delta_{y(n+1)} / \Delta_{y(n)} &= 1 - r_y \Delta t Y_p (1 + \mu_y \Delta_{x(n)} / \Delta_{y(n)}).\end{aligned}\quad (3.42)$$

Итерационная процедура сходится, если выполняются ограничения

$$\left| \Delta_{x(n+1)} / \Delta_{x(n)} \right| < 1 \quad \text{и} \quad \left| \Delta_{y(n+1)} / \Delta_{y(n)} \right| < 1 \quad (3.43)$$

или с учетом (3.42):

$$\begin{aligned}0 < r_x \Delta t X_p (1 + \mu_x \Delta_{y(n)} / \Delta_{x(n)}) < 2, \\ 0 < r_y \Delta t Y_p (1 + \mu_y \Delta_{x(n)} / \Delta_{y(n)}) < 2.\end{aligned}\quad (3.44)$$

Считая $\Delta_{x(n)}$ и $\Delta_{y(n)}$ величинами одного порядка и выбирая менее благоприятный случай, когда они одного знака, из (3.44) определяем ограничения на Δt_A :

$$\begin{aligned}\Delta t_A < 2 / r_x X_p (1 + \mu_x), \\ \Delta t_A < 2 / r_y Y_p (1 + \mu_y).\end{aligned}\quad (3.45)$$

Учитывая, что устойчивость процесса нарушается, если нарушается хотя бы одно из ограничений (3.45), окончательно имеем

$$\Delta t_A < \min \left(\frac{2}{r_x X_p (1 + \mu_x)}, \frac{2}{r_y Y_p (1 + \mu_y)} \right). \quad (3.46)$$

Для рассмотренной выше четвертой точки равновесия максимальное и минимальное значения оценки величины шага интегрирования Δt_A можно определить, когда μ_x или μ_y равны нулю. Так, при $\mu_x = 0$ и $r_x = r_y = 1$ имеем $X_p = 1$; $Y_p = Y^* - \mu_y$.

При приближении значения μ_y к Y^* величина $Y_p \rightarrow 0$. Из (3.46) в этом случае получаем ограничение на шаг интегрирования $\Delta t_A < 2$. При $\mu_y = 0$ получаем $X_p = 1 - \mu_x Y^*$; $Y_p = Y^*$.

При приближении значения μ_x к $1/Y^*$ величина $X_p \rightarrow 0$ и ограничение на Δt_A принимает вид $\Delta t_A < 2/Y^*$.

При $Y^* = 1,2$ имеем $\Delta t_A < 1,67$, а для $Y^* = 1,5$ — соответственно $\Delta t_A < 1,33$, что практически совпадает с предельными значениями Δt_A , приведенными на рис. 3.17 и полученными в результате вычислительного эксперимента.

3.5. ГАРМОНИЧЕСКИЙ ОСЦИЛЛЯТОР

Колебательные движения механических систем широко распространены в технике: качания маятников, движения поршней двигателей внутреннего сгорания, колебания струн, стержней и пластин, вибрации двигателей, фундаментов и множество других подобных процессов. Много внимания уделяется механическим и электрическим колебаниям в стандартном курсе физики средних учебных заведений.

Содержательная постановка задачи

Тело, лежащее на гладкой горизонтальной плоскости, прикреплено к неподвижной стене пружиной. Исследовать колебательные движения тела. Масса тела и жесткость пружины известны.

Концептуальная постановка задачи

Примем следующие предположения:

- объектом исследования является поступательно движущееся тело массой m , принимаемое за материальную точку;
- движение тела подчиняется второму закону Ньютона;
- тело находится под действием трех сил (рис. 3.18): силы тяжести mg , реакции поверхности N и силы упругости F_e пружины. Так как поверхность гладкая, то силой трения пренебрегаем;

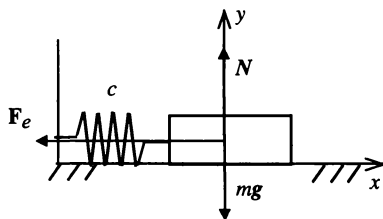


Рис. 3.18. Схема задачи

- тело совершает прямолинейные колебательные движения, так как сила тяжести mg уравнивается силой реакции поверхности N ;
- в равновешенном состоянии центр масс тела находится в положении с координатами (x_p, y_p) ;
- при малом растяжении пружины величину возникающей в ней силы упругости можно представить линейной зависимостью (закон Гука) $F_e = c\Delta x$, где $\Delta x = x - x_p$ — растяжение пружины (отклонение тела от положения равновесия x_p), c — жесткость пружины. Направлена сила в сторону положения равновесия.

Принимая, что в некоторый момент пружину растянули на величину x_0 и сообщили телу скорость v_0 , требуется определить координату и скорость тела как функции времени.

Математическая постановка задачи

С математической точки зрения имеем задачу Коши

$$m \frac{dv}{dt} = F_e = -c(x - x_p); \quad \frac{dx}{dt} = v \quad (3.47)$$

при следующих начальных условиях:

$$x(0) = x_0, \quad v(0) = v_0. \quad (3.48)$$

Требуется найти решение данной задачи.

Решение задачи

Введем обозначение для производных по времени:

$$\ddot{x} = \frac{d}{dt} \left(\frac{dx}{dt} \right) = \frac{dv}{dt} = \dot{v}.$$

Тогда, принимая $x_p = 0$, систему (3.47) можно заменить обыкновенным дифференциальным уравнением

$$\ddot{x} + k^2 x = 0, \quad (3.49)$$

где $k^2 = c/m$. Уравнение (3.49) является линейным дифференциальным уравнением второго порядка без правой части и называется

дифференциальным уравнением свободных колебаний материальной точки.

Будем искать решение данного уравнения в виде

$$x = \exp(r t). \quad (3.50)$$

Подставляя (3.50) в (3.49), получим

$$r^2 + k^2 = 0. \quad (3.51)$$

Уравнение (3.51) называется *характеристическим*, имеющим в данном случае два мнимых корня:

$$r_{1,2} = \pm ki. \quad (3.52)$$

Общее решение уравнения (3.49) в этом случае можно привести к виду

$$x = C_1 \cos(kt) + C_2 \sin(kt). \quad (3.53)$$

Подставляя начальные условия (3.48), получим $x_0 = C_1 \cos(0) + C_2 \sin(0)$, $v_0 = -kC_1 \sin(0) + kC_2 \cos(0)$ или $C_1 = x_0$, $C_2 = v_0/k$. Окончательно имеем

$$x = x_0 \cos(kt) + (v_0/k) \sin(kt). \quad (3.54)$$

Вводя амплитуду a и начальную фазу α , запишем

$$x = a \sin(kt + \alpha), \quad (3.55)$$

где $a = \sqrt{x_0^2 + v_0^2/k^2}$, $\sin \alpha = x_0/a$, $\cos \alpha = v_0/ka$.

Частоту k называют *собственной частотой* колеблющейся системы, или *частотой ее свободных колебаний*. Таким образом, модель (3.47), (3.48) описывает гармоническое колебательное движение с частотой k и периодом

$$T = \frac{2\pi}{k} = 2\pi \sqrt{\frac{m}{c}}, \quad (3.56)$$

зависящими не от начальных условий, а только от жесткости пружины c и массы тела m .

Свойство независимости частоты или периода колебаний от начальных условий (свойство *изохронности* колебаний) связано с ли-

нейностью восстанавливающей силы (оно было открыто Галилеем). В случае нелинейной восстанавливающей силы свойство изохронности не наблюдается.

Качественный анализ задачи

Для случая $x_p = 0$ систему (3.47) можно переписать в следующем виде:

$$\dot{v} = -k^2 x, \quad \dot{x} = v. \quad (3.57)$$

Имеем две фазовые переменные x и v , изменяющиеся в интервалах $x \in [-a, a]$, $v \in [-ka, ka]$. Приравняв правые части в (3.57) нулю, находим точку равновесия

$$x^* = v^* = 0. \quad (3.58)$$

Матрица линеаризации имеет вид

$$A = \begin{bmatrix} \frac{\partial(-k^2 x)}{\partial v} & \frac{\partial(-k^2 x)}{\partial x} \\ \frac{\partial v}{\partial v} & \frac{\partial v}{\partial x} \end{bmatrix} = \begin{bmatrix} 0 & -k^2 \\ 1 & 0 \end{bmatrix}.$$

Собственные значения находим из уравнения

$$\det \begin{bmatrix} -\lambda & -k^2 \\ 1 & -\lambda \end{bmatrix} = \lambda^2 + k^2 = 0. \quad (3.59)$$

Корни этого уравнения чисто мнимые. Следовательно, точка равновесия (x^*, v^*) является центром с интегральной кривой в виде эллипса. Найденное выше решение (3.55) позволяет построить интегральную кривую. Уравнения $x = a \sin(kt + \alpha)$, $v = ka \cos(kt + \alpha)$ можно рассматривать как параметрические уравнения эллипса с полуосями a и ka .

Численное исследование модели

Решение рассматриваемой задачи можно получить численно, заменяя производные их приближенными разностными аналогами и переходя к системе разностных уравнений (схема Эйлера):

$$v_{n+1} = v_n - k^2 x_n \Delta t, \quad x_{n+1} = x_n + v_n \Delta t. \quad (3.60)$$

На рис. 3.19 приведены графики изменения скорости v и смещения x на интервале времени от $100T$ до $101T$ при различных значениях шага интегрирования Δt . Аналитическое решение практически совпадает с графиком 1 на рис. 3.19. Как можно видеть из приведенных результатов, увеличение шага интегрирования приводит к возрастанию амплитуды колебаний. Причем даже при шаге в 10^{-5} периода колебаний на интервале времени интегрирования в 100 периодов погрешность вычислений становится заметной. Таким образом, амплитуда колебаний при интегрировании по схеме Эйлера зависит от величины шага интегрирования Δt и времени процесса T_{\max} . Следует отметить, что величина шага интегрирования в рассматриваемом случае не приводит к заметному изменению фазы и частоты колебаний.

В качестве характеристики точности процедуры численного интегрирования можно использовать величину

$$\Delta = \frac{A(\Delta t, T_{\max}) - a}{a} 100\%,$$

где a – амплитуда, определяемая из аналитического решения; $A(\Delta t, T_{\max})$ – максимальная амплитуда, полученная численной процедурой с шагом Δt к моменту времени T_{\max} процесса. На рис. 3.20 приведены диаграммы изменения точности интегрирования Δ в зависимости от величины шага Δt и времени T_{\max} интегрирования.

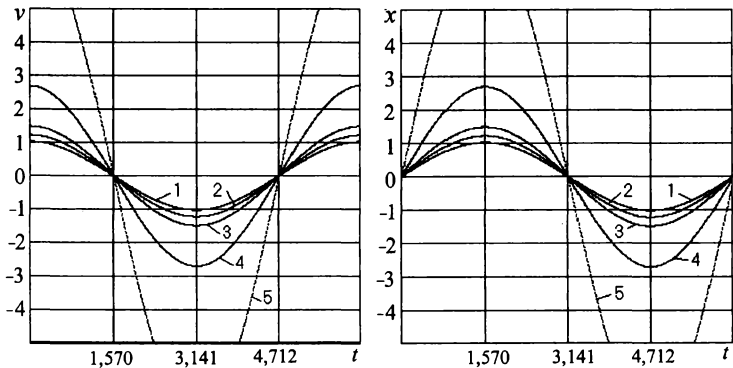


Рис. 3.19. Влияние величины шага по времени Δt на точность решения
 Δt : 1 – $10^{-5}T$; 2 – $10^{-4}T$; 3 – $5 \cdot 10^{-3}T$; 4 – $2 \cdot 10^{-3}T$; 5 – $10^{-3}T$;
 $k = 1$, $\alpha = 0^\circ$, $T = 2\pi$, $a = 1$

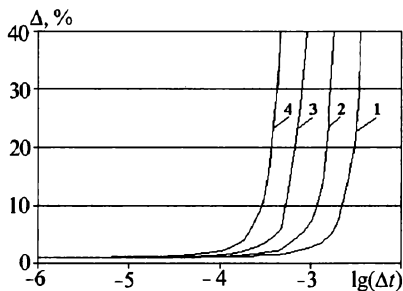


Рис. 3.20. Зависимость точности интегрирования Δ по схеме Эйлера от величины шага интегрирования Δt и времени интегрирования T_{\max} : 1 – до $50T$, 2 – до $100T$, 3 – до $200T$, 4 – до $400T$; $k = 1$, $a = 0^\circ$, $T = 2\pi$, $\alpha = 1$

Можно отметить, что с увеличением времени интегрирования шаг интегрирования должен уменьшаться. Например, если нас интересует поведение материальной точки на интервале от 0 до $400T$, то шаг интегрирования не должен быть больше $10^{-4}T$.

Точность интегрирования существенно зависит от метода интегрирования. К соотношениям (3.47) можно применить модифицированную схему Эйлера. В этом случае на шаге интегрирования движение точки считается не равномерным, а равноускоренным, т.е. при вычислении нового положения точки используется средняя на шаге интегрирования скорость движения. Тогда соотношения (3.60) имеют такой вид

$$\begin{aligned} v_{n+1} &= v_n - k^2 x_n \Delta t, \\ x_{n+1} &= x_n + 0,5(v_n + v_{n+1})\Delta t. \end{aligned} \quad (3.61)$$

На рис. 3.21 приведены диаграммы изменения точности интегрирования Δ в зависимости от величины шага Δt и времени T_{\max} интегрирования для модифицированной схемы Эйлера.

Сравнивая эти кривые с диаграммами на рис. 3.20, можно отметить увеличение точности интегрирования с использованием модифицированной схемы примерно в 2 раза. Так, например, кривая 4 на рис. 3.21 практически совпадает с кривой 3 на рис. 3.20, т.е. при использовании модифицированной схемы Эйлера можно применять по сравнению с обычной схемой Эйлера или в 2 раза больший шаг интегрирования или в 2 раза длиннее интервал интегрирования.

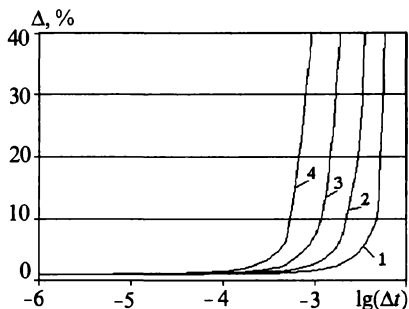


Рис. 3.21. Зависимость точности интегрирования Δ для модифицированной схемы Эйлера от величины шага интегрирования Δt и времени интегрирования T_{max} :
 1 – до $50T$, 2 – до $100T$, 3 – до $200T$, 4 – до $400T$;
 $k = 1$, $a = 0^\circ$, $T = 2\pi$, $a = 1$

Усложним исходную задачу введением дополнительной силы сопротивления движению. Такая ситуация возможна, если тело погружено в жидкость (рис. 3.22, а) или если в рассматриваемую схему добавлен масляный демпфер, гасящий колебания (рис. 3.22, б). В последнем случае вязкое масло при перетекании внутри цилиндра создает дополнительное сопротивление движению.

Наличие сил вязкого трения приводит к необходимости добавления новой гипотезы при концептуальной постановке задачи. Подобную гипотезу можно сформулировать следующим образом: сила вязкого трения прямо пропорциональна скорости тела и направлена против направления его движения: $F_c = -\mu v$ (μ – коэффициент

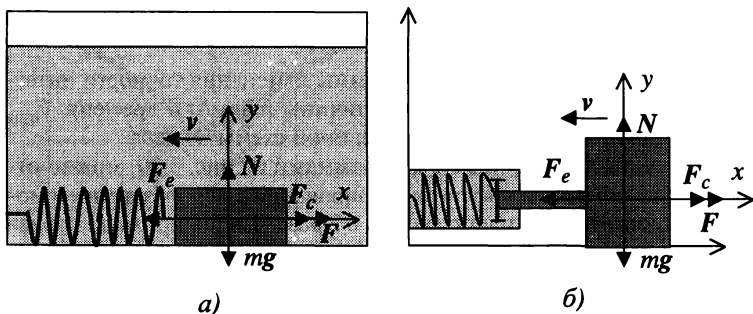


Рис. 3.22. Схемы конструкций с вязкими средами

сопротивления, величина постоянная). С учетом этой гипотезы соотношения (3.47) в математической постановке следует заменить уравнениями следующего вида:

$$\frac{dv}{dt} = -k^2(x - x_p) - 2n v; \quad \frac{dx}{dt} = v, \quad (3.62)$$

где $k^2 = c/m$; $2n = \mu/m$.

Качественный анализ задачи

Для случая $x_p = 0$ систему (3.62) можно переписать в виде

$$\dot{v} = -k^2x - 2nv, \quad \dot{x} = v. \quad (3.63)$$

Приравняв правые части в (3.63) нулю, находим точку равновесия

$$x^* = v^* = 0. \quad (3.64)$$

Матрица линеаризации имеет вид

$$A = \begin{bmatrix} -2n & -k^2 \\ 1 & 0 \end{bmatrix}.$$

Собственные значения находим из уравнения

$$\det \begin{bmatrix} -2n - \lambda & -k^2 \\ 1 & -\lambda \end{bmatrix} = \lambda^2 + 2n\lambda + k^2 = 0. \quad (3.65)$$

Корни уравнения комплексно-сопряженные, но не чисто мнимые. Следовательно, точка равновесия (x^*, v^*) является фокусом.

Решение задачи

Систему обыкновенных дифференциальных уравнений (3.62) при условии $x_p = 0$ можно свести к уравнению второго порядка

$$\ddot{x} + 2n\dot{x} + k^2x = 0. \quad (3.66)$$

Характеристическое уравнение имеет вид, аналогичный (3.65):

$$r^2 + 2nr + k^2 = 0 \quad (3.67)$$

с корнями

$$r_{1,2} = -n \pm \sqrt{n^2 - k^2}. \quad (3.68)$$

В зависимости от соотношения n и k имеем три случая:

1) $n < k$ – *затухающее колебательное движение*. В этом случае имеем комплексно-сопряженные корни характеристического уравнения $r_{1,2} = -n \pm ik_n$, где $k_n = \sqrt{k^2 - n^2}$. Решение задачи можно привести к виду

$$x = ae^{-nt} \sin(k_n t + \alpha), \quad (3.69)$$

где

$$a = \sqrt{x_0^2 + \frac{(v_0 + nx_0)^2}{k_n^2}}, \quad \operatorname{ctg} \alpha = \frac{v_0 + nx_0}{x_0 k_n}. \quad (3.70)$$

На рис. 3.23 показаны изменения координаты и скорости тела, полученные для условий $m = 1$ кг, $c = 2500$ Н/м, $x_0 = 0$ м, $v_0 = 1$ м/с, $\mu = 10$ Н·с/м.

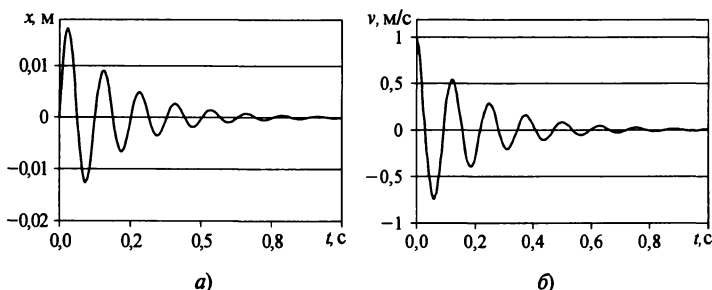


Рис. 3.23. Координата (а) и скорость (б) центра масс тела в вязкой среде

На рис. 3.24 представлена фазовая траектория для данного случая. В отличие от предыдущей задачи наличие сопротивления (диссипации механической энергии) приводит к тому, что фазовая траектория оказывается спиралью с уменьшающимся радиусом, стремящимся к нулю, т.е. точка равновесия является устойчивым фокусом.

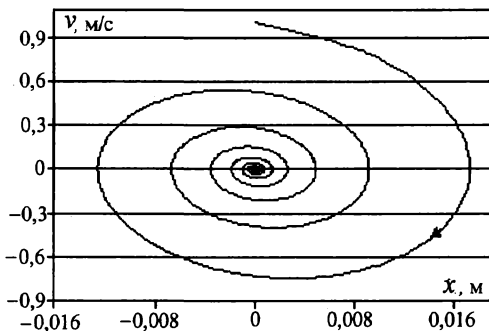


Рис. 3.24. Фазовая траектория движения точки в вязкой среде

Период колебаний

$$T = \frac{2\pi}{k_n} = \frac{T_0}{\sqrt{1 - (n/k)^2}}, \quad (3.71)$$

где $T_0 = 2\pi/k$ – период свободных колебаний при отсутствии сил сопротивления. Как следует из закона движения (3.69), амплитуда колебаний при наличии сил сопротивления с течением времени убывает. Натуральный логарифм отношения двух последовательных амплитуд называется *логарифмическим декрементом (затухания)*:

$$\delta = \ln(a_i/a_{i+1}) = \pi n/k_n. \quad (3.72)$$

2) $n > k$ – *апериодическое движение*. При достаточно большом сопротивлении, когда $k_a^2 = n^2 - k^2 > 0$, решение примет вид

$$x = e^{-nt} \left(x_0 \operatorname{ch} k_a t + \frac{v_0 + nx_0}{k_a} \operatorname{sh} k_a t \right). \quad (3.73)$$

Движение в этом случае не будет носить колебательный характер. На рис.3.25 показаны различные случаи апериодического движения в зависимости от начальной скорости v_0 точки. Случаи а) и б) на рис. 3.25 относят к апериодическому движению первого рода, а случай в) – второго рода.

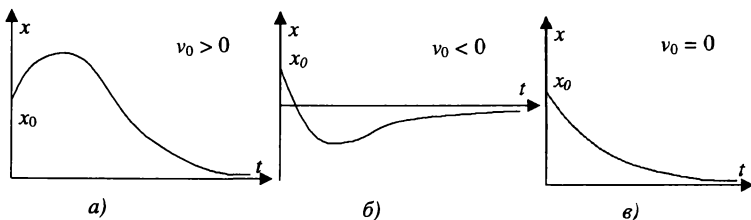


Рис. 3.25. Виды аperiodического движения при $x_0 > 0$

3) $n = k$ – *предельное аperiodическое движение*. Решение в этом случае примет вид

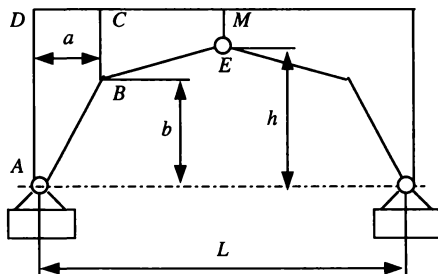
$$x = e^{-nt} [x_0 + (v_0 + nx_0)t]. \quad (3.74)$$

Как и в предыдущем случае, движение не будет колебательным. Вид траектории движения точки в зависимости от соотношения значений x_0 и v_0 аналогичен траекториям, представленным на рис. 3.25.



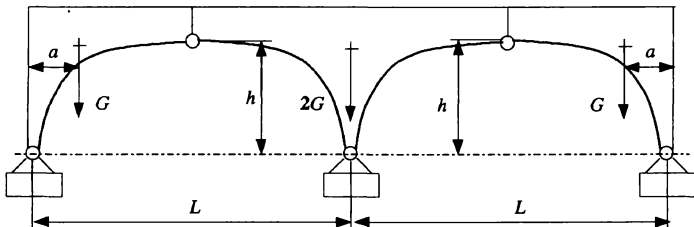
Задания для самостоятельного выполнения

1. Арка моста состоит из двух одинаковых половинок. К опорам каждая из половинок крепится с помощью цилиндрических шарниров, а между собой – одним шаровым шарниром. Длина пролета моста L , высота пролета h , ширина моста w . Каждая из половинок состоит из двух элементов $ABCD$ и $BEMC$, поперечное сечение которых является прямоугольной трапецией, одинаковой по всей ширине моста. Требуется найти такие размеры a и b , при которых горизонтальные реакции опор были бы минимальны. Удельная плотность материала арок моста γ .

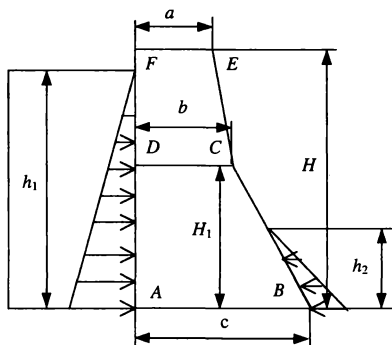


Для определения координат центра тяжести составной плоской фигуры воспользуйтесь соотношением $X_C = \frac{\sum x_i F_i}{\sum F_i}$, где x_i – координаты центра тяжести i -го элемента составной фигуры, F_i – площадь i -го элемента составной фигуры. Вес отдельного трапецевидного элемента определите как произведение удельного веса на ширину моста w и площадь F_i .

2. Проведите статистический анализ двухпролетного моста. Как и в рассмотренном выше примере арки крепятся к двум опорам с помощью цилиндрических шарниров. Между собой арки соединены одним шаровым шарниром. Требуется вычислить усилия в узлах крепления арок.



3. Поперечное сечение бетонной плотины, перегораживающей реку, состоит из двух прямоугольных трапеций $ABCD$ и $DCEF$. Высота плотины $H = 5$ м, глубина воды до плотины $h_1 = 4$ м, после плотины $h_2 = 1$ м. Давление воды возрастает прямо пропорционально ее глубине. Удельный вес бетона $\gamma = 22,5$ кН/м³. Определите размеры a , b , c и H_1 , при которых масса 1 м длины плотины была бы минимальной, а плотина под давлением воды не опрокидывалась. При оценке опрокидывания крепление плотины к берегам и дну реки не учитывайте.



4. Постройте экономическую модель спроса и предложения в предположении их линейной зависимости от цены. Проведите анализ изменения цен в зависимости от начальной цены при следующих исходных

данных: $a = 3$, $b = 2$, $c = 6$, $g = 8$. Определите, в каких пределах может изменяться начальная цена.

5. Постройте экономическую модель спроса и предложения в предположении, что справедливы зависимости

$$s_{n+1} = ap_n^m - b, \quad d_{n+1} = g \exp(-cp_{n+1}),$$

где $a, b, c, g > 0$; $0 < m < 1$. Проведите анализ изменения цен в зависимости от начальной цены при следующих исходных данных: $a = 3$, $b = 2$, $c = 1$, $g = 8$, $m = 0,5$. Определите, как влияет значение начальной цены на сходимость решения.

6. Пусть в некоторой местности обитают две популяция животных, причем животные одной популяции относятся к хищникам, а другой – к травоядным, служащим пищей для хищников. Для описания подобных систем «хищник–жертва» французский математик Вольтерра в 30-е годы XX века предложил следующую математическую модель:

$$\frac{dx}{dt} = ax - bxy, \quad \frac{dy}{dt} = cy - dy,$$

где x – численность популяции жертв, y – численность популяции хищников.

Предложите систему гипотез, на основании которых Вольтерра записал свою математическую модель. Какая модель (Мальтуса или Ферхюльста) была использована для описания изменения численности жертв? Проведите качественный анализ этой системы уравнений. Выполните анализ численной схемы Эйлера для системы «хищник–жертва».

7. Используя модель Ферхюльста для описания поведения жертв, предложите свой вариант математической модели «хищник–жертва». Проведите качественный анализ полученной системы уравнений. Выполните анализ численной схемы для данной системы.

8. При изучении развития эпидемии некоторого заболевания обычно выделяют три группы людей: x – группа людей, восприимчивых к данному заболеванию, но еще не заразившаяся им; y – группа уже больных или инфицированных людей, которые могут выступать разносчиками болезни; z – группа людей, невосприимчивых к этой болезни или получившие иммунитет после перенесенного заболевания. Один из вариантов математической модели развития эпидемии может быть записан в следующем виде:

$$\frac{dx}{dt} = -ax, \quad \frac{dy}{dt} = ax - by, \quad \frac{dz}{dt} = by.$$

Предложите систему гипотез для обоснования данной модели. Поясните смысл коэффициентов a и b . Проведите качественный анализ полученной системы уравнений. Выполните анализ численной схемы для этой системы. Предложите другие варианты моделей эпидемии с учетом:

а) изменения общей численности населения, связанные с рождениями и естественными смертями;

б) смертности от данного заболевания;

в) непостоянства доли заболевших людей.

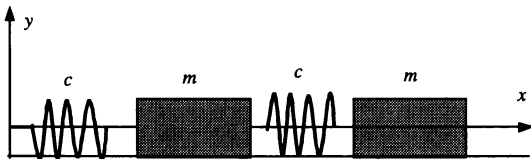
9. Одна группа медиков ищет спонсоров для исследований стоимостью в 1 млн долл. и рассчитывает за один год получить вакцину, которая позволит уменьшить коэффициент a (см. предыдущий пример) на 25%. Другая группа медиков предлагает за один год тоже за 1 млн долл. найти лекарство, которое увеличит на 25% коэффициент b . Если у спонсора есть только 1 млн долл., то какую группу медиков он должен поддержать?

10. Для модели свободных колебаний тела получите разностные соотношения с применением схемы Рунге-Кутты. Разработайте алгоритм решения системы полученных разностных уравнений и реализуйте его на персональном компьютере. Постройте диаграммы точности интегрирования Δ от величины шага интегрирования Δt и времени интегрирования T_{\max} . Сравните полученные результаты с приведенными выше результатами по другим схемам интегрирования.

11. Разработайте алгоритм численного решения задачи о движении маятника при наличии силы вязкого сопротивления и реализуйте его на персональном компьютере. Оцените величину подходящего шага интегрирования Δt в зависимости от времени T_{\max} для различных схем интегрирования.

12. Разработайте алгоритм решения задачи о вынужденных колебаниях маятника при наличии силы вязкого сопротивления и реализуйте его на персональном компьютере. Оцените величину подходящего шага интегрирования Δt . Усложните модель, принимая, что точка подвеса маятника совершает гармонические колебания по вертикали. Исследуйте, как влияет частота и амплитуда колебаний подвеса на поведение маятника.

13. Сформулируйте концептуальную и математическую постановку для модели, описывающей свободные колебания системы, включающей два тела массой m , соединенных пружинами жесткостью c . Разработайте алгоритм численного решения данной задачи. Оцените величину подходящего шага интегрирования Δt в зависимости от времени T_{\max} для различных схем интегрирования. Постройте траектории движения тел в фазовом пространстве.



14. Лодку массы m оттолкнули от берега пруда и, разогнав, отпустили при некоторой начальной скорости v_0 . Необходимо исследовать движение лодки в предположении, что сила сопротивления движению прямо пропорциональна скорости. Коэффициент сопротивления движению μ . Получите аналитическое и численное решения задачи. Оцените величину подходящего шага интегрирования Δt для различных схем интегрирования.

Глава 4

СТРУКТУРНЫЕ МОДЕЛИ

Как было отмечено в гл. 2, одним из важных этапов математического моделирования является переход от содержательной (технической) и концептуальной постановки задачи к математической, т.е. переход с технического языка (или — «языка природы») описания исследуемого объекта к языку математических формул и уравнений. Однако на практике такой переход бывает совсем не просто осуществить. В некоторых случаях это связано со сложностью самого исследуемого объекта, в других — с отсутствием необходимой информации об объекте или соответствующего математического аппарата. Поэтому часто возникает необходимость «разбить» поставленную задачу на несколько более простых подзадач, которые имеют известные решения или которые можно решить с помощью апробированных методов.

Для этой цели удобно использовать методы структурного моделирования, позволяющие еще на стадии постановки упростить решаемую задачу путем исследования внутренней структуры рассматриваемого объекта, изучения свойств отдельных элементов объекта и связей между ними. При этом следует отметить, что важное значение имеет способ разбиения исследуемого объекта на отдельные элементы и установление связей между ними, которые зависят от целей моделирования. Очевидно, что такое разбиение неоднозначно. Чем сложнее исходная задача, тем больше элементов необходимо вводить в рассмотрение и учитывать огромное количество связей между ними. Это обычно приводит к многоуровневым структурным схемам древовидной формы с различной иерархией элементов, которую необходимо учитывать при математическом моделировании. При этом структурные схемы, связи подсистем и их элементов несложно изобразить графически, что делает анализируемые объекты более понятными, «прозрачными».

Важно отметить, что при установлении различных связей между элементами могут возникать новые свойства построенной модели, которыми не обладает каждый из элементов в отдельности. Эти так называемые «системные» свойства модели, появляющиеся при объединении отдельных элементов в одно целое, играют важную роль в моделировании сложных систем, так как позволяют исследовать новые качества объекта с помощью вычислительного эксперимента. Поэтому важно знать общие правила построения структурных моделей, которые приводятся в данной главе и иллюстрируются на нескольких примерах, взятых из различных областей целенаправленной деятельности человека.

4.1. ЧТО ТАКОЕ СТРУКТУРНАЯ МОДЕЛЬ?

Системой является все, что мы хотим различать как систему.

Б. Гейне.

Очень часто для достижения практических целей возникает необходимость рассматривать исследуемый объект как совокупность отдельных элементов, связанных (взаимодействующих) между собой некоторым образом, в то же время взаимодействующих с окружающим миром как нечто целое. В этом случае исследуемый объект удобно представить в виде системы, а при его моделировании использовать методы *системного анализа*.

Напомним основные понятия системного анализа, которые будут использоваться в дальнейшем [83]. Одним из основополагающих понятий системного анализа является понятие искусственной системы, которую определим следующим образом.

Система есть совокупность взаимосвязанных элементов, выделенных из среды и взаимодействующих с окружающей средой как целое для достижения поставленной цели.

Следует отметить, что важным признаком для выделения системы из среды является возможность определения взаимодействия этой системы с окружением независимо от поведения ее отдельных элементов (именно это подразумевается под словами «взаимодействующая ... как целое»). Выделяет систему из среды исследователь, который отделяет ее элементы от среды в соответствии с поставленной целью. Под средой здесь понимается совокупность всех объектов, изменение свойств которых влияет на

систему, а также тех объектов, чьи свойства изменяются в результате поведения системы.

Из приведенных определений видно, насколько важна роль исследователя – он формулирует цели, выделяет систему и определяет среду. При этом сам может отнестись к среде и строить изолированные системы, включить себя в систему и строить ее с учетом своего влияния на ее функционирование (адаптивные системы), а также выделить себя и из системы, и из среды, рассматривая систему как открытую или развивающуюся. В принципе исследователя можно не рассматривать как элемент системы или среды, но, по мнению некоторых авторов [21], дополнительное введение исследователя помогает при построении систем и их классификации.

Для описания систем в системном анализе рассматриваются четыре основные модели. Если внутреннее строение системы неизвестно (или не интересует исследователя), то применяется модель «черного ящика». В этой модели системы отсутствуют (или не используются в явной форме) сведения о внутреннем содержании «ящика» (поэтому он и называется «черным»), а только задаются входные и выходные связи со средой. Обычно это сводится к заданию двух множеств входных и выходных параметров, но никаких соотношений между ними не задается. Примером модели «черного ящика» может служить экспериментальное исследование некоторого сложного объекта, когда экспериментатор, изменяя входные параметры объекта, получает на выходе различные его характеристики.

Очевидно, что исследование внутреннего устройства системы невозможно с помощью модели «черного ящика»; применение последней можно считать оправданным лишь на самых ранних этапах исследования нового объекта. Для этого необходимы более развитые модели. Одной из таких моделей является модель состава системы, описывающая, из каких элементов и подсистем состоит данная система. При этом *элементами* системы называются те ее части, которые полагаются неделимыми; части, состоящие более чем из одного элемента, называются *подсистемами*. Например, если в качестве системы рассмотреть автомобиль, то ее подсистемой можно считать систему управления, а элементами последней – руль, педали и т.д.

Сложность построения модели состава системы состоит в ее неоднозначности. Это можно объяснить следующим образом. Во-первых, понятие «элементарности» можно определять по-разному. Во-вторых, модель состава (как и любая другая модель) является целевой и для отличающихся целей один и тот же объект может по-

требовать различного разбиения на части. В-третьих, всякое разбиение целого на части является относительным. Например, тормозную подсистему автомобиля можно отнести как к ходовой части, так и к подсистеме управления.

В большинстве случаев модели состава системы оказывается недостаточно для ее описания. Мало знать состав системы, кроме этого необходимо установить связи между отдельными элементами, которые называются *отношениями*. Совокупность необходимых и достаточных для достижения цели отношений между элементами называется *моделью структуры системы*. Основной сложностью при описании структуры (списка отношений) является обоснование конечного числа связей, наиболее существенных по отношению к рассматриваемой цели. Например, при моделировании механической системы, движущейся в околоземном пространстве, обычно не учитываются силы взаимного притяжения отдельных материальных точек (элементов), но учитывается сила притяжения к Земле (отношения).

Следует отметить, что структура системы является абстрактной моделью, так как рассматривает только связи (отношения) между элементами, а не сами элементы (понятно, что на практике говорить об отношениях без элементов просто не имеет смысла). Однако в некоторых случаях модель структуры теоретически может быть исследована отдельно, если, например, отношения заданы в виде математических формул или уравнений [76].

Теперь, имея три формальные модели системы – «черного ящика», состава и структуры – и объединив их, можно получить еще одну модель, которую называют *структурной схемой* системы, или моделью «белого ящика» [83]. Данная модель включает все элементы системы, все связи между элементами внутри системы и связи системы (или ее отдельных элементов) с окружающей средой (входы и выходы системы), как изображено на рис. 4.1. Заметим, что под «всеми» элементами и связями понимаются, конечно, значимые с точки зрения цели и задач разрабатываемой модели.

Следует отметить, что структурная схема системы является формальной моделью, отделенной от содержательного наполнения. Это позволяет рассматривать ее как особый математический объект и исследовать его свойства. Такой объект называется *графом*. Он состоит из обозначений элементов произвольной природы, называемых вершинами, и обозначений связей между ними, называемых *ребрами*.

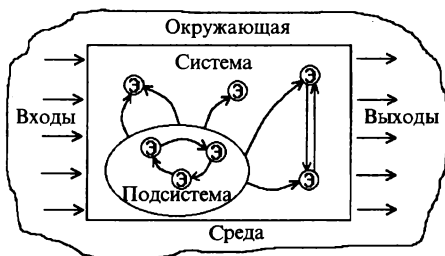


Рис. 4.1. Структурная схема системы

На рис. 4.2 приведен пример графа, у которого вершины обозначены в виде кружочков, а ребра – в виде линий. Стрелки указывают на несимметричность некоторых связей. Такой граф называется *ориентированным*. Каждая пара вершин может быть соединена с любым числом ребер. Чтобы ввести другие различия между ребрами, кроме несимметричности, им приписывают различные веса. Такие графы называются *взвешенными*. В качестве весов могут выступать различные характеристики графа, например длина ребра, число каналов электросети, тип покрытия в сети автомобильных дорог и т.д.

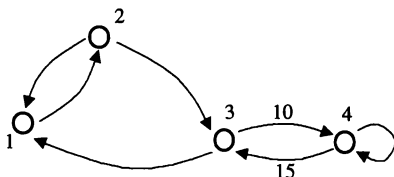


Рис. 4.2. Пример графа

В настоящее время для графов разработана целая теория, имеющая многочисленные приложения [34, 61, 104]. Графы могут изображать любые структуры. Некоторые графы получили специальные названия: линейные, древовидные (иерархические), сетевые, матричные и т.д. Пример сетевого графа приведен на рис. 4.3.

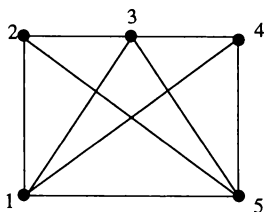


Рис 4.3. Сетевой граф

Как отмечено выше, граф (структурная схема) является формальной моделью, которую необходимо наполнить конкретным содержанием. Только пос-

ле этого структурная схема становится *структурной моделью* исследуемого объекта. Например, если в качестве вершин графа, изображенного на рис. 4.3, считать цилиндрические шарниры, а ребер – прямолинейные стержни, то получим структурную модель стержневой конструкции, широко применяемую при моделировании в строительной механике [3]. Если же в качестве вершин рассматривать узлы связи, а за ребра принять линии связи, то получим структурную модель сети электросвязи [65].

Таким образом, можно дать следующее определение структурной модели.

Структурная модель системы – это совокупность конкретных элементов данной системы, необходимых и достаточных отношений между этими элементами и связей между системой и окружающей средой.

Рассмотрим пример построения структурной модели, поясняющий это определение. Пусть требуется построить структурную модель абсолютно твердого тела, совершающего поступательное движение под действием приложенной силы (рис. 4.4). Напомним, что абсолютно твердое тело при движении не изменяет форму и размеры. Поэтому такое тело можно представить как совокупность материальных точек (элементов), соединенных прямолинейными невесомыми недеформируемыми стержнями.

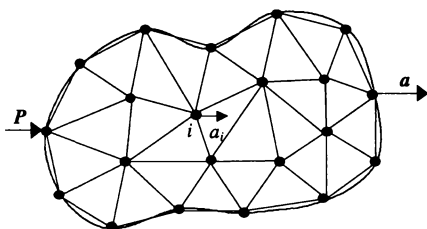


Рис. 4.4. Структурная модель абсолютно твердого тела при поступательном движении

Сила P выступает в качестве воздействия внешней среды на тело, а откликом на это воздействие (выходным параметром) служит ускорение тела a (или любой его точки вследствие их равенства при поступательном движении). Согласно второму закону Ньютона,

$$a = (m_T)^{-1} P, \quad (4.1)$$

где m_T – масса тела.

Таким образом, для того чтобы данная структурная модель правильно описывала поступательное движение тела, необходимо и достаточно, чтобы выполнялось условие

$$\sum_{i=1}^N m_i = m_T, \quad (4.2)$$

где m_i — масса i -го элемента, N — число элементов. Из последнего соотношения следует, что число элементов и их распределение внутри тела не имеют значения. Другое дело, когда необходимо описать вращательное движение твердого тела вокруг некоторой заданной оси под действием приложенного момента сил. В этом случае распределение элементов внутри объема тела должно быть таким, чтобы выполнялось условие равенства моментов инерции реального тела и его структурной модели относительно заданной оси Z :

$$I_Z = \sum_{i=1}^N m_i r_i^2, \quad (4.3)$$

где I_Z — момент инерции тела относительно оси Z , r_i — расстояние i -й точки до оси Z . Из этого равенства вытекают необходимые и достаточные условия на отношения между элементами данной структурной модели (длины и ориентации стержней, массы материальных точек), что хорошо видно из рис. 4.5.

Еще больше усложнится структурная модель в случае описания движения деформируемого, например упругого, тела. Тогда вместо

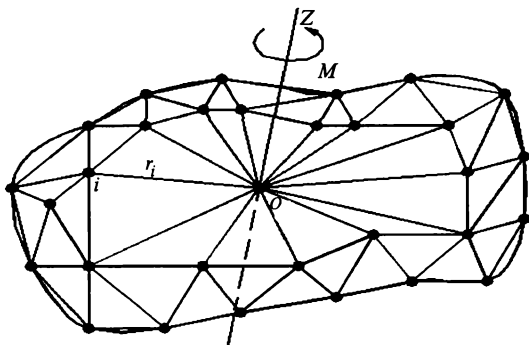


Рис. 4.5. Структурная модель вращающегося тела

недеформируемых стержней в качестве связей между элементами (материальными точками) могут выступать пружинки с различными упругими свойствами (рис. 4.6).

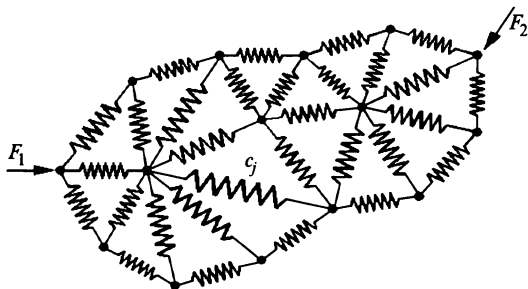


Рис. 4.6. Структурная модель упругого тела

В этой структурной модели для получения необходимых и достаточных отношений между элементами кроме распределения масс по объему тела необходимо учесть распределение жесткостей c_j пружинок, чтобы совокупность последних описывала упругие свойства реального тела.

Из этого примера видно, что структурное моделирование позволяет описывать поведение довольно сложных систем. При этом чем сложнее система, тем структурное моделирование становится все более эффективным (а в некоторых случаях — просто необходимым). Например, при описании поступательного движения твердого тела структурная модель, в принципе, не нужна, так как в этом случае тело можно представить в виде одной материальной точки, масса которой равна массе тела. В других рассмотренных выше примерах структурная модель помогает описать поведение достаточно сложных механических систем с помощью взаимодействующих простейших элементов. Для некоторых механических систем структурное моделирование является едва ли не единственным способом описания их поведения. Это относится, например, к моделированию поведения структурно-неоднородных материалов (композитов, полимеров, керамик и т.д.) [62, 102].

Следует отметить, что той информации, которая содержится в структурной схеме системы, очень часто бывает недостаточно для исследования. Поэтому графы надо рассматривать как вспомогательный инструмент при моделировании. Главным же при струк-

турном моделировании является установление конкретных функциональных связей между входными, внутренними и выходными параметрами. Поэтому без достаточно широкого арсенала методов математического моделирования здесь также не обойтись. В большинстве случаев приходится составлять уравнения, описывающие поведение каждого элемента структуры с учетом взаимодействия всех элементов. Это приводит к необходимости решения систем уравнений с большим числом переменных. Однако с появлением современной вычислительной техники такие подходы к моделированию сложных систем становятся все более распространенными.

В заключение данного параграфа кратко остановимся на классификации структурных моделей и примерах их применения.

Структурные модели бывают четырех видов: пространственные, временные, физические и иерархические.

Пространственные структуры обычно используют для описания геометрии исследуемого объекта и расположения в пространстве его отдельных элементов. Такие структуры хорошо описываются с помощью сетевых и матричных графов, вершины которых указывают места расположения элементов, а ребра — расстояния между ними или другие условия соединения. Пример пространственной структурной модели — структурная схема телефонной сети некоторого населенного пункта, в которой вершинами являются узлы связи, а ребрами — линии связи с указанием, например, числа каналов связи [106]. Другим примером пространственной структуры может служить конечно-элементная сетка, с помощью которой описывается геометрия исследуемого объекта при его, например, прочностном расчете методом конечных элементов [43].

Временные структурные модели широко используются в сетевом и календарном планировании [46], а также в теории массового обслуживания [50]. Во временных структурах в качестве элементов выступают этапы происходящего процесса или состояния системы в некоторый момент времени. Отношениями здесь служат условия перехода от одного этапа к другому или из одного состояния системы в другое. Например, на производстве широко применяют так называемые сетевые графики (технологические карты). Они представляют собой графы, вершинами которых служат необходимые производственные операции, а с помощью ребер указывается последовательность и длительность этих операций. При моделировании систем массового обслуживания удобно применять структурную схему «гибели и размножения», которая представляет собой линейный граф (последовательный набор состояний системы, вы-

тянутый в одну цепочку). Считается, что система обслуживает случайный поток заявок, поступающих в нее. Тогда отношениями между элементами системы (ее состояниями в различные моменты времени) служат условия поступления новой заявки, которые можно характеризовать, например, интенсивностью соответствующих случайных потоков событий [18].

Физические структурные модели применяются для описания сложных физических свойств исследуемого объекта с помощью простых структурных элементов. Пример моделирования упругих свойств тела был приведен выше. Другие примеры применения физического структурного моделирования будут приведены в конце данной главы.

Для моделирования управляемых систем широко применяются *иерархические структурные схемы*, предполагающие наличие нескольких уровней обработки информации и принятия решений. Основная задача иерархической структуры – распределение функций обработки информации и принятия решений между отдельными элементами. На рис. 4.7 приведен пример двухуровневой веерной иерархической структуры управления системой. В данной системе существует один привилегированный элемент, который имеет возможность управлять остальными элементами. Этот привилегированный элемент обычно называется *центром* (Ц), а остальные элементы – *производителями* ($\Pi_i, i=1, \dots, n$).

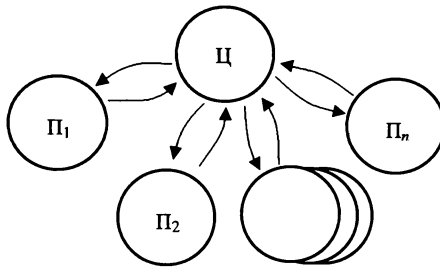


Рис. 4.7. Двухступенчатая веерная структурная схема

Отношениями в этой модели служат условия обмена информацией, денежными и материальными ресурсами между центром и производителями. Следует отметить, что в приведенной структурной схеме отношения между производителями (горизонтальные связи) отсутствуют. Неравноправие элементов системы проявляется

ся в том, что центр назначает правила формирования воздействий на производителей и тем самым имеет возможность направлять в нужное для него русло действия нижних элементов. Данную схему несложно обобщить до многоуровневой веерной структуры, которая широко используется в экономике.

В заключение можно отметить следующее. Как было показано, структурные модели нашли широкое применение в различных областях целенаправленной деятельности человека. В некоторых случаях они помогают построить модель исследуемой системы или процесса, а иногда являются единственно эффективным инструментом при моделировании. Однако остается вопрос: «Как правильно строить структурные модели?» Постараемся дать на него ответ.

4.2. СПОСОБЫ ПОСТРОЕНИЯ СТРУКТУРНЫХ МОДЕЛЕЙ

Не упрощайте жизнь, она и так проста:

В ней всюду и всему отведены места.

Не усложняйте жизнь, она и так сложна:

Чтоб разобраться в ней, вторая жизнь нужна.

В. Кривилев

При структурном моделировании широко применяются методы *анализа* и *синтеза*. С помощью методов анализа производится разделение рассматриваемого объекта на части и исследование каждой из этих частей в отдельности. Методы синтеза, наоборот, служат для соединения частей в целое. Следует отметить, что при структурном моделировании методы анализа и синтеза необходимо применять совместно. Важно не просто разбить целое на отдельные элементы, но и соединить эти элементы таким образом, чтобы они снова образовали единое целое. С этой точки зрения синтез является завершающим этапом анализа, так как только после этого этапа можно объяснить целое через его части – в виде структуры целого. Это связано с тем, что при анализе теряются важные свойства объекта как целого (разобранный автомобиль не поедет), так и отдельных его элементов (оторванный руль «не рулит»). Поэтому, как отмечал один из ведущих специалистов по системному анализу Р. Акофф, результатом анализа является лишь вскрытие структуры системы, знание о том, как система работает, но не понимание того, почему и зачем она это делает. Только после син-

теза можно объяснить поведение системы, рассматривая каждый элемент и его роль через призму всей системы.

Таким образом, анализ и синтез нельзя рассматривать как отдельные методы. Они дополняют друг друга и при структурном моделировании должны применяться совместно. Только в этом случае построенная структурная модель будет отражать основные свойства исследуемого объекта согласно поставленным целям.

Теперь рассмотрим методы анализа и синтеза более подробно. Начнем с методов анализа, которые нашли широкое применение в науке и практике. В математике давно с успехом применяются такие аналитические методы, как разложение функций в ряды, спектральный анализ, дифференциальное и интегральное исчисление; в физике — методы молекулярной динамики; на производстве — конвейерная технология изготовления.

Как было отмечено выше, основной операцией при анализе является разделение целого на части. В дальнейшем эту операцию будем называть *декомпозицией* и понимать под ней метод разложения системы на отдельные элементы. В результате декомпозиции исходная система распадается на подсистемы, задача — на подзадачи и т.д. При необходимости операция декомпозиции может повторяться несколько раз, что приводит к древовидным структурам системы.

Основной проблемой при декомпозиции является ее неоднозначность. Понятно, что одну и ту же систему можно разбить на различные подсистемы в зависимости как от опыта исследователя, так и от применяемой методики анализа. Поэтому в системном анализе существуют специальные критерии для обоснования процесса декомпозиции. Одним из таких критериев является полнота декомпозиции, которая в свою очередь связана с полнотой модели системы, взятой в качестве исходной при декомпозиции.

Как было указано в разд. 4.1, в системном анализе рассматриваются только четыре формальные модели системы: «черного ящика», состава, структуры и «белого ящика». Очевидно, что для декомпозиции подходят только последние три модели, которые позволяют рассматривать систему через взаимодействие ее отдельных элементов. Поэтому одной из проблем системного анализа является накопление наборов полных формальных моделей (полнота здесь рассматривается относительно поставленной цели) для различных исследуемых систем, получивших в теории искусственного интеллекта название *фреймов* [83].

Известны фреймы для некоторых информационных, организационных и социальных систем [83]. Однако построение и обосно-

вание фреймов для произвольной системы остается сложной задачей, от решения которой во многом зависит успех декомпозиции. Рассмотрим, например, педагогический процесс в высшем учебном заведении (вузе). Фреймом при декомпозиции в данном случае может служить формальная модель деятельности человека, предложенная еще К. Марксом в «Капитале» для анализа процесса труда (рис. 4.8). В качестве элементов здесь выделены: субъект деятельности, объект, на который направлена деятельность, и средства, используемые в процессе деятельности, а также все возможные связи между ними и окружающей средой. Используя эту формальную модель, можно построить модель педагогического процесса в вузе, вариант которой изображен на рис. 4.9. Здесь субъектом выступает преподаватель вуза, объектами – студенты, а средствами – методические, информационные и технические средства обучения (учебные программы изучаемых предметов, методические и учебные пособия, лабораторная база и т.д.). Окружающая среда описывается с помощью трех элементов: школы, вуза и министерства, которые оказывают существенное влияние на организацию учебного процесса (понятно, что число элементов, входящих в приведенную структурную модель, может быть гораздо больше).

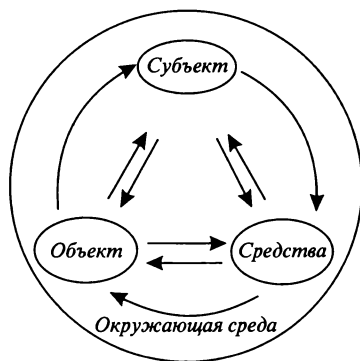


Рис 4.8. Полная формальная модель деятельности человека

Однако при декомпозиции необходимо учитывать и другой критерий – простоту, который требует сокращения размеров древовидной структуры. Таким образом, при декомпозиции должен быть принят некий компромисс между полнотой и простотой, который может быть достигнут в том случае, если в структурную модель



Рис 4.9. Формальная модель педагогического процесса в вузе

включаются только элементы, существенные по отношению к цели анализа.

Число уровней декомпозиции (уровней древовидной структуры) выбирается из следующих соображений. Декомпозиция по каждой из ветвей древовидной структуры ведется до тех пор, пока не приведет к получению элементов системы, не требующих дальнейшего разложения. Такие составляющие называются *элементарными*. Отметим, что понятие элементарности составляющей системы должно быть конкретизировано в каждом рассматриваемом случае отдельно. При этом могут быть использованы как формализованные (с помощью критериев), так и неформализованные (с помощью экспертов) способы. Например, в некоторых случаях многомерную задачу механики сплошной среды удастся разложить на последовательность одномерных задач, имеющих простое (аналитическое) решение. Тогда каждую из этих одномерных задач можно считать элементарной частью исходной системы.

Часть системы, которую нельзя считать элементарной на основании выбранных критериев, подлежит дальнейшей декомпозиции. При этом могут использоваться различные фреймы. Если исследователь «перебрал» все фреймы, но не достиг элементарности на какой-либо ветви древовидной структуры, то вводятся новые элементы в модель, взятую в качестве основания, и декомпозиция про-

должается по ним. Подробная блок-схема алгоритма декомпозиции приведена в [83].

Следует отметить, что в результате декомпозиции будет реализован лишь первый этап структурного моделирования, а именно – этап анализа. После этого этапа удастся разделить исследуемую систему на отдельные элементы (или исходную задачу – на более простые подзадачи). Однако, как было отмечено выше, на поведение каждого элемента нужно смотреть с точки зрения целей всей системы. Другими словами, полученная совокупность элементов кроме внешней целостности (т.е. определенной обособленности от окружающей среды) должна обладать и внутренней целостностью.

Внешняя целостность хорошо описывается моделью «черного ящика», а внутренняя – связана с моделью структуры системы, т.е. установлением отношений между элементами. Для этого используется операция *агрегирования* – объединение нескольких элементов в единое целое. Результатом агрегирования является система, которую называют *агрегатом*. Свойства агрегата не являются только совокупностью свойств его отдельных элементов. Агрегат может обладать такими свойствами, которых нет ни у одного из его элементов, взятых в отдельности. Другими словами, объединение элементов в систему влечет появление нового качества, которое не могло появиться без этого объединения. Такое «внезапное» появление новых качеств у агрегата получило название *эмерджентности* (от англ. emergent – внезапно возникающий). Следует отметить, что новые свойства возникают благодаря конкретным связям между элементами. Другие связи могут дать другие новые свойства агрегата.

Хорошей иллюстрацией свойства эмерджентности является пример, предложенный М. Арбибом [83]. Пусть имеется некоторый цифровой автомат S , увеличивающий на 1 любое целое число, поступающее на его вход. При последовательном соединении двух автоматов в цепочку это свойство не изменяется. Если же соединить два таких автомата последовательно в кольцо (рис. 4.10), то в полученном агрегате обнаружится новое свойство: он генерирует возрастающие последовательности на выходах А и В, причем одна последовательность состоит из четных, а другая – из нечетных чисел. Другим ярким подтверждением свойства эмерджентности может служить пример из материаловедения. Известно, что тип кристаллической решетки (способ соединения атомов) определяет твердость материала. При этом твердость получаемого агрегата, состоящего из одинаковых элементов, может различаться в десятки тысяч раз (графит и алмаз).

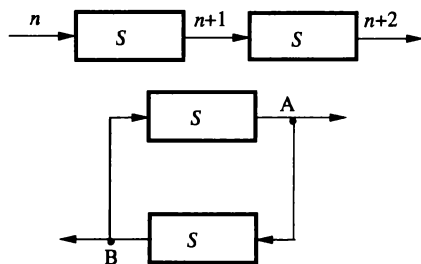


Рис. 4.10. Пример вычислительного агрегата

Возникновение качественно новых свойств при агрегировании есть частное, но яркое проявление одного из законов диалектики — закона перехода количества в качество. При этом считается, что чем больше свойства агрегата отличаются от свойств его элементов, тем выше организованность системы. Кибернетик У. Эшби доказал, что у системы тем больше возможностей в выборе поведения, чем сильнее степень согласованности поведения ее элементов. Высшая степень проявления согласованности поведения элементов системы — самоорганизация системы, изучением которой занимается относительно молодая междисциплинарная область знаний — *синергетика* [81] (от греч. *synergos* — вместе действующий).

Таким образом, как следует из вышеизложенного, при агрегировании большое значение имеет установление связей между элементами, т.е. выбор модели структуры. Значит, в самом общем виде агрегирование можно определить как установление отношений на заданном множестве элементов. Такое установление отношений может быть проведено различными способами: построением математических зависимостей, структурированием, статистической обработкой, классификацией и т.п. В результате получают различные агрегаты, основными из которых являются следующие [83]: конфигуратор, классификатор, оператор, статистик и структура. Рассмотрим эти агрегаты более подробно.

Конфигуратором называется такой агрегат, который состоит из качественно различных языков описания исследуемого объекта и обладает тем свойством, что число этих языков минимально, но необходимо для выполнения заданной цели. Следует отметить, что конфигуратор является содержательной моделью высшего возможного уровня. Перечислив языки, на которых будет вестись описание системы, мы тем самым определяем тип системы и ее основные свойства.

Например, в радиотехнике для описания одного и того же прибора используется следующий конфигуратор: блок-схема, принципиальная схема и монтажная схема. Этот конфигуратор полностью описывает рабочие характеристики прибора. Однако если кроме цели производства радиоаппаратуры ставится цель ее сбыта, то в конфигуратор необходимо добавить язык рекламы (маркетинг, дизайн, цена и т.п.).

В инженерной графике для описания поверхности любого трехмерного тела в качестве конфигуратора используются совокупность трех ортогональных проекций. Число их нельзя уменьшить и нецелесообразно увеличивать.

А какой конфигуратор применяется при математическом моделировании? Выбор языка зависит от вида модели. Понятно, что основным языком для математической модели — язык математических формул. Однако, как было отмечено в гл. 1 и 2, важными этапами математического моделирования являются содержательная и концептуальная постановки задачи. На этапе содержательной постановки осуществляется словесная постановка задачи на том языке, на котором она формулируется заказчиком. На этапе концептуальной постановки выполняется запись задачи на языке тех областей знаний, которые используются при моделировании рассматриваемого объекта. Поэтому конфигуратором в данном случае можно считать содержательную, концептуальную и математическую постановки задачи.

В качестве *классификатора* выступает агрегат, устанавливающий отношения эквивалентности между элементами системы, т.е. описывающий условия образования классов.

Говорят, что на множестве A определено отношение R , если по некоторому правилу составлены упорядоченные пары элементов, находящихся в отношении. При этом пишут aRb , $a, b \in A$.

Отношение R на A , удовлетворяющее аксиомам:

- 1) aRa (элемент a эквивалентен самому себе),
- 2) $aRb \Rightarrow bRa$ (если элемент a эквивалентен элементу b , то элемент b эквивалентен элементу a),
- 3) $aRb, bRc \Rightarrow aRc$ (если элемент a эквивалентен элементу b и элемент b эквивалентен элементу c , то элемент a эквивалентен элементу c),

называется отношением эквивалентности. Оно разбивает множество элементов системы на классы.

Следующим типом агрегата является *оператор*, который ставит в соответствие некоторому набору отдельных элементов один элемент.

Наиболее распространенный в математическом моделировании вид оператора — *функция*. Этот вид оператора появляется, если агрегируемые элементы измеряются в числовых шкалах. В таком случае можно задать отношение на множестве элементов в виде числовой функции многих переменных f , которая и является агрегатом, т.е.

$$f : R^n \rightarrow R, \quad (4.4)$$

где R^n — n -мерное евклидово пространство; R — вещественная ось. Здесь элементом отображаемого пространства R^n является n -мерный вектор переменных системы $x = (x_1, x_2, \dots, x_n)$, характеризующий ее поведение.

Приведенный вид функции — один из простейших. В общем случае области определения и значений функции могут относиться к более сложным пространствам и множествам. Конкретное задание функции $f(x)$ связано с построением математической модели рассматриваемой системы. Поэтому на выбор функции накладываются ограничения, вытекающие из содержательной постановки задачи, т.е. этот выбор не является свободным. В тех же (достаточно редких) случаях, когда оператор-функция является вполне адекватной математической моделью всей системы, свобода выбора функции, агрегирующей набор внутренних переменных, вообще отсутствует. Такой случай имеет место, например, когда закономерности природы удается с достаточной степенью адекватности отобразить безразмерными степенными одночленами физических размерных величин. При этом можно утверждать, что если удалось построить безразмерный степенной одночлен из размерных физических величин, образующих конфигурацию определенного явления, то установлен физический закон данного явления. Это легко можно показать на примере второго закона Ньютона, описывающего поступательное движение твердого тела. Безразмерный одночлен здесь имеет вид

$$m_1 a / F = 1, \quad (4.5)$$

что подтверждает правомерность полученного закона.

К сожалению, построить функциональную зависимость, адекватно описывающую поведение сложной системы, очень трудно,

а иногда практически невозможно. Гораздо проще установить функциональные зависимости между отдельными элементами системы. В этом случае оператор будет представлять собой некоторую (часто нелинейную) систему уравнений. Как правило, внутренними переменными системы являются не числа, а функции одного или нескольких аргументов. Тогда выходными параметрами могут выступать также функции, или функционалы. Например, для динамических систем, описывающих процессы и явления, изменяющиеся во времени, связь между внутренними параметрами $x(t)$ и выходными параметрами системы $y(t)$ в операторной форме имеют вид

$$y(t) = f(x(t)), \quad t \in [0, T], \quad (4.6)$$

где оператор f обычно представляет собой систему дифференциальных уравнений, T – время протекания процесса. Пример построения подобного оператора для механической системы, состоящей из нескольких тел, будет приведен ниже.

При математическом моделировании сложных систем построить оператор f бывает совсем не просто. Это связано со многими причинами. Основной из этих причин можно считать недостаток информации о характере и механизмах взаимодействия между отдельными элементами системы. Например, эти взаимодействия могут носить случайный характер, закон которого нам неизвестен. В этом случае говорят, что моделирование ведется в условиях неопределенности, а оператор f может быть найден только с некоторой ограниченной точностью, например с точностью до конечного числа параметров $\theta = (\theta_1, \theta_2, \dots, \theta_n)$:

$$y(t) = f(x(t), \theta), \quad t \in [0, T]. \quad (4.7)$$

Обычно считается, что параметры θ носят случайный характер или могут быть определены в ходе самого моделирования с помощью методов идентификации [57]. В этом случае используются системы с обратной связью [83]. (Более подробно о способах моделирования в условиях неопределенности будет сказано в следующей главе.) Назовем оператор, который задается с помощью алгоритма, реализующего некоторый набор правил, *имитатором*. Примеры построения агрегата-имитатора приведены в гл. 7.

Отдельно при агрегировании рассматривается ситуация, когда все параметры, описывающие поведение элементов системы, являются случайными величинами. В этом случае вводится понятие *агрегата-статистика*, определяющего отношения на множестве слу-

чайных параметров системы. Для его построения используются функции выборочных значений случайных величин, в качестве которых широко используются функции распределения вероятностей или плотности распределения вероятностей случайных событий [28].

На практике используются достаточные и оптимальные статистики. Достаточными статистиками называются такие агрегаты, которые извлекают всю полезную информацию об интересующем нас параметре из совокупности наблюдений. Например, для систем, результат деятельности которых нас интересует только в среднем, достаточным статистиком может служить математическое ожидание выходной случайной величины. Однако на практике достаточные статистики применяются редко, так как при таком агрегировании потери информации неизбежны. Чаще применяются оптимальные статистики – такие агрегаты, которые позволяют свести потери информации к минимуму. Например, математическое ожидание и дисперсия случайной величины в совокупности являются оптимальным статистиком для многих технических систем. Более подробно способы описания случайных величин при моделировании будут рассмотрены в гл. 5.

Последним из рассматриваемых видов агрегатов, но не последним по частоте применения при структурном моделировании является структура системы, т.е. агрегат, устанавливающий типы связей между отдельными элементами системы. Наиболее широко подобный вид агрегирования применяется при моделировании технических, информационных и организационных систем. Например, в материаловедении в качестве структуры материала используются разные модели кристаллических решеток, устанавливающие типы связи между атомами и симметричные свойства кристалла. В информационных системах применяется структура в виде первичной сети, указывающей направление и интенсивность передачи информации. В организационных системах структура описывает иерархию в процессе принятия решений и распределение власти и ответственности (ответственность за принятые решения).

После рассмотрения основных подходов к построению структурных моделей перейдем к примерам, иллюстрирующим эффективность этих подходов при моделировании различных систем и процессов.

4.3. ПРИМЕРЫ СТРУКТУРНЫХ МОДЕЛЕЙ

*Я слышу и забываю.
Я вижу и запоминаю.
Я делаю и понимаю.*
Конфуций

Рассмотрим четыре примера построения математических моделей, использующих элементы структурного моделирования. Еще один пример, иллюстрирующий эффективность применения структурных схем при моделировании систем массового обслуживания, будет приведен в гл. 5.

Пример 4.1. *Моделирование системы «Солнце—Земля—Луна».*

Содержательная постановка задачи. Требуется определить и исследовать траектории движения центров Земли и Луны вокруг Солнца с учетом сил взаимодействия между планетами.

Концептуальная постановка задачи. Для решения поставленной задачи достаточно исследовать динамику системы, состоящей из трех элементов, взаимодействующих между собой (рис. 4.11). Взаимодействием элементов системы с окружающей средой (другими планетами Солнечной системы и далекими звездами) вследствие малости их влияния можно пренебречь. Также не будем учитывать влияние формы планет на исследуемые траектории. Тогда с большой степенью точности можно считать, что линии действия равнодействующих сил притяжения между планетами проходят через центры планет. Кроме того, поскольку масса Солнца во много раз больше массы Земли и Луны вместе взятых, действием сил притяжения Луны и Земли на движение Солнца также можно пренеб-

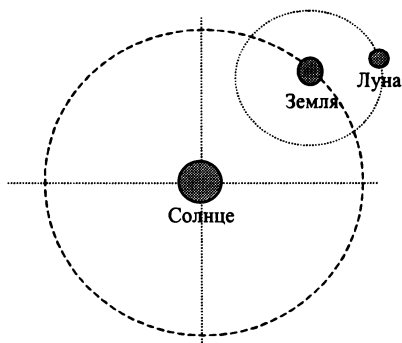


Рис. 4.11. Система «Солнце—Земля—Луна»

речь. Все это позволяет в качестве моделей каждого элемента системы выбрать материальную точку с заданной массой. При этом одну точку можно считать условно неподвижной (Солнце), а две другие – движущимися в одной плоскости под действием центральных сил, определяемых согласно закону всемирного тяготения. В этом случае для произвольного момента времени структурная схема системы будет иметь вид, изображенный на рис. 4.12.

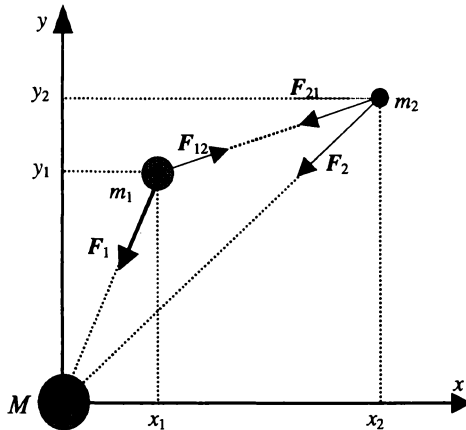


Рис. 4.12. Структурная схема системы двух тел в поле центральных сил

Теперь с учетом того, что положение точки на плоскости в любой момент времени описывается с помощью двух координат, определение траекторий движения точек 1 и 2 данной системы сводится к нахождению функций координат от времени $x_i(t)$, $y_i(t)$, $i = 1, 2$, которые и будут выходными параметрами построенной структурной модели.

Осталось установить функциональную связь между выходными и входными параметрами системы, к которым, в частности, относятся массы планет и расстояния между их центрами. Для этого в соответствии с выбранной структурной схемой достаточно записать дифференциальные уравнения движения двух материальных точек, находящихся в поле центральных сил. Однако это необходимо делать в рамках следующего этапа математического моделирования, а именно – математической постановки исходной задачи.

Математическая постановка задачи. Требуется определить функции координат от времени $x_i(t)$, $y_i(t)$, $i = 1, 2$, описывающие движение двух материальных точек, находящихся под действием четырех центральных сил, указанных на структурной схеме (рис. 4.12).

Используя второй закон Ньютона, уравнения движения для каждой материальной точки в проекциях на оси координат Ox можно записать в следующей форме:

$$\begin{cases} m_1 a_{1x} = -\gamma \frac{m_1 M}{r_1^3} x_1 - \gamma \frac{m_1 m_2}{r_{12}^3} (x_1 - x_2), \\ m_1 a_{1y} = -\gamma \frac{m_1 M}{r_1^3} y_1 - \gamma \frac{m_1 m_2}{r_{12}^3} (y_1 - y_2), \end{cases} \quad (4.8a)$$

$$\begin{cases} m_2 a_{2x} = -\gamma \frac{m_2 M}{r_2^3} x_2 - \gamma \frac{m_1 m_2}{r_{12}^3} (x_2 - x_1), \\ m_2 a_{2y} = -\gamma \frac{m_2 M}{r_2^3} y_2 - \gamma \frac{m_1 m_2}{r_{12}^3} (y_2 - y_1), \end{cases} \quad (4.8b)$$

где $r_1 = \sqrt{x_1^2 + y_1^2}$, $r_2 = \sqrt{x_2^2 + y_2^2}$, $r_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$,

$\gamma = 6,672 \cdot 10^{-11}$ н·м²/кг – гравитационная постоянная.

Заменяя проекции ускорений a_{ix} , a_{iy} (производных соответствующих проекций скоростей по времени) и скоростей v_{ix} , v_{iy} (производных координат точек) разностными аналогами и выполняя алгебраические преобразования, приходим к системе алгебраических уравнений

$$\begin{cases} v_{1x}(t + \Delta t) = v_{1x}(t) - \gamma M x_1(t) \Delta t / r_1^3(t) - \gamma m_2 (x_1(t) - x_2(t)) \Delta t / r_{12}^3(t), \\ v_{1y}(t + \Delta t) = v_{1y}(t) - \gamma M y_1(t) \Delta t / r_1^3(t) - \gamma m_2 (y_1(t) - y_2(t)) \Delta t / r_{12}^3(t), \\ x_1(t + \Delta t) = x_1(t) + v_{1x}(t) \Delta t, \\ y_1(t + \Delta t) = y_1(t) + v_{1y}(t) \Delta t, \\ r_1(t) = \sqrt{x_1^2(t) + y_1^2(t)}, \\ v_{2x}(t + \Delta t) = v_{2x}(t) - \gamma M x_2(t) \Delta t / r_2^3(t) - \gamma m_1 (x_2(t) - x_1(t)) \Delta t / r_{12}^3(t), \\ v_{2y}(t + \Delta t) = v_{2y}(t) - \gamma M y_2(t) \Delta t / r_2^3(t) - \gamma m_1 (y_2(t) - y_1(t)) \Delta t / r_{12}^3(t), \\ x_2(t + \Delta t) = x_2(t) + v_{2x}(t) \Delta t, \\ y_2(t + \Delta t) = y_2(t) + v_{2y}(t) \Delta t, \\ r_2(t) = \sqrt{x_2^2(t) + y_2^2(t)}, \\ r_{12}(t) = \sqrt{(x_1(t) - x_2(t))^2 + (y_1(t) - y_2(t))^2}, \end{cases} \quad (4.9)$$

к которой добавляются начальные условия для величин $x_1(0)$, $y_1(0)$, $v_{1x}(0)$, $v_{1y}(0)$ и $x_2(0)$, $y_2(0)$, $v_{2x}(0)$, $v_{2y}(0)$. Решая численно систему (4.9) с учетом начальных условий, получаем таблицу значений координат и скоростей планет в зависимости от времени. Траектории движения Земли и Луны, находящихся в поле центральных сил, приведены на рис. 4.13.

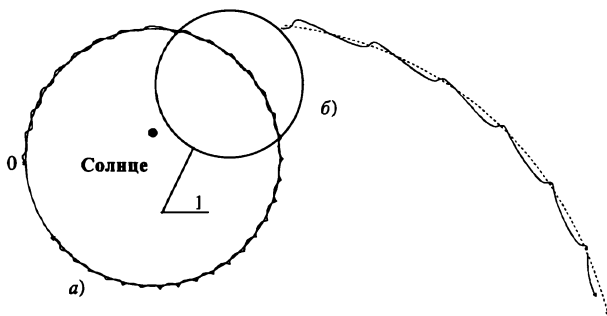


Рис. 4.13. Траектории движения Земли и Луны в поле центральных сил (а); фрагмент 1 траектории движения Луны в окрестности траектории Земли (б)

Пример 4.2. Структурная модель упругопластического тела.

Многие тела при деформировании проявляют упругие, пластические и вязкие свойства. Для моделирования подобных сложных систем удобно использовать структурные элементы, описывающие в отдельности тот или иной физический механизм деформирования [48].

Упругий механизм деформирования, как было показано в разд. 4.1, можно моделировать с помощью структурного элемента в виде линейной или нелинейной пружины (рис. 4.14,а). На диаграмме $\sigma - \epsilon$ (напряжение—деформация) угол наклона прямой (или модуль упругости тела E) зависит от жесткости пружины s .

Пластический механизм может быть смоделирован с помощью структурного элемента сухого трения (P), изображенного на рис. 4.14,б. Из диаграммы $\sigma - \epsilon$ видно, что пластическое деформирование может начаться только при достижении некоторого напряжения σ_s , которое называется пределом текучести и значение которого ассоциируется с величиной коэффициента сухого трения f .

Вязкий механизм можно изобразить с помощью структурного элемента вязкого трения (рис. 4.14,в), сопротивление перемещению которого зависит от коэффициента вязкого трения ν и скорости перемещения.

Теперь, рассматривая различные комбинации соединения структурных элементов, можно строить структурные модели тела с различными деформационными свойствами. Таким образом, в данном случае тело моделируется с помощью агрегата, состоящего из нескольких соединенных между собой структурных элементов. Свойства агрегата в значительной степени зависят от способа соединения элементов (структурной схемы).

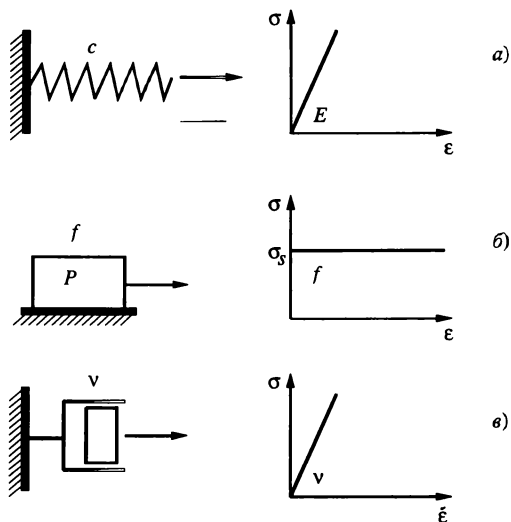


Рис. 4.14. Структурные элементы деформирования

На рис. 4.15 приведено несколько простейших агрегатов. Как видно, путем перестановки последовательно соединенных двух структурных элементов можно получить агрегаты с различными свойствами. Первый из них (рис. 4.15,а) будет обнаруживать лишь пластические свойства (модель упрочняющегося жестко-пластического тела), а второй (рис. 4.15,б) — упругие и пластические свойства (модель идеального упругопластического тела).

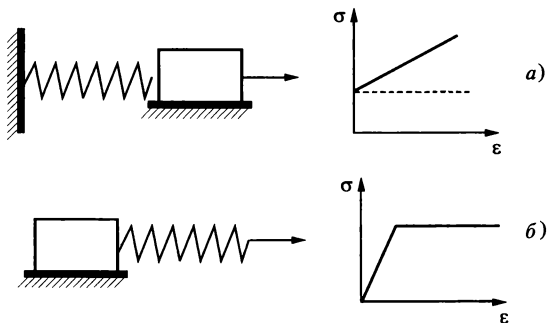


Рис. 4.15. Виды упругопластических агрегатов

С помощью различного соединения структурных элементов можно получать и другие интересные свойства агрегатов. Например, при последовательном соединении вязкого и упругого элементов в них возникают одинаковые усилия (напряжения), но различные перемещения и скорости перемещений (деформации и скорости деформаций). Такая модель называется *телом Максвелла* (рис. 4.16,а). При параллельном соединении этих же структурных элементов свойства агрегата изменяются (рис. 4.16,б); элементы будут испытывать одинаковые деформации (скорости деформаций), но различные напряжения (*тело Фойхта*).

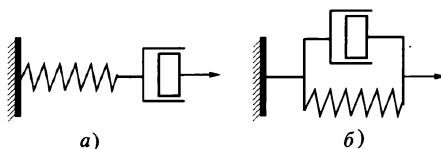


Рис. 4.16. Структурные модели: а – Максвелла; б – Фойхта

Следует отметить, что все приведенные выше структурные модели описывают только одноосное напряженно-деформированное состояние, хотя с помощью более сложных структур можно моделировать и сложное (трехосное) нагружение [48].

После построения структурной модели тела можно получить для этого тела определяющие соотношения, устанавливающие функциональную зависимость между мерами напряженного и деформированного состояний всего агрегата, что является одной из самых сложных задач в механике деформируемого твердого тела (в частности, в теории пластичности) [85]. В соответствии с таким подходом вначале обычно строится модель материала, описывающая одноосное нагружение. С этой целью на основе физического анализа и имеющихся экспериментальных данных определяется набор элементарных структурных элементов, отвечающих за тот или иной механизм деформирования. Затем строится структурная схема, включающая различные способы соединения введенных структурных элементов (т.е. получения агрегатов). Записываются одноосные определяющие соотношения, константы которых уточняются из экспериментов (на одноосное нагружение). После этого с помощью дополнительных гипотез осуществляется переход к многомерным моделям.

Рассмотрим пример получения определяющих соотношений, приведенный в [48].

Содержательная постановка задачи. Построить определяющие соотношения для упругопластического упрочняющегося изотропного тела, учитывающие упругие, пластические и вязкие механизмы деформирования.

Концептуальная постановка. Для построения связи между мерами напряженного и деформированного состояний (и/или их скоростями) уп-

ругопластического тела рассмотрим для произвольного момента процесса деформирования двумерную структурную модель тела, включающую все виды структурных элементов и условно изображенную на рис. 4.17. Отметим, что в данной схеме внешние упругие элементы моделируют упругое деформирование, а внутренний упругий элемент описывает эффект деформационного упрочнения. Вязкие элементы моделируют скоростное упрочнение (или разупрочнение) материала при деформировании. Вследствие того, что тело считается изотропным, упругие и вязкие свойства тела в обоих направлениях совпадают. Это можно легко смоделировать, задав одинаковые характеристики структурных элементов (c и ν).

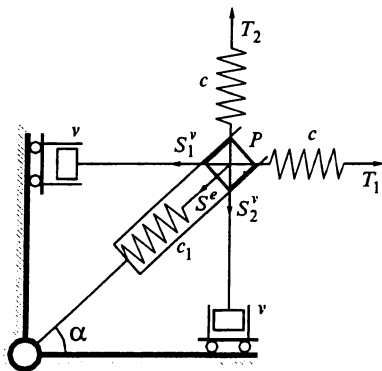


Рис. 4.17. Двумерная структурная модель упругопластического упрочняющегося тела

Обозначим усилия, действующие на внешние упругие элементы, через T_1 и T_2 , усилия во внутренних вязких элементах – через S_1^v и S_2^v , а усилие во внутреннем упругом элементе – через S^e . Перемещения во внешних упругих элементах обозначим через l_1 и l_2 , перемещения пластического элемента – через q_1 и q_2 . Требуется установить функциональную зависимость между внешними воздействиями (усилиями T_1 и T_2) и откликом системы (ее перемещениями). Другими словами, необходимо определить функциональную зависимость вида $\epsilon = F(\sigma)$, где ϵ – некоторая мера деформаций; σ – мера напряженного состояния; F – неизвестный оператор, вид которого требуется определить. При этом будем считать, что полная деформация складывается из упругой, пластической и вязкой деформаций.

Математическая постановка задачи. Найти функциональную зависимость (вид оператора F) между мерами напряженного и деформированного состояний упругопластического упрочняющегося тела. Для этого вос-

пользуемся построенной структурной моделью. Для внешних упругих элементов можно записать

$$l_1 = \frac{1}{c} T_1; \quad l_2 = \frac{1}{c} T_2. \quad (4.10)$$

Очевидно, что перемещение пластического элемента P (поршень в цилиндре с сухим трением) будет происходить, когда результирующие усилия достигнут некоторого предельного значения

$$(T_1 - S_1)^2 + (T_2 - S_2)^2 = K^2, \quad K = \text{const}, \quad (4.11)$$

где K — предельное значение силы сухого трения, а внутренние усилия определяются следующим образом:

$$S_1 = S_1^y + S^e \cos \alpha; \quad S_2 = S_2^y + S^e \sin \alpha; \quad \tan \alpha = T_2/T_1.$$

Приращение перемещения элемента пластичности происходит по направлению равнодействующей, т.е.

$$\frac{\Delta q_1}{T_1 - S_1} = \frac{\Delta q_2}{T_2 - S_2}. \quad (4.12)$$

Сила натяжения во внутренней пружине связана с перемещениями следующим образом:

$$S^e = c_1 q. \quad (4.13)$$

Перемещения вязких элементов связаны с усилиями следующими соотношениями:

$$S_1^y = v \dot{q}_1; \quad S_2^y = v \dot{q}_2. \quad (4.14)$$

Соотношения (4.10)–(4.14) полностью описывают механическое поведение построенной структурной модели. Переходя от усилий к напряжениям, а от перемещений к деформациям, их можно переписать в форме, принятой в теории пластичности [48]. Предварительно заметим, что из условия (4.11) следует, что существует функция нагружения g , с помощью которой можно установить момент возникновения пластических деформаций. При этом функция нагружения имеет вид окружности постоянного радиуса R . Тогда условие (4.11) можно переписать в виде

$$g(\sigma - S) = R, \quad R = \text{const}. \quad (4.15)$$

Здесь под S понимается тензор внутренних напряжений. Это условие известно в теории пластичности как условие пластичности Мизеса [48].

Соотношения (4.12) показывают, что вектор скоростей пластических деформаций должен быть направлен по нормали к поверхности нагружения g , что соответствует известному ассоциированному закону пластического течения [48]:

$$d\epsilon^P = d\lambda \frac{\partial g}{\partial \sigma}, \quad (4.16)$$

где $d\lambda$ – некоторый множитель. Тогда определяющие соотношения, соответствующие построенной структурной модели, имеют вид:

$$\begin{aligned} d\epsilon &= d\epsilon^e + d\epsilon^P, \\ d\epsilon^e &= \frac{1}{E} d\sigma, \quad d\epsilon^P = d\lambda \frac{\partial g}{\partial \sigma}, \\ g(\sigma - S) &= R, \quad R = \text{const}, \\ dS &= E_1 d\epsilon^P, \quad S = v \frac{d\epsilon^v}{dt}. \end{aligned} \quad (4.17)$$

Здесь под E_1 понимается модуль упрочнения материала, зависящий от жесткости внутренней пружины и коэффициента вязкого трения.

Система (4.17) полностью описывает поведение упругопластичного тела с учетом эффектов деформационного и скоростного упрочнения. При этом, изменяя параметры системы (c , c_1 , v и f), можно получать различные диаграммы деформирования (рис. 4.18). Например, задавая жесткость внутренней пружины и коэффициент вязкого трения пренебрежимо малыми, получим модель идеальноупругопластического тела (кривая 1 на рис. 4.18). Если, наоборот, эти коэффициенты задать достаточно большими

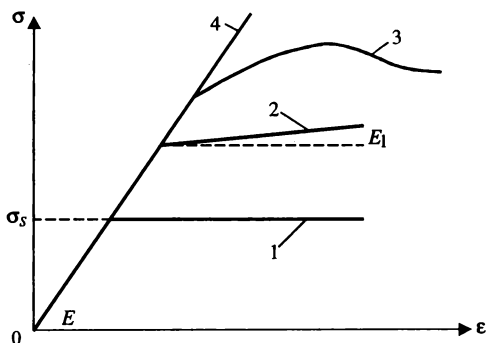


Рис. 4.18. Различные диаграммы деформирования

ми, то пластический элемент «включаться» не будет, и получим модель упругого тела (прямая 4). В том случае, когда коэффициент вязкого трения мал, а жесткости внешних пружин выше жесткости внутренней пружины, имеем модель упругопластического тела с линейным упрочнением (кривая 2). И, наконец, можно подобрать коэффициент вязкого трения и жесткость внутренней пружины таким образом, что получим модель упругопластического тела с нелинейным упрочнением (кривая 3).

Пример 4.3. *Оптимальное распределение инвестиций в системе с иерархической структурой.*

В качестве примера моделирования системы с иерархической структурой рассмотрим задачу, приведенную в [75].

Содержательная постановка задачи. Пусть задано некоторое объединение N промышленных предприятий (акционерное общество, трест, синдикат и т.п.), выпускающих однотипную продукцию. Известно, что в данном объединении принята система управления в виде двухступенчатой всеерной иерархической структуры (см. рис. 4.7). Во главе объединения находится Совет директоров, который принимает решения и который в дальнейшем будем называть центром. В распоряжении центра находится инвестиционный фонд, объем средств которого равен U . Требуется найти такое оптимальное распределение инвестиций между предприятиями, при котором общий объем продукции, выпускаемой всем объединением, будет максимальным.

Концептуальная постановка задачи. В данном случае система представляет собой совокупность $N + 1$ равноправных элементов. При заданной схеме управления центр является привилегированным элементом, хотя сам продукции не производит и результат его деятельности определяется результатом производственной деятельности производителей $\Pi_i, i = \overline{1, N}$. Обозначим через P_i объем продукции, выпускаемой i -м предприятием. В соответствии с содержательной постановкой задачи, целевая функция центра имеет вид

$$J_0 = \sum_{i=1}^N P_i. \quad (4.18)$$

При данной схеме управления центр не может назначать объемы производства $P_i, i = \overline{1, N}$, он лишь косвенно влияет на них, учитывая цели производителей. Будем считать, что объем производства P_i зависит от основных фондов производителей x_i и численности рабочих l_i , т.е.

$$P_i = f_i(x_i, l_i), \quad i = \overline{1, N}. \quad (4.19)$$

Существует несколько моделей, описывающих производство с помощью различных производственных функций f_i . Используем функцию

Кобба–Дугласа, получившую широкое распространение в экономических моделях производства [75]. Она имеет вид

$$P_i = \alpha_i (x_i + u_i)^{k_i} l_i^{1-k_i}, \quad k_i \in (0, 1), \quad i = \overline{1, N}, \quad (4.20)$$

где α_i , k_i – характеристики i -го предприятия; u_i – количество инвестиций, выделяемых i -му предприятию центром на развитие основных фондов.

Будем считать, что целью производителей является максимизация прибыли J_i , которая равна стоимости произведенной продукции за вычетом затрат на ее производство. При этом затраты состоят только из фонда оплаты труда. Тогда доход i -го предприятия можно представить в виде

$$J_i = c_i P_i - \omega_i l_i, \quad (4.21)$$

где c_i – цена единицы продукта; ω_i – средняя ставка заработной платы.

Положим также, что величина фондов i -го предприятия x_i фиксирована, а выделяемые центром инвестиции u_i направляются на увеличение объема производства. Тогда объем продукции P_i зависит только от величины l_i , которая является управляющим параметром i -го производителя.

Отметим, что в данном случае выбор l_i зависит от u_i , а величина u_i , выделяемая центром, зависит от P_i . В этой взаимосвязи и заключается сложность поставленной задачи. Структурная схема системы показана на рис. 4.19. С помощью этой схемы можно значительно упростить исходную задачу, разбив ее на $N + 1$ подзадачу на этапе концептуальной постановки.

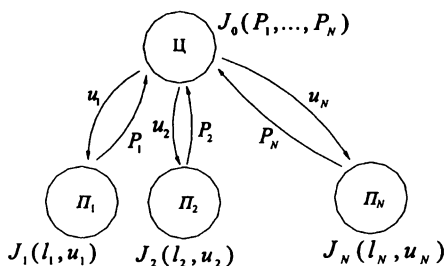


Рис. 4.19. Иерархическая структурная схема системы неравноправных элементов

Математическая постановка задачи. Найти такие \bar{u}_i , \bar{l}_i , $i = \overline{1, N}$, которые сообщают максимальные значения целевым функциям центра и производителей:

$$J_0 = \sum_{i=1}^N P_i(u_i, l_i) \rightarrow \max_{u_i}, \quad (4.22)$$

$$J_i = c_i \alpha_i (x_i + u_i)^{k_i} l_i^{1-k_i} - \omega_i l_i \rightarrow \max_{l_i}, \quad i = \overline{1, N}, \quad (4.23)$$

при ограничении

$$\sum_{i=1}^N u_i = U. \quad (4.24)$$

Задача (4.22)–(4.24) представляет собой многокритериальную задачу оптимизации, которая в общем случае может не иметь решения [75]. Однако в данном случае благодаря принятой структурной схеме управления можно разделить эту задачу на совокупность $N + 1$ однокритериальных задач.

Будем считать, что центр некоторым образом (пока неоптимальным) осуществил распределение инвестиций между предприятиями. Тогда при фиксированных значениях u_i нетрудно найти решения N задач (4.23) из условия

$$\frac{\partial J_i}{\partial l_i} = 0, \quad i = \overline{1, N}.$$

Оптимальное число рабочей силы для i -го предприятия определяется соотношением

$$\bar{l}_i = \left[\frac{c_i \alpha_i}{\omega_i} (1 - k_i) \right]^{1/k_i} (x_i + u_i). \quad (4.25)$$

Отметим, что в решении (4.25) неизвестной величиной является оптимальный объем инвестиций, для определения которого центр должен решить следующую задачу оптимизации:

$$J_0(u_i) = \sum_{i=1}^N P_i(u_i, \bar{l}_i) \rightarrow \max_{u_i},$$

$$\sum_{i=1}^N u_i = U. \quad (4.26)$$

Подставив решение (4.25) в целевую функцию J_0 , получим

$$J_0 = \sum_{i=1}^N \alpha_i (x_i + u_i)^{k_i} \bar{l}_i^{1-k_i} = \sum_{i=1}^N \alpha_i (x_i + u_i)^{k_i} \left\{ \left[\frac{c_i \alpha_i}{\omega_i} (1 - k_i) \right]^{1/k_i} (x_i + u_i) \right\}^{1-k_i} =$$

$$= \sum_{i=1}^N \alpha_i (x_i + u_i) \left[\frac{c_i \alpha_i}{\omega_i} (1 - k_i) \right]^{\frac{1-k_i}{k_i}}. \quad (4.27)$$

Из (4.27) видно, что целевая функция центра $J_0(u_i)$ линейна относительно u_i . Тогда решение задачи (4.26) очевидно. Необходимо вычислить все коэффициенты перед u_i , $i = \overline{1, N}$, в формуле (4.27) и отдать весь инвестиционный фонд тому предприятию, у которого этот коэффициент будет больше.

Если целевая функция J_0 будет иметь более сложный вид или в качестве модели производства выбраны более сложные функции, такое простое решение задачи (4.26) получить не удастся. Тогда, используя решения для $\tilde{I}_i(u_i)$, можно найти оптимальные значения инвестиций \hat{u}_i , рассматривая данную систему как *имитационную*.

Имитируя действия центра, можно направлять различные объемы инвестиций в предприятия, вычисляя при этом значения \tilde{I}_i , P_i и J_0 по соответствующим формулам. То распределение инвестиций, которое даст наибольшее значение целевой функции J_0 , можно считать оптимальным. Конечно, при таком имитационном подходе можно перебрать только ограниченное число распределений инвестиций. Однако при наличии современной вычислительной техники это число может быть достаточно большим, особенно если необходимо удостовериться в правильности принятого решения. Более подробно имитационный подход будет рассмотрен в гл. 7.

Пример о баскетболисте (продолжение). Как было отмечено в гл. 2 при концептуальной постановке данной задачи, в качестве объекта исследования в общем случае должна рассматриваться система: игрок—мяч—шит—кольцо. Если целью моделирования является описание движения мяча в полете, то взаимодействием его с другими элементами системы можно пренебречь. Если же в качестве цели выступает оценка попадания мяча в кольцо, то учет взаимодействия необходим [126]. Это связано с тем, что мяч в кольцо может попасть не «чисто», а после отскока от щита и/или дужки кольца. При этом отскок последовательно может осуществляться несколько раз подряд до попадания мяча в кольцо (рис. 4.20). Поэтому необходимо моделировать все стадии полета мяча, к которым можно отнести следующие: полет мяча до кольца или щита (траектория 1 на рис. 4.20), взаимодействие мяча со щитом (2), отскок от щита (3), взаимодействие с дужкой кольца (4) и отскок от кольца (5).

Моделирование первой стадии полета мяча проведено в гл. 2, поэтому здесь оно рассматриваться не будет, хотя выходные параметры модели (координаты и скорость центра мяча в момент касания кольца или щита) будут использоваться в качестве входных данных для последующих моде-

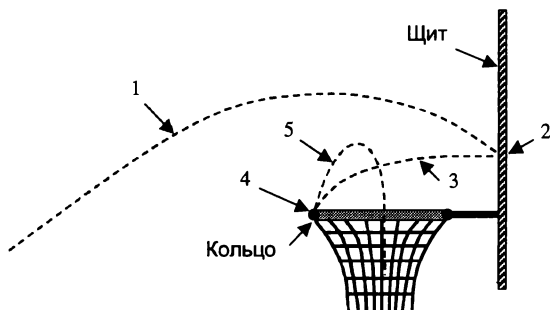


Рис. 4.20. Одна из возможных траекторий попадания мяча в кольцо

лей. Следует только отметить, что при моделировании полета мяча до кольца пренебрегалось возможным вращением мяча вокруг собственной оси. Это было обосновано пренебрежимо малым влиянием вращения мяча на траекторию его полета. Другая ситуация возникает при моделировании стадии взаимодействия мяча с кольцом или щитом. Направление и величина скорости вращения будут влиять на направление сил трения в период взаимодействия мяча с кольцом или щитом и, соответственно, на траекторию отскока и вероятность попадания мяча в кольцо. Кстати, этим пользуются опытные баскетболисты, закручивая мяч при броске в нужную сторону. А вот рекомендации по правильному выбору параметров закрутки мяча должны следовать из результатов моделирования.

Перейдем к рассмотрению этапа взаимодействия брошенного мяча с кольцом или щитом. Будем считать баскетбольный щит и дужку кольца абсолютно жесткими телами, влиянием формы мяча, дужки кольца и щита на взаимодействие пренебрежем, т.е. контакт считается точечным. Так как в данном случае модели взаимодействия мяча со щитом и кольцом одинаковы, то в дальнейшем будем рассматривать только модель взаимодействия мяча с кольцом. Мяч будем моделировать вязко-упругим телом. Упругие свойства его зависят от давления накачанного воздуха, а вязкие свойства — от материала и толщины покрышки мяча (рис. 4.21).

Положение центра масс мяча (точка C) в период взаимодействия задается координатой r и углом φ (здесь, как и в гл. 2, движение рассматривается в плоскости XOY). Начальное значение угла зависит от траектории полета мяча до кольца (пунктирная линия 1 на рис. 4.21).

Рассмотрим силы, действующие на мяч. Кроме силы тяжести mg , со стороны кольца на мяч действует сила контактного взаимодействия, проекция которой на нормаль n , согласно принятой структурно-механической модели мяча, может быть записана в следующем виде: $N = c(R_m - r) - v\dot{r}$, где c — коэффициент жесткости мяча; v — коэффициент демпфирования мяча; R_m — радиус недеформированного мяча; r — расстояние от центра мяча до точки контакта с поверхностью кольца.

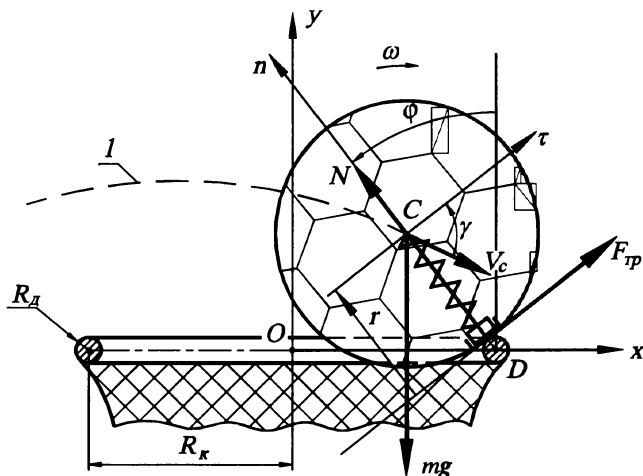


Рис. 4.21. Модель взаимодействия мяча с кольцом

В касательной плоскости действует сила сухого трения Кулона $F_{\text{тр}} = fN$, где f – коэффициент трения, зависящий от шероховатости покрышки мяча и дужки кольца. Направление силы трения зависит от скорости вращения ω и скорости центра мяча v_c .

Перейдем к математической постановке задачи взаимодействия мяча с кольцом. Так как все точки мяча при взаимодействии с кольцом движутся в плоскостях, параллельных плоскости моделирования XOY , то движение мяча можно считать плоским. В данном случае за время T удара о кольцо мяч поворачивается вокруг дужки кольца на некоторый угол $\Delta\phi$ и деформируется. Расстояние от центра мяча в процессе деформирования вначале уменьшается на Δr , а затем увеличивается до первоначального значения R_M . При этом угловая скорость вращения ω вокруг центра масс за счет сил трения изменяется на некоторую величину $\Delta\omega$. Вращательное движение мяча вокруг дужки кольца будем считать переносным, а поступательное движение вместе с центром масс при деформировании – относительным. Тогда систему отсчета πt , связанную с центром масс мяча, нельзя считать инерциальной. Используя теоремы о движении центра масс и об изменении кинетического момента, а также вводя переносную $F_e^{\text{ин}}$ и кориолисову $F_k^{\text{ин}}$ силы инерции, можно записать дифференциальные уравнения движения мяча:

$$\begin{cases} ma_C^{\text{от}} = N + F_{\text{тр}} + mg + F_e^{\text{ин}} + F_k^{\text{ин}}, \\ I_M \dot{\omega} = F_{\text{тр}} r, \end{cases}$$

где $a_C^{\sigma\tau}$ – относительное ускорение точки C ; I_M – момент инерции мяча относительно оси, проходящей через центр масс C . Переносная сила инерции направлена по нормали n и равна $F_e^{\text{ин}} = m(r + R_d)\dot{\varphi}^2$. Сила инерции Кориолиса направлена вдоль оси τ и равна $F_k^{\text{ин}} = 2m\dot{\varphi}\dot{r}$. Проецируя первое уравнение на оси n , τ и учитывая, что $a_{nC}^{\sigma\tau} = \ddot{r}$ и $a_{\tau C}^{\sigma\tau} = \varepsilon(r + R_d) = \ddot{\varphi}(r + R_d)$, запишем математическую постановку задачи в следующем виде.

Найти зависимости от времени $r(t)$, $\varphi(t)$, $\omega(t)$, $t \in [0, T]$, из решения следующей системы дифференциальных уравнений

$$\begin{cases} m\ddot{r} = N - mg \cos \varphi + m(r + R_d)\dot{\varphi}^2, \\ m(r + R_d)\ddot{\varphi} = mg \sin \varphi - F_{\text{тр}} - 2m\dot{\varphi}\dot{r}, \\ I_M \dot{\omega} = F_{\text{тр}} r \end{cases} \quad (4.28)$$

при начальных условиях

$$\begin{aligned} \varphi(0) &= \varphi_0, \quad \omega(0) = \omega_0, \quad r(0) = R_M, \\ \dot{\varphi}(0) &= -\frac{v_C \cos \gamma}{(R_d + R_M)}, \\ \dot{r}(0) &= -v_C \sin \gamma. \end{aligned} \quad (4.29)$$

Здесь R_d – радиус дужки кольца; v_C – скорость центра масс мяча в момент касания дужки.

Видно, что с математической точки зрения задача взаимодействия мяча с кольцом свелась к задаче Коши для системы ОДУ (4.28) с начальными условиями (4.29). Задача решалась с применением численного подхода – модифицированного метода Эйлера (см. разд. 3.5). Результаты получены при следующих исходных данных: радиус кольца $R_k = 0,233$ м; радиус дужки кольца $R_d = 0,008$ м; коэффициент трения $f = 0,5$; момент инерции мяча $I_M = 0,0055$ кг·м²; коэффициент вязкого деформирования мяча $\nu = 15$ Н·с/м; коэффициент жесткости мяча $c = 52820$ Н/м.

Результаты исследования попадания мяча в корзину удобно представить в виде диаграммы, приведенной на рис. 4.22. На диаграмме черные точки соответствуют попаданиям мяча в корзину, белые – нет.

Из приведенных результатов видно, что область попадания, в отличие от результата, полученного в гл. 2 без учета удара о щит и кольцо, распадается на две широкие С-образные подобласти и несколько узких подобластей подобной формы. Широкая подобласть Б включает «чистое» попадание мяча в корзину. Полоса Д характеризует подобное Б попадание

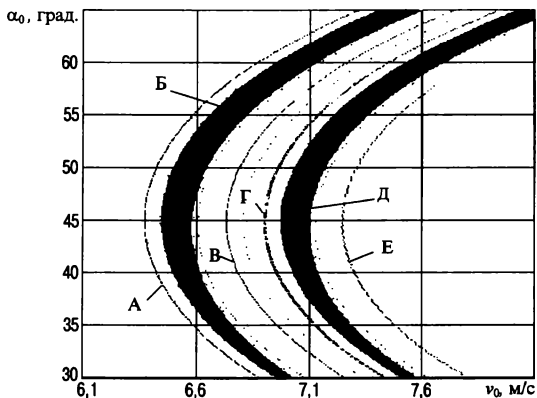


Рис. 4.22. Области попадания в зависимости от v_0 и α_0 при броске со штрафной линии

ние, но после отскока мяча от щита. Узкие подобласти свидетельствуют о разнообразных попаданиях после двух, трех и т.д. отскоков мяча от кольца и щита, показанных на рис. 4.23. Полоска А соответствует траектории, показанной на рис. 4.23,а, Б – 4.23,б, В – 4.23,в и т.д.

Подобласти Б и Д имеют наибольшую площадь и являются определяющими при анализе попаданий мяча в кольцо. Ширина Δv_0 этих подо-

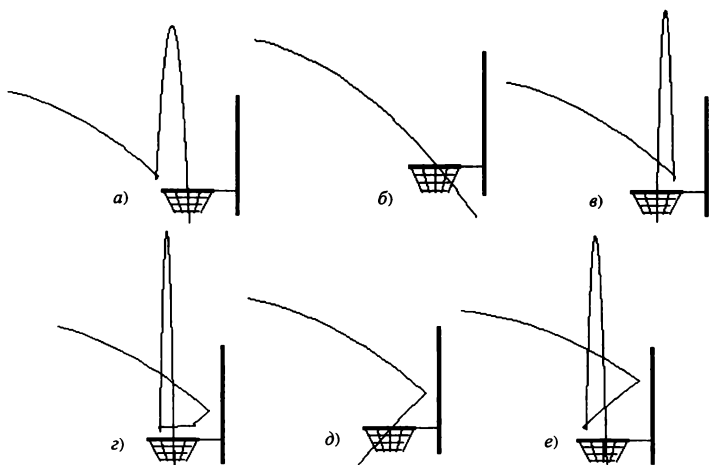


Рис. 4.23. Различные траектории мяча при попадании в корзину

бластей изменяется в зависимости от угла бросания α_0 , расстояния до корзины, высоты бросания $\Delta h = y_0 - y_k$, где y_k – высота центра корзины, y_0 – высота бросания.

На рис. 4.24 приведено изменение ширины подобластей Б и Д при броске со штрафной линии в зависимости от угла и высоты бросания. Как можно видеть, низкорослым игрокам следует использовать навесные траектории. Но даже при броске с уровня корзины угол броска не должен быть меньше 20° . Это значение меньше ограничения в $32,2^\circ$, полученного в гл. 2 из решения задачи «чистого» попадания без учета касания и отскока от дужки кольца. Оптимальный угол бросания находится в интервале 50° – 70° . В этом случае разброс по скорости броска составляет $0,1$ – $0,2$ м/с. При углах бросания более 70° попасть от шита практически невозможно (ширина подобласти Д быстро стремится к нулю, см. рис. 4.24, б), а для «чистого» попадания (рис. 4.24, а) при достаточно широкой подобласти Б и углах бросания более 70° начальная скорость должна быть более 15 м/с.

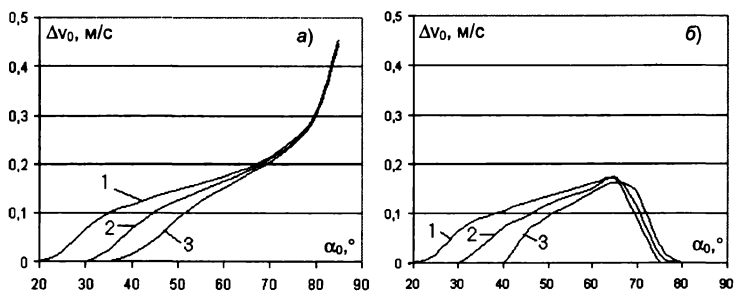


Рис. 4.24. Ширина подобластей Б (а) и Д (б) при различных значениях α_0 и высоты броска (1 – $\Delta h = 0$ м; 2 – $\Delta h = -0,5$ м; 3 – $\Delta h = -1$ м)

Что касается закрутки мяча при броске, то можно дать следующие рекомендации, вытекающие из решения задачи. В том случае, когда бросок производится в дальнюю дужку кольца, мяч следует закручивать в направлении, противоположном тому, которое показано на рис. 4.21. Тогда возникающая сила трения будет помогать «сваливаться» мячу в корзину, что увеличивает вероятность попадания. Однако влияние вращения мяча на попадание его в корзину наблюдается лишь в узком диапазоне начальных траекторий и скоростей мяча. Поэтому использовать такую технику бросания на практике довольно сложно.

? Вопросы для самопроверки

1. Что такое искусственная система? В каком случае совокупность отдельных элементов будет системой, а в каком – нет? Приведите примеры.
2. Когда на практике можно применять модель «черного ящика»? Какие основные недостатки имеет данная модель?
3. В чем сложность построения модели «белого ящика»? Какие виды неопределенностей вы знаете?
4. Что такое структурная схема системы? Чем отличается граф от сети?
5. В чем заключается основная особенность иерархических структурных схем? Где применяются иерархические структуры?
6. Почему синтез считают завершающим этапом анализа?
7. Что такое фрейм? Является он формальной или содержательной структурной моделью?
8. Сформулируйте условия окончания декомпозиции системы.
9. Что такое агрегат? Может ли агрегат обладать новыми свойствами по сравнению с теми, которыми обладают составляющие его элементы?
10. Какой агрегат называется оператором? Переменные и параметры оператора. Какие виды неопределенности оператора вы знаете?



Задания для самостоятельного выполнения

1. Постройте модели состава и структуры семьи при: а) матриархате; б) равноправии супругов.
2. Постройте структурную модель системы управления, которая реализована в вашем учебном заведении.
3. Постройте структурную модель спортивной команды (например, хоккейной или футбольной).
4. Постройте структурную модель автомобиля.
5. Постройте структурную модель солнечной системы.
6. Постройте агрегат-классификатор численных методов решения систем алгебраических уравнений.
7. Постройте одномерную структурную модель упрочняющегося упругопластического тела, для которого задана диаграмма σ - ϵ приведенного вида (σ_s , α , β считать известными). Постройте определяющие соотношения, соответствующие вашей структурной модели.
8. Для приведенной в примере 4.3 модели материала проведите расчеты для нагружения при различных траекториях нагружения (пропорциональное увеличение нагрузок, циклическое нагружение по пропорциональным траекториям и траекториям в виде окружностей).

Глава 5

МОДЕЛИРОВАНИЕ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

Нет ничего более противного разуму и природе, чем случайность.

Цицерон

Развитие современных технологий во всех сферах деятельности человека и разработка соответствующих моделей приводят к необходимости учета максимально возможного количества информации об исследуемом объекте. При этом вопросы моделирования сложных процессов и явлений чаще всего формулируются и обсуждаются на профессиональном языке (искусственном или подмножестве естественного), отражающем специфику области исследования. Следствием этого является использование в процессе моделирования качественных элементов: описаний, понятий и отношений с неопределенными или нечеткими границами, высказываний с многозначной шкалой истинности и т.д.

В тех случаях, когда информация существенно неопределенная, задание строгих границ «волевым» порядком или искусственное введение однозначности означает не что иное, как огрубление исходных данных, и может приводить к получению пусть четкого, но неверного результата. Исследование и учет однозначности (определенности) или неоднозначности (неопределенности) всех параметров и отношений, описывающих исследуемое явление, представляет собой необходимый и важнейший элемент математического моделирования.

Известные закономерности, описывающие процессы и явления объективного мира, можно условно разделить на две группы: од-

нозначно определенные (детерминированные) и находящиеся в условиях неопределенности.

К первой группе относят те закономерности, которые по заданным с определенной точностью характеристикам воздействий позволяют установить вполне определенный (детерминированный) отклик (реакцию) исследуемого объекта. Например, материальное тело падает с некоторой высоты. При заданной точности определения начальных условий и действующих на тело внешних сил можно однозначно, с определенной точностью установить его скорость при соприкосновении с Землей, время полета и т.д. С математической точки зрения эти закономерности описываются на основе аксиом традиционной математикой с использованием вполне определенных величин.

Вторая группа закономерностей описывает случайные события (такие, которые при заданном комплексе условий могут протекать по-разному при одних и тех же условиях). Например, при бросании игрального кубика нельзя заранее однозначно сказать, какая цифра выпадет. Если попытаться учесть природу этих закономерностей как явлений, находящихся в условиях неопределенности, то надо иметь в виду, что описание этой неопределенности может быть разным в зависимости от количества и качества имеющейся информации.

Часто граница, отделяющая случайное событие от неслучайного, очень размытая. Одна из концепций случайности (которая преобладала до начала XX столетия) состояла в том, что если при описании исследуемого объекта предусмотреть все связанные с ним «детали», то никакой случайности не будет. Однако в настоящее время принято придерживаться другой концепции. Вернемся к примеру с падением материального тела. При полете последнего необходимо учитывать температуру окружающей среды, скорость ветра, положение относительно поверхности Земли и другие факторы, которые имеют неоднозначный характер и могут, в свою очередь, влиять друг на друга. Поэтому в «чистом виде» однозначно определенных процессов (явлений), наверное, нет, т.е. при описании достаточно сложных процессов закономерности всегда носят стохастический характер.

5.1. ПРИЧИНЫ ПОЯВЛЕНИЯ НЕОПРЕДЕЛЕННОСТЕЙ И ИХ ВИДЫ

*...– Все неверно, – сказала Гусеница.
– Да, не совсем верно, – робко согласилась
Алиса, – некоторые слова не те.
– Все не так, от самого начала и до самого
конца, – строго проговорила Гусеница.*

Льюис Кэррол

При решении задач математического моделирования (задачи проектирования, описания различных технологических процессов, выбора оптимальных параметров и т.д.) уже на стадии концептуальной постановки необходимо задуматься над тем, насколько однозначно определены параметры, в терминах которых осуществляется математическое описание объекта моделирования. На этом этапе необходимо определить для каждого параметра, можно ли считать его однозначно определенным или ему присуща некоторая неопределенность. Причем неопределенными могут быть не только параметры, но и связи между ними. Неопределенность понимается в том смысле, что соответствующие характеристики рассматриваемой системы находятся в условиях приближения и неполноты информации. Эта неопределенность может быть связана, с одной стороны, с тем, что параметры могут изменяться случайным образом, а с другой стороны, с тем, что они могут адаптироваться (изменяться, устанавливаться) в процессе функционирования исследуемой системы.

К наиболее значимым причинам появления неопределенности можно отнести следующие:

- показатели системы практически всегда зависят от большого количества различных факторов, причем часть из них может быть даже неизвестна исследователю;
- при построении модели обычно ограничиваются отбором наиболее существенных (по мнению субъекта или в силу объективных обстоятельств) переменных, что, конечно, приводит к огрублению модели;
- математические погрешности, возникающие при линеаризации модели или использовании разложения в ряд при ограничении на число членов ряда; ошибки измерений и погрешности при проведении эксперимента и т.п.

В общем случае все причины возникновения неопределенности можно разбить на две основные группы: *субъективные* и *объек-*

тивные. Субъективные причины обусловлены некоторыми частными, нерегулярно повторяющимися явлениями, поэтому их достаточно сложно учесть при решении прикладных задач. Объективные причины чаще всего связаны с физическими особенностями исследуемого явления.

Например, если рассматривать задачу исследования некоторого технологического процесса, то к субъективным причинам можно отнести квалификацию работников, проводящих и регламентирующих исследуемый процесс, их навыки, реакцию, время адаптации и т.д. К объективным причинам появления неопределенности для такого типа задач можно отнести:

- физико-механические свойства поставляемых материалов (в частности, предел текучести, модуль Юнга, коэффициенты теплопроводности, теплоемкости, теплоотдачи и т.д.);
- анизотропию свойств;
- поля остаточных напряжений;
- геометрические характеристики заготовок (форма и размеры);
- характер износа инструмента и т.д.

В свою очередь каждая из указанных объективных причин появления неопределенности может быть обусловлена целым рядом предпосылок. Так, неоднородность свойств материала, с одной стороны, определяется как неоднородность по объему, обусловленная особенностями технологического процесса (отливки, прокат, армированные и порошковые композиты и прочие), а с другой — как неоднородность партий поставляемых заготовок. При решении прикладных задач для устранения неопределенностей обычно вводится предположение о принятии в качестве физико-механических характеристик некоторых предельных или средних значений (из возможных диапазонов). На наш взгляд, подобное предположение является весьма спорным в силу нелинейности исследуемых процессов и сложного характера взаимодействия отдельных частей объекта между собой. Поэтому возникает необходимость учета распределения соответствующей неопределенной величины.

В зависимости от полноты описания неопределенность можно разбить на три основные группы: неизвестность, недостоверность и неоднозначность [82]. Рассмотрим группы описания неопределенности более подробно (рис. 5.1).

Неизвестность — это начальная стадия описания неопределенности, при которой информация полностью отсутствует.

Недостоверность — это вторая стадия описания неопределенности, которая для различных стадий сбора информации может

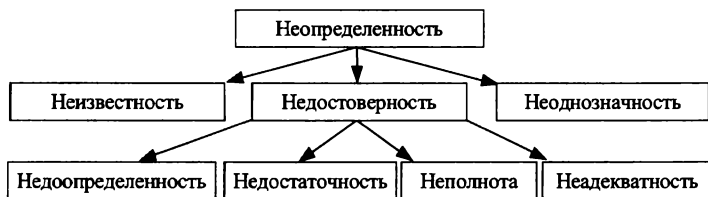


Рис. 5.1. Виды описания неопределенности

классифицироваться как неполнота, недостаточность, недоопределенность и неадекватность. *Неполнота* характеризуется тем, что собрана не вся возможная информация; *недостаточность* – собрана не вся необходимая информация. *Недоопределенность* – для некоторых элементов определены не их точные описания, а лишь множества, которым эти описания принадлежат [78]; *неадекватность* – ряд элементов исследуемого объекта описан по аналогии с уже имеющимися описаниями подобных элементов, т.е. имеет место так называемое «замещающее» описание, которое не всегда удовлетворяет целям исследования.

Дальнейший анализ неопределенности, учет новых факторов, определяющих исследуемое явление, может привести либо к устранению неопределенности (все элементы описаны однозначно), либо к неоднозначности.

Неоднозначность – это конечная (по полноте возможного описания) степень неопределенности, когда вся возможная информация собрана, но полностью необходимое описание не получилось.

Причины возникновения неоднозначности могут быть лингвистические и физические. (рис. 5.2).

Физическая неопределенность связана либо с наличием нескольких возможностей, каждая из которых *случайным* образом может стать реальностью, либо с *неточностью* вычислений или измерений. Таким образом, физическая неопределенность связана или с *физической сущностью* исследуемого явления, или с его *измеряемыми* проявлениями.

Лингвистическая неопределенность связана с использованием некоторого естественного языка. Она порождается, с одной стороны, множественностью значений слов (понятий и отношений) – *полисемией* (греч. polysēma – многозначность), с другой – *неоднозначностью смысла фраз*.

Можно выделить два вида полисемии: омонимию и нечеткость.

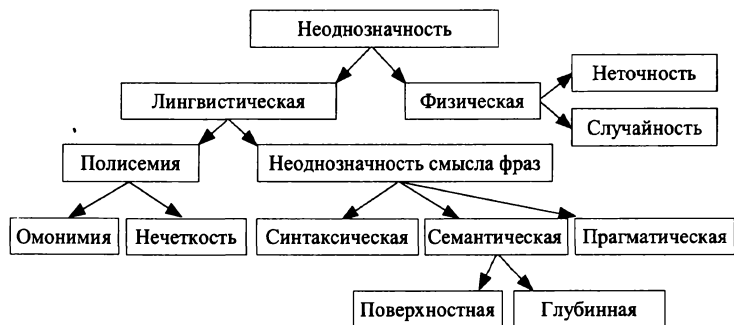


Рис. 5.2. Причины возникновения неоднозначности

Омонимия (греч. *homonymia* – одноименность) характеризуется тем, что одним и тем же словом можно характеризовать различные физические объекты. Например: *коса* – это вид побережья, инструмент или прическа. Если же объекты описания сходны по сути, но описывают некоторое множество понятий, то ситуацию относят к *нечеткости*. Например, понятие *несколько шагов*. Это может быть два шага, три шага, четыре шага и т.д.

Рассматривая источники неоднозначности смысла фраз, можно выделить синтаксическую, семантическую и прагматическую неоднозначность. При *синтаксической* неопределенности уточнение синтаксиса позволяет понять смысл фразы. Пример: «*казнить нельзя помиловать*» – «*казнить, нельзя помиловать*» или «*казнить нельзя, помиловать*». *Семантическая* неопределенность бывает *поверхностная* и *глубинная*. В первом случае отдельные слова понятны, но неясен смысл фразы («*голубые зеленые мысли яростно спят*»), во втором случае непонятны и все отдельные слова («*глокая куздра штеко будланула бокра и курдячит бокренка*»). *Прагматическая* неопределенность связана с совместным проявлением синтаксической и семантической неопределенностей. Она чаще всего проявляется при работе с незнакомыми объектами и, возможно, в незнакомой (например, языковой) среде [87].

Как уже отмечалось, на стадии концептуальной постановки задачи необходим детальный анализ степени неопределенности всех характеристик системы и связей между ними. При переходе к математической постановке задачи перед исследователем встают непростые вопросы: каким типом переменных описать те или иные параметры и как описать связь между этими параметрами?

Если цели исследования предполагают однозначное описание явления (процесса) или если параметры системы и связи между ними определены единственно возможным образом, то в этом случае применяется четкое описание, т.е. все характеристики считаются детерминированными и связи между соответствующими четкими переменными – однозначными. В противном случае в зависимости от целей исследования и требуемой полноты описания можно использовать различные математические подходы представления неопределенностей. Отметим, что усложнение модели (например, иерархичность) также может привести к необходимости использования других типов описания переменных, характеризующих исследуемое явление.

Математически неопределенность может быть описана стохастически, статистически, с позиций теории нечетких множеств, а также интервально (рис. 5.3). Отмеченные формы описания перечислены по возрастанию степени неопределенности. Рассмотрим физический смысл этих неопределенностей.

Стохастическое описание используется тогда, когда неопределенные параметры имеют вероятностный (случайный) характер.

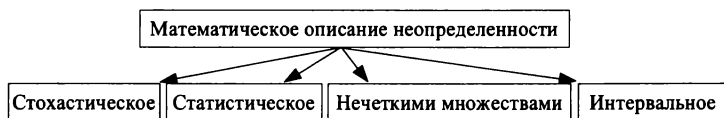


Рис. 5.3. *Формы описания неопределенности*

При этом необходимо, чтобы был определен закон распределения таких случайных параметров. Стохастическим описанием занимается теория вероятностей и теория случайных процессов [28, 49].

Статистическое описание является, по существу, частным случаем стохастического описания. Эту форму описания применяют, когда заданы только выборочные оценки каких-либо характеристик случайной величины или наборы значений некоторых случайных параметров. Статистическим описанием занимается математическая статистика [89].

При описании с позиций *нечетких множеств* неопределенный параметр задается некоторым множеством возможных его значений, характеризующихся той или иной степенью принадлежности (с помощью так называемой *функции принадлежности*) объекту, описываемому этим нечетким множеством [39,82]. Функция принадлежности может принимать значения от 1 (полная принадлеж-

ность) до 0 (полная непринадлежность). Интерпретацией функции принадлежности является *субъективная мера* того, насколько полно элемент (параметр) соответствует понятию, смысл которого описывается нечетким множеством. Этим описанием занимается теория нечетких множеств [39,82].

Интервальное описание можно использовать, когда неопределенные параметры заданы только диапазонами возможных значений (верхней и нижней границами), причем параметр может принимать любое значение внутри интервала и ему нельзя приписать никакой вероятностной меры. Интервальное описание является предметом исследования интервальной математики [4].

Следует отметить, что зависимость математического подхода к описанию переменных от полноты имеющейся информации весьма условна. Так, при численной реализации тех или иных алгоритмов моделей на ЭВМ даже для детерминированных переменных (неявным образом) используется аппарат интервальных вычислений, так как расчеты на ЭВМ ведутся с интервальными величинами. Поэтому считать, что интервальное описание переменных «менее определенное», чем стохастическое, наверное, нельзя. Однако с этим утверждением можно согласиться, если интервальное описание вводится уже на стадии концептуальной и математической постановок задач.

Многообразие форм описания неопределенностей приводит к различным особенностям постановки и решения соответствующих задач.

5.2. МОДЕЛИРОВАНИЕ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ, ОПИСЫВАЕМОЙ С ПОЗИЦИЙ ТЕОРИИ НЕЧЕТКИХ МНОЖЕСТВ

...Вряд ли можно считать, что мозг в сравнении с современными вычислительными машинами не имеет определенных преимуществ...

Главное из этих преимуществ, — по-видимому, способность мозга оперировать с нечетко очерченными понятиями.

Н.Винер

При решении сложных технических, экономических, технологических, социальных и других задач мы сталкиваемся с тем, что чем сложнее система, тем менее мы способны дать точные и в то

же время имеющие практическое значение суждения о ее поведении. Такая ситуация в [130] определяется термином «*принцип несовместимости*». Следствие из этого принципа кратко можно выразить так: «*Чем глубже мы анализируем реальную задачу, тем неопределеннее становится ее решение*». Именно в этом смысле точный количественный анализ поведения сложных систем для практического исследования реальных задач, по-видимому, недостаточен. В [24,39,56,73,82,130] предлагается подход, который опирается на предпосылку о том, что элементами исследования являются не числа, а некоторые нечеткие множества, для которых переход от «*принадлежности к классу*» к «*непринадлежности*» не скачкообразен, а непрерывен. В основе такого подхода лежит не традиционная двузначная или даже многозначная логика, а логика с нечеткой истинностью, нечеткими связями и нечеткими правилами вывода. Этот подход имеет три отличительные черты:

- 1) в нем используются так называемые «лингвистические» переменные вместо числовых переменных или в дополнение к ним;
- 2) простые отношения между переменными описываются с помощью нечетких высказываний;
- 3) сложные отношения описываются нечеткими алгоритмами.

Отметим, что с математической точки зрения предложенный подход как метод описания неопределенности лежит между описаниями с позиций теории вероятностей и математической статистики (в этом случае параметры системы, имеющие вероятностный, случайный характер, определяются некоторыми распределениями) и с позиций *интервальной математики*, при котором характеристики задаются диапазонами возможных значений (верхними и нижними границами).

Подобный тип задач чаще всего имеет место в том случае, когда концептуальная постановка задачи сформулирована в виде некоторого неопределенного высказывания типа «*Если A , то B* » ($A \Rightarrow B$), в котором A и B можно описать нечеткими множествами.

Прежде чем перейти к подробному обсуждению предлагаемого подхода, приведем некоторые основные положения.

Нечеткое множество — это математическая модель класса с нечеткими или, иначе говоря, размытыми границами.

В этом понятии учитывается возможность постепенного перехода от принадлежности к непринадлежности элемента множеству. Иными словами, элемент может иметь степень принадлежности множеству между полной принадлежностью (1) и полной непри-

надлежностью (0). Если степень принадлежности обозначить μ , то $\mu \in [0,1]$.

Введем некоторые основные понятия и определения [24,39,56,73,82,130]. С математической точки зрения нечеткое множество A можно определить следующим образом:

нечетким множеством A в U называется совокупность пар вида $(u, \mu_A(u))$, где $u \in U$, а $\mu_A(u)$ – функция принадлежности элементов нечеткого множества A , $\mu_A : U \rightarrow [0, 1]$; U – некоторое множество (в обычном смысле) элементов, которое называется универсальным множеством. (Понятие *множества* считается первичным и не определяется. По существу, *множество* – это совокупность элементов любого вида). Для каждого элемента $u \in U$ функция принадлежности определяет степень его принадлежности той совокупности элементов, которая формализуется данным нечетким множеством. Математически нечеткое множество определяется следующим образом:

$$A = \bigcup_{u \in U} \mu_A(u) / u.*$$

Например, пусть универсальное множество и совокупность функций принадлежности описываются выражениями: $U = (a, b, c, d, e, f)$ и $M = (0; 0,5; 1)$. В этом случае M – совокупность возможных значений функции принадлежности. При этом одна из возможных форм записи нечеткого множества A может быть представлена в следующем виде:

$$A = (0,0/a; 1,0/b; 0,5/c; 0,0/d; 0,5/e; 0,0/f).$$

Рассмотрим более подробно, что понимается под функцией принадлежности. Спектр мнений по этому вопросу чрезвычайно широк. Если под функцией принадлежности понимать «меру благоприятствия» соответствующих элементов тому понятию, которое формализуется данным нечетким множеством, то в этом случае функция принадлежности отождествляется с понятием вероятности (определение понятия вероятности дано ниже).

* Знак «/» в данном случае используется для отделения значения функции принадлежности от соответствующего элемента – носителя нечеткого множества.

В [130] предполагается, что функция принадлежности — это некоторое *невероятностное субъективное измерение неточности*, что она отлична от плотности вероятности и от функции распределения вероятности. Иногда под функцией принадлежности понимают возможность или полезность того или иного события.

В данной работе, так же как в [39], под значением функции принадлежности $\mu_A(x)$ нечеткого множества A для любого $x \in X$ будем понимать вероятность того, что *лицо, принимающее решение (ЛПР)*, отнесет элемент x к множеству A . В случае, когда A — некоторое понятие естественного языка, а x — элемент множества объектов, обозначаемых понятием A , $\mu_A(x)$ есть вероятность того, что ЛПР использует A в качестве имени объекта x . Необходимо отметить, что:

- приведенная интерпретация, которую будем называть вероятностной, не исключает других (в том числе невероятностных);
- элемент x , как следует из определения, уже предъявлен ЛПР, а ЛПР и решает задачу отнесения элемента к нечеткому множеству A .

В качестве конкретного примера применения аппарата теории нечетких множеств для математического моделирования некоторого явления рассмотрим следующую задачу.

Пример 5.1. Пусть справедливым считается следующее высказывание: «Если дорога скользкая — езда опасная, в противном случае — не опасная». Необходимо определить, в каком случае езда будет более опасной: если дорога *не очень скользкая* или дорога *очень не скользкая*?

Отметим, что в качестве примера выбрано простейшее утверждение, которое легко формализуется, ответ на него почти очевиден. В реальных задачах операций может быть десятки и сотни и ответы на поставленные вопросы не так тривиальны.

Для математической постановки представленной задачи необходимо ввести ряд определений и некоторые простейшие операции с нечеткими множествами, что и делается ниже.

Носителем нечеткого множества A ($Supp A$ или $S(A)$) называется множество (в обычном смысле), определяемое как

$$Supp A = \{u/u \in U, \mu_A(u) > 0\}.$$

Нечеткое отношение $R: X \rightarrow Y$ представляет бинарное отношение нечетких множеств X и Y ; R следующим образом описывается с помощью функции принадлежности двух переменных:

$$R = \bigcup_{(x,y) \in X \times Y} \mu_R(x,y)/(x,y).$$

где функция принадлежности двух переменных в зависимости от постановки задачи показывает предпочтение или сходство элементов первого и второго нечетких множеств.

Сравнивая определения нечеткого множества и нечеткого отношения, можно видеть, что последнее – это нечеткое множество с векторной базовой переменной. В зависимости от того, для чего используются бинарные нечеткие отношения, вводятся нечеткие отношения *сходства* и нечеткие отношения *предпочтения*.

Для примера рассмотрим бинарное отношение сходства. Предположим, что $X = \{\text{яблоко}, \text{груша}\}$, $Y = \{\text{айва}, \text{апельсин}\}$. Сходство будем оценивать по степени сладости зрелых фруктов (функция принадлежности в этом примере выбирается субъективно; в данном случае полагают, что по степени сладости груша и апельсин наиболее близки, а груша и айва – наименее близки друг другу).

Бинарное нечеткое отношение сходства между элементами множеств X и Y можно записать в виде:

$$\begin{aligned} \text{сходство} = \{ & 0,8/(\text{яблоко}, \text{айва}); & 0,6/(\text{яблоко}, \text{апельсин}); \\ & 0,2/(\text{груша}, \text{айва}); & 0,9/(\text{груша}, \text{апельсин}) \}. \end{aligned}$$

Для удобства записи в теории нечетких множеств нечеткие отношения обычно представляются в виде так называемой *матрицы отношений*. В данном случае имеем:

$$R = \begin{bmatrix} 0,8 & 0,6 \\ 0,2 & 0,9 \end{bmatrix},$$

где элемент R_{ij} равен значению функции $\mu_R(x, y)$ для i -го элемента X и j -го элемента Y .

В реальных задачах часто возникает такая ситуация. Между множествами $X \rightarrow Y$ и $Y \rightarrow Z$ имеются бинарные нечеткие отношения. Необходимо установить бинарное нечеткое отношение между множествами $X \rightarrow Z$. Эта ситуация требует введения операции произведения отношений. Пусть R – отношение $X \rightarrow Y$, а S – отношение $Y \rightarrow Z$, тогда отношение $X \rightarrow Z$ определит произведение $R \circ S$, которое в теории нечетких множеств определяется как *максиминное произведение* следующего вида:

$$R \circ S = \bigcup_{x,z \in (X,Z)} \max(\min(\mu_R(x,y), \mu_S(y,z)))/(x,z).$$

При выполнении обычного произведения матриц элемент матрицы-произведения, стоящий в i -й строке и k -м столбце, равен сумме произведений соответственных элементов i -й строки первой матрицы и k -го столбца второй матрицы. По существу, максиминное произведение определяется как обычное произведение матриц [35], где вместо операции умножения вводится \min , а вместо операции сложения — \max .

Пусть, например,

$$R = \begin{bmatrix} 0,3 & 0,8 \\ 0,6 & 0,9 \end{bmatrix}, \quad S = \begin{bmatrix} 0,5 & 0,9 \\ 0,4 & 1,0 \end{bmatrix},$$

тогда

$$\begin{aligned} R \circ S &= \begin{bmatrix} 0,3 & 0,8 \\ 0,6 & 0,9 \end{bmatrix} \circ \begin{bmatrix} 0,5 & 0,9 \\ 0,4 & 1,0 \end{bmatrix} = \\ &= \begin{bmatrix} \max(\min(0,3; 0,5); \min(0,8; 0,4)) & \max(\min(0,3; 0,9); \min(0,8; 1,0)) \\ \max(\min(0,6; 0,5); \min(0,9; 0,4)) & \max(\min(0,6; 0,9); \min(0,9; 1,0)) \end{bmatrix} = \\ &= \begin{bmatrix} 0,4 & 0,8 \\ 0,5 & 0,9 \end{bmatrix}. \end{aligned}$$

Нечеткой (лингвистической) переменной называется совокупность (кортеж) вида (X, U, \tilde{X}) , где X — наименование нечеткой переменной; $U = \{u\}$ — область ее определения (обычное множество);

$\tilde{X} = \bigcup_{u \in U} \mu_{\tilde{X}}(u)/u$ — нечеткое множество на U , описывающее числовые значения нечеткой переменной X .

Если обратить внимание на структуру наименования лингвистической переменной, то можно отметить, что в общем случае это составной термин, представляющий сочетание некоторых элементарных терминов. Эти элементарные термины можно разбить на четыре основных категории:

- *первичные* — символы специальных нечетких подмножеств, например молодой, старый и т.д.;
- *отрицание НЕ и союзы И, ИЛИ*;
- *неопределенности* типа: очень, слабо, более или менее и т.д.;
- *маркеры* чаще всего это вводные слова.

Пример. Пусть нечеткое множество описывается составным термином: «по мнению окружающих, это был не очень сильный и совсем не высокий человек». Выделим основные категории:

- первичные термины – сильный, высокий (человек);
- отрицание НЕ и союзы И;
- неопределенности – очень, совсем;
- маркеры – по мнению окружающих.

Отрицание НЕ, союзы И, ИЛИ, неопределенности типа *очень, весьма, больше, меньше* и др., которые входят в определение значений лингвистических переменных, могут рассматриваться как символы различных операций, определенных на нечетких подмножествах U . Рассмотрим наиболее существенные из них.

Пусть A и B – нечеткие множества; $S(A)$, $S(B)$ – их носители. Обычно вводятся два набора определений основных операций над нечеткими множествами: *максиминный* (mm) и *вероятностный* (p) [79,111]. Объединением нечетких множеств A и B в U называется нечеткое множество $A \cup B$ с функцией принадлежности вида:

$$\mu_{A \cup B}(u) = \begin{cases} \max(\mu_A(u), \mu_B(u)), & u \in U, & -(mm), \\ \mu_A(u) + \mu_B(u) - \mu_A(u) \cdot \mu_B(u), & u \in U, & -(p). \end{cases}$$

Объединение соответствует союзу ИЛИ. Таким образом, если X и Y – символы нечетких множеств, то $X \cup Y = (X \text{ или } Y)$.

Пример. Пусть $A = (0,2/1; 0,5/2; 1,0/3)$, $B = (0,8/1; 0,4/2; 0,5/3; 0,8/4)$. Тогда:

$$A \cup B = (0,8/1; 0,5/2; 1,0/3; 0,8/4) - (mm);$$

$$A \cup B = (0,84/1; 0,7/2; 1,0/3; 0,8/4) - (p).$$

Пересечением нечетких множеств A и B в U называется нечеткое множество $A \cap B$ с функцией принадлежности вида:

$$\mu_{A \cap B}(u) = \begin{cases} \min(\mu_A(u), \mu_B(u)), & u \in U, & -(mm), \\ \mu_A(u) \cdot \mu_B(u), & u \in U, & -(p). \end{cases}$$

Пересечение соответствует союзу И. Таким образом, $X \cap Y = (X \text{ и } Y)$.

Пример. Пусть $A = (0,2/1; 0,5/2; 1,0/3)$, $B = (0,8/1; 0,4/2; 0,5/3; 0,8/4)$. Тогда:

$$A \cap B = (0,2/1; 0,4/2; 0,5/3; 0,0/4) - (mm);$$

$$A \cap B = (0,16/1; 0,2/2; 0,5/3; 0,0/4) - (p).$$

Дополнением нечеткого множества A называется нечеткое множество \bar{A} с функцией принадлежности:

$$\mu_{\bar{A}}(u) = 1 - \mu_A(u), \quad u \in U, \quad -(mm, p).$$

Определение дополнения соответствует отрицанию НЕ, т.е.

$$\bar{X} = (\text{не } X) = \bigcup_U (1 - \mu_X(x)) / x.$$

Пример. Пусть $A = (0,2/1; 0,5/2; 1,0/3)$. Тогда $\bar{A} = (0,8/1; 0,5/2; 0,0/3)$.

Декартово произведение $A_1 \times A_2 \times \dots \times A_n$ нечетких множеств $A_i \in U_i$, $i = \overline{1, n}$, определяется как нечеткое множество A в декартовом произведении $U = U_1 \times U_2 \times \dots \times U_n$ с функцией принадлежности вида:

$$\mu_A(u) = \min \{ \mu_{A_1}(u_1), \dots, \mu_{A_n}(u_n) \}, \quad u = (u_1, \dots, u_n) \in U.$$

Обычным множеством α -уровня нечеткого множества A называется

$$S_\alpha = \{ u : u \in U, \mu_A(u) \geq \alpha \},$$

где $\alpha \in [0, 1]$.

Для определения арифметических операций $\otimes = \{ +, -, \cdot, / \}$ в [24] приведен так называемый *принцип обобщения* Л.А.Заде.

Пусть A и B – два нечетких множества. Тогда нечеткое множество $D = A \otimes B$ определяется функцией принадлежности

$$\mu_D(u) = \Theta[\mu_A(u), \mu_B(u)],$$

где

$$\Theta[\mu_A(u), \mu_B(u)] = \begin{cases} \sup_{\substack{a \otimes b = u \\ a \in S_A, b \in S_B}} \min(\mu_A(u), \mu_B(u)), & -(mm), \\ \sup_{\substack{a \otimes b = u \\ a \in S_A, b \in S_B}} (\mu_A(u) \cdot \mu_B(u)), & -(p). \end{cases}$$

Теперь арифметические операции $\otimes = \{ +, -, \cdot, / \}$ можно определить следующим образом: $A \otimes B = \bigcup_U \mu_D(u) / (a \otimes b)$, если операция $(a \otimes b)$ определена.

Пример операций над множествами. Пусть $A = (0,2/1; 0,5/2; 1,0/3)$, $B = (0,8/1; 0,4/2; 0,5/3; 0,8/4)$. Тогда

$$A + B = \sup_{\substack{a \otimes b = u \\ a \in S_A, b \in S_B}} (0,2/2; 0,2/3; 0,2/4; 0,2/5; 0,5/3; 0,4/4; 0,5/5; 0,5/6; 0,8/4; 0,4/5; 0,5/6; 0,8/7) = (0,2/2; 0,5/3; 0,8/4; 0,5/5; 0,5/6; 0,8/7) - (mm);$$

$$A + B = \sup_{\substack{a \otimes b = u \\ a \in S_A, b \in S_B}} (0,16/2; 0,08/3; 0,1/4; 0,16/5; 0,4/3; 0,2/4; 0,25/5; 0,4/6; 0,8/4; 0,4/5; 0,5/6; 0,8/7) = (0,16/2; 0,4/3; 0,8/4; 0,4/5; 0,5/6; 0,8/7) - (p).$$

Степень нечеткого множества A называется нечеткое множество A^α с функцией принадлежности

$$\mu_{A^\alpha}(u) = (\mu_A(u))^\alpha, \quad u \in U, \quad \alpha > 0.$$

При $\alpha = 2$ получаем операцию *концентрирования* (CON): $CON(A) = A^2$.

В результате применения этой операции к множеству A снижается степень нечеткости описания, причем для элементов с высокой степенью принадлежности это уменьшение относительно мало, а для элементов с малой степенью принадлежности – относительно велико.

При $\alpha = 0,5$ получаем операцию *растяжения* (DIL): $DIL(A) = A^{0,5}$. Эта операция увеличивает степень нечеткости исходного нечеткого множества.

Неопределенность удобно определить через некоторые основные операции (особенно операции *степень*, CON, DIL). Покажем, как это можно сделать для неопределенности *очень*. Аналогичным образом можно установить неопределенности: *больше*, *меньше*, *много*, *слабо*, *вроде*, *вполне* и др.

В обычном использовании неопределенность *очень* не имеет четко определенного значения и действует как усилитель, генерируя подмножества того множества, к которому она применяется. Аналогичным образом действует операция концентрирования, поэтому *очень* u , где u – некоторый термин, может быть определено как квадрат u , т.е.

$$\text{очень } u = u^2 = \bigcup_U \mu_U^2(u)/u.$$

Например, если $u = \text{маленький возраст} = (1,0/1; 0,8/2; 0,6/3; 0,4/4; 0,2/5)$, тогда *очень маленький возраст* $= (1,0/1; 0,64/2; 0,36/3; 0,16/4; 0,04/5)$. Рассматриваемый как оператор, *очень* может действовать сам на себя. Так, например, *очень очень* $u = (\text{очень } u)^2 = u^4$.

Заметим, что порядок следования элементарных терминов в составном термине существенно влияет на результат. Так, например:

$$u = \text{очень не точно} = (\overline{\text{точно}})^2 \text{ и } u = \text{не очень точно} = \overline{(\text{точно})^2}$$

– не одно и то же. С другой стороны, *не очень точно* может быть записано по-разному, хотя результат будет один и тот же:

$$u = \text{не очень точно} = \overline{\text{очень точно}} = \overline{(\text{точно})^2}.$$

Вернемся к нашему примеру о скользкой дороге.

Высказывание «*если A, то B, иначе C*» (где A, B, C – нечеткие подмножества, при этом A обязательно из U , а B и C могут быть определены как на U , так и на V в зависимости от формулировки задачи; напомним, что U и V – универсальные множества, на которых определены нечеткие множества A, B, C) в терминах декартова произведения можно определить следующим образом:

$$\text{если } A, \text{ тогда } B, \text{ иначе } C \stackrel{\text{def}}{=} A \times B \cup \bar{A} \times C.$$

По существу, мы получили некоторое неопределенное отношение $R (R: U \Rightarrow V)$. Далее (по условию задачи) мы хотим определить значения некоторых подмножеств из V , которые определяются заданными подмножествами из U . Для решения этой и подобного типа задач в [130] сформулировано *составное правило вывода*, которое имеет следующий вид.

Если R – неопределенное отношение $U \Rightarrow V$ и x – неопределенное подмножество U , тогда неопределенное подмножество $y \in V$, которое индуцируется подмножеством x , дается композицией x и R , т.е. $y = x \circ R$, где « \circ » – максиминное произведение.

В рассматриваемом примере 1 необходимо определить x_1 – «*дорога не очень скользкая*», x_2 – «*дорога очень не скользкая*», и, решив соответствующие задачи вывода, получим подмножества y_1 и y_2 .

Для сравнения полученных нечетких множеств строится некоторая четкая функция $H(y_1, y_2)$ от нечетких аргументов, которая

называется *индексом ранжирования*. Значения индекса для конкретной пары нечетких чисел (являющихся, по существу, нечеткими множествами) дает основание решить вопрос о том, какое из двух нечетких чисел больше (или – с какой степенью больше). Простейший индекс ранжирования имеет вид [24]:

$$H_{\frac{1}{2}}^1(A, B) = \sup_{a>b} \min(\mu_A(a), \mu_B(b)),$$

где \sup – обозначение точной верхней границы множества. При этом если $H_{\frac{1}{2}}^1(A, B) > H_{\frac{1}{2}}^1(B, A)$, то $A > B$.

Отметим, что приведенный индекс ранжирования в качестве наибольшего выбирает то нечеткое число, пик функции принадлежности которого соответствует большему значению носителя. Данный индекс ранжирования является детерминированным, так как для сравнения нечетких чисел используются однозначно определенные представители этих нечетких чисел. Недостатком детерминированных индексов ранжирования является то, что они не учитывают вид и форму функции принадлежности сравниваемых нечетких чисел. Чтобы избежать такого ограничения, в теории нечетких множеств введены так называемые интегральные индексы ранжирования. Рассмотрим один из них:

$$H_2(A, B) = H_+(A) - H_+(B), \quad H_+(A) = \sum_{\Delta\alpha} M(A^\alpha) \cdot \Delta\alpha,$$

где A^α – α -уровневое подмножество нечеткого множества A ;

$M(A^\alpha) = \frac{\inf_{a \in A^\alpha} a + \sup_{a \in A^\alpha} a}{2}$, где \inf – точная нижняя граница множества.

При этом если $H_2(A, B) \geq 0$, то $A \geq B$. Отметим, что интегральные индексы ранжирования дают более точный результат решения, чем детерминированные.

Пример. Пусть $A = (0,1/1; 0,2/2; 0,8/3; 0,4/4; 0,2/5)$; $B = (0,4/1; 0,5/2; 0,9/3; 0,5/4; 0,1/5)$. Тогда (рис. 5.4)

$$\begin{aligned} H_+(A) &= H_+(A)_1 + H_+(A)_2 + H_+(A)_3 + H_+(A)_4 = \\ &= \frac{1+5}{2} \cdot 0,1 + \frac{2+5}{2} \cdot 0,1 + \frac{2,33+4}{2} \cdot 0,2 + \frac{3+3}{2} \cdot 0,4 \approx 2,48; \end{aligned}$$

$$H_+(B) = \frac{1+5}{2}0,1 + \frac{2+4}{2}0,4 + \frac{3+3}{2}0,4 = 2,7;$$

$$H_2(A, B) = H_+(A) - H_+(B) = 2,48 - 2,7 = -0,22 < 0 \Rightarrow A < B.$$

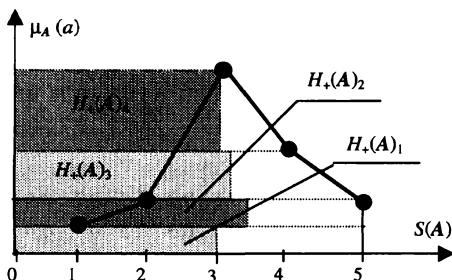


Рис. 5.4. Иллюстрация к определению индексов ранжирования

Продолжение примера 5.1.

Введем нечеткие множества скользкая дорога и опасная езда. Первое нечеткое множество можно определить через коэффициент трения скольжения. Если использовать коэффициенты 0,01; 0,02; 0,03; 0,04; 0,05; 0,06, то можно записать:

$$A = \text{скользкая дорога} = \\ = (0,8/0,01; 1,0/0,02; 0,6/0,03; 0,4/0,04; 0,2/0,05; 0,1/0,06).$$

Для определения степени опасности можно ввести коэффициенты опасности I–V. Например, I – это легкие ушибы, ссадины; II – переломы конечностей и т.д. При этом

$$B = \text{опасная езда} = (0,1/I; 0,3/II; 0,5/III; 0,7/IV; 0,9/V).$$

Теперь можно определить все необходимые для решения задачи нечеткие множества:

$$\text{Не скользкая дорога} = \bar{A} = (0,2/0,01; 0,0/0,02; 0,4/0,03; 0,6/0,04; 0,8/0,05; 0,9/0,06);$$

$$\text{Не опасная езда} = \bar{B} = (0,9/I; 0,7/II; 0,5/III; 0,3/IV; 0,1/V);$$

$$X_1 = \text{не очень скользкая дорога} =$$

$$= (\bar{A})^2 = (0,36/0,01; 0,0/0,02; 0,64/0,03; 0,84/0,04; 0,96/0,05; 0,99/0,06);$$

$X_2 = \text{очень не скользкая дорога} =$

$$= (\bar{A})^2 = (0,04/0,01; 0,0/0,02; 0,16/0,03; 0,36/0,04; 0,64/0,05; 0,81/0,06);$$

Определим нечеткое отношение $R: R = A \times B \cup \bar{A} \times \bar{B}$. Если

$$A \times B = \begin{bmatrix} 0,1 & 0,3 & 0,5 & 0,7 & 0,8 \\ 0,1 & 0,3 & 0,5 & 0,7 & 0,9 \\ 0,1 & 0,3 & 0,5 & 0,6 & 0,6 \\ 0,1 & 0,3 & 0,4 & 0,4 & 0,4 \\ 0,1 & 0,2 & 0,2 & 0,2 & 0,2 \\ 0,1 & 0,1 & 0,1 & 0,1 & 0,1 \end{bmatrix}; \quad \bar{A} \times \bar{B} = \begin{bmatrix} 0,2 & 0,2 & 0,2 & 0,2 & 0,1 \\ 0,0 & 0,0 & 0,0 & 0,0 & 0,0 \\ 0,4 & 0,4 & 0,4 & 0,3 & 0,1 \\ 0,6 & 0,6 & 0,5 & 0,3 & 0,1 \\ 0,8 & 0,7 & 0,5 & 0,3 & 0,1 \\ 0,9 & 0,7 & 0,5 & 0,3 & 0,1 \end{bmatrix},$$

то

$$R = A \times B \cup \bar{A} \times \bar{B} = \begin{bmatrix} 0,1 & 0,3 & 0,5 & 0,7 & 0,8 \\ 0,1 & 0,3 & 0,5 & 0,7 & 0,9 \\ 0,4 & 0,4 & 0,5 & 0,6 & 0,6 \\ 0,6 & 0,6 & 0,5 & 0,4 & 0,4 \\ 0,8 & 0,7 & 0,5 & 0,3 & 0,2 \\ 0,9 & 0,7 & 0,5 & 0,3 & 0,1 \end{bmatrix}.$$

Окончательно получим:

$$\begin{aligned} Y_1 &= X_1 \circ R = \\ &= \left(\frac{0,36}{0,01}; \frac{0,0}{0,02}; \frac{0,64}{0,03}; \frac{0,84}{0,04}; \frac{0,96}{0,05}; \frac{0,99}{0,06} \right) \circ \begin{bmatrix} 0,1 & 0,3 & 0,5 & 0,7 & 0,8 \\ 0,1 & 0,3 & 0,5 & 0,7 & 0,9 \\ 0,4 & 0,4 & 0,5 & 0,6 & 0,6 \\ 0,6 & 0,6 & 0,5 & 0,4 & 0,4 \\ 0,8 & 0,7 & 0,5 & 0,3 & 0,2 \\ 0,9 & 0,7 & 0,5 & 0,3 & 0,1 \end{bmatrix} = \\ &= (0,9/\text{I}; 0,7/\text{II}; 0,5/\text{III}; 0,6/\text{IV}; 0,6/\text{V}); \end{aligned}$$

$$\begin{aligned} Y_2 &= X_2 \circ R = \\ &= \left(\frac{0,04}{0,01}; \frac{0,0}{0,02}; \frac{0,16}{0,03}; \frac{0,36}{0,04}; \frac{0,64}{0,05}; \frac{0,81}{0,06} \right) \circ \begin{bmatrix} 0,1 & 0,3 & 0,5 & 0,7 & 0,8 \\ 0,1 & 0,3 & 0,5 & 0,7 & 0,9 \\ 0,4 & 0,4 & 0,5 & 0,6 & 0,6 \\ 0,6 & 0,6 & 0,5 & 0,4 & 0,4 \\ 0,8 & 0,7 & 0,5 & 0,3 & 0,2 \\ 0,9 & 0,7 & 0,5 & 0,3 & 0,1 \end{bmatrix} = \\ &= (0,81/\text{I}; 0,7/\text{II}; 0,5/\text{III}; 0,36/\text{IV}; 0,36/\text{V}). \end{aligned}$$

Сравним полученные результаты для Y_1 и Y_2 между собой, для чего воспользуемся индексом ранжирования $H_5^1(Y_1, Y_2)$:

$$H_5^1(Y_1, Y_2) = \sup (\min(0,9_I; 0,81_I), \min(0,7_{II}; 0,81_I; 0,7_{II}), \min(0,5_{III}; 0,81_I; 0,7_{II}; 0,5_{III}), \min(0,6_{IV}; 0,81_I; 0,7_{II}; 0,5_{III}; 0,36_{IV}), \min(0,6_V; 0,81_I; 0,7_{II}; 0,5_{III}; 0,36_{IV}; 0,36_V)) = \sup (0,81; 0,7; 0,5; 0,36; 0,36) = 0,81;$$

$$H_5^1(Y_2, Y_1) = \sup (\min(0,81_I; 0,9_I), \min(0,7_{II}; 0,9_I; 0,7_{II}), \min(0,5_{III}; 0,9_I; 0,7_{II}; 0,5_{III}), \min(0,36_{IV}; 0,9_I; 0,7_{II}; 0,5_{III}; 0,6_{IV}), \min(0,36_V; 0,9_I; 0,7_{II}; 0,5_{III}; 0,6_{IV}; 0,6_V)) = \sup (0,81; 0,7; 0,5; 0,36; 0,36) = 0,81.$$

Использование детерминированного индекса ранжирования не позволило сделать вывод о том, какая дорога более опасная. Уточним решения, используя для сравнения интегральный индекс ранжирования. При этом для оценки степеней опасности введем следующие числовые значения: I – 1, II – 2, III – 3, IV – 4, V – 5 (это необходимо сделать, чтобы как-то численно оценивать элементы носителя. Можно было бы пойти по другому пути – «опасность» езды оценивать по степени ущерба для здоровья человека, например, I – 5%, II – 10% и т.д.).

$$H_+(Y_1) = \frac{1+5}{2} \cdot 0,6 + \frac{1+2}{2} \cdot 0,2 + \frac{1+1}{2} \cdot 0,2 = 2,3.$$

Таким образом, *не очень скользкая дорога* более опасна, чем *очень не скользкая дорога*. Отметим, что интегральный индекс ранжирования дает более точный результат по сравнению с детерминированным индексом, так как учитывает весь спектр распределений нечетких множеств. Приведенный пример является достаточно простым, имеет тривиальный ответ и служит только иллюстрацией применения теории нечетких множеств. Примеры использования этой теории для разработки математических моделей сложных процессов и явлений читатель может найти в работах [56, 78].

Как уже отмечалось, подход с позиций теории нечетких множеств к математическому моделированию явлений или процессов наиболее предпочтителен в тех областях знаний, в которых трудно или даже невозможно получить точные количественные оценки. К таким разделам можно отнести экологию, метеорологию, педагогику, психологию, социологию и т.п.

Рассмотрим еще один пример решения задачи с позиций теории нечетких множеств в метеорологии.

Пример 5.2. Многолетние погодные наблюдения показали, что если атмосферное давление высокое, то на следующий день погода будет солнечной, в противном случае – пасмурной. Необходимо определить какая будет погода, если:

➤ атмосферное давление не очень высокое;

➤ атмосферное давление очень не высокое.

Оценить, в каком случае будет «более солнечно».

Введем нечеткие множества «высокое давление», «солнечная погода» и «пасмурная погода».

Пусть $A = \text{высокое давление} = (0,2/740; 0,4/750; 0,6/760; 0,8/770; 1/780)$ (элементы носителя числового нечеткого множества A измеряются в миллиметрах ртутного столба).

Для определения погоды введем нечисловые нечеткие множества с элементами носителя: «пасмурно», «пасмурно с прояснением», «переменная облачность», «ясно» и «солнечно». Для сокращения записей соответствующие элементы носителя обозначим: I–V. Например I – пасмурно, II – пасмурно с прояснением и т.д. При этом:

$B = \text{солнечная погода} = (0,2/\text{I}; 0,3/\text{II}; 0,5/\text{III}; 0,7/\text{IV}; 1/\text{V})$,

$C = \text{пасмурная погода} = (1/\text{I}; 0,8/\text{II}; 0,4/\text{III}; 0,2/\text{IV}; 0,1/\text{V})$.

Теперь можно определить все необходимые для решения задачи нечеткие множества:

не высокое давление $= \bar{A} = (0,8/740; 0,6/750; 0,4/760; 0,2/770; 0/780)$;

$X_1 = \text{не очень высокое давление} =$

$= (\bar{A})^2 = (0,96/740; 0,84/750; 0,64/760; 0,36/770; 0/780)$;

$X_2 = \text{очень не высокое давление} =$

$= (\bar{A})^2 = (0,64/740; 0,36/750; 0,16/760; 0,04/770; 0/780)$;

Определим нечеткое отношение $R: R = A \times B \cup \bar{A} \times \bar{C}$.

$$R = A \times B \cup \bar{A} \times \bar{C} = \begin{bmatrix} 0,8 & 0,8 & 0,4 & 0,2 & 0,2 \\ 0,6 & 0,6 & 0,4 & 0,4 & 0,4 \\ 0,4 & 0,4 & 0,5 & 0,6 & 0,6 \\ 0,2 & 0,3 & 0,5 & 0,7 & 0,8 \\ 0,2 & 0,3 & 0,5 & 0,7 & 1 \end{bmatrix}.$$

Окончательно получим:

$Y_1 = X_1 \circ R = (0,8/\text{I}; 0,8/\text{II}; 0,5/\text{III}; 0,6/\text{IV}; 0,6/\text{V})$;

$Y_2 = X_2 \circ R = (0,64/\text{I}; 0,64/\text{II}; 0,4/\text{III}; 0,36/\text{IV}; 0,36/\text{V})$.

Сравним полученные результаты для Y_1 и Y_2 между собой, для чего воспользуемся индексом ранжирования $H_3^1(Y_1, Y_2)$:

$$H_3^1(Y_1, Y_2) = 0,64; \quad H_3^1(Y_2, Y_1) = 0,64.$$

Применение детерминированного индекса ранжирования не позволило сделать вывод о том, в каком случае будет «более солнечно». Уточним решение, используя для сравнения интегральный индекс ранжирования. При этом можно положить, что I – соответствует не более 5% дневного времени, когда солнце не затянато тучами, II – 15%, III – 40%, IV – 80%, V – 100% (это необходимо сделать, чтобы численно оценивать элементы носителя): $H_+(Y_1) = 33,3$; $H_+(Y_2) = 24,6$.

Таким образом, если давление не очень высокое, то на следующий день будет «более солнечно», чем в том случае, когда давление – очень не высокое.

Пример 5.3. Для задачи о баскетболисте, постановка которой рассмотрена в гл. 2, подход с позиций теории нечетких множеств может помочь в обосновании используемых приемов бросания мяча, уточнении методики подготовки баскетболиста и т. д. Рассмотрим одну из возможных задач – вопрос об оценке точности попадания мяча в кольцо при броске со штрафной линии.

Пусть известно, что если начальная скорость v_0 мяча будет *средняя*, то точность попадания будет *высокой*, в противном случае – *не высокой*. Оценим, в каком случае точность попадания будет выше: когда начальная скорость *больше* средней или когда начальная скорость *меньше* средней.

Для решения этой задачи необходимо ввести неопределенности типа *больше* (средней) и *меньше* (средней).

Функцию принадлежности неопределенности *больше* (средней) определим следующим образом:

$$\mu_A^*(u) = \begin{cases} 0,2, & u < u^*, \\ 1 - \mu_A(u), & u \geq u^*, \end{cases}$$

где u^* – элемент носителя нечеткого множества с максимальной степенью принадлежности. Введением такого вида функции принадлежности мы увеличиваем возможность выбора больших значений носителя. Теперь функцию принадлежности неопределенности *меньше* (средней) можно найти так:

$$\mu_A^*(u) = \begin{cases} 1 - \mu_A(u), & u < u^*, \\ 0,2, & u \geq u^*. \end{cases}$$

Такой вид функции принадлежности увеличивает возможность выбора меньших значений носителя.

Введем нечеткие множества *средняя скорость* и *высокая точность*. Пусть

$$A = \text{средняя скорость} = (0,1/5,5; 0,4/6; 0,8/6,5; 0,6/7; 0,3/7,5)$$

(элементы носителя числового нечеткого множества A измеряются в м/с);

$$B = \text{высокая точность} = (0,2/20; 0,3/40; 0,5/60; 0,7/80; 1/100)$$

(элементы носителя числового нечеткого множества B измеряются в %).

Теперь можно определить все необходимые для решения задачи нечеткие множества:

$$\bar{A} = (0,9/5,5; 0,6/6; 0,2/6,5; 0,4/7; 0,7/7,5);$$

$$\bar{B} = (0,8/20; 0,7/40; 0,5/60; 0,3/80; 0/100);$$

$$X_1 = \text{скорость больше средней} = (0,2/5,5; 0,2/6; 0,2/6,5; 0,4/7; 0,7/7,5);$$

$$X_2 = \text{скорость меньше средней} = (0,9/5,5; 0,6/6; 0,2/6,5; 0,2/7; 0,2/7,5).$$

Определим нечеткое отношение $R: R = A \times B \cup \bar{A} \times \bar{B}$. Имеем

$$R = \begin{bmatrix} 0,8 & 0,7 & 0,5 & 0,3 & 0,1 \\ 0,6 & 0,6 & 0,5 & 0,4 & 0,4 \\ 0,2 & 0,3 & 0,5 & 0,7 & 0,8 \\ 0,4 & 0,4 & 0,5 & 0,6 & 0,6 \\ 0,7 & 0,7 & 0,5 & 0,3 & 0,3 \end{bmatrix}.$$

Окончательно получаем:

$$Y_1 = X_1 \circ R = (0,7/20; 0,7/40; 0,5/60; 0,4/80; 0,4/100);$$

$$Y_2 = X_2 \circ R = (0,8/20; 0,7/40; 0,5/60; 0,4/80; 0,4/100).$$

Сравним результаты для Y_1 и Y_2 между собой, для чего воспользуемся индексом ранжирования $H_3^1(Y_1, Y_2)$:

$$H_3^1(Y_1, Y_2) = 0,7; \quad H_3^1(Y_2, Y_1) = 0,7.$$

Использование детерминированного индекса ранжирования не позволило сделать однозначный вывод. Уточним решения, применив для сравнения интегральный индекс ранжирования:

$$H_+(Y_1) = 36; \quad H_+(Y_2) = 38.$$

Таким образом, точность попадания будет выше, если начальная скорость *меньше* средней по сравнению с броском с начальной скоростью *больше* средней.

5.3. МОДЕЛИРОВАНИЕ В УСЛОВИЯХ СТОХАСТИЧЕСКОЙ НЕОПРЕДЕЛЕННОСТИ

...Истинной логикой для этого мира является исчисление вероятностей, занимающееся нахождением величин вероятностей, которые учитывает или должен учитывать любой здравомыслящий человек.

Дж. Кларк Максвелл

При математическом описании модели любого сколько-нибудь сложного процесса или явления всегда возникает задача учета случайности. Сейчас уже трудно установить, кто первый поставил вопрос, пусть и в несовершенной форме, о количественном измерении возможности появления случайного события. В течение длительного периода исследователи ограничивались рассмотрением разного рода игр, особенно игр в кости, поскольку их изучение позволяет ограничиваться простыми и «прозрачными» математическими моделями. Однако следует заметить, что многие отлично понимали то, что позднее, в 1657 г., было прекрасно сформулировано Х.Гюйгенсом: «... Я полагаю, что при внимательном изучении предмета читатель заметит, что имеет дело не только с игрой, но здесь закладываются основы очень интересной и глубокой теории». И эта теория, изучающая количественные закономерности случайных явлений в однородных массовых процессах, была создана и названа теорией вероятностей.

За последние десятилетия неизмеримо выросла роль, которую играет теория вероятностей в современном естествознании. После того, как молекулярные представления о строении вещества получили всеобщее признание, стало неизбежным широкое использование теории вероятностей в физике и химии. Заметим, что с точки зрения молекулярной физики каждое вещество состоит из огромного числа малых частиц, находящихся в непрерывном движении и в процессе этого движения взаимодействующих друг с другом. Естественно, что при таких условиях обычные для физических теорий методы математических исследований становятся бессмысленными.

Теория вероятностей развилась из потребностей практики: в абстрактной форме она отражает закономерности, присущие случайным событиям массового характера. Эти закономерности играют исключительно важную роль в физике, химии, различных технических дисциплинах, экономике, военном деле и т.д.

До недавнего времени теория вероятностей представляла собой еще не сложившуюся математическую науку, в которой основные понятия были недостаточно корректно определены. Сказанное относится к так называемой классической теории вероятностей. В 1900 г. на II Международном математическом конгрессе в Париже немецкий ученый Д.Гильберт сделал доклад, в котором указал 23 важнейшие проблемы, стоящие перед математикой. Одна из них — аксиоматическое построение теории вероятностей. В настоящее время эта проблема успешно решена. Наиболее удачный подход к построению аксиоматической теории вероятностей предложил А.Н. Колмогоров. Этот подход тесно связывает теорию вероятностей с современной теорией меры, а также теорией множеств.

Иллюстрацию моделирования в условиях стохастической неопределенности проведем на примере задачи о баскетболисте, постановка которой рассмотрена в гл. 2. Там же было приведено аналитическое решение для дальности (2.11) и точности (2.12) броска баскетболиста в зависимости от начальной скорости v_0 мяча, угла бросания α_0 и расстояния x_k до кольца в момент броска.

В реальной ситуации баскетболист не может выполнить хотя бы два полностью одинаковых броска, т.е. с одинаковыми начальными параметрами v_0 и α_0 . Это связано с субъективными и объективными причинами. Как-то: дрогнула рука, изменилось положение тела, произошел вдох или выдох, сменилось эмоциональное состояние и т.п. (субъективные причины); изменилось направление движения воздуха, температура окружающей среды и т.д. (объективные причины). Поэтому для более полного отражения реальной действительности при моделировании необходимо учитывать весь спектр возможных разбросов параметров. Если выполняется серия бросков одним мячом, то неопределенными будут только начальные параметры бросания. Данные величины могут быть представлены как интервальные, статистические, стохастические параметры или характеристики, описываемые нечеткими множествами. С наименьшей степенью неопределенности влияние изменения этих параметров можно учесть с позиций стохастической неопределенности. При этом необходимо знать законы распре-

деления соответствующих случайных величин. Введем некоторые основные определения.

Опыт – это осуществление какого-либо комплекса условий, который может быть воспроизведен сколь угодно большое число раз.

Под *событием* будем понимать результат опыта или наблюдения. События могут быть *элементарными* (неразложимыми) и *составными* (разложимыми).

Элементарное событие – это событие, которое происходит в результате единичного опыта. *Составное событие* – это совокупность элементарных событий.

Пример. Игральный кубик подбрасывается 2 раза. Пусть составное событие определено следующим образом: «сумма выпавших цифр равна 6». Тогда элементарными будут события: «5+1», «4+2», «3+3», «2+4» и «1+5». Любые другие сочетания не относятся к рассматриваемому составному событию.

Генеральной совокупностью называют совокупность событий, которые могут быть реализованы в результате бесконечного числа однотипных опытов. *Выборочной совокупностью* или просто *выборкой* называют совокупность случайно отобранных событий из генеральной совокупности.

Объемом совокупности называют число событий этой совокупности.

Случайной величиной будем называть величину, которая в результате опыта (наблюдения, испытания) может принять одно из возможных значений, но какое именно – неизвестно.

Вероятность некоторого события – это мера его «благоприятствия».

События будем называть *равновозможными*, если мера их «благоприятствия» одинакова. В реальных условиях, когда число опытов *конечно*, мера «благоприятствия» определяется не вероятностью, а *частотой*. Пусть событие A наблюдалось в m опытах из n опытов (испытаний). Тогда частота события A ($W(A)$) определяется формулой

$$W(A) = m/n. \quad (5.1)$$

Если n достаточно велико, то «работает» одна из предельных теорем (закон больших чисел – теорема Бернулли) [28], и может быть записано приближенное равенство:

$$p(A) \approx W(A), \quad (5.2)$$

где $p(A)$ – вероятность события A .

Отметим, что по принятому в теории вероятности соглашению вероятность произвольного события полагается изменяющейся от 0 до 1. При этом нулевая вероятность соответствует *невозможному событию* (которое никогда произойти не может), а единичная — *достоверному событию* (которое обязательно произойдет).

Изложенный в данном разделе подход к описанию стохастической неопределенности базируется на классической теории вероятностей. В настоящее время общепринятым является аксиоматическое построение теории вероятностей, базирующееся на аксиомах А.Н.Колмогорова [14]. Сформулируем основные положения этого подхода (для удобства чтения на примере стохастического эксперимента).

Пусть Ω — пространство элементарных событий любой природы (по существу, Ω — это достоверное событие, а \emptyset (пустое множество) — это невозможное событие). Предположим, что в Ω выделена система подмножеств U , удовлетворяющая следующим условиям:

A1. $\Omega \in U$;

A2. Если $A \in U$, то $\bar{A} = \Omega \setminus A \in U$;

A3. Из того, что $A_i \in U$, $i = 1, 2, \dots, k$, следует, что $\bigcup_{i=1}^k A_i \in U$ и $\bigcap_{i=1}^k A_i \in U$.

Система подмножеств U , удовлетворяющая (A1–A3), называется алгеброй событий. Алгебру U называют также *кольцом*, поскольку в U определены только две операции (сложение и умножение), не выводящие из U . Алгебра U является *кольцом с единицей*, так как $\Omega \in U$ и $A \cap \Omega = \Omega \cap A = A$ для любого $A \in U$.

Добавим к A1–A3 еще одну аксиому:

A4. Из того, что $A_i \in U$, $i = 1, 2, \dots$, следует, что $\bigcup_{i=1}^{\infty} A_i \in U$ и $\bigcap_{i=1}^{\infty} A_i \in U$,

тогда система подмножеств U , удовлетворяющая (A1–A4), называется σ -алгеброй событий.

Таким образом, алгебра есть класс множеств, замкнутый относительно конечного числа операций дополнения, объединения и пересечения; σ -алгебра есть класс множеств, замкнутый относительно счетного числа этих операций.

Если задано множество Ω и какая-нибудь алгебра или σ -алгебра U его подмножеств, то говорят, что задано *измеримое пространство* (Ω, U) .

Осталось ввести понятие вероятности события.

Вероятность на (Ω, U) есть числовая функция, определенная на множествах из U и обладающая следующими свойствами:

P1. $p(A) \geq 0$ для любого $A \in U$;

P2. $p(\Omega)=1$;

P3. Аксиома конечной аддитивности. Если последовательность событий $A_i \in U$, $i = 1, 2, \dots, k$, такова, что $A_i \cap A_j = \emptyset$ при $i \neq j$, $\bigcup_{i=1}^k A_i \in U$, то
$$p\left(\bigcup_{i=1}^k A_i\right) = \sum_{i=1}^k p(A_i);$$

P4. Аксиома счетной аддитивности. Если последовательность событий $A_i \in U$, $i = 1, 2, \dots, k$, такова, что $A_i \cap A_j = \emptyset$ при $i \neq j$, $\bigcup_{i=1}^k A_i \in U$, то
$$p\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} p(A_i).$$

Тройка (Ω, U, p) называется вероятностным пространством.

Отметим, что аксиоматика Колмогорова является непротиворечивой, так как реализуется на объектах, где справедлива классическая теория вероятностей, и не полной, так как не указывает способа, как приписывать событиям вероятность. Однако это скорее не недостаток, а достоинство. Дискуссии о том, что следует понимать под вероятностью, связаны с желанием увязать определение вероятности с ее «физической» природой. Ввиду сложности последней такие попытки всегда наталкивались на существенные трудности не только математического, но и философского характера.

Появление аксиоматики Колмогорова как бы отделяет математическую сторону вопроса от всего остального. При таком подходе физическая трактовка понятия вероятности появляется уже в виде теоремы (усиленный закон больших чисел), в силу которой частота появления некоторого события при неограниченном повторении независимых испытаний сближается (в строго определенном смысле) с вероятностью этого события.

Таким образом, можно сказать, что при аксиоматическом подходе построение вероятностного пространства (Ω, U, p) является основным этапом в создании математической модели того или иного эксперимента. При этом произвол хотя и существует, но он по существу всегда определяется постановкой конкретных задач.

Рассмотренная выше аксиоматика Колмогорова оказалась исключительно плодотворной для развития новых разделов теории вероятностей, в частности – теории случайных процессов, примеры которых будут приведены в следующем параграфе.

Полностью охарактеризовать случайную величину можно законом распределения. Необходимо различать случайные величины (СВ), принимающие лишь отдельные, изолированные возможные значения (дискретные СВ), и случайные величины, возможные значения которых сплошным образом заполняют некоторый промежуток (непрерывные СВ).

Законом распределения случайной величины называют любое правило (таблицу, функцию), позволяющее находить вероятности всевозможных событий, связанных со случайной величиной (например, вероятность того, что она примет такое-то значение или попадет на такой-то интервал). Если СВ X имеет данный закон распределения, то про нее говорят, что она «распределена» по этому закону (или «подчинена» этому закону распределения).

Дискретной называют случайную величину, которая принимает отдельные, изолированные возможные значения с определенными вероятностями. Число возможных значений дискретной случайной величины может быть конечным или счетным.

Непрерывной называют случайную величину, которая может принимать все значения из некоторого конечного или бесконечного промежутка. Очевидно, что число возможных значений непрерывной СВ бесконечно (несчетно).

Рассмотрим некоторые простейшие характеристики случайной величины. К таким характеристикам относятся математическое ожидание, дисперсия, мода и медиана случайной величины.

Математическим ожиданием дискретной СВ называют сумму произведений всех ее возможных значений на их вероятности.

Математическое ожидание является средневзвешенным значением СВ, в которое каждое значение входит с «весом», равным соответствующей вероятности. Таким образом, математическое ожидание представляет собой центр тяжести системы материальных точек, координаты которых суть всевозможные значения СВ, а «массы» равны вероятностям этих значений.

Пусть СВ X принимает только значения x_1, x_2, \dots, x_n , вероятности которых соответственно равны p_1, p_2, \dots, p_n . Тогда математическое ожидание $M(X)$ определится формулой

$$M(X) = \sum_{i=1}^n x_i \cdot p_i. \quad (5.3)$$

Пример 5.4. Найти математическое ожидание случайной величины, заданной таблицей значений:

X	3	5	2
p	0,1	0,6	0,3

Решение: $M(X) = 3 \cdot 0,1 + 5 \cdot 0,6 + 2 \cdot 0,3 = 3,9$.

Математическое ожидание – важная, но не единственная характеристика усредненного значения СВ. Усредненное значение может характеризоваться также модой и медианой.

Модой случайной величины называется ее наиболее вероятное значение (то, для которого вероятность достигает максимума).

Медианой случайной величины X называется такое ее значение x , для которого $p(X < x) \approx p(X > x)$.

Пример 5.5. Найти математическое ожидание, моду и медиану СВ, заданной таблицей значений:

X	0	1	2	3	4
p	0,2	0,3	0,4	0,05	0,05

Решение

Математическое ожидание:

$$M(X) = 0 \cdot 0,2 + 1 \cdot 0,3 + 2 \cdot 0,4 + 3 \cdot 0,05 + 4 \cdot 0,05 = 1,45.$$

Мода: $M_x(X) = 2$. Медиана: $x = 1$.

Математическое ожидание, мода и медиана СВ по разному характеризуют ее усредненное значение. Эти различия становятся понятными из следующего примера. Если некоторая фирма хочет оценить необходимое количество денег для выдачи зарплаты, ее интересует средняя зарплата сотрудника, т.е. математическое ожидание зарплаты сотрудника, рассматриваемой как случайная величина. При устройстве на работу будущего сотрудника должна интересовать не средняя зарплата, а ее мода – заработная плата, получаемая большинством сотрудников. Сопоставление собственной заработной платы с медианой заработной платы говорит сотруднику о том, принадлежит ли он к хорошо или плохо оплачиваемой части сослуживцев.

На практике часто требуется оценить среднее отклонение возможных значений случайной величины от ее среднего значения. Характеристикой такого отклонения служит дисперсия случайной величины.

Дисперсией дискретной случайной величины называют математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

$$D(X) = M(X - M(X))^2. \quad (5.4)$$

Пример 5.6. Найти дисперсию СВ, заданной в примере 5.4.

Решение:

$$M(X) = 3,9;$$

$$D(X) = ((3-3,9)^2 \cdot 0,1 + (5-3,9)^2 \cdot 0,6 + (2-3,9)^2 \cdot 0,3 = \\ = 0,081 + 0,726 + 1,083 = 1,89.$$

Использование дисперсии не всегда бывает удобно, так как ее размерность равна квадрату размерности СВ. Поэтому для оценки отклонения возможных значений случайной величины вокруг ее среднего значения часто используют среднее квадратическое отклонение.

Средним квадратическим отклонением случайной величины X называется квадратный корень из дисперсии

$$\sigma(X) = \sqrt{D(X)}. \quad (5.5)$$

Для примера 5.6: $\sigma(X) = \sqrt{1,89} \approx 1,375$.

Для случайных величин в большинстве случаев невозможно составить перечень всех возможных значений. Поэтому необходимо ввести новые числовые характеристики этих величин. В качестве таких характеристик чаще всего используют функцию распределения и плотность распределения вероятностей (непрерывной СВ).

Функцией распределения случайной величины X называют функцию $F_X(x)$, представляющую собой вероятность того, что СВ X в результате испытания примет значение, меньшее x , т.е.

$$F_X(x) = P(X < x). \quad (5.6)$$

Геометрически это равенство означает, что $F_X(x)$ есть вероятность того, что случайная величина примет значение, которое на числовой оси изображается любой точкой, лежащей левее точки x .

Пусть известно статистическое распределение частостей количественного признака X . В этом случае удобно пользоваться эмпирической функцией распределения.

Эмпирической функцией распределения (функцией распределения выборки) называют функцию $F_X^*(x)$, определяющую частоту события, заключающегося в том, что СВ X в результате испытания примет значение, меньшее x , т.е.

$$F_X^*(x) = W(X < x). \quad (5.7)$$

Таким образом, эмпирическая функция распределения выборки служит для оценки теоретической функции распределения генеральной совокупности.

Непрерывные случайные величины можно исследовать с помощью еще одной функции – плотности распределения вероятностей.

Плотность распределения $f_X(x)$ вероятностей непрерывной с. в. X представляет собой первую производную от функции распределения $F_X(x)$ и характеризует скорость изменения функции распределения этой случайной величины.

Для наглядности представления выборки строят различные графики статистического распределения, в частности – полигон и гистограмму.

*Полигоном частот называют ломанную, отрезки которой соединяют точки $(x_1; n_1), (x_2; n_2), \dots, (x_k; n_k)$, где $n_i, i = 1, \dots, k$, – число наблюдений (частоты), при которых отмечалось значение признака, равное x_i . Для построения полигона частот на оси абсцисс откладывают варианты x_i , а на оси ординат – соответствующие им частоты n_i . Отметим, что сумма всех частот равна объему выборки. Точки $(x_i; n_i)$ соединяют отрезками прямых и получают *полигон частот*.*

Пример 5.7. Построить полигон частот для следующего распределения:

x_i	1	4	5	7
n_i	20	10	14	6

Решение показано на рис. 5.5.

Если вместо n_i выбрать частоты w_i , то получим *полигон относительных частот*.

В случае *непрерывного признака* целесообразно строить *гистограмму*, для чего интервал, в котором заключены все наблюдаемые

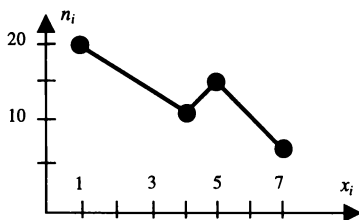


Рис. 5.5. Полигон частот

значения признака, разбивают на несколько частичных интервалов длиной h и находят для каждого частичного интервала n_i – сумму частот вариантов, попавших в i -й интервал.

Гистограммой частот называют ступенчатую функцию, состоящую из прямоугольников, основаниями которой служат частичные интервалы длиной h , а высоты равны отношению n_i/h (плотность частоты). Для построения гистограммы частот на оси абсцисс откладывают частичные интервалы, а над ними проводят отрезки, параллельные оси абсцисс на расстоянии n_i/h .

Площадь i -го частичного прямоугольника равна $h \cdot n_i/h = n_i$, т.е. сумме частот i -го интервала. Следовательно, *площадь гистограммы частот равна сумме всех частот, т.е. объему выборки.*

Если в качестве высот в гистограмме частот выбрать отношение w_i/h (плотность частоты), то получим *гистограмму относительных частот*. Отметим, что *площадь гистограммы относительных частот равна сумме всех относительных частот, т.е. единице.*

Пример 5.8. Построить гистограмму частот по распределению выборки объема $n = 100$, указанному в таблице.

Номер интервала	Частичный интервал $x_i - x_{i+1}$	Сумма частот n_i	Плотность частоты n_i/h_i
1	1 – 5	10	2,5
2	5 – 9	20	5,0
3	9 – 13	50	12,5
4	13 – 17	12	3,0
5	17 – 21	8	2,0

Решение приведено на рис. 5.6.

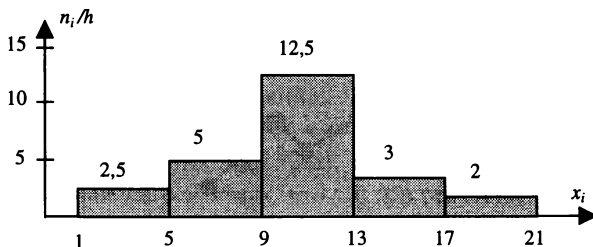


Рис. 5.6. Гистограмма частот

При решении конкретных задач встречаются как дискретные, так и непрерывные случайные величины. В качестве примеров рассмотрим одно дискретное (пуассоновское) и три непрерывных (равномерное, показательное и Гаусса) распределения случайных величин.

Распределение Пуассона. Найти вероятность того, что при большом числе повторных испытаний, в каждом из которых вероятность появления события очень мала и равна p (возможные значения события: $0, 1, 2, \dots, k, \dots$), событие произойдет ровно k раз. Закон распределения вероятностей:

$$P_{\lambda}(X = k) = \frac{\lambda^k}{k!} \exp(-\lambda), \quad k = 0, 1, 2, \dots, \quad (5.8)$$

где λ – параметр распределения, $k!$ (читается ка-факториал) представляет собой произведение всех натуральных чисел от 1 до k включительно ($k! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot k$).

Математическое ожидание и дисперсия в этом случае равны:

$$M(X) = D(X) = \lambda.$$

Показательное распределение. Показательным (экспоненциальным) называют распределение вероятностей непрерывной случайной величины X , которое описывается плотностью:

$$f_X(x) = \begin{cases} 0 & \text{при } x < 0, \\ \lambda \exp(-\lambda x) & \text{при } x \geq 0, \end{cases} \quad (5.9)$$

где $\lambda > 0$ – константа (рис. 5.7).

Математическое ожидание и дисперсия в этом случае равны:

$$M(X) = D(X) = 1/\lambda.$$

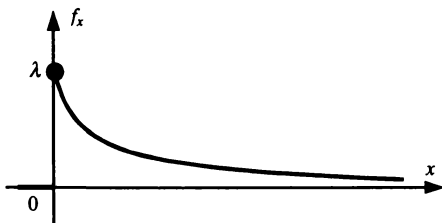


Рис. 5.7. Плотность распределения для показательного закона

Вероятность попадания в интервал (α, β) непрерывной случайной величины X , распределенной по показательному закону,

$$p(\alpha < X < \beta) = \exp(-\lambda\alpha) - \exp(-\lambda\beta).$$

Показательное распределение широко применяется в теории случайных процессов. Например, отказы техники, появление бракованных изделий на конвейере, поступление больных в стационар и т.п. образуют поток повторяющихся однотипных событий. Интервалы времени между соседними событиями в подобном потоке можно считать случайными величинами, которые при некоторых допущениях хорошо описываются показательным распределением.

Равномерное распределение на $[a, b]$. Распределение вероятностей называется равномерным, если на интервале, которому принадлежат все возможные значения случайной величины, плотность распределения сохраняет постоянное значение. Плотность равномерного распределения имеет вид (рис. 5.8):

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases} \quad (5.10)$$

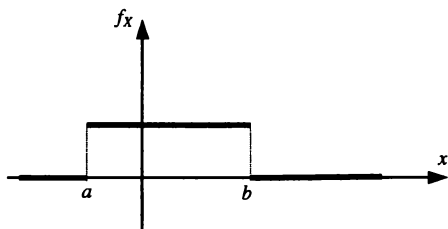


Рис. 5.8. Плотность распределения для равномерного закона

Распределение в интервале $(0, 1)$ будем называть *стандартным равномерным распределением*. Математическое ожидание и дисперсия в этом случае определяются как

$$M(X) = (a+b)/2, \quad D(X) = (b-a)^2/12.$$

Вероятность попадания в интервал (α, β) непрерывной случайной величины X , распределенной по равномерному закону (при $\alpha \in [a, b]$, $\beta \in [a, b]$),

$$P(\alpha < X < \beta) = (\beta - \alpha)/(b - a).$$

Нормальное распределение (распределение Гаусса). Нормальным называется распределение случайной величины, которое полностью определяется математическим ожиданием $M(X) = m$ и средним квадратическим отклонением $\sigma(X) = \sigma$ и описывается плотностью (рис. 5.9):

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right). \quad (5.11)$$

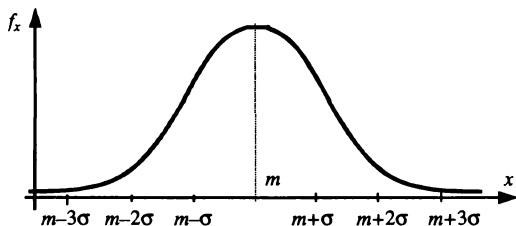


Рис. 5.9. Плотность распределения для нормального закона

Заметим, что при нормальном распределении справедливо *правило трех сигм*: вероятность попадания реализаций нормально распределенной случайной величины в интервал $m \pm 3\sigma$ составляет более 99,7%.

Нормальное распределение с параметрами $m = 0$ и $\sigma = 1$ называется *стандартным нормальным распределением*.

Вероятность попадания в интервал (α, β) непрерывной случайной величины X , распределенной по нормальному закону,

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - m}{\sigma}\right) - \Phi\left(\frac{\alpha - m}{\sigma}\right),$$

где $\Phi(x) = \frac{2}{\sqrt{2\pi}} \int_0^x \exp(-t^2) dt$ — функция Лапласа (интеграл вероятностей, его значения приведены во всех учебных пособиях по теории вероятностей).

Нормальное распределение играет очень важную роль в теории вероятности и широко применяется в математической статистике. Например, любой количественный параметр, полученный в результате измерения, следует рассматривать как случайную величину. На конкретное значение параметра при данном измерении

вливают независимо один от другого множество трудно учитываемых или даже неизвестных факторов: положение и состояние измерительного инструмента, состояние измеряемого объекта, флуктуации параметров окружающей среды (температура, влажность, плотность, скорость движения и т.п.). Все это приводит к различиям в значениях измеряемого параметра в разных измерениях. Оказалось, что ошибка измерения ведет себя как случайная величина, подчиняющаяся нормальному закону распределения.

Роль нормального распределения важна не только при проведении измерений. При сложении случайных величин, распределенных по произвольному закону, полученные суммы представляют собой случайные величины, распределенные по закону, близкому к нормальному. То есть последний играет роль некоторого предельного распределения, к которому стремятся распределения сумм произвольных случайных величин. Например, сумма 12 равномерно распределенных чисел с достаточной для практики точностью позволяет получить нормально распределенную случайную величину.

Моделирование в условиях стохастической неопределенности широко используется в таких разделах теории вероятностей, как дисперсионный анализ, регрессионный анализ и ковариационный анализ [114]. По существу, математическая модель исследуемого явления в рамках этих разделов имеет один и тот же формализм, который можно представить следующим образом.

Пусть имеется n наблюдений или измерений. Эти наблюдения рассматриваются как n случайных величин, $i = 1, 2, 3, \dots, n$, которые являются линейными комбинациями известных факторов x_{ij} , $i = 1, 2, 3, \dots, n$; $j = 1, 2, 3, \dots, k$, с k неизвестными постоянными β_j , $j = 1, 2, 3, \dots, k$, и ошибок u_i , $i = 1, 2, 3, \dots, n$:

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + \dots + x_{ik}\beta_k + u_i, \quad i = 1, 2, 3, \dots, n.$$

Целью *дисперсионного анализа* является получение выводов относительно ошибок u_i , $i = 1, 2, 3, \dots, n$ и некоторых $\{\beta_j\}$, причем выводов, которые остаются справедливыми независимо от значений других $\{\beta_j\}$ и «исключить» которые было бы более желательно, чем «оценить». При этом x_{ij} , $i = 1, 2, 3, \dots, n$; $j = 1, 2, 3, \dots, k$ являются целыми числами, равными обычно 0 или 1 (качественными «переменными-указателями» [114]). Если $\{x_{ij}\}$ не «переменные-указатели», а «пробегают» непрерывные множества значений, то мы получаем *регрессионный анализ*. В случае, когда среди $\{x_{ij}\}$ есть переменные двух видов, мы имеем *ковариационный анализ*. Природа неизвестных параметров $\{\beta_j\}$ может быть различна. Они могут быть или неизвестными постоянными, или случайными величинами, или часть из них — случайные величины, а часть — постоянные.

Другими словами, можно сказать, что различие трех видов анализа заключается в том, что в дисперсионном анализе все факторы исследуются *качественно*, в регрессионном анализе все факторы являются количественными и исследуются *количественно*, тогда как в ковариационном анализе одна часть рассматриваемых факторов исследуется *качественно*, а другая часть — *количественно*.

Как было отмечено в гл. 1, при реализации математических моделей могут использоваться как аналитические, так и алгоритмические методы исследования полученных математических соотношений. При применении аналитических методов выходные переменные модели могут быть представлены в виде аналитических выражений, для которых возможно использование универсальных преобразований случайной величины [28]. Однако подобные преобразования требуют глубокого знания специальных разделов теории вероятности и применимы в относительно простых случаях. Для сложных аналитических выражений, а также для алгоритмических методов анализа модели при оценке закона распределения случайной величины, зависящей от ряда случайных параметров, используют метод Монте-Карло [38]. Этот метод можно рассматривать как метод приближенного моделирования случайных величин с целью оценки характеристик их распределений. Идея моделирования случайных явлений, как известно, очень стара и, по мнению некоторых авторов, восходит ко времени Древнего Вавилона и Ветхого Завета. Современная формулировка метода статистического моделирования и само название «метод Монте-Карло» была предложена Дж.Нейманом совместно с С.М. Уламом в 40-х годах нынешнего столетия.

Под моделированием случайной величины в методе Монте-Карло подразумевается ее конструирование при помощи некоторого случайного механизма. После получения ряда реализаций интересующей нас случайной величины решается типичная задача математической статистики — оценка различных характеристик (математического ожидания, дисперсии и т.д.). Применение метода Монте-Карло возможно лишь с помощью ЭВМ.

Узловой проблемой для метода Монте-Карло является моделирование случайных чисел r , равномерно распределенных в интервале $(0,1)$. Существующие таблицы случайных чисел неудобно использовать для решения «больших» задач на ЭВМ, тем более что еще никем не было замечено преимущество таблиц перед так называемыми «псевдослучайными числами». Уже давно разрабатыва-

ются «физические датчики» случайных чисел, например, с ЭВМ соединяется счетчик числа α -частиц, вылетающих из некоторого радиоактивного источника за фиксированный промежуток времени. Они применяются сравнительно редко по двум причинам:

- невозможно воспроизвести выборочную последовательность для повторения расчетов;
- практически нельзя гарантировать устойчивую удовлетворительную постоянную работу датчика.

Как правило, для решения задач методом Монте-Карло используют рекуррентные формулы, производящие регулярную последовательность чисел, являющуюся для внешнего наблюдателя случайной и удовлетворяющую основным требованиям, накладываемым на «настоящие» случайные числа. Подобную последовательность принято называть *последовательностью псевдослучайных чисел*. Для задания такой последовательности достаточно выбрать ее начальное число. Многочисленные методики получения реализаций псевдослучайных чисел можно найти в [2, 31, 38, 54]. Простейший метод представлен в [2] и заключается в следующем. Выбирается рекуррентная последовательность

$$r_0 \leq c, \quad r_k = (mr_{k-1} + n) \bmod c, \quad (5.12)$$

где \bmod – операция вычисления остатка при целочисленном делении.

Соотношение (5.12) при заданных целых константах m , n , c позволяет получать последовательность из c целых чисел. Например, пусть $c = 8$, $m = 5$, $n = 3$. Примем $r_0 = 1$. Тогда имеем следующий бесконечно повторяющийся ряд значений: 1, 0, 3, 2, 5, 4, 7, 6, 1, ...

В данной последовательности период равен 8, и каждое значение от 0 до 7 на периоде появляется ровно один раз. Цикл из 8 чисел слишком мал. В [31] предлагаются константы для генерации перестановки чисел от 0 до 65525: $m = 25173$; $n = 13849$; $c = 65536$.

Подбор констант для получения псевдоравномерной последовательности – нетривиальная задача, требующая проверки гипотез о равномерности, независимости и однородности получаемых реализаций. Достаточно полный анализ алгоритмов, используемых для генерации псевдослучайных чисел, содержится в [54].

При реализации на ЭВМ значение r_0 генерируют с использованием таймера машины. В этом случае для различных моментов времени запуска программы получают различные фрагменты псевдослучайной последовательности, смещенные друг относительно

друга. Для получения одной и той же последовательности достаточно в качестве r_0 выбрать одно и то же число, например $r_0 = 0$.

Соотношение (5.12) позволяет получить последовательность целых чисел, распределенных равномерно в интервале от 0 до $(c - 1)$. Для нахождения числа, равномерно распределенного в интервале $(0, 1)$, необходимо r_k разделить на величину периода:

$$R_k(0, 1) = r_k / c. \quad (5.13)$$

Тогда для произвольного интервала (a, b) имеем следующее соотношение:

$$R_k(a, b) = a + (b - a)R_k(0, 1). \quad (5.14)$$

Для получения псевдонормальной последовательности можно воспользоваться следствием из основной теоремы теории вероятности [28], согласно которому при сложении случайных чисел найденные суммы представляют собой последовательность, распределение которой близко к нормальному закону. Как уже отмечалось выше, при проведении практических расчетов для определения k -го значения последовательности стандартного нормального распределения достаточно использовать сумму из 12 равномерно распределенных чисел [2]:

$$N_k(0, 1) = \sum_{i=1}^{12} R_i(0, 1) - 6. \quad (5.15)$$

Для нормального распределения со средним значением m и среднеквадратическим отклонением случайное число находим, используя (5.15):

$$N_k(m, \sigma) = m + \sigma N_k(0, 1). \quad (5.16)$$

Для получения псевдопоказательной последовательности можно воспользоваться преобразованием значения псевдоравномерного числа [2]:

$$E_k(\lambda) = -\frac{\text{Ln}[R_k(0, 1)]}{\lambda}. \quad (5.17)$$

Иллюстрацию особенностей анализа моделей в условиях неопределенности проведем на примере задачи о баскетболисте, постановка которой приведена в гл. 2.

Задача о баскетболисте. Несколько изменим содержательную постановку задачи о баскетболисте. Пусть начальная скорость v_0 мяча распределена по нормальному закону со средним значением $M(v_0) = 6,525$ м/с и среднеквадратическим отклонением $\sigma(v_0) = 0,25$ м/с. В соответствии с правилом трех сигм практически все значения v_0 в этом случае (более 99,7%) находятся в интервале $M(v_0) \pm 3\sigma(v_0)$, т.е. $v_0 \in [5,775; 7,275]$ м/с. Величину угла бросания α_0 примем также нормально распределенной с параметрами $M(\alpha_0) = 45^\circ$ и $\sigma(\alpha_0) = 5^\circ$. В этом случае дальность L , точность Δ броска, а также ширину d коридора «чистого» попадания следует рассматривать как случайные величины. В результате исследования полученной модели требуется:

- построить гистограмму относительных частот попаданий для модели (2.13), учитывающей сопротивление воздуха;
- исследовать изменение вероятности попадания мяча в корзину при броске со штрафной линии при различных значениях математического ожидания начальной скорости и угла бросания.

Исследование провести для случая «чистого» попадания мяча в корзину.

Для анализа вероятности попадания p_{Π} воспользуемся методом Монте-Карло. Проведем первоначально 16 вычислительных экспериментов. Используя (5.16), получим последовательность из 16 значений v_0 и α_0 . Затем по (2.13) и (2.14) вычислим дальности бросания L , соответствующие каждому значению скорости и угла; по (2.12) определим значение точности броска Δ_i , а по (2.15) – ширину коридора «чистого» попадания d_i (табл. 5.1).

Таблица 5.1

Результаты численных экспериментов

№	1	2	3	4	5	6	7	8
v_0	6,32	6,57	5,95	6,09	6,60	6,37	6,76	6,41
α_0	47,04	48,01	42,86	44,00	44,00	45,26	50,42	46,87
Δ_i	-0,265	0,037	-0,704	-0,539	0,1	-0,197	0,221	-0,152
d_i	0,061	0,064	0,049	0,052	0,052	0,056	0,069	0,061
		Попал						
№	9	10	11	12	13	14	15	16
v_0	6,69	6,22	6,38	6,30	6,34	6,89	6,57	6,00
α_0	52,20	43,83	59,24	36,04	41,65	48,59	39,36	51,45
Δ_i	0,066	-0,376	-0,686	-0,456	-0,247	0,441	-0,013	-0,742
d_i	0,073	0,052	0,085	0,021	0,044	0,065	0,036	0,072
	Попал						Попал	

При попадании мяча в корзину величина модуля Δ_i меньше ширины d_i коридора попадания. Из 16 бросков этому условию удовлетворяют только три. Отношение числа попаданий к общему числу бросков, или частота $w_{\text{п}}$ попаданий, равна 0,1875, или 18,75%. Однако выборка из 16 бросков не является достаточно *представительной*, так как в последующих 100 сериях по 16 бросков число попаданий изменялось от 0 до 6 попаданий в серии, т.е. частота попаданий колебалась от 0 до 0,375. Границы изменения относительной частоты $w_{\text{п}}$, вычисленные по 100 сериям бросков, в зависимости от объема выборки показаны на рис. 5.10. Можно видеть, что с увеличением объема выборки разброс значений оценки $w_{\text{п}}$ уменьшается, стремясь к некоторому предельному значению в районе 0,123. Таким образом, вероятность попадания со штрафной линии при принятых исходных данных составляет 12,3%, т.е. в среднем только 12 бросков из 100 попадают в цель.

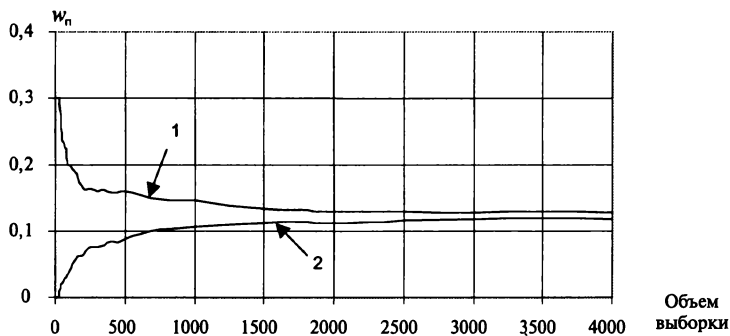


Рис. 5.10. Изменение предельных значений относительной частоты $w_{\text{п}}$ в зависимости от объема выборки:
1— верхнее значение; 2— нижнее значение

На рис. 5.11 изображены гистограммы относительных частот $w(\Delta)$, построенные как по выборке из табл. 5.1, так и по выборке из 2000 опытов. Для наглядности отрезками прямых линий соединены значения относительных частот в средних точках интервалов. Анализ приведенных гистограмм позволяет отметить, что с увеличением объема выборки графики статистических параметров получают более сглаженными и достоверными.

На рис. 5.12 показано распределение вероятности попадания при выполнении штрафного броска в зависимости от величины математического ожидания начальной скорости v_0 и угла бросания α_0 . При построении распределения вероятностей исследуемый диапазон начальной скорости и угла бросания покрывался сеткой с шагом 0,2 м/с для скорости и 2°

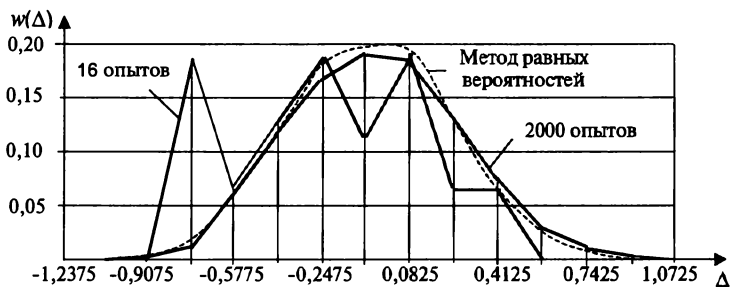


Рис. 5.11. Гистограмма относительных частот для точности броска

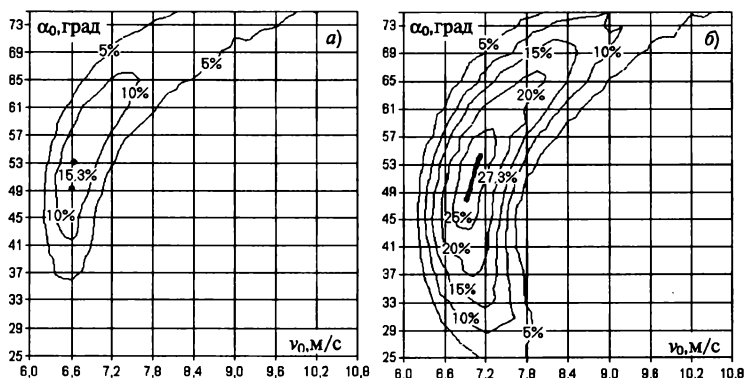


Рис. 5.12. Вероятность попадания в зависимости от начальных условий:
а — чистое попадание; б — с учетом отскока от щита и кольца

угла бросания. Для каждого узла сетки строилась выборка попаданий из 2000 опытов. Величина среднеквадратического отклонения во всех опытах принималась постоянной: $\sigma(v_0) = 0,25$ м/с и $\sigma(\alpha_0) = 5^\circ$. Величины математического ожидания соответствовали значениям скорости и угла бросания в конкретном узле сетки.

На рис. 5.12,а показано распределение вероятности чистого попадания. При оценке попадания использовалась модель (2.13), учитывающая сопротивление воздуха. Можно отметить, что максимум вероятности попадания 15,3% достигался, если величины математических ожиданий скорости и угла принимали значения 6,6 м/с и 50° соответственно.

Для сравнения на рис. 5.12,б приведено распределение вероятности попадания с учетом отскоков мяча от щита и кольца. В этом случае ис-

пользовалась модель взаимодействия мяча с кольцом и щитом, рассмотренная в гл. 4. Максимум вероятности попадания составил 27,3% при математических ожиданиях скорости 6,9 м/с и угла 50°. Оптимальный угол бросания по сравнению с чистым попаданием практически не изменился, оптимальная начальная скорость стала чуть больше. Увеличение среднего значения начальной скорости в данном случае можно расценивать как смещение точки прицеливания при броске от центра корзины ближе к щиту. Таким образом, если игрок выполняет штрафной бросок и нацеливается не на чистое попадание мяча в кольцо, а с возможностью отскока его от щита, то вероятность попадания при таком броске увеличивается почти в 2 раза.

К существенным недостаткам метода Монте-Карло можно отнести необходимость проведения большого числа вычислительных экспериментов с целью получения надежных оценок случайных переменных модели. Например, при нормальном законе распределения случайных параметров примерное число вычислений исследуемых переменных (необходимое число опытов) при построении гистограммы относительных частот можно определить из оценки отклонения относительной частоты от постоянного значения вероятности:

$$P\left(\left|\frac{m}{N} - P\right| \leq \varepsilon\right) = 2\Phi\left(\varepsilon \sqrt{\frac{N}{P(1-P)}}\right), \quad (5.18)$$

где $\Phi(\cdot)$ – функция Лапласа; ε – величина отклонения (ε -полоска); P – оцениваемая величина вероятности.

Величина отклонения ε определяет точность построения гистограммы. Например, при $\varepsilon = 0,05$ оцениваемая величина p лежит в интервале, составляющем $\pm 5\%$ от значения относительной частоты. Если в результате вычисления по (5.18) значение вероятности получилось равным 0,91, то это означает, что истинное значение p лежит в ε -полоске с вероятностью 91%. Задаваясь вероятностью $p(\varepsilon)$ попадания в ε -полоску и величиной точности ε , из соотношения (5.18) можно получить оценку необходимого числа опытов:

$$N = \left[\frac{\Phi^{-1}(p(\varepsilon)/2)}{\varepsilon} \right]^2 P(1-P).$$

Из анализа этого соотношения можно заключить, что наибольшее число опытов необходимо при $P = 0,5$, и оно обратно пропорционально квадрату точности ε . Например, при $\varepsilon = 0,05$ и

$P(\epsilon) = 0,95$ требуется провести 384 опыта, а с точностью $\epsilon = 0,01$ – 9600 опытов.

В случае, если число случайных параметров модели невелико, для построения гистограммы относительных частот можно предложить более эффективный приближенный метод, который назовем *методом равных вероятностей*. Поясним суть данного метода (рис. 5.13).

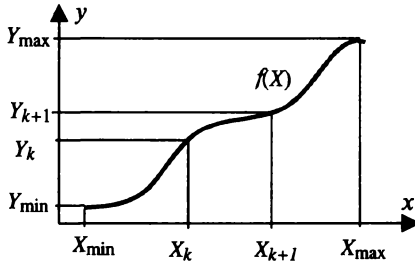


Рис. 5.13. Иллюстрация к методу равных вероятностей

Пусть известна функция $Y=f(X)$ и задан закон распределения случайной величины X . Требуется построить гистограмму для величины Y . Разобьем область значений X на ряд интервалов таким образом, чтобы в каждом из них функция $f(X)$ была монотонна. В этом случае вероятность попадания значений X в интервал $[X_k, X_{k+1}]$ равна вероятности попадания Y в соответствующий интервал $[Y_k, Y_{k+1}]$:

$$P(X_k \leq x \leq X_{k+1}) = P(Y_k \leq y \leq Y_{k+1}) = P(f(X_k) \leq x \leq f(X_{k+1})).$$

Тогда среднюю плотность величины Y на интервале $[Y_k, Y_{k+1}]$ можно вычислить следующим образом:

$$P(Y_k \leq Y \leq Y_{k+1}) = \frac{1}{|Y_{k+1} - Y_k|} \int_{X_k}^{X_{k+1}} p(X) dx, \quad (5.19)$$

где $p(X)$ – плотность распределения величины X .

Теперь необходимо заменить непрерывные случайные величины их дискретными аналогами на интервалах. Точность метода повышается с увеличением числа интервалов разбиения области значений величины X . Если функция на области значений X немонотонна, то с учетом *несовместности* X , т.е. невозможности од-

новременной реализаций двух различных значений X , значение плотности распределения для фиксированного $Y = \bar{y}$ определяется суммированием плотностей по всем интервалам для данного \bar{y} .

Если число случайных параметров функции больше одного, то с учетом независимости параметров одного от другого величину плотности при фиксированном $Y = \bar{y}$ можно найти как произведение плотностей, полученных для отдельных параметров:

$$p(Y = \bar{y}) = \prod_{k=1}^M p_{X_k}(Y = \bar{y}), \quad (5.20)$$

где M – число случайных параметров.

Пример о баскетболисте (продолжение). Построим гистограмму относительных частот с применением метода равных вероятностей. В модели использованы два случайных параметра: ν_0 и α_0 , имеющие заданное нормальное распределение. Область возможных значений для ν_0 и α_0 выберем в соответствии с правилом «трех сигм». Разобьем выбранные области на 12 интервалов по $0,5\sigma$. Определим значения ν_0 и α_0 на границах интервалов. Получим два массива по 13 значений ν_0 и α_0 . Используя модель (2.13), вычислим значения точности попаданий Δ для всех сочетаний ν_0 и α_0 . В итоге получим массив значений Δ размером 13 на 13 элементов. Вероятности попаданий значений случайной величины X в заданные интервалы для стандартного нормального распределения, вычисленные с помощью функции Лапласа, приведены в табл. 5.2.

Таблица 5.2

X	[0,0; 0,5]	[0,5; 1,0]	[1,0; 1,5]	[1,5; 2,0]	[2,0; 2,5]	[2,5; 3,0]
P	0,1915	0,1498	0,0919	0,0440	0,0166	0,0049

Используя данные, приведенные в таблице, а также соотношения (5.19) и (5.20), можно получить распределение относительной частоты для точности попадания D . Полученная при этом гистограмма показана на рис.5.11 штриховой линией. Как видно, она практически совпадает с гистограммой выборки, построенной по результатам 2000 опытов (более сглаженная). Для ее построения потребовалось провести всего 169 опытов.

5.4. МОДЕЛИРОВАНИЕ МАРКОВСКИХ СЛУЧАЙНЫХ ПРОЦЕССОВ

Человек, имеющий одни часы, твердо знает, который час. Человек, имеющий несколько часов, ни в чем не уверен.

Закон Сегала

Любые данные, полученные в результате временных наблюдений реального физического явления, можно классифицировать как *процесс*. При этом если процесс можно описать явными математическими формулами в терминах детерминированных переменных, то это будет *детерминированный процесс*. В противном случае, если любое наблюдение дает только один вариант из множества возможных, получим *случайный процесс*. Таким образом, *случайный процесс* – это случайная величина, зависящая от времени.

Введем некоторые основные определения [11, 18, 89].

Выборочной функцией (реализацией, траекторией) случайного процесса будем называть конкретную реализацию описываемого случайного явления. Тогда *случайный процесс* – это совокупность выборочных функций, которые может дать исследуемое случайное явление.

Сечением случайного процесса будем называть случайную величину, характеризующую исследуемое случайное явление в некоторый момент времени, т.е. это временной срез *случайного процесса*.

Обычно простую математическую модель, позволяющую получить решение в явном (аналитическом) виде, найти не удается. Однако в некоторых особых случаях ее удастся построить. Подобная ситуация, например, имеет место тогда, когда исследуемая операция представляет собой (точно или приближенно) так называемый *марковский случайный процесс*.

Одним из основных признаков, по которым классифицируют случайные процессы, является отношение зависимости между сечениями случайного процесса.

Случайный процесс называется *марковским*, если для любого момента времени t_0 вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент t_0 и не зависят от того, когда и как *система* пришла в это состояние.

Приведем пример простейшего марковского процесса. Пусть система – это счетчик Гейгера, на который время от времени по-

падают космические частицы; состояние системы в момент времени t характеризуется показанием счетчика — числом частиц, пришедших до этого момента. Пусть в момент t_0 индикатор счетчика показывает S_0 . Вероятность того, что в момент $t > t_0$ индикатор покажет то или другое число частиц S_1 (или не менее S_1), разумеется, зависит от S_0 , но не зависит от того, в какие именно моменты приходили частицы до момента t_0 .

В реальных условиях, в сущности, любой процесс можно рассматривать как марковский, если все параметры из «прошлого», от которых зависит «будущее», включить в «настоящее».

Для марковского процесса будущее зависит от прошлого только через настоящее.

Хотя на практике марковские процессы в чистом виде обычно не встречаются, нередко приходится иметь дело с процессами, для которых влиянием «предыстории» можно пренебречь. При изучении таких процессов применяют марковские модели. Наиболее простыми являются марковские случайные процессы с дискретными состояниями и дискретным или непрерывным временем.

Марковский процесс называется *процессом с дискретными состояниями*, если его возможные состояния S_1, S_2, S_3, \dots можно заранее перечислить (перенумеровать), и переход системы из состояния в состояние переходит «скачком», практически мгновенно. Например, пылесос может находиться в трех состояниях: включенном, выключенном и в состоянии поломки. Процесс называется *процессом с непрерывным временем*, если моменты возможных переходов из состояния в состояние не фиксированы заранее, а неопределенны, случайны, т.е. если переход может осуществиться в принципе в любой момент. Если процесс марковский и моменты возможных переходов из состояния в состояние фиксированы, то этот процесс с дискретным временем называется *марковской цепью*.

Мы будем рассматривать только марковские процессы с дискретными состояниями и непрерывным временем.

Пример 5.9. Техническое устройство S состоит из двух узлов, каждый из которых в случайный момент времени может выйти из строя (отказ), после чего мгновенно начинается ремонт узла, тоже продолжающийся заранее неизвестное, случайное время. Графическое изображение возможных состояний системы и порядок переходов называется *графом состояний системы* (рис. 5.14). Возможные состояния системы: S_0 — оба узла исправны; S_1 — первый узел ремонтируется, второй исправен; S_2 — второй узел ремонтируется, первый исправен; S_3 — оба узла ремонтиру-

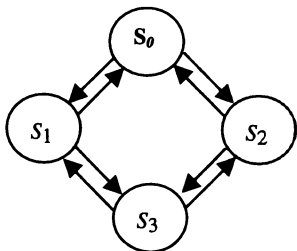


Рис. 5.14. Граф состояний

ются. Переходы системы S из состояния в состояние происходят практически мгновенно, в случайные моменты выхода из строя того или иного узла либо окончания ремонта.

Чтобы построить математическую модель марковского процесса с дискретными состояниями и непрерывным временем необходимо ввести понятие потока событий.

Потоком событий называется последовательность однотипных ситуаций, следующих одна за другой в какие-либо случайные моменты времени. Например: поток отказов вычислительной техники; появление бракованных изделий.

Ситуации, образующие поток, сами вероятностью не обладают; вероятностями обладают другие, основанные на них события. Например: исходная ситуация – попадание на участок времени Δt некоторого события. Случайное событие – попадание на участок времени Δt ровно двух событий.

Введем некоторые характеристики потока событий. Важнейшей характеристикой является *интенсивность* потока λ – среднее число событий, приходящихся на единицу времени. Интенсивность потока может быть как постоянной ($\lambda = \text{const}$), так и зависящей от времени t .

Поток событий называется *регулярным*, если события следуют одно за другим через определенные равные промежутки времени. Поток событий называется *стационарным*, если его вероятностные характеристики не зависят от времени. В частности, интенсивность λ стационарного потока должна быть постоянной.

Поток событий называется *потоком без последействия*, если для любых двух непересекающихся участков времени τ_1 и τ_2 число событий, попадающих на один из участков, не зависит от того, сколько событий попало на другой. Например, в приведенном выше примере со счетчиком Гейгера поток поступления частиц можно считать потоком без последействия.

Поток событий называется *ординарным*, если события в нем появляются поодиночке, а не группами по несколько событий сразу. Например, поток событий – прохождение автомобилей мимо наблюдателя по дороге с одnorядным движением. Поток поступления вагонов на железнодорожную станцию не является ординарным, так как вагоны прибывают на станцию составами.

Поток событий называется *простейшим* (или *стационарным пуассоновским*), если он обладает сразу тремя свойствами: *стационарен, ординарен и не имеет последствий*.

Простейший поток играет среди других потоков особую роль. Это связано с тем, что при наложении (суперпозиции) достаточно большого числа независимых, стационарных и ординарных потоков (сравнимых между собой по интенсивности) получается поток, близкий к простейшему.

Для простейшего потока с интенсивностью λ интервал времени Δt между соседними событиями имеет *показательное распределение* с плотностью

$$f(\Delta t) = \lambda e^{-\lambda \Delta t} \quad (\Delta t > 0). \quad (5.21)$$

Величина λ в (5.21) называется *параметром показательного закона*.

Если ввести так называемый *коэффициент вариации* для случайного интервала времени T , равный отношению среднеквадратического отклонения к математическому ожиданию ($v_T = \sigma_T / m_T$), то можно сделать вывод о том, что для простейшего потока событий коэффициент вариации интервалов между событиями равен единице, так как для показательного распределения $\sigma_T = m_T = 1/\lambda$.

Очевидно, что для регулярного потока событий ($\sigma_T = 0$), коэффициент вариации $v_T = 0$. Для большинства потоков событий, встречающихся на практике, коэффициент вариации интервалов между событиями заключен между 0 и 1 и может служить некоторой мерой *степени регулярности потока*.

Простейший поток является наименее регулярным потоком.

Величина числа событий N за время Δt для простейшего потока подчиняется распределению Пуассона:

$$P_{N=k} = \frac{(\lambda \Delta t)^k}{k!} e^{-\lambda \Delta t}. \quad (5.22)$$

В частности, вероятность не появления на интервале времени Δt ни одного события составляет $P_{N=0} = e^{-\lambda \Delta t}$; более одного события — $P_{N>1} = 1 - e^{-\lambda \Delta t}$.

В случае достаточно малого промежутка времени ($\Delta t \ll 1$) величина экспоненты в (5.22) $e^{-\lambda \Delta t} \approx 1$. Тогда вероятность появления на таком малом интервале времени ровно одного события

$$P_{N=1} = p_{\Delta t} = \lambda \Delta t, \quad \Delta t \ll 1, \quad \Delta t > 0. \quad (5.23)$$

Величину $p_{\Delta t}$, входящую в (5.23), называют *элементом вероятности*. Заметим, что элемент вероятности для простейшего потока не зависит от того, сколько событий и когда появилось ранее.

Поток событий называется *рекуррентным (потоком Пальма)*, если он стационарен, ординарен, а интервалы времени между событиями T_1, T_2, T_3, \dots представляют собой независимые случайные величины с одинаковым *произвольным распределением*. Очевидно, что простейший поток представляет собой частный случай рекуррентного потока, когда интервалы между событиями имеют показательное распределение.

Рассмотрим моделирование марковского случайного процесса [18]. Для этого составим и решим так называемые *уравнения Колмогорова* – особого вида дифференциальные уравнения, в которых неизвестными (искомыми) функциями являются вероятности состояний. Покажем на конкретном примере, как эти уравнения составляются.

Назовем вероятностью i -го состояния вероятность $p_i(t)$ того, что система в момент времени t находится в состоянии S_i . Так как число состояний конечно и система должна находиться хотя бы в одном из них, то в любой момент времени справедливо следующее так называемое *нормировочное условие*:

$$\sum_{k=1}^N p_k(t) = 1, \quad (5.24)$$

где N – число состояний системы.

Пусть система S имеет четыре состояния: S_1, S_2, S_3, S_4 ; размеченный *граф состояний* [18] показан на рис. 5.15. Переход системы из состояния i в состояние j представляет собой простейший поток интенсивностью λ_{ij} .

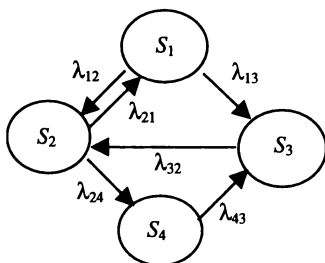


Рис. 5.15. Граф состояний

Рассмотрим одну из вероятностей состояния, например $p_1(t)$ – вероятность того, что в момент t система будет в состоянии S_1 . Придадим t малое приращение Δt и найдем $p_1(t + \Delta t)$ – вероятность того, что в момент $(t + \Delta t)$ система будет в состоянии S_1 . Это

может произойти, очевидно, в двух случаях: либо в момент t система уже была в состоянии S_1 , а за время Δt не вышла из него: либо в момент t система была в состоянии S_2 , а за время Δt перешла из него в состояние S_1 .

Найдем вероятность развития событий по первому варианту. Система находится в состоянии S_1 . В этом случае возможен переход в состояние S_2 или S_3 . Это несовместные события. Вероятность появления одного из двух несовместных событий равна сумме вероятностей: $(\lambda_{12} + \lambda_{13})\Delta t$. Тогда вероятность того, что система останется в состоянии S_1 , равна $1 - (\lambda_{12} + \lambda_{13})\Delta t$. Так как вероятность оказаться в состоянии S_1 к моменту времени t у системы равна $p_1(t)$ и последующие переходы не зависят от того, каким образом система пришла к этому состоянию, то вероятность первого варианта равна произведению вероятностей $p_1(t)[1 - (\lambda_{12} + \lambda_{13})\Delta t]$.

Рассуждая аналогично, получим вероятность второго варианта $p_2(t)\lambda_{21}\Delta t$. Так как первый и второй варианты несовместны (реализуется один из них), то вероятность для системы находиться к моменту времени $(t + \Delta t)$ в состоянии S_1 равна:

$$p_1(t + \Delta t) = p_1(t)[1 - (\lambda_{12} + \lambda_{13})\Delta t] + p_2(t)\lambda_{21}\Delta t,$$

откуда $[p_1(t + \Delta t) - p_1(t)]/\Delta t = \lambda_{21}p_2(t) - (\lambda_{12} + \lambda_{13})p_1(t)$.

Устремив Δt к нулю и переходя к пределу, получим дифференциальное уравнение для $p_1(t)$:

$$\frac{dp_1}{dt} = \lambda_{21}p_2 - (\lambda_{12} + \lambda_{13})p_1.$$

Аналогичным образом можно составить остальные три уравнения. В итоге будем иметь систему четырех обыкновенных дифференциальных уравнений Колмогорова с четырьмя неизвестными функциями времени p_1, p_2, p_3, p_4 :

$$\begin{cases} \frac{dp_1}{dt} = \lambda_{21}p_2 - (\lambda_{12} + \lambda_{13})p_1, \\ \frac{dp_2}{dt} = \lambda_{12}p_1 + \lambda_{32}p_3 - (\lambda_{24} + \lambda_{21})p_2, \\ \frac{dp_3}{dt} = \lambda_{31}p_1 + \lambda_{43}p_4 - \lambda_{32}p_2, \\ \frac{dp_4}{dt} = \lambda_{24}p_2 - \lambda_{43}p_4. \end{cases} \quad (5.25)$$

Заметим, что любое из них можно отбросить, заменив соотношением (5.24).

Чтобы решить уравнение Колмогорова и найти вероятности состояний, прежде всего необходимо задать начальные условия. Если мы точно знаем, что в начальный момент система находится в состоянии S_1 , то для этого момента $p_1(0) = 1$, а все остальные начальные вероятности равны нулю. Теперь полученную систему линейных дифференциальных уравнений можно решить аналитически или численно. В результате решения находим все вероятности состояний как функции времени. Типичное распределение вероятностей показано на рис. 5.16.

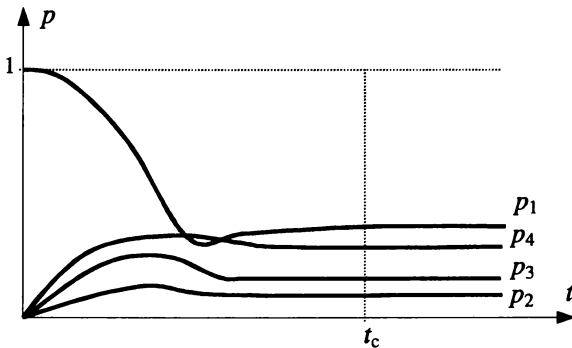


Рис.5.16. Вероятности состояний

Характерной особенностью является то, что начиная с некоторого момента времени t_c вероятности состояний практически перестают зависеть от времени, принимая некоторое постоянное стационарное значение.

В теории случайных процессов доказывается, что если число состояний системы конечно и из каждого из них можно (за конечное число шагов) перейти в любое другое, то в этом случае существуют так называемые *финальные вероятности* (пределы вероятности состояний при $t \rightarrow \infty$, не зависящие от начального состояния системы).

Финальные вероятности будем обозначать теми же буквами, что и сами вероятности состояния (p_1, p_2, p_3, p_4), но теперь это уже не функции времени, а постоянные числа. Финальную вероятность состояния S_i можно рассматривать как среднее относительное время пребывания системы в этом состоянии. Например, если система S имеет три состояния S_1, S_2, S_3 и их финальные вероятности

сти равны 0,1; 0,2 и 0,7, то это значит, что в предельном стационарном режиме система в среднем одну десятую часть времени проводит в состоянии S_1 , две десятых – в состоянии S_2 и семь десятых – в состоянии S_3 .

Чтобы найти финальные вероятности, воспользуемся тем, что теперь p_1, p_2, p_3, p_4 – постоянны, т.е. их производные по времени равны нулю. Теперь из уравнений Колмогорова получим однородную систему линейных алгебраических уравнений. Однако в связи с тем, что в уравнениях нет свободных членов, неизвестные определяются с точностью до произвольного множителя. Поэтому необходимо воспользоваться нормировочным условием (5.24). При этом одно (любое, как правило, наиболее сложное по виду) из уравнений можно отбросить (оно вытекает как следствие из остальных).

Рассмотренную выше модель случайных процессов удобно применить к моделированию так называемых *систем массового обслуживания* (СМО). Под СМО понимают системы, на вход которых подается случайный поток однотипных заявок (событий), обрабатываемых одним или несколькими однотипными *каналами* (устройствами). Спектр систем, которые можно рассматривать как СМО, очень широк (табл. 5.3). СМО могут быть одноканальными и многоканальными, с очередями или без них. Разработкой и изучением математических моделей для СМО занимается специальная дисциплина, называемая *теорией массового обслуживания* [18]. Как правило, вопросы моделирования СМО особенно остро стоят при проектировании новых систем. Само проектирование выпол-

Таблица 5.3

Примеры СМО

<i>СМО</i>	<i>Канал</i>	<i>Поток событий</i>
Телефонная станция	Телефонная линия	Звонки
Ремонтная мастерская	Мастер	Заявки на ремонт
Железнодорожная касса	Кассир	Пассажиры
Участок цеха	Отдельный станок	Детали на обработку
Больница (стационар)	Больничная койка	Больные
Перекресток	Направление переезда	Автомашины
Магазин	Продавец	Покупатели
Морской порт	Причал	Суда под разгрузку

няется с учетом так называемых *параметров эффективности* системы, в качестве которых могут выступать:

- среднее число заявок, обслуженных в единицу времени;
- среднее число занятых каналов;
- средняя длина очереди;
- среднее время ожидания в очереди;
- среднее число заявок, получивших отказ в обслуживании и т.д.

Например, если известна интенсивность поступления заготовок на участок цеха, то требуется определить, какое число станков и какой производительности необходимо использовать, чтобы участок работал эффективно, т.е. загрузка станков была выше 50–60% (станки и рабочие не простаивали), и в то же время участок справлялся с потоком заявок (не тормозил работу других участков). Основная сложность заключается в том, что имеется несколько вариантов решений. Так, можно взять меньшее число высокопроизводительных станков, но, как правило, такие станки более дорогостоящие, требуют более квалифицированного обслуживания, более дорогих запасных частей. В то же время, в данном случае число самих рабочих будет меньше. Напротив, менее производительные станки дешевле при приобретении и в обслуживании, но для их установки, возможно, потребуются обширные производственные помещения. Исследование математической модели подобной системы позволит более обоснованно принять решение о числе и типе станков.

Рассмотрим модель n -канальной СМО с отказами. Подобную схему можно применить к телефонной станции (если все каналы заняты, то звонок не обслуживается), зенитную батарею (если все пушки «захватили» свои цели, то новые самолеты остаются свободными от огня) или стационар больницы (если все койки заняты, то больные вынуждены обращаться в другие больницы).

Примем следующие предположения:

- все каналы однотипны;
- время обслуживания заявки в канале случайно и образует простейший поток интенсивностью μ ;
- поступающий поток заявок интенсивностью λ будем считать простейшим;
- система может иметь $n + 1$ состояний: S_0 – все каналы свободны; S_1 – занят один канал; S_2 – заняты два канала и т.д.;
- если все каналы системы заняты, то очередная заявка не будет удовлетворена.

Так как мы предположили, что входной и выходной потоки простейшие, то система может переходить из одного состояния в другое тоже последовательно: из S_0 в S_1 ; из S_1 в S_0 или в S_2 и т.д. (рис. 5.17).

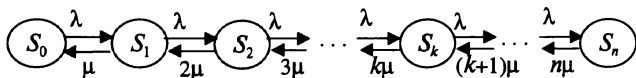


Рис. 5.17. Граф состояний

Составим уравнения Колмогорова для финальных вероятностей:

$$\begin{aligned}
 S_0 : \mu p_1 - \lambda p_0 &= 0 \Rightarrow p_1 = \frac{\lambda}{\mu} p_0, \\
 S_1 : 2\mu p_2 + \lambda p_0 - (\lambda + \mu) p_1 &= 0 \Rightarrow p_2 = \frac{\lambda^2}{2\mu^2} p_0, \\
 S_2 : 3\mu p_3 + \lambda p_1 - (\lambda + 2\mu) p_2 &= 0 \Rightarrow p_3 = \frac{\lambda^3}{2 \cdot 3\mu^3} p_0, \\
 &\dots \\
 S_{k-1} : k\mu p_k + \lambda p_{k-2} - (\lambda + (k-1)\mu) p_{k-1} &= 0 \Rightarrow p_k = \frac{\lambda^k}{k! \mu^k} p_0.
 \end{aligned} \tag{5.26}$$

Введем понятие *приведенной интенсивности* заявок $\alpha = \lambda/\mu$, характеризующее среднее число поступивших заявок за среднее время обслуживания одной заявки. Используя условие нормировки (5.24), можно найти вероятность p_0 :

$$p_0 = \left(1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + \dots + \frac{\alpha^n}{n!} \right)^{-1} = \left(\sum_{k=0}^n \frac{\alpha^k}{k!} \right)^{-1}. \tag{5.27}$$

Вероятность остальных состояний можно выразить через p_0 :

$$p_k = \frac{\alpha^k}{k!} p_0. \tag{5.28}$$

Соотношения (5.27) и (5.28) называются *формулами Эрланга*.

Если все каналы заняты, то очередная заявка не будет обслужена системой и получит отказ. Поэтому *вероятность отказа системы*

$$P_{\text{отк}} = P_n = \frac{\alpha^n}{n!} P_0. \quad (5.29)$$

Вероятность обслуживания заявки

$$Q = 1 - P_{\text{отк}} = 1 - \frac{\alpha^n}{n!} P_0. \quad (5.30)$$

Абсолютной пропускной способностью (A) называют среднее число заявок, обслуженных системой в единицу времени:

$$A = \lambda Q. \quad (5.31)$$

Среднее число загруженных каналов

$$N_3 = p_1 + 2p_2 + \dots + np_n. \quad (5.32)$$

Можно ввести коэффициент загрузки одного канала как отношение среднего числа загруженных каналов к их общему числу:

$$K_3 = N_3/n. \quad (5.33)$$

Пример 5.10. Рассматривается автоматическая телефонная станция (АТС) на 5 линий связи. Интенсивность поступающего потока заявок составляет 2 вызова в мин. Средняя длительность одного звонка равна 1 мин. Требуется определить вероятность отказа и среднее число занятых каналов.

Итак, дано: $n = 5$; $\lambda = 2$ выз./мин.; $t_{\text{обс}} = 1$ мин. По средней длительности одного звонка можно определить $\mu = 1/t_{\text{обс}} = 1$. Тогда $\alpha = \lambda\mu = \lambda t_{\text{обс}} = 2$. Для определения параметров АТС построим табл. 5.4:

Таблица 5.4

Характеристики АТС

k	$\alpha^k / k!$	p_k	$k p_k$
0	1,000	0,138	0,000
1	2,000	0,275	0,275
2	2,000	0,275	0,550
3	1,333	0,183	0,549
4	0,667	0,092	0,368
5	0,267	0,037	0,185
	$p_0 = 0,138$	$\Sigma p_k = 1,000$	$N_3 = 1,927$

Вероятность отказа $P_{\text{отк}} = 0,037$, или 3,7%, т.е. 37 заявок из 1000 в среднем получают отказ. Вероятность, что все каналы свободны, составляет $p_0 = 0,138$, т.е. 13,8% времени все каналы простаивают одновременно. Среднее число загруженных каналов $N_3 = 1,927$. Коэффициент загрузки одного канала составляет $k_3 = 0,39$, т.е. в среднем один канал работает лишь 39% времени, а 61% – простаивает.

Рассмотрим модель n -канальной СМО с бесконечной очередью. Подобную схему можно применить к железнодорожной кассе, участку цеха, ремонтной мастерской и т.п. Примем следующие предположения:

- все каналы однотипны;
- время обслуживания заявки в канале случайно и образует простейший поток интенсивностью μ ;
- поступающий поток заявок интенсивностью λ будем считать простейшим;
- если все каналы системы заняты, то очередная заявка ставится в конец очереди;
- система может иметь, вообще говоря, бесконечное, но счетное число состояний: S_0 – все каналы свободны; S_1 – занят один канал; S_2 – заняты два канал; S_n – заняты все каналы; S_{n+1} – одна заявка в очереди; S_{n+m} – m заявок в очереди, и т.д.

Граф состояний системы приведен на рис. 5.18. Составляя уравнения Колмогорова для финальных вероятностей аналогично СМО с отказами, получим:

$$p_k = \frac{\alpha^k}{k!} p_0, \quad k = 1, 2, \dots, n,$$

$$p_k = \frac{\alpha^n}{n!} \left(\frac{\alpha}{n}\right)^{k-n} p_0, \quad k > n. \quad (5.34)$$

Используя условие нормировки (5.24), можно найти вероятность p_0 :

$$p_0 = \left(1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + \dots + \frac{\alpha^n}{n!} \left(1 + \frac{\alpha}{n} + \frac{\alpha^2}{n^2} + \frac{\alpha^3}{n^3} + \dots \right) \right)^{-1}. \quad (5.35)$$

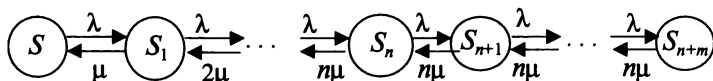


Рис. 5.18. Граф состояний

Несложно заметить, что бесконечная сумма в скобках выражения (5.35) образует геометрическую прогрессию со знаменателем α/n . Геометрическая прогрессия сходится, если ее знаменатель меньше единицы:

$$\alpha/n < 1. \quad (5.36)$$

Если $\alpha/n \geq 1$, то длина очереди в системе стремится к бесконечности. При выполнении условия (5.36) бесконечную сумму в (5.35) можно заменить суммой ряда геометрической прогрессии

$$1 + \frac{\alpha}{n} + \frac{\alpha^2}{n^2} + \frac{\alpha^3}{n^3} + \dots = \frac{1}{1 - \alpha/n} = \frac{n}{n - \alpha}.$$

Тогда (5.35) можно записать в виде [80]:

$$p_0 = \left(1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + \dots + \frac{\alpha^{n+1}}{n!(n - \alpha)} \right)^{-1}. \quad (5.37)$$

Среднее число загруженных каналов $N_3 = \alpha$. Коэффициент загрузки каналов $K_3 = \alpha/n$. Среднее число заявок в очереди

$$L_{оч} = \sum_{k=1}^{\infty} k p_{n+k} = \frac{\alpha^{n+1} p_0}{n \cdot n! (1 - \alpha/n)^2}. \quad \text{Среднее время ожидания в очереди}$$

$$W_{оч} = L_{оч} / \lambda.$$

Пример 5.11. По конвейеру в бункер поступают заготовки двух видов: *A* и *B*. Интенсивность поступления деталей одинакова и составляет $\lambda_A = \lambda_B = 0,45$ заготовок в минуту. Для обработки заготовок используются два станка, которые могут обрабатывать как детали *A*, так и детали *B*. Средняя трудоемкость обработки заготовок одинакова и составляет $t_{об} = 2$ мин. Требуется определить среднюю длину очереди и время ожидания в очереди.

Итак, имеем: $n = 2$; $\lambda = \lambda_A + \lambda_B = 0,9$ дет./мин.; $\mu = 1/t_{об} = 0,5$; $\alpha = \lambda/\mu = \lambda t_{об} = 1,8 < 2$. Тогда:

$$p_0 = (1 + 1,8 + 1,8^2/2! + 1,8^3/2!(2 - 1,8))^{-1} = 19^{-1} = 0,0525 \quad (5,25\%);$$

$$N_3 = \alpha = 1,8 \text{ канала};$$

$$K_3 = N_3/n = 0,9 \quad (90\%);$$

$$L_{оч} = 1,8^3(0,0525/(2 \cdot 2!(1 - 1,8/2)^2)) = 0,306/0,04 = 7,65 \text{ деталей};$$

$$W_{оч} = 7,65/0,9 = 8,54 \text{ мин.}$$

Предположим, что поступило предложение специализировать каждый станок только на один вид детали. В результате получили две одноканальные системы с бесконечной очередью:

$$n = 1; \lambda = \lambda_A + \lambda_B = 0,45 \text{ дет./мин.}; \mu = 1/t_{\text{обс}} = 0,5; \alpha = \lambda/\mu = \lambda t_{\text{обс}} = 0,9 < 1;$$

$$p_0 = 1 - \alpha = 0,1 \text{ (10\%); } K_3 = \alpha = 0,9 \text{ (90\%);}$$

$$L_{\text{оч}} = 0,9^2 (0,1/(1 - 0,9))^2 = 0,81/0,1 = 8,1 \text{ детали};$$

$$W_{\text{оч}} = 8,1/0,45 = 18 \text{ мин.}$$

Итак, как показывают результаты, специализация станков только ухудшила параметры эффективности системы: загрузка станков осталась прежней, но время ожидания в очереди возросло более чем в 2 раза. Число деталей в бункере также увеличилось более чем в 2 раза, так как в первом случае их общее среднее число не превышало 7,65 деталей, а во втором их среднее число как вида А, так и вида В стало по 8,1 детали, т.е. всего 16,2 детали. Объяснение полученным результатам кроется в случайной природе входного потока, поскольку поступление заготовок не является регулярным – в отдельные периоды больше поступает заготовок типа А, в другие – В. При отсутствии специализации станки «подстраховывают» друг друга, что приводит к более быстрому обслуживанию заявок. При специализации возможны ситуации, когда один из станков имеет большую очередь, а второй в это время простаивает. Поэтому применение специализации оправдано лишь в том случае, если это приводит к существенному сокращению времени обработки заявки. Например, в рассматриваемом случае специализация будет выгодна, если время обработки заготовки станет в среднем меньше 1,828 мин., т.е. снизится по крайней мере на 8,6%.



Вопросы для самопроверки

1. Сформулируйте основные причины появления неопределенностей. Какие из них являются субъективными, а какие – объективными?
2. Как различается неопределенность в зависимости от полноты и качества описания?
3. Какие основные причины возникновения неоднозначности вы знаете?
4. Приведите примеры лингвистических и физических неоднозначностей.
5. Как описывается неопределенность математически?
6. Приведите примеры математического описания неопределенностей для различных физических явлений.
7. Когда в задаче математического моделирования применяется стохастическое описание переменных?
8. В чем сущность аксиоматического подхода к построению теории вероятностей? Сформулируйте аксиомы А.Н. Колмогорова.

9. Можно ли ввести понятие плотности распределения вероятностей для дискретной случайной величины?

10. Объясните различие между модой, медианой и математическим ожиданием.

11. Какова вероятность того, что вторая карта, произвольно вытянута из колоды (36 карт), побьет первую?

12. Вероятность попадания в мишень каждого из двух стрелков равна 0,3. Стрелки стреляют по очереди, причем каждый делает по два выстрела. Попавший в мишень первым получает приз. Найти вероятность того, что приз стрелки получат.

13. Один раз бросают две игральные кости. Случайная величина S – сумма выпавших очков. Определить математическое ожидание и дисперсию S .

14. Случайная величина X задана функцией распределения

$$F_X(x) = \begin{cases} 0, & x < 1, \\ \frac{x-1}{2}, & x \in [1, 3], \\ 1, & x > 3. \end{cases}$$

Вычислите вероятности попадания случайной величины X в интервалы (1.5;2.5) и (2.5;3.5).

15. Вычислите математическое ожидание, дисперсию и постройте график плотности и функции распределения для случайной величины, равномерно распределенной на отрезке $[0,1]$.

16. В чем сущность метода равных вероятностей?

17. В чем сущность метода Монте-Карло?

18. Может ли простейший поток быть регулярным? Почему?

19. Может ли простейший поток быть нестационарным?

20. При каких допущениях процесс поступления больных в стационар можно считать простейшим потоком?

21. Наличие в системе очереди – это хорошо или плохо? Что плохо, если в СМО длинная очередь? Что плохо, если в СМО очередь образоваться не успевает?

22. Пусть интервал времени между поступлением заявок в СМО с отказами подчиняется равномерному закону. Можно ли для анализа системы применять соотношения Эрланга?

23. Чему равна вероятность отказа в обслуживании для системы с бесконечной очередью?

24. Почему при $\lambda = \mu$ длина очереди в СМО с бесконечной очередью может расти бесконечно? Поясните на примере.

25. Поясните на примере, что означает фраза «вероятность отказа в обслуживании составляет 0,01»?

26. Система имеет три состояния. Будет ли верно условие $p_1 = p_2 = p_3 = 0,5$?

27. В каких случаях выгоднее взять универсальный станок, а в каких – специализированный?



Задания для самостоятельного выполнения

1. Сформулируйте и запишите нечеткие отношения предпочтения между элементами множеств X и Y , Y и Z , если:

$X = \{\text{лес, кирпич, пенобетон}\}$, $Y = \{\text{железо, шлакобетон, брус}\}$,

$Z = \{\text{гипсоблоки, ракушечник, бетон}\}$.

Определите произведение этих отношений.

2. Приведите примеры и запишите числовые и нечисловые нечеткие и лингвистические переменные.

3. Для универсального множества $U = (a, b, c, d, e, f, g)$ и нечетких подмножеств

$A = (0,0/a; 0,3/b; 0,7/c; 1,0/d; 0,0/e; 0,2/f; 0,6/g)$,

$B = (0,3/a; 1,0/b; 0,5/c; 0,8/d; 1,0/e; 0,5/f; 0,6/g)$,

$C = (1,0/a; 0,5/b; 0,5/c; 0,2/d; 0,0/e; 0,2/f; 0,9/g)$

найдите:

а) $A \cap B$; б) $A \cup B$; в) $A \circ \bar{B}$; г) $(A \cup \bar{B}) \circ C$;

д) $\overline{(A \circ B)} \circ \bar{C}$; е) $A \oplus B$; ж) $\overline{A \oplus B}$; з) $(A \circ \bar{A}) \cup A$;

и) A/B ; к) $A - B$; л) $A \times B \times C$; м) $A \oplus C$.

4. Для нечетких подмножеств из упр. 3 вычислите:

а) $A \cdot B$; б) $B + C - A$; в) $A/C + B$; г) $(A - C)/B$;

д) $(A - B)/(A - C)$; е) $A \oplus B \oplus C$; ж) $A \cdot (B \oplus C)$.

5. Определите обычные подмножества α -уровня для нечеткого множества $A = (0,7/a; 0,5/b; 1,0/c; 0,2/d; 0,6/e)$, если:

а) $\alpha = 0,1$; б) $\alpha = 0,6$; в) $\alpha = 0,8$; г) $\alpha = 0,9$.

Запишите разложение нечеткого множества A .

6. Введите правила для определения понятий *чрезмерно* и *достаточно*.

7. Сравните нечеткие числа

$A = (0,2/3; 0,8/4; 0,4/5; 0,2/6)$ и $B = (0,1/3; 0,95/4; 0,3/5)$.

Для сравнения используйте все известные индексы ранжирования.

8. Концептуальная постановка задачи состоит в следующем: два истребителя из противоборствующих воздушных армий руководствуются стратегиями:

А: если снарядов *мало*, то вероятность поражения противника *малая*, иначе *не малая*.

В: если снарядов *не мало*, вероятность поражения противника *большая*, иначе *не большая*.

Известно, что:

мало снарядов = $(0,8/3; 0,4/15; 0,3/30)$,

малая вероятность = $(0,1/0,9; 0,5/0,5; 0,8/0,1)$,

большая вероятность = $(0,8/0,9; 0,5/0,5; 0,3/0,2)$.

Число снарядов *не очень мало*. Кто победит?

9. На участке имеется N однотипных станков. Производительность одного станка такова, что в среднем в течение смены на нем можно обработать μ деталей. Детали поступают на обработку партиями по M штук. Партии поступают неравномерно и их поток близок к пуассоновскому. Интенсивность поступления составляет λ партий в смену. Постройте граф состояний данной СМО и запишите систему уравнений Колмогорова для финальных вероятностей.

10. Мастерская по ремонту имеет складское помещение на M мест. В мастерской работают N мастеров, обеспечивающих среднее время ремонта T . Интенсивность поступления заявок λ . Если все мастера заняты и склад заполнен, то заявка отклоняется. Постройте граф состояний данной СМО и запишите систему уравнений Колмогорова для финальных вероятностей.

11. Определите требуемое число коек в стационаре больницы, если среднее время выздоровления одного больного составляет 21 день. Новые больные не принимаются, если все койки в стационаре заняты. Поток поступления больных близок к пуассоновскому с интенсивностью 3 человека в день. Вероятность отказа не должна быть выше 5%.

12. Определите среднюю длину очереди в кассу магазина, если среднее время обслуживания одного покупателя составляет 0,3 мин. Поток покупателей близок к пуассоновскому с интенсивностью 3 покупателя в минуту. Сколько необходимо установить касс, если интенсивность потока возрастет в 5 раз? Средняя длина очереди при этом не должна превышать 10 человек.

Глава 6

ЛИНЕЙНЫЕ И НЕЛИНЕЙНЫЕ МОДЕЛИ

Главное — это из множества проблем выбрать наиболее простые, решение которых позволит выработать допускающие обобщения концепции.

Д. Гильберт

Большинство рассмотренных в предыдущих главах примеров относятся к линейным моделям. В определенном смысле подобное построение книги следует истории развития естественных наук и математики, в которых вплоть до относительно недавнего времени (30–40-е годы XX века) в качестве объектов исследования практически безраздельно господствовали линейные процессы и явления (точнее, процессы и явления, описываемые линейными операторными уравнениями). Вероятно, обусловлено это было инструментарием, доступным научным сотрудникам тех времен, главными из которых являлись аналитические методы.

Аналитические методы, без сомнения, и в настоящее время являются одними из самых эффективных и притягательных для исследователей, однако круг доступных им задач весьма узок. Особенно больших усилий требуют аналитические решения нелинейных задач, которые все чаще возникают в различных научных областях, в технике и технологиях. Большинство нелинейных задач требуют для своего решения применения численных методов и мощных ЭВМ, доступ к которым ученые получили в 50–60-е годы XX века.

Следует отметить, что в ряде случаев проведение численных экспериментов позволяет исследователям «догадаться» о возмож-

ном виде аналитического решения, подсказывает направление поиска, подход, метод установления последнего. Кроме того, в последние годы весьма быстрыми темпами развиваются аналитические пакеты.

В этой главе на примерах будет показано, что могут и что не могут описать линейные модели явлений и процессов, происходящих в нашем, в общем-то, нелинейном мире. Для того чтобы распознать признаки линейности или нелинейности изучаемого явления, ученый должен владеть культурой (набором «штампов») нелинейной науки, с некоторыми понятиями и идеями которой мы попытаемся ознакомить читателя. Эти соображения определили некоторую специфику построения главы, где процедуре построения моделей уделяется меньше внимания, чем характерным признакам, сопровождающим линейные и нелинейные явления. Свои «рассказы» мы постарались выдержать в прикладном духе, чтобы первоначальное назначение математики, «созданной для достижения победы над противником», не отходило далеко на задний план. Авторы смогли написать только о малой части области науки, называемой «нелинейной наукой», более полную картину читатель сможет найти в популярных книгах [59, 60, 67] и многих других цитированных там работах.

6.1. О ЗАКОНЕ ГУКА И ГРАНИЦАХ ЛИНЕЙНОСТИ

Приступая к созданию физической теории, необходимо отбросить все предшествующие физические модели, а также основанную на них физическую «интуицию».

П. Дирак

С одной из первых линейных моделей – законом Гука

$$F = -kx, \quad (6.1)$$

где F – растягивающая сила, приложенная к концам образца; x – относительное перемещение его концов; k – некоторая постоянная для данного образца (материала), мы знакомимся, только начиная изучать физику. Этот закон в виде анаграммы «ceiiinosssttuv» был помещен Робертом Гуком на свободном месте в конце его работы о гелиоскопе в 1676 г. Два года позже в трактате «О

восстановительной способности или об упругости» анаграмма была им расшифрована: «*ut tensio sic vis*» – «какова сила, таково растяжение». Гук отмечал, что он впервые открыл этот закон еще в 1660 г., но от опубликования его удержало стремление защитить свое изобретение спиральных часовых пружин. Для их конструирования Гук использовал факт независимости частоты колебаний от их амплитуды, вытекающий из линейности связи (6.1).

Упругость в современном понимании означает существование однозначной монотонно возрастающей функции, связывающей напряжение $\sigma = F/S$ (S – площадь поперечного сечения образца) и деформацию $\epsilon = x/l$ (l – длина образца):

$$\sigma = f(\epsilon), \quad f(0) = 0. \quad (6.2)$$

Функция f в общем случае – нелинейная. Нелинейными упругими свойствами обладают, например, высокоэластичные резины. Резиновый шнур можно растянуть в десять раз ($\epsilon = 0,9$), а затем отпустить, после чего он восстановит свою длину. При таких больших степенях деформации не следует ожидать линейности связи $\sigma(\epsilon)$. Гук изучал длинные металлические проволоки, которые подвергал малым деформациям ($\epsilon \approx 0,001$), и обнаружить нелинейность реакции не удалось. Это означает, что при таких ϵ первый член разложения функции (6.2) (полагая ее аналитической) в степенной ряд

$$\sigma = \frac{\partial f}{\partial \epsilon} \epsilon + \frac{1}{2!} \frac{\partial^2 f}{\partial \epsilon^2} \epsilon^2 + \dots \quad (6.3)$$

значительно превосходит все остальные. Соотношение (6.2) в этом случае можно переписать в виде

$$\sigma = E\epsilon, \quad (6.4)$$

где E – модуль упругости *материала* при его одноосном растяжении-сжатии.

Открытие Гуком упругого закона вызвало весьма большой интерес научного сообщества; опыты Гука демонстрировались самому королю Англии Карлу II. Поспешно обобщая свои результаты, Гук заявлял: «Линейную зависимость можно наблюдать не только в этих телах, но и любых других упругих материалах, будь то металл, дерево, камни, спекшаяся глина, волос, шелк, кость, сухожилия, стекло и тому подобное».

Тем не менее не все ученые того времени разделяли оптимизм по поводу универсальности линейного закона. В 1687 г. Яков Бернулли провел опыты со скрипичной струной, изготовленной из кетгута (cat gut – кошачья кишка), показавшие нелинейность реакции. Готфрид Вильгельм Лейбниц, прочтя трактат Гука, в письме Христиану Гюйгенсу в 1691 г. высказал осторожное недоверие линейному закону и попросил его прислать эксперименты, которые тот проводил по этому предмету. Гюйгенс признался, что он согласен с результатами Гука, но только если пружины будут растянуты незначительно. И только в 1849 г. Британская королевская комиссия, назначенная для исследования применения железа для железнодорожных сооружений, «отменила» закон (6.4) в пользу параболической зависимости

$$\sigma = A\epsilon - B\epsilon^2.$$

Заключение комиссии основывалось на почти двадцатилетнем детальном изучении данных экспериментальных исследований на образцах из железа.

И все же факт, что при достаточно малых деформациях между последними и напряжениями в металлах была обнаружена линейная зависимость, привел к господству идей *линейной* теории упругости в течение почти трех столетий. Исключительное значение результатов этой теории подчеркивал Джеймс Фредерик Белл: «Если бы в 17 веке для твердых тел наблюдались исключительно нелинейные зависимости между напряжениями и деформациями, то большинство достижений в физике и технике, имевших место за прошедшие 200 лет, задержалось бы на несколько столетий» [9]. Механики-естествоиспытатели получили возможность продвигаться вперед, развивая линейную теорию упругости, а вслед за ней и математический аппарат, пригодный для построения моделей многих других явлений и процессов.

Из этой истории мы можем видеть, что линейная модель упругости занимает только определенную нишу в теории упругости. Этот факт в естествознании – достаточно общий. Как правило, любая линейная теория справедлива в ограниченных пространственных и временных рамках и при малой интенсивности воздействий на изучаемую систему. Например, архитектор без ущерба для своего дела может не принимать во внимание кривизну земной поверхности, поскольку размеры рассматриваемого им участка малы по сравнению с радиусом Земли (≈ 6400 км). Точно так же конструктор авиационной или космической техники в своей работе

вполне сможет обойтись ньютоновской классической механикой. не прибегая к теории относительности, поскольку последняя «вступает в силу» при скоростях, соизмеримых со скоростью света ($c \approx 300\,000$ км/с). При меньших скоростях v нелинейностью в основных законах можно пренебречь с точностью до малых порядка $(v/c)^2$.

Если же мы изучаем явление, не являющееся линейным в интересующих нас рамках, то, взяв в качестве первого приближения линейную модель, мы можем не получить даже качественного его описания. Нелинейность в таких явлениях принципиальна. Об этом свидетельствует целый ряд фактов, некоторые из них приведены в настоящей главе.

6.2. ПОЛЯ, СПЛОШНЫЕ СРЕДЫ И УРАВНЕНИЯ МАТЕМАТИЧЕСКОЙ ФИЗИКИ. ЛИНЕЙНЫЕ УРАВНЕНИЯ И ПРИНЦИП СУПЕРПОЗИЦИИ

*Часть, в некотором смысле, больше целого...
Общее утверждение должно доказываться легче,
чем любой его частный случай*

Дж. Сильвестр

Ранее мы говорили о линейной *функции*. Более богатые модели дают нам уравнения математической физики, которые также могут быть линейными или нелинейными. Можно сказать, что *математическая физика* — это наука, изучающая математические модели физических явлений. Уравнения математической физики — это чаще всего дифференциальные уравнения в частных производных, описывающие процессы в пространстве и времени. Таким образом, искомые функции должны иметь не менее двух аргументов; дифференциальные уравнения для функции одного аргумента относятся к классу обыкновенных дифференциальных уравнений. В общем случае мы рассматриваем процесс в области (возможно, неограниченной) трехмерного евклидова пространства, т.е. на такой области определено физическое поле, которое изменяется со временем. Одно из первых уравнений этого типа появилось благодаря Пьеру Симону Лапласу. Лаплас с недоверием относился к закону Всемирного тяготения Ньютона: как планеты, находясь на колоссальных дистанциях, могут действовать друг на друга? Лапласу казалась реальнее концепция контактной передачи

усилия (короткодействие) от одной планеты к другой, посредством эфира. Введя в рассмотрение эту субстанцию, поле, заполняющую трехмерное пространство, он получил для нее уравнение, которое сейчас называют *уравнением Лапласа*. Концепция эфира не подтвердилась, но оказалось, что это уравнение описывает многие стационарные процессы в сплошных средах: равновесие линейно-упругого тела, потенциальное течение вязкой жидкости, стационарные распределения электрического и температурного полей в теле и т.д.

Скажем несколько слов о модели *сплошной среды*. Обычно рассматриваемые нами поля определены в области геометрического пространства. Поле при этом оказывается как бы внешним, удобным для наблюдения и непрерывно заполняющим область математическим объектом. Механики имеют дело с полем материала, вложенного в это пространство, причем на этом материале обычно рассматриваются и другие поля (например, температурное). Физики работают, например, с электромагнитным или гравитационным полями, для которых «механический носитель» не требуется. Необходимо различать внешнее пространство и поле, поскольку последнее изменяется во времени (движется, течет, деформируется) относительно внешнего пространства, с которым (будем считать) жестко связана система отсчета наблюдателя. Оказывается, что реальные материальные среды вовсе не непрерывно заполняют пространство: при рассмотрении их вблизи они оказываются разрывными, состоящими из расположенных в пространстве зерен, гранул, молекул, атомов. Следовательно, и измеряемые на них физические поля с точки зрения внешнего геометрического континуума также не могут считаться непрерывными. Какую бы среду и на каких бы масштабах мы не рассматривали, объем, в котором сосредоточено собственно само вещество, оказывается много меньше объема тела. Но, несмотря на то, что все тела «состоят из пустоты», в практически малых объемах всегда оказывается огромное число частиц. Поэтому мы можем «размазать материю», т.е. полагать материальную среду и поле на ней непрерывными на масштабах, много больших по сравнению с характерными масштабами строения среды и осцилляций поля на ней. Применяя аппарат обычного дифференциального исчисления, мы должны осознавать, что работаем с уже осредненными полями (конечно, осредненное поле далеко не всегда полностью характеризует нам исходное поле). В механике модель сплошной среды применяют к газам, жидкостям и твердым телам.

Понятие линейности пришло в область дифференциальных уравнений, по-видимому, из геометрии. В линейном пространстве X , как известно, определены и замкнуты композиции $\alpha x + \beta y$ для любых элементов x, y из X и любых действительных чисел α, β . Рассмотрим пару n -мерных линейных пространств X, Y и отображение L , переводящее элементы X в Y . Оператор L называется линейным, если из соотношений $L(x_1) = y_1, L(x_2) = y_2$ для любых α, β следует $L(\alpha x_1 + \beta x_2) = \alpha y_1 + \beta y_2$.

Но можно рассматривать линейные пространства *функций*, являющиеся бесконечномерными линейными пространствами, а оператор L – дифференциальным. Тогда на дифференциальное уравнение

$$L(x) = f \tag{6.5}$$

можно смотреть как на оператор, отображающий одно функциональное пространство в другое или в себя, если рассматривать пространство бесконечно гладких (аналитических) функций. В качестве L будем рассматривать линейный дифференциальный оператор в частных производных, а в качестве функций $x = x(r), f = f(r)$ – функции аргумента r , четырехмерного пространства-времени.

Нетрудно убедиться, что множество решений однородного дифференциального уравнения $L(x) = 0$ образуют линейное пространство (возможно, состоящее из одной тождественно нулевой функции), являющееся подпространством X и называемое ядром оператора L . Обозначим x_0 произвольный элемент ядра. Очевидно, если x_f – какое-либо решение неоднородного уравнения (6.5), то общее решение (6.5) запишется в виде

$$x = x_0 + x_f.$$

Элемент x_0 конкретизируется из граничных и начальных условий задачи (иначе наша задача может иметь «слишком много» решений, если вообще они есть).

Если мы разобьем функцию в правой части (6.5) $f = f_1 + f_2$, то в силу линейности нашего оператора (в предположении одинаковых граничных условий) решение уравнения запишется в виде $x = x_0 + x_{f1} + x_{f2}$, где x_{f1} и x_{f2} – решения соответственно уравнений $L(x) = f_1$ и $L(x) = f_2$. Можно ожидать, что, представляя функцию в правой части разложением в ряд Фурье, получая частное реше-

ние для каждой его гармоник и составляя аналогичный ряд из частных решений, мы получим частное решение исходного уравнения. Этот путь, использующий *принцип суперпозиции*, действительно приводит к результату.

На этом пути можно пойти еще дальше — научиться решать уравнение, у которого в правой части стоит дельта-функция Дирака:

$$L(x) = \delta(r - r'), \quad (6.6)$$

где r' — некоторая фиксированная, а r — по прежнему произвольная точки пространства-времени. Физики «изобрели» и пользовались дельта-функцией для решения уравнений в частных производных давно. Саму дельта-функцию нет нужды определять, поскольку она используется только на промежуточных этапах решения данных уравнений, — достаточно знать, что для любой «допустимой» функции $f(r)$:

$$\int_{-\infty}^{+\infty} f(r) \delta(r - r') dr = f(r').$$

(О смысле «допустимости» говорить не будем, поскольку здесь нам важно показать идею метода.) Это делает возможным представить правую часть (6.5) как бесконечную интегральную сумму $f(r) = \int f(r') \delta(r - r') dr'$. Тогда, если $x(r) = G(r - r')$ — решение (6.6), то вследствие принципа суперпозиции решением (6.5) будет $x(r) = \int f(r') G(r - r') dr'$.

Функцию $G(r)$ называют функцией Грина оператора L (при определенных граничных условиях).

Как мы видим, для решения линейных уравнений математической физики разработан аппарат, использующий принцип суперпозиции. Продемонстрируем его, решив одномерное волновое уравнение

$$\frac{\partial^2 v}{\partial t^2} - a^2 \frac{\partial^2 v}{\partial x^2} = 0. \quad (6.7)$$

Это уравнение описывает распространение линейных волн $v(x, t)$ различной физической природы. Например, (6.7) можно

вывести как уравнение малых возмущений совершенного идеального баротропного газа (или свободных продольных или поперечных колебаний струны).

6.3. ВЫВОД ВОЛНОВОГО УРАВНЕНИЯ ИЗ ЗАКОНОВ МЕХАНИКИ

Нужно сковывать себя ограничениями — и тогда можно свободно выдумывать

У. Эко

Попутно с выводом волнового уравнения покажем, как строятся сплошносредние модели механики и физики. Экспериментально открыт факт, что физические процессы удовлетворяют законам сохранения массы, количества движения, энергии и др. Для сплошного деформируемого тела закон сохранения массы в локальной форме принимает вид

$$\frac{d\rho}{dt} + \rho \nabla \cdot \mathbf{v} = 0, \quad (6.8)$$

где $\rho = \rho(\mathbf{r}, t)$ — плотность; $\mathbf{r} = r_i \mathbf{k}_i$ — радиус-вектор произвольной геометрической точки; \mathbf{k}_i , $i = 1, 2, 3$ — декартов ортонормированный репер; r_i , $i = 1, 2, 3$ — координаты радиус-вектора,

$\nabla \rightarrow \mathbf{k}_i \frac{\partial}{\partial r_i}$ — набла-вектор (оператор Гамильтона), $\mathbf{v} = \mathbf{v}(\mathbf{r}, t)$ — векторное поле скоростей бесконечно малых материальных частиц (материальных точек). (Очевидно, каждая материальная точка имеет только три степени свободы). Поскольку рассматриваемое тело движется относительно внешней системы координат, для каждой его *материальной точки* \mathbf{R} имеет место закон движения $\mathbf{r} = \mathbf{r}(\mathbf{R}, t)$.

В отличие от теоретической механики в сплошном теле материальных точек оказывается настолько много, что для их идентификации необходим континуум точек внешней системы координат: $\mathbf{r}(\mathbf{R}, 0) = \mathbf{R}$. Таким образом, \mathbf{R} есть радиус-вектор геометрической точки, в которой находилась материальная точка в начальный момент времени. Скорость $\mathbf{v} = \mathbf{v}(\mathbf{r}, t)$ материальной точки, находящейся в момент времени t в месте \mathbf{r} , определяется как $\mathbf{v} = \frac{d\mathbf{r}}{dt}$. Про-

изводя дифференцирование произвольного поля \mathbf{a} , можно получить:

$$\left. \frac{d\mathbf{a}}{dt} \right|_R = \left. \frac{\partial \mathbf{a}}{\partial t} \right|_r + \mathbf{v} \cdot \nabla \mathbf{a}. \quad (6.9)$$

Рассмотрим, например, сплошную среду, *свободно* движущуюся вдоль прямой Ox со скоростью $v = v(x, t)$. Закон движения любой (материальной) точки этой среды $x = \varphi(t) = x_0 + ut$ (u — ее скорость)

удовлетворяет второму закону Ньютона: $\frac{\partial^2 \varphi}{\partial t^2} = 0$. По определению, скорость материальной точки $u = \frac{\partial \varphi}{\partial t} = v(\varphi(t), t)$; дифференцируя по t , получаем уравнение, которому удовлетворяет поле

$v = v(x, t)$: $\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = 0$.

Заметим, что за появление нелинейности $v v_x$ (здесь v_x означает частную производную скорости по координате x) отвечает переход от материальных координат к пространственным (внешним) координатам. Пространственные координаты \mathbf{r} называются *эйлеровыми*, материальные же координаты \mathbf{R} , позволяющие следить за судьбой отдельных материальных точек, не меняющиеся для каждой материальной частицы в течение всего исследуемого процесса, называются *лагранжевыми*.

Балансовое уравнение для плотности массы (6.8) с учетом (6.9) запишется в виде:

$$\frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla \rho + \rho \nabla \cdot \mathbf{v} = 0. \quad (6.10)$$

Другой закон сохранения, необходимый сейчас нам, — количества движения — в локальной форме принимает вид

$$\rho \frac{d\mathbf{v}}{dt} = \nabla \cdot \boldsymbol{\sigma} + \rho \mathbf{f}, \quad (6.11)$$

где \mathbf{f} — вектор массовых сил; $\boldsymbol{\sigma}$ — тензор напряжений. В правой части (6.11) фигурируют силы, действующие на материальную точку, занимающую положение \mathbf{r} . Тензор напряжений отвечает за

контактное силовое воздействие на точку со стороны окружающего материала, и его следует воспринимать как линейный оператор, ставящий в соответствие проходящей через материальную точку плоскости с нормалью \mathbf{n} вектор поверхностной силы \mathbf{t} : $\mathbf{t} = \mathbf{n} \cdot \boldsymbol{\sigma}$. Левая часть (6.11) соответствует скорости изменения количества движения материальной точки; равенство нулю этой величины в каждой точке сплошного тела описывает его статику. Применяя (6.9) к последнему уравнению, получаем

$$\frac{d\mathbf{v}}{dt} + \mathbf{v} \cdot \nabla \mathbf{v} = \frac{1}{\rho} \nabla \cdot \boldsymbol{\sigma} + \mathbf{f}. \quad (6.12)$$

Одного рассмотрения статики или динамики сплошного деформируемого тела обычно оказывается недостаточно для определения физических полей в этом теле (число неизвестных полевых функций оказывается больше числа уравнений (6.12)). Для получения дополнительных уравнений рассматривают геометрическую или кинематическую сторону вопроса, в результате чего появляются дополнительные уравнения. Полученные динамические и кинематические соотношения остается связать замыкающими уравнениями, которые являются наиболее неформализуемыми и зависящими от особенностей рассматриваемой сплошной среды. Эти замыкающие уравнения называются *определяющими соотношениями*. Можно сказать, что они являются «центральными» уравнениями в любой модели.

Будем считать, что среда – идеальный совершенный газ. Первое условие («идеальность») означает $\boldsymbol{\sigma} = -p\mathbf{I}$, p – гидростатическое давление; \mathbf{I} – тензорная единица. Второе условие требует, чтобы газ подчинялся уравнению Менделеева–Клапейрона $p = R\rho T$, где T – абсолютная температура; R – универсальная газовая постоянная. Последнее уравнение и есть определяющее соотношение.

Сузим еще круг рассматриваемых процессов: будем полагать, что последнее принимает форму функции

$$p = p(\rho). \quad (6.13)$$

Такие процессы, называемые *баротропными*, существуют: например, для рассматриваемой среды уравнения изотермического и адиабатического процессов записываются в форме (6.13). Уравнения динамики (6.12) (при $\mathbf{f} \equiv 0$), сохранения массы (6.10) и состояния (6.13) в одномерном случае запишутся в виде

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} + \frac{1}{\rho} \frac{\partial p}{\partial x} = 0, \quad \frac{\partial \rho}{\partial t} + v \frac{\partial \rho}{\partial x} + \rho \frac{\partial v}{\partial x} = 0, \quad p = p(\rho) \quad (6.14)$$

(три уравнения, три неизвестные функции v , p , ρ независимых переменных x и t). Система (6.14) содержит одно алгебраическое и два нелинейных дифференциальных уравнения. Эти дифференциальные уравнения с коэффициентами, зависящими от искомым функций, однако, линейны относительно старших производных (такие уравнения называются *квазилинейными*). Несмотря на название уравнения (6.14) не являются линейными и обладают богатой динамикой. Легко проверить, что тройка функций

$$v \equiv 0, \quad \rho \equiv \rho_0, \quad p \equiv p_0 = p(\rho_0) \quad (6.15)$$

является решением системы (6.14). Данное решение соответствует состоянию покоя газа, причем значение давления (либо плотности) необходимо определить из начальных условий. Получим дифференциальные уравнения эволюции малых возмущений состояния покоя системы (6.14) [63]. Путь, который предлагается, — достаточно общий для линеаризации нелинейных дифференциальных уравнений вблизи какого-либо их решения.

Определим возмущенное решение (6.15), оставляя для возмущений такие же обозначения, как и у возмущенных полей,

$$v = v, \quad \rho = \rho_0 + \rho, \quad p = p_0 + p. \quad (6.16)$$

Возмущения v , p , ρ будем считать малыми. Из (6.14)–(6.16) следует, что

$$\frac{\partial v}{\partial t} + \frac{1}{\rho_0} \frac{\partial p}{\partial \rho} \frac{\partial \rho}{\partial x} = 0, \quad \frac{\partial \rho}{\partial t} + \rho_0 \frac{\partial v}{\partial x} = 0 \quad (6.17)$$

с точностью до членов второго порядка. Функцию (6.13) разложим в ряд вблизи ρ_0 : $p(\rho_0 + \rho) = p(\rho_0) + p'(\rho_0)\rho + O(\rho^2)$. Эксперимент показывает, что $p'(\rho_0) > 0$, поэтому обозначим $p'(\rho_0) = a^2$. Почленным дифференцированием уравнений и удалением подобных систему можно привести к двум уравнениям вида (6.7) для функций v и ρ тех же независимых переменных (x , t):

$$\frac{\partial^2 v}{\partial t^2} = a^2 \frac{\partial^2 v}{\partial x^2}, \quad \frac{\partial^2 p}{\partial t^2} = a^2 \frac{\partial^2 p}{\partial x^2}.$$

Окончательно наша модель приняла вид системы дифференциальных уравнений в частных производных с *постоянными* коэффициентами, т.е. чисто линейных уравнений. Это – обычное дело, если возмущается однородное и стационарное решение. Информация о возмущаемом решении содержится в коэффициенте a .

6.4. РЕШЕНИЕ ВОЛНОВОГО УРАВНЕНИЯ МЕТОДОМ ФУРЬЕ

Я готов биться об заклад, что принцип суперпозиции будет стоять в веках.

Р.Фейнман

Приступим к решению уравнения (6.7)

$$\frac{\partial^2 v}{\partial t^2} - a^2 \frac{\partial^2 v}{\partial x^2} = 0,$$

удовлетворяющему граничным условиям

$$v(0, t) = v(l, t) = 0 \tag{6.18}$$

и начальным условиям

$$v(x, 0) = \phi(x), \quad \frac{\partial v}{\partial t}(x, 0) = \psi(x). \tag{6.19}$$

Эта задача соответствует динамическому поведению газа в тонком канале длиной l (или колебаниям струны с закрепленными концами). Задавая функции (6.19) начального возмущения состояния газа, не надо забывать, что (6.7) справедливо для малых возмущений. Найдем нетривиальное классическое решение краевой задачи (6.7),(6.18),(6.19). Доказано, что это решение существует и единственно. Будем, следуя Эйлеру, искать его в виде, обычном для линейных уравнений математической физики с постоянными коэффициентами:

$$v \sim \exp(kx + ct), \tag{6.20}$$

где k и c могут быть комплексными. Решения типа (6.20) называют *нормальными модами*, или нормальными колебаниями. *Характеристическое (дисперсионное) уравнение* принимает вид

$$c^2 - a^2 k^2 = 0,$$

откуда

$$c = \pm ak. \quad (6.21)$$

Число решений характеристического уравнения оказалось равным порядку производной по t , т.е., вместо (6.20) надо записать

$$v = \exp(ct) [A \exp(cx/a) + B \exp(-cx/a)].$$

Применяя к последнему равенству граничные условия, находим, что при действительном c имеется только тривиальное решение $v \equiv 0$ (покажите). При $c = i\gamma$ условие существования нетривиальной пары A, B примет вид $\sin(\gamma l/a) = 0$, откуда $\gamma = \pi n a/l$, $n = 0, \pm 1, \pm 2, \dots$. С учетом решения характеристического уравнения функция (6.20) приобретает вид

$$v_n = [F_n \sin(\pi n a t/l) + G_n \cos(\pi n a t/l)] \sin(\pi n x/l), \quad n = 0, 1, 2, \dots \quad (6.22)$$

Эти функции суть частные решения нашей краевой задачи. В силу принципа суперпозиции функция

$$v = \sum_{n=0}^{\infty} [F_n \sin(\pi n a t/l) + G_n \cos(\pi n a t/l)] \sin(\pi n x/l) \quad (6.23)$$

также будет ее решением. Подчиним коэффициенты линейной комбинации (6.23) начальным условиям (6.19):

$$\phi(x) = \sum_{n=0}^{\infty} G_n \sin(\pi n x/l), \quad \psi(x) = \sum_{n=0}^{\infty} F_n \frac{\pi n a}{l} \sin(\pi n x/l).$$

Но это есть разложения начальных функций $\phi(x)$ и $\psi(x)$ в ряд Фурье. Коэффициенты этого ряда G_n и $H_n = F_n \frac{\pi n a}{l}$ находятся как

$$G_n = \frac{2}{l} \int_0^l \phi(\xi) \sin(\pi n \xi / l) d\xi, \quad H_n = \frac{2}{l} \int_0^l \psi(\xi) \sin(\pi n \xi / l) d\xi.$$

Как известно, коэффициенты ряда Фурье убывают, и тем быстрее, чем более гладкой является функция. Окончательно решение задачи примет вид

$$v = \sum_{n=0}^{\infty} \left[\frac{2}{\pi n} \int_0^l \psi(\xi) \sin(\pi n \xi / l) d\xi \sin(\pi n a t / l) + \right. \\ \left. + \frac{2}{l} \int_0^l \phi(\xi) \sin(\pi n \xi / l) d\xi \cos(\pi n a t / l) \right] \sin(\pi n x / l).$$

Данный метод, называемый *методом Фурье*, применяется для таких линейных дифференциальных уравнений с постоянными коэффициентами, когда существует дисперсионное соотношение (т.е. для фиксированного k ограничено c), а система частных решений образует полный базис, т.е. способна представить начальное условие задачи. Пара $k, c(k)$ — одна базисная функция (мода), а суммирование по всем модам (по k) определяет решение дифференциального уравнения.

Обратим внимание на то, что частные решения (6.22) краевой задачи представимы в виде

$$v = T(t)X(x), \quad (6.24)$$

т.е. начальный профиль $T(0)X(x)$ со временем только растягивается или сжимается, оставаясь подобным самому себе (автомодельным). Подставляя представление (6.24) в (6.7), получаем

$$T'X - a^2TX'' = 0,$$

откуда после деления на a^2TX находим

$$\frac{1}{a^2} \frac{T'}{T} = \frac{X''}{X} = \lambda. \quad (6.25)$$

Здесь λ , очевидно, не зависит ни от t , ни от x . Одно из уравнений системы (6.25)

$$X'' - \lambda X = 0 \quad (6.26)$$

вместе с условиями

$$X(0) = X(l) = 0, \quad (6.27)$$

вытекающими из (6.18), представляет собой частный случай известной в математической физике задачи Штурма-Лиувилля – задачи нахождения нетривиального (т.е. ненулевого) решения обыкновенного дифференциального уравнения (6.26), удовлетворяющего краевым условиям (6.27). Те значения λ , при которых эта задача имеет нетривиальные решения, называются собственными числами (спектром) задачи, а соответствующие им решения – собственными функциями задачи.

Как видим, задача Штурма-Лиувилля напоминает задачу нахождения собственных векторов t симметричного тензора второго ранга X над конечномерным пространством:

$$(X - \lambda I) \cdot t = 0, \quad t \neq 0.$$

Формальная замена линейного *симметричного* оператора X другим линейным *самосопряженным* оператором – координатной частью линейного дифференциального оператора с краевыми условиями, и приводит к задаче Штурма-Лиувилля (напомним, что функции также являются векторами, т.е. элементами линейного пространства).

Известно, что эта задача приводит к счетному множеству собственных чисел, все они вещественные, среди них имеется наименьший, а точка сгущения расположена в бесконечности, т.е.

$\lim_{n \rightarrow \infty} \lambda_n = +\infty$. Каждому собственному значению соответствует единственная нормированная собственная функция, а полный набор собственных функций образует ортонормированный базис пространства непрерывных функций на отрезке (точнее говоря, речь идет об ортонормированном базисе в гильбертовом пространстве, всюду плотном в $C[a, b]$). Решения задачи (6.26)–(6.27) имеют вид

$$X_n(x) \sim \sin(\sqrt{\lambda_n} x), \quad \lambda_n = \left(\frac{\pi n}{l}\right)^2, \quad n = 0, 1, 2, \dots \quad (6.28)$$

Собственные функции $\sin(\sqrt{\lambda_n} x)$, $n = 0, 1, 2, \dots$ еще называются *модами*, или *гармониками*.

Решая второе уравнение системы (6.25) с учетом краевых условий, получим дискретный спектр и представление:

$$T_n(t) = F_n \sin(\sqrt{\lambda_n} at) + G_n \cos(\sqrt{\lambda_n} at). \quad (6.29)$$

Обратим внимание на то, что элемент спектра определяет не только форму гармоники (6.28), но и частоту колебаний $\sqrt{\lambda_n} a$, называемую *собственной частотой* этой гармоники. Из (6.24) следует, что если выбрать начальное возмущение (6.19) соответствующим (6.28), то в системе будут происходить колебания газа с n -й собственной частотой.

Рассмотренная нами задача также описывает поперечные колебания линейно-упругой струны, шарнирно закрепленной по концам. Заметим, что амплитуда колебаний любой гармоники не зависит от частоты ее колебаний. Этот факт является общим для линейных волновых и колебательных моделей (и упоминался нами в начале главы в связи с открытием Гука).

Задача на нахождение собственных значений и собственных функций обобщается на граничные задачи, поставленные для трехмерных тел. Удивительно, что тело, некоторые свойства которого описываются линейным дифференциальным уравнением в частных производных, имеет набор собственных частот и функций («паспортных данных»), зависящих только от оператора, формы тела и наложенных на его границах условий. Важно, что собственные функции задачи образуют базис некоторого функционального векторного пространства, по которому можно разложить решение (вопрос, всегда ли – оставим открытым). Например, для решения задачи со сложным (но линейным) оператором и некоторыми граничными условиями иногда разумнее решить сначала задачу с более простым оператором и теми же условиями, а полученное разложение далее использовать для основной задачи. Отметим, что если мы решаем линейные уравнения для бесконечных областей, спектр оказывается непрерывным и суперпозиция нормальных мод в представлении начальных условий сводится к интегральной сумме по k .

6.5. О ХАРАКТЕРИСТИКАХ УРАВНЕНИЙ МАТЕМАТИЧЕСКОЙ ФИЗИКИ. РЕШЕНИЕ ВОЛНОВОГО УРАВНЕНИЯ МЕТОДОМ ДАЛАМБЕРА

Долог путь поучений, короток и успешен путь примеров.

Сенека Младший

Здесь нам понадобится понятие характеристик линейного дифференциального оператора. Начнем с простого примера, использующего понятия эйлеровых и лагранжевых координат (см. разд. 6.3). Пусть каждая частица ξ «одномерной» сплошной среды имеет некоторый локальный параметр ϑ , не изменяющийся со временем:

$$\frac{d\vartheta}{dt} = 0. \quad (6.30)$$

Согласно формуле (6.9),

$$\frac{\partial \vartheta}{\partial t} + v(x, t) \frac{\partial \vartheta}{\partial x} = 0. \quad (6.31)$$

Поле скоростей $v(x, t)$ будем считать известным, а $\vartheta(x, t)$ — разыскиваемым. Дифференциальное уравнение (6.31), линейное относительно старших производных с коэффициентами, зависящими только от координаты x и времени t , принадлежит к классу почти линейных уравнений в частных производных первого порядка (по терминологии [42]). Как любое дифференциальное уравнение, оно имеет бесконечное множество решений (в некотором функциональном пространстве), и чтобы выделить из них определенное, следует дополнить это уравнение начальным условием

$$\vartheta(x, 0) = \vartheta_0(x). \quad (6.32)$$

Такая задача называется начальной задачей, или задачей Коши.

Однако решать задачу (6.31)–(6.32) легче всего, принимая во внимание ее происхождение (физическое содержание). Из (6.30) вытекает, что $\vartheta = \vartheta(\xi)$, откуда с учетом (6.32) имеем

$$\vartheta = \vartheta_0(\xi). \quad (6.33)$$

Мы получили (6.33) – решение с точки зрения Лагранжа. Для нахождения поля $\vartheta(x, t)$ нам придется решить еще одно дифференциальное уравнение

$$\frac{dx}{dt} = v(x, t) \quad (6.34)$$

с начальным условием

$$x(0) = \xi. \quad (6.35)$$

Решение этой задачи Коши $x = x(\xi, t)$ – закон движения частицы с именем ξ – в силу свойств континуума можно обратить $\xi = \xi(x, t)$, и окончательно решение с точки зрения Эйлера примет вид

$$\vartheta = \vartheta_0(\xi(x, t)).$$

Интегральные линии уравнения (6.34) на плоскости x, t , т.е. траектории движения частиц при заданном поле скоростей $v(x, t)$, называются характеристиками уравнения с частными производными (6.31) [29]; это не что иное, как координатные линии $\xi = \text{const}$. Очевидно, поле характеристик плотно заполняет плоскость x, t . Результат (6.33) можно понимать так: решения (6.31) должны быть постоянными вдоль каждой из его характеристик. Таким образом, вдоль каждой своей характеристики уравнение в частных производных (6.31) превращается в обыкновенное дифференциальное уравнение.

Посмотрим, сможет ли это дать что-либо для понимания системы (6.17), которую перепишем в виде

$$\frac{\partial v}{\partial t} + \frac{a^2}{\rho_0} \frac{\partial \rho}{\partial x} = 0, \quad \frac{\partial \rho}{\partial t} + \rho_0 \frac{\partial v}{\partial x} = 0. \quad (6.36)$$

Пусть нам известно, что эта система имеет гладкое решение в некоторой области G плоскости x, t . Выберем в этой области точку x_0, t_0 и проведем через нее гладкую кривую γ . Предположим, что нам почему-либо известны значения полей v, ρ вдоль этой кривой. Поставим задачу определения значений этих полей вблизи кривой. Конечно, эта задача есть задача Коши. Принятые требования означают, что нам известны производные полей v, ρ вдоль кривой γ в точке x_0, t_0 :

$$\frac{dt}{d\gamma} \frac{\partial v}{\partial t} + \frac{dx}{d\gamma} \frac{\partial v}{\partial x} = \frac{dv}{d\gamma}, \quad \frac{dt}{d\gamma} \frac{\partial \rho}{\partial t} + \frac{dx}{d\gamma} \frac{\partial \rho}{\partial x} = \frac{d\rho}{d\gamma}. \quad (6.37)$$

Условие разрешимости системы (6.36)–(6.37) относительно $\frac{\partial v}{\partial t}$, $\frac{\partial v}{\partial x}$, $\frac{\partial \rho}{\partial t}$, $\frac{\partial \rho}{\partial x}$ записывается как

$$\begin{vmatrix} 1 & 0 & 0 & \frac{a^2}{\rho_0} \\ 0 & \rho_0 & 1 & 0 \\ \frac{dt}{d\gamma} & \frac{dx}{d\gamma} & 0 & 0 \\ 0 & 0 & \frac{dt}{d\gamma} & \frac{dx}{d\gamma} \end{vmatrix} = a^2 \left(\frac{dt}{d\gamma} \right)^2 - \left(\frac{dx}{d\gamma} \right)^2 \neq 0. \quad (6.38)$$

Это условие в силу принятой нами гипотезы существования решения вблизи кривой заведомо выполняется. Таким образом, из (6.36)–(6.37) находим первые производные, затем точно так же последовательно можно найти смешанные производные высших порядков, что и позволит построить искомое решение. Но если γ имеет уравнение $x \pm at = \text{const}$, то требование о существовании решения может не иметь смысла. Из линейной алгебры известно, что в данном случае решение возможно, если ранг матрицы в (6.38) совпадает с рангом расширенной матрицы

$$\begin{bmatrix} 1 & 0 & 0 & a^2/\rho_0 & 0 \\ 0 & 0 & 1 & \rho_0 & 0 \\ \frac{dt}{d\gamma} & \frac{dx}{d\gamma} & 0 & 0 & \frac{dv}{d\gamma} \\ 0 & 0 & \frac{dt}{d\gamma} & \frac{dx}{d\gamma} & \frac{d\rho}{d\gamma} \end{bmatrix}.$$

Например, должен равняться нулю определитель матрицы:

$$\begin{vmatrix} 0 & 0 & 0 & a^2/\rho_0 \\ 0 & \rho_0 & 1 & 0 \\ \frac{dv}{d\gamma} & \frac{dx}{d\gamma} & 0 & 0 \\ \frac{d\rho}{d\gamma} & 0 & \frac{dt}{d\gamma} & \frac{dx}{d\gamma} \end{vmatrix} = \frac{a^2}{\rho_0} \left(-\rho_0 \frac{dt}{d\gamma} \frac{dv}{d\gamma} - \frac{d\rho}{d\gamma} \frac{dx}{d\gamma} \right) = 0.$$

Вдоль характеристик $x \pm at = \text{const}$ это требование равносильно выполнению соотношений $d\left(v \mp \frac{a}{\rho_0} \rho\right) = 0$. В этом примере приходим к понятию характеристик как к кривым, на которых могут не иметь смысла начальные условия задачи Коши. Характеристиками оказались два семейства прямых $x \pm at = \text{const}$, образующих координатную сетку в плоскости x, t . Эта система координат обладает особыми свойствами: на каждой ее координатной кривой сохраняется постоянной величина соответствующая $v \mp \frac{a}{\rho_0} \rho$. Запишем систему (6.36) в новой системе координат $y = x + at, z = x - at$.

$$\frac{\partial v}{\partial y} + \frac{a}{\rho_0} \frac{\partial \rho}{\partial y} = 0, \quad -\frac{\partial v}{\partial z} + \frac{a}{\rho_0} \frac{\partial \rho}{\partial z} = 0$$

или после преобразований:

$$\frac{\partial^2 v}{\partial y \partial z} = 0, \quad \frac{\partial^2 \rho}{\partial y \partial z} = 0.$$

Любое из уравнений системы легко решается: $(v_y)_z = 0$ означает, что функция v_y постоянна вдоль прямых $y = \text{const}$ на плоскости y, z . Итак, $v_y = \theta(z)$, а значит

$$v = \int \theta(z) dy = f(z) + g(y) = f(x + at) + g(x - at). \quad (6.39)$$

Таким образом, имеем общее решение системы (6.7) с точностью до двух произвольных функций, аргументами которых слу-

жат пространственно-временные комплексы $x \pm at$. Добавляя к системе начальные условия, для удобства принятые в виде $v(x, 0) = \varphi(x)$, $\frac{\partial v}{\partial t}(x, 0) \equiv 0$, а также полагая среду бесконечной $-\infty < x < +\infty$, легко найти f и g и решение задачи

$$v(x, t) = \frac{1}{2}(\varphi(x - at) + \varphi(x + at)). \quad (6.40)$$

Данное решение называется решением типа бегущей волны. Точнее, в решении (6.40) присутствуют две волны одинакового профиля, разбегающиеся в разные стороны от «начальной волны» $\varphi(x)$. Существование такого решения является следствием трансляционной инвариантности задачи по временной и пространственной координатам. Это означает, что существуют два сдвига Δx , $a\Delta t$, переводящие решение само в себя при фиксированном a . То есть снова сталкиваемся с автомодельным решением: профиль каждой из пары волн (6.39) не меняется, меняется только их расположение в пространстве.

Обратим внимание на то, что величина $a = \sqrt{p'(\rho_0)}$ есть скорость распространения малых возмущений в газе, что видно из (6.40). Учитывая, что константа a формально присутствует в волновом уравнении, описывающем различные физические волновые процессы, ее нужно трактовать как скорость распространения волн. Здесь нельзя не вспомнить об одном глубоком выводе, сделанном Дж. К. Максвеллом из открытых им уравнений электромагнитной динамики.

6.6. УРАВНЕНИЯ МАКСВЕЛЛА

*Не бог ли эти знаки начертал?
Таинственен их скрытый дар!*

Гете

В отличие от механического, электромагнитное поле не требует для себя какого-либо носителя типа твердого, жидкого или газообразного тела, оно само является материальным полем и может распространять возмущения в вакууме и с конечной скоростью. Эти идеи в середине прошлого века в свете господствующей

амперовской электродинамики еще казались абсурдными. Практические успехи «старой» электродинамики впечатляли: к тому времени уже действовал телеграфный кабель через Атлантический океан. Однако она вследствие своей внутренней противоречивости напоминала Максвеллу вавилонскую башню. Он не мог принять мгновенного действия на расстоянии (дальнодействия), заложенного, например, в законе Кулона. Возможно, его воодушевляло то, что в 1849 г. А. Физо нашел чудовищное (но конечное) значение для скорости света – более $313 \cdot 10^3$ км/с. Максвелл, еще со студенчества воодушевленный работами М. Фарадея, впервые введшего абстрактные силовые линии в пространстве, придерживался точки зрения, что главным действующим лицом должно являться электромагнитное поле в пространстве. Эту смелую философию он подкреплял аргументами: *«Кому свойства воздуха незнакомы, тому передача силы посредством этой невидимой среды будет казаться столь же непонятной, как и всякий другой пример действия на расстоянии...»*. Отказавшись от некоторых законов, сохранив и обобщив другие, наделив каждую математическую операцию физическим смыслом и в то же время дав каждой физической величине четкую математическую характеристику, Максвелл пришел к своим знаменитым уравнениям электромагнитного поля:

$$\nabla \cdot \mathbf{D} = \rho, \quad \nabla \times \mathbf{E} = -\frac{d\mathbf{B}}{dt}, \quad \nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{H} = \mathbf{j} + \frac{d\mathbf{D}}{dt}, \quad (6.41)$$

где \mathbf{E} – вектор напряженности электрического поля; \mathbf{D} – вектор электрической индукции; \mathbf{B} – вектор магнитной индукции; \mathbf{H} – вектор напряженности магнитного поля; ρ – плотность заряда; \mathbf{j} – ток смещения.

Для уяснения смысла некоторых понятий Максвелл пользовался механическими моделями типа изображенной на рис. 6.1. В «Трактате об электричестве и магнетизме», главном труде Максвелла, такие схемы он уже не использовал (с точки зрения поля они бессмысленны), но именно они позволили ему многое понять. Нужно заметить, что в век Максвелла («век пара») доминировали идеи и методы механики, так что уже после выхода «Трактата» многие пытались выяснить механическое содержание его уравнений. Людвиг Больцман думал, что Максвелл имеет механические модели, которыми он обосновывал свои уравнения, но прячет. После смерти Максвелла Больцман, который им восхищался, излагая на лекциях максвелловскую теорию, предварял изло-

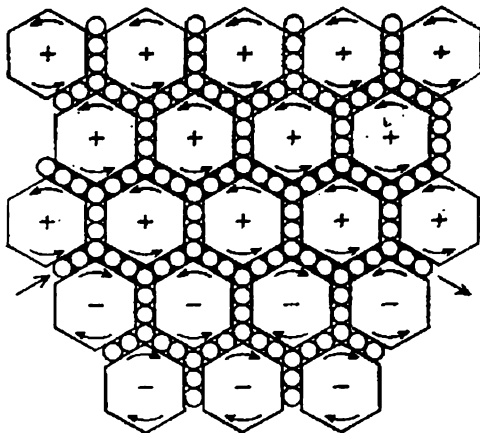


Рис. 6.1. Механическая модель электромагнитного поля [52]

жение эпиграфом из «Фауста»: «Я должен пот тяжелый лить, чтоб научить тому, чего не понимаю сам». Известно, что Максвелл пользовался представлениями Гельмгольца о вихревых движениях идеальной жидкости, но заслуживает изумления смелая аналогия «неосязаемого» континуума с «осязаемым». До сих пор термин «механическое» механики часто заменяют на «материальное», хотя электромагнитные поля материальны ничуть не меньше.

Система уравнений (6.41) — не замкнута, в ней не хватает определяющих соотношений. В простейшем (и имеющем смысл) случае можно записать

$$\mathbf{D} = \mathbf{E}, \quad \mathbf{B} = \mathbf{H}, \quad \mathbf{j} \equiv \mathbf{0}. \quad (6.42)$$

Применяя к (6.41) операцию $\nabla \times$ и используя (6.42) и (6.41), получаем

$$\nabla \times \nabla \times \mathbf{H} = -\frac{d^2 \mathbf{H}}{dt^2}.$$

Далее, пользуясь известным тождеством $\nabla \times \nabla \times \mathbf{a} = \Delta \mathbf{a} - \nabla(\nabla \cdot \mathbf{a})$ и (6.41) вместе с (6.42), окончательно имеем

$$\Delta \mathbf{H} = \frac{d^2 \mathbf{H}}{dt^2} \quad (6.43)$$

($\Delta \equiv \nabla \cdot \nabla$ — оператор Лапласа). Точно такое же уравнение получается и для E .

Но (6.43) — это волновое уравнение в трехмерном случае. Это привело Максвелла к предсказанию существования электромагнитных волн. Если попытаться вычислить скорость распространения этих волн, то получится, что она равна отношению электростатической и электромагнитной единиц заряда. Однако эта величина известная! Ее независимо измерили Кольрауш, Вебер и Столетов и равна она примерно 300 000 км/с, т.е. скорости света (ввиду такой высокой скорости распространения электромагнитные воздействия и принимали за мгновенные).

Максвелл смог придать этому случайному совпадению, «любопытному курьезу» по его словам, глубокий физический смысл: свет есть не что иное, как электромагнитные волны. Последующая всесторонняя экспериментальная проверка уравнений Максвелла блестяще их подтвердила. Г. Герц писал об этих уравнениях: «Трудно избавиться от чувства, что эти математические формулы живут независимой жизнью и обладают своим собственным интеллектом, что они мудрее, чем мы сами, мудрее даже, чем их первооткрыватели, и что мы извлекаем из них больше, чем было заложено в них первоначально» [52].

Молодой голландский физик Гендрик Антуан Лоренц впоследствии пытался применить теорию Максвелла к движущимся телам, и из этого труда появились «преобразования Лоренца» — важнейшая предпосылка создания теории относительности. Трактат Максвелла об электричестве и магнетизме открыл эпоху новой физики.

6.7. О КЛАССИФИКАЦИИ КВАЗИЛИНЕЙНЫХ СИСТЕМ

Нет ничего практичнее хорошей теории.

Л.Больцман

Понятие характеристик позволяет классифицировать уравнения математической физики. Рассмотрим систему уравнений в частных производных первого порядка, символически представимую в виде

$$\frac{\partial u}{\partial t} + C \frac{\partial u}{\partial x} = f. \quad (6.44)$$

Здесь u и f есть n -мерные векторы неизвестных и известных функций соответственно; C – матрица коэффициентов, зависящих от x, t . Эта система – почти линейная [42]. Систему (6.44) сведем к одному уравнению n -го порядка и $n-1$ связям (обратный переход уже делали с системой (6.17)). Поставим уже известную нам задачу отыскания условий разрешимости задачи Коши вблизи кривой g в плоскости x, t . Для нахождения первых производных вблизи этой кривой систему (6.44) следует дополнить уравнениями

$$\frac{dt}{d\gamma} \frac{\partial u}{\partial t} + \frac{dx}{d\gamma} \frac{\partial u}{\partial x} = \frac{du}{d\gamma}. \quad (6.45)$$

Разрешимость этой задачи связана с определителем

$$\left\| \begin{array}{cc} I & C \\ \frac{dt}{d\gamma} I & \frac{dx}{d\gamma} I \end{array} \right\| = \left(\frac{dt}{d\gamma} \right)^n \left\| \begin{array}{cc} I & C \\ I & \frac{dx}{dt} I \end{array} \right\| = (-1)^n \left(\frac{dt}{d\gamma} \right)^n \left\| C - \frac{dx}{dt} I \right\|.$$

Поэтому характеристиками являются кривые, удовлетворяющие уравнениям $dx/dt = k_i(x, t)$, в которых поля коэффициентов локального наклона характеристик $k_i(x, t)$, очевидно, суть собственные числа матрицы C . Спектр произвольной матрицы с действительными коэффициентами в общем случае есть подмножество комплексных чисел. Если спектр действителен и прост в любой точке рассматриваемой области, то система (6.44) называется *гиперболической*. Если спектр комплексный (конечно, если n четно), то система – *эллиптическая*. В других случаях строения спектра система называется *параболической*. Таким образом, только гиперболические системы имеют инвариантную систему координат, состоящую из характеристик.

Слово «инвариантный» означает связанность с самим уравнением, а не с его конкретной записью в той или иной системе координат. Для таких систем имеет смысл задача Коши в открытой области, согласованной с расположением поля характеристик, и не имеют смысла постановки в ограниченных областях.

Гиперболические системы описывают распространение возмущений с конечной скоростью. Примеры таких «чисто линейных» систем уже встречались ранее: (6.17), (6.43). Для эллиптических систем характеристики отсутствуют, и имеют смысл только задачи на замкнутой области (пусть и бесконечной), а условия на гра-

нице называются *краевыми*. Возмущения граничных условий распространяются этими системами мгновенно («диффузионным механизмом»). Пример эллиптической системы – упоминавшееся ранее уравнение Лапласа $\Delta v = 0$.

Для этого уравнения второго порядка различают задачу Дирихле (на границе задано значение функции), задачу Неймана (на границе задана производная по нормали к границе) и смешанную задачу. Параболические системы, имеющие поле характеристик, не образующее систему координат, допускают постановки в открытой области, согласованной с расположением характеристик. То есть краевые («начально-граничные») задачи для параболических систем могут иметь смысл, а могут и не иметь его. Например, уравнение диффузии (теплопроводности) $\Delta v = dv/dt$ с условиями, заданными в области $t \geq 0$, описывает распределение субстанции v по пространству: начальные особенности со временем сглаживаются. Но если поставить задачу на области $t \leq 0$, из первоначально гладкого профиля в течение обратного хода времени неизбежно будут возникать особенности, т.е. решения такой задачи будут неустойчивыми. Задачи, в которых нужно установить «прошлое» по настоящему, относятся к некорректным по Адамару задачам. Корректность по Адамару означает существование, единственность и устойчивость решения (в данном случае по отношению к возмущению начальных условий). Причина таких свойств параболического уравнения диффузии кроется в его неинвариантности по отношению к замене $t \rightarrow -t$ (проверьте, что гиперболическое волновое уравнение (6.17) удовлетворяет условию инвариантности).

Классификация систем с использованием характеристик обобщается на квазилинейные системы (т.е. системы дифференциальных уравнений, линейные относительно старших производных, но содержащие в коэффициентах искомые функции и их производные низших порядков) и на системы с числом независимых переменных больше двух (тогда характеристики являются уже поверхностями). Исследуем, например, систему (6.14), являющуюся квазилинейной, считая начальные «возмущения» конечными.

Введением функции давления P : $dP = dp/(a\rho_0)$ эта система сводится к следующей:

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} + a \frac{\partial P}{\partial x} = 0, \quad a \frac{\partial v}{\partial x} + \frac{\partial P}{\partial t} + v \frac{\partial P}{\partial x} = 0.$$

По образцу, данному выше для квазилинейных систем (добавляя к независимым переменным v), или элементарными сложением и вычитанием уравнений получаем:

$$\frac{\partial}{\partial t}(P+v) + (v+a)\frac{\partial}{\partial x}(P+v) = 0,$$

$$\frac{\partial}{\partial t}(P-v) + (v-a)\frac{\partial}{\partial x}(P-v) = 0.$$

Очевидно (см. (6.31), (6.34)), характеристиками системы будут кривые $dx/dt = v \pm a$, на которых постоянны соответствующие величины $P \pm v$. Здесь в отличие от системы, описывающей малые возмущения, a представляет собой поле местной скорости звука, $\pm a$ – относительные скорости распространения волн по газу. Вдоль первого семейства характеристик распространяется волна разрежения, вдоль второго – волна сжатия [63].

6.8. СВЯЗЬ НЕПРЕРЫВНОГО И ДИСКРЕТНОГО НА ПРИМЕРАХ УРАВНЕНИЯ КОЛЕБАНИЙ СТРУНЫ И УРАВНЕНИЯ ШРЕДИНГЕРА

Лишь первый исчез, как сейчас же в
ином положении

Новый родится за ним, а нам кажется,
– двинулся первый.

Лукреций

Вопросы о движении, переходе постепенных количественных изменений в качественные, появлении у целого свойств, которыми не обладает ни одна из его частей, являются одними из ключевых вопросов современного фундаментального естествознания. Эти вопросы имеют глубокие философские корни. Развитие естественных наук заставляет еще и еще раз возвращаться к ним. Ученых XIX в. поразило наличие волновых свойств у света, который они представляли как поток дискретных частиц. К глубокому пересмотру фундаментальных понятий привело в XX в. создание квантовой механики. Оказалось, что дискретные и непрерывные свойства материи нельзя противопоставлять друг другу, что они неразрывно связаны между собой. Выяснилось и другое важное об-

стоятельность. Анализ многих явлений требует сочетания дискретного и непрерывного подходов. И вопрос о соотношении тех и других свойств при построении теории оказывается далеко не простым. От его успешного решения часто зависит, насколько глубоко нам удастся разобраться в изучаемом объекте.

Рассмотрим задачу, при решении которой был развит ряд ключевых идей. Проведем анализ поведения упругой струны, по которой ударили в начальный момент времени. Остановимся прежде на системе, состоящей из точечного груза массой m , к которому прикреплены две одинаковые упругие горизонтальные нити длиной $l_0/2$, натянутые силой F_0 (сила тяжести отсутствует). При отклонении груза от положения равновесия появляется возвращающая сила, пропорциональная отклонению $F = -2F_0 \sin \alpha \approx 4F_0 u/l_0$, тогда согласно второму закону Ньютона имеем:

$$\frac{d^2u}{dt^2} + \omega^2 u = 0, \quad \omega^2 = 4F_0 / (ml_0). \quad (6.46)$$

Уравнение (6.46) описывает колебания с круговой частотой ω :

$$u = A \sin \omega t + B \cos \omega t,$$

где константы A и B определяются начальными положением и скоростью груза. Мы решили задачу о колебании струны, вся масса которой сосредоточена в центре. Для однородной струны с массой m и длиной l_0 это слишком грубое приближение; разумнее заменить ее набором из N шариков с массой $\mu = m/N$, расположенных на расстоянии $h = l_0/N$ и соединенных нитями, натянутыми силой F_0 (так рассуждали, в частности, при выводе уравнений колебаний струны Иоганн и Даниил Бернулли). Если u_k – отклонение k -го шарика от положения равновесия, то при условии, что разница в отклонениях соседних шариков мала,

$$\begin{aligned} \mu a_k &= \frac{mh}{l_0} \frac{\partial^2 u_k}{\partial t^2} = F \approx F_0 \frac{u_{k+1} - 2u_k + u_{k-1}}{h} \Rightarrow \\ &\Rightarrow \frac{\partial^2 u_k}{\partial t^2} = \frac{F_0 l_0}{m} \frac{u_{k+1} - 2u_k + u_{k-1}}{h^2}. \end{aligned}$$

В пределе при $h \rightarrow 0$ получаем уже известное нам волновое уравнение

$$\frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} = 0, \quad c^2 = F_0 l_0 / m.$$

Этот вывод уравнения впервые был сделан Даламбером, который не только записал указанное уравнение, но и нашел его общее решение (6.39) в виде суперпозиции двух волн

$$u = f(x + ct) + g(x - ct).$$

Однако использовать данное решение для струны конечных размеров непросто. Действительно, после удара по струне вправо и влево идут волны. Они доходят до концов струны, отражаются, идут в обратную сторону — устанавливается некий режим, описывать который с помощью полученной формулы неудобно. Возможен другой путь, предложенный Фурье. Так как струна совершает колебательные движения, решение задачи ищут в виде

$$u = (A \sin \omega t + B \cos \omega t) z(x),$$

что дает в итоге (см. разд. 6.4)

$$u = \sum_{n=0}^{\infty} (A_n \sin \omega_n t + B_n \cos \omega_n t) \sin \frac{\pi n x}{l_0}.$$

Решение нашлось в виде суперпозиции стоячих волн, для которых на длине струны укладывается целое число n полуволн. Собственные значения λ_n определяют, с какой частотой может колебаться струна в конфигурации n -й стоячей волны, форму которой описывает собственная функция $z_n = \sin(\pi n x / l_0)$.

Найденные решения по виду сильно различаются. Равноправность их не очевидна. Этот вопрос стал причиной дискуссии в середине XVIII в. (спор о струне) между Эйлером, Даламбером и Лагранжем. Дискуссия позволила убедиться не только в эквивалентности двух решений (ведь исходная задача имеет решение и оно единственное!), но и лучше разобраться в уравнениях. Решение, полученное Фурье, дает возможность выяснить соотношение непрерывного и дискретного в этой задаче. Если мы имеем автомодельное решение волнового уравнения в виде стоячей волны

$u = c_n(t) \sin(\pi n x / l_0)$, то окажется, что для разных $c_n(t)$ получаются уравнения

$$\frac{d^2 c_n(t)}{dt^2} + \omega_n^2 c_n(t) = 0, \quad n = 0, 1, 2, \dots$$

Но это опять уравнения колебаний. Значит, струна оказывается эквивалентной бесконечному множеству независимых колеблющихся грузов.

Интересно и другое: в непрерывной задаче, описывающей колебания струны, есть дискретный набор собственных частот. Непрерывное и дискретное вновь оказываются тесно связанными.

Глубокая связь дискретного и непрерывного отмечена и в физике микромира, где в одних случаях материю удобно рассматривать как электромагнитную волну, а в других — как поток частиц (квантов). Микрочастицы в некоторых опытах ведут себя как волны, например, испытывая дифракцию и интерференцию. В то же время было установлено, что свет (волна) квантуется, регистрируются дискретные порции света — фотоны. Оказалось, что и здесь довольно гармоничный дуализм может быть описан с помощью *линейного* уравнения математической физики — уравнения Шредингера. Выражение простейшей (плоской) волны, описывающей колебания в пространстве с частотой ω и волновым вектором k , записывается как $\exp(ikx - i\omega t)$.

Сопоставим этой волне частицу с энергией $E = \eta\omega$ и импульсом $p = \hbar k$, где \hbar — постоянная Планка. Выразив ω и k из последних соотношений и подставив их в формулу для волны, получим волновую функцию для частицы с энергией E и импульсом p :

$$\psi(x, t) = \exp\left(i\frac{p}{\hbar}x - i\frac{E}{\hbar}t\right).$$

Эта комплексная функция определяет плотность вероятности нахождения частицы во времени и пространстве. Нетрудно убедиться в справедливости тождеств

$$-\frac{\hbar}{i} \frac{\partial \psi}{\partial t} = E\psi \quad \text{и} \quad -\hbar^2 \frac{\partial^2 \psi}{\partial x^2} = p^2 \psi.$$

Но это означает, что энергия E является собственным значением оператора $-\frac{\hbar}{i} \frac{d\psi}{dt}$, а квадрат импульса p^2 — собственным значением оператора $-\hbar^2 \Delta \psi$, при этом ψ в обоих случаях выступает в качестве собственной функции.

Если частица массы m движется в потенциальном поле $V(x)$, то в силу закона сохранения энергии $E = p^2/2m + V$. Чтобы получить волновой аналог этого соотношения, мы должны заменить E , p^2 и V соответствующими дифференциальными операторами:

$$-\frac{\hbar}{i} \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi + V \psi,$$

откуда и следует фундаментальное уравнение волновой и квантовой механики

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi + V \psi, \quad (6.47)$$

открытое австрийским физиком Эрвином Шредингером в 1926 г. Оно описывает движение частицы в заданном потенциальном поле.

Квадрат амплитуды волновой функции $\int_G \psi(x, t) \psi^*(x, t) dx \equiv P(G, t)$

определяет вероятность $P(G, t)$ нахождения частицы в момент времени t в области G . Для оператора (6.47) можно поставить задачу на собственные значения:

$$-\frac{\hbar^2}{2m} \Delta \psi + V \psi = E \psi, \quad \int_{-\infty}^{\infty} \psi \psi^* dx = 1,$$

в результате чего получим спектр E_n значений энергии частицы, которые она может принимать, двигаясь в заданном потенциальном поле. Соответствующие им собственные функции ψ_n показывают, с какой вероятностью можно обнаружить частицу в разных точках пространства.

Решив уравнение Шредингера для кулоновского потенциала $V = -e^2/r$ (e — заряд электрона, r — расстояние до протона), можно найти энергетические уровни атома водорода, с огромной точностью согласующиеся с экспериментом.

До появления уравнения Шредингера существовала матричная квантовая механика Гейзенберга, использующая в качестве аппарата пространство бесконечномерных векторов с ограниченным скалярным квадратом. Теория же Шредингера использует векторное пространство функций с ограниченным скалярным квадратом

(вводимым через интеграл). С точки зрения функционального анализа эти пространства есть эквивалентные представления гильбертова пространства.

Одна из самых замечательных идей Давида Гильберта заключается в том, чтобы рассматривать пространства функций как евклидовы. Шредингер первым увидел в квантованности состояний аналогию с проблемой собственных значений линейного дифференциального оператора. Однако рассказывают, что еще до открытия Шредингером своего уравнения к Гильберту в Геттинген приезжали физики и задавали ему вопрос о смысле матриц в теории Гейзенберга, на который Гильберт ответил, что обычно такие таблицы появляются при решении некоторых дифференциальных уравнений. Физики решили, что Гильберт просто не понял вопроса, а через год Шредингер открыл свое знаменитое уравнение.

6.9. О ПОЛЬЗЕ ФЕНОМЕНОЛОГИИ ПРИ ПОСТРОЕНИИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

Если наши соображения правдоподобны, не следует стремиться к большему.

Цицерон

Любой современный компьютер способен справиться с весьма сложными моделями. Иногда молодые ученые воспринимают это обстоятельство как приглашение создавать сложные и глубокие модели в любом случае, даже когда на то нет никакой необходимости. Между тем многие из задач могут быть решены без привлечения глубоких знаний о структуре изучаемого объекта, процессах, протекающих на микроскопическом уровне, физических механизмах, управляющих поведением системы, т.е. решены *феноменологически*.

Феноменологические модели связывают только непосредственно наблюдаемые и измеряемые в макроскопических опытах величины, никакие модельные представления о структуре материи при этом не используются (по крайней мере, в явном виде). Неоценимое преимущество этих моделей заключается поэтому в их общности, независимости от деталей, частностей, недостаток же является продолжением достоинств – физический механизм при решении задач остается нераскрытым. Таким моделям бессмысленно задавать вопросы «почему»? Приведем только два известных всем

примера феноменологических моделей — закон всемирного тяготения и равновесную термодинамику, которые наш читатель сможет оценить по-новому, увидев, как здесь работают особенности данного подхода.

Закон всемирного тяготения Ньютона гласит, что любые два тела притягиваются друг к другу с силами, пропорциональными произведению их масс и обратно пропорциональными квадрату расстояния между ними. С помощью этого закона и трех законов динамики Ньютону удалось очень хорошо описать движения планет вокруг Солнца. Закон всемирного тяготения предполагает действие на расстоянии, что довольно необычно. Однако Ньютон мудро воздерживался от каких-либо высказываний относительно физической природы сил тяготения (о которой до сих пор нет единого мнения), поэтому то, что он просто ограничился формулировкой количественных соотношений измеримых в макроопытах величин без всяких фантастических измышлений, определенно поспособствовало прогрессу в физике и технике.

Классическая равновесная термодинамика [91] базируется на трех законах, называемых «началами», также основанных на результатах макроэкспериментов. Далее мы изложим и обоснуем на примере идеальных газов эти начала и в качестве демонстрации их возможностей докажем невозможность построения вечного двигателя.

В классической равновесной термодинамике изучают *равновесные процессы*, т.е. непрерывные изменения равновесных состояний. *Равновесность* состояния изучаемого объема газа означает однородность распределения по этому объему параметров состояния газа, т.е. отсутствие их градиентов и потоков. Для его реализации воздействие на среду должно прилагаться достаточно медленно по сравнению со скоростью установления в ней равновесия, т.е. *квазистатически*.

Рассмотрим газ, который при некоторой температуре T имеет давление P и занимает объем V . Мы, конечно, умеем измерять температуру, объем и давление нашего газа с помощью известных приборов. Из этих трех параметров независимы только два, т.е. состояние газа характеризуется любыми двумя такими параметрами.

Термостатом назовем массивное тело, температура которого не меняется при контакте с газом. Термостат позволит нам поддерживать желаемую температуру газа. Проводя в термостате при заданной температуре испытания над газом, получим кривую *изо-*

термы $P(V)$. Непрерывное изменение температуры термостата дает непрерывную сеть непересекающихся изотерм. Отсюда следует **принцип температуры**: существует функция состояния газа в процессах, протекающих в термостате, — *температура*.

Адиабатом назовем сосуд с теплонепроницаемыми стенками. Испытывая в адиабате газ, построим кривую *адиабаты* $P(V)$. При непрерывной подкачке в адиабат теплоты (приведением в контакт с телом более высокой температуры) получим непрерывную сеть непересекающихся адиабат. Отсюда следует **принцип энтропии**: существует функция состояния газа в процессах, протекающих в адиабате, — *энтропия*.

Сети изотерм и адиабат образуют систему криволинейных координат на плоскости (P, V) . Эти координатные кривые можно локально отградуировать так, чтобы якобиан преобразования от координат (T, S) к координатам (P, V) всюду равнялся бы единице. Тем самым по известным шкалам давления и объема определяются шкалы *абсолютной температуры* T и *абсолютной энтропии* S . Поскольку T и S являются независимыми функциями состояния, последнее можно однозначно описать не только парой P, V , но и другими парами независимых переменных состояния: T, S ; P, T ; V, S ; P, S или V, T .

Определим *совершение элементарной работы* газом как $\delta A = PdV$. Тогда работа определяется следующим соотношением:

$$A = \int_{V_1}^{V_2} PdV.$$

Поскольку давление определяется не объемом, а состоянием газа (объемом и температурой или объемом и энтропией), работа является не функцией состояния, а *функцией процесса*. Чтобы вычислить работу, нужно знать, как в течение процесса изменялась температура или энтропия.

Принцип энергии: существует функция состояния термодинамической системы U , называемая *внутренней энергией*. Значение этой функции может быть изменено двумя независимыми способами: а) совершением работы δA газом или б) подведением теплоты δQ , т.е. приведением газа в контакт с телом, имеющим большую температуру, если в течение такого контакта газом не совершается работа, т.е. $dU = \delta Q - \delta A$.

В адиабате второй канал изменения внутренней энергии исключен, поэтому функция состояния U , для адиабатических процессов совпадающая с работой, есть

$$\delta A_S = PdV_S = -dU_S.$$

Для неадиабатических процессов очевидно

$$dU = -PdV + \alpha dS,$$

где параметр α – условие того, чтобы справа стоял полный дифференциал. Используя принятую ранее нормировку якобиана, получаем

$$dU = TdS - PdV.$$

Из принципа энергии и определения адиабатического потенциала получаем выражение для подведения теплоты в виде

$$\delta Q = TdS.$$

Подведенная теплота находится как

$$Q = \int_{S_1}^{S_2} TdS.$$

Поскольку температура определяется не энтропией, а состоянием газа (энтропией и объемом или энтропией и давлением), подведенная теплота является не функцией состояния, а *функцией процесса*. Чтобы вычислить подведенную теплоту, нужно знать, как в течение процесса изменялись объем или давление.

Трех сформулированных принципов оказывается достаточно, например, для доказательства невозможности построения тепловой машины с предельно высоким коэффициентом полезного действия (КПД), т.е. вечного двигателя второго рода.

Представим себе замкнутый цикл в плоскости (P, V) . Так как T и S – функции состояния и возврат в ту же точку при осуществлении цикла в плоскости (P, V) не меняет состояния, в плоскости (T, S) тоже реализуется цикл. Из принципа энергии следует, что в рассматриваемом цикле

$$\oint dU = \oint \delta Q - \oint \delta A,$$

а так как при осуществлении цикла опять приходим в то же состояние, а U есть функция состояния, то

$$\oint \delta Q = \oint \delta A,$$

т.е. площади циклов в плоскостях (P, V) и (T, S) совпадают. Для выполнения цикла с точки зрения координат (T, S) необходимо получить теплоту $Q_1 > 0$ от нагревателя и отдать теплоту $0 < Q_2 < Q_1$ холодильнику, при этом только разность $Q_1 - Q_2 = \oint \delta Q = \oint \delta A$ превращается в работу. У теплового двигателя КПД $\eta = A/Q_1 = (Q_1 - Q_2)/Q_1$.

Довольно легко показать, что наибольшим КПД обладает цикл Карно, описываемый прямоугольником в плоскости (T, S) со сторонами вдоль координатных осей. Описываем вокруг произвольного цикла в этой плоскости такой прямоугольник, для которого величины полученной и отданной теплоты суть q_1 и q_2 , а для произвольного цикла — $q_1 - \delta q_1$ и $q_2 + \delta q_2$. Тогда

$$\eta = (q_1 - q_2 - \delta q_1 - \delta q_2)/(q_1 - \delta q_1) \leq (q_1 - q_2)/q_1.$$

Для цикла же Карно

$$q_1 = T_1(S_2 - S_1), \quad q_2 = T_2(S_1 - S_2) \quad \text{и} \quad \eta = (T_1 - T_2)/T_1.$$

Как видно, КПД всегда меньше единицы и приближается к ней только в случае, если температуру холодильника устремить к абсолютному нулю. Для доказательства этого факта нам не потребовалось знания каких-либо частных микроскопических механизмов в газах, оно фактически основано на трех правдоподобных основаниях феноменологического характера, т.е. буквально «лежит на поверхности».

Бесспорными преимуществами феноменологического пути описания явлений, кроме универсальности, оказываются простота, обзорность получаемых в итоге моделей и вытекающая отсюда возможность полнее изучить и понять причинно-следственные связи, ими описываемые (общий факт доказывается легче частного).

6.10. АНАЛИЗ ПОДОБИЯ И РАЗМЕРНОСТИ

Чрезмерная глубина лишь путает и затуманивает мысль.

Э. По

Ученые часто сталкиваются с задачей построения математической модели настолько нового, необычного и сложного явления, для которого еще не найден удовлетворительный механизм и не установлены балансовые и определяющие уравнения. В этих случаях главную роль играют экспериментальные методы исследования, позволяющие установить простейшие опытные факты. Постановка же эксперимента требует предварительного теоретического анализа. Возможность такого предварительного анализа дает теория размерности и подобия [97]. Несмотря на свою простоту и эффективность, этот анализ, конечно, требует от исследователя определенного опыта и проникновения в сущность изучаемого явления. Самый замечательный итог его применения заключается в том, что он оказывается достаточным для установления особого класса автомодельных решений математической формулировки задачи, отражающих глубокие физические закономерности изучаемого явления. Сформулируем суть метода и приведем несколько примеров его применения.

Пусть некая зависимость определяет величину a в функции n параметров a_1, \dots, a_n :

$$a = f(a_1, \dots, a_n). \quad (6.48)$$

Если эта зависимость имеет физическое содержание, то соотношение (6.48) должно отражать тот бесспорный факт, что хотя числа a, a_1, \dots, a_n выражают значения соответствующих величин в определенной системе единиц измерения, физическая закономерность, которую это соотношение представляет, не зависит от произвола при выборе единиц измерения. Анализ размерностей основан на физически содержательном утверждении, сформулированном Э. Бакингамом и известном как Π -теорема.

Пусть существует физическая закономерность, выраженная в виде зависимости некоторой размерной величины от размерных определяющих параметров. Эта зависимость может быть представлена в виде зависимости некоторой безразмерной величины от безразмерных комбинаций определяющих параметров. Количество этих безраз-

мерных комбинаций меньше общего числа определяющих параметров на число размерных определяющих параметров с независимыми размерностями.

Следует отметить, что П-теорема интуитивно вполне очевидна и ее неявное использование началось задолго до того, как она была явно сформулирована и формально доказана. В этой связи следует, прежде всего, назвать имена Галилея, Ньютона, Фурье, Максвелла, Рейнольдса, Рэля. Практический выигрыш при обработке экспериментальных данных, даваемый этой теоремой, очевиден.

Пусть для выяснения зависимости некоторой величины a от некоторого определяющего параметра a_i надо измерить эту величину при десяти значениях данного аргумента. Тогда для экспериментального нахождения величины a как функции n определяющих параметров a_1, \dots, a_n следует произвести 10^n экспериментов. Согласно П-теореме, если все величины a, a_1, \dots, a_n выражаются через k независимых размерностей, дело сводится к определению функции $n-k$ безразмерных аргументов Π_1, \dots, Π_{n-k} , для нахождения которой достаточно 10^{n-k} опытов, т.е. в 10^k раз меньше! Трудоемкость установления искомой функции сокращается в данном случае на столько порядков, сколько среди определяющих параметров величин с независимыми размерностями.

Применение П-теоремы включает следующую последовательность действий. Разобьем величины a_1, \dots, a_n на две группы. В первую группу a_1, \dots, a_k включается ряд величин с независимыми размерностями (например, длина, скорость, плотность и т.д.). Во вторую группу a, a_{k+1}, \dots, a_n входят величины, размерности которых можно выразить через размерности величин первой группы. То есть

$$a \sim a_1^p, \dots, a_k^q, \quad a_{k+1} \sim a_1^{p_{k+1}}, \dots, a_k^{q_{k+1}}, \quad \dots, \quad a_n \sim a_1^{p_n}, \dots, a_k^{q_n},$$

где p, q, p_i, q_i – показатели степени. При этом величины

$$\Pi = \frac{a}{a_1^p, \dots, a_k^q}, \quad \Pi_1 = \frac{a_{k+1}}{a_1^{p_{k+1}}, \dots, a_k^{q_{k+1}}}, \quad \dots, \quad \Pi_{n-k} = \frac{a_n}{a_1^{p_n}, \dots, a_k^{q_n}}$$

оказываются безразмерными, так как их значения будут одними и теми же при любом выборе системы единиц измерения. Независимость закономерности, имеющей физический смысл, от выбора

единиц измерения означает, что соответствующее ей соотношение (6.48) можно представить в виде

$$П = \Phi(\Pi_1, \dots, \Pi_{n-k}).$$

Уменьшение числа аргументов упрощает исследование, иногда — существенно. Рассмотрим показательный пример.

В 1909–1911 гг. специалисты в области физической химии Э. Бозе, Д. Рауэрт и М. Бозе опубликовали серию экспериментов. Ими изучалось время τ заполнения сосуда данного объема Q и перепад давления P на концах трубки при стационарном протекании через трубку различных жидкостей: воды, хлороформа, бромформа, ртути. Результаты опытов были представлены в виде серии зависимостей перепада давления от времени заполнения для разных жидкостей (рис. 6.2).

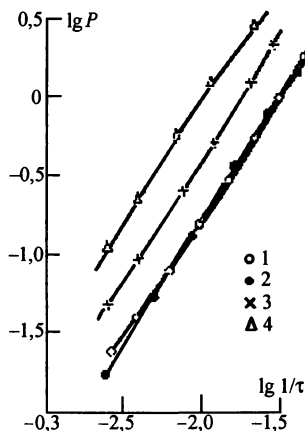


Рис. 6.2. Результаты опытов Бозе-Рауэрта в исходном виде для различных жидкостей: 1 — вода, 2 — хлороформ, 3 — бромформ, 4 — ртуть [7]

Эти работы были замечены Теодором фон Карманом, который исследовал их с точки зрения анализа размерностей. Рассуждения Кармана можно представить следующим образом: перепад давления P на концах трубки должен зависеть от времени заполнения сосуда τ , объема сосуда Q и свойств жидкости (коэффициента вязкости μ и плотности ρ):

$$P = f(\tau, Q, \mu, \rho)$$

(в данном случае $n = 4$). Размерности параметров в классе MLT (масса, длина, время) выражаются следующими соотношениями:

$$[P] = \frac{M}{LT^2}, \quad [\tau] = T, \quad [Q] = L^3, \quad [\mu] = \frac{M}{LT}, \quad [\rho] = \frac{M}{L^3}.$$

Легко увидеть, что определяющие параметры τ , Q , μ имеют независимые размерности, размерность же параметра ρ выражается через размерности первых трех:

$$[\rho] = [\mu][\tau][Q]^{-2/3}.$$

Таким образом, $k = 3$, $n - k = 1$ и анализ размерности дает

$$\Pi = \Phi(\Pi_1), \quad \Pi = \frac{P}{\mu\tau^{-1}}, \quad \Pi_1 = \frac{\rho}{\mu\tau Q^{-2/3}}.$$

Следовательно, в координатах Π , Π_1 все опытные точки должны располагаться на единой кривой. Этот вывод блестяще подтвердился (рис. 6.3). Ясно, что заранее проведенный анализ размерностей мог бы сократить объем экспериментальной работы во много раз.

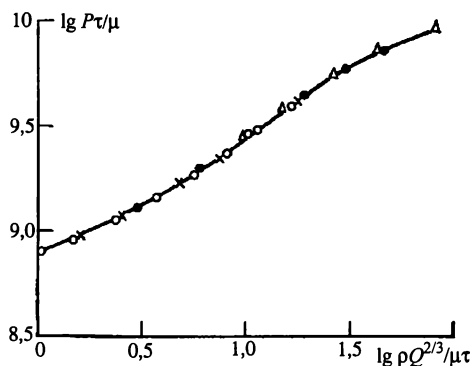


Рис. 6.3. Результаты опытов Бозе-Рауэрта в интерпретации Кармана [7]

Интересен также следующий пример. При атомном взрыве в области, настолько малой, что ее можно считать точкой, быстро (практически мгновенно) выделяется значительная энергия E . От центра взрыва распространяется мощная ударная волна, давление

за которой вначале составляет сотни тысяч атмосфер. Это давление много больше, чем начальное давление воздуха, влиянием которого на первой стадии взрыва можно пренебречь. Таким образом, радиус фронта ударной волны r через промежуток времени τ после взрыва зависит от E , τ и начальной плотности воздуха ρ :

$$r = f(E, \tau, \rho).$$

Таким образом, $n = 3$, а размерности определяющих параметров в классе MLT есть:

$$[E] = \frac{ML^2}{T^2}, \quad [\tau] = T, \quad [\rho] = \frac{M}{L^3}.$$

Легко видеть, что $k = 3$, $n - k = 0$, так что функция Φ в выражении $\Pi = \Phi(\Pi_{n-k}, \dots, \Pi_n)$ не зависит ни от одного из аргументов, т.е. $\Phi \equiv \text{const}$. Далее нетрудно доказать, что $\Pi = r(E\tau^2/\rho)^{-1/5}$, откуда

$$r = \text{const}(E\tau^2/\rho)^{1/5}.$$

Эта формула показывает, что если измерить тем или иным способом радиус ударной волны в разные моменты времени, то в логарифмических координатах $(5/2)\lg r$, $\lg \tau$ экспериментальные точки должны располагаться на прямой

$$(5/2)\lg r = (5/2)\lg(\text{const} \cdot E^{1/5}\rho^{-1/5}) + \lg \tau,$$

имеющей наклон, равный единице.

Это подтвердил в 1950 г. английский механик и физик сэра Джеффри Ингрэм Тэйлор, обработавший кинофильм о распространении огненного шара, снятый во время американских испытаний ядерного взрыва в Нью-Мексико в 1945 г. (рис. 6.4). Как показывает более детальный анализ, значение $\text{const} \approx 1$. Зная это, по экспериментальной зависимости радиуса фронта от времени можно определить энергию взрыва. Публикация Тэйлором этой величины, оказавшейся равной примерно 10^{17} Дж, вызвала в свое время смущение в американских официальных кругах.

Приведем еще один скорее забавный пример применения анализа размерностей: «докажем» с его помощью теорему Пифагора. Площадь прямоугольного треугольника S устанавливается величи-

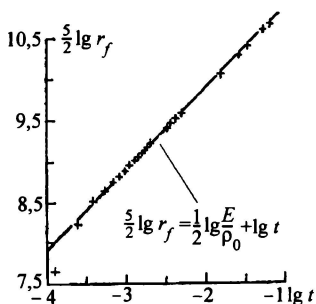
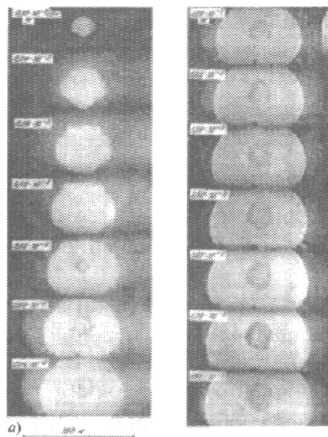


Рис. 6.4. Распространение ударной волны при ядерном взрыве: а – последовательные фотографии огненного шара; б – опытные данные (из кинофильма) располагаются на прямой автомодельного решения [97]

ной его гипотенузы s и для определенности меньшим из его острых углов φ , причем из анализа размерности следует $S = c^2 f(\varphi)$, где f – функция, значения которой безразмерны. Высота, перпендикулярная гипотенузе, разбивает основной треугольник на два подобных ему прямоугольных треугольника. Очевидно, что их площади $S_1 = a^2 f(\varphi)$, $S_2 = b^2 f(\varphi)$, где $f(\varphi)$ – та же, что и в случае основного треугольника. Но поскольку $S_1 + S_2 = S$, то $a^2 f(\varphi) + b^2 f(\varphi) = c^2 f(\varphi)$, откуда следует, что $a^2 + b^2 = c^2$.

6.11. АВТОМОДЕЛЬНОСТЬ

Персонажи обязаны подчиняться законам мира, в котором живут. То есть писатель – пленник собственных предпосылок.

У. Эко

В различных областях математической физики (задачах гидродинамики, переноса тепла, диффузии, химической кинетики) появляются своеобразные частные решения, которые при изменении

времени преобразуются одно из другого по правилу подобия. Для функции пространственной и временной переменных такое решение имеет вид

$$f(x, t) = t^\alpha g(t^\beta x) \quad (6.49)$$

и инвариантно относительно преобразования $t \rightarrow t + bt$, $x \rightarrow x/b^\beta$, $f \rightarrow b^\alpha f$, образующего мультипликативную группу. В геометрии такое преобразование называется *аффинным*. Само решение (6.49) называется *автомодельным* [7, 41].

В ранее рассмотренном автомодельном решении (6.40) волнового уравнения подобие сводится к параллельному переносу начального профиля и не связано с изменением масштаба. В таком решении аргумент $x \pm at = \text{const}$, устойчивый относительно преобразований трансляции $t \rightarrow t + b$, $x \rightarrow x - ab$, образующих аддитивную группу, может быть приведен к виду (6.49) с помощью замены $x = \ln \hat{x}$, $t = \ln \hat{t}$: $\hat{x} \hat{t}^{\pm a}$. Показатель степени $\beta = \pm a$ в (6.49) называется *показателем автомодельности*.

Наиболее важное свойство автомодельного решения состоит в том, что зависимость от аргументов входит через единственный комплекс. Поэтому подставляя (6.49) в уравнение в частных производных, получаем обыкновенное дифференциальное уравнение. Его интегрирование существенно проще нахождения решения начально-граничной задачи.

В сложных нелинейных задачах получение таких решений зачастую остается единственно возможным средством «пробиться» через аналитические трудности и обрести понимание качественных особенностей явления. При обработке опытных данных автомодельность приводит к тому, что, казалось бы, беспорядочное в обычных координатах «облако» опытных точек располагается на единой кривой или поверхности, построенной в автомодельных координатах. Автомодельность привлекает внимание и как глубокий физический факт, свидетельствующий о наличии определенного типа стабилизации исследуемых процессов, имеющей место для достаточно широкого круга начальных и/или граничных условий.

Получим автомодельное решение линейного параболического уравнения диффузии [41]

$$\frac{\partial \rho}{\partial t} = k \frac{\partial^2 \rho}{\partial x^2}, \quad (6.50)$$

где ρ – плотность массы; k – коэффициент диффузии. Используем для построения автомодельного решения анализ подобия и размерности. Решение не может зависеть от естественного произвола в выборе основных единиц измерения и, как следует из анализа размерности, должно представляться в виде функции единственно возможного безразмерного комплекса $s = x^2/(kt)$. Поэтому будем искать решение в виде

$$\rho = t^\alpha g\left(\frac{x^2}{kt}\right), \quad (6.51)$$

т.е. нам остается определить показатель α и функцию g .

Показатель степени находится из закона сохранения масс

$$\int \rho dx = m:$$

$$\int_0^\infty t^\alpha g\left(\frac{x^2}{kt}\right) dx = \int_0^\infty t^\alpha g(s) \sqrt{kt} \frac{ds}{2\sqrt{s}} = \sqrt{k} t^{\alpha+1/2} \int_0^\infty g(s) \frac{ds}{2\sqrt{s}} = \text{const},$$

откуда $\alpha = -1/2$ (почему?).

Подставляя решение (6.51) в (6.50), приходим к соотношению

$$-g(s) - \frac{2x^2}{kt} g'(s) = 4g'(s) + \frac{8x^2}{kt} g''(s) \Leftrightarrow 8sg''(s) + (2s+4)g'(s) + g(s) = 0.$$

Обыкновенные линейные дифференциальные уравнения второго порядка с переменными коэффициентами редко интегрируются в квадратурах, но оператор в нашем случае оказывается представимым в виде произведения двух дифференциальных операторов первого порядка: $(2sD+1)(4D+1)g = 0$, где D означает операцию дифференцирования. Интегрируя уравнение $(2sD+1)f = 0$ по частям, получаем $f = B/\sqrt{s}$. Возвращаясь к функции g , будем иметь для нее уравнение $(4D+1)g = B/\sqrt{s} \Leftrightarrow 4g' + g = B/\sqrt{s}$, решение которого

$$g(s) = C \exp(-s/4) + \frac{B}{4} \exp(-s/4) \int_0^s \frac{1}{\sqrt{z}} \exp(z/4) dz$$

состоит из двух линейно независимых функций:

$$g_1(s) = \exp(-s/4) \int_0^s \frac{1}{\sqrt{s}} \exp(s/4) ds, \quad g_2(s) = \exp(-s/4).$$

Можно показать, что первое из решений приводит к отрицательным значениям для плотности массы на $x \in [-\infty, 0]$, что не имеет физического смысла, а второе решение вполне пригодно:

$$\rho(x, t) = t^{-1/2} \exp\left(-\frac{x^2}{4kt}\right). \quad (6.52)$$

Графики полученного автомодельного решения в некоторые последовательные моменты времени показаны на рис. 6.5. При $t \rightarrow 0$ и $x \neq 0$ имеем $\rho \rightarrow 0$, а при $t \rightarrow 0$ и $x = 0$ имеем $\rho \rightarrow \infty$, что видно из рис. 6.5. В пределе получаем функцию, равную нулю всюду, кроме точки $x = 0$, и бесконечности в этой точке, причем

$$\int_{-\infty}^{+\infty} \rho(x, 0) dx = m, \text{ т.е. уже встречавшуюся нам дельта-функцию:}$$

$$\rho(x, 0) \sim m\delta(x).$$

Так как масса m вещества сохраняется, решение (6.52) описывает эволюцию сгустка частиц, первоначально находящегося в начале координат. В пределе при $t \rightarrow \infty$ получаем равномерно «размазанное» по оси x поле $\rho \rightarrow 0$ (тенденция эта видна из рис. 6.5), однако суммарная масса вещества остается равной m .

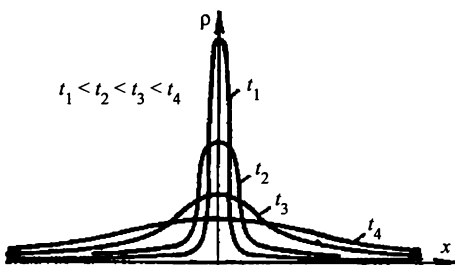


Рис. 6.5. Автомодельное решение уравнения диффузии [41]

В разобранный задаче показатель автомодельности β в (6.49) был найден из соображений размерности. Решения, для которых этот показатель определяется на основании законов сохранения либо размерности, Я.Б. Зельдович отнес к автомодельным реше-

ниям первого рода. Если для нахождения этого показателя требуется решить специальную задачу на собственные значения, то соответствующие решения именуют автомодельными второго рода.

Удивительно то, что в *нелинейных* системах решения такого рода соответствуют некоторому классу начальных условий. Такие системы со временем забывают «случайные черты» начальных условий и стремятся к автомодельному решению (или одному из автомодельных решений, если их несколько). Подробнее это будет рассмотрено ниже.

6.12. САМООРГАНИЗАЦИЯ И СТРУКТУРЫ В НЕЛИНЕЙНЫХ СРЕДАХ

Поистине восхищения заслуживает изобретательность, с которой природа столь тонко и разнообразно маскирует простоту под сложность!

П. Эткинс

Линейное волновое уравнение (6.7) прекрасно описывает распространение звука, когда его громкость невелика (при этом скорость звука не зависит от громкости). Если звук очень громкий, например от взрыва, то может возникнуть ударная волна. Ее скорость зависит от разности давлений за волной и перед ней. Для описания такого поведения требуется формулировать нелинейную модель. Однако получение решения и качественное исследование нелинейной модели требует других подходов, поскольку даже малая нелинейная добавка качественно меняет ситуацию: сумма двух решений уже не удовлетворяет уравнению. Принцип суперпозиции «не работает» и «сшить» общее решение из частных уже не удастся.

Исследуем, к каким качественным изменениям приводит появление нелинейности в простейших математических моделях на примере уравнения теплопроводности, по виду совпадающего с уравнением диффузии (6.50)

$$\frac{\partial T}{\partial t} = k \frac{\partial^2 T}{\partial x^2}, \quad (6.53)$$

где $k > 0$ – коэффициент теплопроводности, $T \geq 0$ – температура. Оно описывает передачу тепла, диффузию частиц, проникновение магнитного поля в плазму и некоторые другие процессы.

Представим, что у нас есть компьютер с соответствующим программным обеспечением, который умеет решать различные уравнения. Остается только задать граничные условия и начальные данные и посмотреть, как ведет себя решение. Будем считать для простоты, что отрезок бесконечен, $-\infty < x < \infty$, причем $T(x \rightarrow \pm\infty) \rightarrow 0$. Пусть в начальный момент до высокой температуры нагрет маленький участок (рис. 6.6,а). На рис. 6.6,а показаны профили температуры в разные моменты времени.

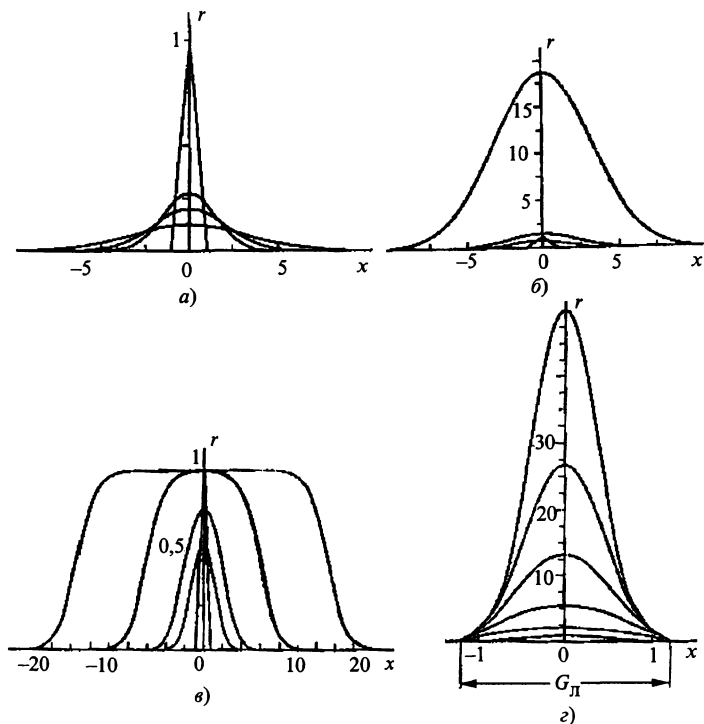


Рис. 6.6. Решения уравнения теплопроводности [60]

Можно видеть, что максимальная температура снижается по закону $A \sim t^{-1/2}$, а полуширина профиля на уровне $A/(2L)$ возрастает как $L \sim t^{1/2}$. Это кажется понятным: от более нагретых участ-

292

ков тепло передается менее нагретым, при этом температура первых уменьшается, а вторых — повышается. Поскольку количество тепла остается неизменным, то $AL \sim \text{const}$.

Теперь рассмотрим среду, в которой происходит процесс горения [59, 60, 67]. При этом в правой части уравнения теплопроводности появляется новый член, описывающий тепловыделение источников. Будем считать, что интенсивность горения пропорциональна температуре. Это приводит к линейному уравнению

$$\frac{\partial T}{\partial t} = k \frac{\partial^2 T}{\partial x^2} + qT, \quad q > 0. \quad (6.54)$$

Решение этого уравнения при тех же начальных условиях представлено на рис. 6.6,б. Как видно, амплитуда A растет, а полуширина меняется по закону $L \sim t^{1/2}$. Таким образом, закономерности очень похожи. В чем причина? Оказывается, уравнение (6.54) можно привести к виду (6.53), если ввести замену переменной $T^* = \exp(qt)T$. Значит, $A \sim \exp(qt)t^{-1/2}$, что совпадает с результатами расчета.

Посмотрим, что произойдет, если источник нелинеен:

$$\frac{\partial T}{\partial t} = k \frac{\partial^2 T}{\partial x^2} + qT - \alpha T^3, \quad q, \alpha > 0. \quad (6.55)$$

Нелинейный член описывает прекращение горения при больших температурах. Это может быть связано с выгоранием топлива или большей ролью эндотермических реакций с ростом T . Уравнение (6.55) возникает не только в теории горения. С его помощью моделируют распространение эпидемий, прохождение импульса по нервному волокну. Посмотрим, как выглядят решения этого уравнения (рис. 6.6,в).

Они разительно отличаются от того, что мы видели раньше: возникает тепловая волна, распространяющаяся с постоянной скоростью, причем амплитуда волны стремится к постоянному значению $(q/\alpha)^{1/2}$. Но известно, что бегущая волна — это автомодельное решение вида $T^*(\pm \eta t)$, сохраняющее свою форму. Причем в отличие от линейного уравнения теплопроводности, где существует бесконечный набор собственных функций T_1, T_2, \dots , здесь функция T^* только одна и определяется свойствами нелинейной среды

(в нелинейных уравнениях собственных функций может оказывать ся конечное число). Роль функции T^* в модели тепловых структур велика. Она определяет локализованные конфигурации в пределах, в которых процессы идут согласованно. Именно поэтому ее во многих работах называют собственной функцией нелинейной среды. Однако в отличие от линейных задач она описывает локализованные процессы и никак не связана с краевыми условиями.

Наличие источников и стоков является типичным в так называемых открытых системах, которые в отличие от замкнутых могут обмениваться с окружающей средой энергией, веществом, информацией. «Забывание» начальных данных, т.е. стремление для целого класса начальных профилей к одному и тому же решению характерно для большого класса открытых нелинейных систем. Причем это решение часто оказывается автомодельным. Такое поведение говорит о возникновении упорядоченности в системе или о самоорганизации. В самом деле, «выход» на автомодельное решение означает уменьшение числа степеней свободы и выделение нескольких основных (параметров порядка), к которым подстраиваются все остальные.

Рассмотрим систему с нелинейным коэффициентом теплопроводности k и нелинейным источником, обладающим следующим свойством: чем больше отклонение от равновесия, тем быстрее идет процесс

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial x} \left(k T^\sigma \frac{\partial T}{\partial x} \right) + q T^\beta, \quad k, q > 0, \quad \beta > \sigma + 1. \quad (6.56)$$

Такие модели характерны для физики плазмы, химической кинетики, экологии, разрушения, пластичности.

Система (6.56) обладает еще более необычными свойствами. Обратим внимание на два принципиальных отличия от всех других решений, обсуждавшихся выше. Профиль температуры оказывается локализованным внутри некоторой области G_λ , вне которой $T(x, t)$ равна нулю (рис. 6.6,з). Из непрерывной среды при этом выделяются ограниченные участки, в пределах которых и происходит горение. Решение существует только в течение ограниченного времени t_f называемого временем обострения. За это время функция $T(x, t)$ в конечной области обращается в бесконечность.

До недавнего времени математики полагали, что такие решения не представляют особого интереса и наличие их говорит о несовершенстве модели. Однако развитие физики плазмы, газовой

динамики, других областей привело к появлению содержательных задач, в которых ведущими оказываются один или несколько наиболее быстрых процессов. Если посмотреть внимательно на решение, то оказывается, что профили температуры в процессе эволюции остаются подобными себе. Возникает вопрос: не определяются ли и они автомодельными решениями? Действительно, на развитой стадии (когда выделяется много больше тепла, чем в начальный момент времени) процесс горения описывается формулой

$$T = g(t)f(x/\varphi(t)). \quad (6.57)$$

Здесь $g(t)$ определяет закон роста амплитуды, $\varphi(t)$ — полуширины, f — форму профиля, $g(t) \rightarrow \infty$ при $t \rightarrow t_f$; g, φ — функции степенного вида:

$$g(t) \sim (t - t_f)^{-1/(\beta-1)}, \quad \varphi(t) \sim (t - t_f)^{(\beta-\sigma-1)/(\beta-1)}.$$

Обратим внимание на то, что в пределах области локализации горение происходит согласованно в разных точках пространства. Выход на решение вида (6.57) говорит о спонтанном возникновении упорядоченности, о формировании локализованных структур. Для этой нелинейной среды также характерно «забывание» деталей начальных данных, что в некотором смысле парадоксально. В среде, где есть только горение и теплопроводность — диссипативный процесс, связанный с рассеянием энергии и обычно уничтожающий всякую упорядоченность, — возникают структуры, сохраняющие свою форму. Чтобы подчеркнуть необычность этого явления, их называют *диссипативными структурами*.

Познакомимся подробнее с тепловыми структурами. Ясно, что у них есть две важные характеристики — время их жизни t_f и область локализации G_λ . Если в среде независимо развиваются две локализованные структуры с разными временами обострения $t_{f2} > t_{f1}$, то практически горение происходит на характерных временах $\approx t_{f1}$, т.е. только в первой, и замирает во второй структуре.

Реально живут в одном мире только структуры с одинаковыми временами обострения. Если попытаться объединить несколько структур так, чтобы их области локализации пересекались, структуры начинают взаимодействовать, возникает волна горения сложной формы, сходящаяся к центру. Однако эта форма с ростом температуры меняется, и в конце концов остается одна простая быстро горящая структура. Возникает вопрос, можно ли по-

строить сложные структуры, сохраняющие в процессе эволюции свою форму. Проблема набора форм, заключенного в различных средах, является очень глубокой. И для древних философов, и для многих физиков-теоретиков, занимающихся элементарными частицами, характерно представление о сложном наборе структур, которые потенциально заключены в среде. Некоторые из них встречаются часто, другие могут появляться только в определенных условиях. Однако и те, и другие характеризуют глубокие внутренние свойства системы.

Посмотрим на собственные функции нелинейной среды с другой точки зрения. Начальные условия можно рассматривать как воздействия на нашу систему. Пусть задача состоит в том, чтобы создать в среде сложную структуру желаемой организации. Оказывается, что для ее решения достаточно задать начальный профиль температуры малой амплитуды, согласованный с собственными функциями нелинейной среды. Для того чтобы воздействие на нелинейную систему было эффективным, оно должно быть согласовано с ее внутренними свойствами. Такое воздействие, даже будучи слабым, окажется действеннее, чем в тысячи раз более сильное, но не согласованное со свойствами системы.

6.13. О НЕЛИНЕЙНЫХ ВОЛНАХ В СПЛОШНЫХ СРЕДАХ

В поисках истины логика нащупывает свой путь по отклонениям от обычного и заурядного.

Э. По

Поскольку большинство ранее рассмотренных примеров составили волны, то нельзя не рассказать о нелинейных волнах на мелкой воде, к которым относят цунами [96]. Цунами — необычайно высокие и мощные волны, обрушивающиеся на побережья океанов. Чаще всего цунами возникают в результате сильных подводных землетрясений, сопровождающихся быстрым образованием на дне океана сбросов, обвалов и оползней. Быстрое смещение блоков горных пород, словно гигантский поршень, выталкивает огромные массы воды в виде горба на поверхность океана. Невероятно страшная волна высотой 30—40 м образовалась при взрыве вулкана Кракатау в 1883 г., в течение нескольких минут она смыла с индонезийских островов 35 тысяч человек.

Цунами необычайно устойчивы — удаляясь от своего (практически точечного) источника на многие тысячи километров, они способны пересечь Тихий океан. Так, Чилийское цунами в 1960 г. нанесло значительный урон Японским и Курильским островам.

Эти волны очень длинные (длина волны составляет 100–300 км) и пологие — даже сильнейшие цунами имеют в открытом океане высоту не более чем 2 м. Суда, которые оказались в открытом океане на пути цунами, обычно не ощущают их. Скорость этих волн — колоссальная, в зависимости от глубины океана может достигать от 400 до 850 км/ч. Высота цунами уменьшается над большими глубинами и увеличивается на мелководье. Скорость, напротив, возрастает на больших глубинах и снижается на малых. Подходя к берегу, длина и скорость волны уменьшаются, а ее высота увеличивается (кинетическая энергия преобразуется в потенциальную), волна принимает резко асимметричную форму и опрокидывает свой гребень вперед, тараном обрушиваясь на берег.

Вероятно, читатель сразу отметил, что некоторые детали указывают на нелинейную природу этих волн (например, существование зависимости скорости от амплитуды). Поэтому строить линейные модели здесь бессмысленно, они не будут адекватными. Вероятно, первым естествоиспытателем, очарованным нелинейной волной, был англичанин Дж. Рассел (кстати, конструктор гигантского океанского лайнера «Грейт Истерн», с которого укладывался первый трансатлантический телеграфный кабель). В 1834 г., прогуливаясь верхом вдоль узкого судоходного канала, он заметил, как от остановившейся баржи отделилось и побежало вперед большое одиночное возвышение («водяной холм»), которое продолжало свой путь, нисколько не меняя своей формы и не снижая скорости, пока, как писал Рассел, *«через одну или две мили погони я не потерял его из виду»*. Это наблюдение вдохновило Рассела провести экспериментальное изучение поверхностных волн на воде, в результате чего были воспроизведены уединенные волны и получено выражение для скорости их распространения

$$v = \sqrt{g(h + \eta)},$$

где g — ускорение свободного падения; h — глубина спокойной жидкости; η — высота гребня волны над поверхностью спокойной жидкости. Заметим: скорость распространения волны зависит от ее амплитуды! Этот результат наблюдательного ученого не был понят многими известными его коллегами и полемика продолжалась

полстолетия, пока голландцы Д. Кортевег и Г. де Фриз не вывели своего уравнения мелкой воды:

$$v_t + 6vv_x + v_{xxx} = 0.$$

Это уравнение — квазилинейное, параболического типа; v — скорость материальной точки жидкости, нижний индекс, как и ранее, обозначает частную производную по соответствующему аргументу.

Далее рассмотрим другое нелинейное уравнение распространения нелинейных волн, обладающее всеми особенностями уравнения для описания цунами, — интегродифференциальное уравнение в частных производных

$$v_t + vv_x + \int_{-\infty}^{\infty} K(x-s)v_s(s,t)ds = 0, \quad (6.58)$$

предложенное в 1967 г. Дж. Уиземом [23, 63]. Но прежде остановимся на важных «атрибутах» нелинейных волн — дисперсии и нелинейности.

Пусть рассматривается линейное дифференциальное уравнение в частных производных по x и t (с постоянными коэффициентами), которое в символической форме имеет вид

$$P(\partial/\partial t, \partial/\partial x)v = 0, \quad (6.59)$$

где P — некоторый полином. Подставим решение в нормальной форме

$$v = A \exp(ikx - i\omega t), \quad (6.60)$$

где k — волновое число (вещественное); ω — частота (комплексное число), в уравнение (6.59):

$$G(\omega, k)A \exp(ikx - i\omega t) = 0,$$

где G — полином. Экспонента не может обращаться в нуль при всех x, t , поэтому нетривиальное решение (6.59) ($A \neq 0$) существует, если k и ω связаны дисперсионным соотношением

$$G(\omega, k) = 0. \quad (6.61)$$

Предположим, что (6.61) позволяет определить ω как вещественную функцию волнового числа k . Для каждой моды (т.е. каж-

дого k) точки постоянной фазы перемещаются с фазовой скоростью

$$c = \omega/k. \quad (6.62)$$

Если отсутствует зависимость фазовой скорости от волнового числа k , т.е.

$$\frac{d^2\omega}{dk^2} = 0, \quad (6.63)$$

то говорят, что волна, описываемая уравнением (6.59), перемещается без дисперсии [1]. В противном случае имеет место дисперсия волн, фазовая скорость их становится различной и поведение решения зависит от того, как волны интерферируют друг с другом. В такой ситуации вводят групповую скорость $c_r = d\omega/dk$, значение которой состоит в том, что по истечении достаточно большого времени каждое волновое число доминирует в решении в области $x \sim c_r t + O(t)$.

Дисперсионное уравнение для волнового уравнения (6.7) было нами уже получено (6.21): $\omega = \pm ak$, откуда $c = a$, следовательно, волны распространяются без дисперсии. Для уравнения диффузии (6.53) имеем $\omega = \kappa k^2$, откуда $c = c_r/2 = \kappa k$ (κ – коэффициент диффузии), т.е. наблюдается дисперсия волн (иное было бы странно для уравнения диффузии). Как уже известно, решения (6.59) могут быть представлены в виде суперпозиции (суммы или интеграла) по k волн вида (6.60). Поэтому смысл дисперсии волн – в «размазывании» с течением времени волнового возмущения по пространству. Недиспергирующие среды переносят это возмущение без изменения его формы.

Сказанное относится к линейным уравнениям и к зависимости $\omega(k)$ (не путать с зависимостью скорости волны от ее амплитуды $c(A)$, существующей только для нелинейных волн). В пространственном случае (т.е. когда независимые переменные дифференциального уравнения суть t, r), нормальная форма решения имеет вид плоской волны $v = A \exp(i\mathbf{k} \cdot \mathbf{r} - i\omega t)$ (\mathbf{k} – волновой вектор). В неоднородных, или нелинейных, средах уравнение (6.59) становится почти линейным, квазилинейным или сильно нелинейным, и попытка дать определение дисперсии волн затруднительна, по-

скольку у упомянутых уравнений, как правило, отсутствуют решения в виде плоских волн.

Для исследования дисперсионных свойств уравнения Уизема рассмотрим вспомогательное линейное уравнение

$$v_t + \int_{-\infty}^{\infty} K(x-s)v_s(s,t)ds = 0. \quad (6.64)$$

Подставляем в него решение вида (6.60):

$$i\omega \exp(ikx) - \int_{-\infty}^{\infty} K(x-s)ik \exp(iks)ds = 0.$$

После деления на $ik \exp(ikx)$ и замены переменной $x-s = y$ под интегралом получаем

$$c(k) = \frac{\omega}{k} = \int_{-\infty}^{\infty} K(y) \exp(-iky) dy.$$

Поскольку правая часть представляет собой преобразование Фурье ядра $K(y)$, выполняя обратное преобразование, запишем

$$K(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c(k) \exp(ikx) dk.$$

Таким образом, задаваясь любой желаемой функцией $c(k)$ (дисперсионной зависимостью), можно определить соответствующее ядро $K(x)$, т.е. модельное уравнение описывает дисперсию волн.

Теперь остановимся на вопросе нелинейности. Нелинейность уравнения Уизема (6.58) связана с конвективным членом vv_x . Простейшее уравнение в частных производных, обладающее этой нелинейностью, имеет вид

$$v_t + vv_x = 0 \quad (6.65)$$

и носит название *уравнения Эйлера* [23]. Снабдим его начальными условиями

$$v(x,0) = v_0(x). \quad (6.66)$$

Но (6.65) нам уже встречалось раньше. Комбинация $v_t + vv_x$ часто присутствует в уравнениях динамики жидкости и газа и, как было показано, появляется при переходе от материальных (лагранжевых) координат к пространственным (эйлеровым). Без такого перехода вместо исходного квазилинейного уравнения вида $v_t + vv_x + L(v) = 0$ имели бы почти линейное уравнение $\dot{v} + \tilde{L}(v) = 0$: избавившись от vv_x , получаем неоднородные коэффициенты в остальной части оператора — $\tilde{L}(v)$. Однако в этом случае будем иметь решение, определенное в материальных (лагранжевых) переменных, координаты которых в пространстве сами требуют определения из системы дифференциальных уравнений (каких?). При описании развитого турбулентного движения второй подход является намного сложнее. К тому же в механике жидкости и газа граничные условия чаще задаются на пространственной (неподвижной) границе области (каналы, резервуары или просто произвольная область пространства, сквозь которую стационарно течет среда).

Ранее было найдено уравнение характеристик для (6.65): $dx/dt = v(x, t)$. Если бы поле $v(x, t)$ было постоянным, то общее решение (6.65) имело бы вид $v = f(x - vt)$. Подстановка такого представления в исходное уравнение с учетом начальных условий дает нам неявное решение

$$v = v_0(x - vt), \quad (6.67)$$

обнаруживающее некоторую зависимость скорости от амплитуды (а не позволит ли нам это описать свойства волн цунами и опрокидывание волн?).

Качественный ответ на вопрос можно получить с помощью характеристик. Пусть начальная функция $v_0(x)$ имеет вид, представленный на рис. 6.7,а. Выберем материальную точку с координатой ξ и построим проходящую через нее характеристику Γ_ξ : $x = \xi + v_0(\xi)t$ с углом наклона $1/v_0(\xi)$ (рис. 6.7,б).

Выберем момент времени t и определим графически из рис. 6.7,б эйлерову координату x_k материальной частицы ξ . Теперь точку $x, v_0(\xi)$ можно отложить на плоскости (x, v) . Выполнив эту процедуру со всеми материальными точками ξ оси, получим профиль $v(x, t)$ волны на плоскости (x, v) .

Обратимся вновь к рис. 6.7, а, б и заметим, что при данном начальном профиле характеристики должны неизбежно пересечься, начиная с некоторого t_p . Это означает, что при $t > t_p$ функция $v(x, t)$ оказывается неоднозначной (рис. 6.7, в).

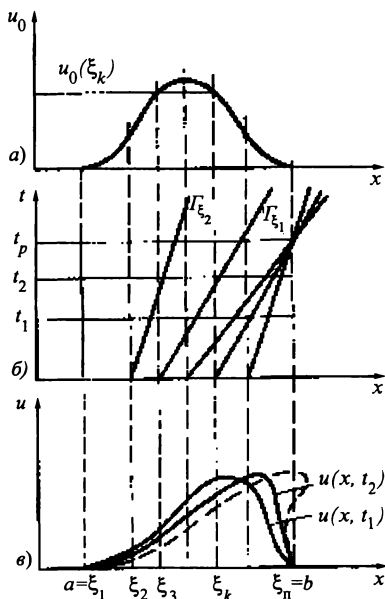


Рис. 6.7. Поведение решения уравнения Эйлера [23]

Из построенных графиков и (6.67) видно, что чем больше амплитуда начальной скорости $v_0(x)$ материальной точки в x , тем с большей скоростью она распространяется. Гребень волны неизбежно опережает ее подошву. То есть уравнение Эйлера описывает опрокидывание волн! Здесь нужно отметить, что интерпретация этого явления зависит от того, что понимать под смыслом решения уравнения. Если допустить неоднозначные решения $v(x, t)$, то столкнемся с опрокидыванием волны, а если считать его однозначным, то уравнение Эйлера опишет разрушение волны, т.е. образование разрыва скорости. Такое решение будет пониматься в ином, неклассическом смысле.

Нетрудно увидеть, что уравнение Уизема содержит в себе оба рассмотренных модельных уравнения. При этом единоборство механизмов опрокидывания (разрушения) и диссипации волн приводит к весьма богатой динамике уравнения. Теоретические исследования показали, что, во-первых, это уравнение имеет решения типа уединенных волн, распространяющихся с неизменной формой и постоянной скоростью. Скорость таких волн монотонно увеличивается с ростом амплитуды, причем существует предельная амплитуда. При этом оказывается, что гребень волны с предельной амплитудой заостряется («стоксово заострение»). Уравнение имеет и периодические решения («рябь»), обладающие теми же особенностями. Во-вторых, существуют режимы, при которых волны разрушаются.

Интересно исследовать процесс набегания волн на берег. Выбирая дисперсионное соотношение (у нас есть такое право) в виде

$$c(k) = \sqrt{g \operatorname{th}(kh) / k},$$

где h – глубина, g – ускорение свободного падения, th – гиперболический тангенс, и вычисляя по нему ядро $K(x)$, а также предполагая берег пологим (т.е. h постоянно на каждом этапе эволюции волн), получаем следующую картину (рис. 6.8). Там, где глубина сравнительно большая, волны являются гладкими, периодическими, бегущими; нелинейность при этом играет лишь роль фактора, сдерживающего дисперсию. При переходе в область меньших глубин задние гребни волн, имеющие большую скорость распространения, начинают догонять передние, в результате чего происходит уменьшение длины волны. Однако энергия волн, ос-

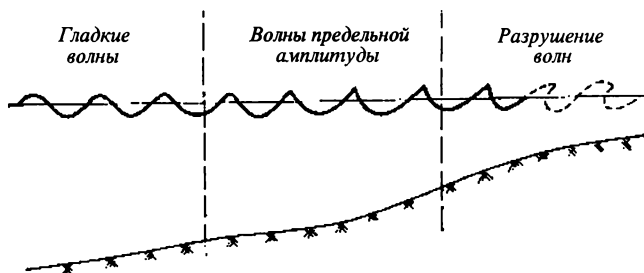


Рис. 6.8. Зависимость профиля волнового решения уравнения Уизема от глубины [23]

таваясь неизменной (а это еще одно замечательное свойство этого уравнения), приводит к увеличению амплитуды волн. Процесс происходит до некоторой предельной амплитуды и сопровождается заострением волн (нелинейность здесь проявляет себя именно в этом). Волны предельной амплитуды, переходя на мелководье, начинают опрокидываться, поскольку при $h \rightarrow 0$ $K(x) \rightarrow 0$, и нелинейность полностью «побеждает» диссипацию.

Необходимо сказать, что здесь нами было предпринято качественное моделирование явления. Мы не выводили уравнения движения волн на поверхности жидкости, а изначально изучали свойства некоторых модельных уравнений, из которых скомбинировано уравнение Уизема. Такой путь, требующий большой физической интуиции, является весьма ценным, но не заменяет собой прямого («честного») пути построения модели: постановки проблемы, принятия гипотез, вывода уравнений и их исследования. Специалист в области математического моделирования должен овладеть обоими подходами.

6.14. ИЕРАРХИЧЕСКИЕ МОДЕЛИ ТУРБУЛЕНТНОСТИ И МНОГОМАСШТАБНЫЕ ФУНКЦИОНАЛЬНЫЕ БАЗИСЫ

*Когда я предстану перед Творцом, то первое,
о чем я спрошу его, будет:*

«В чем же загадка турбулентности?»

Т. фон Карман

Сложные нелинейные системы часто имеют иерархическую структуру, состоящую в том, что элементы этой системы образуют иерархию уровней, а взаимодействия между ними становятся существенными только в том случае, если элементы близки друг другу и «по вертикали», и «по горизонтали». Яркий пример такой системы – развитая турбулентность, характерной особенностью которой является многомасштабность. В турбулентном потоке существуют вихри самых различных размеров – от размеров порядка характерного масштаба потока (диаметра трубы, размера обтекаемого тела, размера сосуда и т.д.), называемого интегральным, или макромасштабом турбулентности, до наименьшего, диссипативного, или микромасштаба, на котором становятся существенными вязкие силы, переводящие подводимую к потоку механичес-

кую энергию в тепловую. Тот факт, что макромасштабы, на которых подводится энергия, и микромасштабы, на которых энергия диссипирует, могут отличаться в тысячи (или даже миллионы) раз и определяет необходимость формирования каскада энергии от больших вихрей к меньшим.

В этом разделе будут рассмотрены модели турбулентности, основанные на идее применения функционального базиса специального типа, наиболее точно соответствующего структуре турбулентных полей. Идея такого базиса впервые была предложена В. Зиминым в конце 70-х годов и состояла в использовании семейства самоподобных функций прогрессивно убывающего масштаба [44]. Базис был назван иерархическим, на его основе построены и исследованы многочисленные модели, также названные иерархическими [45]. В конце 80-х годов в научной литературе появилось слово «вейвлет», а к началу 90-х вейвлет-анализ превратился в самостоятельную, хорошо развитую область математической физики. Идеи, лежащие в основе теории вейвлетов, совпадают с идеями иерархического представления турбулентных полей, и в терминах этой молодой науки иерархические модели — это модели, построенные с помощью вейвлет-представления описываемых полей.

Поскольку цель этой главы состоит в изложении подходов к моделированию сложных нелинейных процессов, то начнем с идей, приведших к иерархическим моделям. В то же время нельзя не остановиться и на формулировке основных положений вейвлет-анализа, который оказывается чрезвычайно полезным при анализе временной и пространственной структуры нелинейных систем.

Иерархический базис для турбулентных полей

Рассматривая численные методы решения уравнений движения жидкости, можно отметить, что чаще всего применяются либо сеточные, либо спектральные методы, либо их комбинация. И те, и другие можно отнести к проекционным методам решения уравнений в частных производных, когда для решения используют проекции всех полей на функциональные базисы.

В сеточных методах функции представлены значениями в точках, плотность которых связана со спектральными свойствами рассматриваемых полей (мелкомасштабные вихри не должны «проваливаться» между точками сетки). Более строго эта связь выражается теоремой Котельникова, согласно которой функция $f(x)$,

спектр которой ограничен пространственной частотой $2\pi/h$, может быть представлена суммой функций отсчетов (синусов), центры которых размещены на сетке с шагом h . Очевидно, что сеточное представление эффективно при описании локальных структур — мелкомасштабный вихрь описывается небольшим числом точек, находящихся в соответствующей области пространства. В то же время для описания даже очень простого по структуре крупномасштабного вихря требуется использование всех базисных функций.

Спектральные методы используют разложение по фурье-гармоникам. В этом случае каждая базисная функция описывает, по сути, систему когерентных вихрей, занимающую все пространство. В таком представлении очень просто описать вихрь, занимающий всю область, или периодическую систему вихрей — и в том, и в другом случае достаточно одной базисной функции. Однако если требуется описать отдельный вихрь, занимающий малую часть рассматриваемой области, то потребуется весь гармонический ряд.

Сеточные методы эффективны при вычислении нелинейных членов, так как позволяют выразить значение в точке через небольшое число соседних точек, но приводят к большим затратам машинного времени при решении уравнения Пуассона, требующего построения итерационного процесса, в который вовлечены все точки области. Спектральные методы, наоборот, делают решение уравнения Пуассона тривиальным, но приводят к очень сложной структуре нелинейных членов.

Проблемы двух функциональных базисов связаны с их локализованностью в физическом и в фурье-пространствах. Сетки строго локализованы в физическом пространстве, но спектр точки (дельта-функции) представляет собой белый шум. Это означает, что функции делокализованы в пространстве Фурье. Обратная ситуация возникает при разложении в ряды Фурье. Каждая гармоника представляет строго одну частоту, но соответствующая ей функция занимает все физическое пространство.

В турбулентном потоке сосуществуют вихри самого различного масштаба, но наиболее эффективные взаимодействия происходят между вихрями (структурами), близкими и в физическом, и в фурье-пространстве. Первое очевидно: чтобы вихри взаимодействовали, они должны перекрываться в пространстве. Второе утверждение составляет основу концепции каскадных процессов — взаимодействуют вихри сравнимых размеров (если размеры несопоставимы, то маленькие вихри просто переносятся большими без обмена энергией). Это заставляет обратиться к поиску специаль-

ных функций, более точно соответствующих структуре турбулентного потока.

Еще один недостаток использования рядов Фурье состоит в низкой информативности высоких частот. Хорошо понятен смысл рассмотрения вихрей с характерным размером $L, L/2, L/3, \dots$, но отдельное описание масштабов $L/957, L/958, L/959, \dots$ мало оправдано. Это соображение наводит на мысль о необходимости использования функций, масштаб которых изменяется прогрессивно, — такое соотношение получается при равномерном разбиении пространства масштабов в логарифмическом представлении. В теории турбулентности важную роль играет и идея масштабного подобия. Это значит, что искомым базис должен быть составлен из подобных функций.

Суммируя сказанное, можно сформулировать требования, которым должен удовлетворять функциональный базис, предназначенный для описания турбулентных потоков:

- функции базиса должны быть локализованы и в физическом, и в фурье-пространствах;
- функции должны быть подобны и описывать иерархию вихрей прогрессивно убывающих масштабов;
- мелкомасштабные вихри должны переноситься в поле вихрей большего масштаба;
- при подстановке в уравнения Навье—Стокса (уравнения движения линейно-вязкой жидкости) функциональный базис должен приводить к слабосвязанной динамической системе.

Попробуем построить базис, удовлетворяющий этим требованиям. Построения будем проводить для двумерного случая, так как это упрощает иллюстрацию результатов и запись функций.

Итак, имеем двумерное пространство $r = (x, y)$ и соответствующее ему пространство волновых векторов $k = (k_x, k_y)$. Фурье-плоскость разобьем на кольцевые зоны (рис. 6.9) таким образом, что для зоны с номером N

$$k_N < |k| < k_{N+1}, \quad k_N = \pi 2^N, \quad N = 0, \pm 1, \pm 2, \dots \quad (6.68)$$

Каждая кольцевая зона включает, таким образом, одну октаву волновых чисел (напомним, что октавой называется интервал, в пределах которого частота изменяется в 2 раза).

Рассмотрим поле скорости $v(t, x, y)$ и представим его в виде

$$v(t, x, y) = \sum_N v_N(t, x, y), \quad (6.69)$$

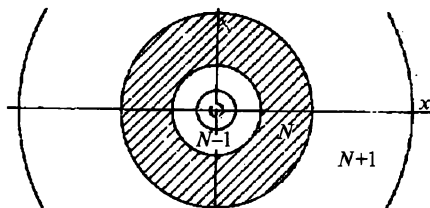


Рис. 6.9. Разбиение фурье-плоскости на кольцевые зоны

где каждая функция v_N есть результат фильтрации в фурье-плоскости по соответствующему кольцу (6.68):

$$v_N(t, x, y) = \iint v(t, x', y') g_N(x - x', y - y') dx' dy'. \quad (6.70)$$

Здесь $g_N(r)$ есть функция, фурье-образ которой $\tilde{g}_N(k)$ локализован в кольце:

$$\tilde{g}_N(k) = \begin{cases} 1 & \text{в кольце } N, \\ 0 & \text{вне кольца } N. \end{cases} \quad (6.71)$$

В силу определения операции фильтрации (6.70)–(6.71) $\int v_N v_M dr \sim \delta_{NM}$ и, следовательно, энергия распадается на сумму

$$E = \sum_N E_N = \sum_N \frac{1}{2} \int v_N^2 dr.$$

Таким образом, проведена первая часть построения – исходное поле разбито по масштабам. На втором этапе нужно провести разбиение полученных полей v_N на сумму функций, каждая из которых характеризует поле скорости данного масштаба только в определенной области пространства:

$$v_N = \sum_n a_{Nn}(t) f_N(r - r_{Nn}),$$

где $f_N(r)$ – базисные функции масштаба N ; r_{Nn} – радиус-вектор центра вихря (функции). Функции $f_N(r)$ должны быть подобны, взаимно ортогональны и обеспечивать разряженную матрицу нелинейных взаимодействий X_{NnMmLl} в уравнении

$$d_i a_{Nn} = \sum X_{NnMmLl} a_{Mm} a_{Ll} + \dots,$$

получаемся при проектировании уравнений Навье–Стокса (уравнений движения линейно-вязкой жидкости) на функциональный базис. Структура связей, описываемых матрицей X_{NnMmLl} , показана на рис. 6.10. Каждый горизонтальный слой (ярус) соответствует совокупности функций определенного масштаба, а все вместе описывают иерархическую систему, давшую название моделям этого типа.

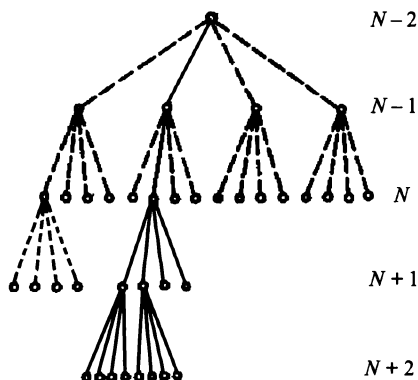


Рис. 6.10. Структура связей, описываемых матрицей X_{NnMmLl}

На свойствах и возможностях иерархических моделей останавливаться не будем, но поговорим о функциях, используемых при их построениях. Увы, удовлетворить всем приведенным требованиям к базисным функциям не удастся. Задача имеет решение в такой постановке только в одномерном случае. Одномерный базис не представляет интереса с точки зрения описания турбулентности, но его построение имеет методический интерес, и мы его проведем.

Одномерный иерархический базис

Рассмотрим функцию $f(x)$, для которой существует преобразование Фурье:

$$\tilde{f}(\gamma) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i \gamma x} dx.$$

Ось волновых чисел γ (напомним, что $k = 2\pi\gamma$) разбиваем на октавы $\gamma_N = 2^N$ (рис. 6.11) и вводим функции

$$\tilde{f}_N(\gamma) = \begin{cases} \tilde{f}(\gamma) & \gamma_N < \gamma < \gamma_{N+1}, \\ 0 & \text{вне зоны.} \end{cases} \quad (6.72)$$

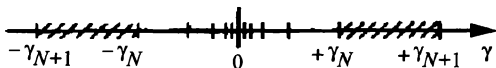


Рис. 6.11. Разбиение на октавы оси волновых чисел

Очевидно, что $\tilde{f}(\gamma) = \sum \tilde{f}_N(\gamma)$. Полученные функции \tilde{f}_N обладают замечательным свойством — они допускают периодическое продолжение на всю ось γ с периодом $2\gamma_N$ (рис. 6.12):

$$\tilde{F}_N(\gamma) = \begin{cases} \tilde{f}_N(\gamma - 2(m-1)\gamma_N) & (2m-1)\gamma_N < \gamma < 2m\gamma_N, \\ \tilde{f}_N(\gamma - 2(m+1)\gamma_N) & 2m\gamma_N < \gamma < (2m+1)\gamma_N. \end{cases}$$

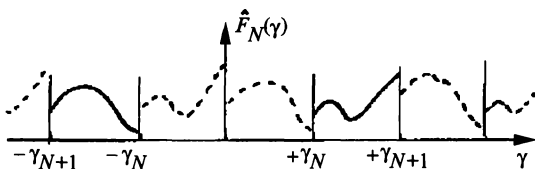


Рис. 6.12. Периодическое продолжение функций

Это позволяет разложить функции $\tilde{F}_N(\gamma)$ в ряд Фурье

$$\tilde{F}_N(\gamma) = \sqrt{h_N} \sum_n A_{Nn} e^{-2\pi i h_N n \gamma}, \quad (6.73)$$

где $h_N = 1/(2\gamma_N)$. Функции $h_N e^{-2\pi i h_N n \gamma}$ образуют полный базис в классе функций \tilde{F}_N , а те же функции, определенные внутри зоны (6.72), — полный базис в классе функций \tilde{f}_N .

Чтобы получить вид базисной функции в физическом пространстве, нужно взять обратное преобразование Фурье. Получается функция вида

$$f_{Nn}(x) = \frac{1}{\sqrt{h_N}} \frac{\sin\left[\frac{\pi}{2h_N}(x - h_N n)\right]}{\frac{\pi}{2}(x - h_N n)} \cos\left[\frac{3\pi}{2h_N}(x - h_N n)\right]. \quad (6.74)$$

Вид функции (6.74) для $n = 0$ показан на рис. 6.13. Эти функции известны в математике как функции Литлвуда–Пелли. Функции медленно убывают в физическом пространстве ($f_{Nn}(x) \sim x^{-1}$), что является результатом обрыва функций в пространстве Фурье. Все базисные функции взаимно ортогональны, т.е.

$$\int f_{Nn}(x) f_{Nm}(x) dx = \delta_{NM} \delta_{nm},$$

что следует из ортогональности функций в фурье-пространстве и инвариантности скалярного произведения двух функций относительно преобразования Фурье. Коэффициенты в разложении (6.73) определяются формулой

$$A_{Nn} = \int f(x) f_{Nn}(x) dx.$$

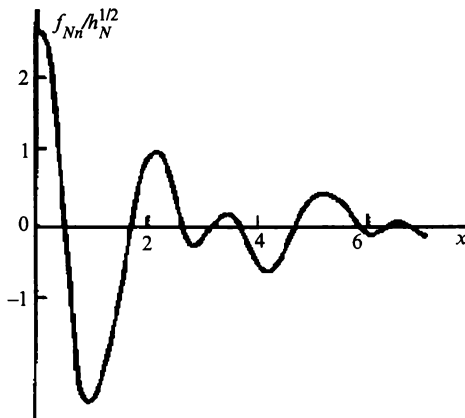


Рис. 6.13. Вид функции (6.74) для $n = 0$

Базисные функции имеют двойную индексацию. Большой индекс отвечает за масштаб, малый — за положение функции в пространстве. Увеличение индекса N на единицу сжимает функцию вдвое, увеличение индекса n на единицу сдвигает функцию вдоль оси x на величину h_N .

Двумерный базис

Простейший способ получения двумерного базиса состоит в определении двумерной функции как произведения одномерных:

$$f_{NnMm}(x, y) = f_{Nn}(x)f_{Mm}(y).$$

Однако такие функции не являются изотропными и не удовлетворяют требованию подобия. Последнее обстоятельство не оставляет надежд на получение простой динамической системы для коэффициентов разложения.

Исходя из локальной изотропии мелкомасштабной турбулентности и стремления получить базис, образованный разномасштабными, но однотипными функциями, можно построить относительно простой, но «не совсем ортогональный» базис. Для этого поле скорости представляется в виде $v(t, x, y) = \sum_{Nn} a_{Nn}(t)v_{Nn}(r - r_{Nn})$, где a_{Nn} — зависящая от времени амплитуда; v_{Nn} — осесимметричная базисная функция, у которой большой индекс отвечает за масштаб, а малый — за положение в пространстве; r_{Nn} — радиус-вектор центра функции.

Используем введенное выше разбиение спектральной плоскости на расширяющиеся кольцевые зоны (6.68) и определим базисную функцию так, чтобы ее фурье-образ был равен константе в пределах соответствующего кольца:

$$\hat{v}_{Nn}(\gamma) = \begin{cases} \frac{i(\gamma \times e)}{\sqrt{3\pi^3 \gamma^2}} 2^{-N} \exp(-2\pi i \gamma r_{Nn}) & \gamma_N < |\gamma| < \gamma_{N+1}, \\ 0 & \text{вне зоны,} \end{cases} \quad (6.75)$$

где e — единичный вектор, перпендикулярный рассматриваемой плоскости. Экспоненциальный множитель задает сдвиг центра вихря в физическом пространстве. Числовой коэффициент выбран из условия нормировки.

Чтобы получить вид функции в физическом пространстве, нужно взять обратное преобразование Фурье от (6.75). Соответствующие вычисления дают

$$v_{Nn}(r - r_{Nn}) = \frac{2^{-N}(s \times e)}{\sqrt{3\pi}} \left(\frac{J_0(2s) - J_0(s)}{s} \right), \quad (6.76)$$

где $s = \pi 2^N |r - r_{Nn}|$; $J_0(s)$ — функция Бесселя.

Мы оставили без внимания вопрос о числе базисных функций и об их распределении в пространстве. Плотность функций в физическом пространстве можно оценить исходя из принципа неопределенности. Если области локализации в r и k пространствах имеют, соответственно, размеры Δr и Δk , то, требуя $\Delta r \Delta k = 2\pi$, получаем, что плотность функций заданного масштаба ρ_N связана с площадью области локализации функции в пространстве Фурье ΔS_k как

$$\rho_N = \frac{\Delta S_k}{4\pi^2} = \frac{3\pi}{4} 2^{2N}. \quad (6.77)$$

При вычислении (6.77) учитывалось, что ΔS_k есть площадь кольцевой области (6.68). Формула (6.77) отражает тот факт, что число вихрей при переходе от масштаба к масштабу возрастает в 4 раза (естественно, что в трехмерном случае это отношение будет равно восьми).

Вопрос о распределении функций в пространстве более сложен. Формулируя требования к базису, мы хотели воспроизвести структуру турбулентного потока, в котором мелкие вихри переносятся крупными. Это означает, что радиус-вектор центра функции должен подчиняться уравнению

$$d_l r_{Nn} = \sum_{M < N} \sum_m a_{Mm} v_{Mm}(r_{Nn} - r_{Mm}). \quad (6.78)$$

Подчеркнем, что суммирование в (6.78) ведется только по масштабам, большим данного.

Введенный таким образом базис ортогонален по индексу N , поскольку в фурье-пространстве функции различного масштаба занимают неперекрывающиеся области. Неортогональность функций по малому индексу, отвечающему за положение вихрей в пространстве, можно оценить путем вычисления интеграла $\int v_{Nn} v_{Mm} dr$ для

двух вихрей одного масштаба, расположенных друг от друга на расстоянии $\rho_N^{-1/2}$, равном среднему расстоянию между вихрями данного масштаба. Такая оценка дает для функций (6.76) значение порядка 0,1.

6.15. ВЕЙВЛЕТЫ

Ошибки в исследованиях обычно проистекают из свойственной человеческому разуму склонности недооценивать или же преувеличивать значение исследуемого предмета из-за неверного определения его удаленности от нас.

Э. По

В самых разных областях науки возникают задачи, связанные с анализом пространственных полей со сложной, многомасштабной структурой либо временных сигналов с меняющимся со временем спектральным составом. Эти задачи заставляли исследователей пытаться построить специальные функциональные разложения, близкие по своей идеологии описанному выше иерархическому базису. Центральной идеей всех этих подходов было использование базиса, каждая функция которого характеризует как определенную пространственную (временную) частоту, так и место ее локализации в физическом пространстве (во времени).

Слово «вейвлет» (от англ. wavelet — маленькая волна или рябь) было введено А.Гроссманном и Ж.Морле в 1984 г. в [123], посвященной проблеме анализа сейсмических сигналов, в которых требуется выделить и время (положение) всплеска в сигнале, и его спектральный состав (масштаб). Статья, где были сформулированы основные определения и доказаны основополагающие теоремы, вызвала огромный интерес, и уже к началу 90-х годов вейвлет-анализ превратился в развитую область математической физики, нашедшую широкое применение в задачах анализа временных сигналов, распознавания образов и синтеза изображений, шифровки и дешифровки информации и многих других.

Как уже отмечалось, вейвлеты используются как при анализе временных сигналов, так и при исследовании структуры пространственных полей. Временные ряды представляют собой одномерный сигнал, и все основные идеи проще продемонстрировать на задачах анализа временных последовательностей. По этой причине

забудем на некоторое время о пространственных полях и переключимся на сигналы вида $f(t)$.

Первая попытка построить функциональный базис, состоящий из функций, каждая из которых характеризует пульсации определенной продолжительности в конкретный момент времени, принадлежит А. Хаару (1909). Первые семь функций Хаара, построенные на единичном отрезке, показаны на рис. 6.14. Каждая функция представляет собой пару следующих друг за другом прямоугольных импульсов с разными знаками и одинаковой длительностью. Среднее значение любой функции равно нулю, а совокупность функций образует полный ортонормированный базис. Каждая функция строго локализована в физическом пространстве (во времени), но характеризуется медленно спадающим спектром частот (как $1/\nu$).

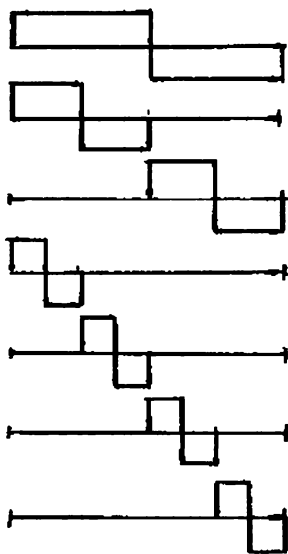


Рис. 6.14. Функции Хаара

Следующим шагом стали функции Литлвуда–Пелли (1937). Именно это семейство функций получается при построении одномерного иерархического базиса. Функции строятся путем вырезания полосы частот в пространстве Фурье. Это дает строгую лока-

лизацию в пространстве частот, но медленное затухание функции в физическом пространстве (во времени): функции описывают осцилляции, амплитуда которых падает как $1/t$.

Важным этапом в развитии идеи локального анализа спектральных (частотных) свойств стало преобразование Габора (1946), называемое также фурье-преобразованием в окнах. Функции Габора представляют собой гармонический сигнал, модулированный функцией Гаусса. Они хорошо локализованы и во времени, и в частотах, но каждая функция Габора характеризуется тремя параметрами: положением центра окна t_0 , шириной окна τ и частотой осцилляций ν (рис. 6.15). При этом функции различного масштаба не являются подобными (имеют различное число осцилляций).

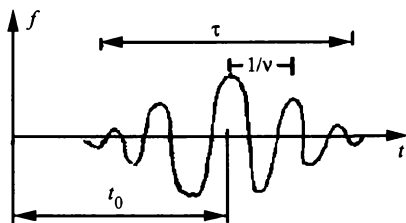


Рис. 6.15. Функция Габора

Вейвлеты объединили в себе два важных свойства: подобие и выраженную локализованность в физическом и фурье-пространствах. Сформулируем требования, которым должно удовлетворять семейство функций, чтобы быть вейвлетами.

1. *Допустимость.* Функция $\psi(t)$, которую будем называть *анализирующим вейвлетом* (употребляют также термин *материнский вейвлет*), должна иметь нулевое среднее значение:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0. \quad (6.79)$$

2. *Подобие.* Все функции семейства получаются из анализирующего вейвлета путем масштабного преобразования и сдвига,

$$\psi_{a,b}(t) = \psi\left(\frac{t-b}{a}\right).$$

Таким образом, вейвлеты образуют двухпараметрическое семейство функций, в котором параметр a отвечает за масштаб (растяжение) функции, а параметр b — за ее положение (сдвиг).

3. *Обратимость.* Вейвлет-преобразование должно быть обратимо, т.е. должно существовать обратное преобразование, однозначно восстанавливающее исходную функцию по ее вейвлет-представлению.

4. *Регулярность.* Функция $\psi(t)$ должна быть хорошо локализована и в физическом пространстве, и в пространстве Фурье.

Согласно последнему требованию, и функции Хаара, и функции Литлвуда — Пелли не попадают под определение вейвлетов. По сути, они являют собой два предельных случая (в одном случае резкие границы в физическом пространстве приводят к бесконечным «хвостам» в пространстве частот и, наоборот, обрыв в пространстве частот дает бесконечные «хвосты» в физическом пространстве).

В отличие от преобразования Фурье, вейвлет-преобразование допускает широкий выбор анализирующей функции. Согласно первому требованию, вейвлет всегда является знакопеременной функцией, включающей обычно небольшое количество осцилляций. Выбор конкретного вида вейвлета зависит от целей проводимого анализа.

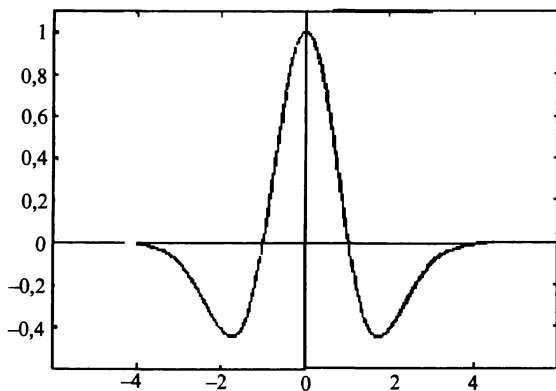
Приведем несколько примеров широко используемых вейвлетов. Простым вещественным вейвлетом, применяемым в задачах, требующих хорошего пространственного разрешения и не требовательных к спектральному разрешению, является вейвлет, получивший название «мексиканская шляпа» (рис. 6.16,а):

$$\psi(t) = (1 - t^2)e^{-t^2/2}. \quad (6.80)$$

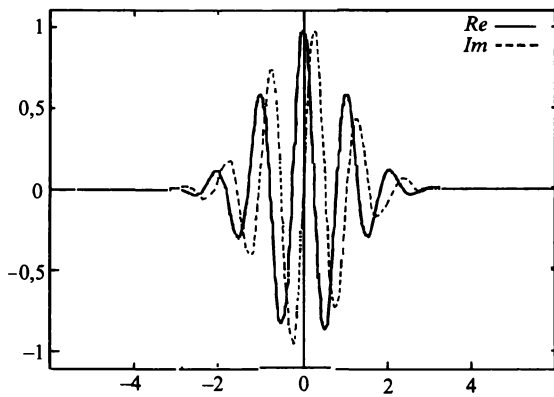
В задачах, требующих лучшего спектрального разрешения, часто используется вейвлет Морле — комплексная функция вида

$$\psi(t) = e^{-t^2/2}e^{i\omega_0 t}. \quad (6.81)$$

На рис. 6.16,б сплошной линией показана его вещественная часть, а штриховой — мнимая. Сама функция (6.81) совпадает с видом функций, используемых в преобразовании Габора, но семейство вейвлетов отличается от функций Габора тем, что один раз выбрав частоту ω_0 для анализирующего вейвлета и задав тем самым число осцилляций, мы в дальнейшем сжимаем или растягиваем функцию как целое, не нарушая подобия отдельных функций семейства.



а)



б)

Рис. 6.16. Графики вейвлетов: а – «мексиканская шляпа»; б – Морле

Непрерывное вейвлет-преобразование

Непрерывное вейвлет-преобразование одномерной функции есть

$$w(a, b) = a^{\kappa} \int_{-\infty}^{\infty} f(t) \psi^* \left(\frac{t-b}{a} \right) dt, \quad (6.82)$$

где $\psi(t)$ — вещественная или комплексная функция, удовлетворяющая требованиям 1–4 (знак * вводится для комплексно сопряженных величин). Если $\hat{\psi}(\omega) = \int \psi(t)e^{-i\omega t} dt$ есть фурье-образ анализирующего вейвлета и выполнено условие

$$C_{\psi} = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty, \quad (6.83)$$

то для преобразования (6.82) существует формула обращения

$$f(t) = \frac{1}{C_{\psi}} \int_0^{\infty} \int_{-\infty}^{\infty} \psi\left(\frac{t-b}{a}\right) w(a,b) \frac{da db}{a^{3+k}}. \quad (6.84)$$

Условие (6.83) эквивалентно условию (6.79), так как интеграл (6.83) расходится при наличии в спектре вейвлета нулевых частот, что равносильно отличному от нуля среднему значению. В определении (6.82) присутствует параметр k — показатель степени масштабного множителя. Конкретный выбор этого параметра зависит от целей анализа. Широко используется нормировка $k = -1$, при которой равные значения вейвлет-коэффициентов $w(a, b)$ соответствуют равным амплитудам пульсаций сигнала, независимо от масштаба пульсаций.

Вейвлет-образ $w(a, b)$ функции $f(t)$ можно выразить и через ее фурье-образ $\hat{f}(\omega)$. Действительно,

$$\hat{f}(\omega) = \frac{1}{C_{\psi}} \int_0^{\infty} \int_{-\infty}^{\infty} \hat{\psi}(a\omega) w(a,b) e^{-i\omega b} \frac{da db}{a^{2+k}}, \quad (6.85)$$

а

$$w(a,b) = \frac{a^{k+1}}{2\pi} \int_{-\infty}^{\infty} \hat{\psi}^*(a\omega) \hat{f}(\omega) e^{i\omega b} d\omega. \quad (6.86)$$

Пользуясь соотношениями (6.85)–(6.86) и теоремой Парсеваля, несложно получить аналог этой теоремы для вейвлет-преобразования

$$\int_{-\infty}^{\infty} f_1(t) f_2^*(t) dt = \frac{1}{C_{\psi}^2} \int_0^{\infty} \int_{-\infty}^{\infty} w_1(a,b) w_2^*(a,b) \frac{dadb}{a^{3+2k}},$$

из которого, в частности, следует

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 d\omega = \frac{1}{C_{\Psi}^2} \int_0^{\infty} \int_0^{\infty} |w(a,b)|^2 \frac{da db}{a^{3+2k}}. \quad (6.87)$$

Напомним, что в фурье-анализе спектральной плотностью энергии является величина $E(\omega) = |\hat{f}(\omega)|^2$ (называемая также спектром энергии) и введем величину

$$M(a) = \int_{-\infty}^{\infty} |w(a,b)|^2 db,$$

которая характеризует интенсивность *всех* пульсаций заданного масштаба. Если в определении вейвлет-преобразования положить $k = -1/2$, то формулу (6.87) можно переписать в виде

$$E = \int_0^{\infty} E(\omega) d\omega = \int_0^{\infty} M(a) \frac{da}{a^2}. \quad (6.88)$$

В этом случае $M(a)$ описывает распределение энергии пульсаций по масштабам и называется *интегральным вейвлет-спектром*. Из сказанного следует, что нормировка $k = -1/2$ должна использоваться, если результаты вейвлет-анализа предполагается сопоставлять с фурье-представлением сигнала. Действительно, если фурье-спектр следует степенному закону $E(\omega) \sim \omega^\alpha$, то при этой нормировке интегральный вейвлет-спектр будет иметь тот же степенной закон $M(a) \sim a^{-\alpha} \sim \omega^\alpha$ (это вытекает из формулы (6.88) с учетом того, что $\omega \sim 1/a$, а $d\omega \sim -da/a^2$).

Вейвлет-преобразование отображает пространство функций одной переменной (время) в пространство функций двух переменных (время и частота, или время и масштаб) и является избыточным. Избыточность непрерывного вейвлет-преобразования выражается в коррелированности вейвлет-коэффициентов, которая тем больше, чем больше рассматриваемый масштаб. Иначе говоря, чем больше масштаб, тем меньше независимых точек в вейвлет-разложении. Этот недостаток устраняется в дискретном вейвлет-представлении (пример тому – рассмотренный выше иерархический базис, в котором число функций геометрически уменьшается с ростом пространственного масштаба).

Преимущество вейвлет-преобразования перед преобразованием Фурье состоит в том, что оно позволяет проследить за изменением спектральных свойств сигнала со временем и указать, какие частоты (масштабы) доминируют в сигнале в каждый конкретный момент времени.

На рис. 6.17, *а, б*, показаны два примера вейвлет-разложения простых временных сигналов с помощью вейвлета Морле (6.81). В верхней части каждого рисунка показан модуль вейвлет-разложения на плоскости (a, b) , а в нижней – фаза. На рис. 6.17, *а* сигнал представляет собой суперпозицию двух гармоник, а в сигнале на рис. 6.17, *б* эти же две частоты появляются последовательно друг за другом. Фурье-преобразования этих двух сигналов практически не отличаются друг от друга, так как спектр Фурье теряет всякую информацию о том, когда какая гармоника присутствовала в сигнале. Вейвлет-анализ позволяет восстановить полную эволюцию спектрального состава сигнала во времени.

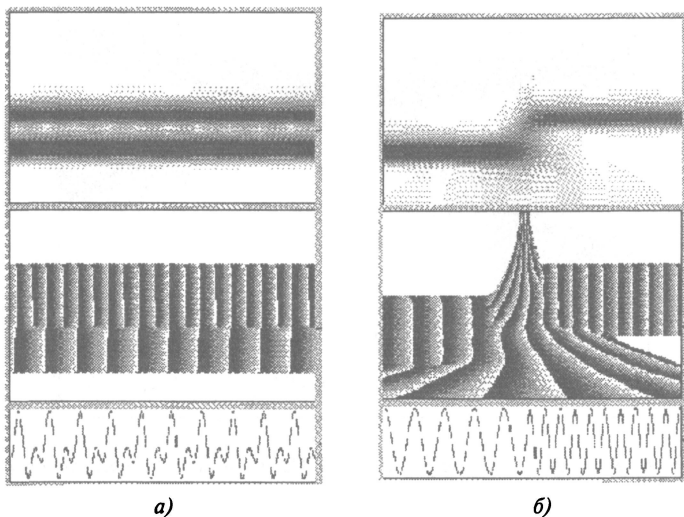


Рис. 6.17. Вейвлет-разложение Морле сигнала, изображенного в нижней части рисунка

Общее представление о спектрально-временной структуре сигнала можно получить по распределению модуля вейвлет-преобразования. Ширина полосы при разложении гармонического сигнала

ла характеризует спектральное разрешение используемого анализирующего вейвлета. Распределение фазы вейвлет-преобразования менее информативно, особенно для сложных сигналов. В то же самое время именно фаза дает наиболее точную информацию об особенностях (сингулярностях) в сигнале. Так, на рис. 6.17,б можно видеть, что именно по распределению фазы можно с большой точностью идентифицировать момент смены частоты.

На рис. 6.18 показан результат вейвлет-разложения сигнала, представляющего собой суперпозицию двух гармонических составляющих с непрерывно меняющимися частотами (снова использован вейвлет Морле). Сам сигнал показан в нижней части рисунка, модуль вейвлет-разложения – в верхней части. Вейвлет-представление позволяет получить точный вид эволюции частоты каждого из двух сигналов.

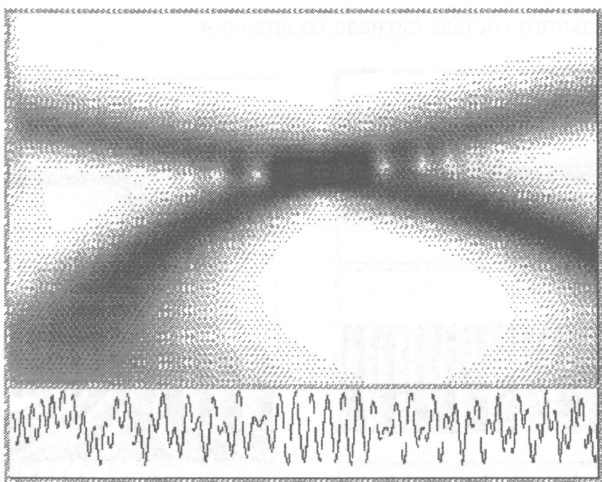


Рис. 6.18. Модуль вейвлет-разложения Морле сигнала, график которого приведен в нижней части рисунка

На рис. 6.19 дан пример использования действительного вейвлета типа (6.80). В качестве сигнала использован тот же временной ряд, что и в примере на рис. 6.17,б (удвоение частоты гармонических колебаний). В этом случае результатом преобразования является действительная величина, модуль которой показан на

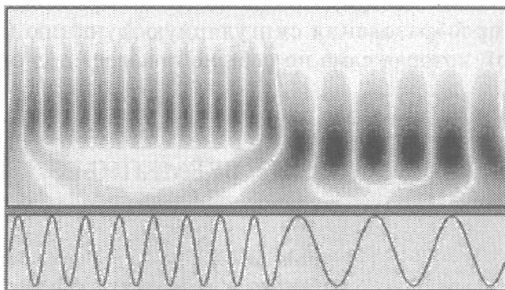


Рис. 6.19. Вейвлет-разложение сигнала, показанного в нижней части рисунка, с использованием вейвлета «мексиканская шляпа»

рисунке. Белые полосы на вейвлет-плоскости, неизбежно появляющиеся при работе с вещественными функциями, соответствуют смене знака вейвлет-коэффициентов и содержат, по сути, информацию, которую в комплексном представлении несет фаза.

В заключение отметим важное свойство вейвлет-представления функций – на этапе разложения сигнала по вейвлетам (анализа) и этапе восстановления исходного сигнала по его вейвлет-образу (синтеза) можно использовать различные семейства вейвлетов.

Пусть для анализа используется вейвлет $\psi(t)$, а для синтеза – вейвлет $\phi(t)$. Тогда прямое преобразование по-прежнему описывается выражением (6.82), а формула восстановления сигнала (6.84) примет вид

$$f(t) = \frac{1}{C_{\psi\phi}} \int_0^{\infty} \int_{-\infty}^{\infty} \phi\left(\frac{t-b}{a}\right) w(a,b) \frac{da db}{a^{3+k}}. \quad (6.89)$$

Восстановление (6.89) возможно, если выполнено условие

$$C_{\phi\psi} = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)\hat{\phi}^*(\omega)|}{|\omega|} d\omega < \infty.$$

Это условие «мягче», чем условие (6.83), так как теперь один из двух вейвлетов может и не удовлетворять требованию (6.83) (но при условии, что его «недостатки» компенсирует вейвлет, используемый на втором этапе). Преимущество восстановления по фор-

муле (6.89) состоит в том, что она позволяет применить на одном из этапов преобразования сингулярную функцию (например, δ -функцию), которая сама по себе не попадает под определение вейвлета.

6.16. ВЕЙВЛЕТ-АНАЛИЗ ВРЕМЕННЫХ КОЛЕБАНИЙ

*Ведь каждый день пред нами солнце ходит,
Однако ж прав упрямый Галилей.*

А.С. Пушкин

Спектры периодических (пусть даже достаточно сложных) колебаний имеют линейчатую структуру, описывая конечный набор частот, присутствующих в сигнале. При переходе к стохастическим колебаниям спектры становятся сплошными, а признаком развитой турбулентности служит, как известно, развитый инерционный интервал. Однако это не означает, что в развитых турбулентных течениях отсутствуют выделенные крупномасштабные пульсации. Экспериментальные исследования турбулентной конвекции в замкнутых объемах показывают, что течения на масштабах, сравнимых с размерами самой полости, характеризуются целыми сериями выделенных частот, причем периоды колебаний могут в тысячи раз превышать время оборота жидкости в полости. Эти результаты подкрепляются и наблюдениями за природными системами. Так, Солнце, являющее собой крупнейшую из доступных прямому наблюдению конвективных ячеек (именно конвекция — основной источник движения на Солнце, характеризуется гигантским значением числа Релея), демонстрирует целый набор циклов с периодами от нескольких дней до тысяч лет.

В качестве примера приложения вейвлет-анализа к исследованию временной изменчивости сложных гидродинамических систем рассмотрим результаты анализа солнечной активности по двум характеристикам: вариациям числа групп солнечных пятен и вариациям солнечного диаметра.

О том, что на Солнце есть пятна, знает каждый школьник. Многие знают и о том, что число этих пятен колеблется и достигает максимума примерно каждые 11 лет. Менее известен факт, что число пятен связано с интенсивностью магнитного поля Солнца. Эту связь поясняет рис. 6.20.

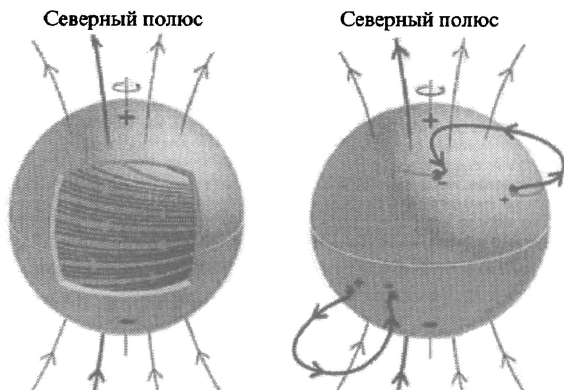


Рис. 6.20. Структура магнитного поля Солнца и ее связь с солнечными пятнами

Магнитное поле Солнца имеет полоидальную компоненту (силовые линии выходят на поверхность вблизи одного полюса и заходят вблизи другого) и более мощную азимутальную — ее силовые линии образуют замкнутые кольца внутри конвективной оболочки Солнца. При увеличении напряженности магнитного поля вследствие неустойчивости на этих магнитных линиях возникают гигантские петли, выходящие за пределы конвективной оболочки. В местах выхода магнитное поле направлено вертикально и подавляет конвективное течение, приносящее горячую плазму из недр Солнца. В результате температура оказывается ниже, чем на остальной поверхности, эта область видна как темное пятно. Чем сильнее магнитное поле, тем больше петель и тем больше пятен видно на поверхности Солнца.

Связь пятен с магнитными полями стала понятна не так давно, но само существование пятен на Солнце в свое время так взволновало человечество, что астрономы начали вести систематический подсчет этих пятен практически с того момента, как Галилей построил первый телескоп (конечно, иногда солнечные пятна можно было наблюдать невооруженным глазом и раньше). Долговременная запись среднемесячных чисел солнечных пятен начинается с наблюдений Галилея в феврале 1610 г., а с октября 1611 г. наблюдения становятся довольно регулярными. Существующий на сегодня ряд данных не имеет в астрономии аналогов по регулярности и продолжительности наблюдений.

График изменения числа солнечных пятен уже столетия привлекает внимание ученых, так как доказано, что многие процессы на Земле связаны с уровнем солнечной активности. Первое, что бросается в глаза при взгляде на график солнечной активности (рис. 6.21) – это череда пиков, каждый из которых охватывает приблизительно 11 лет. Это и есть знаменитый 11-летний солнечный цикл, характеризующий работу «солнечного динамо» – магнитогидродинамического генератора поля. Можно, однако, заметить, что амплитуда циклов непрерывно изменяется, а временами в работе «динамо» возникают сбои. Самый заметный сбой имел место в конце XVII – начале XVIII веков, когда в течение почти 50 лет пятен на Солнце практически не было. Этот период называют минимумом Маундера. Другое заметное ослабление солнечной активности было отмечено в начале XIX века и называется минимумом Дальтона.

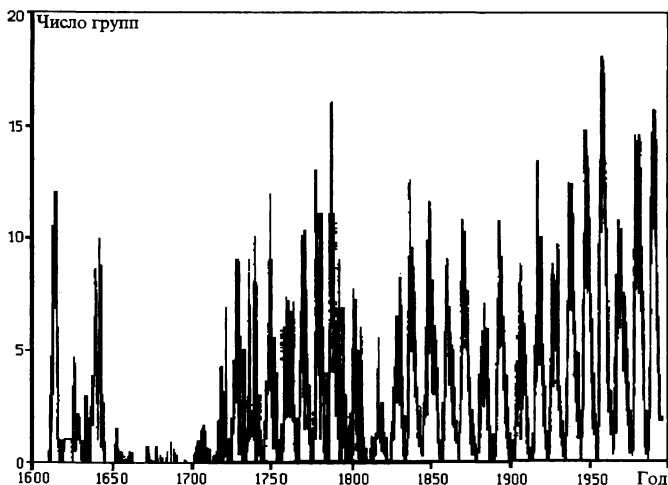


Рис. 6.21. Число групп солнечных пятен

Что нового могут дать вейвлеты в изучении записи числа солнечных пятен, если учесть, что сотни людей уже анализировали этот сигнал самыми разными методами? Для ответа на этот вопрос обратимся к результатам работ [122, 125].

Вейвлет-представление проектирует одномерный сигнал (который был функцией только времени) на плоскость «время–час-

тота» и позволяет увидеть изменение во времени спектральных свойств сигнала. На рис. 6.22 показан полученный с помощью вейвлета Морле модуль вейвлет-преобразования данных, представленных на рис. 6.21. На вейвлет-плоскости 11-летнему циклу соответствует темная горизонтальная полоса. При этом напомним, что идеально ровная горизонтальная полоса соответствовала бы ус-

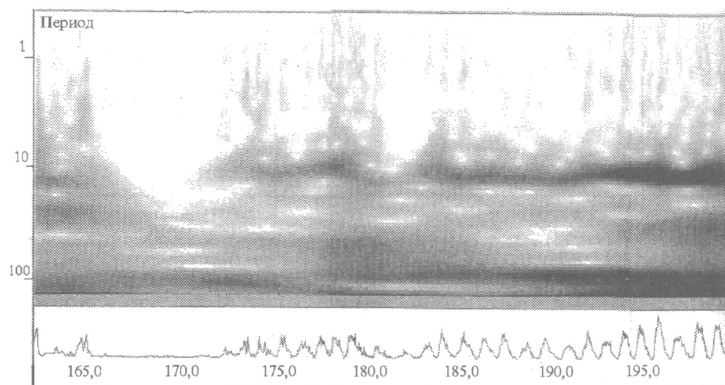


Рис. 6.22. Модуль вейвлет-преобразования Морле данных рис. 6.21

тойчивому периодическому колебанию. Как видно, кроме основного 11-летнего колебания в исследуемой записи присутствует еще одна приблизительно 100-летняя периодичность. Особенно хорошо эти периодичности видны на интегральном вейвлет-спектре (кривая *б* на рис. 6.23). На этом же рисунке для сравнения показан и спектр Фурье того же сигнала (кривая *а*), в котором 11-летний цикл выделяется на фоне сплошного частого пиков. По поводу значимости этих пиков велись споры долгие десятилетия. Сравнивая два спектра на рисунке, еще раз вспомним, что вейвлет-спектр является сглаженной версией спектра Фурье и не дает кратных гармоник при негармоническом характере колебаний.

Вейвлет-анализ позволяет проследить, как меняется длительность номинального 11-летнего цикла со временем, показывая, что 100-летний цикл фиксирует периодические попытки механизма генерации солнечного магнитного поля дать сбой и свернуть с обычных 11-летних колебаний в новый эпизод типа минимума Маундера. Удастся получить и неизвестную ранее количественную закономерность в формировании сбоев в работе солнечного динамо.

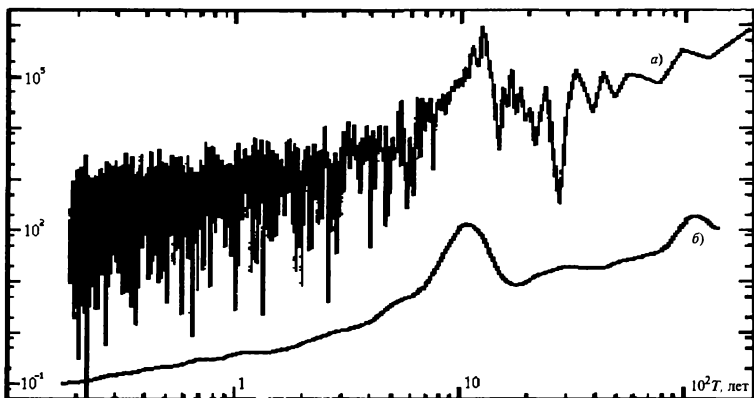


Рис. 6.23. Спектр Фурье (а) и интегральный вейвлет-спектр (б) сигнала, показанного на рис. 6.21

На рис. 6.24 приведен график изменения длины солнечного цикла со временем. Этот график получен путем оцифровки максимума в темной полосе, соответствующей на вейвлет-плоскости 11-летнему циклу. Вертикальными линиями на рисунке отмечены известные наблюдателям периоды снижения солнечной активности. Неожиданный результат состоит в том, что все эти периоды совпадают со спадающими участками на графике $T(t)$. Причем чем выше было значение T перед началом очередного минимума, тем глубже был сам минимум. Это обстоятельство совместно с имеющимся на сегодня значением периода солнечного цикла позволяет сделать вывод, что хотя очередного сбоя в солнечной активности и можно ожидать в начале текущего столетия, однако нового минимума Маундера случиться не должно.

На примере анализа солнечной активности покажем эффективность вейвлет-анализа в фильтрации сигналов и их совместной обработке. В эпоху знаменитого минимума Маундера постоянно измерялась еще одна характеристика Солнца – солнечный диаметр. Вариации видимого солнечного диаметра непрерывно регистрировались в парижской обсерватории с 1683 по 1718 г. (отдельные серии измерений проводились различными астрономами и ранее). Интерес к систематическим измерениям вариаций солнечного диаметра вновь появился только в наше время, и измерения были возобновлены, начиная с 1978 г.

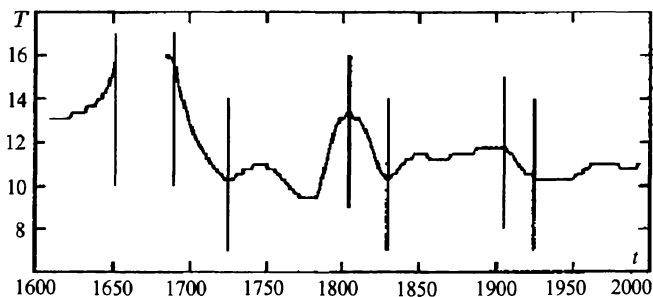


Рис. 6.24. Зависимость изменения длины солнечного цикла от времени

Все результаты измерений представлены на рис. 6.25. Очевидно существенное отличие современных данных от тех, что были выполнены четыре столетия назад. Напрашивается простое объяснение этому факту – качество измерений в то далекое время было намного ниже, что и обусловило высокий уровень пульсаций сигнала (систематическое отличие в уровне сигнала объясняется тем, что видимый диаметр Солнца – величина субъективная и зависит от способа его определения).

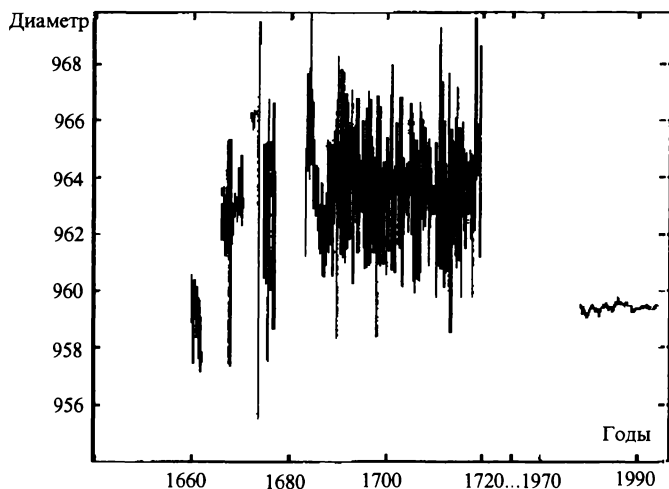


Рис. 6.25. Измерения вариаций солнечного диаметра с 1683 по 1718 гг.

Вейвлеты дают возможность изучить степень коррелированности двух сигналов отдельно на каждом временном масштабе. В сложной системе, какой является Солнце, вполне возможно представить ситуацию, когда какие-либо два сигнала скоррелированы на одних масштабах и практически независимы на других. Определим корреляционную функцию двух сигналов в виде

$$C(a) = \frac{\int w_1(a, b) w_2^*(a, b) db}{\left(\int w_1^2(a, b) db \int w_2^2(a, b) db \right)^{1/2}},$$

где w_1 и w_2 – вейвлет-образы рассматриваемых сигналов. Эта функция показывает, насколько в двух исследуемых сигналах связаны пульсации, характеризуемые определенным периодом, или временем релаксации.

На рис. 6.26 показана корреляционная функция, вычисленная для вариаций числа групп пятен и вариаций диаметра по перекрывающимся интервалам наблюдений. Видно, что на временах порядка 2 лет имеется узкий положительный пик, а на временных масштабах порядка 10 лет и более сигналы становятся строго антикоррелированы (больше пятен – меньше диаметр).

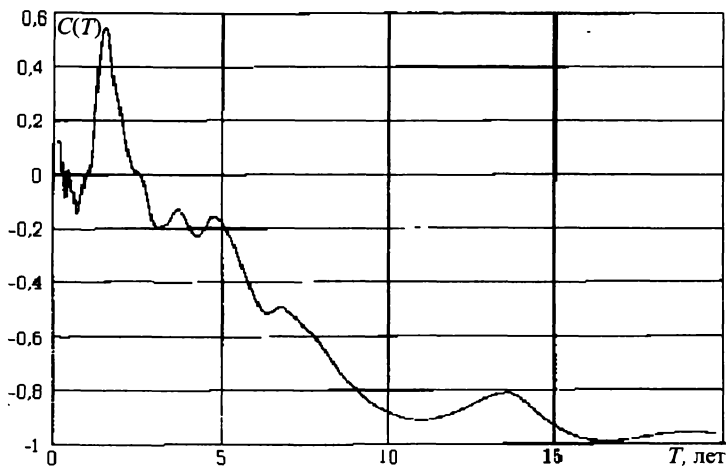


Рис. 6.26. Корреляционная функция, вычисленная для вариаций числа групп пятен и вариаций диаметра по перекрывающимся интервалам наблюдений

Наибольший интерес представляет частота основного (11-летнего) солнечного цикла. Выделяя из вейвлет-представления соответствующий временной масштаб, построим зависимости от времени вейвлет-коэффициентов $w(a, b)$ для $a = 11$ лет. Графики отфильтрованных 11-летних вариаций диаметра $\Delta D(b)$ и числа групп пятен $\Delta N(a)$ для интервала времени 1666–1718 показаны на рис. 6.27. Беспорной научной удачей можно считать тот факт, что наблюдения за изменениями солнечного диаметра начались во время минимума Маундера и продолжались во время выхода из минимума.

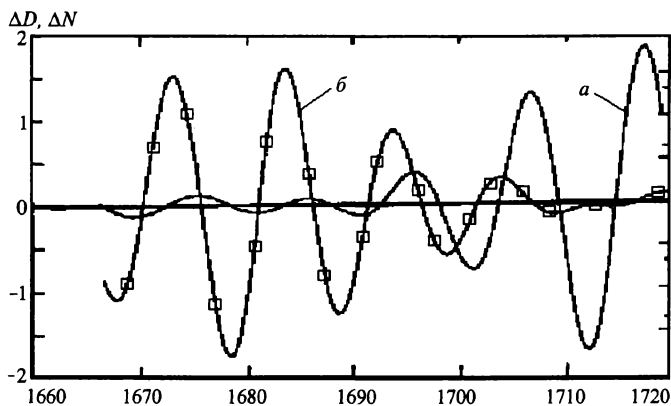


Рис. 6.27. Графики 11-летних вариаций диаметра и числа групп пятен для интервала времени 1666–1718 гг.

Результаты вейвлет-фильтрации данных наблюдений, представленные на рисунке, дали совершенно неожиданный результат, состоящий в том, что 11-летние вариации солнечного диаметра имели наибольшую амплитуду как раз во время глубокого минимума солнечной активности. По мере выхода из минимума вариации числа пятен начинают нарастать, а вариации диаметра спадать. Этот результат дает возможность объяснить разительное отличие современных данных от данных XVIII века: в сравнении с 1718 годом, когда были прекращены измерения диаметра, среднее количество групп пятен возросло примерно на порядок, а в свете полученной закономерности это должно было привести к суще-

ственному снижению интенсивности вариаций диаметра — что и подтверждают современные наблюдения.

Полученный результат заставляет пересмотреть сложившийся взгляд на природу солнечного цикла. 11-летний цикл объясняют, исходя из точки зрения, что он является свойством динамо-процессов. Следуя этой точке зрения, нужно признать, что во время остановки «динамо» исчезнуть должен и этот цикл. Приведенный результат заставляет думать, что природа 11-летнего цикла не связана собственно с динамо-процессом. Механизм его зарождения неясен, но представляется, что он действует независимо от динамо, модулируя активность последнего. Когда динамо не работает, энергия этого процесса выливается в гидродинамическую моду, приводя к 11-летним вариациям диаметра звезды.

6.17. О ФРАКТАЛАХ И ИХ ПРИМЕНЕНИИ

В 1953 году я понял, что прямая линия ведет человечество к упадку. Тирания прямой стала абсолютной. Прямая линия — это нечто трусливое, прочерченное по линейке, без эмоций и размышлений; это линия не существующая в природе. И на этом насквозь прогнившем фундаменте построена наша обреченная цивилизация.

Ф. Хундертвассер

Примеры фракталов

Фракталы окружают нас повсюду. Изрезанные береговые линии, изломанные поверхности горных хребтов, причудливые очертания облаков, раскидистые ветви деревьев, разветвленные сети кровеносных сосудов и нейронов, вспененные потоки бурных рек — все это фракталы. Одни фракталы, типа облаков и горных потоков, постоянно изменяются, другие, подобные деревьям и нейронным сетям, сохраняют свою форму неизменной.

Язык фрактальной геометрии природы оставался непонятым вплоть до появления в 1983 г. книги Мандельброта «Фрактальная геометрия природы». До этого времени естествоиспытатели говорили на языке геометрии Евклида. Идеально регулярные образы — прямая и плоскость, треугольники и пирамиды, окружность и сфера — составляли основу этого языка и всей научной картины

мира. Эту мысль в 1623 г. сформулировал Галилео Галилей: «Философия природы написана в величайшей книге — я разумею Вселенную, — которая всегда открыта перед нашими глазами, но понять ее сможет лишь тот, кто сначала выучит язык и постигнет письмена, которыми она начертана. А написана эта книга на языке математики, и письмена ее — треугольники, окружности и другие геометрические фигуры, без коих нельзя понять по-человечески их слова; без них тщетное кружение в темном лабиринте».

Потребовалось еще 350 лет, прежде чем естествознание обрело качественно новый язык фрактальной геометрии. Вот что об этом пишет первооткрыватель фрактальной геометрии Бенуа Мандельброт: «Почему геометрию называют холодной и сухой? Одна из причин заключается в ее неспособности описать форму облака, горы, дерева или берега моря. Облака — это не сферы, горы — это не конусы, линии берега — это не окружности, и кора не является гладкой, и молния не распространяется по прямой.... Природа демонстрирует нам не просто более высокую степень, а совсем другой уровень сложности. Число различных масштабов длин в структурах всегда бесконечно.

Существование этих структур бросает нам вызов в виде трудной задачи изучения тех форм, которые Евклид отбросил как бесформенные, — задачи исследования морфологии аморфного. Математики, однако, пренебрегли этим вызовом и предпочли все больше и больше отдаляться от природы, изобретая теории, которые не соответствуют ничему из того, что можно увидеть или почувствовать».

Широчайшее распространение фрактальных структур объясняется их разномасштабностью и самоподобием: и большие, и малые масштабы фрактальных структур имеют одинаковый закон построения. Форма фрактальной структуры, разглядываемая в микроскоп с любым увеличением, видится одной и той же. Это геометрическое подобие и есть основной принцип роста всего живого, который называют также иерархическим принципом организации (законы ветвления самой тонкой веточки дерева абсолютно те же, что и для всех его ветвей, и для всего ствола в целом).

Задать фрактальную структуру — значит задать не застывшую, неизменную форму, а принцип роста, закон изменения формы. Как правило, алгоритм построения формы гораздо проще, чем полученная с его помощью форма. Фрактал дает компактный способ описания самых замысловатых форм. Итак, «фрактал не есть конечная форма (фрактал никто никогда не видел, так же как

число π), а есть закон построения этой формы. Фрактал аккумулирует в себе идею роста» [22].

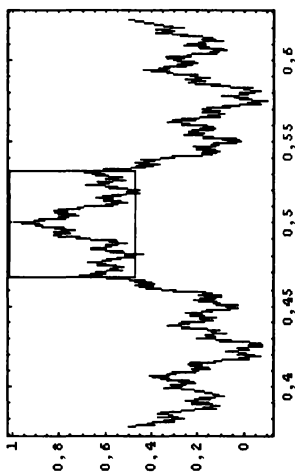
Осознание этой идеи привело к тому, что понятие фрактала стало широко использоваться в научных исследованиях, и было обнаружено большое число задач, в которых фрактальная структура и размерность служат основными характеристиками системы. Например, в турбулентности теория фракталов теснейшим образом связана с теорией масштабной инвариантности Колмогорова. Скорость турбулентного потока (как функция пространственных переменных и времени) — фрактал, аналогичный броуновской кривой, только с иными локальными свойствами.

Е. Федер в основу своей книги [109] положил исследование явлений, имеющих место при вытеснении нефти водой в пористой среде; эти явления приводят к тому, что нефть оказывается запертой в водяных ловушках, что ведет к ее потерям. Если вязкость вытесняемой жидкости больше вязкости вытесняющей, то фронт вытеснения неустойчив и образуются так называемые «вязкие пальцы», имеющие фрактальную структуру.

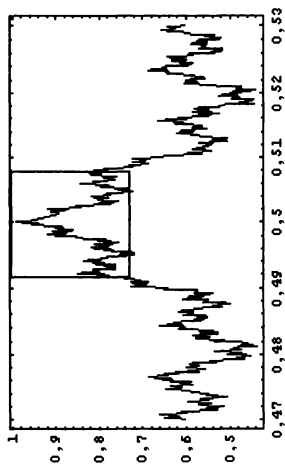
На рис. 6.28 представлена фрактальная кривая, являющаяся графиком так называемой функции Вейерштрасса. Эта функция непрерывна, всюду не дифференцируемая (каждая точка этой прямой — точка излома, производная в которой не существует). Она задается бесконечной суммой тригонометрических функций

$$W(x) = \sum_{n=1}^{\infty} a^n \cos(b^n \pi x), \quad a < 1, \quad b > 1, \quad ab > 1,$$

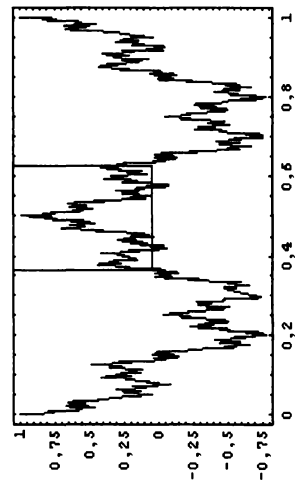
обладает сложной изломанной структурой и является самоподобной: форма функции остается неизменной при растяжении в b раз вдоль оси абсцисс и в $1/a$ раз вдоль оси ординат. На рис. 6.28, *a* показана функция Вейерштрасса при $a = 0,5$ и $b = 4$ и три последовательных увеличения ее центральной части (рис. 6.28, *б, в, г*). Можно видеть, что заключенные в прямоугольник части функции являются точными копиями предыдущего целого, т.е. функция Вейерштрасса является самоподобной. Впервые подобные функции появились в XIX веке, однако они были отвергнуты математиками как «некрасивые». Их не только не хотели изучать — о них не хотели даже говорить. В самом деле, если любая сколь угодно малая часть функции Вейерштрасса повторяет всю функцию в целом, то чему же равна длина ее графика? Очевидно, бесконечности. Вопрос о длине фрактальной линии остался в XIX веке без ответа.



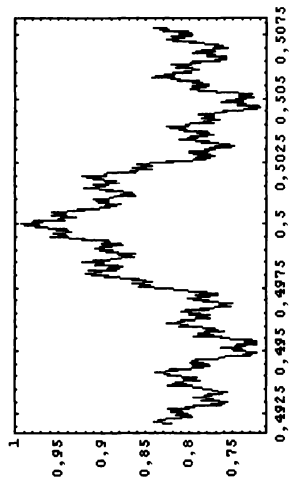
a)



б)



б)



з)

Рис. 6.28. Фрактальная функция Вейерштрасса [22]

Еще раз столкнулись с подобной проблемой во второй половине XX века при измерении длины береговой линии Англии. Эта задача была поручена известному английскому физику Л. Ричардсону. Для ее решения Ричардсон на карте заменил истинную береговую линию замкнутой ломанной, составленной из отрезков прямой длины δ , вершины которой располагались на побережье. Длина ломанной $L(\delta)$ принималась за приближенное значение длины береговой линии, соответствующее выбранному δ . Далее, переходя к пределу при $\delta \rightarrow 0$, ожидали получить истинное значение длины побережья. Из математики известно, что предел длины ломанной при $\delta \rightarrow 0$ для непрерывно дифференцируемой на отрезке $[a, b]$ функции $f(x)$ не зависит ни от длины элемента ломанной δ , ни от способа построения ломанной и равен длине L графика функции $f(x)$, вычисляемой по формуле

$$L = \int_a^b \sqrt{1 + f'(x)^2} dx.$$

Однако береговая линия Англии, в отличие от линий, описываемых гладкими функциями, оказалась настолько изрезанной вплоть до самых малых масштабов карты, что с уменьшением звеньев δ длина ломанной $L(\delta)$ не стремится к конечному пределу, а становится бесконечно большой. Ричардсону удалось установить характер стремления $L(\delta)$ к бесконечности, который выражался степенной функцией:

$$L(\delta) = C\delta^{1-D},$$

где $C = \text{const} > 0$ и $D = \text{const} > 1$.

Для побережья Англии $D \approx 1,24$, поэтому $L(\delta) = C/\delta^{0,24}$, и ясно, что $L(\delta) \rightarrow \infty$ при $\delta \rightarrow 0$. Для других береговых линий получались другие значения D ($1 < D < 2$), причем чем более изрезанной была линия, тем большие значения D соответствовали ей (например, для Норвегии $D \approx 1,5$).

Примером более упорядоченной фрактальной кривой может служить фрактал, открытый в 1904 г. немецким математиком Хельгой фон Кох. Алгоритм построения его очень прост: рассматривается равносторонний треугольник со сторонами единичной длины; каждый прямолинейный элемент делится на три части, на средней части строится меньший равносторонний треугольник и его основание отбрасывается. Предфракталы — фигуры, полученные за четыре первых шага, изображены на рис. 6.29.

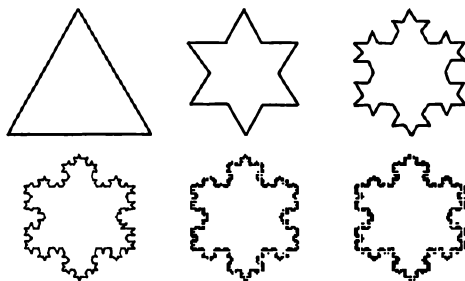


Рис. 6.29. Построение снежинки Кох

Можно вычислить периметр этой фигуры.

На нулевом шаге, т.е. $n = 0$, число элементов $N(\delta_0) = 3$, длина элемента $\delta_0 = 1$, длина кривой $L(\delta_0) = N(\delta_0)\delta_0 = 3$.

На шаге $n = 1$: $\delta_1 = 1/3 \cdot \delta_0 = 1/3$, $N(\delta_1) = 4N(\delta_0) = 4 \cdot 3$, длина кривой $L(\delta_1) = N(\delta_1)\delta_1 = 3 \cdot 4/3$.

На шаге $n = 2$: $\delta_2 = 1/3 \cdot \delta_1 = (1/3)^2$, $N(\delta_2) = 4N(\delta_1) = 3 \cdot 4^2$, длина кривой $L(\delta_2) = N(\delta_2)\delta_2 = 3(4/3)^2$.

На шаге n : всего $N(\delta_n) = 3 \cdot 4^n$ звеньев длиной $\delta_n = (1/3)^n$, тогда $n = -\ln \delta / \ln 3$, длина кривой

$$\begin{aligned}
 L(\delta) &= 3 \cdot (4/3)^n = 3 \exp \left\{ -\frac{\ln \delta \cdot [\ln 4 - \ln 3]}{\ln 3} \right\} = 3 \exp \left\{ -\frac{\ln \delta}{\ln 3} \ln \frac{4}{3} \right\} = \\
 &= 3 \exp^{-1} \left\{ \ln \delta (\ln 4 - \ln 3) / \ln 3 \right\} = \delta^{1-D},
 \end{aligned}$$

где $D = \ln 4 / \ln 3 \approx 1,2628 > 1$.

При $n \rightarrow \infty$, $\delta \rightarrow 0$, следовательно, длина кривой стремится к бесконечности. Множество точек, полученное как предел бесконечного числа итераций процедуры Кох, не являются кривой, для которой длина – удобная мера. Это уже не линия – «длина без ширины», а нечто большее, некая «толстая линия».

Понятие фрактальной размерности

Термин *фрактал* (от лат. *fractus* — изломанный, дробный) ввел в употребление в 1975 г. американский математик Б. Мандельброт, сотрудник исследовательского центра имени Томаса Дж. Уотсона корпорации ИВМ. Фракталами Мандельброт назвал структуры, обладающие двумя признаками: изломанностью и самоподобием (любая часть структуры подобна всему целому). Самоподобный понимается не только в классическом смысле как «линейно увеличенный или уменьшенный», но и в смысле «похожий». Кроме того, эти структуры характеризуются параметром, называемым фрактальной размерностью. В чем смысл этой размерности?

Прежде чем перейти к определению фрактальной размерности, сначала обобщим понятие меры множества. Рассмотрим уже знакомые понятия длины, площади и объема, которые являются мерами линии, поверхности и пространственного тела соответственно. Линия, поверхность, тело — некоторые множества точек в евклидовом пространстве.

Для того чтобы измерить «величину» множества \mathcal{Z} точек в пространстве, покроем множество \mathcal{Z} , например, пробными функциями с характерным размером δ (в качестве такой функции можно выбрать, например, сферу диаметра δ или куб с ребром δ). Центр сферы поместим в какую-нибудь точку множества \mathcal{Z} , тогда все точки, находящиеся от центра сферы на расстоянии $r < \delta/2$, окажутся покрытыми этой сферой. Посчитаем число сфер $N(\delta)$, необходимое для покрытия множества точек \mathcal{Z} , — получим в пределе при $\delta \rightarrow 0$ меру величины множества. Так, покрывая кривую отрезками длины δ , определим ее длину:

$$L(\delta) = N(\delta) \cdot \delta \xrightarrow{\delta \rightarrow 0} L_0.$$

В пределе мера L асимптотически равна длине кривой и не зависит от δ . Если же попытаться вычислить площадь кривой, при этом покрывать ее кругами диаметра δ (или квадратами со стороной δ), то найдем, с учетом предыдущей формулы, что она зависит от δ и в пределе при $\delta \rightarrow 0$ обращается в ноль:

$$A = N(\delta) \delta^2 \xrightarrow{\delta \rightarrow 0} L_0 \delta \xrightarrow{\delta \rightarrow 0} 0.$$

Точно также в ноль обратится и объем кривой:

$$V = N(\delta)\delta^3 \xrightarrow{\delta \rightarrow 0} L_0\delta^2 \xrightarrow{\delta \rightarrow 0} 0.$$

Еще раз подчеркнем, что для обычных кривых при $\delta \rightarrow 0$ площадь и объем стремятся к нулю, т.е. подходящей мерой для кривой является ее длина.

Чтобы вычислить площадь поверхности, покроем ее кругами диаметра δ (или квадратами со стороной δ), тогда

$$A = N(\delta)\delta^2 \xrightarrow{\delta \rightarrow 0} A_0.$$

В пределе мера A при $\delta \rightarrow 0$ равна площади поверхности и не зависит от δ . Если попытаться вычислить длину поверхности, покрывая ее отрезками длины δ , то с учетом предыдущего соотношения получим

$$L = N(\delta)\delta \xrightarrow{\delta \rightarrow 0} A_0\delta^{-1} \xrightarrow{\delta \rightarrow 0} \infty,$$

т.е. длина поверхности будет стремиться к бесконечности. Аналогично можно показать, что объем поверхности будет стремиться к нулю. Таким образом, только площадь является подходящей мерой для поверхности.

Объем тела можно получить в пределе при $\delta \rightarrow 0$, покрывая его сферами диаметра δ :

$$V = N(\delta)\delta^3 \xrightarrow{\delta \rightarrow 0} V_0.$$

Очевидно, что площадь и длина тела в пределе $\delta \rightarrow 0$ будут стремиться к бесконечности, т.е. подходящей мерой для тела является объем (а не длина и площадь).

Однако не для любого множества точек в пространстве можно указанным способом определить меру. Существуют, например, кривые Пеано, изогнутые настолько, что заполняют плоскость. Для их измерения необходимо обобщить понятие меры множества.

До сих пор в качестве пробной функции выбирались отрезок, круг, сфера (или квадрат, куб), т.е. функции вида $h(\delta) = \gamma(d)\delta^d$, и

покрывали множество \mathfrak{Z} , образуя меру $M_d = \sum_{n=1}^{N(\delta)} h(\delta)$. Заметим, что для отрезка, квадрата, куба — $\gamma(d) = 1$, для круга $\gamma(d) = \pi/4$, для сферы $\gamma(d) = \pi/6$.

В общем случае при $\delta \rightarrow 0$ мера M_d обращается в 0 или ∞ в зависимости от выбора величины d – размерности пробной функции:

$$M_d = \sum \gamma(d) \delta^d = \gamma(d) N(\delta) \delta^d \xrightarrow{\delta \rightarrow 0} \begin{cases} 0 & \text{при } d > D, \\ \infty & \text{при } d < D. \end{cases}$$

Величину M_d назовем d -мерой множества \mathfrak{S} .

Теперь можно ввести понятие размерности Хаусдорфа – Безиковича, которое играет ключевую роль в определении фрактала и фрактальной размерности.

Итак, *размерность Хаусдорфа–Безиковича D множества \mathfrak{S} – это такая критическая размерность, при которой мера M_d изменяет свое значение от 0 до ∞ .*

В этом определении важно отметить следующее:

1. При $d = D$ величина M_d может быть конечной, но может быть и 0 или ∞ . Важно, что именно при $d = D$ мера M_d меняется скачком.

2. Приведенное определение размерности – локальное свойство в том смысле, что эта размерность характеризует свойства множества в точке в пределе при исчезающе малом δ , следовательно, фрактальная размерность Хаусдорфа–Безиковича есть локальная характеристика множества.

3. Если множество покрывается шарами неодинакового размера, но диаметры их всех меньше некоторого δ , то d -мера – это нижняя граница, т.е. минимальное значение, получаемое при всех возможных покрытиях.

Вернемся к примеру со звездой Кох. Рассматривая ее периметр как одномерную линию, видим, что его длина стремиться к бесконечности:

$$L = 3\delta^{1-D} \xrightarrow{\delta \rightarrow 0} \infty \text{ при } \delta \rightarrow 0, D = \ln 4 / \ln 3.$$

Таким образом, конечную длину можно получить только в случае, если $d = D = \ln 4 / \ln 3 > 1$. Значит, размерность звезды Кох не целое число $d = 1$, а дробная величина $d = D = \ln 4 / \ln 3 > 1$. Итак, *фрактал имеет дробную размерность.*

С учетом введенной размерности можно привести еще одно определение фрактала, данное Мандельбротом: *фрактал – некоторое множество, размерность Хаусдорфа–Безиковича которого строго больше топологической размерности.*

Примеры классических фрактальных множеств

Триадное канторово множество. Алгоритм построения канторова множества таков: отрезок единичной длины (затравка) делится на три равные части, средняя часть отбрасывается, остаются два отрезка длиной $1/3$ каждый; далее каждый из оставшихся отрезков вновь делится на три части и средние части отбрасываются; оставшиеся четыре отрезка имеют длину $1/9$ каждый. И так далее. Два первых шага построения множества изображены на рис. 6.30. На n -м шаге множество состоит из $N = 2^n$ отрезков длиной $l_i = (1/3)^n$, $i = 1, 2, \dots, N$. Процедура построения повторяется бесконечное число раз. В результате имеем множество точек, называемых канторовой пылью.

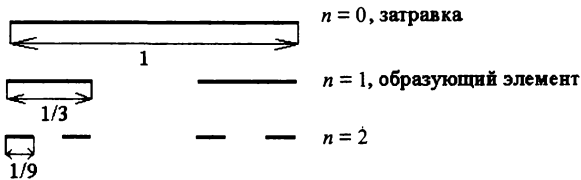


Рис. 6.30. Триадное канторово множество

Чтобы определить размерность канторова множества, покроем его отрезками длиной $\delta = l_i$, при этом окажутся покрытыми все точки множества. Тогда d -мера

$$M_d = \sum_{i=1}^N \delta^d = N(\delta)\delta^d = 2^n (1/3)^{nd} = \delta^{d-D}.$$

Она расходится и стремится к бесконечности при $\delta \rightarrow 0$, если только $d < D = \ln 2 / \ln 3$. Топологическая размерность множества $d = 0$, так как $d < D$, то канторово множество является фракталом.

Пусть отрезки в образующем элементе множества неравны, например $l_1 = 1/4$, $l_2 = 2/5$. Вычислим для этого случая фрактальную длину и размерность. На n -м шаге число отрезков $N = 2^n$, самый короткий отрезок имеет длину $l_1^n = (1/4)^n$, а самый длинный — $l_2^n = (2/5)^n$.

На n -м шаге

$$M_d = \sum_{i=1}^n l_i^d = \sum_{i=1}^n C_n^k l_1^d l_2^{(n-k)d} = (l_1^d + l_2^d)^n.$$

Поскольку $n \rightarrow \infty$ при $\delta \rightarrow 0$, то мера M_d конечна тогда и только тогда, когда $d = D$, где D удовлетворяет соотношению $(l_1^D + l_2^D)^n = 1$. Численно можно найти, что $D = 0,6110$ при $l_1 = 1/4$, $l_2 = 2/5$.

Ковер Серпинского. Алгоритм построения «ковра Серпинского» следующий: единичный квадрат делят на девять равных квадратов, длина стороны которых равна $1/3$, средний квадрат удаляют, а оставшиеся восемь опять делят на девять равных квадратов, средние части вновь удаляют. Построение фрактала на пяти первых шагах показаны на рис. 6.31.

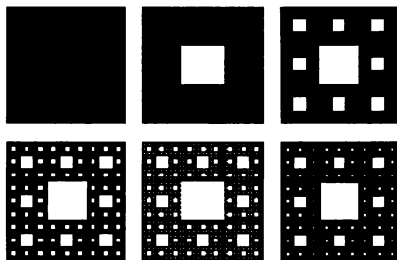


Рис. 6.31. Построение фрактала «ковер Серпинского»

Процедура повторяется бесконечное число раз. По аналогии с предыдущими примерами размерность полученного фрактала $D = \ln 8 / \ln 3$.

Салфетка Серпинского. Для того чтобы построить фрактал, называемый «салфеткой Серпинского» берется правильный треугольник со стороной единичной длины, затем соединяются середины его сторон, при этом исходный треугольник получается разделенным на четыре меньших правильных треугольника со сторонами, равными $1/2$. Далее отбрасывается средний треугольник, а оставшиеся три вновь делятся на четыре равных треугольника со сторонами по $1/4$. Алгоритм повторяется бесконечное число раз. На рис. 6.32 приведены первые пять шагов построения фрактала.

Несложно показать, что фрактальная размерность «салфетки Серпинского» $D = \ln 3 / \ln 2$.

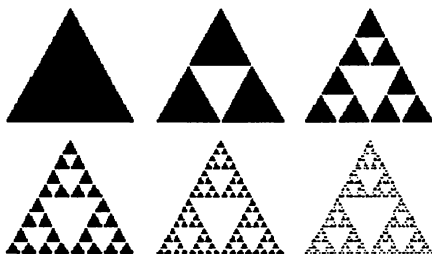


Рис. 6.32. Построение фрактала «салфетка Серпинского»

Подобие и скейлинг

Скейлинг – это преобразования параллельного переноса и изменения масштаба. Такая геометрическая фигура как прямая является инвариантной относительно преобразований параллельного переноса и изменения масштаба. Действительно, прямая, проходящая через точку $x_0 \in R^3$, – это множество \mathcal{S} точек $x \in R^3$. Ее можно задать как $x = x_0 + ta$, где $a \in R^3$, $-\infty < t < +\infty$.

Изменим масштаб в одинаковое для всех точек x_i r раз; новое множество $r(\mathcal{S})$ состоит из точек:

$$x' = r(x_0 + ta) = x_0 + t'a - (1-r)x_0, \text{ где } t' = rt \in (-\infty, \infty).$$

Если сдвинуть $r(\mathcal{S})$ параллельным переносом на $(1-r)x_0$, то получим исходное множество \mathcal{S} .

Аналогичные преобразования можно ввести для плоскости и пространства, т.е. плоскость и пространство являются множествами, инвариантными относительно скейлинга. Круг инвариантен относительно поворотов вокруг центра.

Рассмотрим ограниченное множество \mathcal{S} – отрезок прямой. Он не обладает трансляционной симметрией. Но изменив длину в $r < 1$ раз получим новое множество $\mathcal{S}' = r(\mathcal{S})$ – часть исходного. При определенном выборе масштаба r можно однократно покрыть исходное множество \mathcal{S} непересекающимися \mathcal{S}'_i множествами; в этом случае можно сказать, что \mathcal{S} – самоподобно с коэффициентом подобия r .

Для отрезка единичной длины таким масштабом является $r(N) = 1/N$, для квадрата $r(N) = (1/N)^{1/2}$, для параллелепипеда $r(N) = (1/N)^{1/3}$, где N – натуральное число.

Вообще для множества размерности D_s имеем $r(N) = (1/N)^{1/D_s}$, где D_s – размерность подобия: $D_s = -\ln N / \ln r(N)$.

Таким образом, для рассмотренных ранее фракталов можно вычислить размерность подобия через коэффициент r :

а) *кривая Кох*: коэффициент подобия $r = 1/3$, $N = 4$, размерность подобия $D_s = -\frac{\ln N}{\ln r(N)} = -\frac{\ln 4}{\ln(1/3)} \approx 1,26$;

б) *«салфетка Серпинского»*: $r = 1/2$, $N = 3$, $D_s = -\frac{\ln 3}{\ln 2} \approx 1,59$;

в) *«ковер Серпинского»*: $r = 1/3$, $N = 8$, $D_s = -\frac{\ln 8}{\ln 3} \approx 1,89$.

Следует заметить, что для самоподобных фракталов, каковы-ми являются рассмотренные ранее, фрактальная размерность и размерность подобия совпадают: $D_s = D$.

Множества Мандельброта и Жюлиа

До сих пор рассматривались так называемые линейные фракталы: алгоритм их построения задавался линейной зависимостью. Однако фрактальные множества могут давать алгоритмы, задаваемые нелинейными зависимостями (например, квадратичными). Одним из известнейших нелинейных фракталов является множество Жюлиа. Рассмотрим итерационное квадратичное преобразование на комплексной плоскости: $z_{n+1} = z_n^2 + c$, $c = \text{const}$.

Начальная точка z_0 и значение параметра c в итерационном процессе могут быть выбраны произвольно. Полученная последовательность точек может вести себя трояко: либо начальная точка будет постепенно уходить в бесконечность, либо будет стремиться к некоторой конечной точке комплексной плоскости, либо она не сможет принять определенного значения и будет блуждать по некоторой линии, т.е. существует три различных аттрактора (геометрического места точек притяжения решения).

Граница раздела между конечными аттракторами и бесконечным аттрактором называется *множеством Жюлиа*. Вид множества Жюлиа зависит от параметра c . При $c = 0$ итерационный процесс имеет два аттрактора: ноль и бесконечность, границей между ними является окружность единичного радиуса с центром в начале координат. На рис. 6.34 приведены множества Жюлиа для значений параметра $c = 0,32 + 0,043i$ и $c = -0,11 + 0,67i$. При некоторых зна-

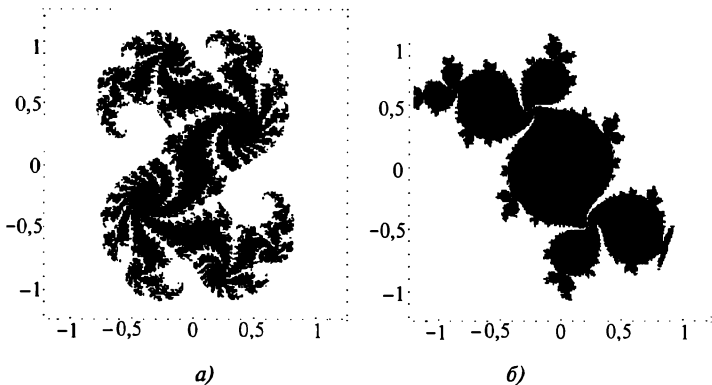


Рис. 6.33. Фрактальное множество Жюлиа для значений параметра:
 $c = 0,32 + 0,043i$ (а); $c = -0,11 + 0,67i$ (б)

чениях параметра с множество Жюлиа перестает быть связным и распадается на множество отдельных мелких частей – такое множество называют пылью Фату. Вид множества Жюлиа весьма разнообразен, однако в этом разнообразии существует определенный порядок: можно исследовать, при каких значениях параметра c множество будет связным, а при каких нет.

Поиски ответа на этот вопрос привели к открытию фрактального множества Мандельброта – множества значений параметра c , при которых последовательность, полученная в итерационном процессе $z_{n+1} = z_n^2 + c$, где $c = \text{const}$, и стартующая из точки $z_0 = 0$, остается ограниченной. Вид множества Мандельброта приведен на рис. 6.34.

Полученное множество симметрично относительно оси абсцисс; ее основу составляет кардиоида с вершинами на оси абсцисс и круг с вершиной на оси абсцисс. Границы множе-

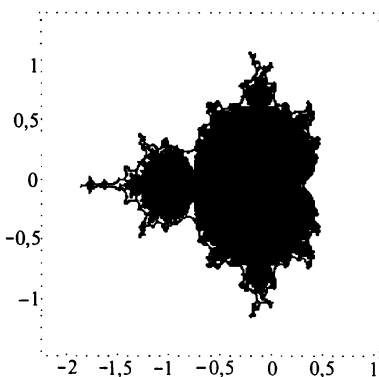


Рис. 6.34. Фрактальное множество Мандельброта

ства покрыты «наростами», которые называют почками, каждая из них подобна всему множеству в целом. На рис. 6.35 приведены увеличенные части границы множества.

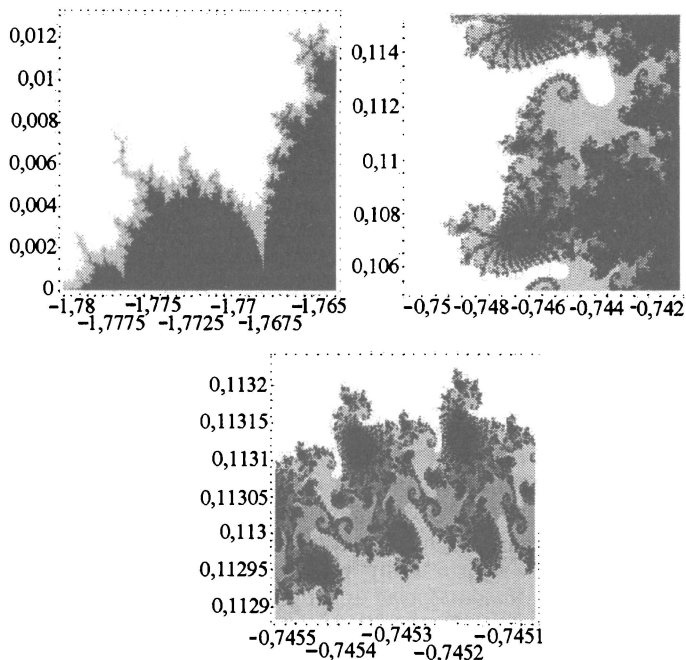


Рис. 6.35. Границы множества Мандельброта (цифры указывают координаты участка границы множества Мандельброта в единичном квадрате)

Фрактальная размерность кластеров

Кластеры — комплексные соединения, в основе молекулярной структуры которых лежит объемная ячейка из непосредственно связанных между собой атомов, играющая роль центрального атома. *Фрактальные кластеры* — это структуры, образующиеся, например, при:

- ассоциации твердых аэрозолей в газе при их диффузионном движении;

- электролизе;
- кристаллизации жидкости на подложке;
- вытеснении жидкостью с меньшей вязкостью жидкости с большей вязкостью (так называемые «вязкие пальцы»);
- течении в пористых средах.

Примером фрактального кластера может служить множество пустот в пористом материале.

Для «классических» фракталов при определении размерности Хаусдорфа–Безиковича нужно, чтобы $\delta \rightarrow 0$. Однако для физических систем существует минимальный характерный размер, например R_0 – радиус атома (молекулы). То есть линию нужно заменить цепочкой молекул (мономеров), поверхность (двумерное множество) – плоским набором мономеров, объем – набором сфер, тогда будем иметь число мономеров:

- в цепи длиной $L = 2R$: $N = R/R_0$, где $R/R_0 \gg 1$;
- в круглом диске площадью $S = \pi R^2$: $N = (R/R_0)^2$;
- в шаре объемом $V = 4\pi R^3/3$: $N = (R/R_0)^3$.

Плотность числа мономеров, которая зависит от упаковки, $\rho \leq \pi\sqrt{3}/2$.

Так как диск и сферу можно покрыть мономерами приближенно, то асимптотическая формула для соотношения числа частиц – размера «кластера», оцениваемого по радиусу R наименьшей сферы, содержащей кластер внутри себя, имеет вид

$$N = \rho (R/R_0)^D, \quad N \rightarrow \infty,$$

где D – размерность кластера; N – масса кластера; ρ – плотность массы. Размерность D не зависит от формы кластера, а ρ зависит от упаковки и формы кластера (для эллипсоидов вращения

$$\rho = (b/a)(\pi\sqrt{2}/3)).$$

Размерность D может быть дробной (фрактальной). Поясним на примере кривой Кох: рассмотрим предфрактал как конструкцию мономеров; мономер включает в себя один образующий элемент (рис. 6.36).

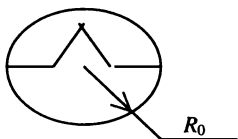


Рис. 6.36. Образующий элемент кривой Кох

При $n = 1$ – первое поколение, число мономеров $N = 4$, радиус кластера

$R = 3^2 R_0$; при $n = 2$ – число мономеров $N = 4^2$, радиус кластера $R = 3^2 R_0$; n -поколение – число мономеров $N = 4^n$, радиус кластера $R = 3^n R_0$, т.е. триадные кластеры Кох удовлетворяют соотношению

$$N = 4^{\ln(R/R_0)/\ln 3} = (R/R_0)^D,$$

где $D = \ln 3 / \ln 4$ – фрактальная размерность кластера, служащая количественной характеристикой того, как кластер заполняет занимаемое им пространство, но не описывающая его форму.

Примером кластера может служить множество точек, полученных в процессе ограниченной диффузией агрегации (ОДА), $D = 1,71$ (такой кластер характерен для структур, динамика которых описывается уравнением Лапласа): модель протекания в пористых средах, «вязкие пальцы» в ячейке Хеле-Шоу, вытеснение одной жидкости другой жидкостью с меньшей вязкостью.

Экспериментальные методы определения фрактальной размерности

Клеточная размерность. Для определения длины береговой линии ее покрывают, например, набором квадратов со стороной $\delta \rightarrow 0$ и подсчитывают их число $N(\delta)$. Далее в двойных логарифмических координатах строят зависимость $\ln N(\delta) - \ln(\delta)$. Так как $M_d \propto N(\delta)\delta^D$, то D определяется по углу наклона графика

$$D = (\text{const} - \ln N(\delta)) / \ln \delta.$$

Отношение длина/ площадь (определение D из соотношения между характеристиками множеств разной топологической размерности). Например, площадь фигуры, ограниченной фрактальной границей, $S \propto R^2$, а длина периметра $L \propto R^D$, где R – характерный размер. Тогда из соотношений $S^{1/2} \propto R \propto L^{1/D}$ следует, что

$$D = (\text{const} + \ln L^2) / \ln S,$$

т.е. фрактальная размерность границы D определяется как тангенс угла наклона зависимости квадрата периметра L^2 от площади S ,

построенной в двойных логарифмических координатах. Причем длина измеряется непосредственно (курвиметром) либо так же, как в задаче о береговой линии.

Более надежны методы, основанные на прямом экспериментальном определении величины $N(\delta)$. Так, если имеется большое число одинаковых фрактальных кластеров, изображение которых фиксировано на прозрачной фотопластинке, то, пропуская через нее световой пучок толщиной r и измеряя интенсивность прошедшего света I , можно использовать формулу

$$D = (\text{const} - \ln I) / \ln r.$$

Модель случайных фракталов для описания растущих дендритных структур

*На спинках блох блошата есть,
Кусают блох они там.
Блошонок у блошат не счесть,
И так indefinite...*

Даниель Дефо

Фотографии дендритов, полученные с помощью сканирующего микроскопа, позволяют различать ветви до восьмого поколения. Анализ структуры наталкивает на мысль использовать для описания дендритной структуры фракталы. Как показывают эксперименты, в процессе роста дендрита происходит оплавление некоторых его ветвей за счет флуктуаций температуры или концентрации элементов сплава, а также обламывание ветвей конвективными потоками жидкого металла, что не позволяет использовать для описания дендритов регулярные фракталы.

Вначале рассмотрим фрактальное описание уже сформировавшейся дендритной структуры (см. рис. 6.37). Будем считать, что частица представляется системой некоторых пространственных d -мерных фигур (элементарных фрагментов). Размеры этих фигур уменьшаются при переходе к следующему поколению. Нулевому поколению соответствует основной ствол частицы, имеющий характерный размер l_0 и объем $v_0 = G_f l_0^d$. Здесь G_f – геометрический фактор, который зависит от типа фигуры, аппроксимирующей дендритные ветви, и не зависит от масштаба длины.

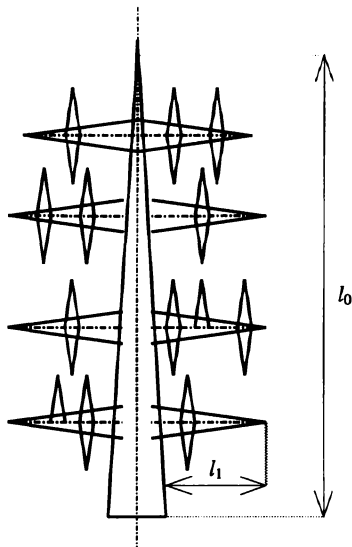


Рис. 6.37. Схема дендритной частицы со случайными геометрическими параметрами

Первое поколение состоит из p_1 ветвей, имеющих характерный размер $l_1 = l_0/k_1$, тогда суммарный объем частицы после первого этапа построения

$$v_1 = v_0 + G_f p_1 (l_0 / k_1)^d. \quad (6.90)$$

Второе поколение состоит из p_2 ветвей, каждая из которых имеет длину $l_2 = l_0/(k_1 k_2)$, а суммарный объем ветвей после второго этапа построения

$$v_2 = v_0 + v_1 + p_1 p_2 G_f \frac{l_0^d}{k_1^d k_2^d};$$

после n -го этапа суммарный объем

$$v_n = G_f l_0^d \left[1 + \frac{p_1}{k_1^d} + \frac{p_1 p_2}{k_1^d k_2^d} + \dots + \left(\frac{p_1}{k_1^d} \dots \frac{p_n}{k_n^d} \right) \right]. \quad (6.91)$$

Здесь $p_1, \dots, p_n, k_1, \dots, k_n$ — независимые случайные величины. Кроме того, считается, что величины p_i ($i = 1, \dots, n$) подчиняются

одному закону распределения, а величины k_i ($i = 1, \dots, n$) – другому. Если $p_i = p = \text{const}$ и $k_i = k = \text{const}$, т.е. имеет место регулярный фрактал, то

$$v_n = G_f l_0^d \frac{(p/k^d)^n - 1}{p/k^d - 1} = G_f l_0^d \frac{(l_0/\eta_n)^{D_f - d} - 1}{p/k^d - 1}, \quad (6.92)$$

где $D_f = \ln p / \ln k$, $\eta_n = l_0 / k^n$.

Найдем математическое ожидание величины v_n для случайного фрактала. Так как p_i и k_i – независимые случайные величины, распределенные по своим законам, то обозначим математическое ожидание соответствующих величин $M(p_i) = \bar{p}$, $M(k_i) = \bar{k}$, а d -й начальный момент $M(k_i^d) = \{k^d\}$ ($i = 1, \dots, n$). Тогда математическое ожидание величины объема, занимаемого дендритом, имеет вид

$$\bar{v}_n = G_f l_0^d \left[1 + \frac{\bar{p}}{\{k^d\}} + \left(\frac{\bar{p}}{\{k^d\}} \right)^2 + \dots + \left(\frac{\bar{p}}{\{k^d\}} \right)^n \right] = G_f l_0^d \frac{(\bar{p}/\{k^d\})^n - 1}{\bar{p}/\{k^d\} - 1}. \quad (6.93)$$

При выводе соотношений (6.92) и (6.93) использовалась формула суммы n членов геометрической прогрессии со знаменателем $\bar{p}/\{k^d\}$. Отметим, что $\bar{p}/\{k^d\} < 1$; это условие гарантирует отсутствие наложений фрагментов фрактала, обеспечиваемое видом соотношения (6.91).

Предположим, что выражение для среднего объема имеет тот же вид, что и для регулярного фрактала (6.92):

$$\bar{v}_n = G_f l_0^d \frac{(l_0/\bar{\eta}_n)^{\bar{D}_f - d} - 1}{\bar{p}/\{k^d\} - 1}, \quad (6.94)$$

где G_f не зависит от масштаба длины и $\bar{\eta}_n = l_0 / (\bar{k})^n$, т.е. предполагается статистическое самоподобие структуры. Определим величину фрактальной размерности \bar{D}_f для нерегулярного фрактала. Сравнивая (6.93) и (6.94), заметим, что

$$\left(\frac{\bar{p}}{\{k^d\}} \right)^n = \left(\frac{l_0}{\bar{n}} \right)^{\bar{D}_f - d}. \quad (6.95)$$

Прологарифмировав соотношение (6.95) и воспользовавшись правилами преобразования логарифмических выражений, получим:

$$n \ln \left(\frac{\bar{p}}{\{k^d\}} \right) = (\bar{D}_f - d) \ln \left(\frac{l_0}{\bar{n}} \right) = (\bar{D}_f - d) \ln (\bar{k})^n = n(\bar{D}_f - d) \ln \bar{k},$$

$$\ln \bar{p} - \ln \{k^d\} = (\bar{D}_f - d) \ln \bar{k},$$

$$\bar{D}_f = \frac{\ln \bar{p}}{\ln \bar{k}} - \frac{\ln \{k^d\}}{\ln \bar{k}} + d = \frac{\ln \bar{p}}{\ln \bar{k}} - \log_{\bar{k}} \left(\frac{\{k^d\}}{(\bar{k})^d} \right).$$

Таким образом, фрактальная размерность случайного фрактала вычисляется по формуле

$$\bar{D}_f = \frac{\ln \bar{p}}{\ln \bar{k}} - \log_{\bar{k}} \left(\frac{\{k^d\}}{(\bar{k})^d} \right) \quad \text{или} \quad \bar{D}_f = \frac{\ln \bar{p}}{\ln \bar{k}} - \frac{\ln \{k^d\}}{\ln \bar{k}} + d. \quad (6.96)$$

Рассмотрим частные случаи для различной топологической размерности. Для фрактальной структуры на базе одномерного объекта $d=1$ (например, для периметра плоской проекции дендритной частицы) фрактальная размерность имеет вид

$$\bar{D}_f = \frac{\ln \bar{p}}{\ln \bar{k}} - \log_{\bar{k}} \left(\frac{\{k^1\}}{(\bar{k})^1} \right) = \frac{\ln \bar{p}}{\ln \bar{k}} - \log_{\bar{k}} 1 = \frac{\ln \bar{p}}{\ln \bar{k}}. \quad (6.97)$$

Для плоской фрактальной структуры $d=2$ (например, для площади плоской проекции дендритной частицы) фрактальная размерность определяется соотношениями:

$$\begin{aligned} \bar{D}_f &= \frac{\ln \bar{p}}{\ln \bar{k}} - \log_{\bar{k}} \left(\frac{\{k^2\}}{(\bar{k})^2} \right) = \frac{\ln \bar{p}}{\ln \bar{k}} - \log_{\bar{k}} \left(\frac{(\bar{k})^2 + D(k)}{(\bar{k})^2} \right) = \\ &= \frac{\ln \bar{p}}{\ln \bar{k}} - \log_{\bar{k}} \left(1 + \frac{D(k)}{(\bar{k})^2} \right) \end{aligned} \quad (6.98)$$

или

$$\bar{D}_f = \frac{\ln \bar{p}}{\ln \bar{k}} - \log_{\bar{k}} \{k^2\} + 2, \quad (6.99)$$

где $D(k)$ – дисперсия случайной величины k . Если k – детерминированная величина, то ее дисперсия равна нулю, и формула (6.98) примет вид выражения для фрактальной размерности самоподобного регулярного фрактала: $D_f = \ln p / \ln k$.

Для объемной фрактальной структуры $d = 3$ (например, для объема дендритной частицы)

$$\bar{D}_f = \frac{\ln \bar{p}}{\ln \bar{k}} - \log_{\bar{k}} \left(\frac{\{k^3\}}{(\bar{k})^3} \right) = \frac{\ln \bar{p}}{\ln \bar{k}} - \log_{\bar{k}} \left(\frac{(\bar{k})^3 + 3\bar{k}D(k) + \{(k - M(k))^3\}}{(\bar{k})^3} \right) \quad (6.100)$$

или

$$\bar{D}_f = \frac{\ln \bar{p}}{\ln \bar{k}} - \log_{\bar{k}} \{k^3\} + 3, \quad (6.101)$$

где $\{(k - M(k))^3\} = M[(k - M(k))^3]$ – третий центральный момент случайной величины k [103].

В задаче кристаллизации необходимо определять суммарный объем растущего дендрита, поэтому параметры дендритной и, соответственно, фрактальной структуры должны зависеть от времени. Пусть растущий дендрит имеет форму, ограниченную параболой и прямой $z = 0$ в двумерном случае и параболоидом вращения и плоскостью $z = 0$ в пространственном; его ось направлена вдоль оси Oz . Вершина дендрита перемещается равномерно вдоль оси Oz со скоростью v . Тогда уравнения, описывающие форму, имеют вид:

$$\begin{aligned} d = 2, \quad z = vt + \alpha/2 - x^2/(2\alpha), \\ d = 3, \quad z = vt + \alpha/2 - r^2/(2\alpha), \quad r^2 = x^2 + y^2, \end{aligned} \quad (6.102)$$

где α – параметр параболы (параболоида) – расстояние от фокуса до директрисы. В качестве характерного размера дендритной частицы выберем расстояние от вершины до начала координат, равное $l_0 = (2vt + \alpha)/2$. Выбор данной формы растущего дендрита обусловлен следующими причинами: во-первых, параболическая форма была получена аналитически в [47] в задаче исследования устойчивости границы раздела фаз; во-вторых, такая зависимость позволяет одновременно описать не только удлинение дендритных ветвей, но и их утолщение. Вычислим площадь (объем), ограниченную параболой (параболоидом) и осью (плоскостью) $z = 0$:

$$d = 2, S(\alpha, t) = 2 \int_0^{\sqrt{\alpha(2vt+\alpha)}} [vt + \alpha/2 - x^2/(2\alpha)] dx = \\ = \frac{2}{3} (2vt + \alpha) \sqrt{\alpha(2vt + \alpha)},$$

$$d = 3, V(\alpha, t) = \int_0^{2\pi} d\varphi \int_0^{\sqrt{\alpha(2vt+\alpha)}} [vt + \alpha/2 - r^2/(2\alpha)] r dr = \\ = \frac{2\pi}{8} \alpha (2vt + \alpha)^2, \quad (6.103)$$

и выделим из этих соотношений геометрический фактор $G_f(\alpha, f)$, соответствующий такой форме дендрита:

$$d = 2, S(\alpha, t) = \left(\frac{2vt+\alpha}{2}\right)^2 \left(\frac{8\sqrt{\alpha}}{3\sqrt{2vt+\alpha}}\right) = I_0^2(t) G_f(\alpha, t), \\ d = 3, V(\alpha, t) = \left(\frac{2vt+\alpha}{2}\right)^3 \left(\frac{2\pi\alpha}{2vt+\alpha}\right) = I_0^3(t) G_f(\alpha, t). \quad (6.104)$$

Заметим, что $G_f(\alpha, f)$ зависит от времени. Это объясняется тем, что рост дендритных ветвей происходит не из-за их равномерного растяжения (преобразования подобия), а за счет перемещения параболы как жесткого целого со скоростью v ; при этом свойство самоподобности сохраняется.

Результаты применения модели случайных фракталов

Проанализируем результаты применения предложенной модели для двумерной задачи. Форма дендритных ветвей задается параболой. Вычислим площадь, занимаемую дендритной частицей в единичном квадрате. Пока не будем рассматривать рост дендрита. Полученные результаты сравним с экспериментальными данными, определенными посредством обработки снимка дендритной частицы (рис. 6.38) с помощью пакета «Image Tool». По снимку определялись площадь, занимаемая дендритной частицей в квадра-

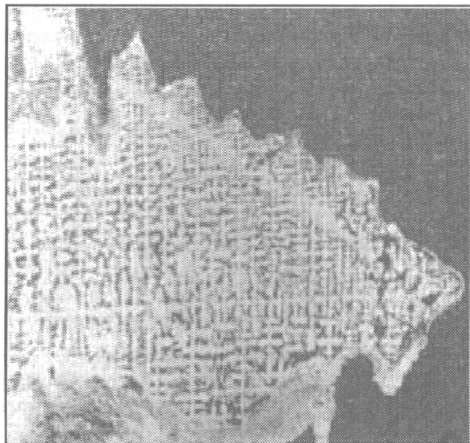


Рис. 6.38. Структура дендритного кристалла в разрезе. Дендрит извлечен из усадочной раковины крупной отливки [110]

те, длины ветвей нулевого, первого и второго порядков, число ветвей следующего порядка на каждой ветви. Были определены интервалы значений названных величин (табл. 6.1). Предполагалось, что величины p и k распределены равномерно. Заметим, что для приведенных значений выполняется соотношение $\bar{p}/\{k^d\} < 1$.

Вычисления проводились для нескольких реализаций случайных величин. По заданным распределениям для величин p и k с помощью генератора случайных чисел составлялись выборки, содержащие по 100 элементов. Затем по выборке определялись среднее значение $M(\cdot)$ и среднее квадратичное отклонение $\sigma(\cdot)$, которые и подставлялись в соотношения (6.94), (6.99) для нахождения занимаемой площади и фрактальной размерности.

Приведем результаты вычислений для пяти реализаций случайных величин. В табл. 6.2 даны параметры модели, средние зна-

Таблица 6.1

Интервалы значений параметров дендритной структуры

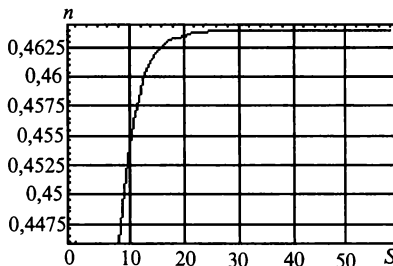
p	k	l_0	S
19–40	2–10	1	0,47±0,03
$R(19; 40)$	$R(2; 10)$	–	–

**Параметры структуры и результаты вычислений
для сформировавшегося дендрита**

№ п/п	$M(p)$	$\sigma(p)$	$M(k)$	$\sigma(k)$	α	νt	S	\bar{D}_f
1	29,716	6,433	5,601	2,278	0,005	1	0,464	1,825
2	29,680	6,417	5,458	2,414			0,424	1,781
3	29,607	6,192	5,769	2,475			0,384	1,771
4	29,353	6,450	5,694	2,130			0,464	1,812
5	30,014	6,047	5,367	2,504			0,402	1,780
Среднее							0,4276	1,794

чения, стандартные отклонения величин p и k ; величины площади и фрактальной размерности, вычисленные согласно предложенной модели. Как видно из таблицы, площадь, занимаемая дендритом, и его фрактальная размерность меняются при различных реализациях случайных величин, однако разброс остается в пределах реальных значений и достаточно близок к замеренным по фотографии.

На рис. 6.39 показано, как изменяется величина S_f в зависимости от числа поколений фрактала. Для каждого n определялись средние значения и стандартные отклонения величин по соответствующим выборкам, по которым и определялись площадь и размерность. Из рисунка можно заметить, что после достижения $n = 25$ дальнейшее увеличение числа поколений фрактала не приводит к заметному изменению площади (эти установившиеся значения и приведены в табл. 6.2). Очевидно, именно эту величину



**Рис. 6.39. Зависимость площади
от числа уровней самоподобия**

ну и нужно принимать за площадь, занимаемую фрактальной частью.

Заметим, что относительную площадь, занимаемую дендритной частью, можно рассматривать как долю твердой фазы для плоской задачи кристаллизации, и предложенная модель дает удовлетворительную оценку этой величины.

Проведем расчет изменения площади, занимаемой дендритом, и его фрактальной размерности со временем. Для этого длина оси нулевого порядка и параметры равномерного распределения $R(a; b)$ числа дендритных ветвей задавались в виде линейных функций времени, а расстояние между вторичными ветвями – некоторой постоянной величиной d_{II} . В табл. 6.3 даны вид этих функций, значения параметров, используемых в расчетах, а также значения площади и размерности в момент $t = 1$. Затем на каждом шаге по времени генерировалась выборка случайных чисел для количества дендритных ветвей с текущими параметрами распределения, определялись ее среднее значение и стандартное отклонение и по соотношениям (6.94), (6.99) вычислялись площадь и фрактальная размерность.

Таблица 6.3

Параметры структуры и результаты вычислений для растущего дендрита

l_0	d_{II}	a	b	ν	t_0	S	\bar{D}_f
$0,1+\nu t$	0,025	$0,3 (l_0/d_{II})$	l_0/d_{II}	1	0,01	0,431	1,752

На рис. 6.40 и 6.41 приведены зависимости соответственно площади и фрактальной размерности растущего дендрита от времени.

Можно показать, что форма элементарных фрагментов фрактала, то есть те фигуры, которые аппроксимируют форму дендритных ветвей, несущественно сказываются на результатах вычислений. Для примера сравним значения, полученные для формы ветвей, ограниченной параболой, равнобедренным треугольником и половиной эллипса. В табл. 6.4 приведены уравнения кривых, площади и геометрические факторы фигур, ограниченных соответствующими кривыми и прямой $z = 0$.

Параметры выбирались таким образом, чтобы совпадали точки пересечения этих кривых и координатных осей: меньшая полуось

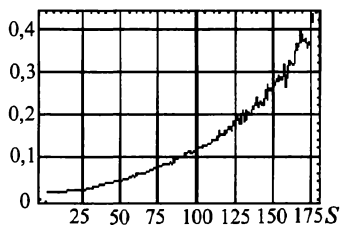


Рис. 6.40. Зависимость площади растущего дендрита от времени

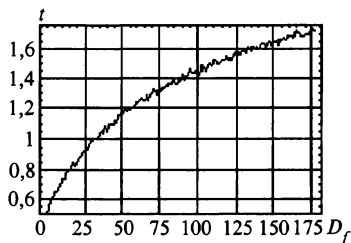


Рис. 6.41. Зависимость фрактальной размерности растущего дендрита от времени

Таблица 6.4

Параметры кривых, аппроксимирующих дендритные ветви

Тип фигуры	Уравнение кривой	Площадь	Геометрический фактор $G_f(\alpha, t)$
Парабола	$z = vt + \alpha/2 - x^2/(2\alpha)$	$\frac{2}{3}(2vt + \alpha)\sqrt{\alpha(2vt + \alpha)}$	$\frac{8\sqrt{\alpha}}{3\sqrt{2vt + \alpha}}$
Треугольник	$z = \begin{cases} \frac{2vt + \alpha}{2} \left(1 - \frac{x}{\sqrt{\alpha(2vt + \alpha)}}\right), & x \geq 0 \\ \frac{2vt + \alpha}{2} \left(1 + \frac{x}{\sqrt{\alpha(2vt + \alpha)}}\right), & x < 0 \end{cases}$	$(2vt + \alpha)\sqrt{\alpha(2vt + \alpha)}$	$\frac{2\sqrt{\alpha}}{\sqrt{2vt + \alpha}}$
Эллипс	$z = \frac{2vt + \alpha}{2} \sqrt{1 - \frac{x^2}{\alpha(2vt + \alpha)}}$	$\frac{\pi}{4}(2vt + \alpha)\sqrt{\alpha(2vt + \alpha)}$	$\frac{\pi\sqrt{\alpha}}{\sqrt{2vt + \alpha}}$

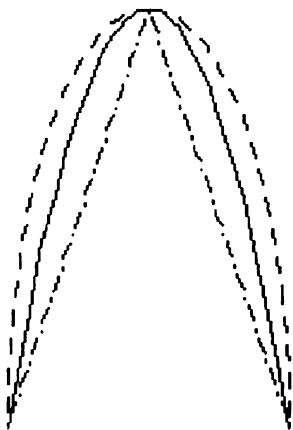


Рис. 6.42. Кривые, аппроксимирующие форму дендритной ветви

эллипса равна $\sqrt{\alpha(2vt + \alpha)}$, большая полуось – l_0 ; основание равнобедренного треугольника равно 2, высота равна 10 (как показано на рис. 6.42). Сплошная линия соответствует параболе, штриховая – верхней половине эллипса, штрихпунктирная – равнобедренному треугольнику. В этом случае характерная длина по-прежнему имеет вид $l_0 = (2vt + \alpha)/2$.

По сути, независимость от типа фигуры элементарного фрагмента фрактала заложена изначально в соотношениях (6.94) и (6.96). Площади дендритных частей будут отличаться лишь на константу, а размерность вообще не изменится. Сказанное подтверждают результаты расчетов, приведенные в табл. 6.5.

Таблица 6.5
Сравнение результатов для разных типов фигур

Тип фигуры	S	\bar{D}_f
Парабола	0,465	1,812
Треугольник	0,349	
Эллипс	0,548	

Как и следовало ожидать, наибольшую площадь занимает дендритная частица, ветви которой аппроксимируются половиной эллипса, наименьшую – частица с ветвями, аппроксимируемыми равнобедренными треугольниками.

6.18. НЕЛИНЕЙНЫЕ МОДЕЛИ ДНК

Малые группы атомов, слишком малые, чтобы проявлять точные статистические законы, играют главенствующую роль в весьма упорядоченных и закономерных процессах внутри организма. Они управляют видимыми признаками большого масштаба, которые организм приобретает в течение своего развития; они определяют важные особенности его функционирования, и во всем этом проявляются весьма отчетливые и строгие биологические законы.

Э.Шредингер

Структура и физические свойства ДНК

Поскольку реальные механизмы многих биологических процессов, связанных с ДНК, до настоящего времени не поняты, биологи обращают все большее внимание на структурные аспекты с целью объяснения основных механизмов функционирования ДНК – транскрипции, репликации и денатурации. Принимая во внимание тот факт, что симметрия ДНК при ее формировании может быть связана с дефектами, эффекты дисторсии в структуре двойной спирали ДНК могут также рассматриваться как зарождение и эволюция новых структурных дефектов. Двойная спираль ДНК представляет собой уникальный полимер, в котором молекулярные цепи образуют так называемую В-конформацию, впервые установленную по рентгеновским дифракционным картинам. В соответствие с результатами Уотсона и Крика В-структура ДНК имеет следующие основные черты [117]:

1. В состав входят две полинуклеотидные цепи (рис. 6.43), обвивающие с правосторонним поворотом общую ось, образуя двойную спираль с диаметром $\sim \lambda$. Двойные цепи являются антипарал-



Рис. 6.43. Структура двойной спирали

лельными и обернуты таким образом друг относительно друга, что не могут быть разделены без взаимного расплетения (эффект плектонемической закрутки). Основания покрывают поверхность спирали, которая обвита сахаро-фосфатными цепями на периферии, причем так, что достигается минимизация сил отталкивания между заряженными фосфатными группами.

2. Плоскости оснований почти перпендикулярны оси спирали; каждая пара расположенных друг против друга оснований имеет водородную связь и образует при этом плоскую пару оснований (base pair). Взаимодействие оснований посредством водородных связей (complementary base pairing phenomenon) и является причиной специфического объединения двух цепей в двойную спираль. Идеальная В-ДНК спираль содержит 10 пар оснований (base pairs, pb) на один поворот (поворот спирали на 1 pb составляет 36°), что обеспечивает шаг спирали (rise per turn) и пространственную периодичность угла закрутки соответственно $h = 34 \text{ \AA}$ и $\omega_0 = 2\pi/h = 1.85 \text{ nm}^{-1}$. Ароматические основания имеют ван-дер-ваальсовский масштаб $3,4 \text{ \AA}$, и, как следствие этого, наблюдаются эффекты упаковки (base stacking).

3. Персистентные длины поворотных и изгибных мод представляют естественные масштабы для термических флуктуаций (приблизительно $15-20 h$). ДНК является достаточно жестким полимером вследствие растяжения соответствующих пар аминокислот химическими и водородными связями. Термические персистентные длины ДНК на изгиб и вращение составляют соответственно 150 bp (50 nm) и 220 bp (75 nm).

Замечательной чертой структуры ДНК, по Уотсону-Крику, является аккомодация двух типов пар оснований: остаток аденина объединяется с остатком тимина и каждый остаток гуанина — с остатком цитозина. Разнообразие протеинов обеспечивает специфические микроскопические изменения в двойной спирали.

4. При пространственных масштабах в тысячи пар оснований ($\sim 0,5 \text{ мкм}$) ДНК *in vivo* организованы в топологически независимые петли. Молекулы ДНК с размерами $2-10 \text{ kbp}$ ($0,6-3,3 \text{ мкм}$) представляют собой «связанные» множества. Основная часть генов (prokariotic cells) организована как единичные замкнутые ДНК с размерами $\sim 2 \text{ Mbp}$.

Статистические модели ДНК. Структурные переходы в ДНК

Имеет место формальная аналогия между механизмами неустойчивости при фазовых переходах (*melting instability*) в однородной структуре ДНК и фазовыми переходами в однородных системах. Эффекты плавления (денатурации) в растягиваемой двойной цепи ДНК (рис. 6.44) – одна из фундаментальных проблем в статистической физике биологических систем – обнаруживают различные типы критического поведения, наблюдаемые в других термодинамических системах [119, 129].

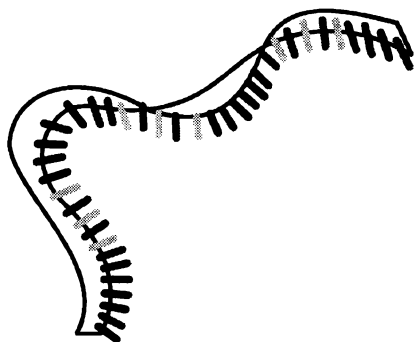


Рис. 6.44. Распад ДНК

Ранние модели термического распада ДНК, т.е. разделения двух цепей при нагреве [121], были основаны на моделях типа Изинга, в которых параметр порядка, характеризующий взаимодействие оснований, сводился к расстоянию между последними. Современные исследования связывают эти переходы с нелинейными механизмами зарождения конечно-амплитудных флуктуаций [120]. Анализ закономерностей скейлинга указанных переходов был проведен в [128] с использованием асимметричного потенциала, содержащего пологий участок для больших значений перемещений, что соответствует распаду двойной спирали при повышенных температурах. Этот тип потенциала включал также взаимодействие между последовательными основаниями в длинноволновом приближении. В соответствии с этим описанием локальная неустой-

чивость плавления возникает в «мягком» режиме, что не подтверждается экспериментальными наблюдениями динамики переходов, которые носят резкий характер.

Обобщение данной модели, предложенное в [127], связано с введением параметра жесткости упаковки (stacking parameters), который, учитывая взаимодействие последовательных пар оснований, позволил описать резкий характер перехода.

Исследования однородных ДНК также показали, что указанная модель дает возможность описать последовательный распад (денатурацию), наблюдаемый в эксперименте. В дополнение установлено, что в рамках модели возможна генерация «энтропийного барьера», с которым связывается «сужение» перехода. При некоторых значениях параметра жесткости модель демонстрирует признаки переходов первого рода с конечным значением энтропии плавления, нарушением непрерывности для части связей (что типично для ДНК), расходимость продольных и поперечных корреляционных масштабов.

Степени свободы молекулы ДНК

Степени свободы, описывающие дисторсии в молекуле ДНК, могут быть разделены на переменные, характеризующие конформации центральной оси, и переменные, определяющие внутренний поворот двойной спирали [124]. Положение оси ДНК может быть описано пространственной кривой, заданной параметром длины вдоль оси ДНК и радиусом-вектором точки $r(s)$. Касательная $t \equiv \partial_s r$ задается единичным вектором t , положение которого в пространстве определяется двумя углами или локальной кривизной и дисторсией.

Поворот измеряется как пространственная интенсивность поворота пар оснований относительно центральной оси, которая для неискаженной структуры ДНК задается как ω_0 . Отклонение в интенсивности поворота от ω_0 может быть описано с использованием скалярного поля $\Omega(s)$. Это дает в целом три степени свободы каждой координате, что необходимо для описания бесконечно-малых поворотов координатных осей, связанных с парами оснований при перемещении последних вдоль молекулы.

Модель Пейрара-Бишопа

Проблема распада (denaturation) ДНК вызывает большой интерес в контексте процессов, лежащих в основе механизмов транскрипции – важнейшего свойства молекулярных биологических систем. Эксперименты на разбавленных растворах ДНК убедительно демонстрируют существование термически инициируемых процессов денатурации, сопровождающихся разрывом связей пар оснований и образованием локализованных мод дисторсии (opening of base pairs) при некоторой критической температуре T_m [117,129]. Показано, что процесс денатурации может быть описан в рамках переходов первого рода при контролируемых внешних параметрах – температуре и внешнем закручивающем моменте. Качественно эти переходы аналогичны переходам «жидкость–газ», которые контролируются температурой и давлением.

Для описания поведения однородной ДНК предложен потенциал [120]

$$H = \sum_n \left[\frac{p_n^2}{2m} + W(y_n, y_{n-1}) + V(y_n) + Dhay_n \right],$$

где m – приведенная масса пар оснований; y_n определяет трансляционную моду, обусловленную растяжением водородных связей, соединяющих два основания n -й пары, $p_n = m(dy_n/dt)$. Связь между последовательными парами оснований описывается соотношением

$$W(y_n, y_{n-1}) = \frac{K}{2} \left[1 + \rho \exp(-\delta(y_n + y_{n-1})) \right] (y_n - y_{n-1})^2,$$

где K, ρ, δ – постоянные.

Этот выбор потенциала взаимодействия обусловлен выводом, что энергия упаковки в основном определяется свойством пар оснований, а не свойствами одиночных оснований. Эффективная константа связи изменяется от значения $(1+\rho)K$ до K , когда одна из двух взаимодействующих пар оснований раскрывается. Это предположение находится в качественном соответствии с реальными признаками взаимодействия оснований в ДНК.

Взаимодействие двух оснований в паре описывается одноточечным потенциалом

$$V(y_n) = D(e^{-\alpha y_n} - 1)^2,$$

где D , α — постоянные.

Последний член в выражении для потенциала описывает влияние продольных внешних усилий h .

Относительная простота постановки позволяет осуществлять моделирование в широком временном интервале, что необходимо для оценки экспериментальных условий.

За единицы измерений времени и силы в [120] приняты соответственно $1,021 \cdot 10^{-14}$ с и 1602 пн, в соответствии с которыми модель имеет следующие параметры:

$$D = 0,04 \text{ эВ}; \quad K = 0,04 \text{ эВ/Å}; \quad a = 4,5 \text{ Å}; \quad \alpha = \text{Å}^{-1};$$

$$m = 300 \text{ а.м.е.}; \quad r = 0,5; \quad \delta = 0,35$$

и дает значение температуры денатурации 360 К.

Вращательная мода в ДНК

В последние десятилетия с целью более глубокого понимания механизмов денатурации интенсивно исследуется [118, 124, 129] реакция молекулы ДНК на внешнее закручивающее напряжение. С биологической точки зрения появление закручивающих напряжений в живой клетке не является необычным и может существенно влиять на функционирование ДНК. Было показано, что денатурация описывается в рамках фазового перехода первого рода, когда контролирующими параметрами являются температура и приложенная величина внешнего вращения. Эта теория дает объяснение В-DNA фазового перехода в экспериментах с единичными молекулами.

Эта модель воспроизводит спираль Уотсона и Крика (В-DNA) и схематически представлена на рис. 6.45. В плоскости, перпендикулярной оси спирали, для каждой пары оснований ($n = 1, \dots, N$) вводится полярная система координат r_n и ϕ_n . Сахаро-фосфатное основание моделируется жесткими стержнями с фиксированной длиной L , в то время как расстояние h_n между плоскостями оснований ($n - 1$) и n может флуктуировать.

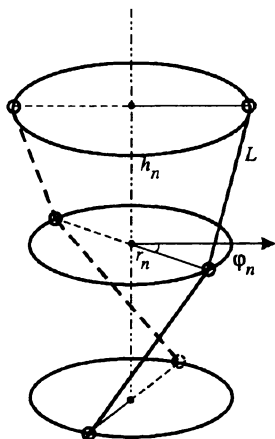


Рис. 6.45. Модельное представление структуры ДНК

Потенциальная энергия V в переменных r_n, φ_n записывается в форме

$$V = \sum_n [V_m(r_n) + V_s(r_n, r_{n-1}) + V_b(r_n, r_{n-1}, \theta_n)],$$

где $\theta_n = \varphi_n - \varphi_{n-1}$ — угол закручивания спирали.

Первый член $V_m(r_n) = D(e^{-a(r-R)} - 1)^2$ представляет энергию водородной связи для заданной пары оснований. В данной модели пара оснований с диаметром $r > r_d = R + 6/a$ рассматривается как открытая. Глубина потенциала D (обычно порядка 0,1 эВ), зависит от типа пары оснований (АТ — аденин-тимин или ГС — гуанин-цитозин), а также от прочности ионной связи.

Второе слагаемое описывает сдвиговые усилия, которые оказывают сопротивление скольжению одного основания относительно другого, характеризуются потенциалом упаковки (stacking potential)

$$V_s(r_n, r_{n-1}) = Ee^{-b(r_n + r_{n-1} - 2R)}(r_n - r_{n-1})^2.$$

Вследствие уменьшения молекулярной упаковки с появлением открытых пар, сдвиговый фактор экспоненциально уменьшается и становится пренебрежимо малым на расстояниях, превышающих $\approx 5b^{-1} = 10 \text{ \AA}$, которые близки к диаметру пар оснований.

Третье слагаемое описывает флуктуации аксиального расстояния h_n между соседними парами оснований и описывается упругим потенциалом

$$V_b(r_n, r_{n-1}, \theta_n) = K[h_n - H]^2 = \\ = K \left[\sqrt{L^2 - r_n^2 - r_{n-1}^2 + 2r_n r_{n-1} \cos \theta_n} - H \right]^2.$$

Геликоидальные структуры возникают при $H < L$: в B -конформации $r < r_d$ минимум потенциала $V_b = 0$ достигается при ненулевом угле поворота (см. рис. 6.45).

Нелинейные локализованные моды дисторсии в структуре ДНК

Модель Пейрара-Бишопа позволила провести разнообразные исследования локализованных мод в структуре ДНК. Приближенное решение в рамках данной структуры гамильтониана для одномерной ДНК решетки, показало существование периодических и пространственно-локализованных решений, получивших название РВ-бризеры Пейрара-Бишопа [127]. Эти коллективные моды были исследованы применительно к механизмам термической денатурации ДНК. На кривой «плавления» они проявляются в виде локализованных мод тонкой структуры, с динамикой которых связываются основные механизмы транскрипции. Эти локализованные моды, являясь устойчивыми в малоамплитудном пределе, могут распространяться вдоль молекулярной цепи в течение конечного интервала времени с незначительным падением энергии и без изменения формы. Имея очень малую амплитуду, движущиеся РВ-бризеры не могут рассматриваться как распространяющиеся активирующие дисторсии, но они играют роль предвестников, формирующих так называемые считывающие домены (transcription bubble) в термализованной цепи как спонтанную локализацию энергии.

Бризеры обладают способностью к росту, аккумулируя энергию малых возмущений до возникновения больших доменов денатурации и в завершение — полного разделения цепей. Природа этих флуктуаций связана с широкой вариацией скоростей, с которыми растет число денатурированных пар оснований. Такие большие флуктуации, соответствующие «дыханию ДНК», наблюдаются

экспериментально. Важным следствием этих экспериментов является установление факта, что изолированный бризер останавливается, если его фронт достигает «холодной» области. Таким образом, настоящий результат показывает, что разрывы водородных связей не являются чисто механическим процессом, важную роль играют термические флуктуации.



Задания для самостоятельного выполнения

1. Уравнение $(1 - M_\infty^2) \frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = 0$, где φ – потенциал скоростей,

M_∞ – число Маха (отношение скорости потока газа к скорости звука в этом газе), описывает поведение малых возмущений однородного плоского стационарного потока газа. Исследовать тип уравнения в случаях дозвукового и сверхзвукового потоков газа. Определить, к какому качественному изменению ведет переход через скорость звука.

2. Определить тип уравнения Гельмгольца $\Delta u + u = 0$ (u – скалярное поле), дополнить его согласующимися с его типом условиями. Определить, при каких условиях соответствующая задача имеет периодическое решение.

3. Определить период пульсаций звезды (самогравитирующего газового шара) через систему определяющих параметров: M – массу звезды, R – радиус звезды и γ – гравитационную постоянную.

4. Построить модель движения сплошного потока автомашин по бесконечно длинной дороге, движущихся свободно, «накатом». Исследовать, какие качественные эффекты способна описывать построенная модель.

5. Выбрать, исходя из соображений пространственной или спектральной локализованности, для каждой из приведенных ниже задач, в которых может быть эффективно использован вейвлет-анализ, тип анализирующего вейвлета («мексиканская шляпа» или вейвлет Морле):

- имеется периодический сигнал, частота которого слегка «плавает» ($\Delta\omega \approx 5\%$). Требуется отслеживать вариации частоты;
- датчик дефектоскопа регистрирует дефекты в стальных прутках. Требуется изучить статистику дефектов (распределение по размерам и равномерность распределения по длине прутков);
- по электрокардиограмме требуется исследовать аритмию работы сердца, т.е. изучить вариации длительности отдельных импульсов и интервалов между ними;
- требуется создать компьютерную программу, реализующую нотную запись мелодии, регистрируемой микрофоном.

МОДЕЛИРОВАНИЕ С ИСПОЛЬЗОВАНИЕМ ИМИТАЦИОННОГО ПОДХОДА

Развитие моделей, использующих имитационный подход, связано с необходимостью исследования очень сложных систем, встречающихся на практике. Аналитические и численные методы, позволяющие провести наиболее полное исследование математической модели объекта, применимы далеко не для всех систем. При построении аналитических моделей многих сложных систем исследователю зачастую приходится идти на серьезные упрощения, чтобы получить представление о некоторых общих свойствах моделируемой системы, например оценить устойчивость стационарного состояния системы. Моделируемая система может оказаться настолько сложной, а поведение ее так многообразно и непредсказуемо, что принятая система гипотез будет способна привести не только к существенным количественным, но и качественным отличиям результатов моделирования от поведения системы в реальных условиях. При этом повысить степень адекватности модели будет нельзя по многим причинам: вследствие неразвитости существующих аналитических и численных методов, невозможности построения аналитического описания поведения отдельных элементов системы или взаимодействия между элементами и т.д. В [66] перечислены ситуации, когда исследователю рекомендуется применять модели, имитирующие поведение реального объекта:

- 1) если не существует законченной постановки задачи исследования и идет процесс познания объекта моделирования или отдельных его элементов;

2) если аналитические методы имеются, но математические процедуры трудно реализуемы, сложны и трудоемки;

3) когда кроме оценки влияния параметров сложной системы желательно осуществить наблюдение за поведением отдельных компонентов этой системы в течение определенного периода времени;

4) когда имитационный подход оказывается единственным способом исследования сложной системы из-за невозможности наблюдения явлений в реальной обстановке;

5) когда необходимо контролировать протекание процессов в сложной системе путем замедления или ускорения явлений в ходе имитации;

6) при подготовке специалистов и освоении новой техники, когда имитатор обеспечивает возможность приобретения необходимых навыков в эксплуатации новой техники;

7) когда изучаются новые ситуации в сложных системах, о которых мало что известно. В этом случае имитация служит для предварительной проверки новых стратегий и правил принятия решений перед проведением экспериментов на реальной системе;

8) когда основное значение имеет последовательность событий в проектируемой сложной системе и модель используется для предсказания узких мест в функционировании системы и других трудностей, связанных с добавлением в систему новых элементов.

Имитационный подход, в частности, оправдан, если вопросы, на которые должна ответить модель, относятся не к выяснению фундаментальных законов и причин, определяющих динамику реальной системы, а к анализу поведения системы, как правило, выполняемому сугубо в практических целях [30].

7.1. ОСОБЕННОСТИ МОДЕЛЕЙ, ИСПОЛЬЗУЮЩИХ ИМИТАЦИОННЫЙ ПОДХОД

Страх перед возможностью ошибки не должен отвращать нас от поисков истины.

Гельвеций

Суть подхода, используемого при разработке имитаторов, состоит в том, что процесс функционирования сложной системы представляется в виде определенного алгоритма, реализуемого на ЭВМ. Для систем, состоящих из множества элементов, приходит-

ся строить модель не только всей системы, но и модели отдельных элементов. Как и для аналитического подхода, разработка имитатора ведется с использованием некоторой совокупности гипотез. Изменение даже одной гипотезы для одного из элементов системы может привести к необходимости пересмотра всей модели системы и поиску новых методов исследования (именно поэтому аналитические и численные подходы к моделированию сложных систем применяют после длительного и всестороннего изучения поведения как всей системы, так и отдельных ее элементов). Имитационный подход позволяет *«максимально использовать всю имеющуюся в распоряжении исследователя информацию о системе»* [30].

Например, пусть требуется построить модель популяции живых существ с учетом взаимодействия с конкурирующими видами, хищниками и окружающей природой. Объектом моделирования в данном случае является сложная система, состоящая из живых существ разного вида, взаимодействующих друг с другом. При некоторых ограничениях можно описать изменение численности животных с помощью системы обыкновенных дифференциальных уравнений. Однако биологи, наблюдающие за реальной биосистемой, могут описать поведение отдельных особей только в виде совокупности правил и приближенно контролировать общую численность различных животных. Использовать эти наблюдения для определения констант, входящих в систему уравнений, достаточно сложно. В то же время построить алгоритм, реализующий указанную таблицу правил поведения одной особи, гораздо проще. Если известны правила взаимодействия с особями другого вида и с окружающей природой, то несложно разработать алгоритм, имитирующий поведение всей системы. Если в процессе дальнейшего наблюдения за реальной биосистемой будут уточняться или изменяться некоторые правила поведения или взаимодействия особей, то их учет можно выполнить, изменив соответствующий алгоритм, без существенной модификации всей модели. Таким образом, на самом раннем этапе исследований реальной системы можно получить ее рабочую модель, которая будет достаточно легко совершенствоваться по мере накопления экспериментального материала.

Имитаторы, как правило, используются для моделирования сложных динамических систем. При этом приходится моделировать не только структуру системы, но и время ее функционирования. При моделировании обычно используются три представления времени:

➤ *реальное время* моделируемой системы;

- *модельное время*, по которому организуется синхронизация событий в системе;
- *машинное время* имитации, отражающее затраты ресурсов времени ЭВМ на организацию имитации.

Для сложных систем возможна ситуация, когда различные события в различных компонентах происходят одновременно с точки зрения реального времени. Однако подавляющее большинство существующих современных вычислительных устройств не поддерживают распараллеливание вычислений. Поэтому возникает необходимость введения модельного времени, с помощью которого реализуется *квазипараллельная работа* компонент имитатора. Приставка «квази» в данном случае вводится для подчеркивания последовательного характера обработки событий в имитаторе, одновременно возникающих в разных компонентах реальной системы. В отличие от реального времени модельное изменяется не непрерывно, а пошагово. При этом величина шага по времени может быть фиксированной или переменной. При фиксированном шаге изменение модельного времени происходит всегда на одну и ту же величину. В случае переменного шага его величина соответствует интервалу времени между соседними событиями в системе. На практике большее распространение получил способ переменного шага.

Еще одной особенностью имитационного подхода является относительная простота учета стохастической неопределенности исходных параметров моделирования. Метод Монте-Карло достаточно хорошо подходит для моделирования параметров имитатора. Использование преобразования случайных величин позволяет получать распределения случайных параметров, соответствующие практически любому закону распределения случайных величин. Работа с имитатором представляет собой вычислительный эксперимент, осуществляемый на ЭВМ, который во многом сродни эксперименту с реальной системой. В связи с такой особенностью имитатор обычно дает ответы на вопросы лишь в статистическом смысле, что следует признать неизбежным при работе со сложными системами и более соответствующим поведению реальных объектов.

Построение имитатора можно представить как технологический процесс, многие этапы которого аналогичны этапам, рассмотренным в гл. 2, однако имеются и определенные отличия. Всего в этом случае можно выделить восемь технологических этапов [66]:

1. *Содержательное описание объекта моделирования*: формулируются основные вопросы о поведении сложной системы, ответы на которые требуется получить; определяется объект имитации; устанавливаются границы и ограничения моделирования; выбираются показатели для сравнения эффективности вариантов системы.

2. *Концептуальная модель системы*: на основе содержательного описания определяется общий замысел модели, выдвигаются основные гипотезы, фиксируются сделанные допущения. Концептуальная модель сложной системы представляет собой упрощенное алгоритмическое отображение реальной системы. Согласно [16], сложная система расчленяется на конечное число частей (декомпозиция системы), сохраняя при этом связи, обеспечивающие их взаимодействие. Полученные части при необходимости вновь расчленяются до тех пор, пока не получатся элементы, удобные для математического или алгоритмического описания. В результате этого сложная система представляется в виде многоуровневой «конструкции» взаимосвязанных элементов, объединенных в подсистемы (подмодели) различных уровней. При этом стремятся к тому, чтобы получаемые подмодели отвечали реально существующим фрагментам системы.

В состав концептуальной модели входят: уточненное описание объекта моделирования, свободное от всего того, что не представляет интереса для имитации поведения системы; список параметров и переменных моделирования; критерии эффективности функционирования вариантов системы; список используемых методов обработки результатов имитации и перечисление способов представления результатов моделирования.

3. *Формальное описание объекта моделирования*: построение исследователем формального представления алгоритмов поведения компонентов сложной системы и отражение вопросов взаимодействия компонентов между собой. Для составления формального описания используется один из трех видов формализации: аппроксимация явлений функциональными зависимостями, алгоритмическое описание процессов в системе и смешанное представление. После составления формального описания выполняют проверку правильности функционирования имитатора, используя классические модели, достоверность которых доказана. При этом выясняют следующие вопросы: позволяет ли имитатор решить поставленные задачи моделирования, насколько полна предложенная схема модели и отражает ли она фактическую последовательность разви-

тия процессов в реальной системе. На этом же этапе выполняется выбор вычислительных средств, которые обеспечили бы исследователю легкость программирования, минимальные затраты на моделирование, доступность выбранной ЭВМ, быстрое получение результатов.

4. *Конструирование имитатора*: преобразование формального описания в описание имитатора. Данный этап, как правило, вводится для сложных систем, которые нельзя представить в виде агрегативной схемы (см. гл. 4) или системы массового обслуживания. Здесь же прорабатываются вопросы, связанные с синхронизацией частей компонентов модели в модельном времени, организацией сбора статистики, заданием начальных условий моделирования, планированием процесса имитации отдельных вариантов системы, проверкой условий окончания моделирования, обработкой результатов имитации.

5. *Программирование и отладка модели*: этот этап практически не отличается от соответствующего этапа, рассмотренного в гл. 2 для аналитических и численных моделей. Отладочный процесс обязательно включает как независимую отладку отдельных компонентов, так и комплексную отладку программы для всей модели. Данный этап предполагает разработку технической документации на программную реализацию модели.

6. *Испытание и исследование модели*: проверка правильности алгоритма моделирования исследуемого объекта в ходе имитации его поведения; определение степени адекватности модели и объекта исследования. Под адекватностью программной реализации имитатора понимают степень совпадения с заданной точностью векторов характеристик поведения объекта и модели. При отсутствии адекватности проводят *калибровку* модели, т.е. уточняют алгоритмы как отдельных компонентов, так и взаимодействия компонентов. Эта операция может включать и уточнение формального описания компонентов.

Исследование свойств имитатора предполагает оценку точности и устойчивости результатов имитации явлений, а также определение чувствительности значений критериев качества к изменению параметров модели.

Под *точностью имитации* явления понимают оценку влияния стохастических элементов на функционирование модели сложной системы.

Устойчивость результатов моделирования характеризуется сходимостью контролируемых параметров к определенным величинам

при увеличении времени моделирования варианта сложной системы, что обычно наблюдается на практике в системах с конечным числом состояний.

Стационарность режима моделирования характеризует некоторое установившееся равновесие процессов в модели системы, при котором дальнейшее увеличение времени моделирования не приводит к получению новой информации.

Чувствительность модели определяется величиной максимального приращения значений выбранного критерия качества, вычисляемого по статистикам моделирования, при последовательном изменении параметров моделирования на всем диапазоне.

Следует отметить, что без оценки точности, устойчивости и чувствительности имитатора нельзя рассчитывать на доверие к нему со стороны как разработчика, так и заказчика.

7. *Эксплуатация имитатора.* Этап начинается с составления плана экспериментов, позволяющего исследователю получить максимум информации при минимальных затратах на проведение вычислений и обработку результатов. При этом необходимо выполнить обязательное статистическое обоснование плана экспериментов. Планирование эксперимента представляет собой процедуру выбора минимального числа и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с заданной точностью. После составления плана экспериментов приступают к их реализации. Итогом работы являются результаты экспериментов на модели.

8. *Анализ результатов моделирования:* всесторонний анализ полученных результатов с целью получения рекомендаций по проектированию системы или ее модификации. На результаты анализа и их интерпретацию существенное влияние может оказывать способ представления данных на ЭВМ. Так, использование средств компьютерной графики и мультипликации может оказать существенную помощь на данном этапе.

Успех или неудача проведения имитационных экспериментов с моделями сложных систем существенным образом зависит от инструментальных средств, используемых для моделирования, т.е. от набора аппаратно-программных средств, представляемых пользователю-разработчику или пользователю-исследователю имитатора. В настоящее время существует большое количество специальных языков создания имитаторов на ЭВМ, которые называют *языками моделирования*. Перед разработчиком возникает проблема выбора языка, наиболее эффективного для целей моделирования конкретной системы.

Языки моделирования заслуживают пристального внимания [101], так как, во-первых, число существующих языков и систем моделирования превышает несколько сотен, поэтому необходимо научиться ориентироваться в них. Во-вторых, почти каждый новый язык моделирования является не только средством, облегчающим доведение концептуальной модели до готовой машинной моделирующей программы, но и представляет собой новый способ «видения мира», т.е. построения моделей реальных систем. К наиболее известным языкам моделирования систем с дискретными событиями относят SIMULA, SIMSCRIPT, GPSS, SOL, CSL. Достаточно полный обзор разработанных к середине 90-х годов языков моделирования, их классификация, обсуждение достоинств и недостатков, сравнение между собой имеется в [101]. Описание, изучение и применение языков моделирования является предметом отдельного пособия. Поэтому для примеров, рассмотренных ниже, языки моделирования не используются.

В заключение еще раз остановимся на достоинствах и недостатках имитационного подхода. При достаточно глубоком знании поведения реальной системы и правильном представлении феноменологической информации в имитаторе последний характеризуется, вообще говоря, большей близостью к реальной системе, чем аналитические и численные модели. Такая близость в значительной степени обусловлена блочным принципом построения имитатора, который позволяет выполнять верификацию каждого блока до его включения в общую модель, а также благодаря тому, что имитатор может использовать зависимости более сложного характера, которые трудно (а иногда невозможно) учесть с помощью математических соотношений. В то же время создать хороший имитатор поведения сложной системы, как правило, труднее, дольше и дороже, чем аналитическую модель. Это связано с обязательным присутствием в цикле создания имитатора этапа программирования, стоимость, сложность и длительность которого может быть значительной. Кроме того, необходимость использования ЭВМ соответствующего класса в значительной степени осложняет использование модели. И, наконец, срок жизни имитатора, как правило, ограничен сроком использования ЭВМ, для которых была разработана его программная реализация.

Учитывая, что время разработки достаточно сложного имитатора составляет несколько лет, а период существенного обновления вычислительной техники сократился до 10 лет, — все это яв-

ляется серьезным недостатком моделей, основанных на имитационном подходе. Быстрое развитие средств вычислительной техники, совершенствование языков и технологий разработки имитаторов позволяет надеяться, что отмеченные недостатки будут со временем устраняться.

7.2. ИМИТАТОР СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ

Вся наука является не чем иным, как усовершенствованием повседневного мышления.

А. Эйнштейн

В гл. 5 рассматривались математические модели систем массового обслуживания (СМО). Под СМО понимают системы, на вход которых подается случайный поток однотипных заявок (событий), обработка которых выполняется одним или несколькими однотипными каналами (устройствами). Как правило, построить аналитическую модель подобной системы можно при соблюдении следующих условий:

- Система должна быть относительно простой. Например, участок цеха можно представить как относительно простую многоканальную СМО (с очередью или без нее); в то же время число и форма связей для цеха в целом (или для всего предприятия) могут быть достаточно сложными и трудно учитываемыми в аналитической модели.
- Потоки событий между элементами системы должны быть простейшими или близкими к ним (коэффициенты вариации потоков должны быть близки к единице). Если поток в системе не является простейшим и не описывается показательным распределением, то вероятности состояний нельзя описывать уравнениями Колмогорова.

При невыполнении данных условий аналитическое моделирование используется лишь в некоторых частных случаях, когда после принятия дополнительных гипотез возможно разрешение возникающих сложностей. Если эти сложности при аналитическом моделировании являются непреодолимыми, то представляется целесообразным применение моделей, использующих имитационный подход.

Пример 7.1. Одноканальная СМО с отказами. Рассмотрим имитатор одноканальной СМО с отказами (см. рис. 7.1). Пусть на вход системы поступает случайный поток заявок, интервал времени Δt между которыми является случайной величиной, распределенной по равномерному закону в интервале от t_1 до t_2 ($\Delta t = R(t_1, t_2)$). Время выполнения Δw заявки в канале также случайно и подчинено равномерному закону: $\Delta w = R(w_1, w_2)$. Требуется определить вероятность отказа в обслуживании и коэффициент загрузки канала.

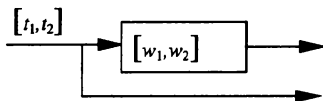


Рис. 7.1. Схема СМО

Выполним декомпозицию данной системы. В качестве отдельных элементов можно выделить следующие.

1. Источник заявок, который характеризуется:

- общим числом N_{\max} сгенерированных заявок за весь интервал наблюдений;
- числом $N(t)$ заявок, сгенерированных к моменту времени t ;
- законом распределения интервалов времени Δt между появлением заявок;
- временем t_n появления текущей заявки.

В некоторых случаях при рассмотрении неоднородных потоков источник может генерировать заявки разного типа (сложности), характеризуемые своим законом распределения.

2. Канал, который характеризуется:

- своим состоянием $S(t)$ (занят или свободен в момент времени t);
- числом N_w обслуженных заявок;
- суммарным временем T_w нахождения в занятом состоянии;
- производительностью Δw (закон распределения времен выполнения заявок определенного типа);
- временем W_n окончания обслуживания n -й заявки.

Имитация работы системы происходит в системном времени t . Введенные параметры для элементов СМО позволяют ответить на поставленные вопросы о параметрах эффективности данной системы. Зная число N_w обслуженных заявок в момент времени t из их общего числа N , можно оценить вероятность отказа $P_{\text{отк}}(t)$ ее частотой $W_{\text{отк}}(t)$:

$$P_{\text{отк}}(t) \approx W_{\text{отк}}(t) = (N(t) - N_w(t)) / N(t). \quad (7.1)$$

Коэффициент загрузки K_3 канала можно оценить, зная время работы T_w канала:

$$K_3(t) = T_w(t) / t. \quad (7.2)$$

Для разработки имитатора используем переменный шаг по времени. В результате работы имитатора происходит генерация N_{\max} заявок с заданными статистическими параметрами, имитирующими случайный процесс поступления заявок в реальной системе. Очевидно, что при увеличении N_{\max} (объема статистической выборки) увеличивается достоверность статистических оценок параметров эффективности системы. Алгоритм, реализующий имитатор, представлен на рис. 7.2. Запись этого алгоритма на псевдокоде приведена далее в Алгоритме 7.1.

На рис. 7.3 показано изменение параметров эффективности в зависимости от системного времени при следующих исходных данных: $w_1 = t_1 = 1$; $w_2 = t_2 = 3$; $N_{\max} = 200$.

Можно отметить, что с увеличением времени наблюдения за системой (т.е. с ростом объема выборки N_{\max} числа рассмотренных заявок) оценки параметров эффективности стремятся к некоторым предельным значениям. Указанные оценки являются значениями параметров эффективности, которые соответствуют финальным вероятностям рассматриваемой системы при заданных исходных данных. Так, для рассматриваемого случая оценка вероятности отказа $P_{\text{отк}}$ стремится к значению 0,34, а оценка коэффициента загрузки K_3 — к значению 0,66.

На практике при построении имитаторов сложных систем, как правило, предусматривается возможность наблюдения за величиной оценок параметров эффективности в зависимости от системного времени или числа сгенерированных заявок. Вычислительный эксперимент с имитатором прекращается при выходе наблюдаемых оценок на некоторое стационарное значение. Автоматический останов вычислительного эксперимента можно ввести в алгоритм имитатора, если предусмотреть вычисление величины максимального изменения отслеживаемых параметров за некоторый промежуток системного времени, начиная от текущего момента. Если модуль величины максимального изменения стал меньше некоторого наперед заданного малого положительного числа, то эксперимент с имитатором можно прекращать, а полученные значения отслеживаемых параметров считать решением задачи.

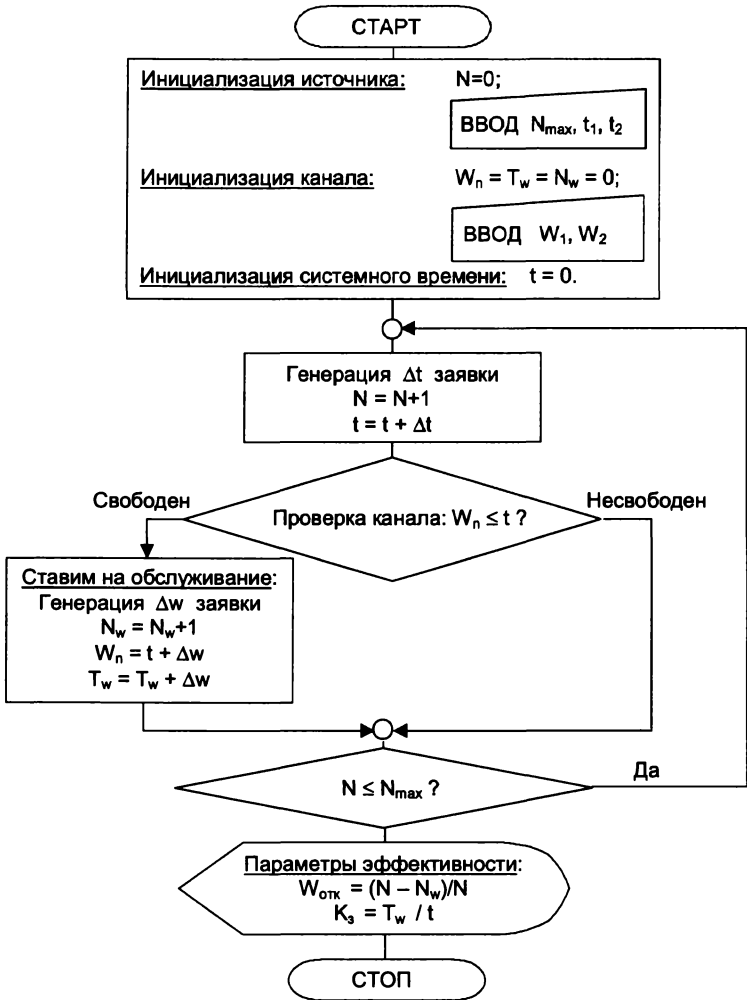


Рис. 7.2. Блок-схема имитатора

Алгоритм 7.1

programm Имитатор одноканальной СМО с отказами.

Данные: N_{\max} – максимальное число заявок;
 t_1, t_2 – границы разброса интервала времени между заявками;
 w_1, w_2 – границы разброса интервала времени обработки;
 ran – случайное число.

Результаты: $P_{\text{отк}}$ и K_3 – вероятность отказа и коэффициент загрузки.

procedure Random генерация случайного числа $R(0,65535)$

Данные: ran – случайное число

Выход: новое значение случайного числа.

start

$\text{ran} := (25173 r + 13849) \bmod 65536^*$

return

start

```
ran := 0           N := 0;
N_max := 200      t_1 := 1;           t_2 := 3
w_t := 0;         T_w := 0;          N_w := 0
w_1 := 1;        w_2 := 3
t := 0;
while (N ≤ N_max)  цикл по заявкам
  Random
  Δt := t_1 + (t_2 - t_1) * ran / 65536
  t := t + Δt
  N := N + 1
  if wt ≤ t then проверка канала
    Random
    Δw := w_1 + (w_2 - w_1) * ran / 65536
    N_w := N_w + 1
    W_t := t + Δw
    T_w := T_w + Δw
  end if
end while
W_отк := (N - N_w) / N
K_3 := T_w / t
```

stop

* Операция mod – вычисление остатка при целочисленном делении.
Например: $7 \bmod 3 = 1$.

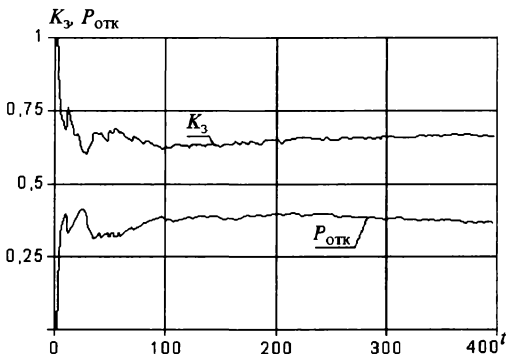


Рис. 7.3. Изменение параметров эффективности в зависимости от системного времени

7.3. КЛЕТОЧНЫЕ АВТОМАТЫ

Нельзя доверять моделированию без знания теории.

Эддингтон

Автоматом называют устройство (или совокупность устройств), которое без непосредственного участия человека выполняет процессы приема, преобразования и передачи энергии, материалов или информации в соответствии с заложенной в него программой. В данном случае нас интересуют автоматы (рис. 7.4), выполняющие дискретное преобразование информации, для кото-

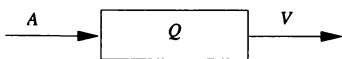


Рис. 7.4. Автомат

рых заданы множества A входных сигналов, Q внутренних состояний и V выходных сигналов, а также две функции: δ — функция переходов и λ — функция выходов. Первая определяет, в какое состояние q' (из множества возможных состояний Q) перейдет автомат, если он находится в состоянии q и на вход поступил сигнал $a \rightarrow \delta(q, a) = q'$, а вторая показывает, какой при этом образуется выходной сигнал v (из множества выходных сигналов V): $\lambda(q, a) = v$. Множество сигналов и состояний дискретны, и, кроме того, дискретно множество моментов, в которые поступают

входные сигналы, выдаются выходные сигналы и меняются состояния. В этом случае число возможных значений аргументов функций переходов и выходов также конечно, и эти функции можно задавать таблично. Подобные автоматы называются *конечными*.

Клеточные автоматы (КА) являются частным случаем [107] конечных автоматов, используемых для моделирования динамического поведения однородных сред. При этом пространство и время считаются дискретными, а физические величины в каждой точке моделируемой среды могут принимать конечное множество дискретных значений. Для КА существует достаточно развитая теория. В ее основу легли работы Дж. фон Неймана, который в 1948 г. ввел в науку само понятие «клеточный автомат» при разработке первой компьютерной модели биологического самовоспроизводства. Клеточный автомат можно представить, как регулярную решетку (или «таблицу») ячеек («клеток»), каждая из которых может находиться в конечном числе возможных состояний, например 0 или 1. Состояние системы полностью определяется значениями переменных в каждой клетке. Важными особенностями клеточных автоматов являются следующие:

- состояние каждой ячейки обновляется в результате выполнения последовательности дискретных постоянных шагов во времени (или тактов);
- переменные в каждой ячейке изменяются одновременно («синхронно»), исходя из значений переменных на предыдущем шаге;
- правило определения нового состояния ячейки зависит только от локальных значений ячеек из некоторой окрестности данной ячейки.

Широко используются [101] автоматы для моделирования динамического поведения двумерных сред. Каждой частице такой среды, занимающей некоторое пространство, ставится в соответствие элементарный автомат, имеющий форму квадрата (реже треугольника или шестиугольника) и именуемый *клеткой*. Совокупность клеток образует *клеточное пространство*, в котором функционирует КА. Как отмечалось выше, отдельная клетка имеет конечный набор состояний Q , а выходные сигналы клеток есть номера их состояний. Функция переходов δ определяется текущим состоянием q клетки, а также состоянием ее окружения.

В зависимости от свойств моделируемого объекта выбирают различные типы окрестностей (рис. 7.5). В простейших случаях в

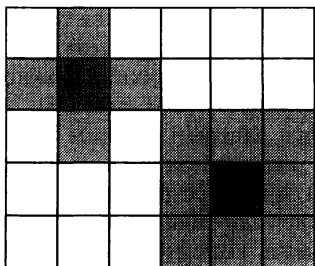


Рис. 7.5. Типы окрестностей

качестве окрестности принимают четыре или восемь ближайших клеток. Однако возможны ситуации, когда состояние клетки зависит от состояния более удаленных клеток. Чтобы исключить особый вид окрестности для клеток, лежащих в крайних рядах, область замыкают, т.е. для клеток крайнего левого столбца соседями слева считают клетки крайнего правого столбца, и наоборот. Аналогичное замыкание выполняют для

клеток верхнего и нижнего ряда области. В результате можно считать, что рассматриваемая совокупность клеток лежит на поверхности тора (рис. 7.6). Подобный прием в практике моделирования называется заданием *границных условий циклического типа*.

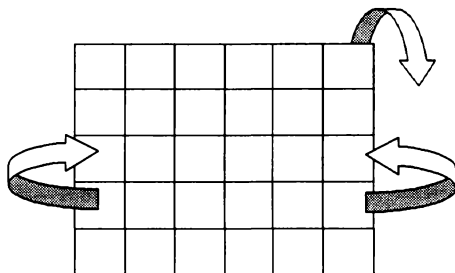


Рис. 7.6. «Сворачивание» области

Функцию переходов δ для клетки принято называть множеством *правил КА*. Наиболее известным примером КА считается автомат «Жизнь», придуманный в 1970 г. Дж. Конвеем. Множество правил для данного автомата состоит в следующем:

- клетка может находиться в двух состояниях – пассивном и активном;
- в качестве окрестности рассматривается восемь соседних клеток;
- если в окрестности пассивной клетки две активных, то данная клетка также становится активной («рождается»);
- если в окрестности активной клетки три или более активных клеток, то она становится пассивной («умирает»).

В начальный момент времени дается некоторое распределение активных и пассивных клеток. Используя полученное распределение, определяем состояние клеток на следующем шаге по времени и т.д. В зависимости от вида начального распределения приходим к различным эволюциям состояний клеток. Например, случайное начальное распределение приводит к случайной эволюции последующих состояний. На рис. 7.7 показаны некоторые состояния автомата размером 256×256 ячеек, полученные для начального состояния, когда активные клетки занимали квадратную область. На втором шаге активными остаются клетки только вблизи углов квадрата. Затем они образуют четыре структуры, которые начинают развиваться, образуя симметричный узор.

Пример простейшего алгоритма КА «Жизнь», записанный на псевдокоде, приведен в алгоритме 7.2.

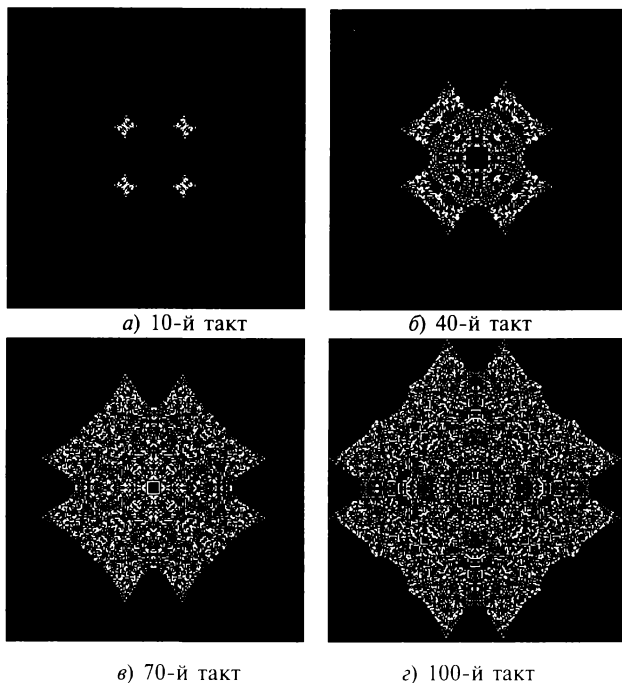


Рис. 7.7. Клеточный автомат «жизнь» (начальная область — квадрат)

Алгоритм 7.2

programm КА "Жизнь". Клеточное пространство 256 на 256 клеток.

Данные: X_{\max}, Y_{\max} – размер клеточного пространства;
 $KA[0...1, 1... X_{\max}, 1... Y_{\max}]$ – два состояния клеток КА;
 k, m – номер старого и нового состояния КА;
 T_{\max} – рассматриваемое число тактов;
 ran – случайное число.

procedure Random генерация случайного числа $R(0, 65535)$

Данные: r – случайное число

Выход: новое значение случайного числа.

start

$ran := (25173 r + 13849) \bmod 65536$

return

start

$ran := 0; \quad X_{\max} = 256; \quad Y_{\max} = 256; \quad T_{\max} = 500;$

for $x = 1$ **to** X_{\max} **do** инициализация массивов состояний

for $y = 1$ **to** Y_{\max} **do**

$KA_{0xy} = 0$

Random

$KA_{1xy} = ran \text{ div } 32735^*$

Next y

Next x

$m = 1; \quad k = 0$

for $n = 1$ **to** T_{\max} **step** 1 цикл на 500 тактов

for $x = 1$ **to** X_{\max} **do** цикл по клеткам

if $x > 1$ **then** $L = x - 1$ **else** $L = X_{\max}$ номер левой клетки

if $x < X_{\max}$ **then** $R = x + 1$ **else** $R = 1$ номер правой клетки

for $y = 1$ **to** Y_{\max} **do**

if $y > 1$ **then** $U = y - 1$ **else** $U = Y_{\max}$ номер верхней клетки

if $y < Y_{\max}$ **then** $D = y + 1$ **else** $D = 1$ номер нижней клетки

$S = KA_{kLU} + KA_{k0U} + KA_{kRU} + KA_{kLy} + KA_{kRy} + KA_{kLD} + KA_{k0D} + KA_{kRD}$

case S **of** смена состояния клетки

2 : **if** $KA_{k0y} = 0$ **then** $KA_{mxy} = 1$ **else** $KA_{mxy} = 0$

3..8 : **if** $KA_{k0y} = 1$ **then** $KA_{mxy} = 0$ **else** $KA_{mxy} = 1$

end case

Next y

Next x

Вывод на экран информации о новом состоянии клеток

$m = (m+1) \bmod 2; \quad k = (k+1) \bmod 2$

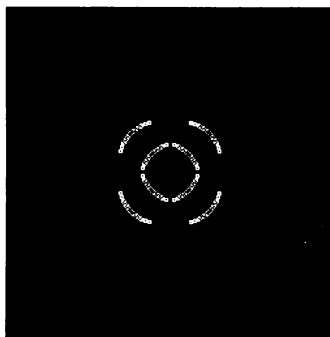
next n

stop

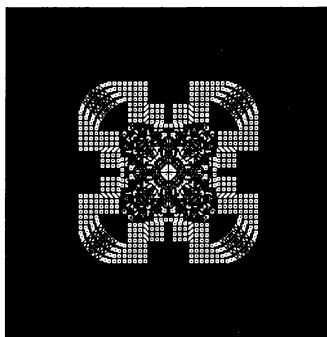
* Операция div – вычисление целой части при целочисленном делении. Например: $7 \text{ div } 3 = 2$.

Автомату «Жизнь» посвящено достаточно много исследований [101]. Как правило, это первый клеточный автомат, с которым встречается исследователь, начинающий знакомство с данным классом алгоритмических моделей. Автомат прославился своими многочисленными очаровательными узорами и способностью заставлять людей тратить огромное количество машинного времени на поиски более интересных узоров.

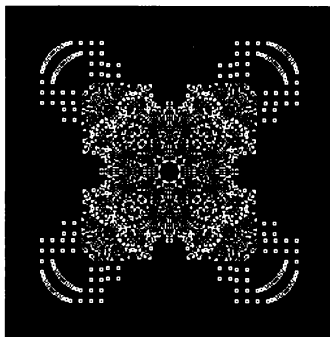
Можно придумать достаточно много вариаций КА «Жизнь». Например, на рис. 7.8 приведены состояния КА, реализующего следующее множество правил:



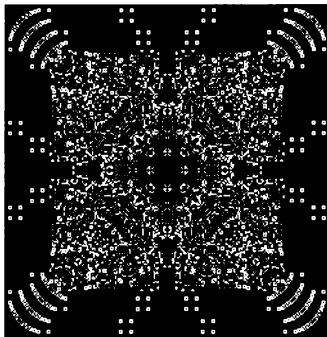
а) 10-й такт



б) 40-й такт



в) 70-й такт



г) 94-й такт

Рис. 7.8. Клеточный автомат « N из 8»
($N = 1$; начальная область – круг)

- клетка может находиться в двух состояниях: пассивном и активном;
- в качестве окрестности рассматриваются восемь соседних клеток;
- клетка становится активной, если в ее окрестности находится ровно N активных клеток;
- если число активных клеток в окрестности не равно N , то клетка становится пассивной.

Рассмотренные примеры КА довольно сложно использовать в качестве имитаторов каких-то реальных объектов. Однако оказалось, что КА прекрасно подходят для моделирования поведения так называемых активных сред. Активные среды характеризуются непрерывным распределенным притоком энергии от внешнего источника и ее диссипацией [64]. Благодаря тому, что через каждый физически малый элемент среды протекает поток энергии от источника к термостату, этот элемент выводится из состояния равновесия и приобретает способность совершать автоколебания, быть триггерным (бистабильным) либо возбудимым. Если такие отдельные элементы локально связаны между собой и формируют распределенную активную среду, в последней наблюдается образование различных стационарных или зависящих от времени пространственных структур. Эти процессы лежат в основе явлений самоорганизации, происходящих в активных средах. В качестве примеров активных сред можно назвать некоторые растворы с протекающими в них специфическими химическими реакциями, нервные волокна, горючие среды и т.д.

Выделяют три типа простейших активных элементов: бистабильные, возбудимые и автоколебательные.

Бистабильный (или триггерный) элемент обладает двумя устойчивыми состояниями, в каждом из которых он может находиться неограниченно долго. Если внешние воздействия превышают некоторое пороговое значение, то элемент может сменить свое состояние. Одним из вариантов бистабильного элемента можно считать элементарный автомат, применяемый в КА «Жизнь» и « N из 8».

Возбудимый (или мультивибраторный) элемент имеет несколько состояний, но устойчивым к достаточно слабым воздействиям является только одно из них. Если воздействие превышает некоторый порог, то возбудимый элемент совершает ряд переходов из состояния в состояние и лишь затем возвращается к первоначальному устойчивому состоянию. Пример среды из возбудимых эле-

ментов — нервное волокно. Если раздражение превышает некоторый уровень, то нервная клетка переходит в возбужденное состояние, создавая в свою очередь раздражение для соединенных с ней клеток и переводя их в возбужденное состояние. Затратив энергию на возбуждение соседей, клетка переходит в состояние восстановления (рефрактерности), когда она не может воспринимать новые раздражения и возбуждаться. Затем клетка возвращается в исходное состояние покоя, готовая к передаче новых импульсов. Подобную модель нервных клеток предложили в 1946 г. Н.Винер и А.Розенблут [64].

Автоколебательный элемент работает подобно «вечному двигателю», автономно совершая циклические переходы из одного состояния в другое. Внешние воздействия могут замедлить или ускорить эти переходы, но не могут остановить их.

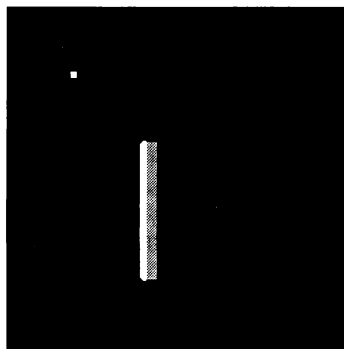
Рассмотрим пример КА «Нейронная сеть», который имитирует явления в однородной двумерной нейронной сети, состоящей из возбудимых элементов. Поведение КА основано на следующей совокупности правил:

- клетка может находиться в трех состояниях: покоя, активном и состоянии восстановления;
- в качестве окрестности рассматриваются восемь соседних клеток;
- переход в состояние активности зависит от некоторого параметра, называемого *уровнем активатора*. В возбужденном состоянии клетки уровень активатора равен 1. В других состояниях он распадается на $A\%$ за такт;
- если клетка была в покое и общее количество активатора в восьми соседних и в данной клетке превысило порог активации T_a , то клетка возбуждается на T_v тактов;
- через T_v тактов клетка переходит в состояние восстановления на T_p тактов, а затем в состояние покоя.

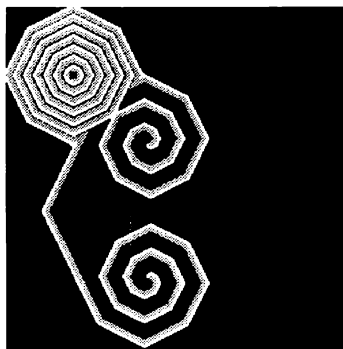
На рис. 7.9 показаны состояния КА «Нейронная сеть» при следующих параметрах: $A = 30\%$; $T_a = 3$; $T_v = 5$; $T_p = 8$. Начальное распределение состояния клеток задано плоским фронтом. В левом верхнем углу действует периодический источник возбуждения (*нейсмекер*) с периодом 15 тактов.

Более подробно вопросы моделирования активных сред с помощью клеточных автоматов рассмотрены в [64].

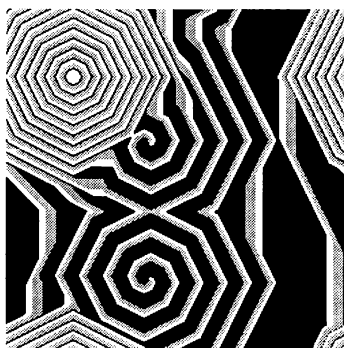
Еще одним интересным примером использования КА являются биологические системы. Представим сосуд с питательной средой, содержащий колонию одноклеточных (амебы, инфузории и



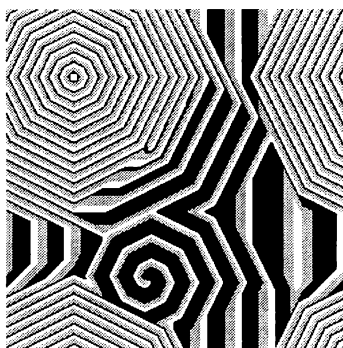
а) 1-й такт



б) 100-й такт



в) 200-й такт



г) 300-й такт

Рис. 7.9. Клеточный автомат «Нейронная сеть»

т.п.). Одноклеточные посредством простейших приспособлений (например, ресничек) могут перемещаться. Необходимые компоненты для своей жизнедеятельности эти организмы берут из окружающей среды. Питательная среда, обедняясь в процессе деятельности организмов, имеет возможность за определенный период времени полностью восстанавливать свои свойства. Например, «питательность» среды могут обеспечивать более простейшие живые существа, служащие кормом для рассмотренных выше одноклеточных. Каким образом будет эволюционировать описанная система «организмы – питательная среда» с течением времени?

Так как рассматриваемая система конечна в силу ограниченности сосуда с питательной средой, то в качестве параметров можно использовать N – численность одноклеточных и P – степень питательности раствора. Эволюционные уравнения для подобной системы известны в литературе [70] как уравнения Лотки–Вольтера для системы «хищник–жертва»:

$$\begin{aligned} \frac{dP}{dt} &= (r_p - c_1 N - bP)P, \\ \frac{dN}{dt} &= (-r_n + c_2 P)N, \end{aligned} \quad (7.3)$$

где r_p – скорость восстановления питательности раствора; r_n – скорость гибели одноклеточных; b – степень замедления скорости восстановления питательности раствора; c_1 и c_2 – коэффициенты взаимного влияния питательности раствора и численности одноклеточных. Графики изменения $N(t)$ и $P(t)$, полученные из решения уравнений Лотки–Вольтера (7.3), имеют вид затухающих колебаний (рис. 7.10). При затухании система стремится к положению равновесия при значениях параметров

$$P_p = \frac{r_n}{c_2}, \quad N_p = \frac{r_p c_2 - b r_n}{c_1 c_2}. \quad (7.4)$$

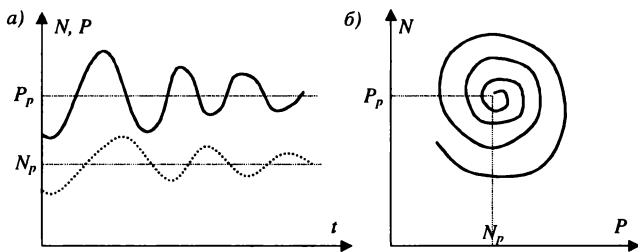


Рис. 7.10. Зависимость численности и степени питательности раствора от времени (а); фазовая диаграмма (б)

На рис. 7.10,б приведена фазовая диаграмма, показывающая изменение численности одноклеточных в зависимости от питательности среды. Аналитическая модель (7.3) позволяет получить точ-

ные значения параметров в положении равновесия. Сложность использования данной модели заключается в трудностях, связанных с выбором коэффициентов модели для реальной системы.

Рассмотрим построение КА для подобной биологической системы. Примем следующую совокупность правил:

1. Клеточное пространство КА образует поле 256×256 клеток.
2. Окрестность клетки КА составляют восемь соседних клеток.
3. Каждой клетке соответствует значение p степени питательности раствора (энергоемкости), которое может изменяться от 0 до p_{\max} .
4. Прирост Δp питательности (энергоемкости) раствора клетки за такт времени выполняется следующим образом:

$$\Delta p = \begin{cases} 0 & p = p_{\max}, \\ r_p & p < p_{\max}, \end{cases} \quad (7.5)$$

где r_p — скорость прироста питательности.

5. Общий запас энергии питательного раствора определяется суммарной питательностью (энергией) всех клеток и не может быть более

$$P_{\max} = 256^2 p_{\max}. \quad (7.6)$$

6. Общий запас энергии КА будем характеризовать относительной величиной

$$P = \sum_{x=1}^{256} \sum_{y=1}^{256} p_{xy} / P_{\max}. \quad (7.7)$$

7. Клетка КА может быть свободной или содержать не более одного одноклеточного или другого живого организма.

8. Отдельная особь одноклеточного черпает энергию из питательного раствора клетки, в которой она находится, снижая его питательность и повышая свой запас энергии на Δp_N за такт.

9. Максимально возможное количество энергии, запасаемое одноклеточным, не превышает p_N .

10. На свои нужды отдельная особь затрачивает Δe_N энергии за такт.

11. Особь всегда старается перейти на соседнюю свободную клетку, выбирая направление перехода случайным образом.

12. Время жизни отдельной особи составляет L тактов.

13. Если время жизни особи превысило продолжительность жизни для данных организмов или запас энергии снизился до нуля, особь умирает.

14. Начиная с возраста L_3 тактов особь считается зрелой и может производить себе подобных, затрачивая Δr_N энергии при каждом

делении дополнительно. При этом старая особь переходит на свободную соседнюю клетку, а новая остается в старой. Если свободных клеток в окрестности нет, то деления не происходит.

15. Начальное распределение особей по клеточному пространству подчинено равномерному закону распределения. Начальное число особей составляет $N_0\%$ максимально возможного, равного $N_{\max} = 256^2 = 65536$. Общее число особей в КА будем характеризовать абсолютной величиной n или относительной величиной $N = n/N_{\max}$.

Рассмотрим КА со следующими исходными данными:

Питательный раствор: $p_{\max} = 10$; $r_p = 1$.

Одноклеточный организм:

$$N_0 = 30\%; \quad L = 15; \quad L_3 = 3;$$

$$\Delta p_N = 5; \quad p_N = 35; \quad \Delta e_N = 2; \quad \Delta r_N = 3.$$

На рис. 7.11 приведены: *a* – зависимости P и N от времени (номера такта); *б* – фазовая диаграмма и *в* – типичное распределение особей в клеточном пространстве. Как можно видеть зависимости P и N от времени носят колебательный затухающий характер, стремясь к состоянию равновесия ($P_p \approx 18,7$; $N_p \approx 56,06$). Колебания P и N изменяются в противофазе, что является характерным для подобных систем. Период колебаний составляет 22 такта. Фазовая диаграмма представляет собой спиральную линию, сходящуюся к равновесной точке. Распределение особей в клеточном пространстве носит равномерный характер с незначительными областями сгущения и разрежения.

Интересно посмотреть, каким образом сказывается на эволюции системы изменение правил. Заменяем 11-е правило следующим: *особь всегда старается перейти на соседнюю свободную клетку с наибольшим уровнем энергоемкости. Если ячейки в окрестности имеют меньший запас энергии, то особь остается в прежней клетке.*

На рис. 7.11 также показаны: *г* – зависимости P и N от времени; *д* – фазовая диаграмма; *е* – типичное распределение особей в клеточном пространстве, отвечающее такому «умному» поведению особей. Как можно видеть, уже через 100 тактов система приходит к состоянию динамического равновесия (средние значения составляют $P_p \approx 16,8$; $N_p \approx 52,4$, что несколько меньше соответствующих значений для «глупого» поведения особей).

В дальнейшем параметры системы колеблются относительно средних значений с небольшой амплитудой. На рис. 7.11,₂ показано увеличенное изображение колебаний величины относительной

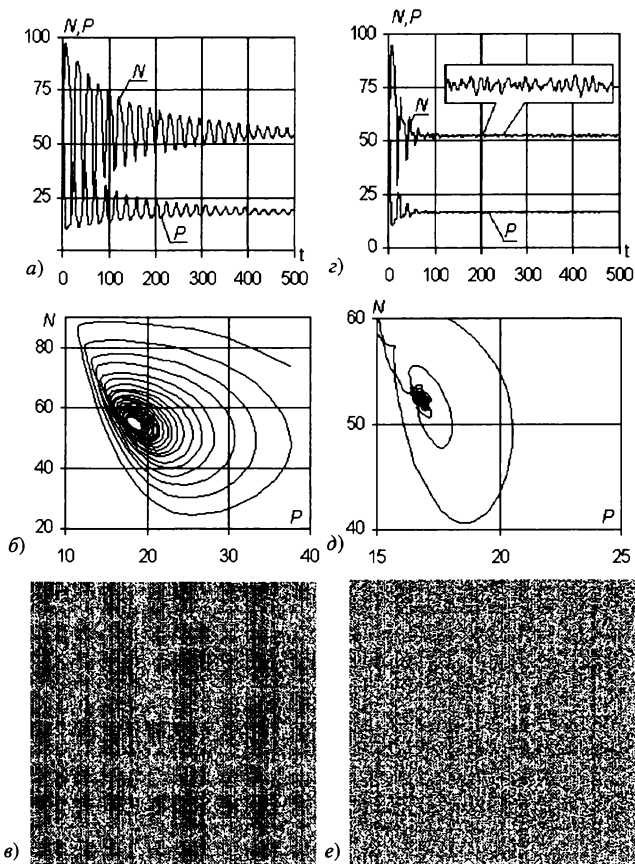


Рис. 7.11. КА «питательная среда – одноклеточные»

численности особей в состоянии динамического равновесия. Сравнивая рис. 7.11,в и рис. 7.11,е, можно отметить, что распределение особей по клеточному пространству при «умном» поведении более равномерное, чем в случае «глупого» поведения. Данное обстоятельство можно объяснить тем, что «умные» особи более жизнеспособны. Они стремятся перейти на клетки с большей энергией, создавая более равномерное заполнение клеточного пространства и повышая устойчивость всей системы.

Следует отметить, что вариант системы с «глупым» поведением особой также достигает состояния динамического равновесия, но за значительно больший период времени.

Что произойдет с поведением системы, если в нее поместить еще один тип организмов, для которых одноклеточные сами служат пищей? Для описания поведения подобных «хищников» добавим к рассмотренным выше 15 правилам следующие:

16. «Хищник» может нападать на свою «жертву», если она находится в его окрестности, т.е. в одной из восьми соседних клеток. При этом «жертва» погибает с вероятностью P_N , а «хищник» перемещается в ее клетку, повышая запас своей энергии на величину энергетической ценности «жертвы». Вероятность гибели «жертвы» не равна 1, так как предполагается, что она имеет некоторые средства защиты. Если «жертва» остается жива, то погибает «хищник».

17. Если в окрестности нет одноклеточных, то «хищник» переходит на соседнюю клетку, выбирая направление случайным образом.

18. Характеристиками «хищника» являются:

- максимально возможное количество запасаемой энергии p_X ;
- энергия Δe_X , затрачиваемая на свои нужды за такт времени;
- время жизни L_X хищника в тактах;
- возраст зрелости L_{X3} ;
- энергия Δr_X , затрачиваемая при делении;
- энергетическая ценность E_N поедаемого одноклеточного;
- вероятность P_N остаться в живых при нападении на одноклеточного.

19. Если время жизни превысило максимально возможное или запас энергии снизился до нуля, то «хищник» умирает.

20. Начальное распределение «хищников» по клеточному пространству подчинено равномерному закону распределения. Начальное число «хищников» составляет $X_0\%$ максимально возможного N_{\max} . Общее число «хищников» в КА будем характеризовать абсолютной величиной x или относительной величиной $X = x/N_{\max}$.

На рис. 7.12 приведены результаты для КА со следующими исходными данными:

Питательный раствор: $p_{\max} = 10$; $r_p = 1$.

Одноклеточный организм:

$N_0 = 30\%$; $L = 15$; $L_3 = 3$;
 $\Delta p_N = 5$; $p_N = 35$; $\Delta e_N = 2$; $\Delta r_N = 3$.

«Хищник»:

$X_0 = 5\%$; $L = 10$; $L_3 = 5$;
 $p_X = 20$; $\Delta e_X = 4$; $\Delta r_X = 5$; $E_N = 10$; $P_N = 93\%$.

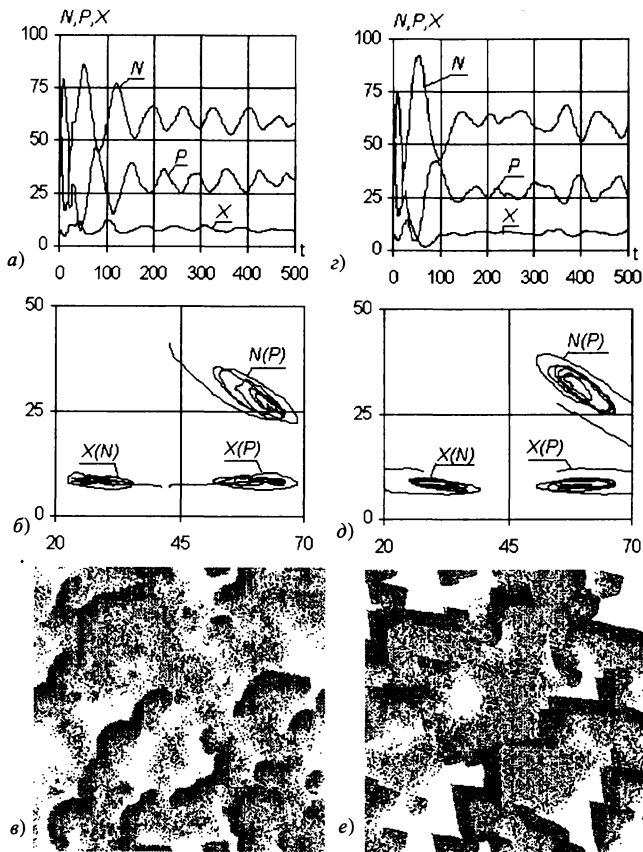


Рис. 7.12. КА «питательная среда – одноклеточные – хищники»

Как можно видеть, параметры для питательной среды и одноклеточных совпадают с рассмотренными выше. На рис. 7.12 приведены результаты расчетов зависимости P , N и X от времени (а, с), фазовые диаграммы (б, д) и типичное распределение особей в клеточном пространстве (е, ф), отвечающее «глупому» (а–в) и «умному» (с–е) поведению одноклеточных.

Прежде всего следует отметить отсутствие равномерного распределения одноклеточных по клеточному пространству. Появляются подобласти четырех типов:

- свободные от одноклеточных и хищников (белого цвета);
- практически полностью занятые одноклеточными, соответствующие надвигающемуся на свободную подобласть фронту одноклеточных (темно-серый цвет);
- с малой плотностью одноклеточных, расположенные после прохождения фронта и похожие на распределение особей при отсутствии «хищников» (светло-серый цвет);
- занятые одноклеточными и «хищниками», расположенные на границе областей с малой плотностью и свободными областями (черные точки – «хищники», серые – одноклеточные).

Фронты постоянно перемещаются в направлении свободного пространства. Интересно отметить, что для «умного» поведения одноклеточных характерно наличие прямолинейных или клиновидных фронтов. При этом острие клина направлено в сторону движения фронта. Отсутствие «хищников» перед фронтом можно объяснить наличием у одноклеточных средств защиты. Высокая плотность особей внутри фронта и, как следствие, частые встречи «хищника» с множеством одноклеточных практически не оставляют шансов для его выживания. Поэтому «хищники» следуют за фронтом, нападая на особей из области с малой плотностью.

Изменения коснулись также зависимостей для параметров КА. Во-первых, в значительной степени увеличился период колебаний значений параметров. Однако определение величины периода колебаний в данном случае затруднено. Рассматривая зависимости параметров КА от времени как некоторые одномерные сигналы, применим вейвлет-анализ (см. гл. 6). Исследования колебаний численности от времени проводились как с использованием вейвлета «мексиканская шляпа», так и Марле. Результаты по диаграммам энергетического спектра при использовании обоих типов вейвлетов оказались близки как в качественном, так и в количественном отношении. Поэтому далее приводятся результаты, полученные с использованием только «мексиканской шляпы».

На рис. 7.13 приведены результаты вейвлет-анализа зависимости $X(t)$ (см. рис. 7.12,г, численность хищников для «умных» одноклеточных). При анализе использовались значения, начиная с 100-го такта, соответствующие более близким к равновесным состояниям КА. Оценку периода колебаний можно выполнить по диаграмме энергетического спектра.

Как можно видеть, график на диаграмме имеет один максимум. Удвоенное значение масштаба, соответствующее максимуму, можно принять за оценку периода колебаний.

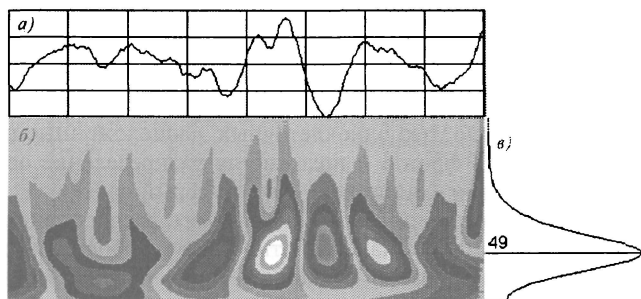


Рис. 7.13. Результаты вейвлет-анализа зависимости $X(t)$ с рис. 7.12, г: а – профиль сигнала; б – картина вейвлет-коэффициентов; в – диаграмма энергетического спектра

На рис. 7.14 приведены диаграммы энергетического спектра. Как можно видеть, добавление для КА «питательная среда—одно-

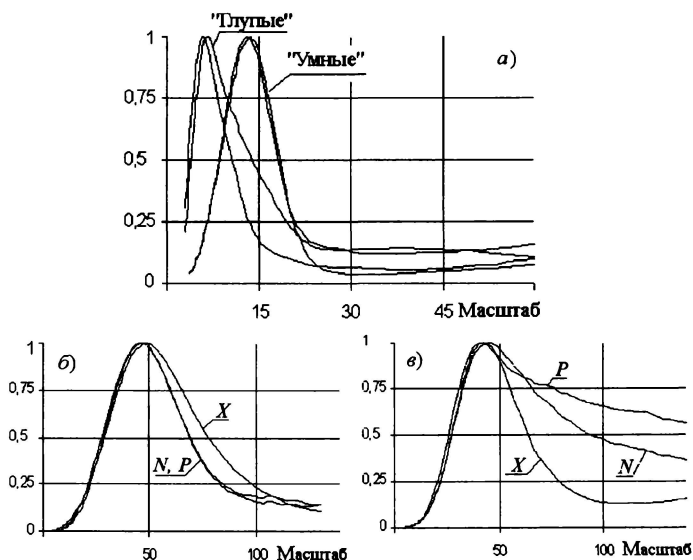


Рис. 7.14. Диаграммы энергетического спектра: а – КА «питательная среда—одноклеточные»; б и в – КА «питательная среда—одноклеточные—хищники»; (б – «глухие»; в – «умные»)

клеточные» способности выбора направления перехода для одно-клеточных («ума») приводит к почти двукратному увеличению периода колебаний (с 11–12 до 26–28 тактов). Добавление в систему «хищников» увеличивает период колебаний почти до 100 тактов.

Как правило, моделируемые сложные системы представляют собой открытые термодинамические системы (т.е. обменивающиеся с внешней средой веществом, импульсом, энергией, информацией и т.д.), способные к диссипации (рассеянию) подводимой энергии. Единоробство этих двух механизмов приводит к появлению диссипативных структур из начально случайного «бульона» субстанции. Можно сказать, что подчас поразительно красивые структуры возникают для более эффективного противодействия внешним воздействиям, а в симметрию «рисунка» вносят вклад симметрии свойств среды и внешних воздействий.

7.4. МОДЕЛИРОВАНИЕ ДИСЛОКАЦИЙ В МЕТАЛЛЕ

Некоторые вещи недоступны человеческому уму, но мы не знаем какие.

«Закон Джеффи»

Самоорганизация дислокаций в модели клеточных автоматов

Рассмотрим моделирование явлений самоорганизации дефектов кристаллического строения, наблюдаемых в пластически деформируемых металлах. Природой пластического течения металлов (в особенности — с гранецентрированной кубической решеткой) при обычных температурах является движение образующихся в них при деформировании в огромном количестве линейных дефектов кристаллического строения — дислокаций. Учет этих дислокаций в одночастичном приближении в моделях пластичности объяснил многие особенности пластического течения. Эти теории применимы для описания процессов на ранних стадиях пластического течения, когда дислокаций в кристалле не так много, чтобы учитывать их многочастичное взаимодействие. Достаточно тонкие наблюдения фиксируют на последующих стадиях временную нестабильность и пространственную локализацию пластического тече-

ния, отражающие возникновение режимов самоорганизации дислокаций, приводящих к структурным переходам. Для описания этих процессов нужны модели, трактующие пластичность и описывающие ряд эффектов пластичности как результат коллективных явлений в ансамблях дислокаций. Одним из решений этой задачи является применение имитационного подхода, описывающего на основе техники клеточных автоматов поведение большого количества дискретно распределенных взаимодействующих дислокаций.

Дислокации могут быть краевыми, винтовыми и смешанного типа. Дислокацию характеризуют два вектора: единичный ξ , направленный вдоль линии дислокации (на рис. 7.15,б за плоскость рисунка ортогонально ей) и особый (ненулевой только для линейных дефектов типа дислокаций) – вектор Бюргерса \mathbf{b} . Краевые дислокации представляют собой края лишних атомных полуплоскостей, и для них $\xi \perp \mathbf{b}$ (рис. 7.15,б); для винтовых дислокаций $\xi \parallel \mathbf{b}$, для смешанных – взаимная ориентация векторов ξ и \mathbf{b} меняется вдоль линии дислокации. В зависимости от того, как расположена лишняя атомная полуплоскость – ниже плоскости залегания или выше, т.е. в зависимости от знака вектора Бюргерса, краевые дислокации условно разделяют на положительные и отрицательные и обозначают символами \perp и \top . Линии дислокаций из геометрических соображений могут обрываться только на поверхности кристалла (но не внутри него) или на других дислокациях, образуя узлы дислокационной сетки. При обычных температурах основным механизмом движения краевых дислокаций является их скольжение в плоскости залегания (вдоль горизонтали, см. рис. 7.15).

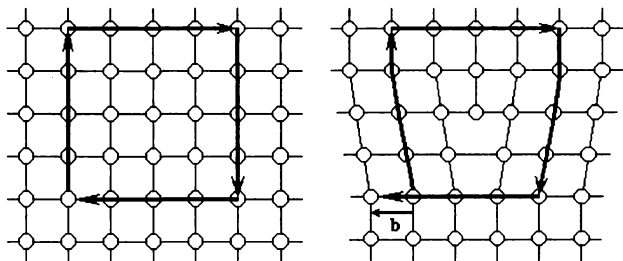


Рис. 7.15. Контуры Бюргерса: а – в идеальном кристалле, б – в кристалле с краевой дислокацией (мера несовпадения – вектор Бюргерса \mathbf{b})

При встрече краевых дислокаций противоположных знаков выстраивается полная атомная плоскость – дислокации аннигилируют.

Теоретически рассчитано, что для сдвига одной части кристалла относительно другой в бездефектном кристалле необходимо приложить касательное усилие, большее $\mu/(2\pi)$, где μ – модуль сдвига. При наличии же краевой дислокации достаточно приложения ничтожного напряжения Пайерлса $\tau_* = 10^{-5} - 10^{-4} \mu$ для незначительного смещения атомов в ядре дислокации, в результате которого «перекидываются» атомные связи и происходит скольжение краевой дислокации, реализующее пластический сдвиг одного блока кристалла относительно другого.

Эта схема объясняет исключительную роль именно краевых дислокаций в процессе неупругого деформирования твердых тел. Поскольку существует силовой порог (сухое трение), который необходимо преодолеть для того, чтобы «сдвинуть» дислокацию, то движение дислокаций диссипативно, и деформирование вследствие движения даже одной дислокации является необратимым. При пластическом деформировании за счет процессов аннигиляции и образования новых дислокаций меняется собственная энергия ансамбля дислокаций, а за счет движения дислокаций – энергия их взаимодействия. Таким образом, кристалл с ансамблем взаимодействующих дислокаций представляет собой довольно сложную термодинамическую систему, которая содержит как источники энергии, так и каналы диссипации подводимой энергии деформирования.

В результате сложных внутренних процессов в пластически деформируемом кристалле образуются дислокационные структуры, которые в зависимости от условий нагружения проходят несколько стадий – от хаотической структуры клубков дислокаций разных знаков до трехмерной ячеистой структуры. Примером дислокационной структуры при циклическом деформировании является строение устойчивой полосы скольжения, содержащей двумерные структуры – мультипольные дислокационные стенки (рис. 7.16).

В реальных металлах с гранцентрированной кубической решеткой краевые дислокации, как правило, наблюдаются в сочетании с поверхностными дефектами упаковки. Длина вектора Бюргера краевой дислокации не может быть меньше межатомного расстояния, иначе после скольжения дислокации не будет восстановлена правильная структура кристалла. Если же вдоль части AA'

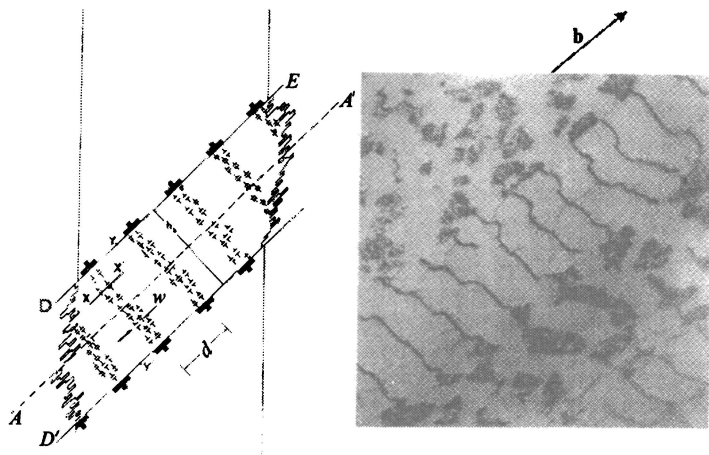


Рис. 7.16. Модель идеальной устойчивой полосы скольжения и микрофотография (просвечивающая электронная микроскопия) нерегулярной структуры с мультипольными стенками*

плоскости скольжения совершить сдвиг на вектор \mathbf{b}' , меньший вектора \mathbf{b} решетки, то линия A (рис. 7.17, *a*) будет границей области, по которой прошел сдвиг \mathbf{b} , а линия A' – границей сдвига \mathbf{b}' . На линии A получается скачок вектора сдвига $\mathbf{b}'' = \mathbf{b} - \mathbf{b}'$, и дислокация с вектором Бюргера \mathbf{b} расщепляется на две частичные дислокации с векторами Бюргера \mathbf{b}' и \mathbf{b}'' . В кристалле на полосе AA' нарушается правильная укладка структуры. Укладка атомов при этом может отличаться от равновесной, но существуют величины сдвигов \mathbf{b}' и \mathbf{b}'' , обеспечивающие минимум энергии в кристалле с локальным нарушением укладки. Такие поверхностные дефекты называются *дефектами упаковки*. Энергия дефекта упаковки является фундаментальной характеристикой кристалла, меняющейся от десятков до сотен миллиджоулей на квадратный метр. Соответственно изменяется ширина расщепленных дислокаций, об-

* Pedersen O.B. Fatigue and associated microstructural aspects; J. Lepinoux et al. (eds.). Multiscale phenomena in plasticity. Kluwer Academic Publishers, 2000. P. 83-97.

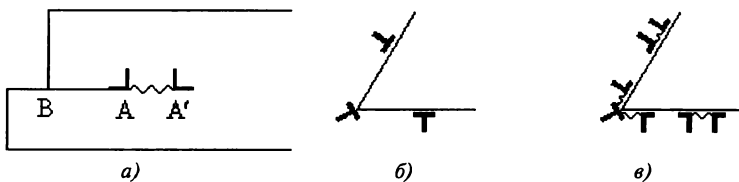


Рис. 7.17. Расщепленная дислокация (а) и реакции между дислокациями из пересекающихся плоскостей (б, в)

ратно пропорциональная энергии дефекта упаковки, от десятков до долей нанометров.

Дислокационными реакциями называются перестройки дефектной структуры кристалла, в результате которых одна или несколько дислокаций с векторами Бюргерса b_1, b_2, \dots превращаются в другие дислокации с векторами Бюргерса b'_1, b'_2, \dots . Полный вектор Бюргерса в результате реакций сохраняется. Примером реакции является аннигиляция дислокаций противоположных знаков.

Другой простейшей реакцией является расщепление полной дислокации на частичные. В кристалле с гранцентрированной кубической решеткой происходят простые реакции дислокаций, лежащих в плоскости (111), с векторами Бюргерса $b_1 = [110], b_2 = [101], b_3 = [011]$ (рис. 7.17,б), образующими равносторонний треугольник: $b_1 + b_2 \rightarrow b_3$. Если же в реакции участвуют расщепленные дислокации, то в результате реакции возникает прочный барьер Ломера-Коттрелла, блокирующий обе пересекающиеся плоскости скольжения (рис. 7.17,в). Эти барьеры играют большую роль в образовании прочных структур при множественном скольжении.

Плоская модель, описывающая самоорганизацию дислокаций с помощью техники клеточных автоматов, построена для дислокаций трех равнонаклоненных систем скольжения. Область моделирования, в дальнейшем называемая ячейкой, разбивается на большое число одинаковых клеток (сторона ячейки содержит несколько сотен клеток). Время «идет дискретными шагами». Рассматриваются прямолинейные краевые дислокации, бесконечно протяженные в третьем направлении. Все дислокации могут быть либо расщепленными, либо нет. Если дислокации расщеплены, то в результате дислокационных реакций появляются барьеры, в противном случае – нерасщепленные дислокации. «Переползание» расщепленных дислокаций невозможно.

На плоскости моделирования дислокации представлены своими следами. Клетка может быть занята дислокацией любой из трех систем скольжения любого знака, источником дислокаций или барьером (рис. 7.18), т.е. клетка должна обладать симметрией 3-го порядка. Для этого идеально подходят равносторонние шестиугольники (рис. 7.19,а). Учитывается дальнедействующее силовое взаимодействие всех дислокаций. Принимаются периодические граничные условия для взаимодействия и перемещения дислокаций, в силу последних отдельная дислокация в результате прямолинейного движения после нескольких итераций возвращается в исходное положение.

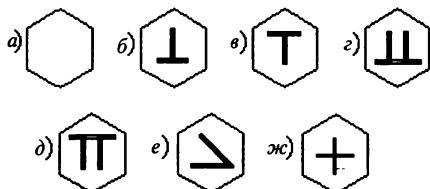


Рис. 7.18. Возможные состояния клетки: а – пустая клетка, б, в – положительная и отрицательная полные краевые дислокации, г, д – положительная и отрицательная расщепленные краевые дислокации, е – дислокационный барьер, перекрывающий соответствующие плоскости скольжения, ж – источник дислокаций из некоторой плоскости скольжения

Поскольку рассматриваются три равнонаклоненные системы скольжения, то при задании прямоугольной формы ячейки дислокации разных систем скольжения оказываются в различных условиях – дислокации горизонтальной и наклонных систем скольжения пройдут пути неодинаковой длины. Если ячейка имеет гексагональную форму, этого не происходит – разные системы скольжения оказываются равноправными. Периодичность граничных условий соответствует сворачиванию пространства в тор. Для того чтобы пространство оставалось плоским, при расчете взаимодействия учитывается несколько образов рассматриваемой ячейки – текущая ячейка представляется окруженной несколькими слоями ее образов. Для плотного покрытия плоскости ячейка и ее образы должны частично перекрываться (рис. 7.19,б). Соседние стороны ячейки периодичности (ее образы заполняют плоскость без перекрытия) содержат разное число клеток (рис. 7.19,в).

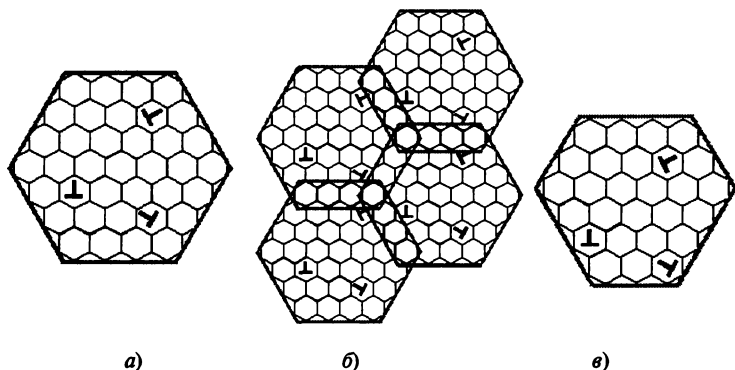


Рис. 7.19. Разбиение пространства: а – ячейка делится на гексагональные клетки, б – периодические граничные условия и разделение плоскости на одинаковые ячейки, в – форма ячейки периодичности

Дислокации разных систем взаимодействуют дальнедействующими полями согласно решению, полученному для линейной изотропной упругой среды:

$$F_{\tau} = \frac{mb^2}{2p(1-n)} \frac{x((x^2 - y^2) \cos(2\alpha) - 2xy \sin(2\alpha))}{(x^2 + y^2)^2},$$

$$F_n = -\frac{mb^2}{2p(1-n)} \frac{y(x^2 + y^2) + 2x^2y \cos(2\alpha) + x(x^2 - y^2) \sin(2\alpha)}{(x^2 + y^2)^2}, \quad (7.8)$$

где F_{τ} – сила, вызывающая скольжение; F_n – переползание дислокации с координатами $\{x, y\}$ из наклонной плоскости за счет взаимодействия с дислокацией из начала координат горизонтальной системы скольжения; α – угол наклона нормали наклонной системы к вертикали. Сила, действующая на выбранную дислокацию, определяется с помощью окрестности дальнего действия. Последняя имеет форму правильного шестиугольника (рис. 7.20) – в нее по-

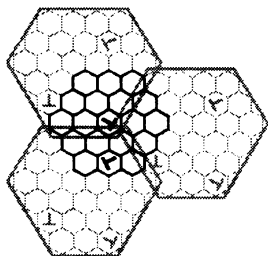


Рис. 7.20. Окрестность дальнего действия выбранной дислокации (показана темной линией в центре)

падают все рассматриваемые дислокации или их образы, кроме образов выбранной дислокации. Равнодействующая сила взаимодействия с дислокациями складывается с однородной приложенной нагрузкой. По найденным величинам напряжений и пороговому закону дислокациям приписывается состояние движения. Скольжение и переползание (рис. 7.21) дислокаций реализуются независимо в отдельных проходах по консервативной схеме, обеспечива-

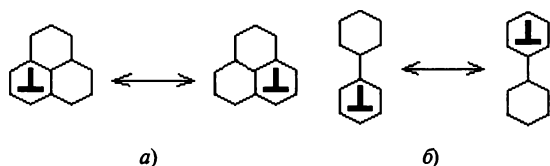


Рис. 7.21. Схема движения краевой дислокации произвольной плоскости залегания: а – скольжение, б – переползание (Правила читаются в зависимости от направления движения слева направо или справа налево)

ющей сохранение субстанции, с использованием локальной окрестности.

Скольжение дислокаций разных систем на каждой итерации происходит одновременно (параллельно) и реализуется с помощью трехклеточной локальной окрестности за четыре прохода (рис. 7.22). На каждом проходе перебираются поочередно все объекты (дислокации, источники, барьеры) и окружаются, в зависимости от но-

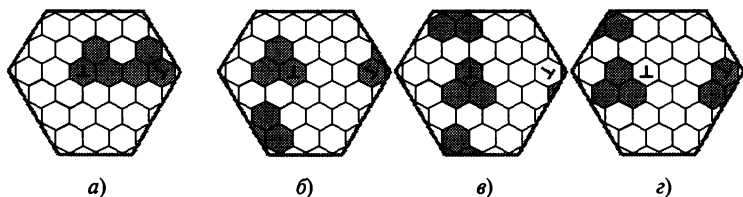


Рис. 7.22. Проходы, реализующие одновременное скольжение всех дислокаций

мера прохода, соответствующей окрестностью (темные клетки на рис. 7.22). Видно, что окрестности с разных проходов перекрываются – это необходимо для того, чтобы все дислокации со стату-

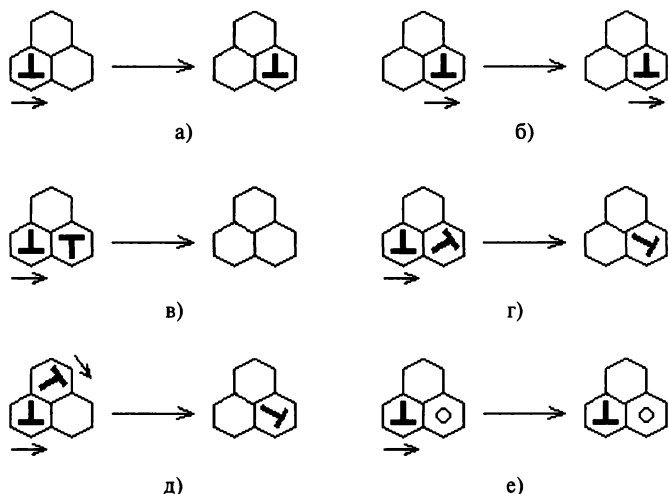


Рис. 7.23. Основные правила скольжения: а – свободное скольжение дислокации, б – отложенное скольжение, в – аннигиляция, г, д – реакция дислокаций (в случае расщепленных дислокаций образуется барьер соответствующей системы скольжения), е – остановка дислокации при встрече барьера, источника или дислокации, реакция с которой невозможна.

Правила инвариантны относительно одновременной смены знаков у объектов в клетках, отражения относительно вертикальной оси (движение в противоположном направлении) и поворота на $\pm 120^\circ$ (другие системы скольжения)

сом движения (например, рис. 7.23,б) могли сдвинуться после всех четырех проходов.

Второй проход отличается от первого смещением окрестностей на одну клетку по горизонтали, т.е. ситуация рис. 7.23,б первого прохода на втором будет соответствовать ситуации рис. 7.23,а. Третий и четвертый проходы (рис. 7.23,в и г) отличаются от первого и второго соответственно смещением на одну клетку по вертикали. Окрестности в течение каждого прохода не перекрываются, что обеспечивает консервативность субстанции при движении. Полагается, что реакция двух дислокаций разных систем происходит только в том случае, когда сумма векторов Бюргерса реагирующих дислокаций принадлежит третьей системе. При взаимодействии нерасщепленных дислокаций в результате реакции появля-

ется новая нерасщепленная дислокация, а в случае расщепленных — дислокационный барьер.

Реализация переползания проводится по очереди для разных систем скольжения с помощью двухклеточной окрестности в два (четный и нечетный) прохода (рис. 7.24), обеспечивающих консервативность движения. Темным цветом показан способ выбора окрестности для одной и той же дислокации для проходов разной четности. В расчетах перебираются все дислокации, и в зависимости от прохода и положения текущей дислокации выбирается окрестность. Аннигиляции и дислокационных реакций при переползании не происходит.

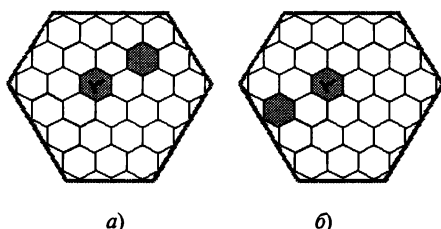


Рис. 7.24. Проходы, реализующие переползание дислокаций одной из систем

После реализации движения всех дислокаций управление передается блоку размножения дислокаций. Последнее происходит в однородно распределенных по области моделирования источниках Франка-Рида в том случае, когда результирующая сила в источнике превышает сумму порога активации источника и сопротивления Пайерлса tb . При этом в плоскости залегания источника справа и слева от него выбрасывается пара дислокаций противоположных знаков, принадлежащих той же системе скольжения: положитель-

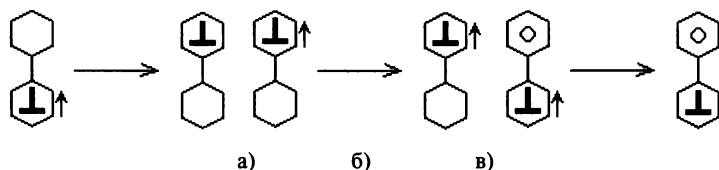


Рис. 7.25. Правила для переползания дислокаций:
а — свободное движение; б — отложенное движение; в — остановка

ная дислокация в направлении действия силы, отрицательная – в противоположном направлении. Положение и число источников в ходе процесса не меняются. При определении дальнедействующих напряжений источники воспринимаются как дислокации соответствующих систем скольжения. В ближкодействии по локальным правилам источники являются препятствиями для движения дислокаций.

После размножения перебираются дислокационные барьеры (если дислокации расщеплены) и проверяются усилия в них. Если усилие в барьере достаточно велико (например, от скопления, запертого барьером), частичные дислокации барьера из вторичных систем поджимаются к головной дислокации и барьер схлопывается в нерасщепленную дислокацию, которая может переместиться на следующей итерации движения, после чего она вновь расщепляется.

Скольжение отдельной дислокации из какой-либо клетки в соседнюю вызывает единичное приращение пластической деформации. В силу периодических граничных условий в результате скольжения одиночная дислокация, возвращаясь в свое начальное положение, пересекает прямоугольник, соответствующий ее плоскости скольжения (рис. 7.26).

Если рассматриваются дислокации только одной системы скольжения, то представление области моделирования шестиугольником эквивалентно представлению области соответствующим

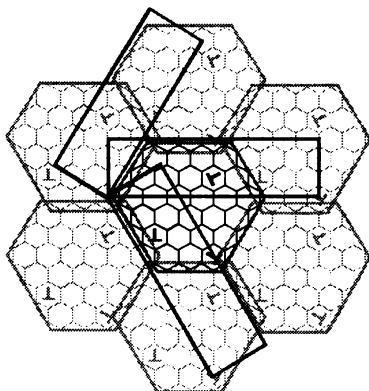


Рис. 7.26. Прямоугольные ячейки периодичности различных систем скольжения (изображены темными линиями)

этой системе прямоугольником (как это делается в динамической модели). Шестиугольная форма ячейки дает единое представление области для всех трех систем.

Пусть L – длина большей стороны шестиугольника, тогда длина и высота прямоугольника (рис. 7.26) суть $3(L-1)$ и $\sqrt{3}(L-1)/2$. Сдвиговая деформация от скольжения на одну клетку одной дислокации некоторой системы вычисляется как $\gamma = \frac{2}{3\sqrt{3}} \frac{lb}{(L-1)^2} = \frac{2}{3\sqrt{3}} \frac{1}{(N-1)^2} \frac{b}{l}$, где l – расстояние между центрами соседних клеток из одного слоя; N – число клеток на большей стороне шестиугольника. Тензор осредненных пластических деформаций

$$\epsilon^{pl} = \left[\sum_{k=1}^3 \gamma_k (\mathbf{m}_k \mathbf{n}_k + \mathbf{n}_k \mathbf{m}_k) \right] / 2, \quad (7.9)$$

где k – номер системы скольжения; \mathbf{m}_k – единичный вектор, направленный вдоль линии скольжения; \mathbf{n}_k – единичный вектор нормали к линии скольжения; γ_k – суммарный пластический сдвиг по k -й системе. Изменение объема в результате переползания дислокаций не учитывается.

Принимается гипотеза аддитивности однородных упругих и осредненных по области пластических деформаций $\epsilon = \epsilon^{el} + \epsilon^{pl}$. Это соответствует структурной схеме материала (см. разд. 4.3) в виде последовательного соединения двумерных упругого и пластического элементов. В плоском случае это можно представить заключением пластического элемента в упругую матрицу (при параллельном соединении пластический элемент помещается на упругую подложку). Упругий элемент моделирует кристаллическую решетку. Пластические деформации определяются модулем клеточных автоматов.

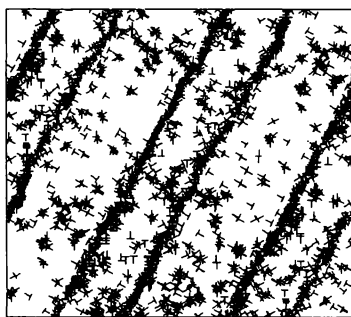
При мягком нагружении образца изменяются растягивающие и касательные компоненты $\sigma_{xx}(t)$, $\sigma_{yy}(t)$ и $\tau_{xy}(t)$ тензора напряжений. В соответствии со значениями напряжений на текущей итерации изменяется конфигурация дислокаций и пересчитывается суммарная (упруго-пластическая) деформация образца. При жестком нагружении накладываемое приращение деформаций представляется как $\Delta\epsilon = \Delta\epsilon^{el} + \Delta\epsilon^{pl}$, т.е. изменение деформаций кристаллической решетки $\Delta\epsilon^{el} = \Delta\epsilon + \Delta\epsilon^{pl}$. Изменение отклика, т.е. напряжения, снимаемого в упругом элементе после изменения

пластической составляющей полной деформации вследствие движения дислокаций, определяется итерационно из самосогласованного уравнения

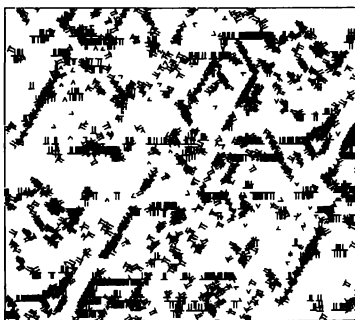
$$\Delta\sigma = \Pi : \Delta\epsilon^{el} = \Pi : (\Delta\epsilon - \Delta\epsilon^{pl}(\Delta\sigma)), \quad (7.10)$$

где тензор четвертого ранга Π определяет линейно-упругие свойства решетки. Решая итерационно уравнение (7.10), можно построить диаграмму деформирования при нагружении по любой двумерной траектории в пространстве деформаций. Приращение пластической составляющей полной деформации определяется по текущей итерации в модели клеточных автоматов.

Вычислительный эксперимент позволил описать образование и развитие дислокационной структуры (микроуровень) в материале и одновременно построить диаграмму деформирования этого материала (макроуровень). При нагружении по окружности в пространстве деформаций (непропорциональное циклическое деформирование) материала с высокой энергией дефекта упаковки, которой соответствуют нерасщепленные дислокации, после трех циклов была получена дислокационная структура, состоящая из наклонных мультипольных стенок (рис. 7.27, а). Для низкой величины энергии дефекта упаковки структура состоит из многочисленных планарных скоплений дислокаций, запертых дислокационными барьерами (рис. 7.27, б). На диаграммах нагружения мо-



а)



б)

Рис. 7.27. Дислокационные структуры при непропорциональном деформировании в материале с высокой (а) и низкой (б) энергией дефекта упаковки

менты образования и перестройки дислокационной структуры сопровождаются локальными падениями напряжений (временная нестабильность), появляющимися при небольших скоростях деформирования (рис. 7.28). Эти падения наблюдаются также в достаточно тонких натуральных экспериментах и известны под названием эффекта Портевена-Ле Шателье.

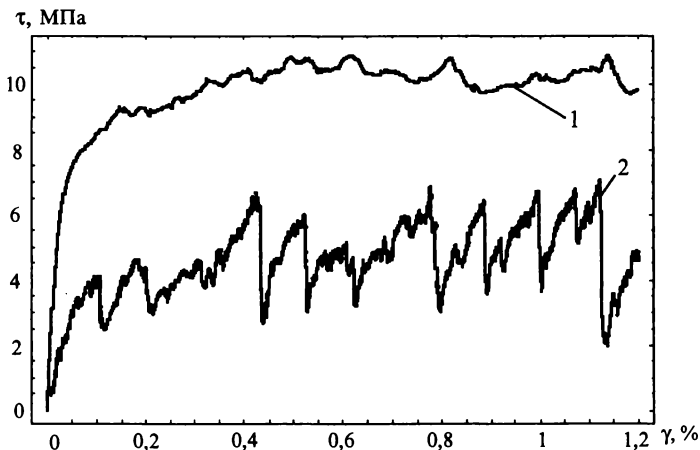


Рис. 7.28. Диаграммы деформирования с относительно высокой (кривая 1) и низкой (кривая 2) скоростями деформирования

При деформировании по пропорциональной (рис. 7.29) и непропорциональной (рис. 7.30) циклическим траекториям в осях «сдвиговая деформация – растягивающая деформация» в пространстве напряжений наблюдается расширение петли, свидетельствующее об упрочнении материала. Последнее связано с образованием в материале дислокационной структуры. В случае низкой энергии дефекта упаковки (все дислокации расщеплены) заметное упрочнение наблюдается уже в первой части цикла (рис. 7.30). Более сильное упрочнение в кристаллах с низкой энергией дефекта упаковки вызвано сильными дислокационными реакциями (образованием барьеров Ломера–Коттрелла) при геометрически необходимом взаимодействии определенных систем скольжения в ходе непропорционального нагружения.

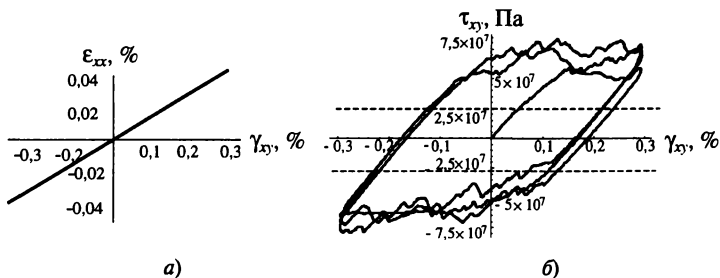


Рис. 7.29. Пропорциональное циклическое деформирование: а – вид траектории в пространстве деформаций, б – диаграмма деформирования (пунктирной линией отмечен порог Пайерлса; дислокации не расщеплены)

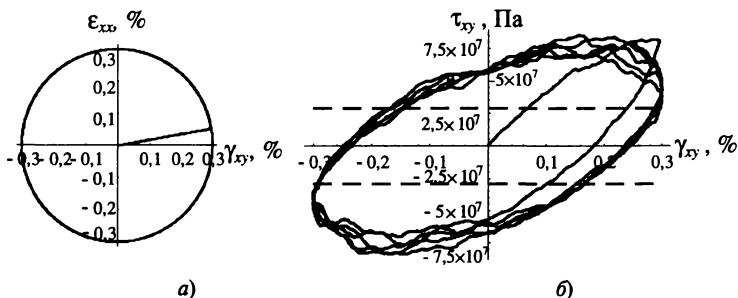


Рис. 7.30. Непропорциональное деформирование: а – вид траектории в пространстве деформаций; б – диаграмма деформирования (штриховой линией отмечен порог Пайерлса; дислокации расщеплены)

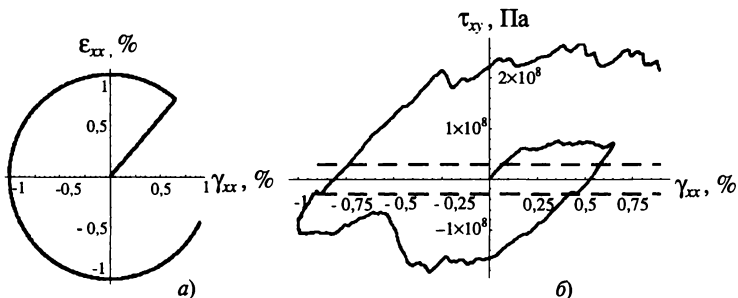


Рис. 7.31. Непропорциональное деформирование: а – вид траектории в пространстве деформаций, б – диаграмма деформирования (пунктирной линией отмечен порог Пайерлса; дислокации расщеплены)



Вопросы для самопроверки

1. Каким образом соотносятся между собой реальное, системное и модельное время?
2. Может ли системное время быть не равно реальному времени? Почему?
3. В каких случаях обосновано применение имитаторов?
4. В чем принципиальные отличия аналитических моделей и имитаторов?
5. В чем схожесть реального и вычислительного с использованием имитатора экспериментов?
6. В чем достоинства и недостатки моделирования системного времени с постоянным и переменным шагами?
7. Назовите отличия технологии создания имитаторов от аналитических моделей.
8. Почему имитаторы можно отнести к разновидности математических моделей?
9. В каких случаях имитатор СМО является предпочтительным по отношению к аналитическому подходу?



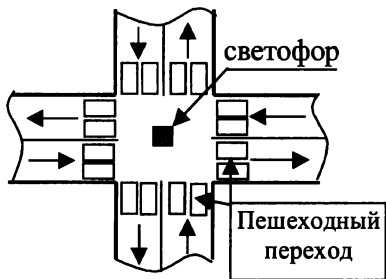
Задания для самостоятельного выполнения

1. Разработайте имитатор одноканальной СМО с ограниченным временем ожидания в очереди. Предполагается, что поток заявок, поступающий в систему, является простейшим. Поступившая в систему заявка может сразу поступить на обслуживание, если канал свободен, либо встать в очередь. Если время пребывания заявки в очереди превысило некоторое предельное для нее значение, она покидает систему необслуженной. Интервал времени Δt между поступлением двух соседних заявок, максимальное время пребывания заявки в очереди Δh и время ее выполнения Δw описываются случайными величинами, подчиненными показательному распределению с заданными средними значениями. Имитатор должен выполнять оценку следующих параметров эффективности СМО:

- среднего времени пребывания заявки в очереди;
- вероятности выполнения заявок;
- коэффициента загрузки системы.

2. Аналогично заданию 1 разработайте имитатор n -канальной СМО с ограниченным временем пребывания в очереди.

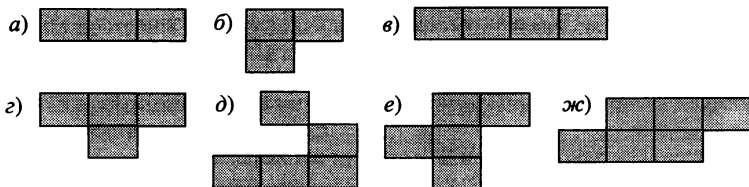
3. Разработайте имитатор регулируемого светофором перекрестка дороги. Движение в одном направлении считайте однополосным. При движении через перекресток разрешено движение прямо, повороты налево



и направо. Перекресток представьте как многоканальную систему с очередями. Поток машин, подходящих к перекрестку, считайте простейшим. При пересечении перекрестка интервал времени между машинами следует принять случайной величиной, подчиненной равномерному закону распределения. Исходные данные получите в результате наблюдения за реальным перекрестком.

При проведении наблюдений следует в журнал наблюдений записать моменты времени подъезда автомобиля к перекрестку, его въезда на перекресток, направление и момент выезда с перекрестка. Для получения достоверных оценок параметров проследите за проездом перекрестка не менее 100 автомобилями.

4. Разработайте программу, реализующую клеточный автомат «Жизнь». Состояние клеточного пространства выведите на экран в графическом режиме. Исследуйте эволюцию КА для следующих начальных состояний, задающих расположение «живых» клеток:



5. Разработайте клеточный автомат «Лишайники», поведение которого подчинено следующим правилам:

- клетка может находиться в активном или пассивном состоянии;
- клетка становится активной, если в восьми соседних клетках находится N_1 , N_2 или N_3 активных клеток;
- если число активных клеток в окрестности не равно N_1 , N_2 или N_3 , то клетка становится пассивной.

6. Разработайте клеточный автомат «Дюны», поведение которого подчинено следующим правилам:

- клетка может находиться в активном и пассивном («спрятанном») состоянии;
- если клетка была активна и из восьми соседних клеток более N активны, то клетка «прячется». Время нахождения в «спрятанном» состоянии равно W тактов;

➤ если время «прятания» закончилось и в окрестности не более M активных клеток, то клетка вновь становится активной.

7. Разработайте клеточный автомат «Термит», поведение которого подчинено следующим правилам:

- клетка может находиться в пассивном или активном состоянии. В начальный момент все ячейки пассивны, «Термит» расположен в центральной клетке и направлен вверх;
- автомат «Термит» переходит на соседнюю клетку, и если она активная, то делает ее пассивной и поворачивает налево на 90° . Если клетка была пассивна, «Термит» делает ее активной и поворачивает направо на 90° .

Приложение 1

ЯЗЫК ФОРМАЛЬНОГО ОПИСАНИЯ АЛГОРИТМОВ

Метод решения любой задачи можно представить в виде некоторого алгоритма, понимаемого как последовательность действий, в результате которых достигается необходимый результат. Сам алгоритм может быть выражен различными способами. Исходя из требований возможной простоты, удобочитаемости и логической ясности естественно в качестве описания алгоритма выбрать такой способ, который был бы сопоставим с описанием на языке программирования и содержал конструкции естественного языка. При этом важно, чтобы описание алгоритма проводилось в соответствии с общими принципами и правилами структурного программирования. Такому требованию удовлетворяет язык проектирования программ PDL, русская версия которого носит название псевдокода. Хотя язык PDL имеет неформальный характер, т.е. не рассчитан на трансляцию, тем не менее перевод записи алгоритма с языка PDL на язык программирования не составляет труда. Ниже приводится подробное описание всех соглашений.

Комментарии. Набираются курсивом. В начале алгоритма помещаются комментарии, кратко описывающие назначение алгоритма и определяющие используемые переменные.

Константы. Все константы считаются десятичными числами, если в комментариях не оговорено иное. Например, все числа 3; -5; 3,14159; $1,5 \cdot 10^{-3}$ являются константами.

Переменные. Переменная – это имя, используемое для хранения значения, которое может изменяться. Имена переменных для большей наглядности могут содержать как прописные, так и строчные символы латинского алфавита. Например, *Pi*, *Fi*, *x*, *y*.

Массивы. Массив – имя для индексированной совокупности величин. Для имен массивов действуют те же соглашения, что и для переменных. Обращение ко всему массиву осуществляется по его имени, а к индивидуальным элементам – по имени массива с указанием списка индексов. Индексы могут принимать как положительные, так и отрицательные значения. При описании массива в квадратных скобках указываются минимальные и максимальные значения каждого индекса.

Пример: A_{25} ; B_{350} — обращение к элементам массива;
 A ; B — обращение ко всему массиву;
 $A[-1 \dots 5, 0 \dots 10]$ — описание массива.

Оператор присваивания. Знак « $=$ » используется для присвоения переменной, расположенной в левой части оператора, значения выражения из правой части.

Арифметические, логические операции и операции отношений. Арифметические операции над целыми числами:

+ сложение; — вычитание;
· умножение; / деление;

div — вычисление целой части числа (например, $7 \text{ div } 2 = 3$);

mod — вычисление остатка от деления (например, $7 \text{ mod } 2 = 1$).

Арифметические операции над вещественными числами:

+ сложение; — вычитание;
· умножение; / деление.

Логические операции **and** и **or**, как это здесь показано, выделяются жирным шрифтом. Операции отношений:

= равно; \neq не равно;
< меньше; > больше;
 \leq меньше или равно; \geq больше или равно.

Пример: $(a < 5)$ **and** $(a \neq 0)$

Операторы start и stop. Используются для обозначения начала и конца работы алгоритма.

Составной оператор. Может содержать произвольное число операторов и выделяется операторными скобками:

```
begin  
  [ произвольное число операторов ]  
end
```

Условный оператор. Условное выполнение оператора может быть представлено в двух формах:

```
if (условие) then  
  [операторы, выполняемые, если условие истинно]  
else  
  [операторы, выполняемые, если условие ложно]  
end if
```

или

```
if (условие) then  
  [операторы, выполняемые, если условие истинно]  
end if
```

Пример: **if** $(a < 0)$ **then** $x := x + 1$
 else $x := x - 1$
 end if

Оператор выбора. Имеет следующий вид:

```
case x of
  x1 : [операторы, выполняемые, если x = x1 ]
  x2,x3,x4 : [операторы, выполняемые, если x = x2, x3 или x4 ]
  x5 ... x6 : [операторы, выполняемые, если x5 ≤ x ≤ x6 ]
  . . .
  xn : [операторы, выполняемые, если x = xn ]
else [операторы, выполняемые, если x не совпадает ни с одной
      из меток ]
end case
```

Пример:

```
case i of
  1 : x:=x+1      выполняется, если i = 1
  2,4,6 : x:=1    выполняется, если i = 2, 4 или 6
  10...15 : x:=0  выполняется, если 10 ≤ i ≤ 15
end case
```

Оператор цикла с параметром:

```
for (переменная цикла) = (нач.значение) to (кон.значение) step (шаг
цикла)
  [выполняемые операторы]
next (переменная цикла)
```

Пример:

```
for i = -5 to 9 step 2
  x := x+i
next i
```

Оператор цикла с предусловием.

```
while (условие)
  [выполняемые операторы]
end while
```

Цикл выполняется, пока условие истинно.

Пример:

```
while (i<5)
  x := x+2
  i := i+1
end while
```

Информационные сообщения и сообщения об ошибке. Для вывода сообщений используются функции **Error** (<пояснение>) и **Attention** (<текст сообщения>)

Процедуры. Отдельные блоки алгоритма удобно оформлять в виде самостоятельных процедур:

```
procedure <название процедуры>
Данные: . . .
Результат: . . .
start
. . .
return
```

Вызов процедуры выполняется с помощью следующего оператора:
do <название процедуры>

Можно использовать предопределенные вспомогательные процедуры выхода:

- halt** – завершение программы;
- exit** – завершение процедуры;
- break** – завершение цикла.

В качестве примера приведем запись на псевдокоде алгоритма вычисления корней квадратного уравнения.

Алгоритм П1.1.

procedure Определение корней квадратного уравнения.

Данные: A, B, C – коэффициенты уравнения $Ax^2 + Bx + C = 0$.

Результаты: x_1 и x_2 – значение корней уравнения.

start

$x_1 := 0$

$x_2 := 0$

if ($|A| > 10^{-5}$) **then**

$D := B^2 - 4AC$

if ($D \geq 0$) **then** $x_{1,2} := \frac{-B \pm \sqrt{D}}{2A}$

else Attention(*Нет действительных корней*)

end if

end if

return

Приложение 2

ЧИСЛЕННЫЕ МЕТОДЫ (минимальные сведения)

Раздел математики, в котором разработаны способы приближенного решения различных задач (решения уравнений, интегрирование, дифференцирование, нахождение экстремумов функций и т.д.) не алгебраическим путем и не методами математического анализа, а путем многократных арифметических действий над числами, называется «Численные методы», или «Вычислительная математика». Начало развития численных методов решения задач относится ко временам Ньютона. Однако лишь с появлением вычислительной техники эти методы нашли широкое применение.

Ниже описываются только самые простые методы, неэффективные в отношении скорости вычислений, дающие сравнительно грубые результаты, но зато доступные для понимания. С более совершенными методами можно познакомиться в курсах и пособиях по вычислительной математике.

П2.1. Решение уравнений высоких степеней и трансцендентных уравнений с одним неизвестным

Очень часто при исследовании математических моделей приходится сталкиваться с необходимостью решать уравнения с одной неизвестной. Так, для линейного уравнения

$$ax + b = 0 \quad (\text{П2.1})$$

все сводится к нахождению x по формуле $x = -b/a$. В случае квадратного уравнения – к нахождению x_1 и x_2 по Алгоритму П1.1 из приложения 1. Кубическое уравнение

$$ax^3 + bx^2 + cx + d = 0 \quad (\text{П2.2})$$

всегда может быть приведено к виду

$$y^3 + py + q = 0 \quad (\text{П2.3})$$

путем деления его на a и подстановкой $x = y - b/(3a)$. Корни уравнения (П2.3) могут быть найдены по формуле Кардано:

$$y = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}. \quad (\text{П2.4})$$

Преобразование (П2.2) к виду (П2.3) и нахождение корней с помощью (П2.4) довольно громоздко, поэтому применение вычислительной техники здесь уже вполне рационально. Не менее сложным является решение уравнения четвертой степени, если только оно не сводится к биквадратному. Формул для решения уравнений пятой степени и выше не существует. Уравнения, в которых неизвестное встречается не только в какой-либо степени, но и под знаком других функций ($\sin x$, $\lg x$, e^x и т.п.), называются *трансцендентными*. Кроме самых простых случаев для нахождения корней этих уравнений формул не существует.

Численные методы в совокупности с ЭВМ позволяют решать сравнительно просто как алгебраические уравнения произвольных степеней, так и трансцендентные. Любое уравнение с одним неизвестным может быть приведено к виду

$$F(x) = 0. \quad (\text{П2.5})$$

Корнем такого уравнения будет значение x , при котором $F(x)$ проходит через ноль (рис. П2.1). Разобьем решение уравнения на два этапа: изоляцию корней и уточнение значения изолированного корня.

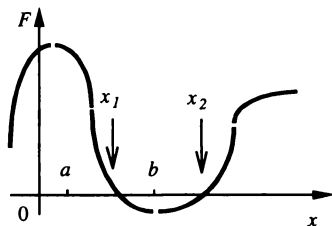


Рис. П2.1. Корни уравнения

Изоляция корней означает определение такого отрезка $[a, b]$, где есть корень, притом только один. Эта задача решается либо путем анализа функции $F(x)$, либо построением таблицы значений функции при различных значениях аргумента (выполнение таблицы функции). Далее построенный по полученной таблице график функции $F(x)$ удобно использовать для приближенного вычисления корней уравнения. Как табуляцию, так и построение графика функции целесообразно выполнять с использованием электронных таблиц (например, электронной таблицы *MS Excel*), не прибегая к написанию собственных программ.

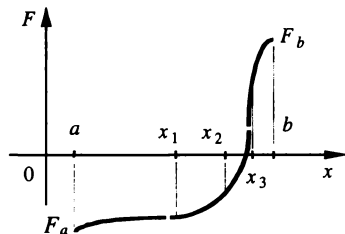


Рис. П2.2. Иллюстрация к методу деления отрезка пополам

Для уточнения значения корня уравнения после его изоляции существуют различные методы. Простейшим является *метод деления отрезка пополам*. Идея метода заключается в следующем. На первом шаге вычисляются значения функции на концах отрезка (рис. П2.2). Затем исходный отрезок $[a, b]$ делят пополам (точка x_1) и вычисляют значение функции в этой точке. Из двух отрезков выбирается тот, для которого значения функции на концах отрезка имеют противоположные знаки (отрезок $[x_1, b]$ на рис. П2.2). Выбранный отрезок снова делят пополам, вновь вычисляют значение функции в полученной точке и осуществляют выбор отрезка. Процесс продолжается до тех пор, пока модуль значения функции в средней точке очередного выбранного отрезка не станет меньше некоторого малого положительного ϵ . Ниже приведен возможный алгоритм данного метода.

Алгоритм П2.1

Procedure Уточнение корней уравнения методом деления пополам.

Данные: $F(x)$ – исследуемая функция;

a, b – исходный исследуемый отрезок; ϵ – точность вычислений.

Результаты: x – значение корня уравнения.

start

$x := a$

if $(F(a) \cdot F(b) < 0)$ **then**

while $(|F(x)| > \epsilon)$ **do**

$x := (a+b)/2$

if $(F(x) \cdot F(a) < 0)$ **then** $b := x$ *новый отрезок* $[a, (a+b)/2]$

else $a := x$ *новый отрезок* $[(a+b)/2, b]$

end if

end while

else **Attention**(*Отрезок задан неверно. Корень не изолирован.*)

return

Более эффективным методом уточнения корней уравнения является *метод хорд*. В отличие от предыдущего метода исследуемый отрезок разбивается на два новых точкой пересечения оси абсцисс с линией,

проходящей через точки A и B графика функции $F(x)$, соответствующие концам отрезка (рис. П2.3). Ниже приведен алгоритм метода хорд.

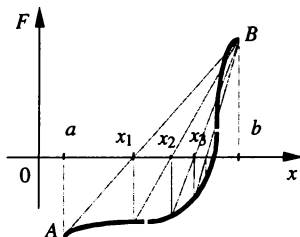


Рис. П2.3. Схема к методу хорд

Алгоритм П2.2

Procedure Уточнение корней уравнения методом хорд.

Данные: $F(x)$ – исследуемая функция; ϵ – точность вычислений.

a, b – исходный исследуемый отрезок, $a < b$.

Результаты: x – значение корня уравнения.

start

$x := a$

if $(F(a) \cdot F(b) < 0)$ then

while $(|F(x)| > \epsilon)$ do

$x := a + (b - a) * |F(a)/(F(b) - F(a))|$

if $(F(x) \cdot F(a) < 0)$ then $b := x$ *новый отрезок [a, x]*
 else $a := x$ *новый отрезок [x, b]*

end if

end while

else Attention(*Отрезок задан неверно. Корень не изолирован.*)

return

П2.2. Решение систем линейных уравнений

Многие задачи математического моделирования сводятся к системе n линейных алгебраических уравнений:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2, \\ \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n. \end{cases} \quad (\text{П2.6})$$

Эту систему можно записать в матричной форме:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{Bmatrix} = \begin{Bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{Bmatrix} \quad (\text{П2.7})$$

или

$$[\mathbf{A}]\{\mathbf{X}\} = \{\mathbf{B}\}. \quad (\text{П2.8})$$

При исследовании многих моделей, связанных, например, с решением задач механики сплошных сред в объемной постановке или задач из области экономики, приходится решать системы, состоящие из нескольких тысяч или десятков тысяч уравнений. В этом случае особое внимание приходится уделять вопросам повышения эффективности и точности применяемых методов решения.

Все методы решения систем линейных уравнений делят на прямые и итерационные.

Прямые методы позволяют за конечное число действий найти точное решение системы уравнений (П2.6), если входная информация (правая часть уравнений $\{\mathbf{B}\}$ и элементы a_{ij} матрицы $[\mathbf{A}]$) заданы точно и вычисления ведутся без округления. Конечно, прямые методы также дают решение с определенной точностью, которая зависит от ошибок округления, т.е. от точности представлений данных в ЭВМ, характера решаемых задач и самого метода решения системы уравнений. К наиболее известным представителям прямых методов можно отнести методы Гаусса и квадратного корня.

Итерационный метод позволяет найти приближенное решение системы путем построения последовательности приближений (итераций), начиная с некоторого начального приближения. К этому классу методов относят методы итераций и Зейделя. При выборе того или иного метода исходят из мощности имеющейся вычислительной техники и требуемой точности получаемого решения. Для больших систем (более 10 тыс. уравнений), как правило, используют итерационные методы.

В качестве примера рассмотрим алгоритм наиболее часто применяемого метода – *метода Гаусса*. Как уже отмечалось, данный метод относится к прямым и позволяет найти решение за конечное число шагов. Алгоритм метода состоит из двух этапов, или ходов: прямого и обратного. При прямом ходе система уравнений приводится к треугольному виду.

Пусть требуется решить следующую систему уравнений:

$$2x_1 + 8x_2 + 4x_3 = 4, \quad (\text{П2.9a})$$

$$4x_1 + x_2 - 2x_3 = -2, \quad (\text{П2.9б})$$

$$8x_1 + 12x_2 + 8x_3 = 12. \quad (\text{П2.9в})$$

Прямой ход. Первый шаг. Исключаем из всех уравнений, кроме первого, неизвестное x_1 . Разделим первое уравнение (П2.9а) на $a_{11} = 2$:

$$x_1 + 4x_2 + 2x_3 = 2. \quad (\text{П2.10})$$

Для исключения x_1 из уравнения (П2.9б), умножим (П2.10) на -4 и сложим с (П2.9б). Аналогично, умножая (П2.10) на -8 и складывая с (П2.9в), исключим из него x_1 :

$$-15x_2 - 10x_3 = -10, \quad (\text{П2.11а})$$

$$-20x_2 - 8x_3 = -4. \quad (\text{П2.11б})$$

В результате имеем систему уравнений второго порядка.

Второй шаг. Исключим x_2 из всех уравнений, кроме первого и второго. Для этого разделим (П2.11а) на -15 :

$$x_2 + 0,67x_3 = 0,67. \quad (\text{П2.12})$$

Умножим (П2.12) на 20 и сложим с (П2.11б):

$$5,33x_3 = 9,33. \quad (\text{П2.13})$$

Третий шаг. Делим (П2.13) на 5,33, в результате имеем $x_3 = 1,75$.

Итак получаем систему:

$$\begin{aligned} x_1 + 4x_2 + 2x_3 &= 2, \\ x_2 + 0,67x_3 &= 0,67, \\ x_3 &= 1,75 \end{aligned} \quad (\text{П2.14})$$

с верхней треугольной матрицей $\begin{bmatrix} 1 & 4 & 2 \\ 0 & 1 & 0,67 \\ 0 & 0 & 1 \end{bmatrix}$.

Обратный ход. Из системы (П2.14) последовательно находим:

3-е уравнение: $x_3 = 1,75$;

2-е уравнение: $x_2 = 0,67 - 0,67x_3 = -0,5$;

1-е уравнение: $x_1 = 2 - 4x_2 - 2x_3 = 0,5$.

Решение системы уравнений (П2.9) найдено:

$$x_1 = 0,5, \quad x_2 = -0,5, \quad x_3 = 1,75.$$

Ниже приведена запись алгоритма метода Гаусса на псевдокоде для системы из n уравнений с n неизвестными.

Алгоритм П2.3.

Procedure Метод Гаусса для системы уравнений $[A]\{X\}=\{B\}$.

Данные: n – число неизвестных;

$A[1\dots n, 1\dots n]$ – квадратная матрица коэффициентов;

$B[1\dots n]$ – правая часть системы уравнений.

Результат: $B[1\dots n]$ – решение системы.

start

for $k:=1$ **to** $n-1$ <= прямой ход

if $|A_{kk}| < 10^{-20}$ **then**

Error (*Элемент главной диагонали равен 0*)

Exit

end if

$B_k := B_k / A_{kk}$ *обработка k-й строки*

for $i := k+1$ **to**

$A_{ki} := A_{ki} / A_{kk}$

next i

for $j := k+1$ **to** n *обработка всех нижележащих строк*

if $|A_{jk}| > 10^{-20}$ **then**

$B_j := B_j - B_k \cdot A_{jk}$

for $i := k+1$ **to** n

$A_{ji} := A_{ji} - A_{jk} \cdot A_{ki}$

next i

$A_{jk} := 0$

end if

next j

next k

$B_n := B_n / A_{nn}$

for $j := n-1$ **to** 1 **step** -1

<= обратный ход

for $i := n$ **to** $j+1$ **step** -1

$B_j := B_j - A_{ji} \cdot B_i$

next i

next j

return

П2.3. Решение обыкновенных дифференциальных уравнений

Математическая постановка многих задач математического моделирования сводится к системе обыкновенных дифференциальных уравнений (ОДУ). Простейшее ОДУ – уравнение первого порядка

$$\frac{dy}{dx} = f(x, y). \quad (\text{П2.15})$$

Решение этих уравнений заключается в нахождении таких функций $y(x)$, которые при подстановке обращают уравнение в тождество. Например, уравнение $dy/dx = ax + b$ имеет решение $y = ax^2/2 + bx + c$.

Если дано начальное условие, например $y|_{x=0} = y_0$, то можно найти значение константы c : $c = y_0$.

Как правило, только в относительно простых случаях решение ОДУ можно получить аналитически, т.е. в виде формул. Численные методы позволяют найти решение *любого* ОДУ, однако ответ получается не в виде аналитической зависимости, а в виде таблицы:

X	X_1	X_2	X_3	...
Y	Y_1	Y_2	Y_3	...

При этом обязательно задание начальных условий. Одним из простейших методов численного интегрирования ОДУ является метод Эйлера.

Метод Эйлера. Пусть дифференциальное уравнение приведено к виду (П2.15), а вид функции $f(x, y)$ известен. Заменим приближенно дифференциалы приращениями:

$$\frac{\Delta y}{\Delta x} = f(x, y), \text{ или } \Delta y = f(x, y)\Delta x. \quad (\text{П2.16})$$

Зададим начальные условия $y|_{x=0} = y_0$ и выберем шаг Δx приращения аргумента. Тогда, используя (П2.16), можно последовательно вычислить значения x и y :

x_0	$x_1 = x_0 + \Delta x$	$x_2 = x_1 + \Delta x$	$x_3 = x_2 + \Delta x$...
y_0	$y_1 = y_0 + f(x_0, y_0) \Delta x$	$y_2 = y_1 + f(x_1, y_1) \Delta x$	$y_3 = y_2 + f(x_2, y_2) \Delta x$...

или

$$x_{k+1} = x_k + \Delta x; \quad y_{k+1} = y_k + f(x_k, y_k)\Delta x. \quad (\text{П2.17})$$

С помощью аналитического метода получаем решение в виде гладкой интегральной кривой уравнения (П2.15) (рис. П2.4).

Использование метода Эйлера позволяет получить последовательность точек, соединив которые можно построить *ломаную Эйлера*. С умень-

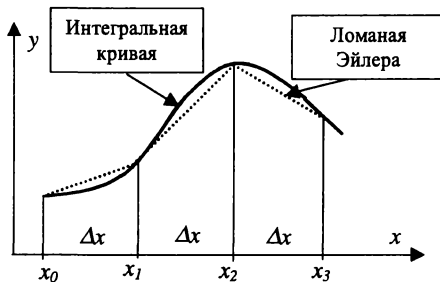


Рис. П2.4

шением шага интегрирования Δx кривая Эйлера приближается к интегральной кривой.

Для метода Эйлера характерны малая точность вычислений и систематическое накопление ошибок. Пример алгоритма, использующего метод Эйлера, для системы уравнений первого порядка приведен в гл. 2.

Примечание. Задачи динамики точки, как правило, приводят к системе двух ОДУ первого порядка по каждой координате. При вычислении новых скоростей точки значения действующих сил берутся на момент начала каждого шага и не учитывается их изменение за шаг. Таким образом, принимается, что на шаге по времени точка движется под действием постоянных сил и с постоянным ускорением.

Это обстоятельство можно учесть при вычислении новой координаты точки. Если уравнение для координат решать после уравнения для скоростей, то в этом случае можно воспользоваться значениями скоростей точки на начало и на конец шага по времени. При равноускоренном движении пройденный путь равен произведению величины интервала времени на среднюю скорость точки на этом интервале. С учетом указанного обстоятельства соотношения для скорости и координаты точки могут быть переписаны следующим образом:

$$v_{k+1} = v_k + \Delta t F(t, x_k, v_k);$$

$$x_{k+1} = x_k + \Delta t (v_{k+1} + v_k) / 2.$$

Приведенная схема вычисления соответствует модифицированному методу Эйлера и дает более точные результаты интегрирования.

Метод Рунге-Кутты. Пусть дифференциальное уравнение приведено к виду (П2.15), а вид функции $f(x, y)$ известен. Выберем шаг интегрирования Δx и для краткости введем обозначения $x_k = x_0 + k\Delta x$ и $y_k = y(x_k)$, $k = 0, 1, 2, 3, \dots$

Рассмотрим числа:

$$\begin{aligned} p_1^{(k)} &= f(x_k, y_k)\Delta x, \\ p_2^{(k)} &= f(x_k + \Delta x/2, y_k + p_1^{(k)}/2)\Delta x, \\ p_3^{(k)} &= f(x_k + \Delta x/2, y_k + p_3^{(k)}/2)\Delta x, \\ p_4^{(k)} &= f(x_k + \Delta x, y_k + p_3^{(k)})\Delta x. \end{aligned} \quad (\text{П2.18})$$

Согласно обычному методу Рунге-Кутты [36], последовательность значений y_k искомой функции y определяется по формуле:

$$y_{k+1} = y_k + \Delta y_k,$$

где

$$\Delta y_k = (p_1^{(k)} + 2p_2^{(k)} + 2p_3^{(k)} + p_4^{(k)})/6, \quad k = 0, 1, 2, 3, \dots \quad (\text{П2.19})$$

Метод Рунге-Кутты обладает значительной точностью и несмотря на свою трудоемкость широко используется при численном решении ОДУ.

Основная сложность при использовании численных методов интегрирования ОДУ – выбор величины шага интегрирования Δx . На практике величину шага Δx выбирают двойным пересчетом: вначале определяют y_k с шагом Δx , а затем с шагом $\Delta x/2$. Если расхождение результатов превысило некоторую малую заданную величину ϵ , то шаг уменьшают еще в 2 раза и повторяют вычисления. Уменьшение производят до тех пор, пока не будет достигнута заданная точность вычислений.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Абловиц М., Сигур Х.* Солитоны и метод обратной задачи. – М.: Мир, 1987. 479 с.
2. *Айвазян С.А., Енюков И.С., Мешалкин Л.Д.* Прикладная статистика: Основы моделирования и первичная обработка данных. – М.: Финансы и статистика, 1983. 471 с.
3. *Александров А.В., Лащенко Б.Я., Шапошников И.И.* Строительная механика. – М.: Стройиздат, 1983. 488с.
4. *Алефельд Г., Херцбергер Ю.* Введение в интервальные вычисления. – М.: Мир, 1987. 358 с.
5. *Алкок Д.* Язык Паскаль в иллюстрациях. – М.: Мир, 1991. 192 с.
6. *Арнольд В.И.* Лекции об уравнениях с частными производными. – М.: Фазис, 1997. 180 с.
7. *Баренблатт Г.И.* Подобие, автомодельность, промежуточная асимптотика, – Л.: Гидрометеиздат, 1982. 255 с.
8. *Бахвалов Н.С., Жидков Н.П., Кобельков Г.М.* Численные методы: учебное пособие. – М.: Наука, 1987. 600 с.
9. *Белл Дж.Ф.* Экспериментальные основы механики деформируемых тел: В 2 ч. Ч. 1. Малые деформации. – М.: Наука, Физматлит, 1984. 600 с.
10. *Белоцерковский О.М., Давыдов Ю.М.* Метод крупных частиц в газовой динамике. – М.: Наука, 1982. 392 с.
11. *Бендат Дж., Пирсол А.* Прикладной анализ случайных данных – М.: Мир, 1989. 540 с.
12. *Биркгоф Г.* Математика и психология. 95 с. М.: Советское радио, 1977.
13. *Блехман И.И., Мышкис А.Д., Пановко Я.Г.* Прикладная математика: предмет, логика, особенности подходов. – Киев: Наукова думка, 1976. 270 с.
14. *Боровков А.А.* Теория вероятностей. – М.: Наука, 1986. 432 с.
15. *Бурсиан Э.В.* Задачи по физике для компьютера. – М.: Просвещение, 1991. 256 с.
16. *Бусленко В.Н.* Автоматизация имитационного моделирования сложных систем – М.: Наука, 1977. 240 с.
17. *Вартофский М.* Модели. Репрезентация и научное понимание. – М., 1988. 37 с.
18. *Вентцель Е.С.* Исследование операций: задачи, принципы, методология. – М.: Наука, 1983. 208 с.
19. *Вентцель Е.С., Овчаров Л.А.* Теория вероятностей и ее инженерные приложения. – М.: Наука, 1988. 480 с.
20. *Вилкас Э.Й.* Оптимальность в играх и решениях. – М.: Наука, 1990. 256 с.
21. *Волкова В.Н., Денисов А.А.* Основы теории систем и системного анализа. – СПб: Изд-во СПбГТУ, 2001. 512 с.
22. *Волошинов А.В.* Математика и искусство. – М.: Просвещение, 2000. 399 с.
23. *Габов С.А.* Введение в теорию нелинейных волн. – М.: МГУ, 1988. 176 с.
24. *Гитман М.Б.* Введение в теорию нечетких множеств и интервальную математику: Ч.1: Применение лингвистической переменной в системах принятия решений. //Учебн. пособие. – Пермь: Изд-во Перм. гос. техн. ун-т, 1998. 45 с.

25. Глинский Б.А., Грязнов Б.С. и др. Моделирование как метод научного исследования. — М.: Наука, 1965. 245 с.
26. Глушков В.М., Иванов А.С., Яненко К.А. Моделирование развивающихся систем. — М.: Наука, 1983. 276 с.
27. Гмурман В.Е. Теория вероятностей и математическая статистика. — М.: Высшая школа, 1977. 479 с.
28. Гнеденко Б.В. Курс теории вероятностей. — М.: Наука, 1988. 448 с.
29. Годунов С.К. Уравнения математической физики. — М.: МГУ, 1971. 416 с.
30. Горстко А.Б. Познакомьтесь с математическим моделированием. — М.: Знание, 1991. 160 с.
31. Грогоно П. Программирование на языке Паскаль. — М.: Мир, 1982. 260 с.
32. Гроссман С., Дж. Тернер. Математика для биологов. — М.: Высшая школа, 1983. 366 с.
33. Гулд Х., Тобочник Я. Компьютерное моделирование в физике: В 2 ч. — М.: Мир, 1990. Ч. 2. 400 с.
34. Давыдов Э.Г. Исследование операций: Учеб. пособие для студентов вузов. — М.: Выш. школа, 1990. 383 с.
35. Данко П.Е., Попов А.Г. Высшая математика в упражнениях и задачах. Ч.1. — М.: Высшая школа, 1974. 416 с.
36. Демидович Б.П., Марон И.А., Шувалова Э.З. Численные методы анализа. Приближение функций, дифференциальные и интегральные уравнения. — М.: Наука, 1967. 368 с.
37. Дзержинский Ф.Я. Язык проектирования структуризованных программ //Алгоритмы и организация решения экономических задач. — М.: Статистика. Вып. 4, 1989. С. 83–95.
38. Ермаков С.М., Михайлов Г.А. Курс статистического моделирования. — М.: Наука, 1976. 320 с.
39. Заде Л.А. Понятие лингвистической переменной и его применение к принятию приближенных решений. — М.: Мир, 1976. 168 с.
40. Замков О.О., Толстомятенко А.В., Черемных Ю.Н. Математические методы в экономике. — М.: МГУ, «ДИС», 1998. 368 с.
41. Зельдович Я.Б., Мышкис А.Д. Элементы математической физики. — М.: Наука-Физматлит, 1973. 352 с.
42. Зельдович Я.Б., Мышкис А.Д. Элементы прикладной математики. — М.: Наука-Физматлит, 1965. 616 с.
43. Зенкевич О. Метод конечных элементов в технике. — М.: Мир, 1975. 412 с.
44. Зимин В.Д. Иерархическая модель турбулентности //Известия АН СССР: Физика атмосферы и океана. 1981. Т.17. № 12. С. 1265–1273.
45. Зимин В.Д., Фрик П.Г. Турбулентная конвекция. — М.: Наука, 1988. 178 с.
46. Зуховицкий С.И., Радчик И.А. Математические методы сетевого планирования. — М.: Наука, 1985. 296 с.
47. Иванцов Г.П. Температурное поле вокруг шарообразного, цилиндрического и иглообразного кристалла, растущего в переохлажденном расплаве // Доклады АН СССР, 1947. Т. 58. № 4. С. 567–570.
48. Ивлев Д.Д., Быковцев Г.И. Теория упрочняющегося пластического тела. — М.: Наука, 1971. 232 с.
49. Ивченко Г.И., Медведев Ю.И. Математическая статистика. — М.: Высшая школа, 1984. 248 с.
50. Ивченко Г.И., Кашистанов В.А., Коваленко И.Н. Теория массового обслуживания. — М.: Высшая школа, 1982. 256 с.

51. *Калиткин Н.Н.* Численные методы. — М.: Наука, 1978. 512 с.
52. *Карцев В.* Джеймс Клерк Максвелл — творец великих уравнений // Наука и жизнь. 1973. № 10. С. 109–119.
53. *Клейнен Дж.* Статистические методы в имитационном моделировании /Пер. с англ. Ю.П.Адлера и др. Вып. 2 — М.: Статистика, 1978. 335 с.
54. *Кнут Д.* Искусство программирования для ЭВМ. Т. 2: Получисленные алгоритмы. — М.: Мир, 1977. 370 с.
55. *Корн Г., Корн Т.* Справочник по математике (для научных работников и инженеров). — М.: Наука, 1978. 832 с.
56. *Кофман А.* Введение в теорию нечетких множеств. — М.: Радио и связь, 1982. 432 с.
57. *Круг Г.К., Сосулин Ю.А., Фатуев В.А.* Планирование эксперимента в задачах идентификации и экстраполяции. — М.: Наука, 1977. 298 с.
58. *Крылов В.И., Бобков В.В., Монастырный П.И.* Вычислительные методы. Т.2. — М.: Наука, 1977. 400 с.
59. *Курдюмов С.П.* Законы эволюции и самоорганизации сложных систем. В кн.: Наука, технология, вычислительный эксперимент. — М.: Наука, 1993. С. 6–32.
60. *Курдюмов С.П., Малинецкий Г.Г., Потапов А.Б., Самарский А.А.* Структуры в нелинейных средах. В кн. Компьютеры и нелинейные явления. — М.: Наука, 1988. С. 5–43.
61. Лекции по теории графов /Емеличев В.А. и др. — М.: Наука, 1990. 384 с.
62. *Лихачев В.А., Малинин В.Г.* Структурно-аналитическая теория прочности. — С-Пт.: Наука, 1983. 471 с.
63. *Лойцянский Л.Г.* Механика жидкости и газа. — М.: Наука-Физматлит, 1973. 848 с.
64. *Лоскутов А.Ю., Михайлов А.С.* Введение в синергетику. — М.: Наука, 1990. 272 с.
65. *Майника Э.* Алгоритмы оптимизации на сетях и графах. — М.: Мир, 1981. 323 с.
66. *Максимей И.В.* Имитационное моделирование на ЭВМ — М.: Радио и связь, 1988. — 232 с.
67. *Малинецкий Г.Г.* Хаос, структуры, вычислительный эксперимент. — М.: Наука, 1997. 255 с.
68. *Мануйлов В.Г.* Разработка программного обеспечения на Паскале. — М.: «ПРИОР», 1996. 240 с.
69. *Марчук Г.И.* Методы вычислительной математики. — М.: Наука, 1980. 536 с.
70. Математическое моделирование /Под ред. Дж.Эндрюса и Р.Мак-Лоуна. — М.: Мир, 1979. 250 с.
71. Математическое моделирование жизненных процессов. — М.: Мысль, 1968. 284 с.
72. *Мигдал А.Б.* О психологии научного творчества. I, II //Наука и жизнь. 1976. № 2. С. 100–107; № 3. С. 100–107.
73. Модели принятия решений на основе лингвистической переменной //А.Н. Борисов, А.В. Алексеев, О.А. Крумберг и др. — Рига: Зинатне, 1982. 256 с.
74. *Моисеев Н.Н.* Математика в социальных науках //Математические методы в социологическом исследовании. М., 1981. 166 с.
75. *Моисеев Н.Н.* Математические задачи системного анализа. — М.: Наука, 1981. 488 с.

76. Мороз А.И. Курс теории систем. — М.: Высшая школа, 1987. 304 с.
77. Мулен Э. Кооперативное принятие решений: Аксиомы и модели. — М.: Мир, 1991. 464 с.
78. Нариньяни А.С. Недоопределенность в системе представления и обработки знаний // Изв. АН СССР. Техн. кибернетика. 1986. № 5. С. 3–28.
79. Неуймин Я.Г. Модели в науке и технике. — Л.: Наука, 1984. 189 с.
80. Новиков О.А., Петухов С.И. Прикладные вопросы теории массового обслуживания — М.: Советское радио, 1969. 400 с.
81. Новое в синергетике. Загадки мира неравновесных структур. — М.: Наука, 1996. 263 с.
82. Обработка нечеткой информации в системах принятия решений // А.Н. Борисов, А.В. Алексеев, Г.В. Меркурьева и др. — М.: Радио и связь, 1989. 304 с.
83. Перегудов Ф.И., Тарасенко Ф.П. Введение в системный анализ. — М.: Высшая школа, 1989. 367 с.
84. Плотинский Ю.М. Теоретические и эмпирические модели социальных процессов. — М.: Логос, 1998. 280 с.
85. Поздеев А.А., Трусов П.В., Няшин Ю.И. Большие упругопластические деформации: теория, алгоритмы, приложения. — М.: Наука, 1986. 232 с.
86. Поля Д. Математика и правдоподобные рассуждения. — М.: Наука, 1975. 464 с.
87. Попов Э.В. Общение с ЭВМ на естественном языке. — М.: Наука, 1982. 360 с.
88. Пуанкаре А. О науке. — М.: Наука, 1990. 736 с.
89. Розанов Ю.А. Теория вероятностей, случайные процессы и математическая статистика. — М.: Наука, 1989. 320 с.
90. Розенблют А., Винер Н. Роль моделей в науке. В кн. Модели в науке и технике. — Л.: 1984. С. 171–175.
91. Румер Ю.Б., Рыбкин М.Ш. Термодинамика, статистическая физика и кинетика. — М.: Наука, Физматлит, 1972. 400 с.
92. Ружилов Ю.И. Решение научно-технических задач на персональном компьютере. — СПб.: КОРОНА, 2000. 272 с.
93. Сабоннадьер Ж.-К., Кулон Ж.-Л. Метод конечных элементов и САПР. — М.: Мир, 1989. 190 с.
94. Самарский А.А. Введение в численные методы. — М.: Наука, 1982. 272 с.
95. Самарский А.А., Гулин А.В. Численные методы: Учеб. пособие. — М.: Наука, 1989. 432 с.
96. Святловский А. Служба цунами // Наука и жизнь. 1973. № 8. С. 38–42.
97. Седов Л.И. Методы подобия и размерности в механике. — М.: Наука-Физматлит, 1987. 432 с.
98. Сивухин Д.В. Общий курс физики. III. Электричество. — М.: Наука-Физматлит, 1983. 688 с.
99. Слоер С. Математические фантазии. — М.: Мир, 1993. 184 с.
100. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики. — М.: Наука, 1969. 512 с.
101. Советов Б.Я., Яковлев С.А. Моделирование систем. — М.: Высшая школа, 1998. 319 с.
102. Соколкин Ю.В., Ташкинов А.А. Механика деформирования и разрушения структурно-неоднородных тел. — М.: Наука, 1984. 115 с.
103. Справочник по теории вероятности и математической статистике/ Королук В.С., Портенко Н.И., Скороходов А.В., Турбин А.Ф. — М.: Наука, 1985. 640 с.

104. *Татт У.* Теория графов. – М.: Мир, 1988. 424 с.
105. Теория выбора и принятия решений: Учеб. пособие /И.М. Макаров, Т.М. Виноградская и др. – М.: Наука, 1982. 328 с.
106. Теория сетей связи. Учебник для вузов / В.Н. Рогинский и др. – М.: Радио и связь, 1981. 192 с.
107. *Тоффолит Т., Марголюс Н.* Машины клеточных автоматов. – М.: Мир, 1991. 280 с.
108. *Уизем Дж.* Линейные и нелинейные волны. – М.: Мир, 1977. 622 с.
109. *Федер Е.* Фракталы. – М.: Мир, 1991. 260 с.
110. *Хворинев Н.И.* Кристаллизация и неоднородность стали. – М.: Машгиз, 1958. 392 с.
111. *Хемминг Р.В.* Численные методы для научных работников и инженеров. – М.: Наука, 1968. 400 с.
112. Численное исследование актуальных проблем машиностроения и механики сплошных и сыпучих сред методом крупных частиц: В 5 т. /Под ред. Ю.М.Давыдова. – М.: Национал. акад. приклад. наук, Междунар. асоц. разработчиков и пользователей метода крупных частиц, 1995. 1595 с.
113. *Шеннон Р.* Имитационное моделирование систем – искусство и наука. – М.: Мир, 1978. 417 с.
114. *Шеффе Г.* Дисперсионный анализ. – М.: Наука, 1980. 512 с.
115. *Штофф В.А.* Моделирование и философия. – М.: 1966. 56 с.
116. *Эйнштейн А.* Физика и реальность. – М.: Наука, 1965. 136 с.
117. *Adams, R.L.T., Knowler, J.T., and Leader, D.P.* The Biochemistry of the Nucleic Acids (11th ed.). – Chapman and Hall. 1992.
118. *Barbi M., Cocco S., Peyrard M., and Ruffo S.* A twist opening model for DNA // J. of Biological Physics. 1999. V. 24. P. 97–114.
119. *Calladine, C.R. and Drew, H.R.* Understanding DNA. – San Diego: Academic Press. 1977.
120. *Dauxois T., Peyrard M.* A nonlinear model for DNA melting// Nonlinear excitation in biomolecules. Ed. M.Peyrard. – Berlin: Springer-Verlag. 1995. P. 127–136.
121. *Dauxois T., Peyrard M.* Entropy driven transition in a one-dimensional system// Phys.Rev.E, 1995. V. 51. № 5. P. 4027–4040.
122. *Frick P., Galyagin D.* et al. Wavelet analysis of solar activity recorded by sunspot groups // Astronomy and Astrophysics. 1997. V. 328. P. 670–681.
123. *Grossmann A., Morlet J.* Decomposition of Hardy functions into square integrable wavelets of constant shape // SIAM J.Math.Analysis. 1984. V. 15. № 4. P. 723–736.
124. *Marko J.F., Siggia E.D.* Statistical mechanics of supercoiled DNA // Physical Review D, 1995 V.52. № 3. P. 2912–2938.
125. *Nemes-Ribes E., Frick P.* et al. Wavelet analysis of Maunder minimum as recorded in Solar diameter data // Comptes Rendues Acad. Sciences Paris, Serie Iib. 1995. V.321. P.525–532.
126. *Okubo H., Hubbard M.* Dynamics of basketball-rim interactions // Proc. of Int. Conference on Sport Engineering.-Oxford: Blackwell Science, 2002. P.660–666.
127. *Peyrard M.* The pathway to energy localization in nonlinear lattices // Physica D. 1998. V.119. P.184–199.
128. *Peyrard M.* Using DNA to probe nonlinear localized excitations? // Europhys. Lett.. 1998. V.44. № 3. P.271–277.
129. Saenger. Principles of Nucleic Acid Structure. – New York: Springer-Verlag, 1984.
130. *Zadeh L.A.* Fuzzy sets // Inform. Contr. 1965. V.8. P. 338–353.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

автомат, 382
автомодельность, 259, 266, 287
агрегат-имитатор, 160
агрегат-классификатор, 158
агрегат-конфигуратор, 157
агрегат-оператор, 159
агрегат-статистик, 160
агрегат-структура, 161
агрегирование, 156
адекватность, 19, 82
аксиоматика Колмогорова, 208
анализ. См. декомпозиция
анализ подобия и размерности, 282
аналогия, 17

Б

богатство модели, 20

В

вейвлет-анализ, 314, 324
вероятность, 207
волновое уравнение, 253, 273
выборка, 207
вычислительный эксперимент, 54

Г

гармоника, 260
гипотеза, 17
гистограмма, 213
граф, 145
граф состояний, 232

Д

декомпозиция, 153
дендриты, 349
детерминированная величина, 182
детерминированные параметры, 44
дискретная случайная величина, 210
дислокации, 399
дислокационные структуры, 401
дисперсия, 211, 299
диссипативные структуры, 295
ДНК, 360

З

задача Штурма-Лиувилля, 260
закон Гука, 246
закон распределения, 210
законы термодинамики, 278
заявка, 235

И

имитатор, 160, 370, 377
имитационная система, 37

имитационный подход, 53, 369
индекс ранжирования, 198
интервальное описание параметров, 45, 188

К

канал, 235
клеточный автомат, 383, 403
когнитология, 32
контроль математической замкнутости, 69
контроль порядков, 68
контроль размерности, 68
корректность, 70

Л

лингвистическая переменная, 193

М

максиминное произведение, 192
марковский случайный процесс, 228
математическое ожидание, 210
медиана, 211
метод Даламбера, 262
метод Монте-Карло, 219
метод равных вероятностей, 226
метод Фурье, 257, 259
мода, 211, 260
моделирование аналоговое, 26
моделирование идеальное, 23, 26
моделирование интуитивное, 27
моделирование математическое, 32
моделирование материальное, 23, 24
моделирование натурное, 24
моделирование научное, 27
модель, 13, 17, 18, 29
модель аналитическая, 50
модель «белого ящика», 145
модель дескриптивная, 48
модель информационная, 36
модель когнитивная, 30
модель концептуальная, 30
модель линейная, 39
модель логико-семантическая, 30
модель математическая, 35
модель нестационарная, 47
модель объяснительная, 30
модель описательная, 30
модель оптимизационная, 49
модель приближенная, 52
модель причинно-следственная, 31
модель прогностическая, 30
модель системы, 144
модель содержательная, 30
модель состава системы, 144

модель стационарная, 47
модель структурно-функциональная, 31
модель структурная, 38, 147, 152
модель структуры системы, 145
модель управленческая, 50
модель формальная, 31
модель «черного ящика», 144
модель численная, 52

Н

недоопределенность, 185
недоверность, 184
неизвестность, 184
неоднозначность, 185
неопределенность параметров, 44
неопределенность прагматическая, 186
неопределенность семантическая, 186
неопределенность синтаксическая, 186
неполнота, 185
нечеткое множество, 190
нечеткое описание параметров, 45
нечеткое отношение, 191
нормальное распределение, 217
нормальные моды, 258

О

обследование объекта, 58, 60
объект исследования, 16
омонимия, 186
оператор модели, 39
определяющие соотношения, 67, 255
опыт, 207
отношения, 145

П

параметры модели, 42
параметры управления, 49
параметры экзогенные, 43
параметры эндогенные, 43
плотность распределения, 213
погрешность, 53, 74
показательное распределение, 215
полигон частот, 213
полисемия, 185
полнота модели, 19
постановка концептуальная, 62
постановка математическая, 66
постановка содержательная, 59
потенциальность модели, 20
поток событий, 230
принцип суперпозиции, 249
простейший поток, 231
простота модели, 20
П-теорема, 282

Р

равномерное распределение, 216
распределение Пуассона, 215

С

свойства моделей, 19
синтез, 152
система, 37, 143
система массового обслуживания, 235
случайное описание параметров, 45
случайное событие, 182
случайный процесс, 228
собственные функции, 260
событие, 182
состояние объекта, 47
спецификация, 78
сплошная среда, 250, 253
среднее квадратическое отклонение, 212
стохастическое описание параметров, 44, 187
структурная схема, 145

Т

теория, 29
техническое задание, 61
точность алгоритма, 74

У

уравнение диффузии, 271
уравнение Кортевега-де Фриза, 298
уравнение Уизема, 298
уравнение Шредингера, 276
уравнения Колмогорова, 233
уравнения Максвелла, 267

Ф

финальные вероятности, 234
фракталы, 332
фрейм, 153
функция принадлежности, 190, 191
функция распределения, 212
фурье-анализ, 258, 309

Х

характеристики уравнений
матфизики, 262, 269

Ч

частость, 207

Э

элемент вероятности, 232
элемент системы, 142
эмерджентность, 156
энтропия, 279
этапы построения модели, 56, 58
эффект Портевена-Ле Шателье, 412

Я

язык моделирования, 375

ОГЛАВЛЕНИЕ

Предисловие	5
Глава 1. Определение и назначение моделирования	11
1.1. Что такое модель?	13
Место моделирования среди методов познания	13
Определение модели	16
Свойства моделей	19
Цели моделирования	21
1.2. Классификация моделей	23
Материальное моделирование	24
Идеальное моделирование	26
Когнитивные, концептуальные и формальные модели	29
1.3. Классификация математических моделей	36
Классификационные признаки	36
Классификация математических моделей в зависимости от сложности объекта моделирования	37
Классификация математических моделей в зависимости от оператора модели	39
Классификация математических моделей в зависимости от параметров модели	42
Классификация математических моделей в зависимости от целей моделирования	48
Классификация в зависимости от методов реализации	50
Вопросы для самопроверки	54
Задания для самостоятельного выполнения	55
Глава 2. Этапы построения математической модели	56
2.1. Обследование объекта моделирования	58
2.2. Концептуальная постановка задачи моделирования	62
2.3. Математическая постановка задачи моделирования	66
2.4. Выбор и обоснование выбора метода решения задачи	72
2.5. Реализация математической модели в виде программы для ЭВМ	78
2.6. Проверка адекватности модели	82
2.7. Практическое использование построенной модели и анализ результатов моделирования	86
Вопросы для самопроверки	90
Задания для самостоятельного выполнения	90
Глава 3. Примеры математических моделей	92
3.1. Статический анализ конструкций	94
3.2. Модель спроса-предложения	100
3.3. Динамика популяций	106
3.4. Модель конкуренции двух популяций	116
3.5. Гармонический осциллятор	128
Задания для самостоятельного выполнения	138
Глава 4. Структурные модели	142
4.1. Что такое структурная модель?	143
4.2. Способы построения структурных моделей	152
4.3. Примеры структурных моделей	162
Вопросы для самопроверки	180
Задания для самостоятельного выполнения	180

Глава 5. Моделирование в условиях неопределенности	181
5.1. Причины появления неопределенностей и их виды	183
5.2. Моделирование в условиях неопределенности, описываемой с позиций теории нечетких множеств	188
5.3. Моделирование в условиях стохастической неопределенности.....	205
5.4. Моделирование марковских случайных процессов	228
Вопросы для самопроверки	241
Задания для самостоятельного выполнения	243
Глава 6. Линейные и нелинейные модели	245
6.1. О законе Гука и границах линейности	246
6.2. Сплошные среды и уравнения математической физики. Линейные уравнения и принцип суперпозиции	249
6.3. Вывод волнового уравнения из законов механики	253
6.4. Решение волнового уравнения методом Фурье	257
6.5. О характеристиках уравнений математической физики. Решение волнового уравнения методом Даламбера	262
6.6. Уравнения Максвелла	266
6.7. О классификации квазилинейных систем	269
6.8. Связь непрерывного и дискретного на примерах уравнения колебаний струны и уравнения Шредингера	272
6.9. О пользе феноменологии при построении математических моделей	277
6.10. Анализ подобия и размерности	282
6.11. Автомодельность	287
6.12. Самоорганизация и структуры в нелинейных средах	291
6.13. О нелинейных волнах в сплошных средах	296
6.14. Иерархические модели турбулентности и многомасштабные функциональные базисы	304
6.15. Вейвлеты	314
6.16. Вейвлет-анализ временных колебаний	324
6.17. О фракталах и их применении	332
6.18. Нелинейные модели ДНК	360
Задания для самостоятельного выполнения	368
Глава 7. Моделирование с использованием имитационного подхода	369
7.1. Особенности моделей, использующих имитационный подход	370
7.2. Имитатор системы массового обслуживания	377
7.3. Клеточные автоматы	382
7.4. Моделирование дислокаций в металле	399
Вопросы для самопроверки	414
Задания для самостоятельного выполнения	414
Приложения	
Приложение 1. Язык формального описания алгоритмов	417
Приложение 2. Численные методы (минимальные сведения)	421
Библиографический список	
Предметный указатель	

Учебное издание

Валерий Николаевич Ашихмин,
Михаил Борисович Гитман, Илья Эрнстович Келлер,
Олег Борисович Наймарк, Валерий Юрьевич Столбов,
Петр Валентинович Трусов, Петр Готлобович Фрик

Введение в математическое моделирование

Учебное пособие

Редактор *Н.Г. Давыдова*
Корректор *Е.В. Комарова*
Оформление *Т. Хрычевой*
Компьютерная верстка *О.Г. Лауровой*

Подписано в печать 26.11.2006. Формат 60х90/16.
Печать офсетная. Бумага офсетная. Печ. л. 27,5.
Тираж 2000 экз. Заказ №1896.

Издательская группа «Логос»
105318, Москва, Измайловское ш., 4

Отпечатано в ОАО «ИПК «Ульяновский Дом печати»
432980, г. Ульяновск, ул. Гончарова, 14

По вопросам приобретения литературы
обращаться по адресу:
105318, Москва, Измайловское ш., 4
Тел./факс: (095) 369-5819, 369-5668, 369-7727
Электронная почта: universitas@mail.ru
Дополнительная информация на сайте:
<http://logosbook.ru>