

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Баламирзоев Насим Диодинович
Должность: И.о. ректора
Дата подписания: 19.08.2023 23:10:06
Уникальный программный ключ:
2a04bb882d7edb7f479cb266eb4aaaaedebee849

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Дагестанский государственный технический университет»

Т.Г. Асланов, У.Р. Тетакаев

Учебное пособие
МОДЕЛИРОВАНИЕ СИСТЕМ

Махачкала

2020

УДК 004.94
ББК 2.18я73
А90

Р е ц е н з е н т ы:

Мамедов Л.К., канд. техн. наук, руководитель НТЦ-4 АО «Азимут»;
Исабекова Т.Э., канд. физ.-мат. наук, заведующая кафедрой прикладной математики и информатики ДГТУ.

А в т о р ы:

Т.Г. Асланов, У.Р. Тетакаев

А90 Моделирование систем: учебное пособие /Т.Г. Асланов, У.Р. Тетакаев;
ДГТУ. – Махачкала, 2020 – 167 с.

Изложены основные методы моделирования систем. Каждая глава содержит теоретический материал, методы решения задач по моделированию систем управления и примеры. Предназначено для студентов, обучающихся по направлению подготовки 27.03.04 «Управление в технических системах» и УГС(Н) 09.00.00. Учебное пособие способствует закреплению теоретических основ курса «Моделирование систем».

УДК 004.94
ББК 2.18я73

Печатается по решению Ученого совета ДГТУ

ISBN 2227-8397

© Асланов Т.Г., Тетакаев У.Р.

© ДГТУ, 2020

Оглавление

| | |
|---|----|
| Введение | 5 |
| 1. Основные определения понятия | 6 |
| 1.1. Назначение моделирования..... | 6 |
| 1.2. Основные определения и понятия теории моделирования..... | 12 |
| 1.3. Области использования моделирования..... | 17 |
| 1.4. Классический и системный подход к моделированию..... | 19 |
| 2. Классификация видов моделирования | 22 |
| 2.1.Классификация моделей по степени абстрагирования модели от оригинала..... | 22 |
| 2.2.Классификация моделей по степени устойчивости..... | 25 |
| 2.3.Классификация моделей по отношению к внешним факторам..... | 25 |
| 2.4.Классификация моделей по отношению ко времени..... | 26 |
| 3. Физическое моделирование | 28 |
| 3.1.Основные определения и понятия физического моделирования..... | 28 |
| 3.2. Критерий подобия..... | 29 |
| 3.3. Динамическое подобие и моделирование явлений..... | 32 |
| 4. Математическое моделирование | 38 |
| 4.1. Классификация математических моделей..... | 38 |
| 4.2. Основные понятия математического моделирования..... | 46 |
| 4.3. Основные этапы математического моделирования..... | 53 |
| 4.4. Этапы построения математической модели..... | 61 |
| 5. Имитационное моделирование | 65 |
| 5.1. Понятие статистического эксперимента..... | 65 |
| 5.2 Область применения и классификация имитационных моделей..... | 69 |
| 5.3. Описание поведения системы..... | 73 |
| 5.4. Моделирование случайных факторов..... | 77 |
| 5.5 Управление модельным временем..... | 86 |
| 5.6. Моделирование параллельных процессов..... | 94 |

| | |
|---|-----|
| 6. Обработка и анализ результатов эксперимента | 111 |
| 6.1. Обработка табличных данных..... | 111 |
| 6.2. Аппроксимация..... | 119 |
| 6.3. Методы численного интегрирования..... | 125 |
| 6.4. Методы решения дифференциальных уравнений..... | 131 |
| 6.5. Методы решения уравнений и систем..... | 137 |
| 6.6. Решение систем нелинейных уравнений..... | 154 |
| 7. Ошибки измерений | 157 |
| Приложение | 161 |
| Библиографический список | 165 |

Введение

Моделирование как прием познания законов природы, тесно связано с развитием человеческих знаний, и применялось человечеством с давних времен, примером такого моделирования является расчет времени солнечных затмений впервые предсказанного Фалесом 585 г. до н.э., определение размера Земли в третьем веке до н.э. Эратосфеном и т.д.

Так с прогрессом в математике и технологиях, человечество все больше прибегало к физическому, математическому и статистическому моделированию.

Появление первых электронных вычислительных машин и зарождение основ кибернетики позволили использовать новые универсальные методы оптимизации для моделирования систем. В данный период получили свое развитие компьютерное и имитационное моделирование. В этот период моделирование приобрело общенаучный характер.

На сегодняшний день при разработке сложных и трудоемких систем используют моделирование для решения практических задач. Моделирование для систем проводят как для отдельных узлов системы, так и системы в целом. Проведение моделирования системы сокращает объем трудоемких и дорогостоящих экспериментов.

Основная задача моделирования различного рода процессов и систем с целью исследования объектов, прогнозирования их поведения или поиска наилучших условий функционирования сводится к расчету анализируемых показателей по математической модели при тех или иных значениях (или функциях) входных величин. Важное значение при этом приобретают вычислительные алгоритмы, с помощью которых можно получить при моделировании решение конкретной математической задачи [1].

Знакомству с идеями и алгоритмами решения наиболее распространенных задач вычислительной математики, применяющихся при математическом моделировании, получению практических навыков их применения и посвящено данное учебное пособие.

1. Основные определения и понятия

1.1. 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 [2].

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.

Рассмотрим место моделирования среди иных 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), на всеобщие, общенаучные и частнонаучные.

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

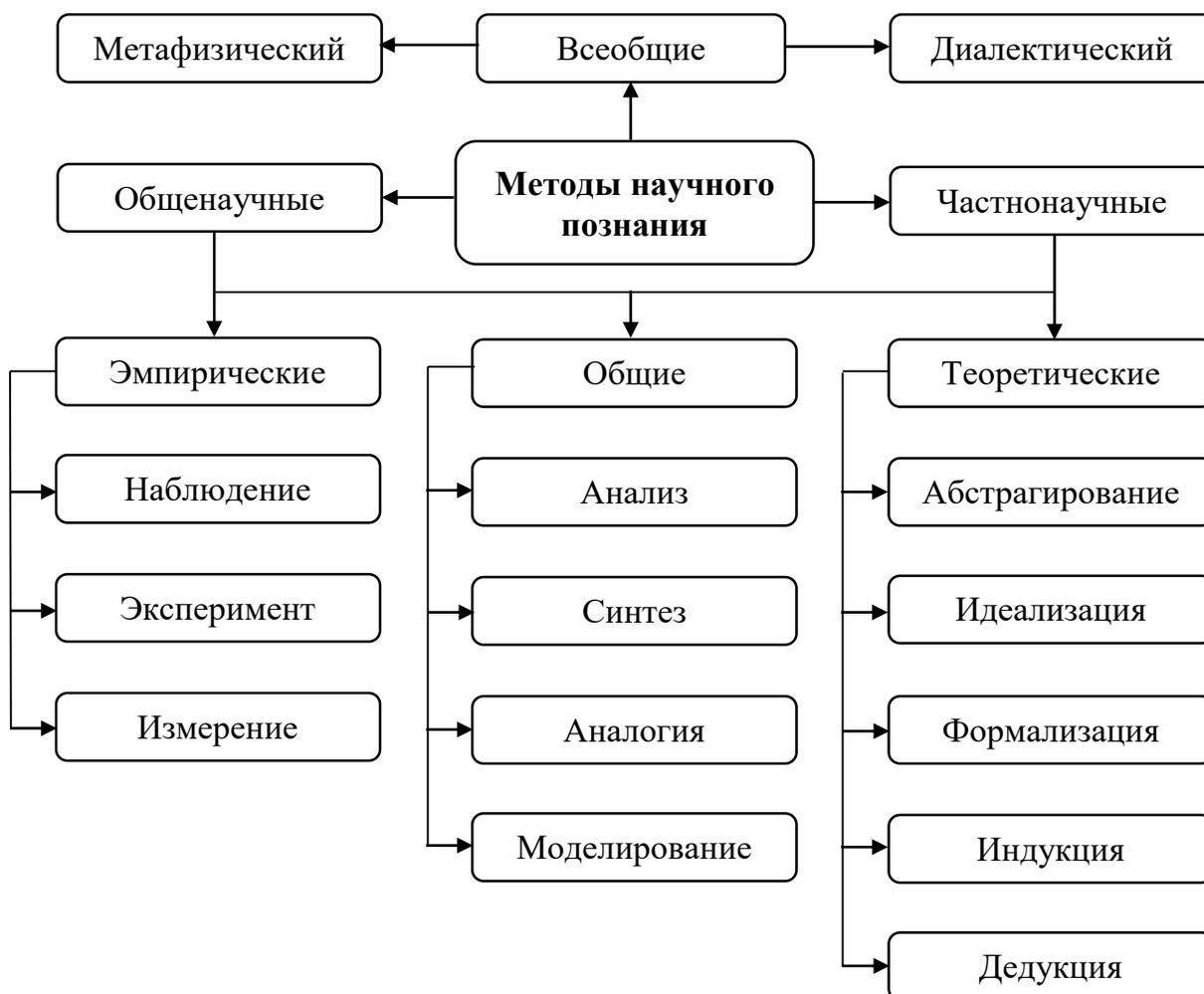


Рис. 1 – Методы научного познания

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 применять 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*

1.2. Основные определения и понятия теории моделирования

Приведем основные определения и понятия, используемые в теории моделирования .

Теория моделирования – теория замещения объекта-оригинала его моделью и исследования свойств объекта на его модели [3].

Модель (объекта – оригинала) (от лат. *modus* – «мера», «объем», «образ») – явление, предмет, установка, знаковое образование или условный образ (описание, схема), находящийся в некотором соответствии с изучаемым объектом и способным заменить его в процессе исследования, давая о нем информацию.

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 [2]:

– 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 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»*.

Любая модель строится и исследуется при определенных допущениях. Модель должна строиться так, чтобы она наиболее полно воспроизводила те качества объекта, которые необходимо изучить в соответствии с поставленной целью. Во всех отношениях модель должна быть проще объекта и удобнее его для изучения. Таким образом, для одного и того же объекта могут существовать различные модели, классы моделей, соответствующие различным целям его изучения. Необходимым условием моделирования

является подобие объекта и его модели. В этом случае мы должны говорить об адекватности модели объекту-оригиналу [4].

Если результаты моделирования подтверждаются и могут служить основой для прогнозирования процессов, протекающих в исследуемых объектах, то говорят, что модель адекватна объекту. При этом адекватность модели зависит от цели моделирования и принятых критериев.

Под *адекватной* моделью понимается модель, которая с определенной степенью приближения на уровне понимания моделируемой системы разработчиком модели отражает процесс ее функционирования во внешней среде. Под адекватностью (от лат. *adaequatus* – приравненный) будем понимать степень соответствия результатов, полученных по разработанной модели, данным эксперимента или тестовой задачи. Если система, для которой разрабатывается модель, существует, то сравнивают выходные данные модели и этой системы. В том случае, когда два набора данных оказываются подобными, модель существующей системы считается адекватной. Чем больше общего между существующей системой и ее моделью, тем больше уверенность в правильности модели системы.

Проверка адекватности модели необходима для того, чтобы убедиться в справедливости совокупности гипотез, сформулированных на первом этапе разработки модели, и точности полученных результатов; соответствует точности, требуемой техническим заданием.

Для моделей, предназначенных для приблизительных расчетов, удовлетворительной считается точность от 10 до 15%, а для моделей, предназначенных для использования в управляющих и контролирующих системах, от 1 до 2 %.

Любая модель обладает следующими свойствами:

- конечностью: модель отображает оригинал лишь в конечном числе его отношений;
- упрощенностью: модель отображает только существенные стороны объекта;

– приближенностью: действительность отображается моделью грубо или приближенно;

– адекватностью: модель успешно описывает моделируемую систему;

– информативностью: модель должна содержать достаточную информацию о системе в рамках гипотез, принятых при построении модели.

Процесс построения, изучения и применения моделей будем называть моделированием.

Моделирование – исследование объекта (явления) с помощью моделей, которые воспроизводят наиболее важные черты оригинала.

Когда говорят о моделировании, обычно имеют в виду моделирование некоторой системы [5].

Система – совокупность взаимосвязанных элементов, объединенных для реализации общей цели, обособленная от окружающей среды и взаимодействующая с ней как целостное целое и проявляющая при этом основные системные свойства. К основным свойствам системы относят [6]: эмергентность (эмерджентность); цельность; структурированность; целостность; подчиненность цели; иерархичность; бесконечность; эргатичность; открытость; необратимость; единство структурной устойчивости и неустойчивости; нелинейность; потенциальная многовариантность актуальных структур; критичность; непредсказуемость в критической области.

1.3. Области использования моделирования

Моделирование широко применяется при решении различных задач, возникающих в сфере обработки данных. Эти задачи могут быть разбиты на следующие три категории [6]:

- 1) Возникающие при выполнении работ по выбору;
- 2) Возникающие при выполнении работ по усовершенствованию;
- 3) Возникающие при выполнении работ по проектированию.

Работы по выбору связаны с проблемами, встречающимися при

проектировании или покупке системы.

Для вычислительной системы, эти проблемы включают выбор способа обработки информации, выбор вычислительной установки из имеющихся вариантов, выбор языка программирования, выбор пакета системных или прикладных программ. В общем случае работы по выбору могут быть определены как работы по выбору наиболее подходящие альтернативы для данного применения среди различных доступных вариантов при помощи некоторых критериев выбора: быстродействия, стоимости и т.д.

Работы по усовершенствованию сводятся к модификации существующих вариантов для усовершенствования их параметров. К ним относятся работы по настройке вычислительной системы (ВС), т.е. подбору ее параметров с целью приспособления к рабочей нагрузке. В эту же категорию входят работы по развитию ВС, состоящие в замене одного или нескольких аппаратных компонентов. Например, ВС может развиваться путем расширения оперативной памяти, добавления центрального процессора или замены его более высокопроизводительным. К другим типам модификации (настройке) относятся переупорядочение информации внутри одного или нескольких накопителей на магнитных дисках или усовершенствование структуры связей между устройствами ввода-вывода и каналами.

При проведении работ по проектированию возникает необходимость в ответе на вопросы, появляющиеся при разработке систем и их компонентов. Части, из которых будет состоять система, могут существовать, но могут и отсутствовать в принципе. В последнем случае они должны быть спроектированы и изготовлены. Синтез первоначальных вариантов системы и их корректировку выполняют разработчики. Но для корректировки и сравнения необходимо иметь возможность этих действий. Такую возможность дают средства моделирования.

Приведенная выше классификация является достаточным условием. Например, определение удовлетворительного размера оперативной памяти может рассматриваться как задача выбора, если сопоставляемые

альтернативные варианты отличаются размером памяти.

1.4. Классический и системный подход к моделированию

При моделировании систем используют два подхода: классический (индуктивный), сложившийся исторически первым, и системный, получивший развитие в последнее время [3, 5].

Классический подход. Исторически первым сложился классический подход к изучению объекта, моделированию системы. Классический подход синтеза модели (М) системы представлен на рис. 2. Реальный объект, подлежащий моделированию, разбивается на подсистемы, выбираются исходные данные (Д) для моделирования и ставятся цели (Ц), отражающие отдельные стороны процесса моделирования. По отдельной совокупности исходных данных ставится цель моделирования отдельной стороны функционирования системы, на базе этой цели формируется некоторая компонента (К) будущей модели. Совокупность компонент объединяется в модель.

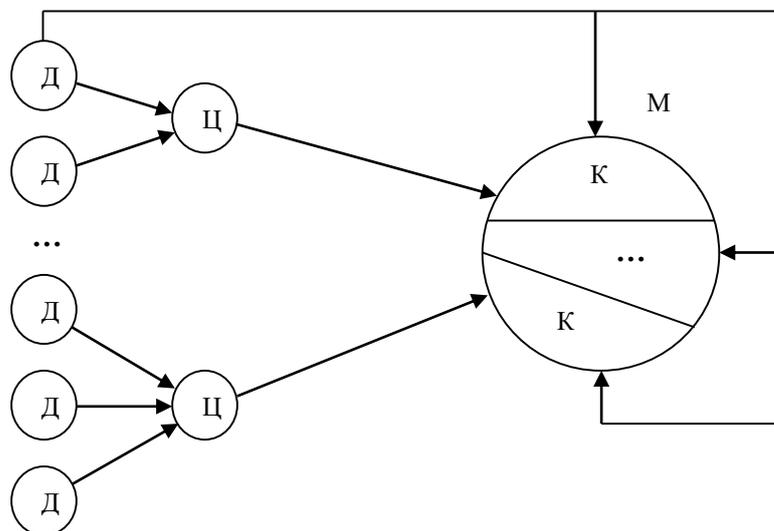


Рис. 2 – Классический подход к построению объекта, изучению модели

Таким образом происходит суммирование компонент, каждая компонента решает свои собственные задачи и изолирована от других частей

модели. Применим подход только для простых систем, где можно не учитывать взаимосвязи между компонентами. Можно отметить две отличительные стороны классического подхода:

- наблюдается движение от частного к общему при создании модели;
- созданная модель (система) образуется путем суммирования отдельных ее компонент и не учитывает возникновение нового системного эффекта.

Системный подход – методологическая концепция, основанная на стремлении построить целостную картину изучаемого объекта с учетом важных для решаемой задачи элементов объекта, связей между ними и внешних связей с другими объектами и окружающей средой. С усложнением объектов моделирования возникла необходимость их наблюдения с более высокого уровня. В этом случае разработчик рассматривает данную систему как некоторую подсистему более высокого ранга.

Например, если ставится задача проектирования АСУ предприятия, то с позиции системного подхода нельзя забывать, что эта система является составной частью АСУ объединением. В основе системного подхода лежит рассмотрение системы как интегрированного целого, причем это рассмотрение при разработке начинается с главного – формулировки цели функционирования. На рис. 3 условно представлен процесс синтеза модели системы на основе системного подхода. Важным для системного подхода является определение структуры системы – совокупности связей между элементами системы, отражающих их взаимодействие.

На основе исходных данных (Д), которые известны из анализа внешней системы, тех ограничений, которые накладываются на систему сверху либо, исходя из возможностей ее реализации, и на основе цели функционирования формулируются исходные требования (Т) к модели системы. На базе этих требований формируются ориентировочно некоторые подсистемы (П), элементы (Э) и осуществляется наиболее сложный этап синтеза – выбор (В) составляющих системы, для чего используются специальные критерии выбора

КВ.

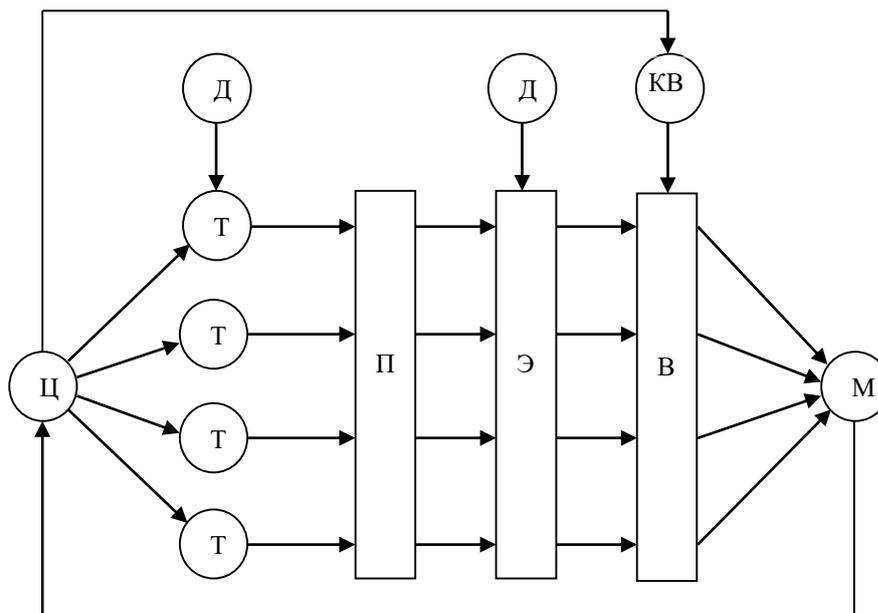


Рис. 3 – Системный подход к построению объекта, изучению модели

При моделировании необходимо обеспечить максимальную эффективность модели системы. Эффективное» обычно определяется как некоторая разность между какими-то показателями ценности результатов, полученных в итоге эксплуатации модели, и теми затратами, которые были вложены в ее разработку и создание.

Существуют структурные и функциональные подходы к исследованию структуры системы и ее свойств. При структурном подходе выявляются состав выделенных элементов системы и связи между ними. При функциональном подходе рассматриваются алгоритмы поведения системы (функции – свойства, приводящие к достижению цели).

2 Классификация видов моделирования

В общем случае все модели, независимо от областей и сфер их применения, бывают трех типов: познавательные, прагматические и инструментальные [4].

Познавательная модель – форма организации и представления знаний, средство соединения новых и старых знаний. Познавательная модель обычно подгоняется под реальность и является теоретической моделью.

Прагматическая модель – средство организации практических действий, рабочего представления целей системы для ее управления. Реальность в них подгоняется под некоторую прагматическую модель. Это, как правило, прикладные модели.

Инструментальная модель – средство построения, исследования и/или использования прагматических и/или познавательных моделей.

Познавательные отражают существующие, а прагматические – хоть и не существующие, но желаемые и, возможно, исполнимые отношения и связи.

Вся остальная классификация моделей выстраивается по отношению к объекту-оригиналу, методам изучения и т. п.

2.1 Классификация моделей по степени абстрагирования модели от оригинала

По степени абстрагирования от оригинала (см. рис. 4) модели могут быть разделены на материальные (физические) и идеальные. К материальным относятся такие способы, при которых исследование ведется на основе модели, воспроизводящей основные геометрические, физические, динамические и функциональные характеристики изучаемого объекта. Основными разновидностями физических моделей являются [7]:

- натурные;
- квазинатурные;
- масштабные;
- аналоговые.

Натурные – это реальные исследуемые системы, которые являются макетами и опытными образцами. Натурные м

одели имеют полную адекватность с системой-оригиналом, что обеспечивает высокую точность и достоверность результатов моделирования; другими словами, модель натурная, если она есть материальная копия объекта моделирования. Например, глобус – натурная географическая модель земного шара.

Квазинатурные (от лат. «квази» – почти) – это совокупность натуральных и математических моделей. Этот вид моделей используется в случаях, когда математическая модель части системы не является удовлетворительной или когда часть системы должна быть исследована во взаимодействии с остальными частями, но их еще не существует либо их включение в модель затруднено или дорого.

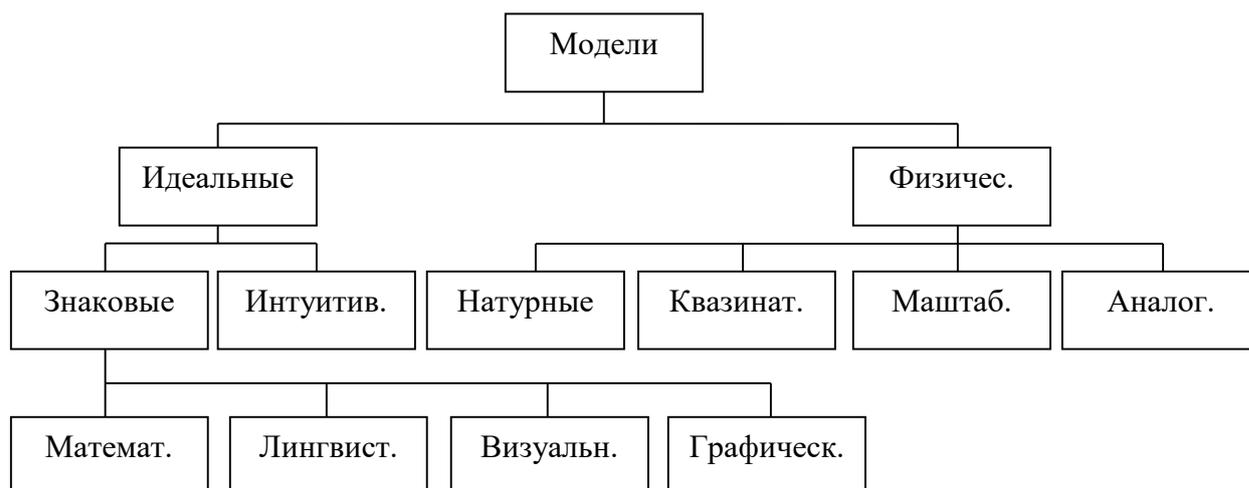


Рис. 4 – Схема классификации моделей по степени абстрагирования от объекта-оригинала

Масштабные модели – это системы той же физической природы, что и оригинал, но отличающиеся от него размерами. В основе масштабных моделей лежит математический аппарат теории подобия, который предусматривает соблюдение геометрического подобия оригинала и модели и соответствующих масштабов для их параметров. Примером масштабного моделирования являются любые разработки макетов домов, а порой и целых

районов при проведении проектных работ при строительстве. Также масштабное моделирование используется при проектировании крупных объектов в самолетостроении и кораблестроении.

Аналоговое моделирование основано на аналогии процессов и явлений, имеющих различную физическую природу, но одинаково описываемых формально (одними и теми же математическими уравнениями, логическими схемами и т.п.). В качестве аналоговых моделей используются механические, гидравлические, пневматические системы, но наиболее широкое применение получили электрические и электронные аналоговые модели, в которых сила тока или напряжение является аналогами физических величин другой природы. Например, является общеизвестным, что математическое уравнение колебания маятника имеет эквивалент при записи уравнения колебаний тока.

Идеальное моделирование носит теоретический характер. Различают два типа идеального моделирования: интуитивное и знаковое.

Под интуитивным моделированием будем понимать моделирование, основанное на интуитивном представлении об объекте исследования, не поддающемся формализации либо не нуждающемся в ней. В этом смысле, например, жизненный опыт каждого человека может считаться его интуитивной моделью окружающего мира.

Знаковым называется моделирование, использующее в качестве моделей знаковые преобразования различного вида: схемы, графики, чертежи, формулы, наборы символов и т. д., включающие совокупность законов, по которым можно оперировать с выбранными знаковыми элементами. Знаковая модель может делиться на лингвистическую, визуальную, графическую и математическую модели.

Модель лингвистическая, – если она представлена некоторым лингвистическим объектом, формализованной языковой системой или структурой. Иногда такие модели называют вербальными, например, правила дорожного движения – языковая, структурная модель движения транспорта и пешеходов на дорогах.

Модель визуальная, – если она позволяет визуализировать отношения и связи моделируемой системы, особенно в динамике. Например, на экране компьютера часто пользуются визуальной моделью объектов, клавиатуры в программе-тренажере по обучению работе на клавиатуре.

Модель графическая, – если она представима геометрическими образами и объектами, например, макет дома является натурной геометрической моделью строящегося дома. Важнейшим видом знакового моделирования является математическое моделирование, классическим примером математического моделирования является описание и исследование основных законов механики И. Ньютона средствами математики.

2.2. Классификация моделей по степени устойчивости

Все модели могут быть разделены на устойчивые и неустойчивые.

Устойчивой является такая система, которая, будучи выведена из своего исходного состояния, стремится к нему. Она может колебаться некоторое время около исходной точки, подобно обычному маятнику, приведенному в движение, но возмущения в ней со временем затухают и исчезают.

В неустойчивой системе, находящейся первоначально в состоянии покоя, возникшее возмущение усиливается, вызывая увеличение значений соответствующих переменных или их колебания с возрастающей амплитудой.

2.3. Классификация моделей по отношению к внешним факторам

По отношению к внешним факторам модели могут быть разделены на открытые и замкнутые.

Замкнутой моделью является модель, которая функционирует вне связи с внешними (экзогенными) переменными. В замкнутой модели изменения значений переменных во времени определяются внутренним взаимодействием самих переменных. Замкнутая модель может выявить поведение системы без ввода внешней переменной. Пример: информационные системы с обратной связью являются замкнутыми системами. Это самонастраивающиеся системы,

и их характеристики вытекают из внутренней структуры и взаимодействий, которые отражают ввод внешней информации.

Модель, связанная с внешними (экзогенными) переменными, называется открытой.

2.4. Классификация моделей по отношению ко времени

По отношению к временному фактору модели делятся на динамические и статические (рис. 5).

Модель называется статической, если среди параметров, участвующих в ее описании, нет временного параметра. Статическая модель в каждый момент времени дает лишь «фотографию» системы, ее срез. Одним из видов статических моделей являются структурные модели.

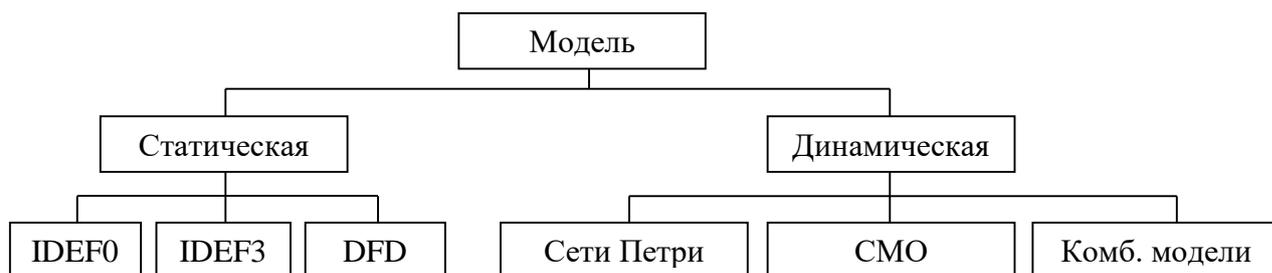


Рис. 5 – Схема классификации математических моделей по отношению ко времени

Динамической моделью называется модель, если среди ее параметров есть временной параметр, т. е. она отображает систему (процессы в системе) во времени.

IDEF0 – Function Modeling – методология функционального моделирования и графическая нотация, предназначенная для формализации и описания бизнес-процессов. Отличительной особенностью IDEF0 является её акцент на соподчинённость объектов. В IDEF0 рассматриваются логические отношения между работами, а не их временная последовательность (Work Flow) [8].

IDEF3 – (англ. Integrated DEfinition for Process Description Capture

Method) – методология моделирования и стандарт документирования процессов, происходящих в системе. Метод документирования технологических процессов представляет собой механизм документирования и сбора информации о процессах. IDEF3 показывает причинно-следственные связи между ситуациями и событиями в понятной эксперту форме, используя структурный метод выражения знаний о том, как функционирует система, процесс или предприятие.

DFD – Data Flow Diagrams – диаграммы потоков данных. Так называется методология графического структурного анализа, описывающая внешние по отношению к системе источники и адресаты данных, логические функции, потоки данных и хранилища данных, к которым осуществляется доступ.

Сети Петри – математический аппарат для моделирования динамических дискретных систем. Сеть Петри есть мультиграф, так как он допускает существование кратных дуг от одной вершины графа к другой. Так как дуги являются направленными, то это ориентированный мультиграф. Вершины графа можно разделить на два множества (позиции и переходы) таким образом, что каждая дуга будет направлена от элемента одного множества (позиций или переходов) к элементу другого множества (переходов или позиций); следовательно, такой граф является двудольным ориентированным мультиграфом.

СМО – система, которая производит обслуживание поступающих в неё требований. Обслуживание требований в СМО производится обслуживающими приборами. Классическая СМО содержит от одного до бесконечного числа приборов.

3. Физическое моделирование

Физическая модель от реальной модели может отличаться как качественно (материалы), так и количественно (масштаб).

3.1. Основные определения и понятия физического моделирования

Физическое моделирование – метод экспериментального изучения различных физических явлений, основанный на их физическом подобии. Метод применяется при следующих условиях [8]:

– исчерпывающе точного математического описания явления на данном уровне развития науки не существует, или такое описание слишком громоздко и требует для расчётов большого объёма исходных данных, получение которых затруднительно;

– воспроизведение исследуемого физического явления в целях эксперимента в реальных масштабах невозможно, нежелательно или слишком дорогостояще (например, цунами).

Метод состоит в создании лабораторной физической модели явления в уменьшенных масштабах, и проведении экспериментов на этой модели. Выводы и данные, полученные в этих экспериментах, распространяются затем на явление в реальных масштабах.

Метод может дать надёжные результаты, лишь в случае соблюдения физического подобия реального явления и модели. Подобие достигается за счёт равенства для модели и реального явления значений критериев подобия – безразмерных чисел, зависящих от физических (в том числе геометрических) параметров, характеризующих явление. Экспериментальные данные, полученные методом физического моделирования распространяются на реальное явление также с учётом критериев подобия.

В широком смысле, любой лабораторный физический эксперимент является моделированием, поскольку в эксперименте наблюдается конкретный случай явления в частных условиях, а требуется получить общие закономерности для всего класса подобных явлений в широком диапазоне

условий. Искусство экспериментатора заключается в достижении физического подобия между явлением, наблюдаемым в лабораторных условиях и всем классом изучаемых явлений.

Некоторые примеры применения метода физического моделирования:

- исследование течений газов и обтекания летательных аппаратов, автомобилей, и т. п. в аэродинамических трубах;
- гидродинамические исследования на уменьшенных моделях кораблей, гидротехнических сооружений и т. п.;
- исследование сейсмоустойчивости зданий и сооружений на макетах;
- изучение устойчивости сложных конструкций, под воздействием сложных силовых нагрузок;
- измерение тепловых потоков и рассеивания тепла в устройствах и системах, работающих в условиях больших тепловых нагрузок;
- изучение стихийных явлений и их последствий.

3.2. Критерий подобия

Критерий подобия – безразмерная величина, составленная из размерных физических параметров, определяющих рассматриваемое физическое явление. Равенство всех однотипных критериев подобия для двух физических явлений и систем – необходимое и достаточное условие физического подобия этих систем [8].

Критерии подобия, представляющие собой отношения одноименных физических параметров системы (например, отношения длин), называются тривиальными и при установлении определяющих критериев подобия обычно не рассматриваются: равенство их для двух систем является определением физического подобия. Нетривиальные безразмерные комбинации, которые можно составить из определяющих параметров, и представляют собой критерии подобия.

Например: «Из каждых 10 яблок 1 гнилое» – отношение количества гнилых яблок к собранным $(1 \text{ яблоко}) / (10 \text{ яблок}) = 0,1 = 10 \%$, и является

тривиальным безразмерным числом.

Всякая новая комбинация из критериев подобия также является критерием подобия, что даёт возможность в каждом конкретном случае выбрать наиболее удобные и характерные критерии. Число определяющих нетривиальных критериев подобия меньше числа определяющих физических параметров с различными размерностями на величину, равную числу определяющих параметров с независимыми размерностями.

Если известны уравнения, описывающие рассматриваемое физическое явление, то критерии подобия для этого явления можно получить, приводя уравнения к безразмерному виду путём введения некоторых характерных значений для каждого из определяющих физических параметров, входящих в систему уравнений. Тогда критерии подобия определяются как безразмерные коэффициенты, появляющиеся перед некоторыми из членов новой, безразмерной системы уравнений. Когда уравнения, описывающие физическое явление, неизвестны, критерии подобия отыскиваются при помощи анализа размерностей, определяющих физические параметры.

Критерий в теоретической механике

Критерий подобия механического движения получается из уравнения, выражающего второй закон Ньютона и называется числом Ньютона:

$$Ne = \frac{Ft^2}{ml},$$

где F – действующая на тело сила,

m – его масса,

t – время,

l – характерный линейный размер.

Критерии подобия в теории упругости

При изучении упругих деформаций конструкции под воздействием

внешних сил основными критериями подобия являются коэффициент Пуассона для материала конструкции:

$$\nu = \left| \frac{\varepsilon_1}{\varepsilon_2} \right|,$$

и критерии

$$\frac{\rho g l}{E}, \frac{F}{E l^2},$$

где $\varepsilon_1 = \frac{\Delta L}{L}$ – относительная продольная деформация,

$\varepsilon_2 = \frac{\Delta d}{d}$ – относительная поперечная деформация,

E – модуль Юнга,

ρ – плотность материала конструкции,

F – характерная внешняя сила,

g – ускорение силы тяжести.

Критерии подобия в гидро-газодинамике

В гидромеханике важнейшими критериями подобия являются:

Число Рейнольдса:

$$Re = \frac{\rho u l}{\mu} = \frac{u l}{\nu}.$$

Определяет, в частности, переход от ламинарного режима к турбулентному.

Число Маха:

$$M = \frac{u}{a^*}$$

Число Фруда:

$$Fr = \frac{u^2}{g l}$$

В этих примерах

ρ – плотность жидкости или газа,

l – характерный размер,

u – скорость течения,

μ – динамический коэффициент вязкости,

$\nu = \mu/\rho$ – кинематический коэффициент вязкости ,

a^* – местная скорость распространения звука в движущейся среде.

Каждый из критериев подобия имеет определенный физический смысл как величина, пропорциональная отношению однотипных физических величин. Так, число Re характеризует отношение инерционных сил при движении жидкости или газа к силам вязкости, а число Fr – отношение инерционных сил к силам тяжести.

Основными критериями подобия процессов теплопередачи между жидкостью (газом) и обтекаемым телом являются число Прандтля $Pr = \nu/a = \mu C_p/\lambda$, число Нуссельта $Nu = \alpha L/\lambda$, число Грасгофа $Gr = bgl^3\Delta T/\nu^2$, а также число Пекле $Pe = ul/a$ и число Стэнтона $St = a/(\rho C_p u)$. Здесь α – коэффициент теплоотдачи, λ – коэффициент теплопроводности, C_p – удельная теплоемкость жидкости или газа при постоянном давлении, $a = \lambda/\rho C_p$ – коэффициент температуропроводности, b – коэффициент объемного расширения, ΔT – разность температур поверхности тела и жидкости (газа). Два последних числа связаны с предыдущими соотношениями: $Pe = Pr Re$, $St = Nu/Pe$.

3.3. Динамическое подобие и моделирование явлений

Теория размерности и подобия имеет большое значение при моделировании различных явлений. Моделирование это есть замена изучения интересующего нас явления в натуре изучением аналогичного явления на модели меньшего или большего масштаба, обычно в специальных лабораторных условиях. Основной смысл моделирования заключается в том, чтобы по результатам опытов с моделями можно было давать необходимые ответы о характере эффектов и о различных величинах, связанных с явлением

в натуральных условиях [9].

В большинстве случаев моделирование основано на рассмотрении физически подобных явлений. Изучение интересующего нас натурального явления мы заменяем изучением физически подобного явления, которое удобнее и выгоднее осуществить. Механическое или вообще физическое подобие можно рассматривать как обобщение геометрического подобия. Две геометрические фигуры подобны, если отношения всех соответственных длин одинаковы. Если известен коэффициент подобия – масштаб, то простым умножением на величину масштаба размеров одной геометрической фигуры получаются размеры другой, ей подобной геометрической фигуры.

Существуют различные способы определения механического или физического подобия. Ниже мы даем определение физического подобия в такой форме, которая необходима для практики и удобна для непосредственных приложений.

Два явления подобны, если по заданным характеристикам одного можно получить характеристики другого простым пересчетом, который аналогичен переходу от одной системы единиц измерения к другой системе.

Для осуществления пересчета необходимо знать «переходные масштабы».

Численные характеристики для двух различных, но подобных явлений можно рассматривать как численные характеристики одного и того же явления, выраженные в двух различных системах единиц измерения. Для всякой совокупности подобных явлений все безразмерные характеристики (безразмерные комбинации из размерных величин) имеют одинаковое численное значение. Нетрудно видеть, что обратное заключение также справедливо, т. е. если все безразмерные характеристики для двух движений одинаковы, то движения подобны.

Совокупность механически подобных движений определяет собой режим движения.

Подобие двух явлений иногда можно понимать в более широком

смысле, принимая, что указанное выше определение относится только к некоторой специальной системе характеристик, полностью определяющей явление и позволяющей находить любые другие характеристики, которые, однако, нельзя получить простым умножением на соответствующие масштабы при переходе от одного к другому «подобному» явлению. Например, в этом смысле два любых эллипса можно считать подобными при использовании декартовых координат, направленных по главным осям эллипсов. Указанным пересчетом можно получить декартовы координаты точек любого эллипса через координаты точек какого-либо одного эллипса (аффинное подобие).

Для сохранения подобия при моделировании необходимо соблюдать некоторые условия. Однако на практике сплошь и рядом эти условия, обеспечивающие подобие явления в целом, не выполняются, и тогда встает вопрос о величине погрешностей (масштабном эффекте), которые возникают при переносе на натуру результатов, полученных на модели.

После установления системы параметров, определяющих выделенный класс явлений, нетрудно установить условия подобия двух явлений.

В самом деле, пусть явление определяется n параметрами, часть из которых может быть безразмерными, а некоторые являются размерными физическими постоянными. Допустим далее, что размерности переменных параметров и физических постоянных выражены через k величин с независимыми размерностями ($k \leq n$).

В общем случае очевидно, что из n величин можно составить не более $n - k$ независимых безразмерных комбинаций. Все безразмерные характеристики явления можно рассматривать как функции от этих $n - k$ независимых безразмерных комбинаций, составленных из определяющих параметров. Следовательно, среди всех безразмерных величин, составленных из характеристик явления, всегда можно указать некоторую базу, т. е. систему безразмерных величин, которые определяют собой все остальные величины.

Определенный соответствующей постановкой задачи класс явлений содержит явления, вообще не подобные между собой. Выделение из него

подкласса подобных явлений осуществляется с помощью следующего условия:

Необходимым и достаточным условием подобия двух явлений будет постоянство численных значений безразмерных комбинаций, образующих базу. Условия о постоянстве базы отвлеченных параметров, составленных из заданных определяющих явление величин, называются критериями подобия.

Если условия подобия выполнены, то для фактического расчета всех характеристик в натуре по данным о размерных характеристиках на модели необходимо знать переходные масштабы для всех соответствующих величин.

Если явление определяется n параметрами, из которых k имеют независимые размерности, то для k величин с независимыми размерностями переходные масштабы могут быть произвольными и их нужно задать или определить условиями задачи, а при экспериментах – из опытов. Переходные масштабы для всех остальных размерных величин легко получить из формул размерности для каждой размерной величины через размерности k независимых, для которых масштабы определены опытом или постановкой задачи.

В задаче об установившемся поступательном плоскопараллельном движении тела в несжимаемой вязкой жидкости все безразмерные величины определяются двумя параметрами: углом атаки α и числом Рейнольдса R . Условия физического подобия – критерии подобия – представляются соотношениями

$$\alpha = \text{const} \text{ и } R = v dp / \mu = \text{const}.$$

При моделировании этого явления результаты опытов с моделью можно переносить на натуру только для одинаковых α и R . Первое условие всегда легко осуществить на практике. Труднее удовлетворить второму условию ($R = \text{const}$), особенно в тех случаях, когда обтекаемое тело имеет большие размеры, как, например, крыло самолета. Если модель меньше натуре, то для сохранения величины числа Рейнольдса необходимо либо увеличивать

скорость обтекающего потока, что практически обычно неосуществимо, либо существенно изменить плотность и вязкость жидкости.

На практике эти обстоятельства вносят большие затруднения при изучении аэродинамического сопротивления. Необходимость постоянства числа Рейнольдса привела к постройке гигантских аэродинамических труб, в которых можно производить продувки самолетов в натуру (рис. 6 и 7), а также труб закрытого типа, в которых циркулирует с большой скоростью сжатый, т. е. более плотный воздух (рис. 8).

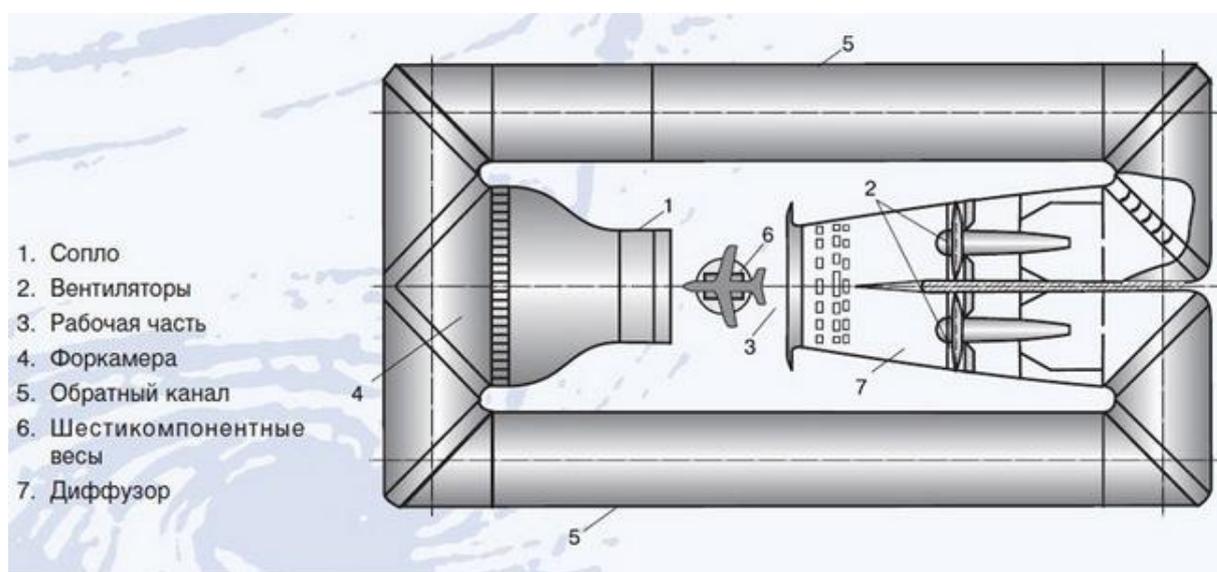


Рис. 6 – Вид сверху аэродинамической трубы ЦАГИ Т-101 для натурального испытания

Специальные теоретические и экспериментальные исследования показывают, что в ряде случаев для обтекаемых тел число Рейнольдса заметно влияет только на безразмерный коэффициент лобового сопротивления и иногда очень слабо влияет на безразмерный коэффициент подъемной силы и на некоторые другие величины, играющие весьма важную роль в различных практических вопросах. Следовательно, различие в значении числа Рейнольдса для модели и явления в натуре в некоторых вопросах не является существенным.



Рис. 7 – Фото самолет Миг-29 в аэродинамической трубе Т-101

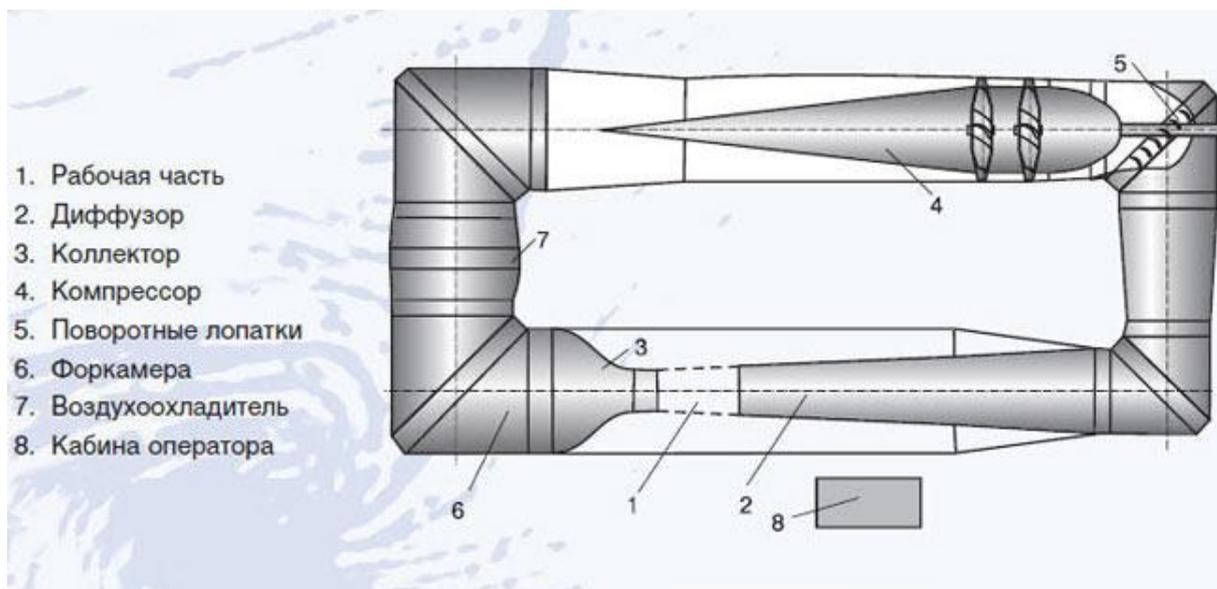


Рис. 8 – аэродинамическая труба Т-106 замкнутого типа, непрерывного действия, переменной плотности

Мы указали условия подобия для движения крыла без учета свойства сжимаемости воздуха, которое несущественно при скоростях, малых по сравнению со скоростью звука.

4. Математическое моделирование

Математическое моделирование – исследование объекта с помощью математической модели, которая воспроизводит наиболее важные черты оригинала.

Математическая модель – формальное описание объекта или явления при помощи функциональных или логических операторных соотношений, алгебраических интегральных и других уравнений, которые могут быть представлены, как в незамкнутой, так и в замкнутой (разрешенной) форме.

4.1. Классификация математических моделей

Математические модели классифицируются [4]:

- по принадлежности к иерархическому уровню;
- характеру отображаемых свойств объекта;
- способу представления свойств объекта;
- способу получения модели;
- форме представления свойств объекта;
- в зависимости от сложности объекта моделирования.

По принадлежности к иерархическому уровню математические модели делятся на модели микроуровня, макроуровня, метауровня (см. рис. 9).



Рис. 9 – Схема классификации математических моделей по принадлежности к иерархическому уровню

Математические модели на микроуровне процесса отражают физические процессы, протекающие, например, при резании металлов. Они описывают процессы на уровне перехода (прохода).

Математические модели на макроуровне процесса описывают технологические процессы.

Математические модели на метауровне процесса описывают технологические системы (участки, цехи, предприятие в целом).

По характеру отображаемых свойств объекта модели можно классифицировать на структурные и функциональные (рис. 10).

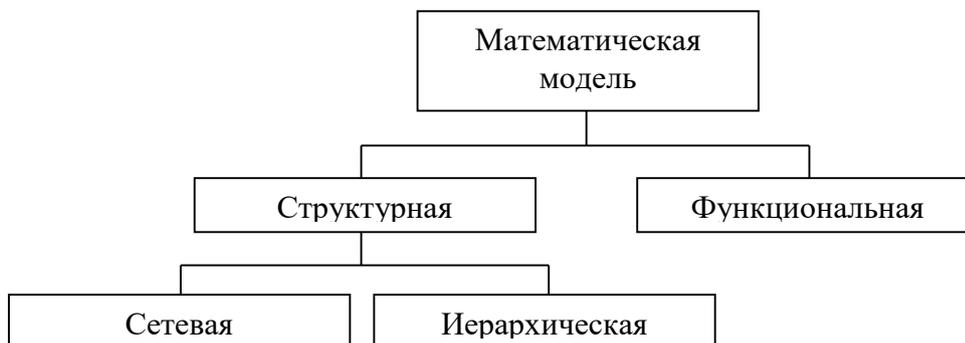


Рис. 10 – Схема классификации математических моделей по характеру отображаемых свойств объекта

Модель структурная, – если она представима структурой данных или структурами данных и отношениями между ними; например, структурной моделью может служить описание (табличное, графовое, функциональное или другое) трофической структуры экосистемы. В свою очередь, структурная модель может быть иерархической или сетевой.

Модель иерархическая (древовидная), – если представима некоторой иерархической структурой (деревом); например, для решения задачи нахождения маршрута в дереве поиска можно построить древовидную модель.

Модель сетевая, – если она представима некоторой сетевой структурой. Эти операции можно представить в виде сетевой модели.

Модель функциональная, – если она представима в виде системы функциональных соотношений. Например, закон Ньютона и модель производства товаров – функциональные.

По способу представления свойств объекта (рис. 11) модели делятся на

аналитические, численные, алгоритмические и имитационные [10].

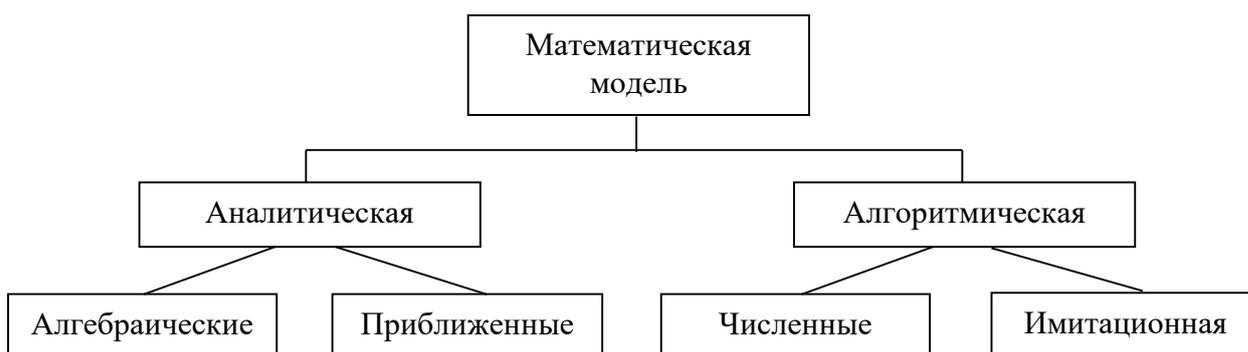


Рис. 11 – Схема классификации математических моделей по способу представления свойств объекта

Аналитические математические модели представляют собой явные математические выражения выходных параметров как функций от параметров входных и внутренних и имеют единственные решения при любых начальных условиях. Например, процесс резания (точения) с точки зрения действующих сил представляет собой аналитическую модель.

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 [2]:

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

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 значение функции с различной степенью точности. Например, учет первых шести членов ряда в разложении показательной функции обеспечивает 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.

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

По способу получения модели делятся на теоретические и эмпирические (рис. 12).

Теоретические математические модели создаются в результате исследования объектов (процессов) на теоретическом уровне. Например, существуют выражения для сил резания, полученные на основе обобщения физических законов. Но они неприемлемы для практического использования, т. к. очень громоздки и не совсем адаптированы к реальным процессам обработки материалов.

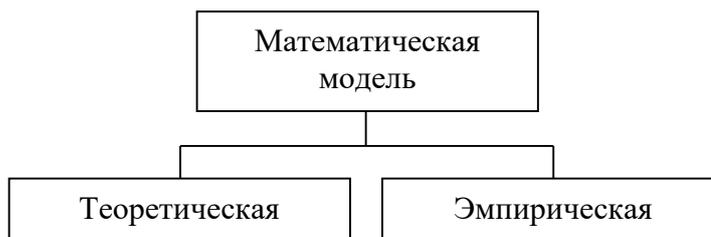


Рис. 12 – Схема классификации математических моделей по способу получения модели

Эмпирические математические модели создаются в результате проведения экспериментов (изучения внешних проявлений свойств объекта с помощью измерения его параметров на входе и выходе) и обработки их результатов методами математической статистики.

По форме представления свойств объекта модели делятся на логические, теоретико-множественные и графовые (рис. 13).



Рис. 13 – Схема классификации математических моделей по форме представления свойств объекта

Модель логическая, если она представима предикатами, логическими функциями, например, совокупность двух логических функций может

служить математической моделью одноразрядного сумматора.

Модель теоретико-множественная, – если она представима с помощью некоторых множеств и отношений принадлежности к ним и между ними.

Модель графовая, – если она представима графом или графами и отношениями между ними.

На рис. 14 приведена классификация математических моделей в зависимости от сложности объекта моделирования



Рис. 14 – Схема классификации 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. 14). 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.

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 по отношению к B системы может отличаться от связи между элементами B по отношению к A .

4.2. Основные понятия математического моделирования

Решение практических задач математическими методами последовательно осуществляется путем формулировки задачи (разработки математической модели), выбора метода исследования полученной математической модели, анализа полученного математического результата. Математическая формулировка задачи обычно представляется в виде геометрических образов, функций, систем уравнений и т.п. Описание объекта (явления) может быть представлено с помощью непрерывной или дискретной, детерминированной или стохастической и другими математическими формами [10].

Теория математического моделирования обеспечивает выявление закономерностей протекания различных явлений окружающего мира или работы систем и устройств путем их математического описания и моделирования без проведения натуральных испытаний. При этом используются положения и законы математики, описывающие моделируемые явления, системы или устройства на некотором уровне их идеализации.

Математическая модель представляет собой формализованное описание системы (или операции) на некотором абстрактном языке, например, в виде совокупности математических соотношений или схемы алгоритма, т. е. такое математическое описание, которое обеспечивает имитацию работы систем или устройств на уровне, достаточно близком к их реальному поведению, получаемому при натуральных испытаниях систем или устройств. Любая математическая модель описывает реальный объект, явление или процесс с некоторой степенью приближения к действительности. Вид математической модели зависит как от природы реального объекта, так и от задач исследования.

Математическое моделирование общественных, экономических, биологических и физических явлений, объектов, систем и различных устройств является одним из важнейших средств познания природы и проектирования самых разнообразных систем и устройств. Известны примеры

эффективного использования моделирования в создании ядерных технологий, авиационных и аэрокосмических систем, в прогнозе атмосферных и океанических явлений, погоды и т.д. [12]

Однако для таких серьезных сфер моделирования нередко нужны суперкомпьютеры и годы работы крупных коллективов ученых по подготовке данных для моделирования и его отладки. Тем не менее, и в этом случае математическое моделирование сложных систем и устройств не только экономит средства на проведение исследований и испытаний, но и может устранить экологические катастрофы – например, позволяет отказаться от испытаний ядерного и термоядерного оружия в пользу его математического моделирования или испытаний аэрокосмических систем перед их реальными полетами.

Между тем математическое моделирование на уровне решения более простых задач, например, из области механики, электротехники, электроники, радиотехники и многих других областей науки и техники в настоящее время стало доступным выполнять на современных ПК. А при использовании обобщенных моделей становится возможным моделирование и достаточно сложных систем, например, телекоммуникационных систем и сетей, радиолокационных или радионавигационных комплексов.

Целью математического моделирования является анализ реальных процессов (в природе или технике) математическими методами. В свою очередь, это требует формализации процесса математической модели, подлежащего исследованию. Модель может представлять собой математическое выражение, содержащее переменные, поведение которых аналогично поведению реальной системы. Модель может включать элементы случайности, учитывающие вероятности возможных действий двух или большего числа «игроков», как, например, в теории игр; либо она может представлять реальные переменные параметры взаимосвязанных частей действующей системы.

Математическое моделирование для исследования характеристик

систем можно разделить на аналитическое, имитационное и комбинированное. В свою очередь, математические модели делятся на имитационные и аналитические.

Аналитическое моделирование

Для аналитического моделирования характерно, что процессы функционирования системы записываются в виде некоторых функциональных соотношений (алгебраических, дифференциальных, интегральных уравнений) [13]. Аналитическая модель может быть исследована следующими методами:

- 1) аналитическим, когда стремятся получить в общем виде явные зависимости для характеристик систем;
- 2) численным, когда не удается найти решение уравнений в общем виде и их решают для конкретных начальных данных;
- 3) качественным, когда при отсутствии решения находят некоторые его свойства.

Аналитические модели удается получить только для сравнительно простых систем. Для сложных систем часто возникают большие математические проблемы. Для применения аналитического метода идут на существенное упрощение первоначальной модели. Однако исследование на упрощенной модели помогает получить лишь ориентировочные результаты. Аналитические модели математически верно отражают связь между входными и выходными переменными и параметрами. Но их структура не отражает внутреннюю структуру объекта.

При аналитическом моделировании его результаты представляются в виде аналитических выражений [12]. Например, подключив RC -цепь к источнику постоянного напряжения E (R , C и E – компоненты данной модели), мы можем составить аналитическое выражение для временной зависимости напряжения $u(t)$ на конденсаторе C :

$$RC \frac{du(t)}{dt} + u(t) = E. \quad (1)$$

Это линейное дифференциальное уравнение (ДУ) и является аналитической моделью данной простой линейной цепи. Его аналитическое решение, при начальном условии $u(0)=0$, означающем разряженный конденсатор C в момент начала моделирования, позволяет найти искомую зависимость – в виде формулы:

$$u(t) = E(1 - e^{-t/RC}). \quad (2)$$

Однако даже в этом простейшем примере требуются определенные усилия для решения ДУ (1) или для применения систем компьютерной математики с символьными вычислениями – систем компьютерной алгебры. Для данного вполне тривиального случая решение задачи моделирования линейной RC -цепи дает аналитическое выражение (2) достаточно общего вида – оно пригодно для описания работы цепи при любых номиналах компонентов R , C и E , и описывает экспоненциальный заряд конденсатора C через резистор R от источника постоянного напряжения E .

Безусловно, нахождение аналитических решений при аналитическом моделировании оказывается исключительно ценным для выявления общих теоретических закономерностей простых линейных цепей, систем и устройств. Однако его сложность резко возрастает по мере усложнения воздействий на модель и увеличения порядка и числа уравнений состояния, описывающих моделируемый объект. Можно получить более или менее обзримые результаты при моделировании объектов второго или третьего порядка, но уже при большем порядке аналитические выражения становятся чрезмерно громоздкими, сложными и трудно осмысляемыми. Например, даже простой электронный усилитель зачастую содержит десятки компонентов. Тем не менее, многие современные системы компьютерной математики,

например, системы символьной математики Maple, Mathematica или среда MATLAB, способны в значительной мере автоматизировать решение сложных задач аналитического моделирования.

Одной из разновидностей моделирования является численное моделирование [14, 15], которое заключается в получении необходимых количественных данных о поведении систем или устройств каким-либо подходящим численным методом, таким как методы Эйлера или Рунге-Кутты [15]. На практике моделирование нелинейных систем и устройств с использованием численных методов оказывается намного более эффективным, чем аналитическое моделирование отдельных частных линейных цепей, систем или устройств. Например, для решения ДУ (1) или систем ДУ в более сложных случаях решение в аналитическом виде не получается, но по данным численного моделирования можно получить достаточно полные данные о поведении моделируемых систем и устройств, а также построить графики описывающих это поведение зависимостей.

Имитационное моделирование

При имитационном моделировании реализующий модель алгоритм воспроизводит процесс функционирования системы во времени. Имитируются элементарные явления, составляющие процесс, с сохранением их логической структуры и последовательности протекания во времени [16].

Основным преимуществом имитационных моделей по сравнению с аналитическими является возможность решения более сложных задач.

Имитационные модели позволяют легко учитывать наличие дискретных или непрерывных элементов, нелинейные характеристики, случайные воздействия и др. Поэтому этот метод широко применяется на этапе проектирования сложных систем. Основным средством реализации имитационного моделирования служит ЭВМ, позволяющая осуществлять цифровое моделирование систем и сигналов.

В связи с этим определим словосочетание «компьютерное

моделирование», которое все чаще используется в литературе. Будем полагать, что компьютерное моделирование - это математическое моделирование с использованием средств вычислительной техники. Соответственно, технология компьютерного моделирования предполагает выполнение следующих действий [13]:

- 1) определение цели моделирования;
- 2) разработка концептуальной модели;
- 3) формализация модели;
- 4) программная реализация модели;
- 5) планирование модельных экспериментов;
- 6) реализация плана эксперимента;
- 7) анализ и интерпретация результатов моделирования.

Содержание первых двух этапов практически не зависит от математического метода, положенного в основу моделирования (и даже наоборот – их результат определяет выбор метода). А вот реализация остальных пяти этапов существенно различается для аналитического и имитационного моделирования.

При имитационном моделировании используемая математическая модель воспроизводит алгоритм («логику») функционирования исследуемой системы во времени при различных сочетаниях значений параметров системы и внешней среды.

Примером простейшей аналитической модели может служить уравнение прямолинейного равномерного движения. При исследовании такого процесса с помощью имитационной модели должно быть реализовано наблюдение за изменением пройденного пути с течением времени.

Очевидно, в одних случаях более предпочтительным является аналитическое моделирование, в других - имитационное (или сочетание того и другого). Чтобы выбор был удачным, необходимо ответить на два вопроса.

С какой целью проводится моделирование?

К какому классу может быть отнесено моделируемое явление?

Ответы на оба эти вопроса могут быть получены в ходе выполнения двух первых этапов моделирования.

Имитационные модели не только по свойствам, но и по структуре соответствуют моделируемому объекту. При этом имеется однозначное и явное соответствие между процессами, получаемыми на модели, и процессами, протекающими на объекте. Недостатком имитационного моделирования является большое время решения задачи для получения хорошей точности.

Результаты имитационного моделирования работы стохастической системы являются реализациями случайных величин или процессов. Поэтому для нахождения характеристик системы требуется многократное повторение и последующая обработка данных. Чаще всего в этом случае применяется разновидность имитационного моделирования - статистическое моделирование (или метод Монте-Карло), т.е. воспроизведение в моделях случайных факторов, событий, величин, процессов, полей [17-19].

По результатам статистического моделирования определяют оценки вероятностных критериев качества, общих и частных, характеризующих функционирование и эффективность управляемой системы. Статистическое моделирование широко применяется для решения научных и прикладных задач в различных областях науки и техники. Методы статистического моделирования широко применяются при исследовании сложных динамических систем, оценке их функционирования и эффективности. Заключительный этап статистического моделирования основан на математической обработке полученных результатов. Здесь используют методы математической статистики (параметрическое и непараметрическое оценивание, проверку гипотез) [20]. Примером параметрической оценки является выборочное среднее показателя эффективности. Среди непараметрических методов большое распространение получил метод гистограмм.

Рассмотренная схема основана на многократных статистических

испытаниях системы и методах статистики независимых случайных величин.

Эта схема является далеко не всегда естественной на практике и оптимальной по затратам. Сокращение времени испытания систем может быть достигнуто за счет использования более точных методов оценивания. Как известно из математической статистики, наибольшую точность при заданном объеме выборки имеют эффективные оценки [21, 22]. Оптимальная фильтрация и метод максимального правдоподобия дают общий метод получения таких оценок [22].

В задачах статистического моделирования обработка реализаций случайных процессов необходима не только для анализа выходных процессов.

Весьма важен также и контроль характеристик входных случайных воздействий. Контроль заключается в проверке соответствия распределений генерируемых процессов заданным распределениям. Эта задача часто формулируется как задача проверки гипотез [21].

Общей тенденцией моделирования с использованием ЭВМ у сложных управляемых систем является стремление к уменьшению времени моделирования, а также проведение исследований в реальном масштабе времени. Вычислительные алгоритмы удобно представлять в рекуррентной форме, допускающей их реализацию в темпе поступления текущей информации [22-24].

4.3. Основные этапы математического моделирования

Первым этапом математического моделирования является постановка задачи, определение объекта и целей исследования, задание критериев (признаков) изучения объектов и управления ими. Неправильная или неполная постановка задачи может свести на нет результаты всех последующих этапов [11].

Вторым этапом моделирования является выбор типа математической модели, что является важнейшим моментом, определяющим направление всего исследования. Обычно последовательно строится несколько моделей.

Сравнение результатов их исследования с реальностью позволяет установить наилучшую из них. На этапе выбора типа математической модели при помощи анализа данных поискового эксперимента устанавливаются: линейность или нелинейность, динамичность или статичность, стационарность или не стационарность, а также степень детерминированности исследуемого объекта или процесса.

Процесс выбора математической модели объекта заканчивается ее предварительным контролем, который также является первым шагом на пути к исследованию модели. При этом осуществляются следующие виды контроля (проверки): размерностей; порядков; характера зависимостей; экстремальных ситуаций; граничных условий; математической замкнутости; физического смысла; устойчивости модели [25].

Контроль размерностей сводится к проверке выполнения правила, согласно которому приравниваться и складываться могут только величины одинаковой размерности.

Контроль порядков величин направлен на упрощение модели. При этом определяются порядки складываемых величин и явно малозначительные слагаемые отбрасываются.

Анализ характера зависимостей сводится к проверке направления и скорости изменения одних величин при изменении других. Направления и скорость, вытекающие из математической модели, должны соответствовать физическому смыслу задачи.

Анализ экстремальных ситуаций сводится к проверке наглядного смысла решения при приближении параметров модели к нулю или бесконечности.

Контроль граничных условий состоит в том, что проверяется соответствие математической модели граничным условиям, вытекающим из смысла задачи. При этом проверяется, действительно ли граничные условия поставлены и учтены при построении искомой функции и что эта функция на самом деле удовлетворяет таким условиям.

Анализ математической замкнутости сводится к проверке того, что математическая модель дает однозначное решение.

Анализ физического смысла сводится к проверке физического содержания промежуточных соотношений, используемых при построении математической модели.

Проверка устойчивости модели состоит в проверке того, что варьирование исходных данных в рамках имеющихся данных о реальном объекте не приведет к существенному изменению решения.

Понятие о вычислительном эксперименте

В настоящее время основным способом исследования математической модели и проверки ее качественных показателей служит вычислительный эксперимент.

Вычислительным экспериментом называется методология и технология исследований, основанные на применении прикладной математики и ЭВМ как технической базы при использовании математической модели. Вычислительный эксперимент основывается на создании математической модели изучаемых объектов, которые формируются с помощью некоторой особой математической структуры, способной отражать свойства объекта, проявляемые им в различных экспериментальных условиях, и включает в себя следующие этапы [25].

1. Для исследуемого объекта строится модель, обычно сначала физическая, фиксирующая разделение всех действующих в рассматриваемом явлении факторов на главные и второстепенные, которые на данном этапе исследования отбрасываются; одновременно формулируются допущения и условия применимости модели, границы, в которых будут справедливы полученные результаты; модель записывается в математических терминах, как правило, в виде дифференциальных или интегро-дифференциальных уравнений; создание математической модели проводится специалистами, хорошо знающими данную область естествознания или техники, а также

математиками, представляющими себе возможности решения математической задачи [15].

2. Разрабатывается метод решения сформулированной математической задачи. Эта задача представляется в виде совокупности алгебраических формул, по которым должны вестись вычисления и условия, показывающие последовательность применения этих формул; набор этих формул и условий носит название вычислительного алгоритма. Вычислительный эксперимент имеет многовариантный характер, так как решения поставленных задач часто зависят от многочисленных входных параметров. Тем не менее, каждый конкретный расчет в вычислительном эксперименте проводится при фиксированных значениях всех параметров. Между тем в результате такого эксперимента часто ставится задача определения оптимального набора параметров. Поэтому при создании оптимальной установки приходится проводить большое число расчетов однотипных вариантов задачи, отличающихся значением некоторых параметров. В связи с этим при организации вычислительного эксперимента можно использовать эффективные численные методы,

3. Разрабатываются алгоритм и программа решения задачи на ЭВМ. Программирование решений определяется теперь не только искусством и опытом исполнителя, а перерастает в самостоятельную науку со своими принципиальными подходами.

4. Проведение расчетов на ЭВМ. Результат получается в виде некоторой цифровой информации, которую далее необходимо будет расшифровать. Точность информации определяется при вычислительном эксперименте достоверностью модели, положенной в основу эксперимента, правильностью алгоритмов и программ (проводятся предварительные «тестовые» испытания).

5. Обработка результатов расчетов, их анализ и выводы [26]. На этом этапе могут возникнуть необходимость уточнения математической модели (усложнения или, наоборот, упрощения), предложения по созданию упрощенных инженерных способов решения и формул, дающих возможности

получить необходимую информацию более простым способом.

Вычислительный эксперимент приобретает исключительное значение в тех случаях, когда натурные эксперименты и построение физической модели оказываются невозможными. Особенно ярко можно проиллюстрировать значение вычислительного эксперимента при исследовании влияния городской застройки на параметры распространения радиосигнала [27]. В связи с интенсивным развитием систем мобильной связи данная задача в настоящее время является особенно актуальной. С целью снижения затрат при частотно-территориальном планировании производится оптимизация частотно-территориального плана с учетом таких факторов как рельеф местности, конфигурация городской застройки, атмосферные воздействия. Кроме этого, с учетом динамичности развития города необходимо постоянное уточнение соответствующих моделей. То, что принято называть уровнем сигнала (средняя напряженность электромагнитного поля) представляет собой результат сложного взаимодействия физических процессов, протекающих при распространении сигнала: прохождение сигнала сквозь здания и сооружения; воздействие на сигнал помех искусственного и естественного происхождения; атмосферная рефракция сигнала; отражения сигнала от зданий и от земной поверхности; потери энергии сигнала в осадках и др. В данном случае окружающую среду можно исследовать, строя соответствующую математическую модель, которая должна позволять предсказывать уровень сигнала при заданной конфигурации застройки, рельефе местности, погодных условиях и т. п. Масштабы среды распространения сигнала настолько грандиозны, что эксперимент даже в одном каком-то регионе требует существенных затрат.

Таким образом, глобальный эксперимент по исследованию распространения сигнала возможен, но не натурный, а вычислительный, проводящий исследования не реальной системы (окружающей среды), а ее математической модели.

В науке и технике известно немало областей, в которых вычислительный

эксперимент оказывается единственно возможным при исследовании сложных систем.

Пригодность математической модели для решения задач исследования характеризуется тем, в какой степени она обладает так называемыми целевыми свойствами, основными из которых являются адекватность, устойчивость и чувствительность.

Оценка адекватности

В общем случае под адекватностью понимают степень соответствия модели тому реальному явлению или объекту, для описания которого она строится. Вместе с тем, создаваемая модель ориентирована, как правило, на исследование определенного подмножества свойств этого объекта. Поэтому можно считать, что адекватность модели определяется степенью ее соответствия не столько реальному объекту, сколько целям исследования. В наибольшей степени это утверждение справедливо относительно моделей проектируемых систем (т. е. в ситуациях, когда реальная система вообще не существует).

Тем не менее, во многих случаях полезно иметь формальное подтверждение (или обоснование) адекватности разработанной модели. Один из наиболее распространенных способов такого обоснования - использование методов математической статистики [20, 28, 29]. Суть этих методов заключается в проверке выдвинутой гипотезы (в данном случае - об адекватности модели) на основе некоторых статистических критериев. При этом следует заметить, что при проверке гипотез методами математической статистики необходимо иметь в виду, что статистические критерии не могут доказать ни одной гипотезы – они могут лишь указать на отсутствие опровержения.

Итак, каким же образом можно оценить адекватность разработанной модели реально существующей системе?

Процедура оценки основана на сравнении измерений на реальной

системе и результатов экспериментов на модели и может проводиться различными способами. Наиболее распространенные из них [13, 25]:

- по средним значениям откликов модели и системы;
- по дисперсиям отклонений откликов модели от среднего значения откликов системы;
- по максимальному значению относительных отклонений откликов модели от откликов системы.

Названные способы оценки достаточно близки между собой, по сути, поэтому ограничимся рассмотрением первого из них. При этом способе проверяется гипотеза о близости среднего значения наблюдаемой переменной Y среднему значению отклика реальной системы Y^* .

В результате N_0 опытов на реальной системе получают множество значений (выборку) Y^* . Выполнив N_M экспериментов на модели, также получают множество значений наблюдаемой переменной Y .

Затем вычисляются оценки математического ожидания и дисперсии откликов модели и системы, после чего выдвигается гипотеза о близости средних значений величин Y^* и Y (в статистическом смысле). Основой для проверки гипотезы является t -статистика (распределение Стьюдента) [20].

Ее значение, вычисленное по результатам испытаний, сравнивается с критическим значением t_{kp} , взятым из справочной таблицы [29]. Если выполняется неравенство $t_n < t_{kp}$, то гипотеза принимается. Необходимо еще раз подчеркнуть, что статистические методы применимы только в том случае, если оценивается адекватность модели существующей системе. На проектируемой системе провести измерения, естественно, не представляется возможным. Единственный способ преодолеть это препятствие заключается в том, чтобы принять в качестве эталонного объекта концептуальную модель проектируемой системы. Тогда оценка адекватности программно реализованной модели заключается в проверке того, насколько корректно она отражает концептуальную модель.

Оценка устойчивости

При проверке адекватности модели как существующей, так и проектируемой системы реально может быть использовано лишь ограниченное подмножество всех возможных значений входных параметров (рабочей нагрузки и внешней среды). В связи с этим для обоснования достоверности получаемых результатов моделирования большое значение имеет проверка устойчивости модели [13]. В теории моделирования – это понятие трактуется следующим образом.

Устойчивость модели – это ее способность сохранять адекватность при исследовании эффективности системы на всем возможном диапазоне рабочей нагрузки, а также при внесении изменений в конфигурацию системы.

Каким образом может быть оценена устойчивость модели? Универсальной процедуры проверки устойчивости модели не существует. Разработчик вынужден прибегать к методам «для данного случая», частичным тестам и здравому смыслу. Часто полезна апостериорная проверка. Она состоит в сравнении результатов моделирования и результатов измерений на системе после внесения в нее изменений. Если результаты моделирования приемлемы, уверенность в устойчивости модели возрастает.

В общем случае можно утверждать, что чем ближе структура модели структуре системы и чем выше степень детализации, тем устойчивее модель. Устойчивость результатов моделирования может быть также оценена методами математической статистики. Здесь уместно вспомнить основную задачу математической статистики, которая заключается в том, чтобы проверить гипотезу относительно свойств некоторого множества элементов, называемого генеральной совокупностью, оценивая свойства какого-либо подмножества генеральной совокупности (то есть выборки). В генеральной совокупности исследователя обычно интересует некоторый признак, который обусловлен случайностью и может иметь качественный или количественный характер.

В данном случае именно устойчивость результатов моделирования

можно рассматривать как признак, подлежащий оценке. Для проверки гипотезы об устойчивости результатов может быть использован критерий Уилкоксона, который служит для проверки того, относятся ли две выборки к одной и той же генеральной совокупности (т. е. обладают ли они одним и тем же статистическим признаком) [20, 29]. Например, в двух партиях некоторой продукции измеряется определенный признак и требуется проверить гипотезу о том, что этот признак имеет в обеих партиях одинаковое распределение; другими словами, необходимо убедиться, что технологический процесс от партии к партии изменяется несущественно. При статистической оценке устойчивости модели соответствующая гипотеза может быть сформулирована следующим образом: при изменении входной (рабочей) нагрузки или структуры математической модели закон распределения результатов моделирования остается неизменным.

Оценка чувствительности

Очевидно, что устойчивость является положительным свойством модели. Однако если изменение входных воздействий или параметров модели (в некотором заданном диапазоне) не отражается на значениях выходных параметров, то польза от такой модели невелика. В связи с этим возникает задача оценивания чувствительности модели к изменению параметров рабочей нагрузки и внутренних параметров самой системы [13].

Такую оценку проводят по каждому параметру модели в отдельности. Основана она на том, что обычно диапазон возможных изменений параметра известен. Данные, полученные при оценке чувствительности модели, могут быть использованы, в частности, при планировании экспериментов: большее внимание должно уделяться тем параметрам, по которым модель является более чувствительной [16].

4.4. 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 [2].

В некоторых случаях необходимо учитывать влияние различных видов излучения, воздействия гравитационных и электромагнитных полей, предыстории 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. 15.

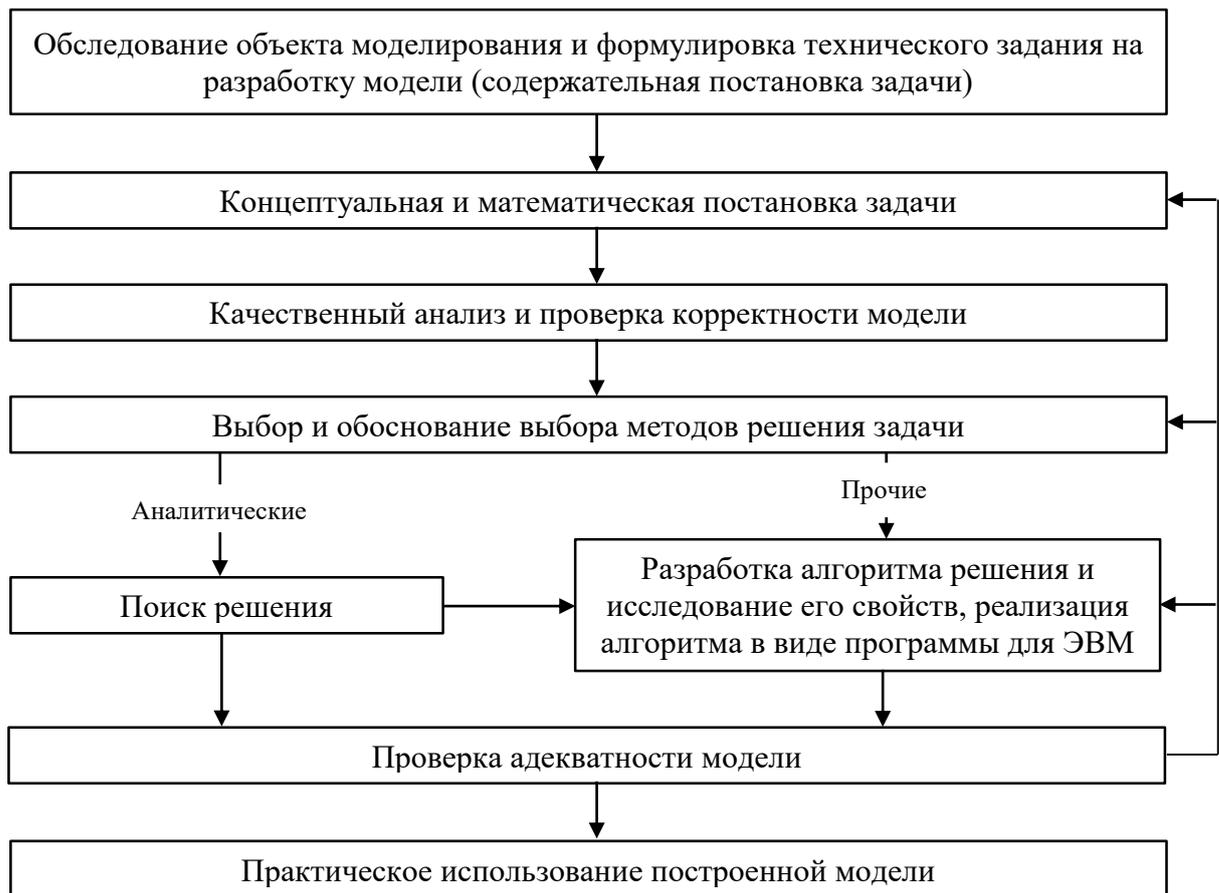


Рис. 15 – Этапы построения математической модели

5. Имитационное моделирование

5.1. Понятие статистического эксперимента

Эксперимент, проводимый с помощью аналитической модели, ничем не отличается от традиционной для математики процедуры вычислений: для конкретных значений параметров модели рассчитывается соответствующая им величина показателя эффективности. И сколько бы раз мы ни повторяли эту процедуру, результат всегда будет один и тот же (при неизменных значениях параметров) [30].

Совсем по-иному обстоит дело в том случае, когда для оценки эффективности системы используется имитационное моделирование, поскольку имитационный эксперимент представляет собой наблюдение за поведением модели под влиянием входных воздействий. При этом часть из них (а может быть, и все) носят случайный характер. В результате такого наблюдения исследователь получает набор экспериментальных данных, на основе которых могут быть оценены характеристики системы. Очевидно, что аналитические модели для проведения имитационного эксперимента не годятся, и здесь нужна специальная имитационная модель.

Такая модель должна отвечать двум основным требованиям:

- отражать логику функционирования исследуемой системы во времени;
- обеспечивать возможность проведения статистического эксперимента.

Методы описания в модели логики функционирования системы будут рассмотрены немного позже. Сейчас остановимся на понятии статистический эксперимент.

В его основе лежит метод статистических испытаний (метод Монте-Карло). Суть его состоит в том, что результат испытания зависит от значения некоторой случайной величины, распределенной по заданному закону. Поэтому результат каждого отдельного испытания также носит случайный характер. Проведя серию испытаний, получают множество частных значений наблюдаемой характеристики (выборку). Полученные статистические данные обрабатываются и представляются в виде численных оценок интересующих

исследователя величин (характеристик системы).

Теоретической основой метода статистических испытаний являются предельные теоремы теории вероятностей (теорема Чебышева, теорема Бернулли, теорема Пуассона). Принципиальное значение предельных теорем состоит в том, что они гарантируют высокое качество статистических оценок при весьма большом числе испытаний. Важно отметить, что метод статистических испытаний применим для исследования как стохастических, так и детерминированных систем.

Еще одной важной особенностью данного метода является то, что его реализация практически невозможна без использования компьютера.

В качестве иллюстрации к изложенному рассмотрим применение метода статистических испытаний для вычисления площади круга заданного радиуса. Данная задача явно относится к классу детерминированных, поскольку весьма сложно представить себе случайные факторы, под влиянием которых площадь неподвижной геометрической фигуры могла бы изменяться. И тем не менее...

Пусть круг имеет радиус $r = 5$, и его центр находится в точке с координатами $(1, 2)$. Уравнение соответствующей окружности имеет вид:

$$(x - 1)^2 + (y - 2)^2 = 25.$$

Для решения задачи методом Монте-Карло впишем круг в квадрат. Его вершины будут иметь координаты $(-4, -3)$, $(6, -3)$, $(-4, 7)$ и $(6, 7)$. Любая точка внутри квадрата или на его границе должна удовлетворять неравенствам $-4 < x < 6$ и $-3 < y < 7$.

При решении задачи естественно исходить из того, что все точки в этом квадрате могут появляться с одинаковой вероятностью, то есть x и y распределены равномерно со следующими плотностями вероятности:

- $f(x) = 1/10$ – для $-4 < x < 6$ и $f(x) = 0$ – в противном случае;
- $f(y) = 1/10$ – для $-3 < y < 7$ и $f(y) = 0$ – в противном случае.

Проведя некоторое количество испытаний (то есть получив множество

случайных точек, принадлежащих квадрату), подсчитаем число точек, попавших внутрь круга или на окружность. Если выборка состоит из n наблюдений и m из n точек попали внутрь круга или на окружность, то оценку площади круга можно получить из соотношения:

$$S_{\text{кр}} = S_{\text{кв}} m/n = 100 m/n.$$

В табл. 1 приведены оценки $S_{\text{кр}}$, полученные для разных значений n , причем для каждого n выполнялось 5 прогонов (точное значение $S_{\text{кр}} = 78.54$).

Прогоны отличаются друг от друга последовательностями случайных чисел, из которых формировались координаты точек (рис. 16).

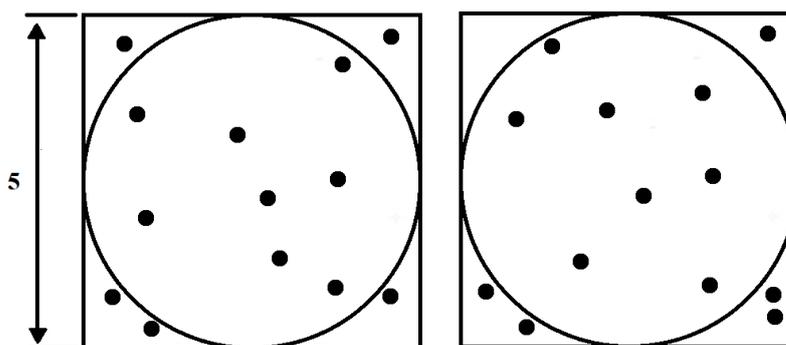


Рис. 16 – Различные результаты двух прогонов при неизменном количестве точек ($n = 13$). Слева показан результат первого прогона, справа – второго

На основании полученных результатов могут быть сделаны выводы, которые справедливы для любого имитационного эксперимента независимо от физической природы и типа моделируемой системы:

- каждый прогон модели можно рассматривать как одно наблюдение в проводимом эксперименте на модели;

- с увеличением продолжительности прогона (то есть продолжительности наблюдения или объема испытаний) отклонение измеряемой величины от ее точного значения уменьшается, поскольку наблюдаемая система переходит в стационарное состояние;

Таблица 1 – Результаты статистических испытаний

| Номер прогона | Оценки площади круга ($S_{кр}$) | | | | |
|---------------|-----------------------------------|------|-------|-------|--------|
| | Объем испытаний (n) | | | | |
| | 100 | 200 | 1000 | 5000 | 10000 |
| 1 | 78 | 79.5 | 78.8 | 78.22 | 78.77 |
| 2 | 70 | 77 | 77.3 | 78.6 | 78.23 |
| 3 | 81 | 79.5 | 80.2 | 77.72 | 78.88 |
| 4 | 70 | 77 | 79.5 | 77.76 | 78.63 |
| 5 | 79 | 77 | 79.8 | 79 | 78.21 |
| Среднее | 75.6 | 78 | 79.12 | 78.26 | 78.544 |
| Дисперсия | 27.3 | 1.88 | 1.29 | 0.3 | 0.1 |

– влияние переходных условий можно уменьшить, если увеличить количество прогонов модели (то есть количество экспериментов);

– существует предел, за которым увеличение продолжительности прогона модели уже не дает существенного повышения точности результата, измеряемой дисперсией.

Основная цель рассмотренного примера – привлечь внимание к тому факту, что имитационное моделирование не ограничивается разработкой модели и написанием соответствующей программы, а требует подготовки и проведения статистического эксперимента. В связи с этим результаты имитационного моделирования следует рассматривать как экспериментальные данные, требующие специальной обработки и анализа. В частности, для любого модельного эксперимента необходимо ответить на перечисленные ниже вопросы.

1) Какова должна быть продолжительность прогона для достижения стационарных условий?

2) Как получить статистически независимые наблюдения?

3) Сколько наблюдений необходимо для обеспечения требуемой

точности?

5.2 Область применения и классификация имитационных моделей

До сих пор речь шла о том, что моделирование можно считать основным инструментом, обеспечивающим принятие своевременных и обоснованных решений. Тем не менее, основной темой книги является не столько сам процесс принятия решений, сколько технология применения имитационного моделирования в различных областях человеческой деятельности. Поэтому пришло время несколько расширить обсуждаемых вопросов и поговорить о том, что же представляет собой полноценная имитационная модель и в каких случаях она может оказаться действительно полезной. Начнем с того, что попытаемся сформулировать характерные черты нашей «героини».

Имитационная модель (ИМ) – это формальное (то есть выполненное на некотором формальном языке) описание логики функционирования исследуемой системы и взаимодействия отдельных ее элементов во времени, учитывающее наиболее существенные причинно-следственные связи, присущие системе, и обеспечивающее проведение статистических экспериментов.

При разработке ИМ остаются справедливыми основные принципы моделирования. В качестве следствия из этого утверждения необходимо отметить два важных обстоятельства:

– взаимосвязь между отдельными элементами системы, описанными в модели, а также между некоторыми величинами (параметрами) может быть представлена в виде аналитических зависимостей (например, при моделировании полета управляемой ракеты отработка поступающих на борт команд может быть описана на уровне логики, а возникающие перегрузки рассчитываются аналитически);

– модель можно считать реализуемой и имеющей практическую ценность только в том случае, если в ней отражены лишь те свойства реальной системы, которые влияют на назначение выбранного показателя

эффективности.

Как было отмечено выше, для ИМ практически отсутствуют ограничения на область их применения (по типу моделируемой системы), и речь может идти только о целесообразности использования ИМ в данной предметной области и об объеме трудозатрат на ее разработку.

Поскольку основой имитационного моделирования является метод статистических испытаний, наибольший эффект от его применения достигается при исследовании сложных систем, на функционирование которых существенное влияние оказывают случайные факторы.

Применение имитационного моделирования целесообразно также в следующих случаях:

- если не существует законченной постановки задачи на исследование и идет процесс познания объекта моделирования;
- если характер протекающих в системе процессов не позволяет описать эти процессы в аналитической форме;
- если необходимо наблюдать за поведением системы (или отдельных ее компонентов) в течение определенного периода, в том числе с изменением скорости протекания процессов;
- при изучении новых ситуаций в системе либо при оценке функционирования ее в новых условиях;
- если исследуемая система является элементом более сложной системы, другие элементы которой имеют реальное воплощение;
- когда необходимо исследовать поведение системы при введении в нее новых элементов;
- при подготовке специалистов и освоении новой техники (в качестве тренажеров).

Приведенный список возможных областей применения имитационных моделей (который одновременно можно рассматривать и как перечень их достоинств) невольно вызывает вопрос: а зачем же в таком случае нужны остальные виды моделей? К сожалению, имитационные модели имеют целый

ряд недостатков. Первый и весьма существенный заключается в том, что разработка ИМ, как правило, требует больших затрат времени и сил. Кроме того, любая имитационная модель сложной системы значительно менее «объективна», чем аналитическая модель, поскольку она, прежде всего, отражает субъективные представления разработчика о моделируемой системе. Причем, бывает достаточно сложно как опровергнуть, так и обосновать адекватность созданной ИМ, особенно если речь идет о проектируемой системе. И, наконец, еще одно обстоятельство. Результаты имитационного моделирования, как и при любом численном методе, всегда носят частный характер. Для получения обоснованных выводов необходимо проведение серии модельных экспериментов, а обработка результатов требует применения специальных статистических процедур.

Каким же образом можно преодолеть указанные недостатки?

Во-первых, современное состояние вычислительной техники и ее программного обеспечения позволило вооружить «модельеров» такими мощными инструментальными средствами моделирования (к которым, в частности, относится пакет M ATLAB), что по мере накопления опыта в их использовании возможные трудозатраты «стремятся к нулю»; то же самое можно сказать о средствах статистического анализа и визуализации полученных результатов.

Во-вторых, «объективность» создаваемой модели может быть обеспечена в том случае, когда разработчик ясно представляет себе, какие именно характеристики исследуемой системы его интересуют: длительность выполнения определенных операций, вероятность перехода системы в некоторое состояние, возможность конфликта между отдельными подсистемами и т.д. Дело в том, что для каждого варианта постановки задачи исследования может быть выбрана соответствующая схема построения модели.

В этом отношении знание существующих схем построения имитационных моделей является весьма полезным.

Наиболее важный признак – *способ представления в модели динамики (движения) системы*. Она может быть описана посредством событий, работ (активностей), процессов и транзактов.

Другой важный признак – *способ изменения модельного времени*. По этому признаку различают моделирование с постоянным шагом и моделирование по особым состояниям.

Все эти понятия являются основополагающими в теории имитационного моделирования.

В большинстве случаев конечной целью моделирования является оптимизация тех или иных параметров системы. Однако, как было отмечено выше, потенциальные возможности имитационного моделирования существенно шире. В зависимости от этапа и назначения проводимых исследований, применяется один из трех наиболее распространенных видов имитационных экспериментов:

- исследование относительного влияния различных факторов на значения выходных характеристик системы;
- нахождение аналитической зависимости между интересующими исследователя выходными характеристиками и факторами;
- отыскание оптимальных значений параметров системы (так называемый «экстремальный эксперимент»).

Вид эксперимента влияет не только на выбор схемы его формализации, но также на построение плана эксперимента и выбор метода обработки его результатов.

С точки зрения организации взаимодействия исследователя с моделью, в ходе эксперимента ИМ делятся на автоматические и диалоговые.

– *Автоматическими* называются ИМ, взаимодействие пользователя с которыми сводится только к вводу исходной информации и управлению началом и окончанием работы моделей.

– *Диалоговыми* называются ИМ, позволяющие исследователю активно управлять ходом моделирования: приостанавливать сеанс моделирования,

изменять значения параметров модели, корректировать перечень регистрируемых данных и т.д.

Классификация имитационных моделей схематично показана на рис. 17.

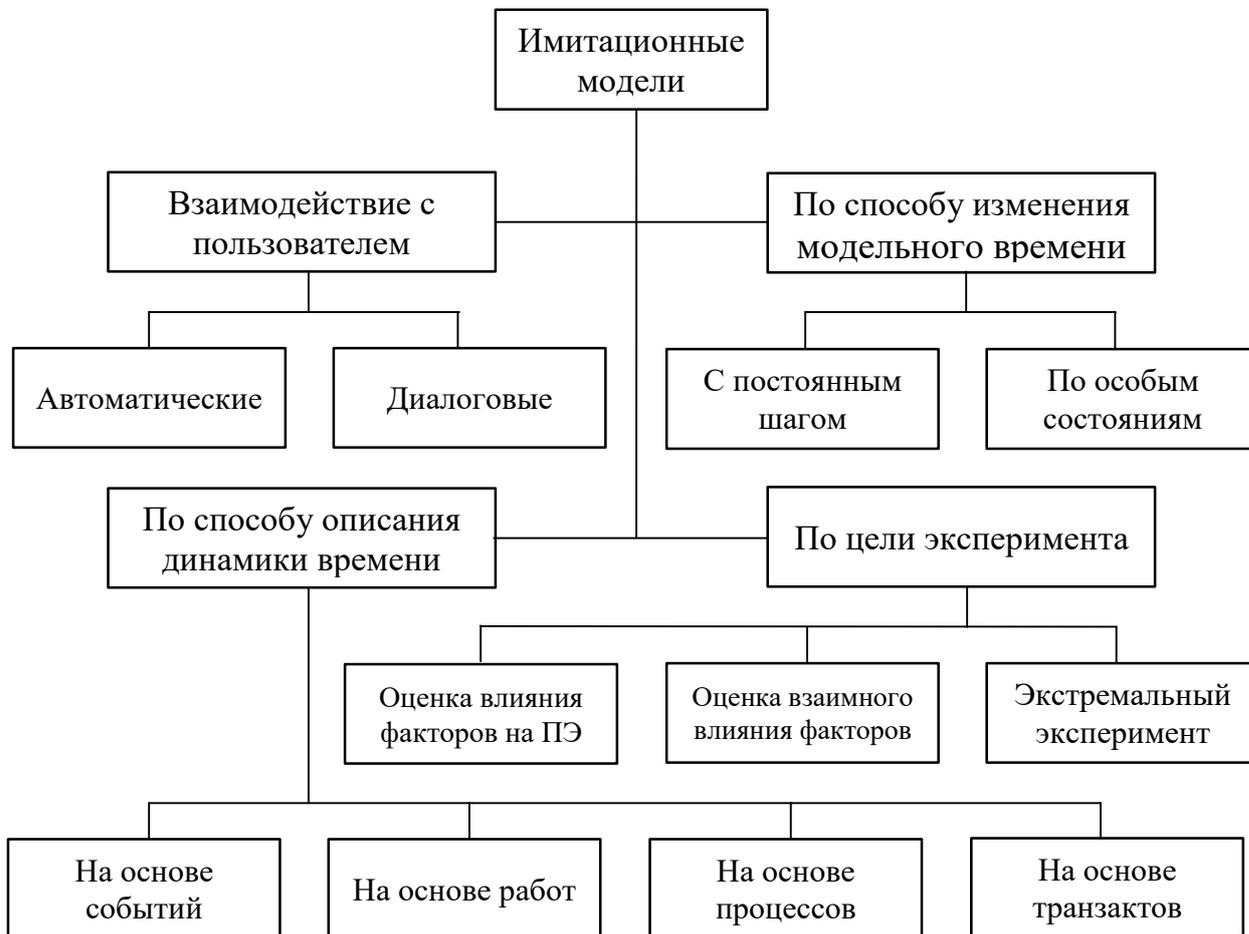


Рис. 17 – Схема классификации имитационных моделей

5.3. Описание поведения системы

Описание динамики системы, или, проще говоря, ее поведения, составляет основу любой имитационной модели. В качестве исходных посылок для решения этой задачи используются результаты, полученные на этапе разработки концептуальной модели системы. К ним относятся:

- определение принадлежности моделируемой системы одному из известных классов;
- описание рабочей нагрузки системы;
- выбор уровня детализации представления системы в модели и ее

декомпозиция.

Все последующие действия исследователя по созданию модели могут быть отнесены к этапу ее формализации, который в общем случае предполагает следующие шаги.

1. Выбор метода представления динамики системы (на основе событий, процессов или транзактов).

2. Формальное (математическое) описание случайных факторов, подлежащих учету в модели.

3. Выбор механизма изменения и масштаба модельного времени.

Начнем с определения понятий «событие», «процесс», «работа» и «транзакт».

При моделировании дискретных систем, ведущую роль играет понятие «события».

Событие представляет собой мгновенное изменение некоторого элемента системы или состояния системы в целом.

Событие характеризуется:

- условиями (или законом) возникновения;
- типом, который определяет порядок обработки (дисциплину обслуживания) данного события;
- нулевой длительностью.

Обычно события подразделяют на две категории:

- *события следования*, которые управляют инициализацией процессов (или отдельных работ внутри процесса);
- *события изменения состояний* (элементов системы или системы в целом).

В том случае, если модель строится с целью изучения причинно-следственных связей, присущих системе, то описание ее поведения в терминах событий является «самодостаточным». Примером такой задачи является проектирование пользовательского интерфейса программного продукта, когда разработчику необходимо оценить взаимную согласованность элементов

интерфейса (кнопок, флажков, переключателей, диалоговых окон и т.д.).

Если же исследователя интересует не только логика смены состояний моделируемой системы, но и временные параметры ее работы, механизм событий служит основой для представления в модели работ, процессов и транзактов.

Работа (активность) – это единичное действие системы по обработке (преобразованию) входных данных.

В зависимости от природы моделируемой системы, под входными данными могут пониматься информационные данные или какие-либо материальные ресурсы. Каждая из работ характеризуется временем выполнения и потребляемыми ресурсами. Соответственно, с помощью моделей, описанных в терминах работ, могут решаться задачи по оценке качества распределения ресурсов системы, ее производительности, надежности и т.д.

При моделировании сложных систем весьма часто имеют место ситуации, когда некоторые работы образуют устойчивую повторяющуюся последовательность. В таких случаях бывает удобнее перейти к описанию модели на основе процессов.

Под *процессом* понимают логически связанный набор работ. Некоторые процессы могут рассматриваться, в свою очередь, как работы в процессе более высокого уровня. И, наоборот, при повышении степени детализации описания системы некоторая работа может быть разделена на более мелкие составляющие и, таким образом, «превратиться» в процесс.

Любой процесс характеризуется совокупностью статических и динамических характеристик.

К *статическим* характеристикам процесса относятся:

- длительность;
- результат;
- потребляемые ресурсы;
- условия запуска (активизации);

– условия останова (прерывания).

В общем случае статические характеристики процесса не изменяются в ходе его реализации, однако при необходимости любая из них может быть представлена в модели как случайная величина, распределенная по заданному закону.

Динамической характеристикой процесса является его состояние (активен или находится в состоянии ожидания).

Моделирование в терминах процессов производится в тех же случаях, что и в терминах работ (то есть когда система оценивается по каким-либо временным показателям, либо с точки зрения потребляемых ресурсов). Например, при оценке производительности вычислительной сети обработка заданий может быть представлена в модели как совокупность соответствующих процессов, использующих ресурсы сети (оперативную память, пространство на жестких дисках, процессорное время, принтеры и т.д.).

При описании моделируемой системы в терминах работ и процессов используются оба вида событий:

- события следования – для отражения временных параметров системы;
- события изменения состояний – для представления логики взаимодействия протекающих в системе процессов (или выполняемых работ).

Еще один способ имитационного моделирования систем основан на использовании понятия транзакта.

Транзакт – это некоторое сообщение (заявка на обслуживание), которое поступает извне на вход системы и подлежит обработке. В некоторых случаях, например, при моделировании автоматизированных систем управления, более удобно проследить функционирование системы именно относительно алгоритма обработки транзакта. В рамках одной ИМ могут рассматриваться транзакты нескольких типов. Каждый транзакт характеризуется соответствующим алгоритмом обработки и необходимыми для его реализации ресурсами системы. Учитывая это, прохождение транзакта по системе можно

в некоторых случаях рассматривать как последовательную активизацию процессов, реализующих его обработку («обслуживание заявки»),

В связи с упоминанием термина «обслуживание заявки» уместно вспомнить о существовании теории массового обслуживания. При разработке и исследовании имитационных моделей на основе транзактов целесообразно использовать методику и показатели, применяемые при анализе систем массового обслуживания.

5.4. Моделирование случайных факторов

Еще раз отметим, что имитационная модель позволяет исследовать поведение различных систем с учетом влияния случайных факторов. Эти факторы, в зависимости от их природы, могут быть отражены в модели как случайные события, случайные величины (дискретные или непрерывные) или как случайные функции (процессы).

Например, если с помощью создаваемой имитационной модели предполагается исследовать надежность вычислительной системы, то возникновение отказа будет представлено в модели как случайное событие. Если же модель предназначена для оценки временных параметров процесса обслуживания клиентов в автомастерской, то интервал времени до появления очередного клиента удобнее всего описать как случайную величину, распределенную по некоторому закону.

Методы генерации случайных чисел

В основе всех методов и приемов моделирования случайных факторов лежит использование случайных чисел, имеющих равномерное распределение на интервале $[0; 1]$.

«Истинно» случайные числа формируются с помощью аналого-цифровых преобразователей на основе сигналов физических генераторов, использующих естественные источники случайных шумов (радиоактивный распад, шумы электронных и полупроводниковых устройств и т.п.).

Случайные числа, генерируемые аппаратно или программно на ЭВМ, называются псевдослучайными. Однако их статистические свойства совпадают со статистическими свойствами «истинно» случайных чисел. В состав практически всех современных систем программирования входят специальные функции генерации случайных чисел, которые обычно называют датчиками или генераторами случайных чисел.

Наиболее простой метод программной генерации случайных чисел – мультипликативный; в его основе лежит следующее рекуррентное соотношение:

$$a_i = (A a_{i-1} + C) \bmod M.$$

Здесь a_i , a_{i-1} – очередное и предыдущее случайные числа соответственно; A , C – константы; M – достаточно большое целое положительное число (чем больше M , тем длиннее неповторяемая последовательность).

Достоинство метода заключается в том, что при одних и тех же значениях параметров, входящих в это выражение, можно полностью воспроизвести эксперимент.

В приложении приведены методы генерации случайных чисел на ЭВМ.

Требования к генераторам случайных чисел

Практика показывает, что результаты имитационного моделирования существенно зависят от качества используемых последовательностей псевдослучайных чисел. Поэтому применяемые в ИМ генераторы случайных чисел должны пройти тесты на пригодность.

Основные анализируемые характеристики генерируемых датчиком последовательностей:

- равномерность;
- стохастичность (случайность);
- независимость.

Рассмотрим методы проведения такого анализа, наиболее часто применяемые на практике.

Проверка равномерности

Проверка равномерности может быть выполнена с помощью гистограммы относительных частот генерируемой случайной величины. Для ее построения интервал $(0, 1)$ разбивается на m равных частей и подсчитывается относительное число попаданий значений случайной величины в каждый интервал. Чем ближе огибающая гистограммы к прямой, тем в большей степени генерируемая последовательность отвечает требованию равномерности распределения (рис. 18).

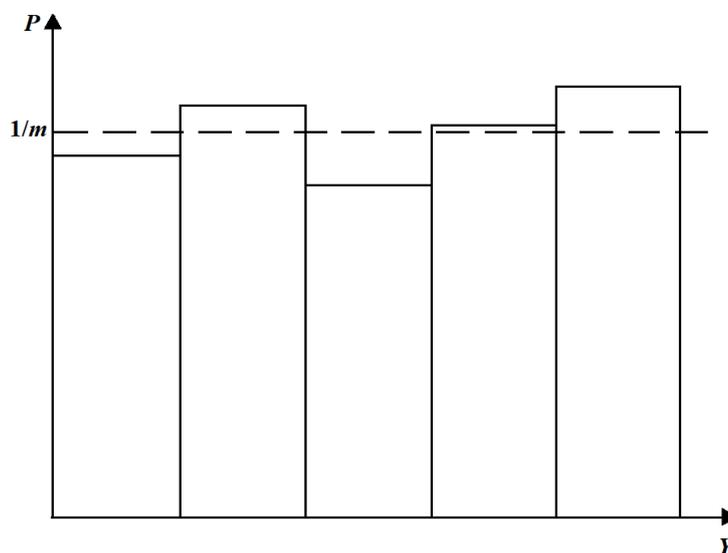


Рис. 18 – Частотная гистограмма последовательности случайных чисел

Проверка стохастичности

Рассмотрим один из основных методов проверки – метод комбинаций. Суть его сводится к следующему. Выбирают достаточно большую последовательность случайных чисел x_i и для нее определяют вероятность появления в каждом из x_i , ровно j единиц. При этом могут анализироваться как все разряды числа, так и только l старших. Теоретически закон появления j единиц в l разрядах двоичного числа может быть описан как биномиальный закон распределения (исходя из независимости отдельных разрядов).

Тогда при длине выборки N ожидаемое число появлений случайных

чисел x_i с j единицами в проверяемых l разрядах будет равно:

$$n_j = NC_l^j p^l(1).$$

Здесь C_l^j – число комбинаций (сочетаний) j единиц в l разрядах; $p^l(1)$ – вероятность появления единицы в двоичном разряде; $p(1)=0.5$.

Для полученной последовательности определяется эта же характеристика. Проверка соответствия реального значения теоретическому выполняется с помощью одного из статистических критериев согласия.

Проверка независимости

Проверка независимости проводится на основе вычисления корреляционного момента. Напомним, что две случайные величины a и b называются независимыми, если закон распределения каждой из них не зависит от того, какое значение приняла другая. Для независимых случайных величин корреляционный момент равен нулю.

Для оценки независимости элементов последовательности поступают следующим образом.

1. Вводят в рассмотрение дополнительную последовательность Y , в которой $y_i = x_{i+t}$, где t – величина сдвига последовательности Y относительно исходной последовательности X .

2. Вычисляют коэффициент корреляции случайных величин X и Y , для чего используются специальные расчетные соотношения.

Еще одна важная характеристика датчика случайных чисел – *длина отрезка аперидичности L* .

Если в основу работы датчика положен мультипликативный метод, то оценить L несложно: она определяется величиной константы M .

Моделирование случайных событий

Для моделирования случайного события A , вероятность которого равна

P_c , достаточно сформировать одно число r , равномерно распределенное на интервале $[0, 1]$. При попадании r в интервал $[0, P_c]$ считают, что событие A наступило, в противном случае не наступило.

Пусть, например, вероятность отказа вычислительной системы составляет 0.3. Чтобы определить, возникнет ли отказ на очередном шаге моделирования, достаточно сгенерировать с помощью датчика одно случайное число r и сравнить его с вероятностью отказа (рис. 19).

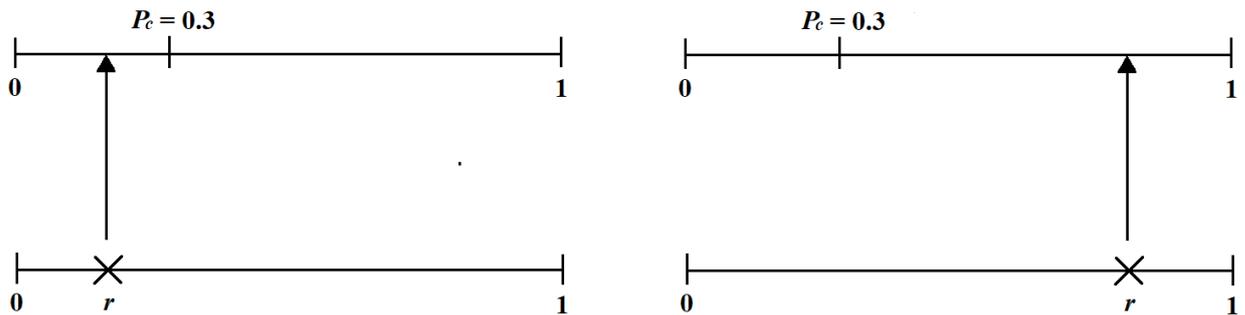


Рис. 19 – Моделирование случайного события: отказ «произошел» (слева), отказ «не произошел» (справа)

Для моделирования полной группы N несовместных событий $A = \{A_1, A_2, \dots, A_N\}$ с вероятностями соответственно P_1, P_2, \dots, P_N также достаточно одного значения r : событие A_i из группы A считается наступившим, если выполняется условие:

$$A = A_i \left| \sum_{i=0}^{n-1} P_i < r < \sum_{i=0}^n P_i. \right.$$

Здесь $n = 1, N$.

Предположим, что в каждый момент времени может происходить обращение только к одному из трех модулей оперативной памяти вычислительной системы. Вероятности обращения к каждому из них P_1, P_2 и P_3 равны соответственно 0.3, 0.5 и 0.2. Чтобы узнать, из какого именно модуля

будут считаны данные, необходимо определить, в какой интервал попадет полученное от датчика случайное число r (рис. 20). Если группа событий A не полна, то вводят фиктивное событие A_{N+1} , с вероятностью P_{N+1} , такой, что сумма вероятностей становится равной 1.

После этого генерируют число r и проверяют указанное выше условие. При $A = A_{N+1}$ считают, что ни одно событие из исходной группы A не наступило.

Для имитации зависимых событий A и B (B зависит от A) необходимо знать безусловную вероятность $P(A)$ события A и условные вероятности $P(B/A)$ и $P(B/\bar{A})$. Сначала описанным выше способом имитируется появление события A , в зависимости от исхода выбирается одна из вероятностей $P(B/A)$ или $P(B/\bar{A})$, и по той же технологии определяется наступление события B .

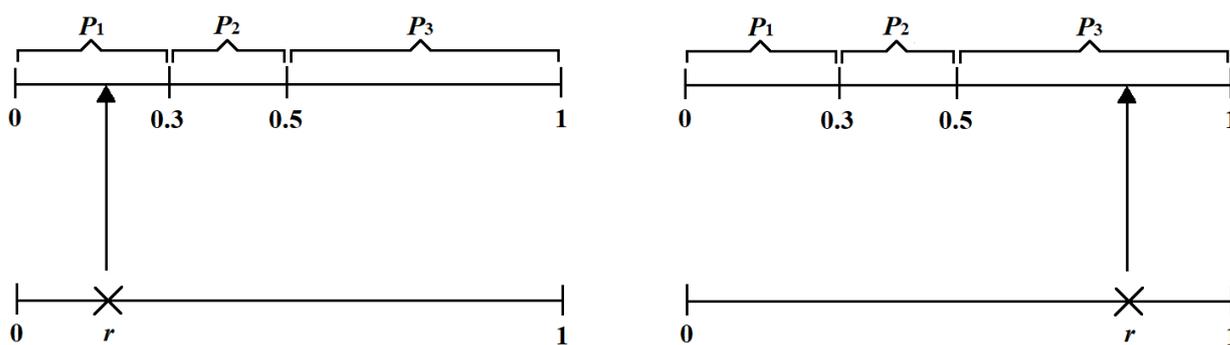


Рис. 20 – Моделирование полной группы из трех несовместных событий – обращение к первому модулю памяти (слева), обращение к третьему модулю памяти (справа)

Моделирование случайных величин

При моделировании дискретных случайных величин наиболее часто используются два метода:

- метод последовательных сравнений;
- метод интерпретации.

Метод последовательных сравнений

Число r последовательно сравнивают со значением суммы $P_1 + P_2 + \dots$,

где P_1 – вероятность наименьшего значения случайной величины Y , P_2 – вероятность второго по величине значения. При первом выполнении следующего условия проверка прекращается, и дискретная случайная величина Y считается принявшей значение

$$r > \sum_{i=1}^n P_i.$$

Процесс можно ускорить, применяя методы оптимизации перебора: дихотомии, ранжирования P и т.д.

Величины P , рассчитывают по функциям распределения вероятности, соответствующим моделируемому закону.

Метод интерпретации

Метод основан на физической трактовке моделируемого закона распределения. Например, биномиальное распределение описывает число успехов в n независимых испытаниях с вероятностью успеха в каждом испытании P и вероятностью неудачи $g = 1 - P$. При моделировании этого распределения с помощью метода интерпретации выбирают n независимых случайных чисел, равномерно распределен на интервале $[0, 1]$, и подсчитывают количество тех из них, которые меньше P . Это число и является моделируемой случайной величиной.

Моделирование непрерывных случайных величин

При моделировании *непрерывных* случайных величин с заданным законом распределения могут использоваться три метода:

- метод нелинейных преобразований;
- метод композиций;
- табличный метод.

Первые два метода требуют от разработчика модели весьма серьезной

математической подготовки и в данной книге не рассматриваются. Третий метод, условно названный «табличным», основан на замене закона распределения непрерывной случайной величины специальным расчетным соотношением, которое позволяет вычислять значение случайной величины по значению случайного числа, равномерно распределенного все на том же интервале $[0, 1]$. Такие соотношения получены практически для всех наиболее распространенных видов распределений и приведены в справочной литературе. В качестве примера ниже даны расчетные соотношения для двух законов распределения – показательного и нормального:

$$x = -\frac{1}{\lambda} \ln(r).$$

Здесь λ – параметр показательного распределения, r – равномерно распределенное случайное число.

$$x = m + s \left(\sum_{i=1}^{12} r_i - 6 \right).$$

Здесь m , s – параметры нормального закона распределения, r_i – равномерно распределенное случайное число.

Технологию моделирования непрерывных случайных величин поясним на двух небольших примерах.

Специалисты в области надежности технических систем доказали, что интервалы времени между отказами распределены по экспоненциальному закону. Чтобы получить в модели величину промежутка времени между двумя соседними отказами, достаточно сгенерировать случайное число r и подставить его в указанное выше выражение для показательного закона распределения; разумеется, при этом должен быть известен параметр распределения λ .

Нормальный закон имеет более широкий диапазон применения, поскольку любая величина, зависящая от большого числа случайных факторов, может считаться распределенной по нормальному закону. Например, если некто играет в «Дартс» и пытается попасть в центр круга (в «бычий глаз»), то координаты точек попадания дротика можно считать распределенными по нормальному закону с параметрами m и s , где m – координаты точки прицеливания, а s характеризует отклонение от нее по каждой из координат. Для определенности предположим, что $m_x = m_y = 0$, а $s_x = s_y = 5$, то есть «бычий глаз» имеет координаты $(0, 0)$ и отклонение составляет 5 см. Чтобы смоделировать координаты очередной точки попадания, следует получить от датчика две последовательности случайных чисел для каждой из координат в отдельности (по 12 чисел в каждой) и подставить их вместе с параметрами закона в выражение для нормального закона распределения:

$$x = 5 \left(\sum_{i=1}^{12} r_i - 6 \right).$$

$$y = 5 \left(\sum_{i=1}^{12} r_i - 6 \right).$$

В одной и той же имитационной модели могут фигурировать различные случайные факторы, одни могут быть представлены как случайные события, другие – как случайные величины; более того, если моделируется поведение достаточно сложной системы, то ее функционирование может быть связано с возникновением нескольких типов событий и учетом большого числа случайных величин, распределенных по различным законам. Если же моделирование всех случайных факторов основано на использовании одного датчика, генерирующего одну «общую» последовательность случайных чисел, то с математической точки зрения их нельзя считать независимыми. В связи с

этим для моделирования каждого случайного фактора стараются использовать отдельный генератор, или, по крайней мере, обеспечивать создание новой последовательности случайных чисел. Во многих специализированных языках и пакетах моделирования такая возможность предусмотрена. Это относится, в частности, и к пакету M ATLAB.

5.5 Управление модельным временем

Виды представления времени в модели

Приступая к изучению механизмов управления модельным временем, уместно поговорить о том, какую роль вообще играет время в имитационном моделировании. При знакомстве с имитационным экспериментом мы отмечали, что он представляет собой наблюдение за поведением системы в течение некоторого промежутка времени. Конечно, далеко не во всех статистических испытаниях фактор времени играет ведущую роль, а в некоторых и вообще может не рассматриваться. Вспомните, например, задачу о вычислении площади круга: полученный результат не зависел от того, сколько долго мы «бомбили» квадрат случайными точками (речь в данном случае не идет о количестве этих точек). Но значительно больше задач, в которых оценка эффективности моделируемой системы напрямую связана с временными характеристиками ее функционирования. К ним относятся упоминавшиеся уже задачи по оценке производительности, некоторые задачи по оценке надежности, качества распределения ресурсов, а также все задачи, связанные с исследованием эффективности процессов обслуживания. Характерной особенностью большинства практических задач является то, что скорость протекания рассматриваемых в них процессов значительно ниже скорости реализации модельного эксперимента. Например, если моделируется работа вычислительного центра в течение недели, вряд ли кому-то придет в голову воспроизводить этот процесс в модели в таком же масштабе времени. С другой стороны, даже те имитационные эксперименты, в которых временные параметры работы системы не учитываются, требуют для своей реализации

определенных затрат времени работы компьютера.

В связи с этим при разработке практически любой имитационной модели и планировании проведения модельных экспериментов необходимо соотносить между собой три представления времени:

- реальное время, в котором происходит функционирование имитируемой системы;

- модельное (или, как его еще называют, системное) время, в масштабе которого организуется работа модели;

- машинное время, отражающее затраты времени Э ВМ на проведение имитации.

С помощью механизма модельного времени решаются следующие задачи:

- отображается переход моделируемой системы из одного состояния в другое;

- производится синхронизация работы компонент модели;

- изменяется масштаб времени «жизни» (функционирования) исследуемой системы;

- производится управление ходом модельного эксперимента;

- моделируется квазипараллельная реализация событий в модели.

Приставка «квази» в данном случае отражает последовательный характер обработки событий (процессов) в ИМ, которые в реальной системе возникают (протекают) одновременно.

Необходимость решения последней задачи связана с тем, что в распоряжении исследователя находится, как правило, однопроцессорная вычислительная система, а модель может содержать значительно большее число одновременно работающих подсистем. Поэтому действительно параллельная (одновременная) реализация всех компонент модели невозможна. Даже если используется так называемая распределенная модель, реализуемая на нескольких узлах вычислительной сети, совсем не обязательно, что число узлов будет совпадать с числом одновременно

работающих компонент модели.

Ранее были названы два метода реализации механизма модельного времени – с постоянным шагом и по особым состояниям.

Выбор метода реализации механизма модельного времени зависит от назначения модели, ее сложности, характера исследуемых процессов, требуемой точности результатов и т.д.

Изменение времени с постоянным шагом

При использовании данного метода отсчет системного времени ведется через фиксированные, выбранные исследователем интервалы времени. События в модели считаются наступившими в момент окончания этого интервала. Погрешность в измерении временных характеристик системы в этом случае зависит от величины шага моделирования Δt .

Метод постоянного шага целесообразно использовать в том случае, если:

- события появляются регулярно, их распределение во времени достаточно равномерно;
- число событий велико и моменты их появления близки;
- невозможно заранее определить моменты появления событий.

Данный метод управления модельным временем достаточно просто реализовать в том случае, когда условия появления событий всех типов в модели можно представить, как функцию времени.

Пусть, например, событие состоит в том, что летящий самолет пересекает некоторый воздушный рубеж, расстояние до которого равно R . Если самолет движется по прямой с постоянной скоростью V , можно вычислять путь, пройденный самолетом, с интервалом времени Δt : $S = S + V \Delta t$. Соответственно, событие считается наступившим, если выполняется условие $S > R$, а момент времени-наступления события принимается равным $n \Delta t$, где n – номер шага моделирования, на котором условие стало истинным.

В общем виде алгоритм моделирования с постоянным шагом

представлен на рис. 21 (t_m – текущее значение модельного времени, T_m – заданный интервал моделирования), а для рассмотренного выше примера с самолетом – на рис. 22.

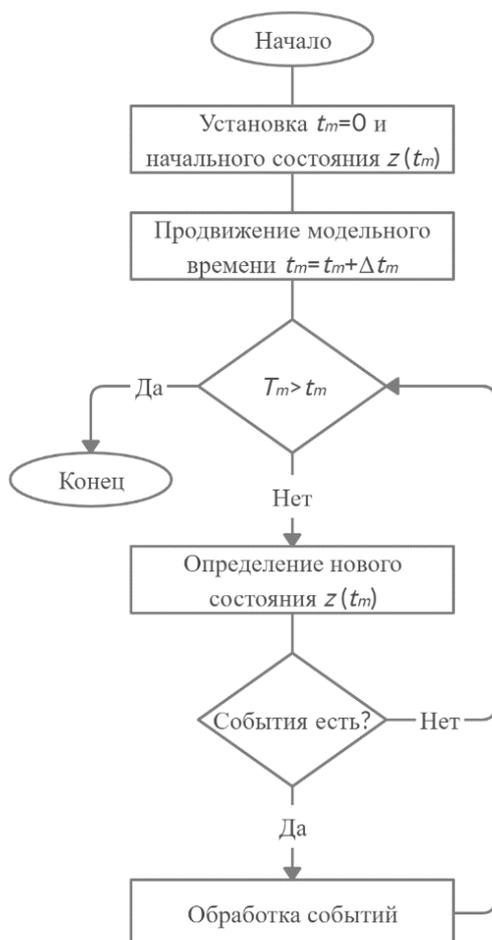


Рис. 21 – Алгоритм моделирования с постоянным шагом

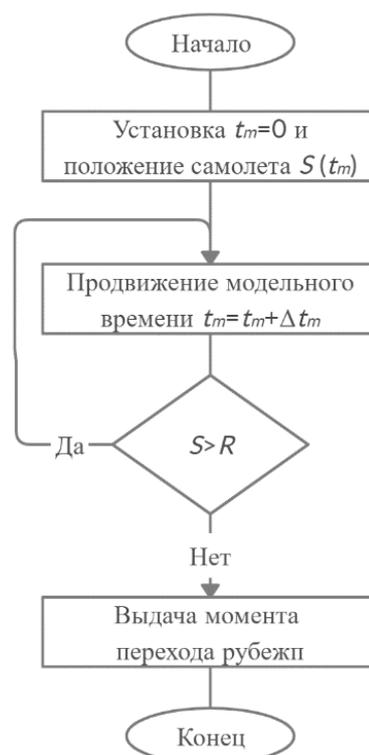


Рис. 22 – Пример моделирования с постоянным шагом

Обратите внимание на то, что в отличие от обобщенного алгоритма, в приведенном примере моделирование завершается не по истечении заданного интервала времени, а при наступлении интересующего нас события. В связи с этим необходимо еще раз подчеркнуть, что при моделировании с постоянным шагом результат моделирования напрямую зависит от величины этого шага. Причем, если шаг будет слишком большим, то результат, скорее всего, будет неверным: момент окончания очередного шага очень редко будет совпадать с

реальным моментом пересечения самолетом заданного рубежа. Такая ситуация показана на рис. 23.

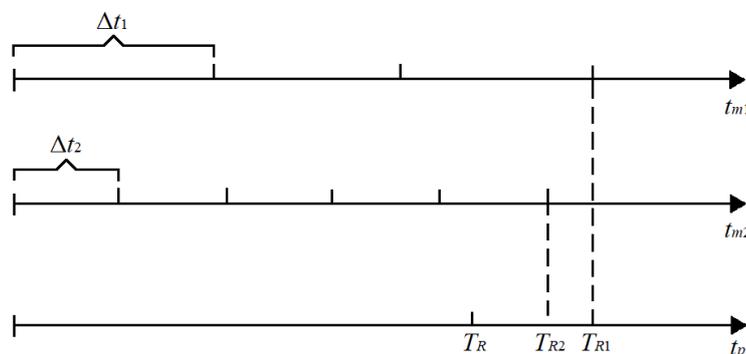


Рис. 23 – Зависимость результата эксперимента от шага модельного времени

На рисунке использованы следующие обозначения:

– t_{m1} – ось модельного времени при использовании шага Δt_1 ;

– t_{m2} – ось модельного времени при использовании шага Δt_2 ;

– t_p – ось реального времени;

– T_R – реальный момент пересечения самолетом рубежа;

– T_{R1} , T_{R2} – моменты пересечения рубежа, полученные для

соответствующих величин Δt .

Приведенный пример призван, кроме всего прочего, обратить внимание читателя на то, что выбор величины шага моделирования является нелегким и очень важным делом. Универсальной методики решения этой проблемы не существует, но во многих случаях можно использовать один из следующих подходов:

– принимать величину шага равной средней интенсивности возникновения событий различных типов;

– выбирать величину шага равной среднему интервалу между наиболее частыми (или наиболее важными) событиями.

Изменение времени по особым состояниям

При моделировании по особым состояниям системное время каждый раз

изменяется на величину, строго соответствующую интервалу времени до момента наступления очередного события. В этом случае события обрабатываются в порядке их наступления, а одновременно наступившими считаются только те, которые являются одновременными в действительности.

Для реализации моделирования по особым состояниям требуется разработка специальной процедуры планирования событий (так называемого календаря событий). Если известен закон распределения интервалов между событиями, то такое прогнозирование труда не составляет: достаточно к текущему значению модельного времени добавить величину интервала, полученную с помощью соответствующего датчика. Пусть, например, за летящим самолетом, фигурировавшим при описании моделирования с постоянным шагом, наблюдает диспетчер. Он вводит информацию о самолете, причем интервалы между вводом двух соседних сообщений являются случайными величинами, распределенными по нормальному закону с заданными параметрами. Иллюстрация к такой ситуации приведена на рис. 24 (T_c – момент ввода очередного сообщения, Δt – случайный интервал).

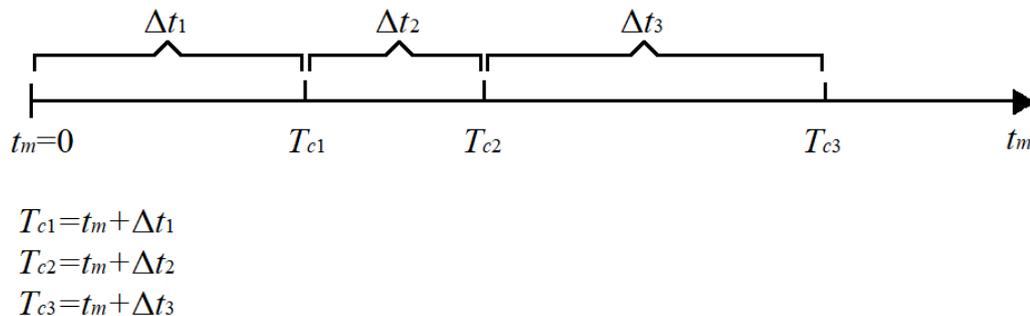


Рис. 24 – Изменение модельного времени по особым состояниям

Если же момент наступления события определяется некоторыми логическими условиями, то необходимо сформулировать эти условия и проверять их истинность для каждого последующего шага моделирования. Практика показывает, что сложности в реализации механизма изменения времени по особым состояниям связаны в первую очередь с корректным описанием таких условий. Трудности еще более возрастают, если в модели

фигурируют несколько типов взаимосвязанных событий. Моделирование по особым состояниям целесообразно использовать, если:

- события распределяются во времени неравномерно или интервалы между ними велики;
- предъявляются повышенные требования к точности определения взаимного положения событий во времени;
- необходимо учитывать наличие одновременных событий.

Дополнительное достоинство метода заключается в том, что он позволяет экономить машинное время, особенно при моделировании систем периодического действия, в которых события длительное время могут не наступать.

Обобщенная схема алгоритма моделирования по особым состояниям представлена на рис. 25 ($t_{\text{соб.}i}$ – прогнозируемый момент наступления i -го события).

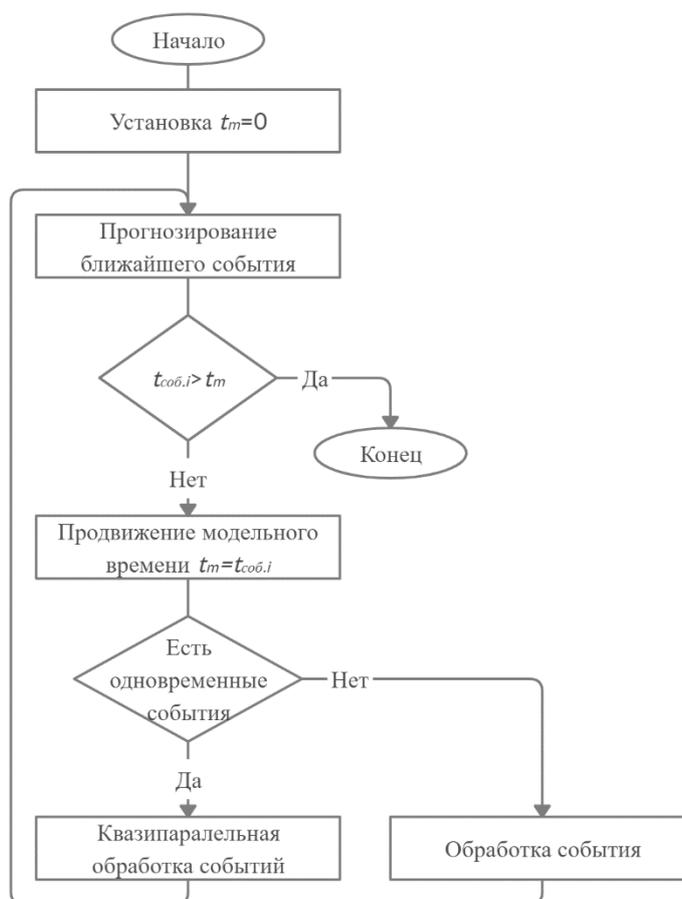


Рис. 25 – Алгоритм моделирования па особым состояниям

Чтобы читатель смог «почувствовать разницу» в использовании двух методов управления модельным временем, вернемся к примеру, с диспетчером. Дополним его следующим условием: необходимо подсчитать число сообщений, которые успеет ввести диспетчер в течение заданного интервала моделирования.

Прежде всего, необходимо ответить на вопрос: что понимать под «особыми состояниями», которые должны влиять на изменение модельного времени? На практике обычно вместо состояний рассматривают события, определяющие смену состояний моделируемого процесса. Для процесса ввода информации диспетчером такой переход выполняется достаточно просто: событие – это ввод очередного сообщения; другими словами, ввод очередного сообщения «продвигает» модельное время на соответствующий интервал. Так, если интервалы между сообщениями подчиняются нормальному закону с параметрами m и s , то очередное i -е значение модельного времени $t_m(i)$ определяется следующим образом:

$$t_m(i) = t_m(i - 1) + \text{norm}(m, s)$$

В этом выражении слагаемое $\text{norm}(m, s)$ означает обращение к генератору случайных чисел, распределенных по нормальному закону.

Алгоритм работы модели приведен на рис. 26 (N – число введенных сообщений).

Подведем итоги изложенному в этом разделе.

– Выбор механизма изменения модельного времени определяет технологию реализации имитационной модели.

– На выбор метода моделирования влияет целый ряд факторов, однако определяющим является тип моделируемой системы: для дискретных систем, события в которых распределены во времени неравномерно, более удобным

является изменение модельного времени по особым состояниям.

– Если в модели должны быть представлены компоненты реальной системы, работа которых измеряется в разных единицах времени, то они должны быть предварительно приведены к единому масштабу.

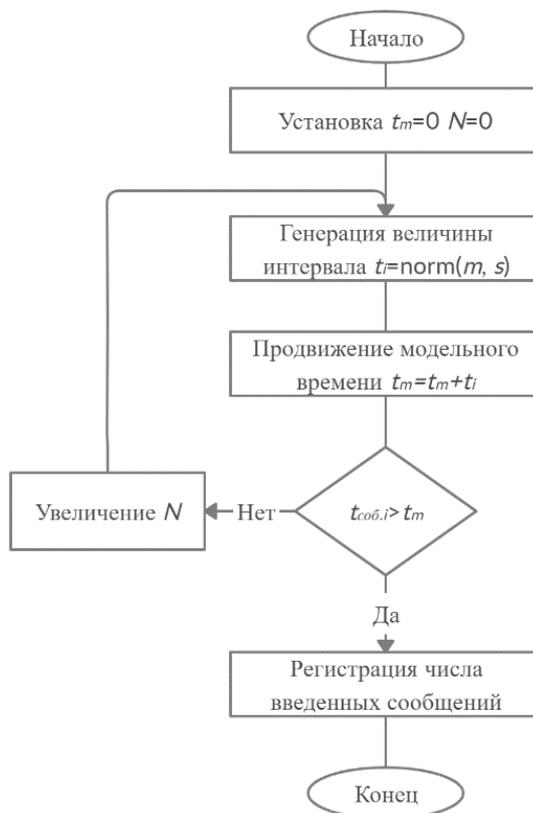


Рис. 26 – Пример работы модели по особым состояниям

5.6. Моделирование параллельных процессов

Практически любая более или менее сложная система имеет в своем составе компоненты, работающие одновременно, или как принято говорить в технических науках, параллельно. Наверняка читатель сам может привести массу примеров таких систем из близкой ему области деятельности. Мы же напомним только один, взятый из монолога М. Жванецкого, посвященного работе ликероводочного завода. Судя по сюжету монолога, все цеха этого предприятия работали абсолютно параллельно и независимо друг от друга.

Итак, если в составе системы имеются компоненты (подсистемы),

выполняющие свои функции одновременно, то можно утверждать, что в такой системе существуют параллельные процессы. Параллельно работающие подсистемы могут взаимодействовать самым различным образом, либо вообще работать независимо друг от друга. Способ взаимодействия подсистем определяет вид параллельных процессов, протекающих в системе. В свою очередь, вид моделируемых процессов влияет на выбор метода их имитации.

Виды параллельных процессов

Асинхронный параллельный процесс – это такой процесс, состояние которого не зависит от состояния другого параллельного процесса (ПП).

Пример асинхронных ПП, протекающих в рамках одной системы, – подготовка и проведение рекламной кампании фирмой и работа сборочного конвейера. Или пример из области вычислительной техники – выполнение вычислений процессором и вывод информации на печать.

Синхронный ПП – это такой процесс, состояние которого зависит от состояния взаимодействующих с ним ПП.

Пример синхронного ПП – работа торговой организации и доставка товара со склада (нет товара – нет торговли).

Один и тот же процесс может быть синхронным по отношению к одному из активных ПП и асинхронным по отношению к другому. Так, при работе вычислительной сети по технологии «клиент-сервер» каждый из узлов сети синхронизирует свою работу с работой сервера, но не зависит от работы других узлов.

Подчиненный ПП создается и управляется другим процессом (более высокого уровня). Весьма характерным примером таких процессов является ведение боевых действий подчиненными подразделениями.

Независимый ПП – процесс, который не является подчиненным ни для одного из процессов.

Скажем, после запуска неуправляемой зенитной ракеты ее полет можно рассматривать как независимый процесс, одновременно с которым самолет

ведет боевые действия другими средствами.

Способ организации параллельных процессов в системе зависит от физической сущности этой системы.

Остановимся несколько подробнее на особенностях реализации параллельных процессов в вычислительных системах (ВС). Это обусловлено следующей причиной.

Разработка и использование любой ИМ предполагает ее программную реализацию и исследование с применением ВС. Поэтому для реализации моделей, имитирующих параллельные процессы, в некоторых случаях применимы механизмы, характерные для выполнения параллельных вычислений.

Вместе с тем, реализация параллельных процессов в ВС имеет свои особенности:

– на уровне задач вычислительные процессы могут быть истинно параллельными только в многопроцессорных ВС или вычислительных сетях;

– многие ПП используют одни и те же ресурсы, поэтому даже асинхронные ПП в пределах одной ВС вынуждены согласовывать свои действия при обращении к общим ресурсам;

– в ВС дополнительно используется еще два вида ПП: родительский и дочерний ПП; особенность их состоит в том, что процесс-родитель не может быть завершен, пока не завершатся все его дочерние процессы.

В силу перечисленных особенностей для организации взаимодействия параллельных процессов в ВС используются три основных подхода:

– на основе «взаимного исключения»;

– на основе синхронизации посредством сигналов;

– на основе обмена информацией (сообщениями).

«*Взаимное исключение*» предполагает запрет доступа к общим ресурсам (общим данным) для всех ПП, кроме одного, на время его работы с этими ресурсами (данными).

Синхронизация подразумевает обмен сигналами между двумя или более

процессами по установленному протоколу. Такой «сигнал» рассматривается как некоторое событие, вызывающее у получившего его процесса соответствующие действия. Часто возникает необходимость передавать от одного ПП другому более подробную информацию, чем просто «сигнал-событие». В этом случае процессы согласуют свою работу на основе *обмена сообщениями*.

Перечисленные механизмы реализуются в ВС на двух уровнях – системном и прикладном.

Механизм взаимодействия между ПП на системном уровне определяется еще на этапе разработки ВС и реализуется в основном средствами операционной системы (частично – с использованием аппаратных средств).

На прикладном уровне взаимодействие между ПП реализуется программистом средствами языка, на котором разрабатывается программное обеспечение. Наибольшими возможностями в этом отношении обладают так называемые языки реального времени и языки моделирования.

Языки реального времени (ЯРВ) – это языки, предназначенные для создания программного обеспечения, работающего в реальном масштабе времени, например, для разработки различных автоматизированных систем управления (предприятием, воздушным движением и т. д.). К ним, в частности, относятся: язык Ада, язык Модуля и практически единственный отечественный язык реального времени — Эль-76 (использовавшийся в многопроцессорных вычислительных комплексах семейства «Эльбрус»).

Методы описания параллельных процессов

Языки моделирования по сравнению с языками реального времени требуют от разработчика значительно менее высокого уровня подготовки в области программирования, что обусловлено двумя обстоятельствами:

– средства моделирования изначально ориентированы на квазипараллельную обработку параллельных процессов;

– механизмы реализации ПП относятся, как правило, к внутренней организации системы (языка) моделирования и их работа скрыта от программиста.

В практике имитационного моделирования одинаково широко используются как процессно-ориентированные языки (системы) моделирования, например SIMULA, так и языки, ориентированные на обработку транзактов (например, язык GPSS). Для тех и других характерны аналогичные методы реализации квазипараллелизма, основанные на ведении списков событий. В процессно-ориентированных системах и используются списки событий следования, а в транзактных системах – списки событий изменения состояний.

Современные языки и системы моделирования, ориентированные на среду многозадачных операционных систем типа Windows, частично используют аналогичные механизмы управления процессами, что делает их применение еще более эффективным. В пакете MATLAB также имеется собственный язык моделирования, и к нему в полной мере можно отнести сказанное выше. Этот язык имеет свои особенности, моделирования параллельных процессов. Тем не менее, во многих случаях оказывается полезным знание общего механизма реализации ПП в языках моделирования.

Рассмотрим его применительно к моделированию на основе транзактов.

В этом случае под событием понимается любое перемещение транзакта по системе, а также изменение его состояния (обслуживается, заблокирован и т.д.)

Событие, связанное с данным транзактом, может храниться в одном из следующих списков.

– *Список текущих событий.* В этом списке находятся события, время наступления которых меньше или равно текущему модельному времени. События с «меньшим» временем связаны с перемещением тех транзактов, которые должны были начать двигаться, но были заблокированы.

– *Список будущих событий.* Этот список содержит события, время

наступления которых больше текущего модельного времени, то есть события, которые должны произойти в будущем (условия наступления которых уже определены, например, известно, что транзакт будет обслуживаться некоторым устройством 10 единиц времени).

– *Список прерываний.* Данный список содержит события, связанные с возобновлением обработки прерванных транзактов. События из этого списка выбираются в том случае, если сняты условия прерывания.

Рассмотрим использование двух первых списков событий в динамике, при моделировании параллельных процессов.

В списке текущих событий транзакты расположены в порядке убывания приоритета соответствующих событий, а при равных приоритетах – в порядке поступления в список.

Каждое событие (транзакт) в списке текущих событий может находиться либо в активном состоянии, либо в состоянии задержки. Если событие активно, то соответствующий транзакт может быть продвинут по системе; если продвижение невозможно (например, из-за занятости устройства), то событие (и транзакт) переводится в состояние задержки.

Как только завершается обработка (продвижение) очередного активного транзакта, просматривается список задержанных транзактов, и ряд из них переводится в активное состояние. Процедура повторяется до тех пор, пока в списке текущих событий не будут обработаны все активные события. После этого просматривается список будущих событий. Модельному времени присваивается значение, равное времени наступления ближайшего из этих событий. Данное событие заносится в список текущих событий. Затем просматриваются остальные события списка. Те из них, время которых равно текущему модельному времени, также переписываются в список текущих событий. Просмотр заканчивается, когда в списке остаются события, времена которых больше текущего модельного времени.

В качестве иллюстрации к изложенному рассмотрим небольшой пример.

Пусть в систему поступают транзакты трех типов, каждый из которых

обслуживается отдельным устройством. Известны законы поступления транзактов в систему и длительность их обслуживания. Таким образом, в системе существуют три параллельных независимых процесса (P_1, P_2, P_3).

Временная диаграмма работы системы при обслуживании одного транзакта каждого типа показана на рис. 27.

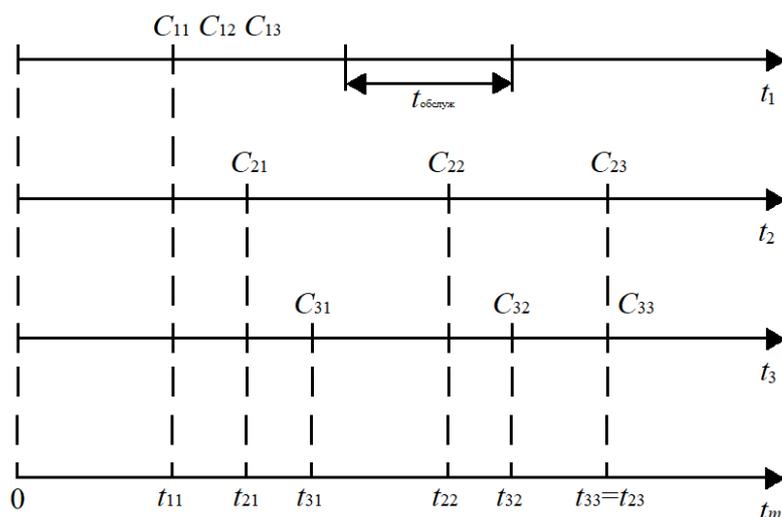


Рис. 27 – Временная диаграмма параллельных процессов

На рисунке события, относящиеся к процессу P_1 , обозначены как C_{1i} , относящиеся к P_2 и к P_3 – соответственно, как C_{2i} и C_{3i} . Время $t_{\text{обслуж}}$ соответствует времени обслуживания транзакта.

Для каждого процесса строится своя цепь событий, однако списки событий являются общими для всей модели. Формирование списков начинается с заполнения списка будущих событий. Как было отмечено выше, в этот список помещаются события, время наступления которых превышает текущее значение модельного времени. Очевидно, что на момент заполнения списка время наступления прогнозируемых событий должно быть известно. На первом шаге $t_{11} = 0$ в список будущих событий заносятся события C_{11} ; C_{21} , C_{31} . Затем событие с наименьшим временем наступления – C_{11} – переносится в список текущих событий; если одновременных с ним событий нет, то оно обрабатывается и исключается из списка текущих событий. После этого вновь корректируется список будущих событий и т.д., пока не истечет заданный

интервал моделирования.

Динамика изменения списков текущих и будущих событий для рассмотренного примера отражена в приведенной ниже таблице.

| Значение модельного времени | Список будущих событий | Список текущих событий |
|-----------------------------|------------------------|------------------------|
| 0 | $C_{11} C_{21} C_{31}$ | 0 |
| T_{11} | $C_{21} C_{31} C_{12}$ | C_{11} |
| T_{21} | $C_{31} C_{12} C_{22}$ | C_{21} |
| T_{31} | $C_{12} C_{22} C_{32}$ | C_{31} |
| T_{12} | $C_{22} C_{32} C_{13}$ | C_{12} |
| T_{22} | $C_{32} C_{13} C_{23}$ | C_{22} |
| T_{32} | $C_{13} C_{23} C_{33}$ | C_{32} |
| T_{13} | $C_{23} C_{33}$ | C_{13} |
| T_{23} | | $C_{23} C_{33}$ |

Многие авторы книг по имитационному моделированию, ставших уже классикой, считают, что знание механизма ведения списков событий просто необходимо разработчику модели; умение проследить в динамике цепь происходящих в модели событий, во-первых, повышает уверенность создателя модели в том, что она работает правильно и, во-вторых, существенно облегчает процесс отладки и модификации модели.

Применение сетевых моделей для описания параллельных процессов

Сведения, приведенные в предыдущем разделе, кроме всего прочего должны были убедить читателя в том, что разработчику имитационной модели необходимо иметь хотя бы минимальные знания и навыки в области программирования. И хотя следующие уроки книги в основном посвящены тому, чтобы развеять это впечатление, тем не менее, существует проблема, которую мы не сможем обойти. Речь идет о том, что этапу программной

реализации модели (то есть ее описанию на одном из языков программирования) должен предшествовать так называемый этап алгоритмизации. Другими словами, прежде чем превратить имитационную модель в работающую компьютерную программу, ее создатель должен воспользоваться каким-то менее формальным и более наглядным средством описания логики работы будущей программы. Разумеется, это требование не является обязательным для всех разработчиков и для всех создаваемых моделей; при наличии достаточного опыта программа не очень сложной модели может быть написана сразу. Однако практика показывает, что человеческие возможности не безграничны, и при моделировании более сложных систем даже опытные разработчики бывают вынуждены немного «притормозить» на этапе алгоритмизации. Для описания логики работы модели могут быть использованы различные средства: либо русский язык (устный или письменный), либо традиционные схемы алгоритмов (как на рис. 2.10, 2.11), либо какие-то другие «подручные» средства. Первые два варианта являются, как правило, наиболее знакомыми и наиболее часто используемыми. Однако если вы попытаетесь описать в виде схемы алгоритма модель даже такой простой системы, которая использовалась в предыдущем примере, то это окажется напрасной тратой времени и сил. Прежде всего потому, что такие схемы совершенно не приспособлены для описания параллельных процессов.

Именно поэтому представляется необходимым познакомить читателя с одним из наиболее элегантных и весьма распространенных средств описания параллельных процессов – *сетями Петри*. Им посвящено достаточно большое число изданий, поэтому мы ограничимся изложением только тех основных сведений, которые необходимы с точки зрения реализации технологии имитационного моделирования параллельных процессов.

Сети Петри

Одно из основных достоинств аппарата сетей Петри заключается в том,

что они могут быть представлены как в графической форме (что обеспечивает наглядность), так и в аналитической (что позволяет автоматизировать процесс их анализа).

При графической интерпретации сеть Петри представляет собой граф особого вида, состоящий из вершин двух типов – *позиций* и *переходов*, соединенных ориентированными дугами, причем каждая дуга может связывать лишь разнотипные вершины (позицию с переходом или переход с позицией). Вершины-позиции обозначаются кружками, вершины-переходы – черточками. С содержательной точки зрения переходы соответствуют событиям, присущим исследуемой системе, а позиции – условиям их возникновения. Таким образом, совокупность переходов, позиций и дуг позволяет описать причинно-следственные связи, присущие системе, но в статике. Чтобы сеть Петри «оживила», вводят еще один вид объектов сети – так называемые *фишки* или *метки* позиций. Переход считается активным (событие может произойти), если в каждой его входной позиции есть хотя бы одна фишка.

Расположение фишек в позициях сети называется *разметкой сети*.

В аналитической форме сеть Петри может быть представлена следующим образом:

$$P = (B, D, I, O, M).$$

В этом выражении представлены следующие параметры:

– $B = b_i$ – конечное непустое множество позиций;

– $D = d_i$ – конечное непустое множество переходов;

– $I: B \times D \rightarrow (0, 1)$ – входная функция (прямая функция инцидентности), которая для каждого перехода задает множество его входных позиций;

– $O: D \times B \rightarrow (0, 1)$ – выходная функция (обратная функция инцидентности), которая для каждого перехода задает множество его выходных позиций;

– M – функция разметки сети, $M: B \rightarrow (0, 1, 2, \dots)$ – ставит каждой позиции сети в соответствие неотрицательное целое число (равное числу меток в данной позиции).

С учетом введенных обозначений необходимое условие срабатывания перехода d_i (для всех входных позиций число находящихся в них меток должно быть не меньше 1) может быть записано следующим образом:

$$\forall b_i \in I(d_j) \{M(b_i) \geq 1\}.$$

Срабатывание перехода d_j изменяет разметку сети $M(B)$ на разметку $M'(B)$ по следующему правилу:

$$M'(B) = M(B) - I(d_j) + O(d_j)$$

То есть переход d_j изымает по одной метке из каждой своей входной позиции и добавляет по одной метке в каждую из выходных позиций. Смену разметки обозначают так:

$$M_0 \xrightarrow{d_j} M'$$

Входная и выходная функции сети Петри (I и O) позволяют описать любую сеть с помощью двух матриц $m \times n$ (матриц входных и выходных позиций), имеющих следующую структуру:

| | | | | |
|---------|---------|---------|---------|---------|
| $I(O)$ | d_1 | d_1 | \dots | d_n |
| b_1 | 0 | 1 | \dots | 1 |
| b_2 | 1 | 1 | \dots | 0 |
| \dots | \dots | \dots | \dots | \dots |
| b_m | 1 | 0 | \dots | 1 |

При анализе сети Петри основное внимание уделяется, как правило, трем направлениям.

– *Проблема достижимости.* В сети Петри с начальной разметкой M_0

требуется определить, достижима ли принципиально некоторая разметка M' из M_0 . С точки зрения исследования моделируемой системы, эта проблема интерпретируется как проблема достижимости (реализуемости) некоторого состояния системы.

– *Оценка живости переходов сети.* Под живостью перехода понимают возможность его срабатывания в данной сети при начальной разметке M_0 . Анализ модели на свойство живости позволяет выявить невозможные состояния в моделируемой системе (например, неисполняемые ветви в программе).

– *Оценка безопасности сети.* Безопасной является такая сеть Петри, в которой ни при каких условиях не может появиться более одной метки в каждой из позиций. Для исследуемой системы это означает возможность функционирования ее в стационарном режиме. На основе анализа данного свойства могут быть определены требования к буферным накопителям в системе.

Итак, достоинства сетей Петри заключаются в следующем:

– позволяют моделировать ПП всех возможных типов с учетом возможных конфликтов между ними;

– обладают наглядностью и обеспечивают возможность автоматизированного анализа;

– позволяют переходить от одного уровня детализации описания системы к другому (за счет раскрытия/закрытия переходов).

Вместе с тем, сети Петри имеют ряд недостатков, ограничивающих их возможности. Основной из них – время срабатывания перехода считается равным нулю, что не позволяет исследовать с помощью сетей Петри временные характеристики моделируемых систем.

E-сети

В результате развития аппарата сетей Петри был разработан ряд расширений. Одно из наиболее мощных – так называемые *E-сети* (evaluation

– «вычисления», «оценка») – «оценочные сети».

В отличие от сетей Петри, в E-сетях:

- имеются несколько типов вершин-позиций: простые позиции, позиции-очереди, разрешающие позиции;
- фишки (метки) могут снабжаться набором признаков (атрибутов);
- с каждым переходом может быть связана ненулевая задержка и функция преобразования атрибутов фишек;
- введены дополнительные виды вершин-переходов;
- в любую позицию может входить не более одной дуги и выходить также не более одной.

В связи с этим любой переход может быть описан тройкой параметров:

$$d_j = (S, t(d_j), p(d_j))$$

Здесь S – тип перехода, $t(d_j)$ – функция задержки, отражающая длительность срабатывания перехода, $p(d_j)$ – функция преобразования атрибутов меток.

Еще одно важное отличие E-сетей от сетей Петри состоит в том, что метки интерпретируются как транзакты, перемещающиеся по сети, а вершины-переходы трактуются как устройства, выполняющие ту или иную обработку транзактов. Следствием такого подхода является требование: ни одна вершина-позиция E-сети не может содержать более одной метки (то есть любая E-сеть изначально является безопасной).

Базовые переходы E-сети описаны ниже.

T-переход («исполнение», «простой переход»). Его графическое представление аналогично представлению вершины-перехода сети Петри (рис. 28, слева). Переход срабатывает при наличии метки во входной позиции и отсутствии ее в выходной позиции. Формально это можно записать так:

$$(1; 0) \text{ — } (0; 1)$$

T-переход позволяет отразить в модели занятость некоторого устройства (подсистемы) в течение некоторого времени, определяемого параметром $t(d)$.

F-переход («разветвление»). Графическое представление приведено на

рис. 28 в центре. Срабатывает при тех же условиях, что и Т-переход:

$$(1; 0, 0) \vdash (0; 1, 1)$$

С содержательной точки зрения, F-переход отображает разветвление потока информации (транзактов) в системе.

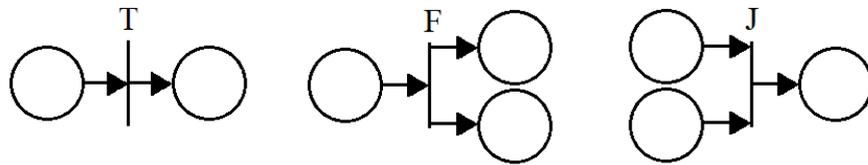


Рис. 28 – Графическое представление переходов E-сети – Т-переход (слева), F-переход (в центре), J-переход (справа)

J-переход («объединение»). Графическое обозначение показано на рис. 28 справа. Переход срабатывает при наличии меток в обеих входных позициях и отсутствии метки в выходной позиции:

$$(1, 1; 0) \vdash (0, 0; 1)$$

Он моделирует объединение потоков или наличие нескольких условий, определяющих некоторое событие.

X-переход («переключатель»). По сравнению с тремя предыдущими типами переходов, он содержит дополнительную управляющую («разрешающую») позицию, которая графически обозначается обычно либо квадратиком, либо шестиугольником (рис. 29, слева).

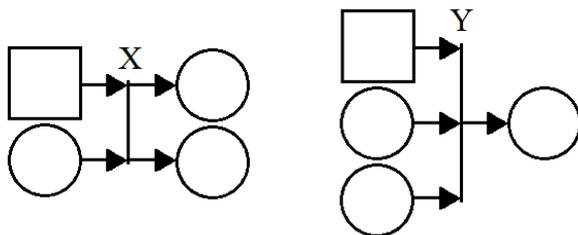


Рис 29 – Графическое представление переходов E-сети, имеющих разрешающую позицию – X-переход (слева), Y-переход (справа)

Логика его срабатывания задается следующими соотношениями:

$$\begin{array}{l} (0; 1; 0, 0) \vdash (0; 0; 1, 0) \\ (0; 1; 0, 1) \vdash (0; 0; 1, 1) \end{array}$$

$$\begin{array}{l|l} (1; 1; 0, 0) & \text{—} (0; 0; 0, 1) \\ (1; 1; 1, 0) & \text{—} (0; 0; 1, 1) \end{array}$$

X-переход изменяет направление потока информации (транзактов). В общем случае разрешающая процедура может быть сколь угодно сложной, но сущность ее работы заключается в проверке выполнения условий разветвления потока (сточки зрения программиста, разрешающая позиция аналогична условной инструкции типа if). Y-переход («выбор», «приоритетный выбор»). Этот переход также содержит разрешающую позицию (рис. 29, справа). Логика срабатывания Y-перехода:

$$\begin{array}{l|l} (0, 1, 1, 0) & \text{—} (0, 0, 1, 1) \\ (0, 1, 0, 0) & \text{—} (0, 0, 0, 1) \\ (0, 0, 1, 0) & \text{—} (0, 0, 0, 1) \\ (1, 1, 1, 0) & \text{—} (0, 1, 0, 1) \\ (1, 1, 0, 0) & \text{—} (0, 0, 0, 1) \\ (1, 0, 1, 0) & \text{—} (0, 0, 0, 1) \end{array}$$

Y-переход отражает приоритетность одних потоков информации (транзактов) по сравнению с другими. При этом разрешающая процедура может быть определена различным образом: как операция сравнения фиксированных приоритетов меток; как функция от атрибутов меток (например, чем меньше время обслуживания, тем выше приоритет). В некотором смысле он работает аналогично инструкции выбора типа case.

Еще раз подчеркнем, что в E-сети все переходы обладают свойством безопасности. Это означает, что в выходных позициях (которые, в свою очередь, могут быть входными для следующего перехода) никогда не может быть более одной метки. Вместе с тем, в E-сетях существуют понятия макроперехода и макропозиции, которые позволяют отображать в модели процессы накопления обслуживаемых транзактов в тех или иных узлах системы, а также расширить логические возможности E-сетей.

Рассмотрим некоторые из них.

Макропозиция *очередь* представляет собой линейную композицию T-переходов; суммарное количество выходных вершин-позиций определяет «емкость» очереди.

Макропозиция *генератор* позволяет представлять в сети источник меток

(транзактов). Если необходимо задать закон формирования меток, то «генератор» может быть дополнен разрешающей позицией.

Поскольку в Е-сети нельзя «накапливать» метки, то вводится макропозиция *поглощения* (или *аккумулятор*).

В целях повышения компактности и наглядности Е-сети для обозначения макропозиций используют специальные символы:

- Q – очередь;
- G – генератор;
- A – аккумулятор.

Аналогичным образом, путем композиции N однотипных переходов могут быть получены макропереходы всех типов: X_N , Y_N , J_N .

Рассмотренные особенности Е-сетей существенно расширяют их возможности для моделирования дискретных систем вообще и параллельных процессов в частности. Ниже приведен пример описания в виде Е-сети мультипрограммной вычислительной системы (рис. 30). Обработка поступающих заданий организована в ней по принципу квантования времени: каждому заданию выделяется равный отрезок (квант) процессорного времени; если задание выполнено, то оно покидает систему, если же времени оказалось недостаточно, то задание встает в очередь и ждет повторного выделения кванта времени.

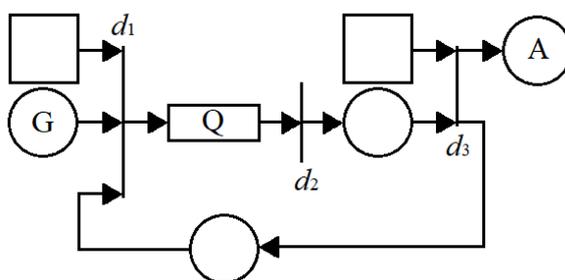


Рис. 30 – Пример описания вычислительной системы в виде Е-сети

На рисунке использованы следующие обозначения:

- d_1 – постановка задания в очередь;

- d_2 – выполнение задания в течение одного кванта времени;
- d_3 – анализ степени завершенности задания.

Помимо очевидных достоинств E-сетей, проявленное к ним внимание объясняется еще и тем, что технология моделирования систем в виде E-сетей весьма эффективно реализуется с помощью инструмента SIMULINK, входящего в состав пакета MATLAB.

6. Обработка и анализ результатов эксперимента Васильков

6.1. Обработка табличных данных

В этом разделе рассматриваются только основные направления обработки данных: интерполяции и аппроксимации, являющиеся базой для решения всех других задач обработки табличных данных, а также методы численного интегрирования. Другие направления, такие, как сглаживание, получение наилучших моделей и т.п., основанные на идеях интерполяции и аппроксимации, здесь не рассматриваются [1].

Интерполяция

Основная задача *интерполяции* – нахождение значения таблично заданной функции в тех точках внутри данного интервала, где она не задана. *Экстраполяция* – несколько более "широкое" понятие, оно сводится к восстановлению функции в точках за пределами заданного интервала. В обоих случаях исходные табличные данные могут быть получены как экспериментально (в этом случае принципиально отсутствуют промежуточные данные без дополнительных работ), так и расчетным путем по сложным зависимостям (в этом случае найти с помощью интерполяции значение сложной функции бывает проще, чем непосредственным вычислением по сложной формуле).

Решение задач интерполяции и экстраполяции обеспечивается построением интерполяционной функции $L(x)$, приближенно заменяющей исходную $f(x)$, заданную таблично, и проходящей через все заданные точки – узлы интерполяции. С помощью этой функции можно рассчитать искомое значение исходной функции в любой точке.

В связи с интерполяцией рассматриваются три основные проблемы:

- 1) выбор интерполяционной функции $L(x)$;
- 2) оценка погрешности интерполяции $R(x)$;
- 3) размещение узлов интерполяции для обеспечения наивысшей возможной точности восстановления функции (x_1, x_2, \dots, x_n) .

Специальные методы интерполяции позволяют определить искомое значение функции без непосредственного прямого построения интерполяционной функции. В принципе все интерполяционные методы, базирующиеся на использовании в качестве интерполяционной функции полиномов, дают одни и те же результаты, но с разными затратами. Это объясняется тем, что полином n -й степени, содержащий $n + 1$ параметр и проходящий через все заданные $n + 1$ точки, – единственный. Кроме того, полином можно представить, как усеченный ряд Тейлора, в который разложили исходную дифференцируемую функцию. Это, пожалуй, одно из главных достоинств полинома как интерполяционной функции. Поэтому чаще первая проблема интерполяции решается выбором в качестве интерполяционной функции именно полинома, хотя могут применяться и другие функции (например, тригонометрические полиномы, другие функции, выбранные из неформальных условий содержательной задачи).

Выбор вида интерполяционной функции является в общем случае важной задачей особенно если помнить, что через заданные точки можно провести любое количество функций (рис. 31). Следует отметить, что существует очевидный способ построения интерполяционной функции: из условия прохождения функции через все точки составляется система уравнений, из решения которой и находятся ее параметры. Однако этот путь далеко не самый эффективный особенно при большом числе точек.

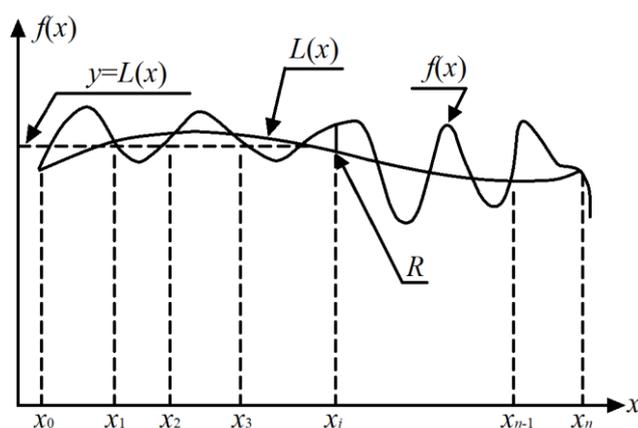


Рис. 31 – Иллюстрация интерполяции

Метод Лагранжа

Пусть известны значения некоторой функции $f(x)$ в $n + 1$ произвольных различных точках $y_i = f(x_i)$, $i = 0, \dots, n$. Для интерполирования (восстановления) функции в какой-либо точке x_i принадлежащей отрезку $[x_0, \dots, x_n]$ необходимо построить интерполяционный полином n -го порядка, который в методе Лагранжа представляется следующим образом:

$$L_n(x) = y_0 Q_0(x) + \dots + y_j Q_j(x) + \dots + y_n Q_n(x),$$
$$Q_j(x) = \frac{(x - x_0) \dots (x - x_{j-1})(x - x_{j+1}) \dots (x - x_n)}{(x_j - x_0) \dots (x_j - x_{j-1})(x_j - x_{j+1}) \dots (x_j - x_n)}.$$

Причем нетрудно заметить, что $Q_j(x_i) = 0$, если $i \neq j$ и что $Q_j(x_i) = 1$, если $i = j$. Если раскрыть произведение всех скобок в числителе (в знаменателе все скобки – числа), то получим полином n -го порядка от x , так как в числителе содержится n сомножителей первого порядка. Следовательно, интерполяционный полином Лагранжа не что иное, как обычный полином n -го порядка, несмотря на специфическую форму записи.

Оценить погрешность интерполяции в точке x из $[x_0, x_n]$ (т.е. решить вторую проблему интерполяции) можно по формуле:

$$R(x) = |f(x) - L_n(x)| \leq \frac{M_{n+1}}{(n+1)!} \prod_{i=0}^n (x - x_i).$$

В формуле $M_{n+1} = \max|f^{(n+1)}(x)|$ – максимальное значение $(n+1)$ -й производной исходной функции $f(x)$ на отрезке $[x_0, x_n]$. Следовательно, для того чтобы оценить погрешность интерполяции, необходима некоторая дополнительная информация об исходной функции (это должно быть понятно, так как через заданные исходные точки может проходить бесчисленное количество различных функций, для которых и погрешность будет разной). Такой информацией является производная $n+1$ порядка, которую не так

просто найти. Отметим также, что применение формулы погрешности возможно, только если функция дифференцируема $n+1$ раз.

Метод Ньютона

Пусть известны значения некоторой функции $f(x)$ в $n+1$ произвольных, попарно не совпадающих точках $y_i = f(x_i)$, $i = 0, \dots, n$. В общем случае интерполяция по формулам Ньютона может производиться для произвольно расположенных узлов интерполяции, но чаще применяется для равномерно расположенных. Поэтому далее рассматривается только случай с равномерным расположением узлов. Тогда $x_{i+1} = x_i + h$, где $h = (x_0 - x_n)/n$.

Метод использует понятие конечных разностей. Конечная разность k -го порядка в i -й точке вычисляется следующим образом:

$$\Delta^k y_i = \Delta^{k-1} y_{i+1} - \Delta^{k-1} y_i,$$

т.е. через конечные разности более низкого порядка. Можно заметить, что при наличии $n+1$ точки $(0, 1, 2, \dots, n)$ конечную разность первого порядка можно вычислить только для первых n точек, конечную разность n -го порядка – только для нулевой точки, конечную разность k -го порядка – только для первых $n-k+1$ точек. Интерполяционный многочлен Ньютона записывается следующим образом:

$$P_n(x) = y_0 + \frac{\Delta y_0(x - x_0)}{h} + \frac{\Delta^2 y_0(x - x_0)(x - x_1)}{2! h^2} + \frac{\Delta^3 y_0(x - x_0)(x - x_1)(x - x_2)}{3! h^3} + \dots + \frac{\Delta^n y_0(x - x_0)(x - x_1)(x - x_2) \dots (x - x_{n-1})}{n! h^n}.$$

Отсюда, нетрудно заметить, что, как и в предыдущем рассмотренном методе, это выражение есть не что иное, как обычный полином n -го порядка

от x , только записанный в другой форме. Это можно доказать, раскрыв все скобки и приведя подобные члены. Конечные разности в выражении – это числовые коэффициенты, вычисленные по заданным точкам.

Часто вводят безразмерную переменную q , показывающую, сколько содержится шагов от x_0 до заданной точки x : $q = (x - x_0)/h$.

В этом случае выражение для интерполяционного полинома запишется так:

$$P_n(x) = y_0 + \Delta y_0 q + \frac{\Delta^2 y_0 q(q-1)}{2!} + \frac{\Delta^3 y_0 q(q-1)(q-2)}{3!} + \dots + \frac{\Delta^n y_0 q(q-1)(q-2) \dots (q-n+1)}{n!}.$$

Обе приведенные формулы носят название первой интерполяционной формулы Ньютона (только записанные в разных переменных – x и q) и рекомендуются для применения при интерполяции вперед (в сторону увеличения x) или при экстраполяции назад (левее x_0).

Существует вторая формула, которая рекомендуется для применения при интерполяции назад (т.е. в конце интервала, но левее x_n) или при экстраполяции вперед, правее x_n . Она записывается следующим образом:

$$P_n(x) = y_n + \Delta y_{n-1} q + \frac{\Delta^2 y_{n-2} q(q+1)}{2!} + \frac{\Delta^3 y_{n-3} q(q+1)(q+2)}{3!} + \dots + \frac{\Delta^n y_0 q(q+1)(q+2) \dots (q+n-1)}{n!}.$$

Погрешность интерполяции можно оценить так же, как и в предыдущем методе. Хотя при использовании относительной переменной q можно указать и другую формулу (которая, естественно, полностью по результату совпадает с предыдущей формулой погрешности):

$$R(x) = |f(x) - L_n(x)| \leq \frac{M_{n+1} h^{n+1}}{(n+1)!} q(q-1)(q-2) \dots (q-n).$$

Использование конечных разностей, которые для дискретных функций являются своеобразными аналогами производных непрерывных функций, помогает находить погрешность интерполяции. Учтем соотношения:

$$f' \approx \frac{\Delta y}{h}; \quad f'' \approx \frac{\Delta^2 y}{h^2}; \quad f''' \approx \frac{\Delta^3 y}{h^3}; \quad \dots; \quad f^{(i)} \approx \frac{\Delta^i y}{h^i},$$

тогда для получения приближенного значения M_{n+1} достаточно иметь несколько (или одну) дополнительных точек x_{n+1}, x_{n+2}, \dots , с использованием которых легко найти максимальное значение конечной разности (или даже одно значение $\Delta^{n+1} y(x_0)$, если есть только одна дополнительная точка) $(n+1)$ -го порядка $\max\{\Delta^{n+1} y(x)\}$.

$$R(x) = |f(x) - L_n(x)| \leq \frac{\max\{|\Delta^{n+1} y(x)|\}}{(n+1)!} q(q-1)(q-2) \dots (q-n).$$

Таким же путем можно поступить и в методе Лагранжа, но при условии, если он реализуется при равномерном шаге.

Метод Чебышева

Этот метод в основном является иллюстрацией решений третьей проблемы интерполяции: выбор узлов интерполяции (если, конечно, это возможно при решении конкретной задачи) для получения при заданном числе узлов минимально возможной погрешности. Применение метода Чебышева позволяет обосновать размещение узлов интерполяции в заданных пределах на числовой оси. Предварительно переменная x преобразуется в переменную z , изменяющуюся от -1 до 1 при изменении x от x_0 до x_n .

$$z = -\frac{x_0 + x_n}{x_n - x_0} + \frac{2x}{x_n - x_0}.$$

В свою очередь переменная z на интервале $[-1, 1]$ формирует узлы следующим образом:

$$z_i = -\cos\left(\frac{i\pi}{n}\right).$$

Из этого соотношения видно, что узлы интерполяции располагаются неравномерно, сгущаясь к концам отрезка $[x_0, x_n]$. В вычислительном плане для интерполяции можно пользоваться интерполяционной формулой Лагранжа.

Метод сплайнов

В ряде случаев возникает задача восстановления не только значений функции, но также ее первой и второй производной. Это, конечно, можно сделать путем дифференцирования интерполяционного полинома, если его получить. Но оказывается, что точность восстановления производных в этих случаях неудовлетворительно низкая. Для решения указанного класса задач успешно применяется *сплайновая интерполяция*.

Наибольшее распространение получила интерполяция с помощью кубических сплайнов. *Сплайн* – это функция, которая на каждом межузловом интервале совпадает с некоторым полиномом, своим для каждого интервала. Полиномы соседних интервалов стыкуются так, чтобы функция была непрерывной. Дополнительно требуют непрерывности нескольких производных (в кубических сплайнах – двух). *Кубический сплайн* склеивается из полиномов третьей степени, которые для i -го участка записываются так:

$$y = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i.$$

Для всего интервала будет соответственно n кубических полиномов, отличающихся коэффициентами a_i, b_i, c_i, d_i . Чаще всего узлы при сплайновой интерполяции располагают равномерно, т.е. $x_{i+1} - x_i = \text{const} = h$ (хотя это и необязательно).

Необходимо найти четыре коэффициента при условии прохождения каждого полинома через две точки (x_i, y_i) и (x_{i+1}, y_{i+1}) , следствием чего являются следующие очевидные уравнения:

$$\begin{aligned} y_i &= d_i, & i &= 0, 1, 2, \dots, n-1, \\ y_{i+1} &= a_i h^3 + b_i h^2 + c_i h + d_i, & i &= 0, 1, 2, \dots, n-1. \end{aligned}$$

Первое условие соответствует прохождению полинома через начальную точку, второе – через конечную точку. Найти все коэффициенты из этих уравнений нельзя, так как условий меньше, чем искомым параметров. Поэтому указанные условия дополняют условиями гладкости функции (т.е. непрерывности первой производной) и гладкости первой производной (т.е. непрерывности второй производной) в узлах интерполяции. Математически эти условия записываются как равенства соответственно первой и второй производных в конце i -го и в начале $(i+1)$ -го участков.

$$\text{Так как } y' = 3a_i(x-x_i)^2 + 2b_i(x-x_i) + c_i \text{ и } y'' = 6a_i(x-x_i) + 2b_i,$$

то,

$$3a_i h^2 + 2b_i h + c_i = c_{i+1}, \quad i = 0, 1, 2, \dots, n-2,$$

$(y'(x_{i+1}))$ в конце i -го участка равна $y'(x_{i+1})$ в начале $(i+1)$ -го),

$$6a_i h + 2b_i = 2b_{i+1}, \quad i = 0, 1, 2, \dots, n-2,$$

$(y''(x_{i+1}))$ в конце i -го участка равна $y''(x_{i+1})$ в начале $(i+1)$ -го).

Получилась система линейных уравнений (для всех участков), содержащая $4n-2$ уравнения с $4n$ неизвестными (неизвестные $a_1, a_2, \dots, a_n, b_1, \dots, b_n$ – коэффициенты сплайнов). Для решения системы добавляют два граничных условия одного из следующих видов (чаще применяют 1):

- 1) $y''(x_0) = y''(x_n) = 0$,
- 2) $y'(x_0) = g_0, \quad y'(x_n) = g_n$,
- 3) $y'(x_0) = y'(x_n) = 0$,
- 4) $y'(x_0) = y'(x_n), \quad y''(x_0) = y''(x_n)$.

Совместное решение $4n$ уравнений позволяет найти все $4n$ коэффициента.

Для восстановления производных можно продифференцировать на каждом участке соответствующий кубический полином. В случае необходимости определения производных в узлах существуют специальные приемы, сводящие определение производных к решению более простой системы уравнений относительно искомым производных второго или первого порядка. К важным достоинствам интерполяции кубическими сплайнами относится получение функции, имеющей минимальную возможную кривизну. к недостаткам сплайновой интерполяции относится необходимость получения сравнительно большого числа параметров.

6.2 Аппроксимация

Основная задача аппроксимации – построение приближенной (аппроксимирующей) функции, в целом наиболее близко проходящей около данных точек или около данной непрерывной функции. Такая задача возникает при наличии погрешности в исходных данных (в этом случае нецелесообразно проводить функцию точно через все точки, как в интерполяции) или при желании получить упрощенное математическое описание сложной или неизвестной зависимости.

Близость исходной и аппроксимирующей функций определяется числовой мерой – критерием аппроксимации (близости). Наибольшее распространение получил квадратичный критерий, равный сумме квадратов отклонений расчетных значений от «экспериментальных» (т.е. заданных), – критерий близости в заданных точках:

$$R = \sum_{i=1}^n \beta_i (y_i - y_i^{\text{расч.}})^2.$$

где y_i – заданные табличные значения функции; $y_i^{\text{расч.}}$ – расчетные значения по аппроксимирующей функции; β_i – весовые коэффициенты, учитывающие относительную важность i -й точки (увеличение β_i – приводит при стремлении уменьшить R к уменьшению прежде всего отклонения в i -и точке, так как это отклонение искусственно увеличено за счет относительно большого значения весового коэффициента).

Квадратичный критерий обладает рядом «хороших» свойств, таких, как дифференцируемость, обеспечение единственного решения задачи аппроксимации при полиномиальных аппроксимирующих функциях.

Другим распространенным критерием близости является следующий:

$$R = \max_i |y_i - y_i^{\text{расч.}}|.$$

Этот критерий менее распространен в связи с аналитическими и вычислительными трудностями, связанными с отсутствием гладкости функции и ее дифференцируемости.

В обоих рассмотренных случаях в качестве значения функции y_i можно брать не только абсолютные, но и относительные значения, например, y_i/y_n и др.

Выделяют две основные задачи:

- 1) получение аппроксимирующей функции, описывающей имеющиеся данные, с погрешностью, не хуже заданной;
- 2) получение аппроксимирующей функции заданной структуры с наилучшей возможной погрешностью.

Чаще всего первая задача сводится ко второй перебором различных аппроксимирующих функций и последующим выбором наилучшей.

Метод наименьших квадратов

Метод базируется на применении в качестве критерия близости суммы квадратов отклонений заданных и расчетных значений. При заданной структуре аппроксимирующей функции $y_i^{\text{расч}}(x)$ необходимо таким образом подобрать параметры этой функции, чтобы получить наименьшее значение критерия близости, т.е. наилучшую аппроксимацию. Рассмотрим путь нахождения этих параметров на примере полиномиальной функции одной переменной:

$$y_i^{\text{расч.}}(x) = a_k x^k + a_{k-1} x^{k-1} + \dots + a_0 = \sum_{j=0}^k a_j x^j.$$

Запишем выражение критерия аппроксимации при $\beta_i = 1$ ($i = 1, 2, \dots, n$) для полиномиального $y_i^{\text{расч}}(x)$:

$$R = \sum_{i=1}^n \left(y_i - \sum_{j=0}^k a_j x_i^j \right)^2.$$

Искомые переменные a_j можно найти из необходимого условия минимума R по этим переменным, т.е. $dR/da_p = 0$ (для $p = 0, 1, 2, \dots, k$). Продифференцируем по a_p (p – текущий индекс):

$$dR/da_p = 2 \sum_{i=1}^n \left(y_i - \sum_{j=0}^k a_j x_i^j \right) (-x_i^p) = 0, \quad p = 0, 1, 2, \dots, k, \quad i = 1, 2, \dots, n.$$

Значения индексов суммирования для простоты пропущены.

После очевидных преобразований (сокращение на два, раскрытие скобок, изменение порядка суммирования) получим

$$dR/da_p = \sum_{i=1}^n y_i(-x_i^p) + \sum_{i=1}^n \sum_{j=0}^k a_j x_i^j x_i^p = \sum_{i=1}^n y_i(-x_i^p) + \sum_{j=0}^k a_j \sum_{i=1}^n x_i^j x_i^p = 0,$$

$$p = 0, 1, 2, \dots, k, \quad i = 1, 2, \dots, n.$$

Перепишем последние равенства.

$$\sum_{j=0}^k a_j \sum_{i=1}^n x_i^j x_i^p = \sum_{i=1}^n y_i x_i^p, \quad p = 0, 1, 2, \dots, k, \quad i = 1, 2, \dots, n.$$

Получилась система $n+1$ уравнений с таким же количеством неизвестных a_j , причем линейная относительно этих переменных. Эта система называется системой нормальных уравнений. Из ее решения находятся параметры a_j аппроксимирующей функции, обеспечивающие $\min R$, т.е. наилучшее возможное квадратичное приближение. Зная коэффициенты, можно (если нужно) вычислить и величину R (например, для сравнения различных аппроксимирующих функций). Следует помнить, что при изменении даже одного значения исходных данных (или пары значений x_i, y_i , или одного из них) все коэффициенты изменят в общем случае свои значения, так как они полностью определяются исходными данными. Поэтому при повторении аппроксимации с несколько изменившимися данными (например, вследствие погрешностей измерения, помех, влияния неучтенных факторов и т.п.) получится другая аппроксимирующая функция, отличающаяся коэффициентами.

Обратим внимание на то, что коэффициенты a_j полинома находятся из решения системы уравнений, т.е. они связаны между собой. Это приводит к тому, что если какой-то коэффициент вследствие его малости захочется отбросить, придется пересчитывать заново оставшиеся. Можно рассчитать количественные оценки тесноты связи коэффициентов. Существует специальная теория планирования экспериментов, которая позволяет обосновать и рассчитать значения x_i , используемые для аппроксимации, чтобы получить заданные свойства коэффициентов (несвязанность, минимальная дисперсия коэффициентов и т.д.)

или аппроксимирующей функции (равная точность описания реальной зависимости в различных направлениях, минимальная дисперсия предсказания значения функции и т.д.).

В случае постановки другой задачи – найти аппроксимирующую функцию, обеспечивающую погрешность не хуже заданной, – необходимо подбирать и структуру этой функции. Эта задача значительно сложнее предыдущей (найти параметры аппроксимирующей функции заданной структуры, обеспечивающей наилучшую возможную погрешность) и решается в основном путем перебора различных функций и сравнения получающихся мер близости. Для примера на рис. 32 приведены для визуального сравнения исходная и аппроксимирующие функции с различной степенью полинома, т.е. функции с различной структурой. Не следует забывать, что с повышением точности аппроксимации растет и сложность функции (при полиномиальных аппроксимирующих функциях), что делает ее менее удобной при использовании.

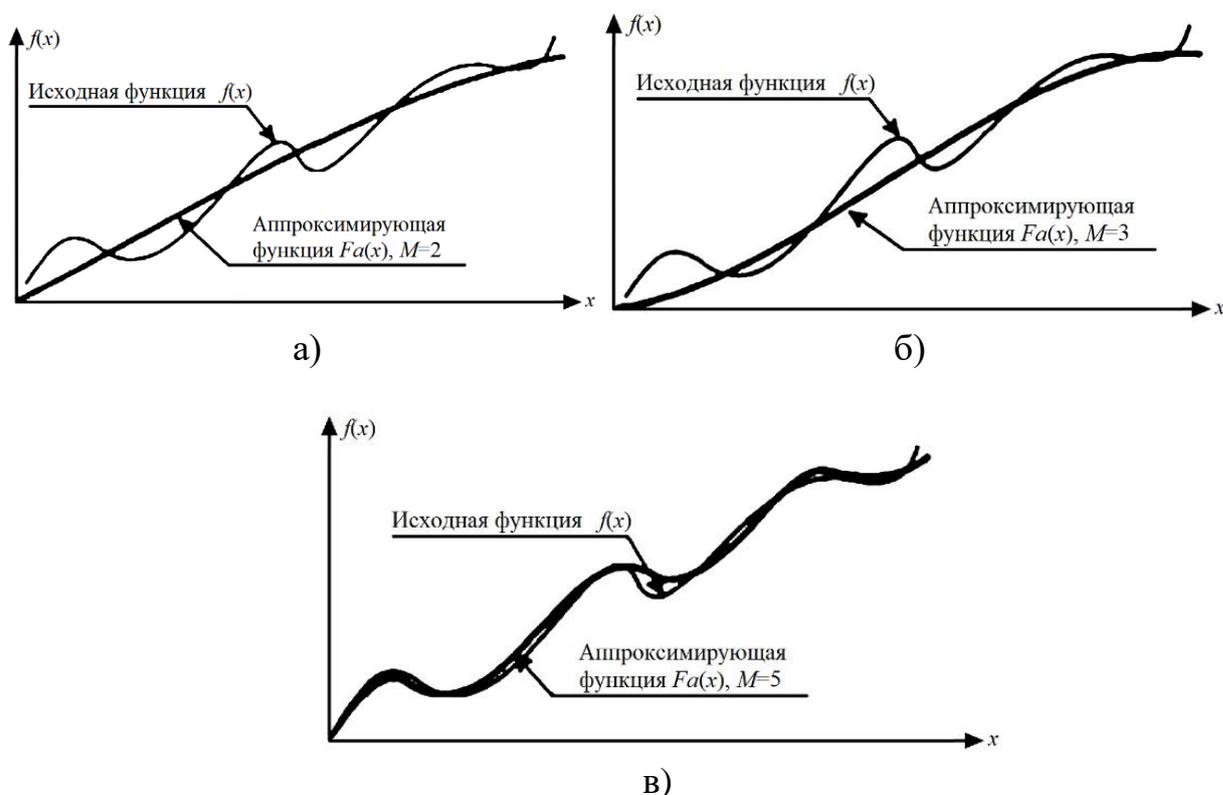


Рис. 32 – Влияние степени аппроксимирующего полинома M на точность аппроксимации: а – $M = 2$, б – $M = 3$, в – $M = 5$

Метод равномерного приближения

Равномерным приближением называется аппроксимация, в которой в качестве критерия близости используется модуль максимального отклонения расчетных и заданных значений. При этом решается задача минимизации этого критерия, т.е.

$$R = \max |y_i - y_i^{\text{расч.}}| \rightarrow \min.$$

Принято считать, что наилучшее равномерное приближение обеспечивает несколько лучшую аппроксимацию, чем среднеквадратичное. Однако теоретические оценки показывают незначительность такого преимущества; оно более значительно для функций с непрерывными старшими производными, не слишком большими по абсолютной величине, при небольших объемах исходных данных. Однако существенный недостаток – отсутствие эффективных вычислительных алгоритмов (кроме прямого поиска $\min R$ методами нелинейного программирования, причем необходимо применять методы, пригодные для не дифференцируемых функций) – делает малоприменимым данный способ аппроксимации.

6.3. Методы численного интегрирования

Все численные методы строятся на том, что подынтегральная функция приближенно заменяется более простой (горизонтальной или наклонной прямой, параболой 2-го, 3-го или более высокого порядка), от которой интеграл легко берется. В результате получаются формулы интегрирования, называемые квадратурными, в виде взвешенной суммы ординат подынтегральной функции в отдельных точках:

$$\int_a^b f(x) dx \approx \sum \omega_i f(x_i).$$

Чем меньше интервалы, на которых производят замену, тем точнее вычисляется интеграл. Поэтому исходный отрезок $[a, b]$ для повышения точности делят на несколько равных или неравных интервалов, на каждом из которых применяют формулу интегрирования, а затем складывают результаты.

Все методы различаются значениями ординат x_i , и весов ω_i .

В большинстве случаев погрешность численного интегрирования определяется путем двойного интегрирования: с исходным шагом (шаг определяется путем равномерного деления отрезка $b - a$ на число отрезков n : $h = (b - a) / n$) и с шагом, увеличенным в 2 раза. Разница вычисленных значений интегралов определяет погрешность.

Сравнение эффективности различных методов проводится по степени полинома, который данным методом интегрируется точно, без ошибки. Чем выше степень такого полинома, тем выше точность метода, тем он эффективнее.

Простейшие методы

К простейшим методам можно отнести методы прямоугольников (левых и правых) и трапеций. В первом случае подынтегральная функция заменяется горизонтальной прямой ($y = c_0$) со значением ординаты, т.е. значения функции соответственно слева или справа участка, во втором случае – наклонной прямой ($y = c_1x + c_0$). Формулы интегрирования при разбиении отрезка $[a, b]$ на n частей с равномерным шагом h соответственно приобретают вид:

– для одного участка интегрирования:

$$I = \int_a^b f(x)dx \approx f(a)(b - a) \approx f(x_0)h,$$

$$I = \int_a^b f(x)dx \approx f(a)(b-a) \approx f(x_1)h,$$

$$\omega_0 = h = \text{const}$$

$$I = \int_a^b f(x)dx \approx \frac{f(a) + f(b)}{2(b-a)} \approx f(x_0)h/2 + f(x_1)h/2,$$

$$\omega_0 = \omega_1 = h/2 = \text{const},$$

где $x_0 = a$, $x_1 = b$, $h = b - a$;

– для n участков интегрирования:

$$I = \int_a^b f(x)dx \approx \sum_{i=0}^{n-1} f(x_i)h \approx \sum_{i=0}^{n-1} \omega_i f(x_i),$$

$$I = \int_a^b f(x)dx \approx \sum_{i=0}^{n-1} f(x_i)h \approx \sum_{i=0}^{n-1} \omega_i f(x_i),$$

$$\omega_i = h = \text{const},$$

$$I = \int_a^b f(x)dx \approx \sum_{i=0}^{n-1} (f(x_i) + f(x_{i+1})) \frac{h}{2} + \sum_{i=0}^{n-1} f(x_i)h + f(x_n) \frac{h}{2} \approx \sum_{i=0}^n \omega_i f(x_i),$$

$$\omega_0 = \omega_1 = h/2 = \text{const}, \quad \omega_1 = \omega_2 = \dots = \omega_{n-1} = h.$$

Нетрудно заметить, что в методе прямоугольников интеграл вычисляется абсолютно точно только при $f(x) = c$ (const), а в методе трапеций – при $f(x)$ линейной или кусочно-линейной.

На рис. 33 для сравнения приведены примеры прямоугольников при различном числе участков. Наглядно видно, что площадь всех прямоугольников на правом рисунке меньше отличается от площади под кривой $f(x)$, чем на левом.

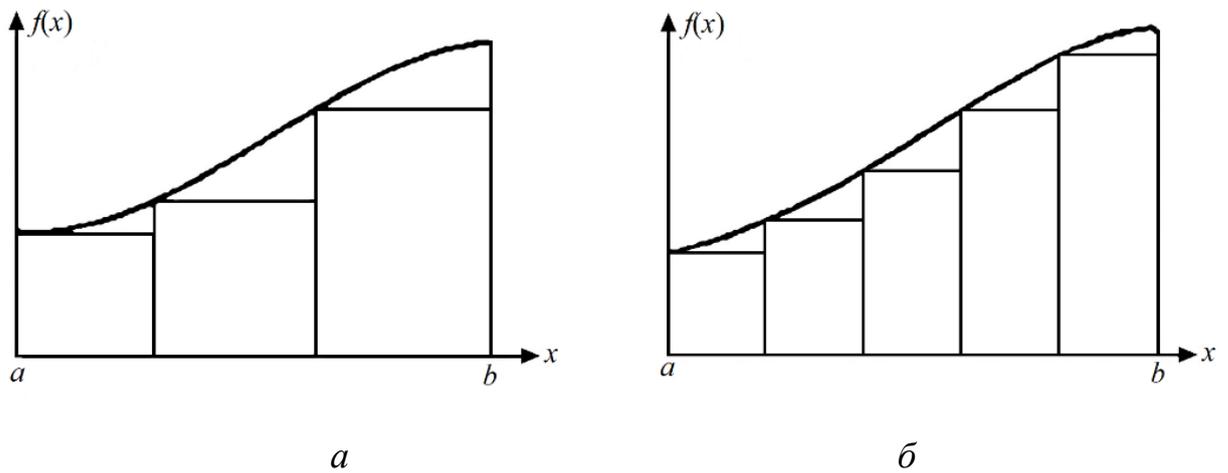


Рис. 33 – Иллюстрация метода левых прямоугольников:
a – с 3 участками разбиения отрезка интегрирования $[a, b]$
б – с 5 участками разбиения отрезка интегрирования $[a, b]$

Метод прямоугольников не находит практического применения в силу значительных погрешностей, что тоже видно из рисунка 6.1

На рис. 34 приведен пример вычисления интеграла методом трапеций. По сравнению с методом прямоугольников метод трапеций более точный, так как трапеция точнее заменяет соответствующую криволинейную трапецию, чем прямоугольник.

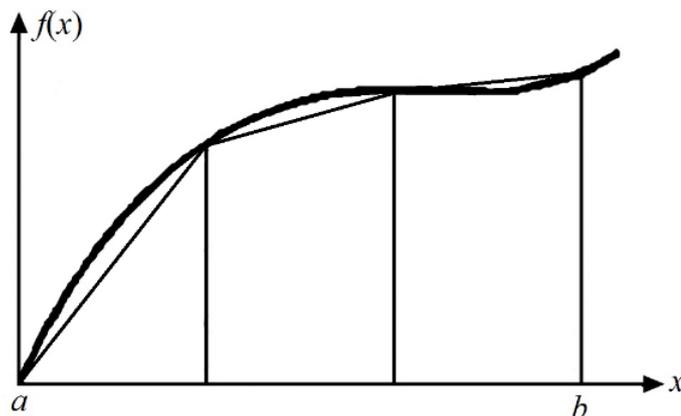


Рис. 34 – Иллюстрация метода трапеций

Погрешность R вычисления интеграла методом трапеций при использовании двойного просчета на практике может быть определена из следующего соотношения:

$$R \leq \frac{|I_n - I_{n/2}|}{3},$$

где I_n и $I_{n/2}$ – соответственно значения интеграла при числе разбиений n и $n/2$. Существуют и аналитические выражения для определения погрешности, но они требуют знания второй производной подынтегральной функции, поэтому имеют только теоретическое значение. С использованием двойного просчета можно организовать автоматический подбор шага интегрирования (т.е. числа разбиений n) для обеспечения заданной погрешности интегрирования (последовательно удваивая шаг и контролируя погрешность).

Метод Симпсона

Этот метод базируется на замене подынтегральной функции квадратичной параболой, которая строится уже не по двум (как прямая в методе трапеций), а по трем точкам на каждом участке. По этим трем точкам (крайние точки участка и средняя точка) строится интерполяционная функция – полином второго порядка, который аналитически интегрируется. Получается следующая расчетная формула:

– для одного участка интегрирования:

$$I = \int_a^b f(x)dx \approx \frac{h}{3}(f(x_0) + f(x_1) + f(x_2)),$$

$$\omega_0 = \omega_2 = \frac{h}{3}, \quad \omega_1 = \frac{4h}{3},$$

где $x_0 = a$, $x_1 = \frac{b+a}{2}$, $x_2 = b$, $h = \frac{b-a}{2}$,

– для n участков интегрирования:

$$I = \int_a^b f(x)dx \approx \frac{hf(x_0)}{3} + \frac{4h}{3} \sum_{i=1}^{n/2} f(x_{2i-1}) + \frac{2h}{3} \sum_{i=1}^{n/2} f(x_{2i}) + \frac{hf(x_n)}{3}$$

$$= \sum_{i=1}^n \omega_i f(x_i),$$

где $\omega_0 = \omega_n = \frac{h}{3}$, $\omega_2 = \omega_4 = \dots = \omega_{n-2} = \frac{2h}{3}$, $\omega_1 = \omega_3 = \omega_{n-1} = \frac{4h}{3}$.

В этой формуле все ординаты с нечетными номерами имеют коэффициент $4h/3$, а с четными – $2h/3$ (кроме нулевого и последнего). При работе с этим методом обязательно разбивают весь интервал на четное число участков.

На рис. 35 приведен пример вычисления интеграла методом Симпсона. По сравнению с методами прямоугольников и трапеций он более точный, что наглядно видно из графика (подынтегральная функция почти совпадает и обеспечивает вычисление интеграла точно, без погрешности при полиноме третьего порядка). Следовательно, этот метод предпочтительнее предыдущих. Количественно оценить погрешность при использовании двойного просчета можно по соотношению

$$R \leq \frac{|I_n - I_{n/2}|}{15},$$

т.е. при увеличении числа разбиений в два раза погрешность падает в 15 раз.

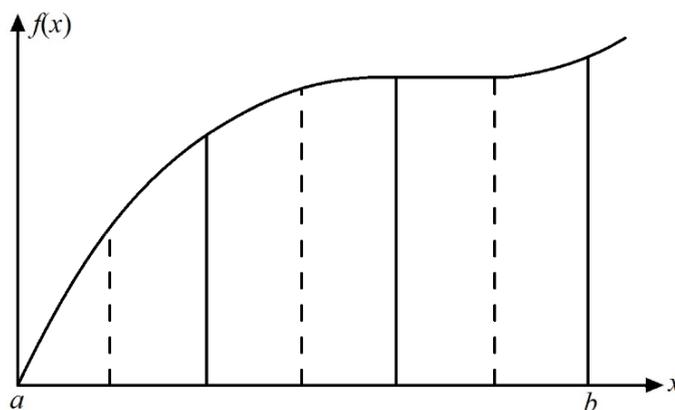


Рис. 35 – Иллюстрация метода Симпсона

Теоретические формулы оценки погрешности содержат производную четвертого порядка от подынтегральной функции, поэтому не имеют практического значения.

Метод Ньютона – Котеса

Данный метод является обобщением предыдущих, построен на аналогичных принципах и предполагает замену подынтегральной функции параболой k -го порядка (а не второго, как в методе Симпсона). Расчетная формула для одного (!) участка выглядит следующим образом:

$$I = \int_a^b f(x) dx \approx (b - a) \sum_{i=1}^k f(x_i) H_i = \sum_{i=1}^k \omega_i f(x_i), \quad \omega_i = (b - a) H_i,$$

где $x_0 = a$, $x_k = b$, $x_i = a + i(b - a)/k$, H_i – коэффициенты Ньютона – Котеса, а k – число используемых ординат на участке (начиная с 0), которые применяются для аппроксимации подынтегральной функции. Естественно, что для замены $f(x)$ параболой третьей степени потребуется уже четыре точки, а четвертой – пять точек. Коэффициенты H_i – не зависят от функции $f(x)$ и определены заранее. Некоторые из них приведены ниже.

$$\begin{array}{llll} k = 1; & H_0 = H_1 = 1/2; & & \\ k = 2; & H_0 = H_2 = 1/6; & H_1 = 2/3; & \\ k = 3; & H_0 = H_3 = 1/8; & H_1 = H_2 = 3/8; & \\ k = 4; & H_0 = H_4 = 7/90; & H_1 = H_3 = 16/45; & H_2 = 2/15; \\ k = 5; & H_0 = H_5 = 19/288; & H_1 = H_4 = 25/96; & H_2 = H_3 = 25/144; \\ \dots & \dots & \dots & \dots \end{array}$$

Все предыдущие формулы являются частным случаем формулы Ньютона – Котеса. В частности, при $k = 1$ получаем метод трапеций (для одного

участка), при $k = 2$ – формулу Симпсона. Не следует брать $k > 10$, так как при больших значениях k алгоритм оказывается неустойчивым. При разбиении всего интервала $[a, b]$ на n участков формулу нужно применять для каждого участка, а результаты сложить.

6.4 Методы решения дифференциальных уравнений

Дифференциальные уравнения очень часто встречаются при построении моделей динамики объектов исследования. Они описывают, как правило, изменение параметров объекта во времени (хотя могут быть и другие случаи). Результатом решения дифференциальных уравнений являются функции, а не числа, как при решении конечных уравнений, вследствие чего методы решения их более трудоемки. Владение методами решения дифференциальных уравнений обязательно при моделировании и очень важно для получения правильного результата.

При использовании численных методов решение дифференциальных уравнений $dy/dx = f(x, y)$ или $y' = f(x, y)$ представляется в табличном виде, т.е. получается совокупность значений y_i и x_i . Решение носит шаговый характер, т.е. по одной или по нескольким начальным точкам (x, y) за один шаг находят следующую точку, затем следующую и т.д. Разница между двумя соседними значениями аргумента $h = x_{i+1} - x_i$ называется *шагом*.

Наибольшее распространение имеют задачи Коши, в которых заданы начальные условия: при $x = x_0$ $y(x_0) = y_0$. Имея их, легко начинать процесс решения, т.е. найти y_1 при x_1 , y_2 – при x_2 и т.д. Задачи другого типа – краевые задачи (например, с конечными условиями или с условиями в промежуточной точке) – решаются специальными приемами, в том числе нередко сведением к другим эквивалентным задачам с начальными условиями.

Выделяют два класса методов решения: *одношаговые* и *многошаговые*. Первый класс методов требует для нахождения следующего значения функции только одной текущей точки, т.е. $y_{i+1} = F[f(x_i, y_i)]$, а второй – нескольких, например, $y_{i+1} = F(y_{i-3}, y_{i-2}, y_{i-1}, y_i)$. Поэтому методы второго класса

не обладают свойством "самостартования", т.е. ими нельзя начать решение задачи Коши, это всегда делается одношаговыми методами. К недостаткам многошаговых методов относится также и невозможность изменения в процессе решения величины шага (так как они используют предыдущие точки с ранее применяемым шагом, а учет меняющегося шага очень сложен и громоздок), что бывает необходимо для повышения эффективности метода. Заметим, что величина шага существенно влияет на точность и скорость решения, поэтому изменение ее в процессе решения – увеличение при медленно изменяющемся решении и уменьшение при быстро изменяющемся – очень важно для эффективности решения. К достоинствам многошаговых методов относят в основном меньший объем памяти компьютера, требующейся для реализации, и возможность теоретической оценки погрешности решения. Представителем класса многошаговых методов являются методы прогноза и коррекции. К классу одношаговых методов относятся методы Эйлера, Рунге – Кутты и др.

Основная идея получения простейших вычислительных алгоритмов в одношаговых методах сводится к разложению искомого решения $y(x)$ в ряд Тейлора в окрестности текущей точки и усечения его. Количество оставленных членов ряда определяет порядок и, следовательно, точность метода. По полученному разложению, зная значение y в точке разложения y_i и производную $f(x_i, y_i)$, находят значение функции y через шаг h : $y_{i+1} = y_i + \Delta y_i$. Если в разложении удерживается большее число членов, то необходимо рассчитывать $f(x_i, y_i)$ в нескольких точках (таким способом избегают необходимости прямого вычисления высших производных, присутствующих в разложении в ряд Тейлора). Расчетные алгоритмы многошаговых методов базируются на построении интерполяционных или аппроксимирующих функций, от которых берется интеграл.

Численными методами решаются не только отдельные уравнения, но и системы уравнений (чаще всего первого порядка), причем большинство методов решения одного уравнения легко распространяются на решение

систем. Дифференциальные уравнения высших порядков вида

$$y^{(n)} = f(x, y, y', y'', \dots, y^{(n-1)})$$

решаются в основном сведением к системе уравнений первого порядка путем замены переменных: $y_1 = y'$, $y_2 = y''$, $y_3 = y'''$ и т.д. При этом дифференциальное уравнение n -го порядка заменяется системой из n уравнений:

$$\begin{aligned}y' &= y_1, \\y'' &= y_2, \\y''' &= y_3, \\&\dots\dots\dots \\y'_{n-1} &= f(x, y, y_1, y_2, \dots, y_{n-1}).\end{aligned}$$

Метод Эйлера и модифицированный метод Эйлера

Метод Эйлера является одним из самых простых методов решения дифференциальных уравнений первого порядка $y' = f(x, y)$. Он в основном используется как учебный, в практических расчетах он дает значительную погрешность. Вычислительный алгоритм представляется следующим образом:

$$y_{i+1} = y(x_i + h) = y_i + \Delta y_i = y_i + hf(x_i, y_i)$$

где h – шаг по x (в общем случае может быть непостоянным). Запускается метод из начальных условий $y(x_0) = y_0$.

На рис. 36 приведена графическая интерпретация метода Эйлера. Величины α_1 и α_2 определяются из условий: $\operatorname{tg} \alpha_1 = f(x_0, y_0)$ и $\operatorname{tg} \alpha_2 = f(x_1, y_1)$, $\Delta y_1 = \Delta x f(x_0, y_0)$, $\Delta y_2 = \Delta x f(x_1, y_1)$ – расчетные значения функции – по соответствующим соотношениям.

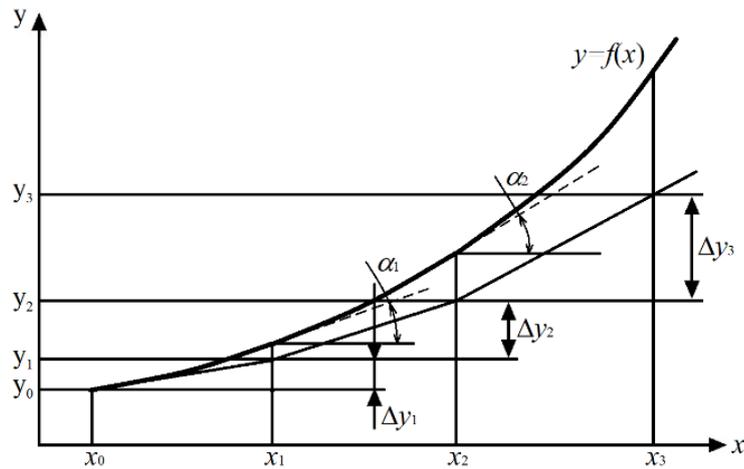


Рис. 36 – Иллюстрация метода Эйлера

Для повышения точности на практике используют модифицированный метод Эйлера второго порядка. Он имеет следующий вычислительный алгоритм:

$$y_{i+1} = y_i + 0.5h[f(x_i, y_i) + f(x_{i+1}, y_{i+1})].$$

Здесь в формуле используется значение $f(x_{i+1}, y_{i+1})$ с еще пока неизвестным значением y_{i+1} . Это значение может быть найдено предварительно, например, по методу Эйлера, а затем использовано в алгоритме. Если же выражение $f(x, y)$ несложное, то можно выразить y_{i+1} из уравнения в явном виде и найти его или решить его относительно y_{i+1} численными методами.

Точность вычислений обычно контролируют двойным просчетом: сначала вычисляют решение уравнения на каком-то текущем шаге h , т.е. находясь в точке x_i и вычисляя значение $y(x_i + h) = y_{i+1}^1$, затем в эту же точку x_{i+1} приходят за два шага по $h/2$, получают y_{i+1}^2 , сравнивают их: если для обоих вариантов различие $|y_{i+1}^1 - y_{i+1}^2|$ в пределах желаемой погрешности, то решение принимают, а если нет, то опять делят шаг на два и т.д., до тех пор, пока не получится приемлемый результат. Однако следует помнить, что при очень маленьком шаге, получающемся в результате его последовательного деления, может значительной оказаться накапливающаяся вычислительная ошибка.

Метод Рунге-Кутты

Существует целая группа методов Рунге – Кутты, среди которых наибольшее распространение получил метод четвертого порядка. Следовательно, он более точен, чем метод Эйлера, который является методом первого порядка. Для расчета одного значения функции необходимо четыре раза вычислять правую часть дифференциального уравнения, а не два, как в модифицированном методе Эйлера второго порядка. Вычислительный алгоритм записывается следующим образом:

$$y_{i+1} = y_i + \frac{k_1 + 2k_2 + 2k_3 + k_4}{6},$$

где

$$\begin{aligned}k_1 &= hf(x_i, y_i); \\k_2 &= hf(x_i + h/2, y_i + k_1/2); \\k_3 &= hf(x_i + h/2, y_i + k_2/2); \\k_4 &= hf(x_i + h, y_i + k_3).\end{aligned}$$

Здесь также для контроля точности можно применять прием двойного просчета.

Метод Милна

Метод Милна относится к многошаговым методам и представляет один из методов прогноза и коррекции. Решение в следующей точке находится в два этапа. На первом этапе осуществляется по специальной формуле прогноз значения функции, а затем на втором этапе – коррекция полученного значения. Если полученное значение y после коррекции существенно отличается от спрогнозированного, то проводят еще один этап коррекции. Если опять имеет место существенное отличие от предыдущего значения (т.е. от предыдущей коррекции), то проводят еще одну коррекцию и т.д. Однако очень часто

ограничиваются одним этапом коррекции.

Метод Милна имеет следующие вычислительные формулы:

– этап прогноза:

$$y_{i+1}^n = y_{i-3} + 4\frac{h}{3}(2f_{i-2} - f_{i-1} + 2f_i),$$

где для компактности записи использовано следующее обозначение $f_i = f(x_i, y_i)$

– этап коррекции:

$$y_{i+1} = y_{i-1} + \frac{h}{3}(f_{i+1} + 4f_i + y_{i+1}^n).$$

Абсолютная погрешность определяется по формуле $\varepsilon \approx |y_{i+1} - y_{i+1}^n| / 29$.

Метод требует несколько меньшего количества вычислений (например, достаточно только два раза вычислить $f(x, y)$, остальные запомнены с предыдущих этапов), но требует дополнительного "расхода" памяти. Кроме этого, как уже указывалось выше, невозможно "запустить" метод: для этого необходимо предварительно получить одношаговыми методами первые три точки.

6.5 Методы решения уравнений и систем

Многие задачи исследования различных объектов с помощью математических моделей, применения их для прогноза или расчета приводят к необходимости решения нелинейных уравнений. Поэтому в данном пособии этому разделу уделено достаточное внимание.

Как правило, процесс решения нелинейного уравнения общего вида: $f(x) = 0$ осуществляется в два этапа. На первом этапе отделяют корни, т.е. находят такие отрезки, внутри которых находится строго один корень. На втором этапе уточняют корень, т.е. находят его значение x^* с предварительно заданной точностью ε . В практических задачах решением называют любое

значение x , отличающееся по модулю от точного значения x^* не более чем на величину ε .

Идеи аналитических методов первого этапа базируются на очевидном свойстве непрерывных функций: корни функции (точки пересечения $f(x)$ с горизонтальной осью) обязательно лежат между соседними экстремумами функции (хотя обратное неверно: между каждой парой экстремумов необязательно находится корень).

Идеи методов второго этапа можно сгруппировать по трем основным направлениям. В первом – поиск корня с заданной погрешностью сводится к перебору всех возможных значений аргумента с проверкой наличия решения. Во втором – поиск корня нелинейной функции заменяется поиском корня той или иной более простой функции (линейной, параболической), близкой к исходной нелинейной; как правило, процесс поиска осуществляется итерационными процедурами (однотипными, последовательно повторяющимися). В третьем – нелинейное уравнение вида: $f(x) = 0$ сводят к одной из форм вида: $g(x) = \varphi(x)$ и стремятся обеспечить равенство левой и правой частей тоже, как правило, с помощью итерационных процедур.

Условием окончания процесса решения уравнения (т.е. получения корня x^* с заданной погрешностью) может быть одно из двух возможных:

$$1) |f(x^*)| \leq \delta,$$

$$2) |x^* - x^k| \leq \varepsilon,$$

где δ , ε – предварительно заданные малые величины, k – номер итерации, т.е. или близость к нулю левой части уравнения, или близость друг к другу двух значений x , между которыми находится решение. Второе условие во многих случаях можно использовать, не зная точного значения корня, путем замены его другим, например: $|x^{k+1} - x^k| \leq \varepsilon'$, при выполнении которого данное условие будет гарантированно выполняться. Условие окончания поиска выбирается исходя из неформальных соображений, и в некоторых случаях применение разных условий может привести к существенно разным результатам. При решении конкретных задач в математическом моделировании важными

являются две цели решения:

1) обеспечение близости к нулю $f(x)$ ($f(x) \approx 0$) как меры выполнения тех или иных балансовых соотношений, тогда не очень важно, при каких именно (в пределах здравого смысла конкретной прикладной задачи) значениях x это равенство справедливо с заданной погрешностью;

2) обеспечение точности нахождения решения x^* , имеющего содержательное значение, при этом $f(x) \approx 0$ является лишь индикатором правильности решения. Отсюда и выбирают условие окончания поиска решения.

Знание особенностей левой части нелинейного уравнения позволяет в ряде случаев, не производя отделения корней, определить число корней (причем отдельно действительных и комплексных), что невозможно в общем случае, а также предельные оценки корней, интервалы существования корней. Это прежде всего касается алгебраических уравнений с действительными коэффициентами (далее для простоты – алгебраических), часто встречающихся в практике. Такие уравнения имеют вид

$$f(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \dots + a_n = 0.$$

Кроме того, с учетом конкретного вида уравнения можно построить более эффективные алгоритмы.

Отделение корней

Отделение корней может производиться графически (путем построения графика функции $f(x)$) или аналитически. Для аналитического отделения корней находят все критические точки функции $f(x)$, т.е. точки, в которых производные равны нулю или не существуют. Это можно сделать численными методами или – в несложных случаях – аналитически. Для этого $f(x)$ дифференцируют, приравнивают производную к нулю и решают полученное уравнение относительно x . Кроме того, определяют все точки, где по тем или

иным причинам (например, знаменатель обращается в нуль, под логарифмом появляется нуль и т.д.) производная может не существовать. В этих (критических) точках или в непосредственной близости от них определяют знак функции $f(x_i)$, т.е. находят $\text{sign } f(x_i)$. Затем строят ряд знаков функции в критических точках, включая в рассмотрение и крайние точки числовой оси $-\infty$ и $+\infty$. Анализируют этот ряд, и по числу смен знаков определяют количество корней (равно числу смен знаков $\text{sign } f(x_i)$) и интервалы, где локализованы эти корни. На левой и на правой границах такого интервала функция $f(x)$ должна иметь разные знаки. В случае необходимости можно дополнительно к критическим точкам использовать и произвольные точки, что позволяет сузить интервал локализации корня. Особенно это надо делать, когда одна из границ интервала находится в бесконечности, так как интервал хотя бы с одной границей в бесконечности не позволит уточнить корни.

Уточнение корней

Рассмотрим методы уточнения корней и раскрывающие их основные идеи. Отметим очевидный момент: при прочих равных условиях тот метод уточнения корней будет более эффективен, в котором результат с той же погрешностью найден за меньшее число раз вычисления функции $f(x)$.

Метод сканирования

Метод предусматривает разделение всего интервала $[a, b]$, где отделен корень, на маленькие отрезки, равные заданной погрешности ε , с последующим вычислением (или определением экспериментально) значений функции $f(x)$ на концах этих отрезков (т.е. в точках, расстояние между которыми не превышает величины ε). Анализируя значения функции, нетрудно выбрать отрезок, где функция меняет знак (или точно равна нулю, что маловероятно). В качестве решения можно взять любую точку – левую (x_i) или правую (x_{i+1}) границу выделенного отрезка, хотя предпочтительнее взять середину этого отрезка $x^* = (x_i + x_{i+1})/2$. В любом случае погрешность решения

не будет превышать заданную погрешность ε , даже при условии, что мы не знаем точного значения решения.

Иногда весь отрезок разбивают на маленькие отрезки величиной 2ε , а затем искомое значение корня берут в середине отрезка, где функция меняет знак. Это непринципиальная разница с основным вариантом, результаты вариантов полностью совпадут и по значению корня, и по затратам на поиск, если в первом сразу взять погрешность вдвое больше необходимой.

Для повышения эффективности метода можно уточнение производить в несколько этапов. На первом этапе задать большое значение ε , найти отрезок, где функция меняет знак (грубо найти корень), затем найденный отрезок еще раз разделить с более мелким шагом, более точно найти корень и т.д. еще несколько этапов (обычно 3...5), после чего удастся найти корень с заданной погрешностью в целом за меньшее число раз вычисления $f(x)$. Метод очевиден и не требует практического пояснения.

Метод деления отрезка пополам

Этот метод можно рассматривать как развитие метода сканирования: величина отрезков, на которые делится весь интервал при многоэтапном применении метода сканирования, становится равной половине исходного отрезка $[a, b]$. В этом случае сначала исходный отрезок делится на две равные части (пополам) и путем сравнения знаков функции на концах каждой из двух половинок (например, по знаку произведения значений функций на концах) определяют ту половинку, в которой содержится решение (знаки функции на концах должны быть разные). Затем найденную половинку опять делят на две равные части, снова выбирают одну из двух половинок, содержащую корень, и т.д. Условием окончания служит заданная малость отрезка, где содержится корень, т.е. аналогично методу сканирования. Существуют и более эффективные алгоритмы, например, выбор точки не в середине отрезка, а ближе к тому краю, в котором значение функции меньше. Но этот вариант не будет работать при немонотонной функции.

Метод половинного деления, как и метод сканирования, очевиден и не требует практического пояснения.

Метод хорд

В этом методе нелинейная функция $f(x)$ на отделенном интервале $[a, b]$ заменяется линейной, в качестве которой берется хорда – прямая, стягивающая концы нелинейной функции. Эта хорда определяется как прямая, проходящая через точки с координатами $(a, f(a))$ и $(b, f(b))$. Имея уравнение хорды $y = cx + d$, можно легко найти точку ее пересечения с горизонтальной осью, подставив в уравнение $y = 0$ и найдя из него x . Естественно, в полученной таким путем точке x_1 , не будет решения, ее принимают за новую границу отрезка, где содержится корень. Через эту точку с координатами $(x_1, f(x_1))$ и соответствующую границу предыдущего интервала опять проводят хорду, находят x_2 и т.д. несколько раз, получая последовательность x_3, x_4, x_5, \dots , сходящуюся к корню. Метод применим только для монотонных функций.

Алгоритм метода зависит от свойств функции $f(x)$. Если $f(b)f'(b) > 0$, то строящаяся на каждом этапе хорда имеет правый фиксированный ("закрепленный") конец, и алгоритм выглядит следующим образом:

$$x_{i+1} = x_i - \frac{f(x_i)}{f(b) - f(x_i)}(b - x_i);$$

при этом последовательность x_1, x_2, \dots будет приближаться к корню слева.

Если $f(a)f'(a) > 0$, то строящаяся на каждом этапе хорда имеет левый фиксированный ("закрепленный") конец, и алгоритм выглядит следующим образом:

$$x_{i+1} = a - \frac{f(a)}{f(a) - f(x_i)}(x_i - a);$$

при этом последовательность x_1, x_2, \dots будет приближаться к корню справа.

На рис. 37 приведен один из вариантов применения метода хорд. В рассматриваемом случае "закрепленным" является правый конец. Приведено пять шагов (пять хорд), при этом к решению приближаемся слева.

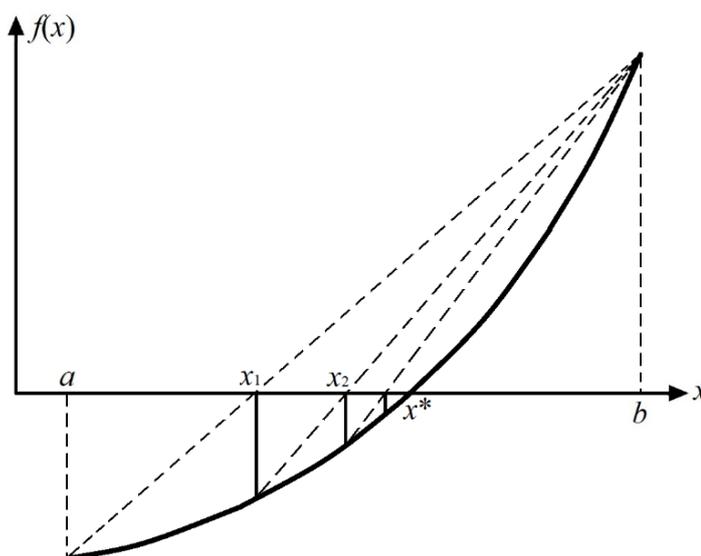


Рис. 37 – Иллюстрация метода хорд

Теоретически доказано, что если первые производные на концах интервала при монотонной и выпуклой функции $f(x)$ не различаются более чем в 2 раза, то справедливо соотношение $|x^* - x_i| < |x_i - x_{i-1}|$ и условием прекращения пополнения последовательности может быть $|x_{i+1} - x_i| \leq \varepsilon$, а в качестве корня принято x_{i+1} (можно также окончить процесс и при достижении $f(x_i) < \delta$, о чем указывалось в концепции методов). На практике указанные условия можно применять и без предварительной проверки производных, отклонение погрешности результата при пологих функциях не будет существенным.

Метод Ньютона (касательных)

Идея, на которой основан метод, аналогична той, которая реализована в методе хорд, только в качестве прямой берется касательная, проводимая в текущей точке последовательности. Уравнение касательной находится по координате одной точки и углу наклона (значение производной). В качестве

начальной точки в зависимости от свойств функции берется или левая точка $x_0 = a$ ($f(a)f''(a) > 0$), или правая точка $x_0 = b$ (если $f(b)f''(b) > 0$). Алгоритм записывается следующим образом:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}.$$

Алгоритм работоспособен при выпуклых и монотонных функциях $f(x)$. Главным теоретическим достоинством метода является квадратичная скорость сходимости, что во многих случаях может привести к сокращению числа вычислений функции при получении решения с заданной погрешностью. В ряде случаев можно применять упрощенный алгоритм, связанный с сокращением числа раз вычисления производных – вместо вычисления производной в каждой очередной точке $f'(x_i)$ использовать значение производной в начальной точке $f'(x_0)$. Следует обратить внимание на следующую особенность метода: последовательность x_1, x_2, x_3, \dots приближается к корню с другой стороны, чем при использовании метода хорд при прочих равных условиях.

На рис. 38 приведен один из вариантов применения метода Ньютона. В рассматриваемом случае процесс начинается с правого конца. К решению приближаемся справа. Условия окончания поиска аналогичны методу хорд.

Комбинированный метод

Данный метод, так же как и предыдущие, базируется на замене нелинейной функции $f(x)$ линейной, но с учетом стремления к корню метода хорд и метода Ньютона с разных сторон; для повышения эффективности использует оба алгоритма одновременно. Один шаг делается методом хорд, а следующий с другой стороны – методом Ньютона. При этом интервал, где содержится корень, сокращается с обеих сторон, что обуславливает другое условие окончания поиска. Поиск можно прекратить, как только разница

между правым и левым концами интервала станет меньше предварительно заданной погрешности ε .

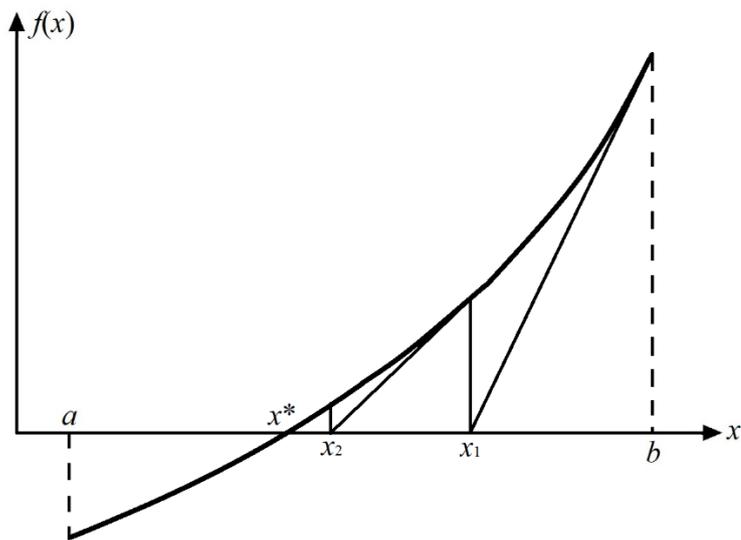


Рис. 38 – Иллюстрация метода Ньютона

Алгоритмы используемых методов следует выбирать с учетом упомянутых выше особенностей функций.

Метод параболической аппроксимации

В этом методе функция $f(x)$ заменяется не линейной, а параболической функцией, что является более точной заменой. Следовательно, метод может обеспечить более быструю сходимость к решению. На первом этапе параболу обычно строят по трем точкам: крайним и средней точкам интервала (a, b) , где отделен корень, т.е. $(a, f(a))$, $((a + b)/2, f((a + b)/2))$, $(b, f(b))$. По полученному уравнению параболы $y = c_2x^2 + c_1x + c_0$ находят приближенный корень (приближенный – потому что парабола приближенно заменяет $f(x)$), для чего решают уравнение $c_2x^2 + c_1x + c_0 = 0$. На втором этапе строят параболу по трем точкам: найденному приближенному корню и двум предыдущим точкам (слева и справа от этой точки), лежащим по разные стороны оси x . Такой вариант выбора точек на практике быстрее приводит к решению по сравнению с вариантом, когда для построения параболы берутся последовательно три

последние точки. Такая процедура повторяется многократно до тех пор, пока величина отрезка, внутри которого находится корень, не будет меньше ε – предварительно заданной погрешности.

На рис. 39 приведен один шаг уточнения корня методом параболической аппроксимации. Начальная парабола проведена через точки a , x_1 , b (здесь x_1 является серединой отрезка $[a, b]$); x_2 – пересечение параболы с осью. Следующая парабола должна проводиться через x_1 , x_2 , b .

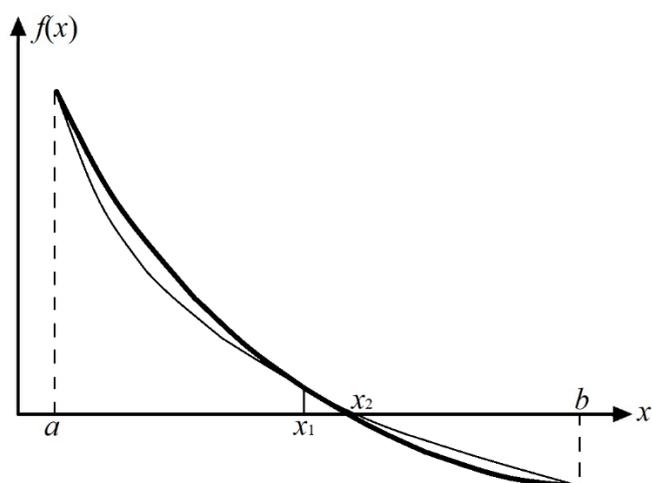


Рис. 39 – Иллюстрация первого этапа метода параболической аппроксимации

Метод простой итерации

Рассматриваемый метод реализует третий подход из представленных в концепции. Предварительно исходное уравнение $f(x) = 0$ преобразуют к виду $\varphi(x) = x$, что является частным случаем более общей структуры $g(x) = f(x)$. Затем выбирают начальное значение x_0 и подставляют его в левую часть уравнения, но $\varphi(x) \neq x$, поскольку x_0 взято произвольно и не является корнем уравнения. Полученное $\varphi(x_0) = x_1$ рассматривают как очередное приближение к корню. Его снова подставляют в левую часть уравнения $\varphi(x_1)$ и получают следующее значение x_2 ($x_2 = \varphi(x_1)$) и т.д., в общем случае $x_{i+1} = \varphi(x_i)$. Получающаяся таким образом последовательность $x_0, x_1, x_2, x_3, x_4, \dots$ при определенных условиях может сходиться к корню x^* (рис. 40).

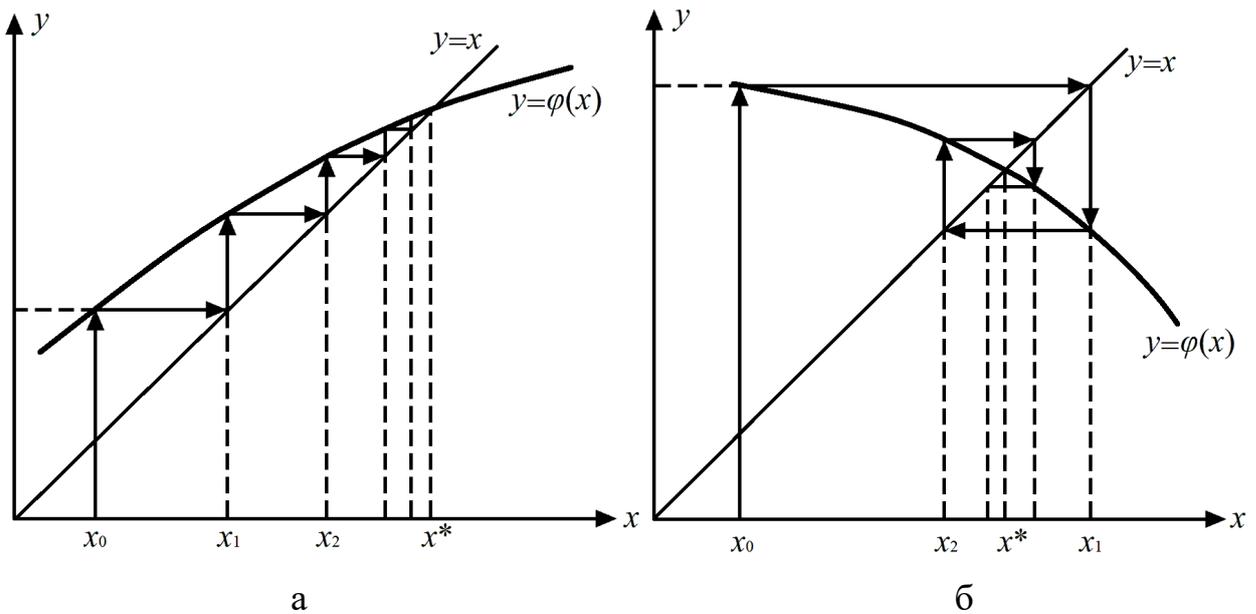


Рис. 40 – Иллюстрация метода итераций для различных ситуаций

а – $0 < \varphi'(x) < 1$; б – $-1 < \varphi'(x) < 0$;

Таким условием является $|\varphi'(x)| \leq 1$ на $[a, b]$, причем чем ближе модуль к нулю, тем выше окажется скорость сходимости к решению. В противном случае последовательность расходится от искомого решения ("метод не сходится"). На рис. 41 приведен один из возможных случаев, когда итерационный процесс не сходится. Видно, что последовательность x_0, x_1, x_2, \dots удаляется от корня x^* . Это всегда будет иметь место в том случае, если тангенс угла наклона $\varphi(x)$ в окрестности корня по модулю больше единицы.

Существуют различные способы преобразования уравнения $f(x) = 0$ к виду $\varphi(x) = x$; одни могут привести к выполнению условия сходимости всегда, другие – в отдельных случаях. Самый простой способ следующий:

$$f(x) + x = 0 + x, \quad f(x) + x = \varphi(x) \Rightarrow \varphi(x) = x,$$

но он не всегда приводит к успеху. Существует другой способ, в соответствии с которым $\varphi(x) = x - f(x)/k$, причем k следует выбирать так, чтобы $|k| > Q/2$, где $Q = \max |f'(x)|$ и знак k совпадал бы со знаком $f'(x)$ на $[a, b]$.

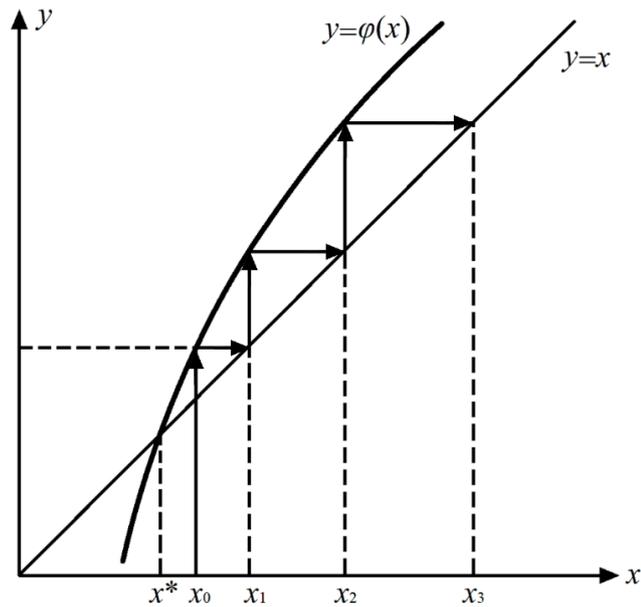


Рис. 41 – Иллюстрация не сходящегося итерационного процесса

Погрешность решения можно оценить из соотношения

$$|x^* - x_i| \leq \frac{q}{1 - q} |x_i - x_{i+1}|, \quad \text{где } q = \max_{[a,b]} \varphi'(x).$$

Вследствие этого для окончания вычислений в методе итераций применяют соотношение $q/(1 - q) |x_i - x_{i+1}| \leq \varepsilon$, где ε – заданная погрешность решения.

Часто используют упрощенное условие окончания поиска $|x_i - x_{i+1}| \leq \varepsilon$, не вычисляя максимальное значение производной, но в этом случае погрешность решения может не соответствовать заданной (т.е. быть больше или меньше).

Определение числа корней алгебраических уравнений

Для уравнения $P_n(x) = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_n = 0$ общее количество действительных корней и комплексных корней (которые всегда в рассматриваемом случае при действительных коэффициентах уравнения будут попарно сопряженными) всегда равно наивысшей степени полинома в левой части уравнения, т.е. в нашем случае – n .

Существует правило Декарта, в соответствии с которым можно оценить

отдельно количество положительных и отрицательных действительных корней: количество действительных положительных корней либо равно числу перемен знака в последовательности коэффициентов уравнения, либо на четное число меньше; количество действительных отрицательных корней либо равно числу перемен знака в последовательности коэффициентов уравнения $P(-x)=0$, либо на четное число меньше (равные нулю коэффициенты в обоих случаях не учитываются). Отсюда следует, что правило Декарта не точно определяет количество корней, а приближенно. Существуют и более строгие методы (например, правило Штурма), которые точно определяют количество действительных корней на любом промежутке числовой оси (а не только положительных и отрицательных).

Предельные оценки и область существования корней алгебраических уравнений

Во многих случаях достаточно знать предельное (максимально возможное) значение корня. Это необходимо в различных приложениях (например, при оценке устойчивости систем управления), а также для более эффективного отделения корней нелинейного уравнения. Эту проблему можно решить, не находя в корни, а воспользовавшись специальными методами, которые кратко рассмотрены ниже.

Метод Лагранжа

Метод сводится к определению верхней границы положительных корней по формуле

$$R = 1 + \sqrt[m]{B/a_0},$$

где m – номер первого по порядку отрицательного коэффициента в полиноме левой части уравнения $P_n(x)$, B – наибольшая и абсолютных величин отрицательных коэффициентов $P_n(x)$; при этом предполагается, что $a_0 > 0$.

Метод Ньютона

Этот метод не дает конструктивного пути для отыскания искомого значения, т.е. не дает более эффективных алгоритмов помимо простого перебора или проверки наугад взятых точек, и формулируется следующим образом: если при каком-то значении $x = \beta$ $P_n(x) > 0$ и все производные $P'(x) > 0$, $P''(x) > 0$, ..., $P^{(n)}(x) > 0$, то значение β является предельным значением положительных корней.

Метод кольца

Метод позволяет находить область существования всех корней алгебраического уравнения, в том числе и комплексных. Действительные корни находятся в интервалах $(-R, r)$ и (r, R) (соответственно отрицательные и положительные). Величины r и R вычисляются по формулам

$$R = 1 + \frac{A}{|a_0|}; \quad r = \frac{1}{1+B/|a_n|},$$

где $A = \max \{|a_1|, |a_2|, \dots, |a_n|\}$; $B = \max \{|a_0|, |a_1|, \dots, |a_{n-1}|\}$.

В случае нахождения области существования всех корней, а не только действительных, строится кольцо с радиусами r и R , внутри которого находятся и действительные корни. Метод дает диапазон существования корней приближенно с определенным запасом.

Метод предельных значений

Существует метод, позволяющий несколько точнее находить область существования корней. Для этого вычисляют верхние границы положительных корней R_i :

для полинома $P_n(x) - R_1$;

для полинома $P_n(-x) - R_2$;

для полинома $x^n P_n(x) - R_3$;

для полинома $x^n P_n(1/x) - R_4$.

Если действительные корни существуют, то они лежат в интервалах $(-R_2, -1/R_4)$ и $(1/R_3, R_1)$.

Уточнение корней алгебраических уравнений

Алгебраические уравнения можно, конечно, решать любым подходящим методом из ранее приведенных, но конкретная и заранее известная структура левой части уравнения позволяет строить более эффективные алгоритмы. Идея уточнения корней алгебраического уравнения методом понижения порядка базируется на возможности выделения в полиноме $P_n(x)$ множителя, содержащего корень. Для действительного корня таким множителем является $x - x_1$, где x_1 – корень, а для пары комплексных корней $\alpha \pm j\omega$ – множитель $x^2 + bx + c$, где $b = -2\alpha$, $c = \alpha^2 + \omega^2$. После такого выделения можно снизить порядок исходного полинома на единицу (или на два) путем деления $P_{n-1} = P_n / (x - x_1)$ (для действительного корня) или $P_{n-2} = P_n / (x^2 + bx + c)$ (для комплексной пары корней), далее для нового полинома опять выделить следующий корень x_2 (или новую пару корней) и т.д. В рассматриваемом алгоритме можно выделить три этапа:

- 1) нахождение приближенного значения корня;
- 2) проверка, является ли найденное значение корнем;
- 3) получение нового полинома степени на единицу меньше, чем предыдущая.

Уточнение действительного корня

При выделении действительного корня процесс его получения носит итерационный характер и реализуется по формуле

$$x_{i+1} = -x_i \frac{a_n}{a_0 x_i^n + a_1 x_i^{n-1} + \dots + a_{n-1} x_i} = \frac{a_n}{a_0 x_i^{n-1} + a_1 x_i^{n-2} + \dots + a_{n-1}}$$

Здесь через x_i и x_{i+1} обозначены соответственно i -е и $(i+1)$ -е приближения

выделяемого корня. Если начальное приближение достаточно близко к искомому корню, то итерационный процесс сходится очень быстро. К сожалению, сходимость метода зависит от условия сходимости

$$\left| 1 + x_1 \frac{P'(x_1)}{P(0)} \right| < 1,$$

которое практически невозможно оценить заранее (x_1 – искомый корень). Кроме того, метод может не сходиться к решению и вследствие отсутствия действительного корня.

По существу, рассматриваемый метод является методом итераций для одного из способов приведения к необходимому виду

$$x = \varphi(x) = x + \frac{xP(x)}{P(0) - P(x)}.$$

Значение приближения x_i на каждой итерации нужно проверять, не является ли оно корнем. Для этого достаточно найти остаток r от деления $P_n(x)$ на $x - x_i$. Если остаток равен нулю (или очень близок к нулю), то текущее значение x_i является корнем, если нет, то итерационный процесс следует повторять. Воспользовавшись теоремой Безу, можно находить остаток, не производя собственно операцию деления. В соответствии с этой теоремой остаток равен значению полинома при $x = x_i$, т.е. $r = P_n(x_i)$. Такой подход существенно сокращает вычисления.

После итерационного выделения одного корня (что контролируется по остатку r) следует понизить порядок полинома, т.е. найти новый полином P_{n-1} , содержащий остальные корни:

$$P_{n-1}(x) = \frac{P_n(x)}{x - x_1}.$$

Коэффициенты b_k нового полинома P_{n-1} находятся по схеме Горнера (не проводя операцию деления), в соответствии с которой

$$b_0 = a_0; \quad b_1 = b_0x_1 + a_1; \quad b_2 = b_1x_1 + a_2; \quad b_3 = b_2x_1 + a_3; \quad \dots$$

$$b_k = b_{k-1}x_1 + a_k; \quad \dots \quad b_{n-1} = b_{n-2}x_1 + a_{n-1};$$

где a_k – коэффициенты полинома $P_n(x)$.

Уточнение комплексной пары корней (метод Хичкока)

Комплексные корни алгебраических уравнений с действительными коэффициентами всегда сопряженные, поэтому для их нахождения достаточно выделить квадратичный множитель вида $x^2 + px + q$, затем уже найти его комплексные корни. В нахождении такого множителя и заключается поиск комплексных корней многими методами, в том числе и рассматриваемым.

Если разделим $P_n(x)$ на $x^2 + px + q$, то получим тождество

$$P_n(x) = (x^2 + px + q)L_{n-2}(x) + xP(p, q) + Q(p, q),$$

где $L_{n-2}(x)$ – частное от деления $P_n(x)$ на квадратичный трехчлен, а $P(p, q)$ и $Q(p, q)$ – многочлены от p и q . Для того чтобы $x^2 + px + q$ был делителем $P_n(x)$, необходимо и достаточно, чтобы $P(p, q) = 0$ и $Q(p, q) = 0$. Если решить систему уравнений

$$P(p, q) = 0, \quad Q(p, q) = 0,$$

то мы получим необходимые значения p и q .

Процесс нахождения p и q носит итерационный характер:

$$p^{i+1} = p^i + h^i, \quad \text{где } h^i = \frac{\Delta p^i}{\Delta^i}, \quad q^{i+1} = q^i + t^i, \quad \text{где } t^i = \frac{\Delta q^i}{\Delta^i},$$

а Δp^i , Δq^i , Δ^i в свою очередь вычисляются через определители так:

$$\Delta p^i = \begin{vmatrix} P'_q(p^i, q^j) & P(p^i, q^j) \\ Q'_q(p^i, q^j) & Q(p^i, q^j) \end{vmatrix}'$$

$$\Delta q^i = \begin{vmatrix} P(p^i, q^j) & P'_p(p^i, q^j) \\ Q(p^i, q^j) & Q'_p(p^i, q^j) \end{vmatrix}'$$

$$\Delta^i = \begin{vmatrix} P'_p(p^i, q^j) & P'_q(p^i, q^j) \\ Q'_p(p^i, q^j) & Q'_q(p^i, q^j) \end{vmatrix}'$$

Значения производных $P'_p(p^i, q^i)$ (для компактности использованы стандартные обозначения $P'_p(p^i, q^i) = dP/dp$, $P'_q(p^i, q^i)$, $Q'_p(p^i, q^i)$, $Q'_q(p^i, q^i)$) находятся по формулам:

$$P'_p(p^i, q^i) = p^i R^i - S^i;$$

$$P'_q(p^i, q^i) = -R^i;$$

$$Q'_p(p^i, q^i) = q^i R^i;$$

$$Q'_q(p^i, q^i) = -S^i.$$

Величины $P(p^i, q^i)$ и $Q(p^i, q^i)$ вычисляются из разложения

$$P_n(x) = (x^2 + p^i x + q^i) L_{n-2}(x) + xP(p^i, q^i) + S(p^i, q^i).$$

а R_i и S_i вычисляются из разложения

$$L_{n-2}(x) = (x^2 + p^i x + q^i) G_{n-4}(x) + xR(p^i, q^i) + S(p^i, q^i).$$

Метод требует сравнительно много вычислений и не всегда может сходиться. Корнями выделенного квадратичного трехчлена на могут оказаться и действительные корни.

После выделения квадратичного сомножителя находят полином более низкой степени $L(x)$, и к нему применяют аналогичный алгоритм для выделения следующей пары комплексных корней.

6.6. Решение систем нелинейных уравнений

Решение системы нелинейных уравнений $f(x)=0$, где x – векторная величина, а $f(x)$ – векторная функция, значительно сложнее, чем решение одного уравнения. Очень сложно отделить корни, поэтому на этом этапе обычно выбирают начальное приближение ближе к потенциально возможному решению. Для получения всех возможных решений чаще всего на практике перебирают различные начальные условия поиска. В основном используют два метода решения систем: метод Ньютона – Рафсона и метод итераций. Реже используются поисковые методы оптимизации, которыми ищут минимум суммы квадратов невязок $f(x)=\varepsilon$ всех уравнений системы, подбирая переменные x_i :

$$R = \sum_i \varepsilon_i^2 \rightarrow \min_{x_i},$$

который будет иметь место только при всех $\varepsilon_i=0$, т.е. соответствующие значения x_i будут являться решением системы.

Метод Ньютона – Рафсона

Метод базируется на разложении функций $f_j(x_1, x_2, \dots, x_n)$ в ряд Тейлора с отбрасыванием всех нелинейных членов разложения (т.е. функции линеаризуют). Затем ставится задача перехода по всем переменным из текущей точки x^i в следующую x^{i+1} , которую считают решением, т.е. находят $h^i = x^{i+1} - x^i$, поэтому полагают $f(x^{i+1})=0$. Получается система линейных уравнений относительно h_j^i ($j = 1, 2, \dots, n$):

$$\frac{df_k(x^i)}{dx_1} h_1^i + \frac{df_k(x^i)}{dx_2} h_2^i + \dots + \frac{df_k(x^i)}{dx_n} h_n^i = -f_k(x^i), \quad k = 1, 2, \dots, n$$

(здесь и далее через d обозначены частные производные), которую решают. С найденными шагами h_j переходят в новую точку по всем переменным $x_j^{i+1} = x_j^i + h_j^i$ ($j = 1, 2, \dots, n$). Однако вследствие линейного приближения новая точка x^{i+1} не является решением, т.е. $f(x^{i+1}) \neq 0$, ее просто считают следующим приближением и повторяют всю процедуру многократно, делая процесс поиска корней итерационным. Сходимость метода довольно высокая особенно если удачно выбрано начальное приближение.

Окончание поиска решения осуществляется по условию малости шагов по всем переменным одновременно, например при выполнении условия $|h_j^i| \leq \varepsilon$ для $j = 1, 2, \dots, n$, где ε – заданная погрешность. Так же как и при решении одного уравнения, можно воспользоваться для окончания поиска условием малости отклонений функций $f(x_1, \dots, x_n)$ от нуля, например условием $\max\{f_j(x^i)\} \leq \varepsilon$, или потребовать выполнения обоих условий сразу.

Недостатком метода является необходимость на каждом шаге вычислять частные производные по всем переменным от всех функций и решать систему линейных уравнений. Существует модификация, не требующая на каждом шаге вычисления производных (используются производные, полученные на первом шаге) но сходимость такого метода существенно ниже.

Метод итераций

В этом методе, как и при решении одного уравнения, предварительно все уравнения приводят к специальному виду $x = p(x)$:

$$x_1 = f_1(x_1, x_2, \dots, x_n),$$

$$x_2 = f_2(x_1, x_2, \dots, x_n),$$

...

$$x_n = f_n(x_1, x_2, \dots, x_n).$$

Сущность метода аналогична соответствующему методу при решении одного уравнения. Берется произвольное начальное значение $x^0 (x_1^0, x_2^0, \dots, x_n^0)$ и подставляется в уравнение. Из каждого уравнения системы находятся новые значения $(x_1^1, x_2^1, \dots, x_n^1)$ т.е. x^1 , которое затем опять подставляется в уравнение и находится x^2 , затем x^3 и т.д. При выполнении условий

$$\sum_{i=1}^n M_{ki} < 1 \quad \text{или} \quad \sum_{k=1}^n M_{ki} < 1, \quad \text{где} \quad M_{ki} = \max \left| \frac{d\varphi_k}{dx_i} \right|,$$

последовательность $x_0, x_1, x_2, x_3, \dots$ сходится к решению. Для обеспечения сходимости метода можно использовать следующий способ преобразования исходной системы к виду, удобному для итераций:

$$x_{i+1} = x_i + \sum_j \alpha_{ij} f_j(x) = \varphi_j(x),$$

где α_{ij} находятся из решения вспомогательной системы уравнений

$$\frac{d\varphi_i(x)}{dx_j} = 0, \quad i, j = 1, 2, 3, \dots, n.$$

Недостатком такого подхода является необходимость большого объема вычислений, так как преобразования следует делать на каждом шаге итераций. Несколько меньшая сходимость может быть получена, если преобразование от $f(x) = 0$ к $x = \varphi(x)$ делать через несколько шагов или один раз.

7. Ошибки измерений

Опыт показывает, что вследствие неточности измерительных приборов, несовершенства наших органов чувств, неполноты наших знаний, трудности учета всех побочных явлений, при многократном повторении одного и того же измерения получаются разные числовые значения изучаемой физической величины. Так бывает, даже если измерения производить в совершенно одинаковых условиях (равноточные измерения). При практическом использовании результатов тех или иных измерений возникает вопрос об истинном значении изучаемой физической величины, о точности измерения.

Термин «точность измерения», т.е. степень приближения результатов измерения к некоторому действительному значению, используется для качественного сравнения измерительных операций. Для количественной оценки используется понятие «погрешность (ошибка) измерений». Эти термины тесно связаны друг с другом: чем меньше погрешность, тем выше точность. Оценка погрешности измерений – одно из важных мероприятий по обеспечению достоверности измерений.

Количество факторов, влияющих на точность измерений, достаточно велико, и любая классификация погрешностей измерений в известной мере условна. На схеме, изображенной на рис. 42, приведена одна из возможных классификаций, которая может служить основой для оценки погрешности измерений. Рассмотрим некоторые виды погрешностей подробнее.

Обозначим через x результат измерения некоторой величины, а через x_0 – истинное значение ее, которое всегда нам неизвестно.

Погрешность измерения – это отклонение результата измерений x от истинного x_0 (действительного) значения измеряемой величины.

В зависимости от формы представления различают абсолютную, относительную и приведенную погрешности измерений.

Абсолютная погрешность измерения определяется как разность $x_0 - x = \Delta x$ между истинным и измеренным значениями физической величины. Абсолютная погрешность может быть положительной или отрицательной в

зависимости от того уменьшен или увеличен результат измерения по отношению к истинному значению.

Относительная погрешность – отношение абсолютной погрешности к истинному значению или к результату измерения. Относительная погрешность чаще всего выражается в процентах.

$$\delta = \pm \frac{\Delta x}{x} 100\% \text{ или } \delta = \pm \frac{\Delta x}{x_0} 100\%.$$

Приведенная погрешность – отношение абсолютной погрешности к нормированному значению x_n , выраженное в процентах $\gamma = \pm \Delta x/x_n 100\%$. В качестве нормированного значения может быть взято, например, максимальное значение x_{\max} измеряемой величины $x_n = x_{\max}$.

Точность измерения – это степень приближения измерения к действительному значению величины.

Достоверность измерения показывает степень доверия к результатам измерения, т.е. вероятность отклонений измерения от действительных значений.

Чтобы повысить точность и достоверность измерений, необходимо уменьшить погрешности.

Погрешности при измерениях возникают вследствие ряда причин:

- несовершенство методов и средств измерений;
- недостаточная тщательность проведения опыта;
- влияние различных внешних факторов в процессе эксперимента;
- субъективные особенности экспериментатора.

В зависимости от характера проявления, причин возникновения и возможностей устранения различают систематическую и случайную составляющую погрешности измерения, а также грубые погрешности (промахи).



Рисунок 42 – Классификация возможных погрешностей

Систематические погрешности (ошибки) – это погрешности, которые сохраняют величину и знак от опыта к опыту, при равноточных измерениях.

Типичными источниками систематических погрешностей бывают:

- несовершенство используемой измерительной аппаратуры;
- несовершенство используемого метода измерений;
- плохая настройка измерительной аппаратуры;
- недостаточное постоянство условий опыта;
- влияние окружающей среды;
- постоянные ошибки экспериментатора;
- неучтенные влияния других параметров.

Систематические погрешности считаются потенциально устранимыми.

Чтобы избежать или уменьшить систематические погрешности необходимо критически относиться к методам исследования, совершенствуя их, применяя более точные приборы, следя за их исправностью и т.д.

Случайные погрешности (ошибки) – это погрешности, изменяющие свою величину или знак от опыта к опыту, при измерениях, выполненных одинаковым образом и при одинаковых условиях. Случайные погрешности обуславливаются большим числом случайных причин, действующих в каждом отдельном измерении различным, неизвестным образом. К числу таких причин относятся случайные вибрации отдельных частей прибора, различные изменения в среде (температурные, оптические, электрические, магнитные воздействия, изменение влажности, колебание воздуха), трение, физиологическое изменение органов чувств экспериментатора (например, утомление) и множество других причин, которые практически невозможно исключить. Предсказать величину случайной погрешности для одного измерения в принципе невозможно. Поэтому приходится повторять измерения до определенного разумного предела, а полученную совокупность экспериментальных результатов обрабатывать с помощью методов теории вероятностей и математической статистики, которые являются основой, так называемой, теории погрешностей.

Промахи или грубые погрешности (ошибки) – это ошибочные измерения или наблюдения, возникающие в результате небрежности при отсчете по прибору или неразборчивой записи показаний, при неправильном включении прибора, или при нарушении условий, в которых должен проводиться опыт (изменение напряжения, загрязнение материала и т.д.). Такие ошибочные данные следует отбросить или сделать повторные (контрольные) измерения.

Если влияния систематических погрешностей и грубых промахов на полученные в эксперименте результаты, так или иначе, можно избежать или уменьшить, то случайные погрешности являются неустранимыми.

Приложение

При проведении компьютерного моделирования часто возникает потребность в случайных числах. Обычно для производства случайных чисел используют некоторый детерминированный алгоритм, в результате работы которого получается последовательность чисел со свойствами, близкими к случайному распределению. В отличие от истинно случайных чисел, случайные числа такого рода называют *псевдослучайными*. Последовательность псевдослучайных чисел является периодической, т. е. в ней можно выделить регулярно повторяющиеся фрагменты. Чаще всего генерируют последовательность равномерно распределенных чисел. Такая последовательность может быть преобразована в другое распределение, например, в нормальное [32].

К псевдослучайным числам предъявляют ряд требований.

1. Псевдослучайные числа должны быть равномерно распределены на некотором интервале.
2. Между ними не должно обнаруживаться никаких корреляций.
3. Период последовательности должен быть большой.
4. Последовательность должна быть воспроизводимой.

Хотя в большинстве систем имеются встроенные генераторы случайных чисел, далеко не всегда они удовлетворяют перечисленным требованиям. В первую очередь это касается периода последовательности и отсутствия корреляций. Например, компьютерное моделирование движения броуновской частицы с использованием встроенных генераторов иногда приводит к странным результатам. Может оказаться, что частица совершает периодическое движение или никогда не попадает в некоторые области пространства. В связи с этим в серьезных работах по компьютерному моделированию используют более надежные генераторы.

Существует целый ряд различных алгоритмов для генерации псевдослучайных чисел. С некоторыми из них мы сейчас познакомимся.

Линейный конгруэнтный генератор

Линейный конгруэнтный генератор используется во многих системах и порождает последовательность целых чисел, равномерно распределенных на интервале от 0 до $m - 1$, используя следующее рекуррентное соотношение

$$R_{i+1} = a R_i + c \pmod{m}$$

где модуль m , множитель a и инкремент c – положительные целые числа. Период последовательности, порождаемой приведенным соотношением не может превышать m . При правильном выборе m , a и c в последовательности присутствуют все числа от 0 до $m - 1$ при любом выборе начального значения последовательности R_0 .

Достоинствами генератора является то, что он работает очень быстро, поскольку требует выполнения только нескольких операций. Один недостаток этого генератора заключается в том, что период последовательности определяется количеством байтов, отводимых для целых чисел в данной системе. Обычно для хранения целых чисел используется 2 байта, поэтому период последовательности составит 32767, что чрезвычайно мало для решения реальных задач. Другой недостаток состоит в том, что сгенерированные числа коррелированы между собой. Если рассматривать k последовательных случайных чисел как координаты точек в k -мерном пространстве, то все такие точки будут лежать на некоторых плоскостях, а не заполнят равномерно весь объем.

Мультипликативный конгруэнтный алгоритм

Этот алгоритм был предложен Льюисом, Гудменом, Миллером в 1969 г. Последовательность случайных чисел задается следующим рекуррентным соотношением

$$R_{i+1} = a R_i \pmod{m}$$

где $a = 7^5 = 16807$, $m = 2^{31} - 1 = 2147483647$. Поскольку на 32-разрядных компьютерах вычисления по вышеприведенной формуле не могут быть выполнены непосредственно, то для вычислений используется следующая формула

$$az \pmod{m} = \begin{cases} az \pmod{q} - r[z/q], & \text{если выражение} > 0, \\ az \pmod{q} - r[z/q] + m, & \text{в остальных случаях,} \end{cases}$$

где $q = 127773$, $r = 2836$, а z – положительное целое число. Квадратные скобки означают целую часть числа. Генератор имеет период $2^{31} - 2 \approx 10^9$.

Ниже приводится пример реализации алгоритма на языке C.

```
#define IA 16807
#define IM 2147483647
#define AM (1.0/IM)
#define IQ 127773
#define IR 2836
#define MASK 123459876
float mcgrand(long*idum)
{
long k;
float ans;
*idum ^=MASK;
k=(*idum)/IQ;
*idum=IA*( *idum-k*IQ) -IR*k;
if (*idum ! 0) *idum += IM;
ans=AM*( *idum);
*idum ^=MASK;
return ans;
```

}

Генератор на основе сдвига регистра

Этот генератор использует следующую идею. Поскольку любое число в компьютере представлено в виде последовательности битов, то получить новое число можно изменив каким-либо образом биты имеющегося числа. В простейшем случае можно просто сдвинуть все биты влево. Тогда старший бит пропадет, а освободившийся младший бит можно заполнить нулем. Повторяя эту операцию можно получить 2^n чисел, где n – число битов, отводимых для хранения числа. Такая операция называется *сдвиг Бернулли* и задается правилом $R_{i+1} = 2 R_i \pmod{1}$.

Можно использовать и более сложные правила преобразования битов. Пусть имеется последовательность битов, пронумерованных от 1 до n и имеющих значения a_1, a_2, \dots, a_n . Будем вычислять значение a_0 по правилу

$$a_0 = a_{18} \wedge a_5 \wedge a_2 \wedge a_1,$$

где \wedge – XOR – исключающее ИЛИ. После этого сдвинем все биты на один, так что a_0 станет a_1 , a_1 станет a_2 , a_n пропадет.

Ниже приводится пример реализации алгоритма на языке C.

```
int srrandom(unsigned long *iseed)
{
    unsigned long newbit;
    newbit = (*iseed >> 17) & 1 ^
    (*iseed >> 4) & 1 ^ (*iseed >> 1) & 1 ^ (*iseed & 1) ;
    *iseed=(*iseed << 1) | newbit;
    return (int) newbit;
}
```

Библиографический список

1. Васильков Ю.В., Василькова Н.Н. Компьютерные технологии вычислений в математическом моделировании : учебное пособие / Ю.В. Васильков, Н.Н. Василькова. – Москва : Финансы и статистика – 2002 – 255 с. - ISBN 5-279-02098-2
2. В.Н. Ашихмин и др. Введение в математическое моделирование : учебное пособие / Под ред. П.В. Трусова. – М.: Логос, 2004. – 440 с.
3. Советов Б.Я., Яковлев С.А. Моделирование систем (3-е изд.). – М.: Высшая школа, 2001. – 420 с.
4. Замятина О.М. Моделирование систем: Учебное пособие. – Томск: Изд-во ТПУ, 2009. – 204 с.
5. Губарев В.В. Системный анализ в экспериментальных исследованиях / В.В. Губарев. – Новосибирск: Изд-во НГТУ, 2000. – Ч. 1. – 99 с.
6. Бадмильсон В.Н. Информатика и вычислительная техника. Моделирование : Курс лекций // Информатика и вычислительная техника. Моделирование URL: <http://www.badmivasil.ru/index.php> (дата обращения: 07.06.2020).
7. Максимей И. В. Имитационное моделирование на ЭВМ. – М.: Радио и связь, 1988.
8. Гусев Ю.И., Карасев И Н., Кольман-Иванов Э.Э. Конструирование и расчет машин химических производств. – М., Машиностроение, 1985. – С. 12-14.
9. Седов Л. И. Методы подобия и размерности в механике, М., 1972
10. Марка Д. А., МакГоуэн К. Методология структурного анализа и проектирования. – М.: МетаТехнология, 1993
11. Васильев К.К., Математическое моделирование систем связи: учебное пособие / - Ульяновск: УлГТУ, 2008. 170 с.
12. Дьяконов В.П., MATLAB 6.5 SP1/7.0 + SIMULINK 5/6. Основы применения / В.П. Дьяконов. - М.: СОЛОН-ПРЕСС, 2005. - 800 С.
13. Гультияев А.В., Визуальное моделирование в среде MATLAB: Учебный курс / А.В. Гультияев. – СПб.: ПИТЕР, 2000. – 432 С.

14. Пригарин С. М., Методы численного моделирования случайных процессов и полей / С. М. Пригарин. – Новосибирск: Изд-во ИВМиМГ СО РАН, 2005. – 259 с
15. Тихонов А.Н., Вводные лекции по прикладной математике: учебное пособие / А.Н. Тихонов, Д.П. Костомаров. – М.: Наука, 1984. – 190 с.
16. Шеннон Р., Имитационное моделирование систем – искусство и наука/ Р. Шеннон; пер. с англ. под ред. Е.К. Масловского. – М.: Мир, 1978. – 418 с.
17. Ермаков С.М., Статистическое моделирование: учебное пособие / С.М. Ермаков, Г.А. Михайлов.– М.: Наука, 1982. – 296 с.
18. Пригарин С.М., Методы численного моделирования случайных процессов и полей / С.М. Пригарин. – Новосибирск: Изд-во ИВМиМГ СО РАН, 2005. – 259 с.
19. Шалыгин А.С., Прикладные методы статистического моделирования / А.С. Шалыгин, Ю.И. Палагин. – Л.: Машиностроение, 1986. – 320 с.
20. Хан Г., Статистические модели в инженерных задачах / Г. Хан, С. Шапиро; пер. с англ. Е.Г. Коваленко, под ред. В.В. Налимова. – М.: Мир, 1969. – 396 с.
21. Васильев К.К., Методы обработки сигналов: учебное пособие / К.К. Васильев. – Ульяновск: УлГТУ, 2001. – 78 с. 22. Тихонов В.И., Статистический синтез и анализ радиотехнических систем и устройств / В.И. Тихонов, В.Н. Харисов. – М.: Радио и связь, 1991. – 608 с.
22. Быков В.В., Цифровое моделирование в статистической радиотехнике / В.В. Быков. – М.: Советское радио, 1971. – 328 с.
23. Полляк Ю.Г., Вероятностное моделирование на электронных вычислительных машинах / Ю.Г. Полляк. – М.: Советское радио, 1971. – 400 с.
24. Полляк Ю.Г., Статистическое машинное моделирование средств связи / Ю.Г. Полляк, В.А. Филимонов. – М.: Радио и связь, 1988. – 176 с. (Серия СТС – вып.30)
25. Основы научных исследований: учебник для вузов / под ред. В.И.

Крутова и В.В. Попова. – М.: Высшая школа, 1989. – 400 с.

26. Сизиков В.С., Устойчивые методы обработки результатов измерений: учебное пособие / В.С. Сизиков. – СПб.: СпецЛит, 1999. – 240 с.

27. Комашинский В.И., Системы подвижной радиосвязи с пакетной передачей информации. Основы моделирования / В.И. Комашинский, А.В. Максимов. – М.: Горячая линия – ТЕЛЕКОМ, 2007. – 176 с.

28. Вентцель Е.С., Теория вероятностей: учебник для вузов / Е.С. Вентцель. 8-е изд., перераб. и доп. – М.: Физматлит, 1999. – 576 с.

29. Вероятностные методы в инженерных задачах: справочник / А.Н. Лебедев, М.С. Куприянов, Д.Д. Недосекин, Е.А. Чернявский. – СПб.: Энергоатомиздат, 2000. – 333 с.

30. Гультияев А. Визуальное моделирование в среде MATLAB : Учебный курс – СПб: Питер, 2000. – 432 с.: ил.

31. Кравченко Н.С. Методы обработки результатов измерений и оценки погрешностей в учебном лабораторном практикуме: учебное пособие / Н.С. Кравченко, О.Г. Ревинская; Национальный исследовательский Томский политехнический университет. – Томск: Изд-во Томского политехнического университета, 2011. – 88 с.

32. Советов Б.Я., Яковлев С.А. Моделирование систем: Учебник для ВУЗов — 3-е юд., перераб. и доп. – М.: Высш. шк., 2001. — 343 с: ил.