

САПР в электрофизике.

Часть I. Основы автоматизации проектирования.

Глава 1. Введение в проблему.

1.1 Мотивация развития САПР

Дисциплина САПР, так же как аналогичное научное направление, возникло сравнительно недавно: в 60^х-70^х годах прошлого столетия (и тысячелетия), но, тем не менее, сравнительно быстро нашло признание и успешно внедряется в практику во всех развитых странах мира, хотя следует отметить, что ввиду новизны этой проблемы и разнообразных сфер её внедрения, представление о ней не всегда однозначны даже среди специалистов. В зарубежной практике это направление определяется как CAD-Computer Aided Design (разработка с помощью компьютера), у французов принято сокращение CAO. Эти сокращения часто применяются для обозначения пакетов прикладных программ, используемых разработчиками в различных аспектах научно-технической деятельности (Auto CAD, P-CAD, Mat CAD и т.п.). Разработки крупных программных комплексов связана с так называемой CASE-технологией (Computer Aided Software Engineering). Все эти направления можно рассматривать как составные части общей проблемы автоматизированного проектирования.

Чтобы понять основные причины интенсивного развития автоматизированного проектирования, необходимо определить задачи, решаемые в процессе проектирования.

Как правило, под проектированием (на первый взгляд) понимается изготовление чертёжно-конструкторской документации (кульман, карандаш, калька и т.п.). Безусловно, изготовление чертежей трудоёмкий, рутинный и длительный процесс, но он является лишь завершающим этапом проектирования. Хотя автоматизация проектирования начиналась со стремления кардинально изменить процедуру изготовления чертёжно-конструкторской документации. Однако следует отметить, что существует понятие функциональное проектирование, составляющее основу проектирования. Оно включает в себя решение трудоемких задач, связанных с определением принципов построения объектов проектирования и оценки свойств на основе исследования процессов их функциони-

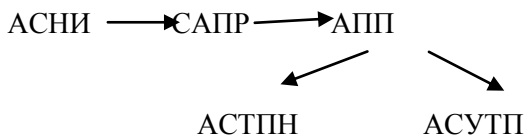
рования. Автоматизация функционального проектирования предполагает решение этих задач с помощью функциональных математических моделей, объектов проектирования на МИКРО, МАКРО и МЕТА уровнях. Именно об этом проектировании в основном будет идти речь в дальнейшем. Вопросы машинной графики, машинной геометрии и т.п. рассмотрены в смежных разделах.

В самом общем виде под автоматизацией проектирования можно понимать технологию использования вычислительной техники для оказания помощи проектировщику при выработке, модификации, анализе и оптимизации проектных решений, как говорил Норберт Винер: «отдайте человеку человеческое, а вычислительной машине машинное. В этом и должна, по-видимому, заключаться разумная линия поведения при организации совместных действий людей и машин».

Существует два научных направления смежных с САПР и так же базирующихся на использовании различных средств вычислительной техники. Начальный этап разработки новых изделий и поиска новых принципов связан с так называемыми АСНИ автоматизированными системами для научных исследований. Эта стадия научного поиска существовала всегда и являлась лидером в деле использования автоматизированных систем для изучения новых идей и использования новых принципов. АСНИ предваряет непосредственное проектирование, являясь более универсальным этапом, поскольку значимые результаты этой деятельности могут быть использованы при проектировании в различных областях техники, являясь составной и очень важной на современном этапе частью научного поиска. В современном понимании это связывают с модным словом инновации (раньше использовалось слово внедрение). С другой стороны на завершающем этапе процесса создания новых изделий, так же связанным с интенсивным (в современных условиях) использованием вычислительной техники находятся так называемые системы АПП - автоматизации производственных процессов. Они включают две функционально-различных и взаимодополняющих системы АСТПП - автоматизированные системы технологической подготовки производства и АСУТП - автоматизированные системы управления технологией производства.

Таким образом, весь цикл создания новых изделий может быть представлен в виде последовательности следующих этапов, в

основе технологии каждого из которых находятся средства вычислительной техники.



В американской литературе в связи с вышесказанным употребляется сокращение CAD/CAM, где CAM- Computer Aided Manufacturing. Это лишний раз указывает, что использование средств вычислительной техники на всех этапах от научного поиска до изготовления готовых изделий является основной тенденцией научно-технического прогресса на современном этапе.

Лидером подобного подхода являются разработчики электронной техники, точнее микроэлектронной техники, где разработка и изготовление схем высокой степени интеграции представляет замкнутый цикл на основе автоматизированных систем- от разработки принципиальных схем, их моделировании и оптимизации, разработки конструкции до технологической подготовки и изготовления готовых изделий.

Выделение автоматизированного проектирования в отдельную дисциплину, как самостоятельного научно-технического направления, связано со значительными изменениями, произошедшими при подходе к технологии проектирования, значительному усложнению этой технологии в связи с появлением широкого спектра новейших технологических средств, математического и программного обеспечения, ориентированного, в том числе и на изготовлении чертёжно-конструктивной документации.

Предмет автоматизации проектирования на современном этапе включает следующее:

- формализация проектных процедур;
- структурирование и формализация процессов проектирования;
- разработка моделей проектируемых объектов;
- разработка методов и алгоритмов проектных задач;
- способы построения технических средств;

- разработка входных языков проектирования;
- и т.п.

Говоря о САПР, следует отметить, что с одной стороны это узко специализированная система. Так САПР электроники отличается от САПР радиотехники и приборостроения, не говоря уже о САПР машиностроения и такой узко специализированной системы автоматизированного проектирования, как разработка пульта управления пилота самолета. Очень специфично проектирование ядерно-энергетических систем и электрофизических установок.

В тоже самое время, отмечая специфику автоматизации проектирования в различных областях техники, необходимо подчеркнуть, что ряд основополагающих положений теории и практики САПР носит универсальный характер, что связано, прежде всего с тем, что основой любой системы автоматизации проектирования являются средства вычислительной техники, являющиеся универсальным инструментом и требующим их подробного изучения. Проблемы использования и принципы построения технических средств, структура программного обеспечения, как универсального, так и специализированного, выбор инструментальных средств разработки прикладных программ, принципы разработки математической модели входных языков носят в значительной степени универсальный характер и практически не зависят от области применения.

В связи с этим в изучении автоматизации проектирования можно выделить два уровня:

- изучения общих, универсальных составляющих САПР, не зависящих от области применения;
- специализированные разделы, отражающие специфику конкретной предметной области;

Всеобщий интерес к автоматизации проектирования и интенсивное его развитие начинается с 60^х-годов прошлого столетия, связан с двумя основными факторами: кризисом в традиционных методах проектирования и революционными событиями, происходящими в электронике и вычислительной технике. Именно в этот период очень своевременно сформировалась новая база, которая позволила преодолеть критическую ситуацию, сложившуюся к тому времени в проектировании.

Следует отметить, что хотя автоматизированное проектирование и представляет одно из основных направлений научно-технического прогресса конца 20^{го} века, оно не связано с какими-то фундаментальными открытиями или новыми достижениями именно в этой области, а целиком и полностью обязано колоссальным технологическим достижениям в смежной области-области новых информационных технологий. Развитие этого направления показало, что не всегда даже очень заманчивые идеи (как то сверхпроводимость или термоядерный синтез) могут привести к положительным практическим результатам, а вдумчивое осознание современных технологических возможностей позволяет решить важнейшие проблемы современности.

Как указывалось выше первая (и основная) причина появления САПР связана с кризисом в проектировании - из-за несоответствия традиционных методов проектирования новым потребностям проектировщиков, постоянно усложняющимся объектами проектирования, появление понятия сложной технической системы (С.Т.С.). Так, в среднем, по сравнению с 50-ми годами сложность объектов проектирования в 60^х-70^х-годах увеличилась в 5-6 раз. Так документация на современный самолёт того времени по весу была примерно равна весу самого самолёта. Очень серьёзные изменения, потребовавшие современно новых подходов к проектированию происходили в электронике, что связано с колоссальным увеличением степени итерации интегральных микросхем (ИМС) (переход от отдельных транзисторов 50^х годов к 100 и более миллионов транзисторов в одной ИМС) и появлением микропроцессоров, степень интеграции в которых приближается к одному миллиарду.

Характерный случай, указывающий на тенденцию значительного усложнения объектов разработки, произошёл на кафедре электрофизических установок МИФИ в 70^х-годах прошлого столетия. Кафедра, на коммерческой основе, выпускала небольшие линейные электронные ускорители для промышленности и медицины.

Для оснащения ускорителей необходимым вспомогательным оборудованием - электронными системами управления, пультами операторов и придания установкам современного дизайна была предпринята попытка привлечь для решения этих задач разработчиков из ГДР. Однако, после составления совместного технико-

экономического обоснования, выяснилось, что стоимость вспомогательного оборудования втрое превышает стоимость основной части установки-самого ускорителя, совместное предприятие не состоялось.

Ситуация с усложнением объектов проектирования и появление СТС привел к ряду проблем, требующих оперативного решения.

Значительно увеличились сроки проектирования. Так составление проекта нового торгового судна требовало 2-4 года, самолёта 4-6 лет, ЭВМ 3-4 года. Создание крупного ускорителя от 4 до 10 лет. Причём подготовки предварительных соображений (предпроектные исследования очень важный этап проектирования) занимал от 2^х до 4^х лет.

В качестве примера ниже приводятся сроки сооружения крупнейших ускорителей того времени

Ускоритель	Годы создания
Синхрофазотрон 10 ГэВ (Дубна)	1952-1957
Проточный синхротрон 26 ГэВ (СРС, ЦЕРН)	1955-1959
Электронный синхротрон 1,3 ГэВ (Томск)	1958-1965
Электронный синхротрон 7,2 ГэВ (DESY <i>ZZZZZZZZZZZZ</i>)	1959-1964
Проточный синхротрон 76 ГэВ (Протвино)	1961-1967
Линейный ускоритель электронов 22 ГэВ (Стэн-фарт, США)	1962-1966
Мезонная фабрика 75-100 МэВ, 100-800МэВ (Лос-Алинос, США)	1962-1972
Проточный синхротрон 200-500 ГэВ (Батавия, США)	1968-1972
Супер протонный ускоритель SPS, 400 ГэВ (ЦЕРН)	1970-1976

Увеличение времени, затрачиваемого на проектирование, само по себе не желательно, так как значительно увеличиваются сроки ввода новых изделий в эксплуатацию, более того в ряде областей техники (в частности в электронике) было невозможно традиционными методами создавать новую продукцию. Как следствие

увеличение сроков проектирования является и то, что идеи, заложенные в проекте, за время его разработки устаревают ещё до ввода в эксплуатацию новых изделий, что связано с опережающими темпами роста научно-технического прогресса. Если 50-е годы прошлого столетия жизненный цикл эксплуатации разработок равнялся профессиональной жизни разработчика-человека, то к 60-70 годам время разработки зачастую превышало время эксплуатации и к моменту создания (выпуска) изделий они оказывались устаревшими. Типичным примером такой ситуации явилась выпускаемая в Советском Союзе единая система вычислительных машин - ЕСВМ. Таким образом, одновременно с увеличением сроков проектирования происходит сокращение времени их жизни.

Ещё одним фактором, приведшим к кризису проектирования в связи с появлением СТС, явилось значительное увеличение лиц занятых в разработке проектов (конструкторов, чертёжников, копировщиков и т.п.) и невысокая престижность конструкторского труда, ввиду рутинного характера большинства проектных процедур. Что в свою очередь приводило к значительному удорожанию проектов.

Вследствие этих причин в цикле создания новой техники: научное исследование-проектирование-производство узким местом явилось не отсутствие научных достижений и даже не изготовление, а медленное освоение имеющихся научных достижений через проектирование.

Вторая причина связана с появлением САПР, позволившая преодолеть кризисные явления этого развития новой технологической базы проектирования, в связи с появлением и бурным развитием новейших информационных технологий на базе средств вычислительной техники.

Чтобы понять суть перемен, произошедших в проектировании и место новых информационных технологий в этом процессе, необходимо рассмотреть детально виды деятельности, присутствующие в этом процессе. На рис. 1 представлен набор

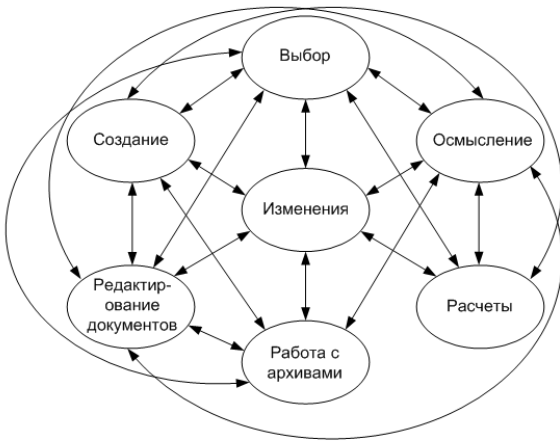


Рис.1. Различные виды деятельности в КБ

разновидностей этой деятельности и взаимосвязь между ними. В нижней части рисунка представлены виды работ, необходимые для выполнения задачи проектирования и их процентное соотношение. Из этого рисунка следует, что весь процесс проектирования представляется как чередование творческих и формальных (рутинных) видов деятельности, причём последние составляют большую его часть (>>90%). К ним относятся – расчёты, хранение, поиск, обработка информации и результатов экспериментов, а так же изготовление технической документации.

Большой ассортимент технических средств современной вычислительной техники, развитие программного обеспечения, автоматизация программирования, неограниченные ресурсы вычислительных систем, интеллектуальные интерфейсы и т.п. предоставляют возможность проектировщикам перенести на компьютер рутинные виды деятельности, значительно повысив эффективность проектирования.

1.2 Основные этапы становления автоматизированного проектирования

Таким образом, автоматизация проектирования с помощью средств вычислительной техники, основной путь, на котором преодолеваются противоречивые трудности, связанные с сокращением сроков проектирования при одновременном увеличении сложности разработок.

Целью автоматизированного проектирования является объединение конструктора и ЭВМ в единую команду для решения задач, способную приходить к поставленным целям в задачах проектирования более эффективно, чем каждая из них, работающих по отдельности. Принципы, лежащие в основе этого разделения, а так же многие другие принципы автоматизированного проектирования (о которых будет сказано ниже) находятся в зависимости от последних достижений в области новых информационных технологий.

Основные задачи, которые ставятся при переходе на автоматизированное проектирование следующие:

- Сократить сроки проектирования;

- Снизить материальные затраты на проектирование;
- Повысить качество проектирования;
- Сократить количество конструкторов и чертёжников, занятых на рутинных операциях;

Хотя автоматизация проектирования - основной путь развития проектирования, но этот путь не тривиален и не всегда все поставленные цели достигаются одновременно:

-На смену кульману приходят достаточно дорогие технические средства, мощные рабочие станции, устройства графического ввода-вывода чертежей, соответствующее программное обеспечение и многое другое;

-Математическая постановка задач для большинства проектных процедур не очевидна, а их последующая алгоритмизация требует разработки оригинальных методов, что в значительной мере определяет содержание теории САПР.

Перечисление проблем, возникающих при внедрении САПР, можно продолжить, но, несмотря на эти трудности, успешное внедрение в практику проектирования подтверждает реальность решения перечисленных проблем.

Уже на ранних стадиях развития новых методов проектирования успешно решались эти задачи в станкостроении, электронной технике, судостроении и т.п. Впечатляющие достижения демонстрируют зарубежные фирмы. Так фирма Boeing способна в течение года запустить в производство новый авиалайнер (к примеру, отечественный авиалайнер ТУ 234 запускался в производство в течении пяти лет).

Можно выделить несколько последовательных этапов развития САПР, тесно связанных с динамикой развития средств вычислительной техники и методов её использования.

Начальный этап связан с использованием вычислительной машины при решении сложных математических задач, возникающих при проектировании. На этом этапе важнейшую роль играло развитие методов вычислительной математики и её внедрение в инженерную практику.

Развитие методов проектирования тесно связано с развитием методики использования вычислительной техники. Решение инженерных математических задач на компьютере (так же как и решение любых других задач) строится по следующей схеме

1. Математическая формулировка задачи.
2. Разработка алгоритма.
3. Выбор численных методов решения задачи.
4. Кодирование алгоритма (программирование).
5. Перфорация (запись программы на различные носители устройств ввода вывода компьютера).
6. Отладка программы и её тестирование.
7. Решение задачи.
8. Обработка результатов.

На начальном этапе использования компьютеров только 7^й пункт выполнялся без участия человека оператора или программиста, а эффективность использования компьютеров была крайне низкой.

Решающую роль в переходе на новый уровень использования компьютеров, имеющий важное значение в автоматизации проектирования, явилась тенденция к универсализации задач и разработка единых подходов к целому классу расчётно-проектных процедур. На начальном этапе этот подход связан с реализацией модульного программирования созданием библиотек по специальности в виде подпрограмм функций и процедур, как на системном уровне, так и в виде внешних процедур, собираемых на стадии компоновки.

Были разработаны библиотеки как для широкого круга применений, как например библиотеки научных подпрограмм на фортране фирмы IBM, так и более специализированные библиотеки для различных научно-технических приложений.

Основная задача, которая ставилась на этом этапе, по возможности, избавить разработчика от рутинной алгоритмизации и кодирования задачи, т.е. исключить ручной труд на этапах 2-5.

Математическая постановка задачи, безусловно, определяется разработчиком. Широкая универсализация задач, позволяющая абстрагироваться от деталей алгоритмизации и программирования (кодирования) увеличивает накладные расходы на ресурсы компьютеров, что при современных темпах развития технических средств не актуально. Гораздо более актуальным является то обстоятельство, что большинство универсальных библиотек не учитывает особенности конкретных приложений и в ряде случаев были не пригодны для практики.

На этом же этапе автоматизации проектирования успешно решались задачи новых способов оформления технической документации, оперативной обработки и отображения результатов проектирования. Это связано с созданием и быстрым совершенствованием специализированных пассивных и интерактивных устройств и систем машинной графики – графопостроителей, дигитайзеров, разнообразных специализированных дисплеев и т.п.. Одновременно развивается специализированное программное обеспечение машинной графики и её более высокая степень - машинная геометрия. Лидерами этого направления были ведущие фирмы США. Так уже в 1963 сотрудники МТИ (Массачусетского технологического института) представили доклад, в котором демонстрировали возможности формирования изображения на экране и манипулирование им в реальном масштабе времени, что явилось началом интерактивной машинной графики (ИМГ). И уже к концу 60^х годов (прошлого столетия). Крупнейшие американские концерны (General Motors, IBM, Lockheed, MC Donal Douglas) включились в создание ИМГ. К началу 70^х годов появилось несколько поставщиков САПР\АПП.

На этом этапе дальнейшее развитие автоматизации проектирования виделось в дальнейшем увеличении мощности компьютеров и развитии специализированной периферии. Однако довольно быстро выяснились ограниченные возможности этого подхода при проектировании сложных технических систем, требующих системного подхода.

Как выразился известный специалист в области информационных технологий академик Моисеев П.П.: «Простое наращивание вычислительной мощности без качественного изменения всей её информационной основы, без надлежащей организации человеческого труда не может быть ожидаемого эффекта, а напротив приводит к «вавилонскому столпотворению»»

Наглядный пример «вавилонского столпотворения» - ситуация, сложившаяся на американской фирме Boeing. В 1970 году на фирме был проведён тщательный анализ результатов около 700 расчётных работ с использованием ЭВМ. Результаты были ошеломляющими. Каждое второе решение расчётной задачи было неверным. В большинстве случаев это не было обнаружено своевременно и результаты пошли в качестве исходных данных в решение

других задач. Дополнительные затраты на устранение каждой ошибки составляли от 2^х до 15 человекомесяцев.

Как показал анализ причин ошибок, они были связаны отнюдь не с низким качеством компьютерных программ, а с низким качеством технической культуры использования вычислительной техники в проектировании. Из-за отсутствия четкой организации вычислений, из-за путаницы в передаче информации, из-за ошибок, внесенных в результаты самими исполнителями. Именно подобного сорта обстоятельства привели к необходимости системного подхода к автоматизации проектирования, т.е. к переходу на новый этап систем автоматизированного проектирования.

1.3. Определение структуры и принцип построения САПР.

По существу современный САПР представляет систему проектирования, в которой произведено рациональное, на данном этапе развития технических и программных средств ЭВМ распределение функций между людьми (проектировщиками) и средствами вычислительной техники. При этом вся процедура проектирования рассматривается с системных позиций, как единый процесс, начинающийся с разработки технического задания и заканчивающийся изготовлением технической документации, подготовленной для запуска на изготовление изделия в производство. ГОСТом введено следующее определение САПР: «САПР-это комплекс средств АПР (автоматизированного проектирования), взаимосвязанных с необходимыми подразделениями проектной организации или коллективом специалистов, выполняющих АПР».

По своему назначению основные подсистемы, входящие в структуру САПР могут подразделяться на:

-Проектирующие, имеющие объектную ориентацию и реализующие определённый этап (стадию) проектирования.

-Обслуживающие, имеющие общесистемное применение и обеспечивающих поддержку функционирования проектирующих систем, а так же оформление, передачу и вывод полученных результатов.

Средства автоматизации проектирования можно сгруппировать по видам обеспечения, которое включает в себя следующее:

Математическое обеспечение (МО), основа которого - алгоритмы, по которым разрабатывается программное обеспечение. Элементы МО в САПР чрезвычайно разнообразны. Среди них - принципы построения функциональных моделей, методы решения алгебраических и дифференциальных уравнений и т.п.

Разработки МО - самый сложный этап создания САПР. По назначению и способам реализации МО САПР делится на 2 части: математические методы и построенные на их основе математические модели, описывающие объекты проектирования;

- Формализованное описание технологии автоматизированного проектирования.

Программное обеспечение (ПО) - совокупность всех программ и эксплуатационной документации к ним, необходимых для выполнения автоматизированного проектирования.

Имеется две части ПО:

-Общесистемное ПО предназначено для организации функционирования технических средств, т.е. для планирования и управления вычислительным процессом.

-Специальное (прикладное ПО) реализует математическое обеспечение для непосредственного выполнения проектных процедур обычно (на современном этапе имеет форму пакетов прикладных программ (ППП), на каждый из которых обслуживает определенный этап процесса проектирования. ППП представляют некоторую операционную среду, работающую под управлением ОС со своими специфическими системными и обрабатываемыми программами, а так же специализированное ПО, реализующее конкретные проектные процедуры.

В настоящее время разработано большое количество ППП, реализующих САПР в разнообразных областях науки и техники - в электронике, машиностроении, математике и т.п.

Информационное обеспечение (ИО) -основу информационного обеспечения САПР составляют данные, которыми пользуется проектировщик в процессе проектирования. Они могут быть представлены в виде документации на различных носителях, содержащих сведения справочного характера.

Различают следующие способы ведения информационного фонда САПР: использование файловых систем и использование

баз, банков данных, создание информационных программ адаптеров.

Техническое обеспечение – это совокупность взаимосвязанных и взаимодействующих технических средств, предназначенных для выполнения автоматизированного проектирования.

Лингвистическое обеспечение – это специальные языковые средства (языки проектирования), предназначенные для описания процедур автоматизированного проектирования и проектных решений – входные языки пакетов прикладных программ. Сюда же следует отнести традиционные инструментальные средства операционных систем – традиционные (императивные) языки программирования.

Методическое обеспечение включает входящие в САПР документы, регламентирующие порядок её эксплуатации, включая описание ППП.

Организационное обеспечение – инструкции, приказы, штатное расписание, квалификационные требования и др. документы, регламентирующие организационную структуру подразделений проектной организации.

Основные принципы построения САПР

-САПР-это человеко-машинная система. Тесное взаимодействие человека и ЭВМ в процессе проектирования - один из принципов построения и эксплуатации САПР. В настоящее время и, по крайней мере, в ближайшие годы создание САПР «не угрожает» монополии человека при принятии узловых решений в процессе проектирования. Человек должен решать в САПР задачи, формализация которых не достигнута и решаются на основе опыта и интуиции человека.

-САПР - иерархическая система - она реализует комплексный подход к автоматизации всех уровней проектирования.

Следует особо подчеркнуть целесообразность комплексного характера САПР, т.к. автоматизация проектирования на одном из уровней при сохранении старых форм проектирования не даёт заметного суммарного эффекта.

-САПР-совокупность информационно согласованных подсистем, что означает - все или большинство последовательностей задач проектирования обслуживается информационно согласованными программами. Две программы называются информационно

согласованными, если все те данные, которые представляют собой объект переработки в обеих программах, входят в числовые массивы, не требующие изменений при переходе из одной программы к другой. Плохая информационная согласованность превращает САПР в совокупность программ. При этом из-за неучёта в подсистемах многих факторов, оцениваемых в других подсистемах, снижает качество проектных решений.

-САПР - открытая и развивающаяся система, т.к. имеется как минимум две причины, по которым она изменяется во времени:

а) Разработка такой сложной системы, как САПР, занимает продолжительное время и экономически выгодно вводить отдельные её части по мере готовности;

б) Прогресс в вычислительной техники и развитие вычислительных методов приводит к накоплению более совершенных методов автоматизации проектирования и необходимость замены старых.

В связи с этим открытость системы предполагает замену отдельных блоков системы без глобальной перделки системы в целом.

-САПР - специализированная система с максимальным использованием унифицированных модулей. Требование высокой эффективности и универсальности, как правило, противоречивы. Высокой эффективности достигают за счёт специализации. При этом число САПР растёт, растёт и их стоимость. Что бы снизить стоимость, целесообразно включить в их состав недорогие унифицированные модули.

§ 1.4 Особенности САПР ЭФУ

Имея общие базовые принципы разработки автоматизированных систем проектирования и традиционный набор видов обеспечения, использующих достижения современных (на данном этапе развития) информационных технологий, каждое направление проектирования разнообразных технических устройств имеет свои характерные и весьма значительные особенности. Эти особенности связаны, прежде всего, с особенностями математического, программного, лингвистического, информационного и т.п. видов обеспечения, отражающих специфику предметной области и типов разрабатываемых изделий. Специфика конкретной предметной области разработки обычно отражается в операционной среде пакетов

прикладных программ автоматизированного проектирования. Так на САПР электроники ориентированы такие широко известные пакеты как P-CAD, OrCAD, MicroCap, Electronics Work Bench и т.п.

Большую популярность и массовое распространение получили ППП AutoCAD фирмы Autodesk RF. Эта мощная операционная среда, включающая в себя большое количество общеинженерных приложений, ориентированных, в основном, на машиностроение, архитектуру и строительство. Существует большое количество ППП, ориентированных на определенные более узкие области техники: САПР в авиастроении, кораблестроении и т.п.

В большинстве проектных организаций общеинженерного профиля большая часть времени тратится на изготовление чертёжно-конструкторской документации, т.е. на черчение. В связи с этим именно эта часть работы изначально понималась проектированием и требовала разработки и внедрения новых технологий с использованием машинной графики, машинной геометрии, создания специализированных пассивных и интерактивных устройств ввода-вывода, систем машинной графики и автоматизированных рабочих мест (АРМ) конструктора.

Разработка программного обеспечения для поддержки этих устройств явилась необходимым атрибутом для создания в дальнейшем пакета прикладных программ автоматизации проектирования (CAD).

Методы расчёта в большинстве общеинженерных дисциплин хорошо изучены, стандартизированы, их результаты, как правило, представлены в виде таблиц, номограмм и т.п. Даже в вакуумной технике основные элементы вакуумных систем определены стандартами. Выбор типа вакуумных прокладок, размеры фланцев и необходимое количество болтов и их размеры для обеспечения надёжного уплотнения определяется из таблиц и нет потребности ни в прочностных и иных расчётах.

Характерным является проектирование в электронике (микроэлектронике) и электротехнике, где основными расчётными соотношениями являются компонентные законы Ома и топологические уравнения Кирхгоффа, а конструирование в основном связано с оптимизацией расположения элементов с целью уменьшения длины соединяющих проводников и проблемы теплоотвода.

Как правило, выше перечисленные изделия выпускаются массово (крупными сериями), а жёсткие требования стандартов связаны с жёсткими требованиями к разработке конструкторской документации.

Особенности автоматизации проектирования в электрофизике связаны с особенностями этого класса установок.

С самого начала разработка ускорителей и других установок электрофизики потребовали проведения сложных вычислительных работ большого объема с привлечением мощных вычислительных систем. С другой стороны не существовало, как токовой, промышленности по производству подобных установок. Разработки, как правило, проводились (и проводятся) физическими исследовательскими институтами и хотя отдельные типы ускорителей для промышленности и медицины получили достаточно широкое распространение и выпускаются рядом зарубежных фирм, общепринятых стандартов на них пока не существует. Среди наиболее характерных особенностей этих установок с точки зрения конструкторской разработки можно отметить следующее:

Мелкосерийность - для промышленности и медицины ускорители выпускаются некоторыми фирмами в единичных экземплярах, либо небольшими сериями (в пределах десятков штук).

Крупные исследовательские ускорительные комплексы уникальны и неповторимы. Это с одной стороны объясняет их высокую стоимость, а с другой отсутствие в необходимости стандартизации.

Очень большое количество принципов ускорения - и как следствие большое количество типов ускорителей. Это электростатические, линейные, индукционные мощные импульсные на основе генераторов Аркадыва-Маркса, линейные высокочастотные, и самые разнообразные циклические ускорители. В них ускоряются самые разнообразные частицы на различные энергии для различных целей.

Сложность теории ускорителей. Каждый тип ускорителей предусматривает свою методику расчёта и требует серьезной научно-исследовательской проработки и большого объема компьютерного численного моделирования. Важнейшее значение при проектировании ЭФУ связано с решением задач синтеза- выбора оптимальных вариантов на различных уровнях проектирования - от вы-

бора типа установки и круга решаемых на ней задач до оптимизации отдельных компонент выбранной установки.

Большое количество общеинженерных систем: Это вакуумная система, система компоновки, установки и юстировки, система автоматизации измерений и управления, энергетическая система и т.п.

Проектирование этих подсистем не имеет непосредственного отношения к САПР ЭФУ и решается в смежных областях техники, имеющих свои системы автоматизированного проектирования. Однако наряду с перечисленными подсистемами, проектирование которых достаточно хорошо проработано и существуют соответствующие системы автоматизированного проектирования, рассматриваемые нами установки имеют четыре подсистемы, характерные только для них и не имеют аналогов проектирования это: ускоряющая, фокусирующая, инжектирующая подсистемы и каналы транспортировки заряженных пучков

Проектирование этих подсистем связано с решением задач синтеза ускоряющих и фокусирующих элементов, через анализ граничных задач электромагнитостатики и электродинамики резонаторных и волноводных структур. Показателем эффективности работы разрабатываемых элементов установок является выходной ток ускоряемых заряженных частиц, определяемый их уравнениями движения в ускоряюще-фокусирующих и транспортирующих электромагнитных полях, создаваемых этими подсистемами. Особую сложность представляют так называемые самосогласованные задачи, в которых ускоряемый ток и заряд частиц оказывают заметное влияние на ускоряюще-фокусирующие поля.

В связи с этим под САПР ЭФУ (или САПР ускорителей заряженных частиц) будем понимать стадию предпроектных исследований, связанную с разработкой математических моделей, их анализом и синтезом для решения вышеперечисленных задач, т.е. разработки ускоряюще-фокусирующих и транспортирующих систем пучков заряженных частиц.

Разработка компьютерных программ для решения перечисленных задач была начата с конца 50^х годов прошлого века во всех крупных ускорительных центрах мира. Эти работы были, как правило, связаны с созданием национальных ускорительных центров. Однако ориентация этих программ на определённые технические

средства, операционные среды и различные версии языков программирования того времени затрудняли переместимость программных продуктов и использования их в других центрах. В дальнейшем, в связи со стандартизацией инструментального программного обеспечения и унификаций операционных систем, эти трудности были преодолены. Наиболее известным примером ускорительного САПР является, разработанный в Европейском центре ядерных исследований (ЦЕРН, Швейцария) фундаментальный программный комплекс MAD, получивший широкое распространение в мире, подробно и качественно документированный, однако для разработки небольших установок, которые используются в прикладных задачах для самых разных целей в промышленности, медицине, сельском хозяйстве и т.п. ППП MAD является с одной стороны избыточным, а с другой не всегда удовлетворяет специфическим требованиям различных установок.

В связи с этим в настоящее время в практике проектирования ускорителей используются узкоспециализированные ППП, о которых будет сказано в дальнейшем.

Фактически, все разработанные и разрабатываемые программные продукты при проектировании небольших установок прикладного характера ориентированы на две задачи:

-Расчет электромагнитных полей в ускоряющих и фокусирующих системах, а так же программы расчёта тепловых полей в этих системах. Эти программы имеют обширную сферу применения, выходящую за пределы электрофизики и разрабатываются специалистами из смежных областей.

-Расчёт динамики ускоряемых частиц в выбранных структурах (и оптимизация структур через анализ динамики). При необходимости учитывается влияние электромагнитных составляющих пучка на ускоряющее поле. Это, как правило, узкоспециализированные программы, ориентированные на определенный класс установок.

Глава 2. Общесистемные вопросы САПР, цели и задачи проектирования.

2.1 Краткие сведения о системном анализе

Системный анализ-это самостоятельное научное направление, которое интенсивно развивается с середины прошлого столетия. Появление его связано с рядом новых проблем, одной из которых является и проектирование, при решении которых оказались не приемлемы традиционные по тем временам методы.

Уже к концу XIX века стало резко увеличиваться число комплексных проектов и задач, требующих участия специалистов различных областей знаний. Постоянное усложнение объектов проектирования, развитие науки, появление проблем большой начальной неопределённости, которые невозможно было решить с помощью формальных математических методов прошлого столетия, требовало развития новых подходов при их решении.

Не меньшее влияние на развитие этого направления в науке вызвало возникновение мировых глобальных проблем, связанных с необратимыми катастрофическими последствиями развития цивилизации - экология и разрушение природных балансов, энергетика и захоронение ядерных отходов и т.п.

В этих задачах всё большее место стал занимать собственно процесс постановки задачи, возросла роль человека, как носителя целостного восприятия, сохранение целостности при расчленении проблемы для облегчения её решения, роль методов активизации интуиции и опыта специалистов различных областей знаний, принимающих участие при решении сложной проблемы.

Для решения подобных комплексных проблем стало широко использоваться понятие «система», и на определенной стадии развития научных знаний теория систем сформировалась в самостоятельную науку, имеющую непосредственное отношение к САПР. В 30-е годы XX века возникла теория открытых систем Л.Фон Бергналафи. Важный вклад в становление системных представлений внес в начале XX века А.А.Богданов, предложивший всеобщую организационную науку тектологию. Потребности практики, почти одновременно со становлением теории систем, привели к возникновению направления, названного исследование операций. Это на-

правление возникло в связи с задачами военного характера. Предметом исследований операций является разработка методов анализа целенаправленных действий (операций) и объективная сравнительная оценка решений (поиск оптимальных решений).

В 60-е годы XX века широкое распространение получил термин «системотехника», предложенный профессором Темниковым из МЭИ, как эквивалент американского «system engineering» (буквальный перевод - системная инженерия), который используется в основном в применении системных методов только в технических приложениях (систематология, аналогичный термин, используемый в философской литературе).

В первые годы становления теории систем наибольшее распространение получил термин системный подход, который использовался как методологическое направление философии, так и в прикладном смысле, как синоним комплексный подход.

Наиболее конструктивным из направлений системных исследований в настоящее время считается системный анализ, занимающийся применением методов и моделей теории систем для принятия решений. Впервые понятие системный анализ было предложено в 1948 г. корпорацией RAND (мозговой центр министерства обороны США) в связи с задачами военного управления. Применяется в тех случаях, когда задача (проблема) не может быть сразу представлена и решена с помощью формальных математических методов. При этом основное внимание уделяется процессу постановки задачи и использованию не только формальных, но и методов качественного анализа. Опирается на основные понятия теории систем и философские концепции, лежащие в основе общесистемных закономерностей.

Рассмотрим некоторые из основных понятий теории систем, которые могут быть нам полезны в дальнейшем: что такое система или сложная техническая система, о которой мы уже упоминали раньше?

Существует несколько десятков определений этого понятия, когда хотят охарактеризовать исследуемый или проектируемый объект, как нечто целое (единое), сложное, о котором невозможно сразу дать представление, показав его, изобразив практически или описав математическим выражением. Тем не менее, можно определить что система (большая система, система большого масштаба,

СТС и т.п.). Эта совокупность двух или более элементов, удовлетворяющих следующим условиям:

-Наличие большого количества взаимосвязанных и взаимодействующих между собой элементов, обладающих свойством совместности;

-Поведение каждого элемента влияет на поведение целого;

-Поведение элементов и их взаимодействие в целом взаимосвязаны;

- Сложность функций, выполняемых системой;

- Возможность разбиения системы на подсистемы, цели функционирования которых подчинены общей цели.

-Наличие управления (часто имеющего иерархическую структуру), разветвленной информационной сети интенсивных потоков информации;

- Наличие взаимодействия с окружающей средой (внешней) и функционирование в условиях воздействия внешних факторов;

При формальном подходе к сложной системе элементом считается объект, не подлежащий дальнейшему расчленению. Внутренняя структура элемента не подлежит изучению. Любая совокупность элементов сложной системы может рассматриваться как подсистема. Обычно подсистемой является некоторая самостоятельно функционирующая часть системы. Правильное выделение подсистемы в СТС часто способствует упрощению расчётов при исследовании и обеспечивает большую наглядность при интерпретации результатов.

Основное свойство системы - это самоорганизация (лобализация) или можно сказать устойчивость. В широком смысле слова под этими словами понимается способность системы возвращаться в состояние равновесия после того, как она была из этого состояния выведена под влиянием внешних сил или способность после воздействия внешней среды путем последовательного изменения своих свойств придти к некоторому устойчивому состоянию (предполагается, что воздействие оказывается в некоторых допустимых пределах.)

Осознание всех выше перечисленных проблем позволило американскому математику Р. Акроффу назвать их основным источником интеллектуальной революции, вызывающей смену эпох. Уходящую эпоху он назвал веком машин, а наступающую веком

систем. Появление этого направления в науке привело к развитию специфических направлений в математике-теории массового обслуживания, теории игр, исследования операций, математическое программирование, дискретная математика, значительно расширилась сфера применения теории вероятностей и математической статистики.

2.2 Блочно-иерархический подход - основной метод исследования сложных систем.

Блочно-иерархический подход не является каким то неожиданным изобретением. Он явился естественной реакцией человечества на сложность в исследовании и описании различных больших по объему, многокомпонентных с большим количеством связей объектов. Его фрагменты в виде блок-схем алгоритмов, структурных схем различных технических объектов и т.п. традиционно использовались для облегчения понимания и разработки в инженерной практике.

Дело в том, что описание технических объектов, должно быть согласованно (по сложности) с возможностями восприятия человека. Однако выполнить это требование в рамках одного описания, не расчлняя его на некоторые составные части, удастся лишь для простых изделий. В СТС, как правило, требуется структурирование описаний и соответствующее расчленение представлений о проектируемом объекте на иерархические уровни и аспекты. Это позволяет распределить работы проектирования между подразделениями проектной организации, что способствует повышению эффективности и труда проектировщиков.

Разделение описаний по степени детализации отображаемых свойств и характеристик объектов лежит в основе блочно-иерархического подхода к проектированию и приводит к появлению иерархических уровней (уровней абстрагирования) в представлении о проектируемом объекте.

На каждом уровне используются свои понятия систем и элементов. При этом элементы верхнего уровня иерархии являются системами для нижнего уровня. Подобное разбиения продолжается вплоть до получения на некотором уровне элементов, описание которых дальнейшему расчленению не подлежит. Такие элементы

называются базовыми. Таким образом, в основе блочно-иерархического подхода используются 2 принципа:

-Принцип иерархичности, означающий структурирование представлений об объектах проектирования по степени детализации описаний;

- Принцип декомпозиции (блочности)- разбиение представлений каждого уровня на ряд составных частей (блоков) с возможностями раздельного (не блочного) проектирования объектов.

Кроме расчленения описаний по степени подробности, отображение свойств проектируемого объекта, порождающей иерархические уровни (которые иногда называют горизонтальные уровни) используется декомпозиция описаний объекта по характеру отображаемых свойств, что приводит к появлению ряда, т.н. аспектов описаний.

В соответствии с ЕСКД вводятся функциональный, конструкторский и технологический аспекты, которые считаются вертикальными уровнями. Решения задач проектирования на этих уровнях называют соответственно функциональным, конструкторским и технологическим проектированием.

Функциональный аспект связан с отображением основных принципов функционирования, характера физических и информационных процессов, протекающих в объекте и находит отображение в принципиальных электрических, кинематических, пневматических и т.п. схемах и сопровождающих их документах.

Конструкторский аспект связан с реализацией результатов функционального аспекта, т.е. определением геометрических форм и их взаимным расположением в пространстве (принципиальная электрическая (логическая) схема - функциональный аспект, печатная плата, реализующая эту схему - конструкторский аспект).

Технологический аспект связан с описанием методов и средств изготовления конструкции.

Все перечисленные аспекты тесно взаимосвязаны. И хотя основные идеи разработки определяются функциональным аспектом, удачное конструктивное решение в значительной степени определяет перспективы внедрения разработки. Ещё большее значение, особенно на современном этапе, имеет технологический аспект. Использование новейших технологических достижений может зна-

чительно повлиять как на функциональный, так и на конструкторский аспекты.

Внутри каждого аспекта, возможно, своё специфическое выделение иерархических уровней.

В рамках рассмотренной блочно-иерархической структуры возможны два подхода к проектированию, которые и реализуются на практике.

Если решение задачи высоких иерархических уровней предшествует решению задач более низких уровней, то проектирование называется нисходящим. Если раньше выполняются этапы, связанные с нижними иерархическими уровнями (т.е. проектирование идет от готовых базовых элементов не подлежащих расчленению и разработке), проектирование называют восходящих.

При всех преимуществах блочно-иерархического подхода к проектированию, позволяющему решать задачи большой размерности, основным недостатком как нисходящего, так и восходящего проектирования является то, что на каждом этапе работа ведется с доконца не определёнными объектами, а решения принимаются в обстановки неполной информации. При проектировании сверху вниз, может оказаться, что базовых элементов на самом нижнем уровне не существует в природе, а их разработка весьма проблематична. Проектирование от базовых элементов (снизу вверх) может привести к тому, что подойдя к самому верхнему уровню оказывается, что невозможно удовлетворить техническому заданию на разрабатываемое изделие.

Тем не менее, более разумной альтернативы блочно-иерархическому подходу не существует, поэтому он является основным методом разработки сложных технических систем.

2.3. Жизненный цикл сложной технической системы (ГОСТ 22487-77)

Если рассматривать проектирование в широком смысле слова, то оно носит пространственно-временной характер. Основные этапы проектирования и задачи, решаемые на этих этапах, принято называть маршрутами проектирования, определяющими жизненный цикл сложной технической системы, который включает следующую последовательность этапов и проектных процедур.

1. Предпроектные исследования.
2. Техническое задание.
3. Техническое предложение.
4. Эскизный проект.
5. Технический проект.
6. Рабочий проект.
7. Изготовление опытного образца.
8. Опытная эксплуатация.
9. Изготовление серийного образца.
10. Промышленная эксплуатация.
11. Модернизация

Предпроектные исследования - выполняются заказчиком от момента появления идеи до того, как будут сформулированы конкретные задачи проектирования. Это этап научно-исследовательской работы (НИР) - поиск принципиальных возможностей построения системы, исследование новых принципов, технических средств, обоснование общих решений.

Техническое задание - так же формулирует заказчик. Устанавливается основное назначение изделия, тактико-технические характеристики, показатели качества и техникоэкономические требования, предъявляемые к изделию, определение этапов разработки. При отсутствии стандартов на изделие разработки технических условий (ТУ) (для серийного и массового производства) определяются технические требования, правила приемки, методы контроля (испытаний), транспортировки и хранения, указания по эксплуатации, гарантии поставки и т.п.

Техническое предложение - предэскизное проектирование выполняет разработчик, содержит:

-Технико-экономическое обоснование, определение стоимости разработки.

-Выяснение вариантов возможных решений.

-Проверка на патентную чистоту и конкурентоспособность (если аналогов нет, то обязательна заявка на изобретение, если не конкурентоспособна- не следует разрабатывать).

Эскизный проект - подлежит разрабатыванию, если это предусмотрено техническим заданием. Разрабатывается с целью установления принципиальных (конструкторских, схемных и др.) решений, дающих общее представление о принципах работы изделия и включает следующие работы:

-Выполнение вариантов возможных решений (их конструктивные разработки);

-Предварительное решение вопросов установки и транспортировки изделий;

-Изготовление и испытание макетов с целью проверки принципов работы (натурные испытания) и т.д. и т.п.

Технический проект - включает совокупность конструкторских документов, содержащих окончательные технические решения, дающие полное представление об устройстве изделия и исходных данных для разработки технической документации.

Технический проект может предусматривать разработку вариантов отдельных частей изделия. В этом случае выбор оптимального варианта производится на основе испытания опытных образцов изделия. Разрабатывается если это предусмотрено техническим заданием и состоит:

-разработка конструкторских решений изделия и его составных частей;

-выполнение необходимых расчётов;

-выполнение необходимых схем: принципиальных схем, схем соединения и т.п.;

-анализ конструкции на технологичность;

-разработка, изготовление и испытание макетов;

Рабочий проект (рабочая документация)- совокупность конструкторских документов, необходимых для изготовления и испытания опытного образца. Объем и содержание рабочей документации определяются видом изделия (деталь, сборная единица, ком-

плекс и т.д.) и масштабом производства - опытное, серийное, массовое.

7,8,9,10,11-этапы жизненного цикла СТС, хотя и не являются непосредственными этапами проектирования, тем не менее, тесно с ним связаны. Прохождение этих этапов, как правило, приводит к очередной итерации процедуры проектирования или отдельных её частей.

Так недостатки, выявленные на этапе опытной эксплуатации (8) (что вполне естественно) приводят к необходимости вносить изменение в проект, проводя дополнительные проектные работы.

Промышленная эксплуатация так же имеет временные ограничения по причинам, рассмотренным в главе 1. В связи с этим наступает новый этап проектирования, связанный с модернизацией изделия.

2.4.Формализация процесса проектирования

Введение, автоматизации в процедуру проектирования на основе средств вычислительной техники, т.е. кибернизация проектирования требует точного формального алгоритма этой процедуры. Очень часто процедура проектирования представляется как изготовление чертежно - технической документации, что на самом деле является лишь завершающим этапом этого процесса - вершиной «айсберга». Этому этапу предшествует решение ряда гораздо более сложных задач, которые приходится выполнять проектировщикам.

Независимо от способа проектирования (автоматизированное или неавтоматизированное) и предметной области объектов проектирования, процедура проектирования включает ряд традиционных (обязательных) этапов. На рисунке 2.1. в рамках рассмотренного выше блочно-иерархического подхода, представлена блок-схема проектирования на отдельном уровне иерархии.

После получения технического задания от предшествующего уровня иерархии наступает этап проектирования, который называется синтез технических объектов. Он включает в себя выбор структуры объекта - структурный синтез, создание модели объекта и выбор его параметров (параметрический синтез).

Если среди вариантов структуры изделия ищется не любой приемлемый вариант, а наилучший (в некотором смысле) задачу

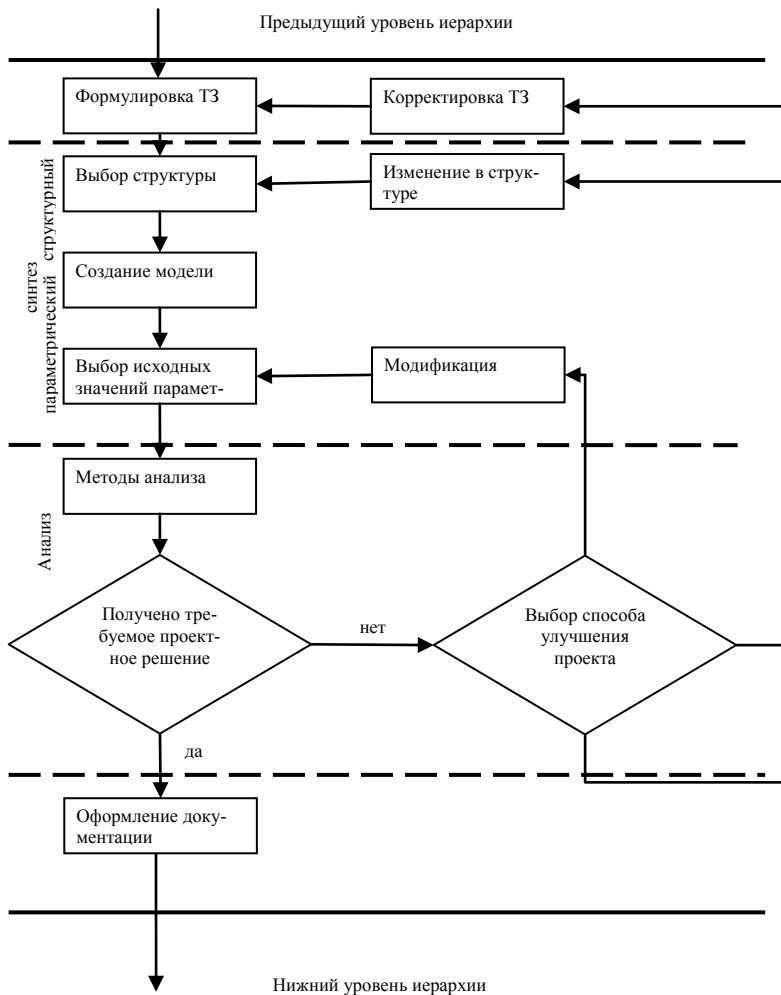


Рис. 2.1. Кибернетическая схема проектирования.

синтеза называют структурной оптимизацией. Выбор параметров структуры, оптимальных с позиции некоторого критерия при за-

данной структуре объекта называют параметрической оптимизацией.

Следующий этап проектирования называется анализ технических объектов. На этом этапе происходит исследование свойств выбранной структуры объекта проектирования (его модели) при различных внутренних параметрах.

Синтез и анализ находятся в диалектическом единстве и не могут существовать друг без друга.

Результаты анализа определяют завершающий этап проектирования - изготовление технической документации.

Если проектное задание не достигнуто, то выбирается способ его достижения. Наиболее простой вариант улучшения проекта - изменение внутренних параметров выбранной структуры объекта (параметрический синтез). Если эти изменения не приводят к желаемому результату, необходимо изменять структуру объекта (структурный синтез). Если и этот способ не приводит к удовлетворению Т.З. необходимо обращаться на верхний уровень иерархии проекта с целью облегчения технического задания.

Завершающий этап проектирования - изготовление - технической документации.

Таким образом, процедура проектирования включает три основных этапа: синтез технических объектов, их анализ и изготовление технической документации. Причем первые две задачи находятся в тесном взаимодействии и в рамках автоматизированного проектирования их основу составляет математическое моделирование.

Синтез технических объектов нацелен на создание новых вариантов объекта проектирования и представляет наиболее сложную математическую задачу, основанную на разделах математики, ориентированных на вычислительные методы требующих очень больших ресурсов вычислительных систем. Наиболее проблемным является структурный синтез. Основу параметрического синтеза составляют методы математического программирования (методы оптимизации), которые интенсивно развиваются на протяжении последних десятилетий.

Анализ технических объектов - это изучение их свойств. Результаты анализа должны ответить на вопрос, какими свойствами обладает исследуемый объект, насколько хорошо он удовлетворяет

предъявляемым требованиям, но непосредственно не содержит в себе рецептов относительно того, что нужно сделать, что бы улучшить объект и выполнить проектное задание. Но несмотря на такой пассивный характер результатов анализа, роль задачи анализа велика, поскольку результаты этого анализа составляют основу реализации задач синтеза. Задачи синтеза связаны с выбором оптимальной стратегии изменения параметров (или структуры) объекта с целью получения результатов, удовлетворяющих техническому заданию (в простейшем случае переборные алгоритмы). Оценка результатов синтеза проводится с помощью методов анализа. Т.е. синтез осуществляется через анализ и невозможен без него. Обычно различают два основных метода анализа технических объектов:

Одновариантный анализ (или детерминированный), который включает следующие разновидности:

-анализ статических состояний (без учёта изменений параметров во времени);

-анализ переходных процессов;

-анализ стационарных видов колебаний;

-анализ частотных характеристик;

-анализ устойчивости.

Многовариантный анализ заключается в исследовании объекта в некоторой окрестности отображаемой точки и он связан с изучением влияний случайных факторов и внешней среды на работу объекта проектирования и включает, как правило, две задачи:

-анализ чувствительности выходных параметров к изменениям внутренней и внешней среды;

-статический анализ, основу которого составляют методы теории вероятностей и математической статистики.

Более подробно методы анализа и синтеза будут рассмотрены в дальнейшем.

Глава 3. Математическое обеспечение САПР (методы многовариантного анализа)

Вопросы математического обеспечения представляют центральную и наиболее сложную часть автоматизации проектирования. Необходимость современных подходов к проектированию потребовало как расширения сферы использования и развития традиционных разделов математики, так и разработку новых, ориентированных на реализацию с помощью средств современной вычислительной техники. Все основные этапы в САПР- анализ, синтез и изготовление технической документации связаны с использованием современных математических методов.

Математическое обеспечение САПР имеет иерархическую структуру и в определённой степени соответствует иерархии задач проектирования сложных технических систем.

Можно выделить два уровня математического обеспечения САПР по аналогии с двумя уровнями (универсальным и специализированным) различных видов обеспечения САПР. Нижний уровень средств математического обеспечения отражает специфику предметной области и используется при анализе разрабатываемых технических систем. Так при расчёте электронных схем основу математического обеспечения определяют законы теории электрических цепей - это законы Ома, Кирхгофа, а так же теория графов и ряд традиционных разделов математики, дифференциальные уравнения, линейная алгебра и т.п. Если говорить о САПР ЭФУ, то здесь основной аппарат управления Максвелла, законы релятивистской механики, разнообразные дифференциальные и интегральные уравнения и т.п. Естественно, что каждый из перечисленных фундаментальных законов физики и математических методов их реализации имеет свои специфические особенности для различных предметных областей, которые и отражаются в особенностях математического обеспечения САПР этих областей. Именно этим особенностям и будет посвящена вторая часть этого учебного пособия – методическое обеспечение САПР ЭФУ.

Однако наряду с вышеуказанным имеется ряд математических методов, которые используются на верхних уровнях иерархии и имеющих достаточно универсальный характер. К таким методам прежде всего относятся методы, применяемые для решения задач

синтеза - различные разделы математического программирования, вычислительная геометрия и геометрическое моделирование и т.п.

При анализе технических объектов достаточно универсальными являются методы многовариантного анализа - статистический анализ (метод Монте-Карло). Этот вид анализа определяется как имитационное моделирование.

Эти методы и будут рассмотрены в настоящей главе, но вначале следует определиться с понятиями модели и моделирования, основополагающими понятиями САПР.

3.1. О моделях и моделировании

Под моделью принято понимать такой материальный или абстрактный объект, который в процессе изучения заменяет объект оригинал, сохраняя важнейшие типичные его черты.

Что касается моделирования, то современные словари дают следующее определение этому термину: «моделирование- исследование каких-либо явлений, процессов или систем путём построения и изучения их моделей, использование этих моделей для определения или уточнения характеристик или рационализации способов вновь конструируемых объектов». А если совсем коротко, то построение модели и исследование объектов на их моделях называется моделированием.

С незапамятных времен при изучении сложных процессов, явлений, конструирования новых сооружений человек применяет модели. Хорошо построенная модель, как правило, доступнее для исследования, нежели реальный объект.

Существует несколько приёмов моделирования, которые можно условно объединить в две большие подгруппы: материальное (предметное) и идеальное (абстрактное) моделирование.

К материальному относятся такие способы моделирования, при которых исследование ведется на основе модели, воспроизводящей основные геометрические, физические, динамические и функциональные характеристики изучаемого объекта.

Основными разновидностями материального моделирования является физическое и аналоговое моделирование.

Физическим моделированием принято называть моделирование, при котором реальному объекту противопоставляется его уве-

личенная или уменьшенная копия, допускающая исследование (как правило, в лабораторных условиях) с помощью последовательного перенесения свойств изучаемых процессов и явлений на объект на основе использования теории размерности физических величин и теории подобия.

Физическое моделирование широко используется в науке и технике как способ разработки и экспериментального изучения на моделях разнообразных механизмов, самолетов (в аэродинамической трубе), судов, тепловых установок и т.п. Этот вид моделирования применяется и при разработке установок в электрофизике. Так, при проектировании 20-ти километрового протонного синхротрона в Протвино, в Радиотехническом институте была создана его уменьшенная (100 метровая) копия, на которой применялись элементы конструкции разрабатываемой установки.

Преимущества физического моделирования перед натурным экспериментом заключается в том, что условия реализации процесса-модели могут значительно отличаться от условий, свойственных оригиналу и их выбирают из удобства и простоты исследования.

Аналоговое моделирование основано на аналогии процессов и явлений, имеющих различную физическую природу, но одинаково описываемых формально (одними и теми же математическими уравнениями). Наиболее простой пример изучение механических колебаний с помощью электрической схемы, описываемой теми же дифференциальными уравнениями. Электрическое моделирование лежит в основе работы машин непрерывного действия - так называемых аналоговых или моделирующих машин (например, дифференциального анализатора, электрического анализатора и др.). Моделью какого-либо процесса при таком «машинном моделировании» можно, очевидно, с равным основанием называть как программу этого процесса, так и саму машину. В 60^х годах прошлого века на кафедре ЭФУ МИФИ для исследования динамики частиц в линейных ускорителях использовалась аналоговая машина γ (гамма). Однако справедливости ради следует отметить, что этот вид моделирования для подобных задач в современных условиях полностью вытеснен цифровым компьютерным моделированием, обладающим гораздо большей точностью и достаточно высоким быстродействием (основное достоинство аналоговой машины). По этим же причинам утратили актуальность и так называемые гиб-

ридные вычислительные системы - объединение аналоговых и цифровых машин.

Необходимо отметить, что в обоих типах материального моделирования (физического и аналогового) модели являлись материальным отражением исходного объекта и были связаны с ними своими геометрическими, физическими и другими характеристиками, причем процесс исследования был тесно связан с материальным воздействием на модель, т.е. состоял в натурном эксперименте с ней. Таким образом, предметное моделирование по своей природе является экспериментальным.

От предметного моделирования принципиально отличается идеальное (абстрактное), которое основано не на материальной, а на идеальной аналогии, мыслимой. Идеальное моделирование носит теоретический характер. Различают два вида идеального моделирования: интуитивное и знаковое.

Под интуитивным моделированием понимается моделирование, основанное на интуитивном представлении об объекте исследования, не поддающимся формализации, либо не нуждающимся в ней. В этом смысле жизненный опыт каждого человека может считаться его индивидуальной моделью окружающего мира. Этот вид моделирования особенно развит у музыкантов, учёных, шахматистов.

Знаковым называют моделирование, использующее в качестве моделей знаковые преобразования какого-либо вида: схемы, графики, чертежи, формулы, наборы символов и т.п., а так же включающее набор знаков, по которым можно оперировать с выбранными знаковыми образованиями и их элементами.

Важнейшим видом знакового моделирования является математическое моделирование, при котором исследование объекта осуществляется посредством модели, сформулированной на языке математики, и использованием тех или иных математических методов. Классическим примером математического моделирования является описание и исследование законов механики И. Ньютона средствами математики.

3.2. Математическое моделирование.

Математическое моделирование есть приближенное описание какого-либо класса явлений внешнего мира с помощью математических моделей, под которой понимается совокупность математических объектов (чисел, символов, переменных, векторов, множеств и т.п.) и отношений между ними, которые адекватно отображают основные свойства проектируемых объектов.

В общем виде математическая модель может быть представлена в виде зависимости (далеко не всегда функциональной):

$$X=F(X,Q)$$

где Y - вектор выходных параметров;

X -вектор входных (внутренних) параметров (параметров элементов);

Q - вектор внешних параметров.

Даже если существует функция $F(X,Q)$, это ещё не означает, что она известна проектировщику. В большинстве случаев связь между внутренними (входными) параметрами и внешними (выходными) параметрами задается в алгоритмической форме, например через численное решение системы уравнений для некоторых параметров, которые характеризуют состояние объекта и не относятся к перечисленным выше параметрам (выходным, внутренним и внешним). Они называются фазовыми переменными (как бы промежуточные параметры). Вектор фазовых переменных задаёт точку в пространстве, называемым фазовым. Фазовые переменные, непосредственно характеризующие запасы энергии в элементах системы именуют переменными состояния. Выходные параметры при таком подходе имеют смысл функционалов, зависимостей фазовых переменных от времени (фазовые переменные возникают только при анализе моделей).

В зависимости от характера отображаемых свойств проектируемого объекта модели можно разделить на структурные (связанные, прежде всего с геометрическим моделированием) и функциональные. Структурные модели отображают только структурные (в частном случае геометрические) свойства объекта. Структурные модели обычно используются в случаях, когда задачи структурного анализа удастся ставить и решать, абстрагируясь от особенностей физического процесса в объекте.

Наибольшую группу представляют так называемые функциональные модели. Они, в основном, ориентированны на методы ана-

лиза. Традиционными и давно разрабатываемыми являются теоретические. Их получают на основе изучения физических закономерностей, структура уравнений и параметры моделей определяют физическим толкованием. Они основаны, как правило, на функциональных физических законах. Так, например, методы математической физики включают те уравнения, которые применяются для построения и изучения математических моделей, описывающих большие классы физических явлений. Теоретические модели по своей структуре разделяют на линейные и нелинейные (в зависимости от линейности или нелинейности входящих в них уравнений), непрерывные или дискретные (в зависимости от непрерывности или конечной счётности их решений), динамические или статические (имеется или нет зависимость от времени).

К сожалению, не всегда условия применимости моделей и методов могут быть найдены и исследованы строгими методами.

Часто применимость того или иного компонента математического обеспечения зависит от конкретных условий, многообразие которых не поддаётся исчерпывающему учёту и классификации.

Особый класс математических функциональных моделей составляют так называемые имитационные, которые иногда называются формальными или даже эвристическими. Развитие этого направления моделирования связано с разработкой сложных технических систем и ограничениями традиционных теоретических моделей. Эти модели получают на основе проявления свойств моделируемого объекта во внешней среде, модель типа «чёрный ящик».

Ниже об этих моделях будет сказано более подробно.

В соответствии с иерархической структурой сложных технических систем (элемент, подсистема, система) математические модели разделяются на три уровня: низшего (микро), среднего (макро) и высшего.

На низшем уровне, называемом так же уровнем проектирования базовых элементов или микроуровне, математические модели должны отражать процессы, протекающие, в общем случае, в трехмерной сплошной среде. Они представляют собой дифференциальные уравнения в частных производных и часто называются распределёнными моделями. Они основаны на известных фундаментальных физических законах (уравнения Ламе для механики упругих сред, Навье-Стокса для гидравлики, уравнения теплопроводности-

термодинамики, уравнения Максвелла- электромагнитные поля). В качестве фазовых переменных в этих уравнениях выступают напряжённости полей и плотности потоков. Точное решение этих уравнений (аналитическое) удаётся получить лишь для ограниченного числа частных случаев, поэтому основная задача, возникающая при моделировании, состоит в построении приближённой дискретной модели. Для этого используются методы конечных разностей, конечных элементов или интегральных граничных элементов. Так как получаемая при дискретизации пространства аппроксимирующая система алгебраических уравнений имеет очень высокий порядок, то при моделировании достаточно сложных технических объектов приходится принимать ряд допущений и переходить к моделированию на среднем (макро) уровне.

Элементами этого уровня являются объекты, которые на микроуровне рассматриваются как системы. Математические модели на данном уровне представляют обыкновенные дифференциальные уравнения, в частных случаях статических задач они превращаются в алгебраические и трансцендентные уравнения. В качестве фазовых переменных выступают токи и напряжения в электрической системе (вместо плотности тока и напряженности полей), в механической системе аналогами этих переменных являются силы и скорости, в гидравлической системе массовые расходы и давление, а в тепловых - тепловые потоки и температура.

Математическую модель системы на этом уровне получают объединением компонентных и топологических уравнений.

Законы функционирования элемента подсистемы задаются компонентными уравнениями, связывающими, как правило, разнообразные фазовые переменные, относящиеся к данному элементу.

Компонентные уравнения могут быть линейными или нелинейными, алгебраическими, обыкновенными дифференциальными уравнениями или интегральными.

Связь между однородными фазовыми переменными, относящимися к различным элементам подсистемы, задается топологическими уравнениями, получаемыми на основании сведений о структуре системы.

В большинстве технических систем можно выделить три типа простейших элементов (компонент):

Элемент типа R- элемент диссипации энергии. На этом элементе, как правило, происходит преобразование энергии в тепловую (электрические сопротивления) трения в механических системах, вязкость в жидкостях и газах и т.п.

Элемент C и L, в которых происходит накопление потенциальной и кинетической энергии.

Для каждой физической подсистемы характерны свои законы, однако для простейших элементов форма выражающих их уравнений оказывается одинаковой.

Так, например, для электрической подсистемы фазовыми переменными являются токи I и напряжения U, а компонентные уравнения для простейших элементов выглядят следующим образом:

Для сопротивления (законы Ома) $I=U/R$;

Для ёмкости $I=C(dU/dt)$, где C- электрическая ёмкость;

Для индуктивности $U=L(dI/dt)$, где L- электрическая индуктивность.

Топологические уравнения в большинстве случаев базируются на уравнениях равновесия и уравнениях непрерывности.

Так, для электрической подсистеме связь между отдельными элементами этой подсистемы устанавливается на основании законов Кирхгоффа.

Уравнения первого закона Кирхгоффа устанавливает равенство нулю суммы токов в узлах системы, т.е. $\sum_{K \in p} I_k = 0$ (уравнения равновесия), где I_k ток в k-ой ветви; p- множество номеров ветвей, инцидентных рассматриваемому узлу.

Из уравнения второго закона Кирхгоффа видно, что сумма падений напряжений на элементах схемы при их обходе по произвольному контуру равна нулю, т.е. $\sum_{j \in q} U_j = 0$ (уравнение непрерывности), где j-номер ветви; U_j - падение напряжения на j-й ветви схемы, входящей в контур; q- множество номеров ветвей, входящих в рассматриваемый контур.

Аналогичные компонентные и топологические уравнения можно привести и для механических, гидравлических (пневматических) и тепловых систем.

Модели высшего уровня или модели технических систем, которые так же называют информационными, применяемые для сложных устройств и комплексов (вычислительные комплексы, радиолокационные станции, ускорительные комплексы и т.п.), функционирование которых рассматривается как цепь событий, происходящих в дискретные моменты времени и заключающихся в изменении состояний элементов.

Для построения математических моделей высшего уровня широко используют математическую логику, теорию массового обслуживания, методы теории автоматического управления, теорию графов и т.п. Очень широкое применение на этом уровне имеет упоминавшееся ранее имитационное моделирование, что представляет собой дословный перевод английского выражения «Simulation modeling». Как легко убедиться, перевод сделан не слишком хорошо, так как в нем содержится тавтология (любое моделирование - есть имитация). Однако этот термин в отечественной практике так широко распространился, что хоть он и неудачен, но маловероятно, что он претерпит изменение.

Суть метода имитационного моделирования состоит в том, что процесс функционирования сложной системы представляется в виде определенного алгоритма, который и реализуется на ЭВМ. По результатам реализации могут быть сделаны те или иные выводы относительно исходного процесса.

Наряду с широко используемыми, традиционными методами математического моделирования (теоретическими) большая роль при проектировании отводится и этому виду, причём его появление, как указывалось выше, связано с возрастающими сложностями объектов, процессов, явлений, ввиду отсутствия в настоящее время математического аппарата, пригодного для исследования. Кроме того, реальные системы подвержены влиянию различных случайных факторов. Учёт этих факторов аналитическим путём представляет весьма большие трудности, зачастую непреодолимые при большом их числе. Этот вид моделирования реализуется на компьютере и отображает внутренние и внешние взаимосвязи моделируемого сложного объекта.

Имитационное моделирование можно трактовать как результат объединения двух подходов: дедуктивного и индуктивного. Дедуктивный подход или подход «сверху вниз» предполагает определение структуры модели на основе знания физических законов (традиционные теоретические модели), которым подчиняются происходящие в системе процессы и построение системы с последующим уточнением параметров. Индуктивный же подход, или подход «снизу вверх» оперирует с экспериментальными данными типа «вход-выход» не используя знания внутренних свойств системы. Типичным примером такого подхода служит экспериментирование с моделью на основе метода Монте-Карло (который рассматривается ниже). При этом имитационное моделирование представляет собой численный эксперимент сродни физическому эксперименту. В ходе эксперимента варьируются экзогенные (внешние) параметры модели, результаты представляются в статистическом смысле, что неизбежно при работе со сложной системой. Основные свойства системы и состояние её параметров определяются на основе длительного наблюдения и набора соответствующих статистических закономерностей.

Поскольку статистическое моделирование является одним из основных приемов имитационного моделирования, носит универсальный характер (его идея очень проста и независима от приложений), широко применяется в технике и физике ниже будет более подробно рассмотрена его методика.

Немного о требованиях, предъявляемых к математическим моделям. Среди общих требований, которые не зависят от типа моделей и вида моделирования, следует отметить следующие:

Точность математической модели - свойство, определяющее степень совпадения предсказанных с помощью модели выходных параметров объекта с истинными значениями этих параметров. Количественная оценка точности затруднена по следующим причинам:

-Реальная модель, как правило, обладает множеством выходных параметров. Поэтому возникает необходимость сведения вектора параметров к скалярному представлению с помощью не всегда достаточно обоснованных методов;

- Модель составляется, как правило, для многократного использования, при анализе различных вариантов объекта. Однако

характер проявления тех или иных свойств объекта зависит от особенностей взаимодействия с внешней средой и другими объектами.

- Истинное значение (которое предполагается заранее известным) обычно отождествляется с экспериментом, однако погрешность экспериментов может оказаться соизмеримой с погрешностями модели, а часто и превышает их.

Экономичность - отождествляется прежде всего затратами машинного времени при реализации на компьютере. Несмотря на постоянный рост современных вычислительных систем, потребность задач (в том числе и проектных) так же возрастает, зачастую превышая возможности машин, в связи с этим разработчикам приходится заботиться об экономичности алгоритмов.

Степень универсальности модели связана с тем, что разработка модели и её адаптация к вычислительной среде достаточно длительный и дорогостоящий процесс, поэтому расширение сферы приложения модели повышает эффективность разработки и снижает стоимость разработанного программного продукта.

Требования точности, универсальности и экономичности модели достаточно противоречивы и требуют принятия оптимальных решений в каждом конкретном случае.

Кроме перечисленных, существует ряд специфических требований к математическим моделям различных типов. Так, например, требования к традиционным теоретическим моделям определены в курсах вычислительной математики (это корректность, устойчивость, сходимости, обусловленность и т.п.) где с ними можно подробно ознакомиться.

3.3 О методах многовариантного анализа.

Методы многовариантного анализа являются в некотором смысле надстройкой над одновариантными моделями, которые отражают специфику конкретных предметных областей. В этом смысле многовариантный анализ достаточно универсален, а его методы практически не зависят от конкретных приложений.

В задачи этого вида анализа входит определение допустимых отклонений внутренних параметров проектируемых объектов от их номинальных значений (полученных с помощью одновариантного анализа) и именующих случайный (стохастический) характер.

Именно такого вида отклонения определяют надежность работы установки и точность поддержания её выходных характеристик. Этот вид анализа предполагает изучение случайных событий и процессов, происходящих в различных системах проектируемых изделий и использование аппарата теорий случайных процессов, математической статистики для оценки их влияния.

При проектировании электрофизических установок и ускорителей указанные отклонения параметров от их номинальных значений могут значительно понизить эффективность использования пучка и даже привести к его потере.

Исследование влияния малых отклонений внутренних (и внешних параметров) на выходные характеристики разрабатываемых изделий предполагает:

- а) Исследование характера (типа) отклонений этих параметров;
- б) Знание статистических характеристик случайных (стохастических) отклонений.

По типу все отклонения могут быть систематические и случайные. Систематические отклонения - постоянные во времени и пространстве. Например снижение напряжения на установке за достаточно длительный промежуток времени. Учёт таких отклонений не представляет больших трудностей и связан с анализом чувствительности входного параметра к изменению внутренних, определяемых в одновариантной модели объекта и в дальнейшем рассматриваться не будет.

Случайные отклонения могут быть пространственными (постоянными во времени) например отклонения в параметрах элементов электронных схем (допуски на сопротивления, конденсато-

ры и т.п.), неточности в изготовлении ячеек ускоряющих волноводов и т.п.

Временные отклонения могут разделяться например на медленные, импульсные (постоянные в пределах одного импульса), внутримпульсные и т.п.

Что же касается статистических характеристик, то при массовом производстве радиодеталей, механической обработке на металлолетунных станках) отклонения, как правило, усечённому (в пределах допуска) закону Гаусса.

Различают следующие методы расчёта допусков:

- 1)Метод наихудшего случая;
- 2)Метод граничных испытаний;
- 3)Метод моментов;
- 4)Метод натуральных испытаний;
- 5)Метод статистических испытаний (метод Монте-Карло).

Ниже будет дана краткая характеристика первых четырёх методов. Метод Монте-Карло, который получил очень широкое распространение в физических исследованиях и техники и не только в аспекте допусков будет рассмотрен более подробно.

1.Метод наихудшего случая характеризуется следующими требованиями: выходной параметр должен находится в пределах установленного поля допусков при наиболее неблагоприятных сочетаниях погрешностей. Существует два способа расчёта погрешностей этим методом:

Первый состоит в том, что погрешность внутренних параметров определяется арифметическим или квадратичным суммированием частичных отклонений, вызванного действием каждого дестабилизирующего фактора в отдельности. По этим суммарным погрешностям определяется погрешность выходного параметра.

При втором способе отдельно определяется частичная погрешность выходного параметра за счёт погрешностей внутренних, вызванных влиянием каждого воздействующего фактора в отдельности. Результирующая погрешность выходного параметра определяется суммированием частичных погрешностей. Например, при ускорении релятивистских электронных сгустков в линейном электронном ускорителе с помощью дифференциального волновода неточность при изготовлении четырёх основных размеров волно-

вода приводит к сдвигу сгустков по фазе ΔQ_i . Этот сдвиг однозначно изменяет набираемую энергию.

В первом случае учёт этих изменений может быть произведён следующим образом:

$$\Delta Q_{\Sigma} = \Delta Q_1 + \Delta Q_2 + \Delta Q_3 + \Delta Q_4$$

$$\text{или } \Delta Q_{\Sigma} = (\Delta Q_1^2 + \Delta Q_2^2 + \Delta Q_3^2 + \Delta Q_4^2)^{1/2}$$

При этом изменения (отклонения) в набираемой энергии оцениваются как

$$\Delta W = f(\Delta Q_{\Sigma})$$

Во втором же случае

$$\Delta W = \sum_{i=1}^4 f(\Delta Q_i)$$

Следует отметить, что оба способа дают одинаковые результаты только при линейной зависимости параметров. Для нелинейных зависимостей при втором способе получаются меньшие погрешности выходного параметра.

Основные недостатки этого метода следующие:

- 1) необоснованно арифметическое и квадратичное суммирование погрешностей параметров. Квадратичное суммирование частичных погрешностей внутренних параметров справедливо только при нормальном законе распределения погрешностей;
- 2) Отсутствует количественная оценка попадания выходного параметра в поле допусков;
- 3) Невозможно оценить случаи появления крайних и средних значений выходного параметра;
- 4) Не позволяет определить причину выхода параметра из поля допуска если на него влияют несколько факторов;

Однако этот метод очень прост и позволяет быстро оценить (хотя и грубо) верхний предел допустимых отклонений внутренних и внешних параметров изделия.

2 Метод граничных испытаний. В этом методе обычно фиксируются параметры ряда элементов. Параметры других элементов варьируются с учётом требований к фиксируемым параметрам.

Из двух варьируемых параметров один, имеющий наибольшее влияние (степень влияния характеризуется частной производной) является основным.

Проводился ряд расчётов, при которых фиксируются ряд значений одного из параметров и в широких пределах проводится изменение основного параметра. С помощью этих расчётов определяются допустимые (граничные) значения изменений этих параметров.

По результатам расчётов строятся графики. По одной из осей всегда откладывается значения основного параметра, по другой – второй изменяемый параметр. Графики представляют область устойчивости работы, внутри которой значения выходного параметра соответствуют требуемым.

По графикам выбирают номинальные значения и допустимые отклонения основного и варьируемого параметров так, чтобы величины выходного параметра находились внутри области устойчивости.

Этот метод используется в практике расчёта ускорителей для определения устойчивости частиц в фазовом пространстве электромагнитных полей.

Достоинство метода состоит в том, что он позволяет определить область устойчивости работы и выбрать номинальные значения внутренних параметров оптимальным образом.

Существенный недостаток метода заключается в том, что он позволяет провести исследование при одновременном изменении только двух элементов. В реальных условиях погрешность выходного параметра является функцией погрешностей всех внутренних параметров. Поэтому значения погрешностей выходного параметра получаются приближёнными.

3. При анализе методом моментов постулируют нормальный закон распределения погрешностей внутренних и выходного параметров. Исходными являются характеристики закона распределения внутренних параметров.

Расчёт точности по методу моментов сводится к определению среднего значения (математического ожидания) определяемого вы-

ходного параметра и его среднего квадратического отклонения или дисперсии

$$\gamma = \varphi(m_{x_1}, m_{x_2}, \dots, m_{x_n}),$$

где m_{x_i} математическое ожидание исходных параметров, γ математическое ожидание выходной функции, а дисперсия

$$D_y = \sum_{i=1}^n (\partial\varphi/\partial x_i)_m * D_{x_i},$$

где $(\partial\varphi/\partial x_i)_m$ производная функции по i -му параметру в точке математического ожидания этого параметра.

Этот метод дает точные результаты только для линейных зависимостей и при нормальном распределении погрешностей. Для нелинейных зависимостей даже при нормальном законе распределения имеют место значительные ошибки.

4. Натурные испытания - этот эмпирический метод оценки параметров. В качестве модели используются несколько образцов. На каждом образце измеряют уровень параметра. Результаты измерений обрабатываются статистически.

Этот метод требует больших затрат и на этапе проектирования неприменим.

3.4 Метод статистических испытаний (Метод Монте-Карло)

Перечисленные выше методы оценки допустимых отклонений внутренних и внешних параметров от их номинальных значений широко используется в практике проектирования и играют важную роль при предварительной (грубой) оценке допусков.

Наиболее эффективным и современным методом является метод Монте-Карло, который получил в последнее время очень широкое распространение благодаря широкому внедрению компьютерных технологий, поскольку его применение требует компьютеров очень высокой производительности. Этот метод имеет гораздо более широкие области применения при проектировании (анализе и синтезе математических моделей) нежели расчёт допусков, о чём будет сказано ниже. Именно с этим методом, прежде всего, связано понятие имитационного моделирования.

Наибольшее распространение метод Монте-Карло получил в исследованиях по физике, в теории массового обслуживания, в теории игр и т.п.

Название метода произошло от города Монте-Карло (княжество Монако), известного своими казино, ибо одним из простейших приборов для генерирования случайных чисел (основы метода Монте-Карло) является рулетка.

Официальной датой рождения методов под общим названием Монте-Карло считают 1949 год, когда появилась статья с названием «метод Монте-Карло» (Metropolis N., Ulam S., the Monte Carlo method, I. Amer. Statistical assoc 1949, 44, №247,335-341.). Создателями этого метода считают американских математиков Дж. Неймана и С. Ульма. В Советском Союзе первые статьи о методе Монте-Карло были опубликованы в 1955-56 годах. Возникновение этого метода связано с именами Дж. Неймана, Г.Кана, Э.Ферми и др. выдающихся учёных. Все они в 40^х-годах работали в Лос Аломосе над Манхэттенским проектом. Разработки и широкое использование этого метода явилось важным инструментом при реализации проекта. Хотя теоретические основы метода Монте-Карло были известны намного раньше, однако до появления вычислительной техники метод не мог считаться универсальным численным методом ввиду трудоёмкости ручного моделирования.

Можно выделить два основных направления применения метода Монте-Карло.

Во-первых, для исследования влияния случайных факторов, естественным образом присутствующих в структуре объекта. Учёт этих факторов в рамках имитационного моделирования имеет очень важное значение при проектировании.

В дальнейшем будет более подробно рассмотрена методика методов Монте-Карло для решения подобных задач.

Во-вторых, этот метод стал активно использоваться для решения детерминированных задач, т.е. задач, модели которых не содержат элемент случайности. В этом проявляется универсальность этого метода. Решение таких задач достигается построением вспомогательных вероятностных моделей, куда в качестве параметров входят подлежащие определению постоянные величины.

В настоящее время созданы вероятностные модели для решения различных математических задач. В их числе: вычисление ин-

тегралов, решение систем линейных алгебраических уравнений, краевых задач для уравнений эллиптического и параболического типов, обращение матриц, нахождение собственных значений и собственных векторов матриц, вычисление кратных интегралов, решение задач поисковой оптимизации и др.

Поскольку основной идеей при решении детерминированных задач методом Монте-Карло является замена детерминированной задачи эквивалентной статистической, то естественно, что при такой замене вместо точного решения задачи получается приближённое, погрешность которого уменьшается с увеличением числа испытаний.

В качестве примера применения метода Монте-Карло для решения детерминированных задач рассмотрим методику вычисления интегралов (площадей, объёмов).

Допустим, требуется вычислить определённый интеграл функции $y=f(x)$

$$J = \int_0^1 f(x) dx$$

Если пределы изменения функции и аргумента произвольные, то путём соответствующей нормировки всегда можно добиться, что бы величина x изменялась в пределах $[0,1]$ (см. рис. 3.1) и функция $f(x)$ так же изменялась в пределах $[0,1]$, при $0 \leq x \leq 1$

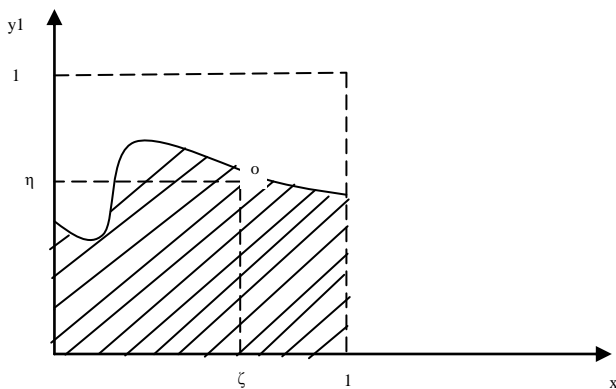


Рис 3.1

Практически, требуется вычислить площадь под кривой, изображённой на рисунке (площадь ограничена кривой $y=f(x)$, осью абсцисс и прямыми $x=0$ и $x=1$).

Для решения этой задачи необходимо иметь два независимых датчика случайных чисел, которые выдают две последовательности случайных чисел x и y , равномерно распределённых в интервале $[0, 1]$. Пусть в квадрат $1, 1$ попала случайная точка o с координатами $[\zeta, \eta]$. Если для этой точки выполняется условие

$$f(\zeta) > \eta,$$

то точка попала под кривую $y=f(x)$.

Проделав этот эксперимент N раз и определив количество точек (L), для которых выполняется это условие, получим

$$L/N \approx P = \int_0^1 f(x) dx$$

Таким образом, задача решена, хотя, в принципе при решении этой задачи можно обойтись одним датчиком случайных чисел.

Однако наибольшее значение метод Монте-Карло приобрёл в САПР при исследовании влияния случайных факторов (временных или пространственных) непременно имеющих место при работе любого изделия. Он является основным методом многовариантного анализа в настоящее время.

Рассмотрим несколько примеров использования этого метода в рамках имитационного моделирования.

1. Определение допусков на номинальные параметры элементов электронных схем, входящих в состав разнообразных приборов (измерителей, генераторов, систем уравнения и т.п.). Предположим, что номинальное значение выходного параметра U требуется поддерживать с точностью не хуже 1% от его номинала. U можно рассчитать, зная параметры их элементов, используя модели теории электрических цепей, о которых говорилось выше.

$$U = f(R_1, R_2, R_3, \dots, R_n, C_1, C_2, C_3, \dots, C_n, \dots \text{ и т.п.})$$

Здесь U может представлять выходной импульс с определёнными техническими заданием параметрами (амплитуда, длительность, требования к фронтам и т.п.).

$R_1, \dots, R_n, C_1, \dots, C_n$ - это сопротивления, конденсаторы и другие электро радиотехнические элементы.

Возникает вопрос, как может повлиять на выходные характеристики импульса отклонения в номинальных значениях параметров элементов схемы и какие, например, нужно брать сопротивления с 5 или 10 процентным разбросом от указанного номинала.

При большом количестве элементов, упомянутые выше методы (наихудшего случая, метод моментов) дают сильно завышенные требования к допускам. Вычислить аналитически распределение U (учитывая, что разброс в серийных элементах схемы, выпускаемых миллионными тиражами, подчиняется нормальному закону) при мало мальски сложной функции $f(x)$ невозможно.

В связи с этим, единственным приемлемым методом для решения этой задачи является метод Монте-Карло. В этом случае поступают следующим образом. Считаем, что элементы, например сопротивления R , случайные величины, имеющие распределения Гаусса с математическим ожиданием M_R , равным номинальному значению, а среднеквадратическое отклонение связано с точностью изготовления каждого параметра (допуском, указанным на рассматриваемый элемент) следующим образом

$$\Delta = 3\sigma$$

Алгоритм расчёта допусков оказывается очень простым: для каждого элемента разыгрывается значение параметра, затем вычисляется значение U_1 . Повторив этот опыт N раз и получив значения U_1, U_2, \dots, U_N , можно построить статистическое распределение (гистограмму) и приближенно рассчитать

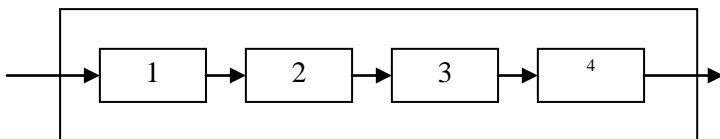
$$MU \approx \frac{1}{N} \sum_{j=1}^N U_j \quad \text{математическое ожидание}$$

$$DU \approx \frac{1}{N-1} * \left[\sum_{j=1}^N U_j^2 - \frac{1}{N} \left(\sum_{j=1}^N U_j \right)^2 \right] \quad \text{дисперсия}$$

2. В качестве ещё одного очень важного примера Монте-Карло рассмотрим расчёт надёжности работы изделия.

Пусть мы хотим оценить время безотказной работы изделия, предполагая, что известны характеристики безотказной работы каждого элемента.

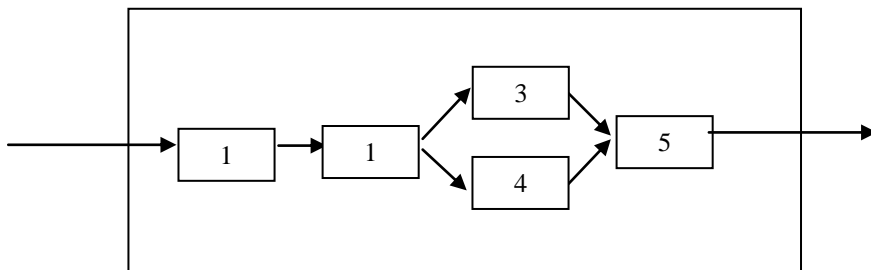
Если считать, что время безотказной работы каждого элемента t_n фиксированная величина, то вычислить время безотказной работы t изделия не составляет труда. Например, для изделия, схематически изображённого на следующем рисунке,



в котором выход из строя одного элемента влечёт за собой выход из строя всего изделия.

$$t = \min(t_1, t_2, t_3, t_4)$$

А для изделия, схематически изображённого на следующем рисунке



в котором один из элементов дублирован

$$t = \min[t_1; t_2; \max(t_3, t_4); t_5]$$

так как если, например, элемент №3 выйдет из строя, то изделие будет продолжать работать на элементе №4.

В действительности, время безотказной работы любого элемента представляет случайную величину Q_k .

Когда мы говорим, что срок службы электрической лампочки 1000 часов, то это лишь среднее значение M_Q величины Q .

Всем известно, что одна лампочка может перегореть быстрее, а другая (такая же) горит дольше.

Если известна плотность распределения вероятности Q_k для каждого из элементов изделия, то можно значительно точнее решить эту задачу используя метод Монте-Карло.

В самом деле для каждого элемента можно разыграть значение случайной величины θ_k , пусть это будет t_k , затем по соответствующей из вышеперечисленных формул можно вычислить t случайной величины Q . Повторив этот опыт многократно (N раз) можно считать, что

$$M_Q \approx \frac{1}{N} \sum_{j=1}^N t_j$$

где t_j значение t , полученное в j -ом опыте.

Необходимо отметить, что вопрос о распределении времени жизни Q_k для отдельных элементов не так уж прост: для наиболее долговечных элементов организация эксперимента (по нахождению закона распределения) затруднительна, так как нужно дожидаться пока выйдет из строя достаточно много элементов.

Наибольшее распространение метод Монте-Карло получил в ядерно-физических исследованиях. Достаточно напомнить, что критическая масса урана в США была определена путём статистического моделирования на электронной вычислительной машине «Эниак».

Вероятностные законы взаимодействия отдельной элементарной частицы (нейтрона, фотона, мезона и др.) с веществом известны. Обычно требуется найти макроскопические характеристики процессов, в которых участвует огромное количество таких частиц: плотности, потоки и т.п. Пожалуй чаще всего метод Монте-Карла используется в нейтронной физике. В качестве примера рассмотрим простейший вариант задачи о прохождении нейтронов сквозь металлическую пластинку.

Пусть на однородную бесконечную пластинку толщиной $h \geq x \geq 0$ падает поток нейтронов с энергией E_0 . Угол падения 90° . При столкновении с атомами вещества, из которого состоит пластинка, нейтроны могут упруго рассеиваться или поглощаться. Предположим, для простоты, что энергия нейтрона при рассеива-

нии не меняется и любое направление «отскока» нейтрона от атома одинаково вероятно. На рисунке 3.2 изображены различные судьбы нейтронов: нейтрон проходит сквозь пластинку (а), нейтрон поглощается (б), нейтрон отражается пластинкой (в).

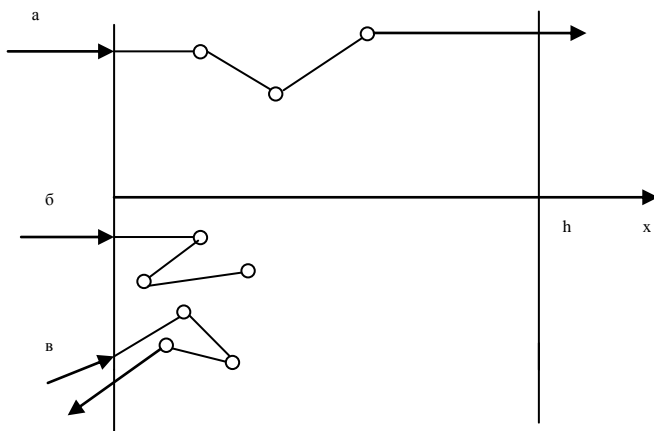


Рис 3.2

Требуется вычислить вероятность прохождения нейтрона сквозь пластинку p^+ , вероятность отражения нейтрона пластинкой p^- и вероятность поглощения нейтрона в пластинке p^0 .

Взаимодействие нейтронов с веществом характеризуется в рассматриваемом случае двумя постоянными $\sum_c + \sum_s$, которые называются сечением поглощения и сечением рассеяния. Индексы с и s- это первые буквы английских слов capture (захват) и scattering (рассеивание).

Сумма этих сечений называется полным сечением

$$\sum = \sum_c + \sum_s$$

Физический смысл этих сечений следующий: при столкновении нейтрона с атомом вещества вероятность поглощения равна \sum_c / \sum , а вероятность рассеивания \sum_s / \sum .

Длина свободного пробега нейтрона λ (то есть длина пути от столкновения до столкновения) - это случайная величина. Она может принимать любые положительные значения с экспоненциальной плотностью распределения вероятностей

$$p(x) = \sum e^{-\Sigma x}$$

Средняя длина свободного пробега (математическое ожидание) []

$$M[\lambda] = 1/\Sigma$$

а формула для λ :

$$\lambda = \frac{1}{\Sigma} \ln \gamma$$

Остается выяснить как выбирать случайное направление нейтрона после рассеивания. Так как задача симметрична относительно оси x , то направление вполне определяется одним углом φ между направлением скорости нейтрона и осью x . Как показано в [] требование равной вероятности любого направления равносильно требованию, что бы косинус этого угла $\mu = \cos(\varphi)$ был равномерно распределен в интервале $(-1, +1)$, а формула для разыгрывания может быть получена, используя преобразования, полученные в []

$$\mu = 2\gamma - 1,$$

которая и применяется для определения угла φ .

Перейдем непосредственно к построению схемы моделирования. Обозначим абсциссы двух последовательных столкновений нейтрона с атомами через x_k и x_{k-1} разыграем длину свободного пробега k -го столкновения по формуле:

$$\lambda_k = -(1/\Sigma) \ln \gamma$$

и вычислим абсциссу следующего столкновения

$$x_{k+1} = x_k + \lambda_k * \mu_k,$$

где $\mu_k = \cos(\varphi_k)$ – направление движения нейтрона после k -го столкновения. Далее проверим условие прохождения нейтрона сквозь пластину

$$x_{k+1} > h$$

При удовлетворении этого условия расчёт траектории заканчивается и добавляется единица к счётчику прошедших частиц. Если это условие не выполняется, то следует разыграть $k+1$ столкновений. Для этого выбираем очередное значение γ и проверяем условие поглощения:

$$\gamma < \frac{\sum p}{\Sigma}$$

при этом расчёт траектории прекращается и добавляется единица к счётчику поглощённых частиц. В противном случае электрон испытал рассеяния в точке x_{k+1} и разыгрывается новое направление скорости движения нейтрона по формуле:

$$\mu_{k+1} = 2\gamma - 1$$

по существу для k -го шага требуется три значения γ . Начальные значения всех траекторий можно выбрать следующие

$$x_0 = 0 \quad \mu_0 = 1$$

В результате N кратного расчёта траектории определяются числа нейтронов N^+ прошедших сквозь пластинку, N^- отразившихся от неё, N^0 - поглощенных веществом пластинку. Соответственно вероятности определяются по формулам:

$$P^+ \approx \frac{N^+}{N}$$

$$P^- \approx \frac{N^-}{N}$$

$$P^0 = \frac{N^0}{N}$$

Блок схема программы представлена на рисунке 3.3, где j - номер траектории ($j=1, \dots, N$), k - номер столкновения. На блок схеме при проверке различных условий имеются два выхода обозначенные 1, если условия выполняются и 0, если условия не выполняются

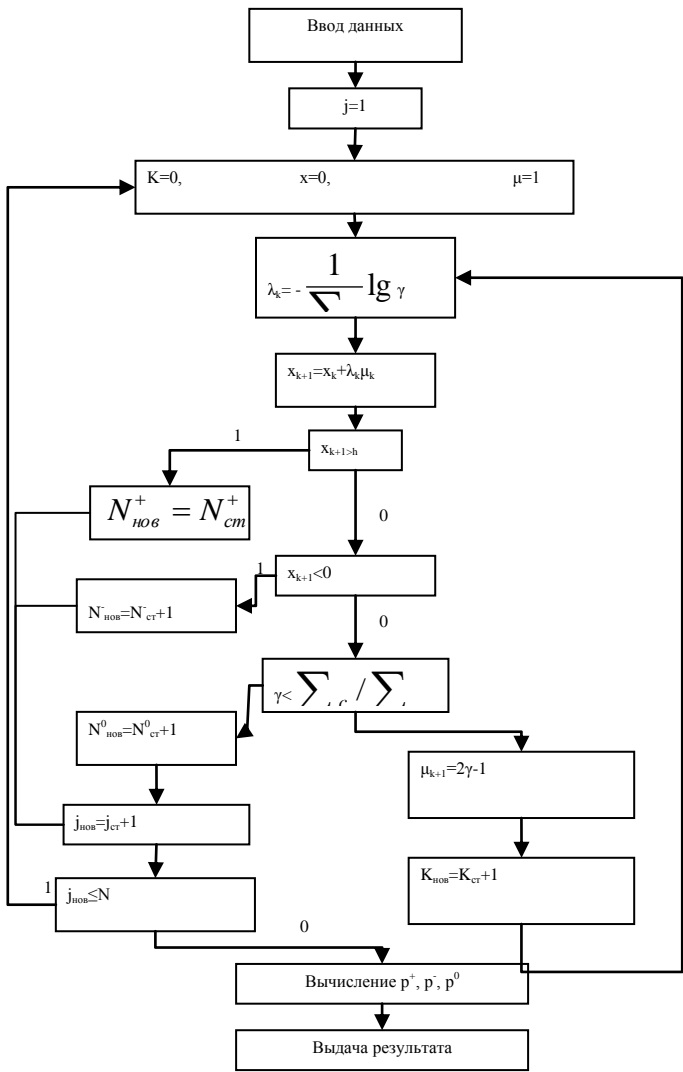


Рисунок 3.3.

В приведенных примерах метод Монте-Карло применяется в задачах самой разной направленности, однако во всех этих задачах можно выделить общие характерные черты моделирования, единые

для любого приложения, связанного с имитационным моделированием.

Общая структура моделирования методом Монте-Карло чрезвычайно проста и может быть представлена в виде следующих схем блоков (рис. 3.4.)

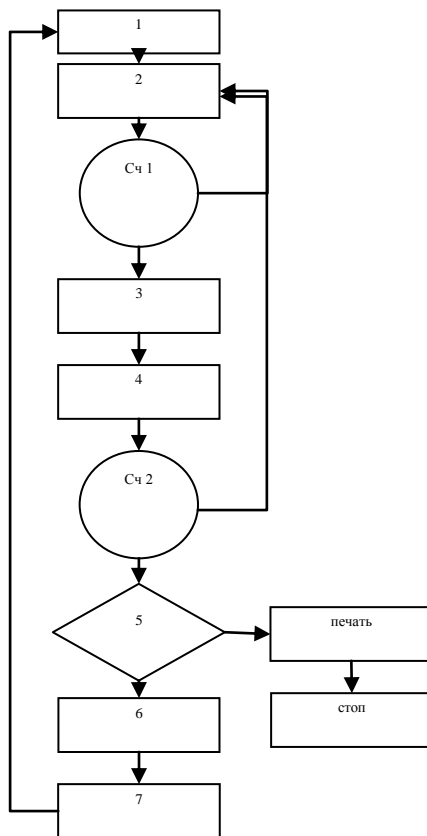


Рис 3.4 Блок схема алгоритма метода Монте-Карло

Блок 1-ввод исходных данных - номинальных значений внутренних и внешних параметров, дисперсии случайных величин (законов распределений) числа испытаний.

Блок 2-генератор случайных (псевдослучайных, квазислучайных) чисел.

Сч1-счётчик, определяющий количество случайных параметров, необходимых при моделировании.

Блок 3-расчёт стационарной модели конкретного приложения (аналитика, системы линейных дифференциальных уравнений, уравнения в частных производных и т.п.)

Блок 4- блок текущего построения гистограммы (статистического распределения)\

Сч2- счётчик общего количества испытаний (задаваемого в блоке 1).

Блок 5- блок сравнения, насколько удовлетворительно знание общего количества испытаний с точки зрения точности получаемых результатов, а так же насколько удовлетворительны заданные требования к внутренним и внешним параметрам системы.

Блок 6-определение параметров, для которых заданные требования к допустимым отклонениям их от номинальных значений нуждаются в коррекции.

Блок 7 формирует команду на изменение входных данных (блок1)- изменение количества испытаний, изменение дисперсии распределения параметра, обладающего наибольшим весом в данный момент и т.п.

Однако не все так просто, реализация рассмотренного алгоритма предполагает решение ряда вероятностно-статистических задач.

1. Получение последовательности случайных чисел. Успех и точность статистического моделирования зависят в основном от двух факторов: качества последовательности случайных чисел и выбора оптимального алгоритма моделирования.

Исходным материалом для формирования случайных последовательностей чисел с различными законами распределения служат случайные числа x_i , имеющие равномерное распределение в интервале $[0,1]$. Таким образом, задача получения случайных чисел обычно разбивается на две. Вначале получают последовательность

случайных чисел, имеющих равномерное распределение. Затем одним из способов преобразования получают из последовательности равномерно распределенных чисел последовательность с нужным распределением.

Другими словами, случайные числа x_i , как возможные значения случайной величины ζ , имеющей равномерное распределение в интервале $[0,1]$ могут быть преобразованы в возможные значения y_i случайной величины η , закон распределения которой требуется.

Существуют два основных пути такого преобразования случайных чисел. Один из них может быть назван прямым. Он состоит в реализации некоторой операции над числом x_i , формирующей число y_i , имеющий требуемый закон распределения. Другой основывается на моделировании условий, соответствующей предельной теоремы теории вероятностей.

Рассмотрим сущность первого пути. Идея построения интересующей нас операции вытекает из следующей теоремы теории вероятностей [*]: Если случайная величина η имеет плотность распределения $f(y)$, то распределение случайной величины

$$\zeta = \int_0^{\eta} f(y) dy$$

является равномерным в интервале $[0,1]$.

На основании этой теоремы можно прийти к следующему правилу. Что бы получить число, принадлежащее совокупности случайных чисел y_i , имеющей функцию плотности распределения $f(y)$ необходимо разрешить относительно y_i уравнение

$$\int_{-\infty}^{y_i} f(y) dy = x_i$$

Варианты реализации этого уравнения представлены в [**].

Геометрическая интерпретация преобразования и графический способ получения случайных чисел с любым законом распределения можно пояснить следующим образом.

На рис 3.4.5 $F(x)$ необходимая нам интегральная функция распределения случайной величины X . Если y является функцией X , то есть $y=F(X)$, то $0 \leq y \leq 1$. Поэтому для получения последовательности случайных чисел, имеющих данную (интегральную)

функцию распределения $F(X)$ необходимо каждое число y с выхода датчика случайных чисел, который формирует числа с равномерным законом распределения в интервале $[0,1]$ подать на нелинейное устройство (аналоговое или цифровое), в котором реализуется функция, обратная $F(x)$, т.е.

$$x=F^{-1}(y)$$

Реализуемая таким образом величина x , будет иметь функцию распределения $F(x)$. Эта процедура может быть использована для графического способа получения случайных величин.

Так, для получения случайной величины x (рис 3.5.) берется число y из последовательности равномерного распределения в интервале $[0,1]$, которое откладывается на оси y , а на оси x считывается соответствующее число x_0 . Повторив неоднократно эту процедуру получим набор случайных чисел, имеющих закон распределения $F(x)$

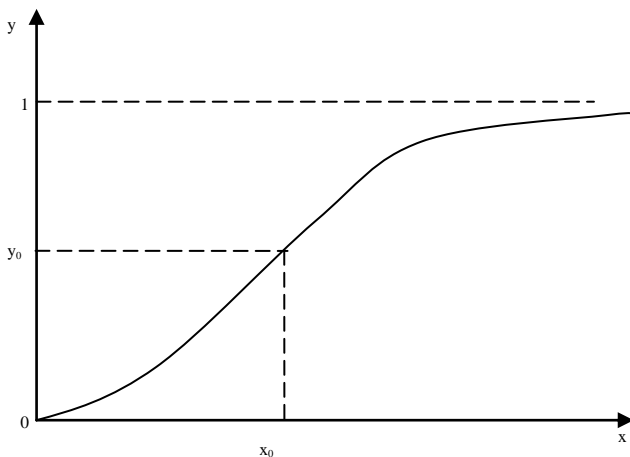


Рис 3.5

Таким образом, получение случайных чисел различных распределений связано с необходимостью получения случайных равномерно распределенных чисел в интервале $[0,1]$. Без преувеличения можно сказать, что наличие простых и экономных способов образования последовательности случайных чисел в интервале $[0,1]$ во многом определяет возможность практического использования метода Монте-Карло.

Напомним основные свойства равномерного распределения. Непрерывная случайная величина η имеет равномерные распределения в интервале $[a, b]$, если её функция плотности равна

$$f(y) = \begin{cases} \frac{1}{b-a} & \text{при } a \leq y \leq b \\ 0 & \text{Вне этого интервала} \end{cases} \quad (3.4.1)$$

Функция распределения случайной величины η имеет вид

$$F(y) = \begin{cases} 0, & y < a \\ \frac{y-a}{b-a}, & a \leq y \leq b \\ 1, & y > b \end{cases} \quad (3.4.2)$$

Математическое ожидание $M[\eta]$ и среднее квадратичное отклонение σ_η соответственно равны

$$M[\eta] = \frac{a+b}{2}, \quad \sigma_\eta = \frac{b-a}{2\sqrt{3}} \quad (3.4.3)$$

В частном случае равномерного распределения в отрезке $[0, 1]$ случайная величина η^* имеет функцию плотности

$$f(y) = \begin{cases} 1 & \text{При } 0 \leq y \leq 1 \\ 0 & \text{Вне этого интервала} \end{cases}$$

Функцию распределения

$$F(y) = \begin{cases} 0 & , y < 0 \\ y & , 0 \leq y \leq 1 \\ 1 & , y > 1 \end{cases} \quad (3.4.4)$$

а математическое ожидание и среднее квадратичное отклонение соответственно равны

$$M[\eta^*] = \frac{1}{2}, \quad \sigma_{\eta}^* = \frac{1}{2\sqrt{3}} \quad (3.4.5)$$

Получение последовательности случайных чисел, равномерно распределенных в интервале [0,1].

Исторически сложилось два способа получения случайных чисел, равномерно распределенных в интервале [0,1].

1. Таблица случайных чисел, которая практически вышла из употребления в настоящее время.
2. Датчики случайных чисел, которые в свою очередь подразделяются на:
 - а) Датчики, использующие физический способ генерирования (датчики истинно случайных чисел);
 - б) Датчики, использующие алгоритмический способ получения, датчики так называемых псевдо-случайных чисел.

При физическом способе получения случайных чисел один из методов состоит в формировании для компьютера дискретной случайной величины, которая может принимать только два значения 0 или 1 с вероятностями

$$P(z_i=0) = 1/2$$

$$P(z_i=1) = 1/2$$

Далее если рассмотреть бесконечную последовательность z_1, z_2, z_3, \dots как значения числа ζ^* вида

$$\zeta^* = z_1 * 2^{-1} + z_2 * 2^{-2} + \dots + z_n * 2^{-n} + \dots,$$

то можно доказать, что случайная величина ζ^* заключённая в интервале $[0,1]$ имеет равномерный закон распределения [*]. Отсюда вытекает способ формирования равномерно распределенной случайной величины. Нужно взять бесконечную последовательность независимых случайных величин z_i и считать их двоичными знаками некоторого числа ζ .

Строго говоря, на цифровой вычислительной машине получить последовательность возможных значений физической величины с равномерным распределением не представляется возможным в силу ограниченного количества двоичных разрядов машинного слова или, другими словами, членов ряда числа ζ^* .

Предположим, что в выбранном нами компьютере числа представляются k двоичными разрядами. Тогда количество несопадающих между собой чисел, каждое из которых можно записать в k разрядную ячейку компьютера, равно 2^k . Поэтому приходится вместо непрерывной совокупности случайных чисел с равномерным распределением в качестве исходной использовать дискретную совокупность 2^k чисел с одинаковыми вероятностями появления любого из них.

Такое распределение иногда называют квазиравномерным, а генераторы, использующие физические способы генерирования, называют генераторами истинно случайных чисел. Однако следует отметить, что расхождения в параметрах распределений (равномерного и квазиравномерного) даже на $16^{\text{ти}}$ разрядной сетке сравнительно небольшое.

Если учесть, что оценить две основные характеристики равномерного распределения математического ожидания (M) и дисперсию (σ) можно используя следующее соотношение [*]:

$$M = \frac{1}{2} * \left(1 - \frac{1}{k}\right), \sigma = \frac{1}{2\sqrt{3}} * \sqrt{1 - \frac{1}{k}}$$

То можно получить следующую таблицу, в которой приводятся измерения значения дисперсии квазиравномерного распределения σ_ζ в зависимости от разрядности сетки и относительная ошибка этого параметра (по отношению к точному значению σ)

К	2	3	5	10	15
σ_z	0,3727	0,3274	0,2979	0,2889	0,2887
σ_z/σ	1,290	1,140	1,030	1,001	1,000

То есть уже на 15-ти разрядах ошибка в оценке дисперсии наблюдается в пятом знаке. Ошибка оценки математического ожидания и того меньше.

На первых вычислительных машинах в качестве генераторов истинно случайных чисел применялись специальные приставки, наиболее часто использующие либо радиоактивные источники, в которых за равные интервалы времени регистрировалось количество испускаемых частиц (чётное $z=0$, нечётное $z=1$), либо шумы электронных ламп (белый шум), в которых за равные промежутки времени регистрировались превышения колебаний напряжения на аноде над номинальных значением и так же чётные превышения оценивались как $z=0$, а нечётные как $z=1$. В современных микропроцессорах «привязка» производится к внутренним процессам кристалла. Так в процессорах x86 сообщалось, что уже у МП Pentium II появился генератор истинно случайных чисел. Основным недостатком генераторов истинно случайных чисел считается невозможность воспроизведения одинаковых последовательностей чисел, т.е. повторяемости одной и той же последовательности, что бывает очень важным свойством генератора в процессе проектирования. Второй способ получения случайных последовательностей связан с генерацией так называемых псевдослучайных чисел непосредственно на вычислительной машине с помощью специальных алгоритмов.

Первый алгоритм для получения псевдослучайных чисел был предложен Дж. Нейманом. Он назывался методом средннего квадрата. Алгоритм этого метода чрезвычайно прост. Берётся четырехзначное число, возводится в квадрат, выбираются 4 цифры из середины возведенного числа и эта процедура повторяется многократно. В качестве случайных чисел используются полученные числа, умноженные на 10^{-4} , т.е. получаются числа в интервале $[0,1]$.

Но этот алгоритм не оправдал себя, поскольку алгоритм выдавал неравномерно большее число малых значений в последовательности.

В настоящее время разработано большое количество генераторов псевдослучайных чисел. Практически во всех наиболее распространённых языках программирования наряду с библиотеками элементарных функций включаются генераторы псевдослучайных чисел.

Однако, наряду с очевидными достоинствами этих генераторов - возможность многократного воспроизведения одних и тех же последовательностей чисел, простота алгоритмов и способа их реализации, основным, и весьма серьёзным недостатком такого способа получения случайных чисел является ограниченность «запаса» псевдослучайных чисел. Имеется в виду повторяемость или цикличность (периодичность) в последовательности случайных чисел, которая наступает значительно раньше, чем генераторы истинно случайных чисел заполняет разрядную сетку компьютера. Точные, аналитические методы оценки периодичности, как правило, отсутствуют, тем более что большинство генераторов используют эвристические алгоритмы. В связи с этим большое значение приобретают методы экспериментальной проверки качества этих алгоритмов. Хотя проверять качество генераторов истинно случайных чисел так же целесообразно из-за возможных сбоев в аппаратуре. Этот вопрос будет рассмотрен в дальнейшем. Предварительно необходимо коснуться методов статистической обработки получаемых случайных последовательностей, которая предваряет проверку их качества.

Понятие статистического ряда и гистограммы.

Наиболее полное представление о последовательности случайных чисел дают их плотности распределения. Описание статистических последовательностей представляется статистическим рядом, а графически в виде гистограммы.

Предположим, что в нашем распоряжении результаты генерации случайной величины X . Разделим весь диапазон наблюдаемых значений X на интервалы (или разряды) и подсчитаем количество значений m_i , приходящее на каждый интервал (разряд). Это число разделим на общее число генерируемых чисел n и найдем частоту (вероятность) соответствующему данному разряду:

$$P_i^* = \frac{m_i}{n}$$

Сумма частот должна быть равна единице, а их последовательность представляется в виде таблицы, называемой статистическим рядом

i	X ₁ , X ₂	X ₂ , X ₃	X _i , X _{i+1}	X _k , X _{k+1}
P _i	P ₁	P ₂		P _i [*]		P _k [*]

Если случайное число попадает на границу интервала, то можно добавлять в каждый интервал по 1/2.

Число разрядов, из которых следует группировать статистический ряд не должно быть слишком большим (при этом проявляются незакономерные колебания), при слишком малом числе разрядов распределения оцениваются слишком грубо.

Практика показывает, что в большинстве случаев рационально выбирать число разрядов 10-20. Размер Случайной последовательности (количество случайных чисел) следует подбирать таким образом, что бы в каждый подинтервал попадало не менее 10 чисел.

Статистический ряд оформляется графически в виде так называемой гистограммы (иногда в виде спектра или полигона) рис . По оси абсцисс откладываются разряды, а по каждому из разрядов, как на основании строится прямоугольник, площадь которого равна частоте попадания случайной величины в данный разряд. Для построения гистограммы необходимо частоту каждого разряда разделить на его длину, а полученное число взять в качестве высоты прямоугольника. Из способа построения гистограммы следует, что полная площадь её равна единицы.

Гистограмма является основным объектом оценки качества генераторов случайных (или псевдослучайных) последовательностей чисел, при использовании критериев согласия.

Критерии согласия

Вопрос о качестве того или иного способа генерации случайных последовательностей равномерно распределенных чисел является общим из основных (если не основным) при проведении статистического моделирования.

Естественно возникает вопрос: объясняются ли естественные расхождения между теоретическим распределением и полу-

ченной (статистической) гистограммой случайными обстоятельствами, связанными с ограниченным числом испытаний или они являются существенными и связаны с низким качеством способа генерации. Для ответа на этот вопрос служат так называемые «критерии согласия».

Идея применения критериев согласия заключается в следующем:

На основании данного статистического материала нам предстоит проверить гипотезу H , состоящую в том, что случайная величина X подчиняется некоторому определенному закону распределения. Этот закон может быть задан в той или иной форме: например, в виде функции распределения $F(x)$ или в виде плотности распределения $f(x)$, или же в виде совокупности вероятностей p_i , где p_i — вероятность того, что величина X попадет в пределы i -го разряда.

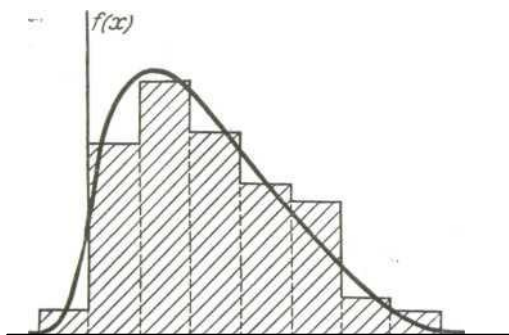


Рис.3.3.1

Так как из этих форм функция распределения $F(x)$ является наиболее общей и определяет собой любую другую, будем формулировать гипотезу H , как состоящую в том, что величина X имеет функцию распределения $F(x)$.

Для того чтобы принять или опровергнуть гипотезу H , рассмотрим некоторую величину U , характеризующую степень расхождения теоретического и статистического распределений. Величина U может быть выбрана различными способами; например, в качестве U можно взять сумму квадратов отклонений теоретических вероятностей P_i от соответствующих частот p_i или же сумму тех же квадратов с некоторыми коэффициентами («весеами»), или же максимальное отклонение статистической функции

распределения $F^*(x)$ от теоретической $F(x)$ и т. д. Допустим, что величина U выбрана тем или иным способом. Очевидно, это есть некоторая *случайная величина*. Закон распределения этой случайной величины зависит от закона распределения случайной величины X , над которой производились опыты, и от числа опытов n . Если гипотеза H верна, то закон распределения величины U определяется законом распределения величины X (функцией $F(x)$) и числом n .

Допустим, что этот закон распределения нам известен. В результате данной серии опытов обнаружено, что выбранная нами мера расхождения U приняла некоторое значение u . Спрашивается, можно ли объяснить это случайными причинами или же это расхождение слишком велико и указывает на наличие существенной разницы между теоретическим и статистическим распределениями и, следовательно, на непригодность гипотезы H ? Для ответа на этот вопрос предположим, что гипотеза H верна, и вычислим в этом предположении вероятность того, что за счет случайных причин, связанных с недостаточным объемом опытного материала, мера расхождения U окажется не меньше, чем наблюдаемое нами в опыте значение u , т. е. вычислим вероятность события:

$$U \geq u.$$

Если эта вероятность весьма мала, то гипотезу H следует отвергнуть как мало правдоподобную; если же эта вероятность значительна, следует признать, что экспериментальные данные не противоречат гипотезе H .

Возникает вопрос о том, каким же способом следует выбирать меру расхождения U ? Оказывается, что при некоторых способах ее выбора закон распределения величины U обладает весьма простыми свойствами и при достаточно большом n практически не зависит от функции $F(x)$. Именно такими мерами расхождения и пользуются в математической статистике в качестве критериев согласия.

Рассмотрим один из наиболее часто применяемых критериев согласия— так называемый «критерий χ^2 » Пирсона.

Предположим, что произведено n независимых опытов, в каждом из которых случайная величина X приняла определенное значение. Результаты опытов сведены в k разрядов и оформлены в виде статистического ряда:

I_i	$X_1; X_2$	$X_2; X_3$	$X_k; X_{k+1}$
p_i^*	p_1^*	p_2^*	p_k^*

Требуется проверить, согласуются ли экспериментальные данные с гипотезой о том, что случайная величина X имеет данный закон распределения (заданный функцией распределения $F(x)$ или плотностью $f(x)$). Назовем этот закон распределения «теоретическим».

Зная теоретический закон распределения, можно найти теоретические вероятности попадания случайной величины в каждый из разрядов:

$$p_1, p_2, p_3, \dots, p_k.$$

Проверяя согласованность теоретического и статистического распределений, мы будем исходить из расхождений между теоретическими вероятностями p_i и наблюдаемыми частотами p_i^* . Естественно выбрать в качестве меры расхождения между теоретическим и статистическим распределениями сумму квадратов отклонений $(p_i^* - p_i)$, взятых с некоторыми «весами» c_i .

$$U = \sum_{i=1}^k c_i (p_i^* - p_i)^2$$

Коэффициенты c_i («веса» разрядов) вводятся потому, что в общем случае отклонения, относящиеся к различным разрядам, нельзя считать равноправными по значимости. Действительно, одно и то же по абсолютной величине отклонение $p_i^* - p_i$ может быть мало значительным, если сама вероятность p_i велика, и очень заметным, если она мала. Поэтому естественно «веса» c_i взять обратно пропорциональными вероятностям разрядов p_i .

Далее возникает вопрос о том, как выбрать коэффициент пропорциональности.

К. Пирсон показал, что если положить

$$c_i = \frac{n}{p_i}$$

то при больших n закон распределения величины U обладает весьма простыми свойствами: он практически не зависит от функции распределения $F(x)$ и от числа опытов n , а зависит только от числа разрядов k , а именно, этот закон при увеличении n приближается к так называемому “распределению χ^2 ”.

При таком выборе коэффициентов c_i мера расхождения обычно обозначается χ^2 :

$$\chi^2 = n \sum_{i=1}^k \frac{(p_i^* - p_i)^2}{p_i}$$

Для удобства вычислений (чтобы не иметь дела с дробными величинами с большим числом нулей) можно ввести n под знак суммы

и, учитывая, что $p_i^* = \frac{m_i}{n}$, где m_i -число значений в i -м разряде,

привести формулу (7.6.3.) к виду:

$$U = \chi^2 = \sum_{i=1}^k \frac{(m_i - np_i)^2}{np_i}$$

Распределение χ^2 зависит от параметра r , называемого числом “степеней свободы” r равно числу

- 1) * Распределением χ^2 с r степенями свободы называется распределение суммы квадратов r независимых случайных величин, каждая из которых подчинена нормальному закону с математическим ожиданием, равным нулю, и дисперсией, равной единице. Это распределение характеризуется плотностью

$$k_r(u) = \begin{cases} \frac{1}{2^{\frac{r}{2}} \Gamma(\frac{r}{2})} u^{\frac{r}{2}-1} e^{-\frac{u}{2}} & \text{при } u > 0 \\ 0 & \text{при } u < 0 \end{cases}$$

Где $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$ -неизвестная гамма-функция

разрядов k минус число независимых условий («связей»), наложенных на частоты p_i^* .

Примерами таких условия могут быть

$$\sum_{i=1}^k p_i^* = 1,$$

если мы требуем только того, чтобы сумма частот была равна единице (это требование накладывается во всех случаях);

$$\sum_{i=1}^k \tilde{x}_i p_i^* = m_x,$$

Если мы подбираем теоретическое распределение с тем условием, чтобы совпадали теоретическое и статическое средние значения;

$$\sum_{i=1}^k (\tilde{x}_i - m_x^*)^2 p_i^* = D_x,$$

если мы требуем, кроме того, совпадения теоретической и статистической дисперсий и т.д.

Для распределения χ^2 составлены специальные таблицы (см. табл 4 приложения). Пользуясь этими таблицами, можно для каждого значения χ^2 и числа степеней свободы r найти вероятность p того, что величина, распределенная по закону χ^2 , превзойдет это значение. В табл. 4 входами являются: значение вероятности p и чисел степеней свободы r . Числа, стоящие в таблице, представляют собой соответствующие значения χ^2 .

Распределение χ^2 дает возможность оценить степень согласованности теоретического и статистического распределений. Будем исходить из того, что величина X действительно распределена по закону $F(x)$. Тогда вероятность p , определенная по таблице, есть вероятность того, что за счет чисто случайных причин мера расхождения теоретического и статистического распределений (7.6.4) будет не меньше, чем фактически наблюдаемое в данной серии опытов значение χ^2 . Если эта вероятность p весьма мала (настолько мала, что событие с такой вероятностью можно считать практически невозможным), то результат опыта следует считать противоречащим гипотезе H_0 о том, что закон распределения величины X есть $F(x)$. Эту гипотезу следует отбросить как неправдоподобную. На-

против, если вероятность p сравнительно велика, можно признать расхождения между теоретическим и статистическим распределениями несущественными и отнести их за счет случайных причин. Гипотезу H о том, что величина X распределена по закону $F(x)$, можно считать правдоподобной или, по крайней мере, не противоречащей опытным данным.

Таким образом, схема применения критерия χ^2 к оценке согласованности теоретического и статистического распределений сводится к следующему:

1) Определяется мера расхождения χ^2 по формуле (7.6.4).

2) Определяется число степеней свободы r как число разрядов k минус число наложенных связей s :

$$r = k - s.$$

3) По r и χ^2 с помощью табл. 4 определяется вероятность того, что величина, имеющая распределение χ^2 с r степенями свободы, превзойдет данное значение χ^2 . Если эта вероятность весьма мала, гипотеза отбрасывается как неправдоподобная. Если эта вероятность относительно велика, гипотезу можно признать не противоречащей опытным данным.

Насколько мала должна быть вероятность p для того, чтобы отбросить или пересмотреть гипотезу, — вопрос неопределенный; он не может быть решен из математических соображений, так же как и вопрос о том, насколько мала должна быть вероятность события для того, чтобы считать его практически невозможным. На практике, если p оказывается меньшим чем 0,1, рекомендуется проверить эксперимент, если возможно — повторить его и в случае, если заметные расхождения снова появятся, пытаться искать более подходящий для описания статистических данных закон распределения.

Следует особо отметить, что с помощью критерия χ^2 (или любого другого критерия согласия) можно только в некоторых случаях опровергнуть выбранную гипотезу H и отбросить ее как явно несогласную с опытными данными; если же вероятность p велика, то этот факт сам по себе ни в коем случае не может считаться доказательством справедливости гипотезы H , а указывает только на то, что гипотеза не противоречит опытным данным.

С первого взгляда может показаться, что чем больше вероятность p , тем лучше согласованность теоретического и статисти-

ческого распределений и тем более обоснованным следует считать выбор функции $F(x)$ в качестве закона распределения случайной величины. В действительности это не так. Допустим, например, что, оценивая согласие теоретического и статистического распределений

по критерию χ^2 , мы получили $p = 0,99$. Это значит, что с вероятностью 0,99 за счет чисто случайных причин при данном числе опытов должны были получиться расхождения большие, чем наблюдаемые. Мы же получили относительно весьма малые расхождения, которые слишком малы для того, чтобы признать их правдоподобными. Разумнее признать, что столь близкое совпадение теоретического и статистического распределений не является случайным и может быть объяснено определенными причинами, связанными с регистрацией и обработкой опытных данных (в частности, с весьма распространенной на практике «подчисткой» опытных данных, когда некоторые результаты произвольно отбрасываются или несколько изменяются).

Разумеется, все эти соображения применимы только в тех случаях, когда количество опытов n достаточно велико (порядка нескольких сотен) и когда имеет смысл применять сам критерий, основанный на предельном распределении меры расхождения при $n \rightarrow \infty$. Заметим, что при пользовании критерием χ^2 достаточно большим должно быть не только общее число опытов n , но и числа наблюдений m_i в отдельных разрядах. На практике рекомендуется иметь в каждом разряде не менее 5—10 наблюдений. Если числа наблюдений в отдельных разрядах очень малы (порядка 1—2), имеет смысл объединить некоторые разряды.

Кроме критерия χ^2 , для оценки степени согласованности теоретического и статистического распределений на практике применяется еще ряд других критериев. Из них мы вкратце остановимся на критерии А. Н. Колмогорова.

В качестве меры расхождения между теоретическим и статистическим распределениями А. И. Колмогоров рассматривает максимальное значение модуля разности между статистической функцией распределения $F^*(x)$ и соответствующей теоретической функцией распределения:

$$D = \max |F^*(x) - F(x)|$$

Основанием для выбора в качестве меры расхождения величины D является простота ее вычисления. Вместе с тем она имеет достаточно простой закон распределения. А.Н. Колмогоров доказал, что какова бы ни была функция распределения $F(x)$ непрерывной случайной величины X , при неограниченном возрастании числа независимых наблюдений n вероятность неравенства

$$D\sqrt{n} \geq \lambda$$

стремится к пределу

$$p(\lambda) = 1 - \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2}$$

Значение вероятности $p(\lambda)$, подсчитанные по формуле (7.6.5.) приведены в таблице 7.6.1.

Таблица 7.6.1.

λ	$P(\lambda)$	λ	$P(\lambda)$	λ	$P(\lambda)$
,0	000	,7	0,771	,4	0,040
,1	000	,8	0,544	,5	0,022
,2	000	,9	0,393	,6	0,012
,3	000	,0	0,270	,7	0,006
,4	0,997	,1	0,178	,8	0,003
,5	0,964	,2	0,112	,9	0,002
,6	0,864	,3	0,068	,0	0,001

Схема применения критерия А. Н. Колмогорова следующая: строится статистическая функция распределения $F^*(x)$ и предполагаемая

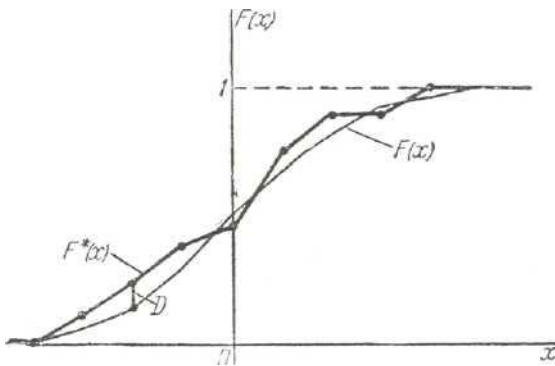


Рис 7.6.2.

теоретическая функция распределения $F(x)$, и определяется максимум D модуля разности между ними (рис. 7.6.2).

Далее, определяется величина

$$\lambda = D\sqrt{n}$$

и по таблице 7.6.1 находится вероятность $P(X)$. Это есть вероятность того, что (если величина X действительно распределена по закону $F(x)$) за счет чисто случайных причин максимальное расхождение между $F^*(x)$ и $F(x)$ будет не меньше, чем фактически наблюдаемое. Если вероятность $P(X)$ весьма мала, гипотезу следует отвергнуть как неправдоподобную; при сравнительно больших $P(\lambda)$ ее можно считать совместимой с опытными данными.

Критерий А. Н. Колмогорова своей простотой выгодно отличается от описанного ранее критерия χ^2 ; поэтому его весьма охотно применяют на практике. Следует, однако, оговорить, что этот критерий можно применять только в случае, когда гипотетическое распределение $F(x)$ полностью известно заранее из каких-либо теоретических соображений, т. е. когда известен не только вид функции распределения $F(x)$, но и все входящие в нее параметры. Такой случай сравнительно редко встречается на практике. Обычно из теоретически соображений известен только общий вид функции $F(x)$, а входящие в нее числовые параметры определяются по данному статистическому материалу. При применении критерия χ^2 это обстоятельство учитывается соответствующим уменьшением числа степеней свободы распределения χ^2 . Критерий А. Н. Колмогорова такого согласования не предусматривает. Если все же применять этот критерий в тех случаях, когда параметры теорети-

ческого распределения выбираются по статистическим данным, критерий дает заведомо завышенные значения вероятности $P(\lambda)$; поэтому мы в ряде случаев рискуем принять как правдоподобную гипотезу, в действительности плохо согласующуюся с опытными данными.

Глава 4. Математическое обеспечение САПР-решение задачи синтеза, методы оптимизации.

Введение в проблему.

Синтез системы - выбор определенного варианта (принятие решения) - основная проблема проектирования. Применение математических методов при решении этой задачи находит всё более широкое применение в различных областях науки и техники. Это сравнительно молодая область математики, ориентированная на компьютерное моделирование, требующая высокопроизводительных средств вычислительной техники. Наибольший успех и широкое внедрение в практику получило решение этих задач в экономике (распределение ресурсов, транспортные задачи и т.д.). Хотя математические модели в этой области сравнительно просты, процедуры поиска оптимальных решений и их определение отнюдь не тривиальны и потребовали оригинальных подходов. Именно в связи с решением подобных задач известный советский математик, академик Конторович Л.В. был удостоен Нобелевской премии по экономике.

Вообще применение математических методов для обоснования решений во всех областях человеческой деятельности занимается раздел научных методов под общим названием исследование операций. Это междисциплинарное направление, возникновение которого предшествовало становлению теории систем и тесно с ней связано.

Изначально, исследование операций возникло в связи с задачами военного характера в министерства обороны США. Предметом исследования является разработка методов анализа целенаправленных действий (операций) и объективная сравнительная оценка полученных решений.

Распространение идей исследования операций совпали с развитием методов математического программирования, которое в отличие от классических математических методов включает некоторые средства постановки задачи, позволяют получать область допустимых решений и различные варианты этих решений.

Особенности исследования операций заключаются в следующем:

а) предполагается разработка нескольких вариантов решений, несколько путей достижения целей, отличных от традиционных;

б) при выборе решения допускается учет не только количественных, но и качественных критериев, что позволяет обеспечивать большее соответствие решаемой задачи реальной действительности и большую объективность.

Основным математическим аппаратом исследования операций является математическое программирование. Это раздел теории оптимизации (теории экстремальных задач), занимающийся изучением и решением задач минимизации (максимизации) функций нескольких переменных на подмножестве векторного пространства векторного пространства заданного в виде системы уравнений или системы неравенств.

Математическое программирование включает следующие разделы:

Линейное программирование применяется для простых линейных задач с линейными ограничениями, нашло широкое применение, главным образом в экономических задачах, методы хорошо разработаны.

Дискретное программирование – применяется для целочисленных решений.

Динамическое программирование – применяется для многоэтапных задач (не оптимальность на каждом этапе может привести к оптимальности в целом).

Нелинейное программирование - для нелинейных задач с нелинейными ограничениями. Существует несколько разновидностей нелинейного программирования – квадратичное, аппроксимируемое квадратичными функциями, выпуклое имеющее один оптимум в рассматриваемой области (унимодальное), невыпуклое – наиболее сложны многоэкстремальные задачи.

Математическая постановка задачи

Цель всех методов – определить оптимальное решение, то есть решение, которое по тем или иным признакам предпочтительнее перед другими. Параметры, совокупность которых образуют решения, называют элементами решения.

Для представления задачи в символическом виде обозначим множество возможных решений с помощью вектора \mathbf{X} , каждую отдельную совокупность решений через \mathbf{X} . Для сравнения между собой различных решений выводится количественный критерий показатель эффективности \mathbf{W} или целевая функция.

Как правило, показатель эффективности зависит от двух факторов:

$$\mathbf{w}=\mathbf{w}(\alpha,\mathbf{x})$$

α – заданная заранее известные факторы (детерминированные или стохастические),

x – одна из совокупностей проектных решений.

Задача выбора совокупности решений формируется следующим образом. При заданном комплексе условий α найти такое решение $x=x^*$, которое обращает показатель эффективности W в максимум.

$$W^*=\max\{W(\alpha,x)\}$$

$$x \in X$$

W^* – максимальное значение $W(\alpha,x)$ взятое по всем решениям из множества возможных X .

В общем виде это классическая математическая задача нахождения максимума функции или функционала (функционалом называется величина, зависящая от вида функции), т.е. решение x включает не только переменные, но и функции и в этом случае $W(\alpha,x)$ является функционалом. Пример из ускорительной техники - выбор функции, описывающий изменения напряженности ускоряющего поля и фазовой скорости электромагнитной волны по длине группирования с целью обеспечения минимального фазового и энергетического разброса сгустка ускоряемых заряженных частиц и максимального набора энергии пучком.

Описанная задача принадлежит к классу так называемых классических математических вариационных задач, давно разрабатываемых математиками. Что бы найти экстремум функции, необходимо продифференцировать модель по всем переменным, приравнять каждое уравнение к нулю и решить полученную систему (в простейшем случае правило Лопиталя - дифференцируем числитель и знаменатель и приравниваем нулю).

Однако использование классических методов вариационного исчисления в ряде случаев на практике затруднено, а иногда и невозможно по ряду причин:

-Когда параметров много, задача решения системы напрямую может оказаться непростой.

-Когда на элементы решения наложены ограничения. Экстремумы часто достигаются отнюдь не в точках нулевых производных (такой точки может не быть вообще, как в линейных задачах). Оптимум может находиться на границе. При этом возникает многомерная вариационная задача при ограничениях, трудно реализуемая традиционными методами.

- В некоторых задачах функция W вообще не имеет производных, например для целочисленного значения аргумента.

Все эти причины делают задачу поиска отнюдь не тривиальной, а классические методы вариационного исчисления малоэффективными. Это и привело к появлению альтернативных, чисто машинных (вычислительных) методов математического программирования. Методов, основанных на использовании модели, как средства анализа и специальных процедур выбора параметров, как средства автоматизированного поиска решения (самый примитивный способ-сканирование - тупой перебор).

Это не строгие, а так называемые эвристические методы, основанные на здравом смысле, интуиции и аналогиях.

Но несмотря на перечисленные трудности представленная задача находится ещё в пределах возможностей вариационного исчисления.

Всё вышеперечисленное применимо в случае одного единственного критерия оптимальности W , на практике большинство задач не может быть оценено с помощью одного критерия и такие задачи называются многокритериальными.

Существует несколько подходов к решению многокритериальных задач, что говорит о том, что ни один из них не является самодостаточным.

Один из наиболее распространённых подходов заключается в сведении нескольких критериев к одному - обобщенному и минимизации этого обобщенного критерия.

Используются два варианта этого подхода:

а) Мультипликативный критерий:

$$F(\bar{x}) = \prod_{j=1}^r y_j / \prod_{k=r+1}^m y_k$$

б) Аддитивный критерий (взвешенный аддитивный)

$$F(\bar{x}) = \sum_{j=1}^r a_j y_j - \sum_{k=r+1}^m a_k y_k$$

Здесь предполагается, что выходные параметры упорядочены: первые r нужно уменьшать, а остальные $m-r$ наоборот увеличивать, a_j - весовой коэффициент, учитывающий относительный вклад соответствующего параметра в общую оценку эффективности. К недостаткам этих критериев относится неучёт технических требований, предъявляемых к выходным параметрам. В результате возможно перевыполнить требования Т.З. к одной группе параметров за счёт невыполнения требований Т.З. к другой - низкую стоимость нельзя компенсировать плохой работоспособностью.

2. Максиминные (минимаксные) критерии оптимизации, исключаяющие недостатки предыдущих критериев.

В этом случае для каждого из критериев u_j вводится (различным образом) запас работоспособности Z_i и проводится максимизация минимального из запасов

$$K = \max \min Z_i(x)$$

Критерий имеет ограниченное применение, поскольку оптимизация ведется только по одному параметру.

Большую ценность для разработчика представляют результаты оптимизации по статистическим критериям, в которых целевой функцией является вероятность выполнения всех заданных условий работоспособности. Однако использование этих критериев требует в процессе оптимизации многократного выполнения статистического анализа методом Монте-Карло, что для достаточно сложных моде-

лей анализа требует использования компьютеров очень высокой производительности.

4.2 Классификация методов решения задач оптимального проектирования (нелинейное программирование).

В силу указанных выше причин задачам нелинейного программирования в настоящее время уделяется значительное внимание. Предложен ряд методов, позволяющих с определенной точностью решения для специального класса задач. Однако, по-видимому, невозможно создать метод, позволяющий получить точное решение общей задачи нелинейного программирования за конечное число шагов. Таблица 4.2.1. представляет обобщенный вариант иерархической структуры методов нелинейного программирования. Безусловно, таблица не претендует на обобщение всех существующих на сегодняшний день методов оптимизации в классе нелинейных задач. Структура таблицы включает основные компонент общей задачи нелинейного программирования и наиболее характерные методы их реализации.

Каждый уровень классификации, приведенный в таблице, соответствует этапу расчёта, либо конкретному методу, применяемому на данном этапе. Пунктирными вертикальными линиями разделены этапы поиска, входящие в качестве составных элементов в процедуру поиска более высокого уровня.

Общая задача невыпуклого программирования (верхний уровень классификации) наиболее сложна и предполагает множество локальных оптимумов (\max , \min). При этом поиск направлен на выявление глобального оптимума (2й уровень). В основе всех алгоритмов глобального оптимума лежат получение и использование информации о расположении локальных оптимумов. Алгоритмы поиска глобальных оптимумов различаются по способу генерации начальных точек. Простейшим является так называемая последовательность независимых испытаний или метод сканирования (обычный переборный алгоритм). Каждое испытание состоит из вычисления в некоторой допустимой точке значения критерия оптимальности и его сравнение со значением, полученным в предыдущих точках. Последовательность испытаний может быть детерминиро-

ванной (например, обход узлов n-мерной сетки в пространстве конструктивных параметров) или стохастической.

Первый подход применим, если независимых переменных немного и невелики пределы, в которых они могут изменяться. Алгоритм прост, но даже при небольшом числе независимых переменных может потребовать колоссальных затрат компьютерного времени. Существуют методы ускорения этого алгоритма, в частности путем построения вероятностных (стохастических) методов выбора начальных точек.

Следующая группа методов поиска глобального оптимума (экстремума) основана на использовании априорной информации (известной до опыта) о характере критерия оптимальности. Обычно эти методы пригодны для специального класса задач, поскольку указанная информация может быть получена на основе исследования специфики проектируемого объекта. Наибольшее распространение получили два априорных условия (критерия): Липшица и сепарабельности.

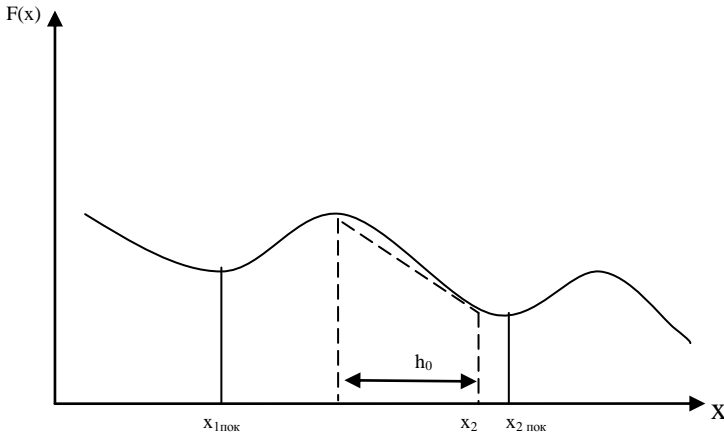
1. Критерий Липшица предлагает полный перебор на неравномерной сетке (при выполнении условия Липшица в области G). Если существует такая постоянная величина C, что для любых двух векторов $X_1, X_2 \in G$ выполняется следующее условие (неравенство):

$$|F(x^1) - F(x^2)| \leq C|x^1 - x^2|$$

Это условие означает, что $F(X)$ убывает и возрастает не быстрее линейной функции с заданным коэффициентом C. По константе C_0 фиксируем шаг и вычисляем значение функции.

Поиск глобального оптимума										
Поиск локального оптимума								Выбор начальных точек		
Методы определения длины шага			Методы представления математической модели				Методы определения направления движения к оптимуму (безусловная оптимизация)			
Параболическая аппроксимация «Золотое сечение» Фибоначчи			Формирование обобщенного критерия оптимальности		Формирование возможных направлений (условная оптимизация)		Стохастический		Детерминированный	
			Барьерные функции		Штрафные функции		«Овражный»		Оценки градиенты	Случайного поиска (слепой поиск)
			Метод проекций в переменной метрике		Линейное локальное моделирование		Переменной метрики		Градиентный	
			Проекционный градиентный метод		«Овражный»		Оценки градиенты		Случайного поиска (слепой поиск)	Прямой
Последовательность испытаний с адаптацией										
Использование априорной информации										
Последовательность независимых испытаний										

Таблица 4.2.1



2. Если выполняется условие сепарабельности, то есть удастся выделить отдельные группы параметров и применить поэтапную оптимизацию по отдельным группам, используя принцип динамического программирования, то задача существенно упрощается.

Наиболее общим и в то же время достаточно эффективным подходом к поиску глобального оптимума является последовательность испытаний с адаптацией. В отличие от обычного подхода, при адаптивном для восполнения недостающей априорной информации активно используется текущая информация. Адаптация в данном случае состоит в последовательном автоматическом совершенствовании процедур поиска (сокращения числа испытаний по мере накопления информации).

На разработку глобальных методов оптимизации направлены усилия многих исследователей, однако имеющиеся к настоящему времени предложения, как из-за повышенной трудоемкости, так и из-за недостаточной надежности определения глобального оптимума (экстремума) еще не получили распространения. Гораздо в большей степени разработаны методы поиска локальных оптимумов, которые с одной стороны составляют основу поиска глобально, а с другой стороны имеют самостоятельное значение в случае унимодальных функций.

Прежде чем переходить непосредственно к изучению методов поиска локальных оптимумов полезно ввести некоторые определения и геометрические представления, облегчающие рассмотрение конкретных процедур поисковой оптимизации унимодальных функций.

Если оптимизация ведется без учета статистического разброса параметров, то соответствующих критерий называется детерминированным критерием, а если разброс учитывается, критерий называется стохастическим. Если при реализации детерминированных критериев вводится вероятностная модель поиска, такой метод называется стохастическим (стохастическая аппроксимация).

При рассмотрении поисковой оптимизации может быть полезно введение некоторых геометрических представлений для исследуемых функций. Геометрическое представление функции целесообразно изображать в виде линий равного уровня, которые определяются уравнением $F(X) = a$ и являются геометрическим местом точек в пространстве управляемых параметров, для которых значение функции постоянно и равно a . Функция представляется в виде набора линий равного уровня в диапазоне допустимых внутренних параметров при размерности пространства $n = 2$ (два исследуемых внутренних параметра). При размерности пространства $n = 3$ функция представляется в виде набора поверхностей с одинаковым значением функции, которые также называются поверхностям отклика. При размерности пространства $n > 3$ имеют место так называемые гиперповерхности отклика, не имеющие геометрической интерпретации. Все дальнейшие рассмотрения будут представлены на двумерных функциях, отображаемых на плоскости.

Ход процесса оптимизации на двумерной функции, отображаемой линиями равного уровня, представляется траекторией поиска — последовательностью отображаемых точек X_k (исходная точка X_0 , X_k — промежуточная точка) в заданную ε -окрестность X^* , соединенных отрезками прямых.

Как следует из таблицы 4.2.1. процедура поиска локального оптимума включает решение трех задач:

- 1) Выбор представления математической модели;
- 2) Выбор процедуры определения направления движения к экстремуму;

3) Определение длины шага по выбранным направлениям движения.

Начало процедуры оптимизации этого уровня предполагает решение первой задачи, которая в приведенной классификации предполагает два способа представления математической модели разрабатываемого объекта.

Первый способ объединяет критерий оптимальности и ограничения в один обобщенный критерий оптимальности, то есть сводит поиск к задаче без ограничений. При приближении к границе допустимой области или при выходе из неё обобщенный критерий начинает резко возрастать и поиск автоматически возвращается в допустимую область. Таким образом снимаются ограничения и становятся возможны для применения методы безусловной оптимизации. В таблице представлены два способа сведения задач с ограничениями к задачам без ограничений: метод штрафных функций и метод барьерных функций (в случае функциональных ограничений).

В первом случае (штрафных, нежестких ограничений) вводится функция $\theta_k(x)=0$ внутри допустимой области параметров и $\theta_k(x)>0$ если не выполняется хотя бы одно ограничение и рассматривается как наложение штрафа.

Обобщенный критерий оптимальности имеет вид

$$\Phi(\bar{x}) = W(\bar{x}) + \theta_k(\bar{x})$$

Где $\theta_k(\bar{x}) = r_k \sum_{i=1}^m [\max\{0, \psi_i(\bar{x})\}]^2$ - как правило набор дельта функций

($r_k > 0$ - штрафной параметр, индекс $k=1, 2, \dots$ - последовательные значения штрафного параметра соответствующие номеру задачи)

Иногда метод штрафных функций называют методом внешней точки, так как поиск может находиться вне допустимой области.

В методе барьерных функций (или методе внутренней точки) поиск может осуществляться только в пределах границ и создается как бы барьер через который функции перейти не сможет.

$\Phi(\bar{x}) = W(\bar{x}) - r_k \sum_{i=1}^m (1/\psi_i(\bar{x}))$ По мере приближения к границе

области, какой-либо из элементов $\psi_i(\bar{x})$ вектора ограничений

стремиться к "0", а функция $\theta_k = -r_k \sum_{i=1}^m (1/\psi_i(\bar{x}))$ неограниченно

возрастает.

Наиболее наглядно этот метод можно продемонстрировать для случая простых (прямых, не функциональных) ограничений. Допустим, что на управляющие параметры (аргументы) положены ограничения $a \leq u_i \leq b$. Введем новую переменную

$x_i = \text{tg} \left[\pi \frac{u_i - (b+a)/2}{b-a} \right]$, которая (как не сложно убедиться) при

приближении U_i к границам а или б устремляется к бесконечности и возвращает параметры в область допустимых значений.

Второй подход при выборе математической модели в процедурах оптимизации и учёта ограничений основан на разделении этапов определения убывания (увеличения) критерия оптимизации и коррекции параметров процедуры в соответствии с ограничивающими условиями на допустимую область поиска. Группа этих методов (условная оптимизация) перечислена в таблице как методы возможных направлений.

Следующие две задачи, которые необходимо решить для минимизации (или максимизации) целевой функции - это выбор направления движения к искомой точке и выбор шага, с которым необходимо двигаться. Хотя по логике следовало бы начать с изучения методов выбора направления движения, а затем уже выбирать каким образом шагать в этих направлениях, мы начнем с последнего, чему есть ряд объяснений.

Во-первых, для одномерных задач выбор шага полностью определяет процедуру оптимизации. Во-вторых, некоторые методы выбора шага являются составной частью многомерных задач оптимизации (т.е. являются основным алгоритмом оптимизации), что будет рассмотрено в дальнейшем.

4.3. Методы одномерного поиска

Простейшим алгоритмом выбора шага является монотонное изменение параметра x с заданным шагом h в сторону убывания критерия оптимальности:

Т.е. $x_k = x_0 + kh, k=1,2,\dots$ после того как $F(\bar{x})$ начинает возрастать для некоторого k_0 производится смена направления движения в обратном направлении с одновременным уменьшением шага вдвое:

$$x_{k_0+2} = x_{k_0+1} - \frac{h}{2} \text{ и т.д.}$$

$$x_{k_0} = x_0 + \sum_{i=1}^{p_k} (-1)^i k_i \frac{h}{2^i}$$

Иногда этот метод называется методом дихотомии. Алгоритм прост, но требует большого числа итераций в тех случаях, когда начальная точка находится далеко от экстремума, а так же при пологой функции критерия оптимальности.

Гораздо более эффективны методы, при которых производится рассмотрение всего диапазона поиска и двухстороннее его сужение.

Поиск экстремума можно сравнить с поиском в озере самого глубокого места. При каждом замере лотом получается новая информация. Если в последующем замере глубина

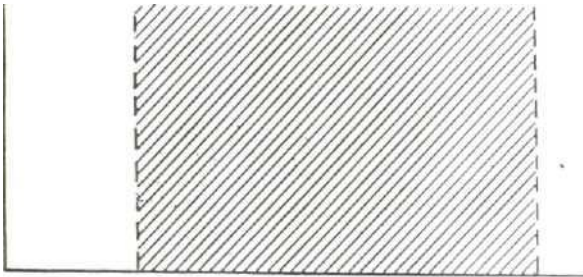


Рис 4.3.1 Интервал неопределенности

оказалась больше, чем в предыдущем, то полученная информация полезна. Напротив, если в последующем замере получена меньшая глубина, чем в предыдущем, то результат не дает полезной информации и затраченные на его получение усилия были напрасными. Разрабатывая методы поиска, стремятся найти экстремум как можно быстрее, сделав как можно меньше бесполезных попыток.

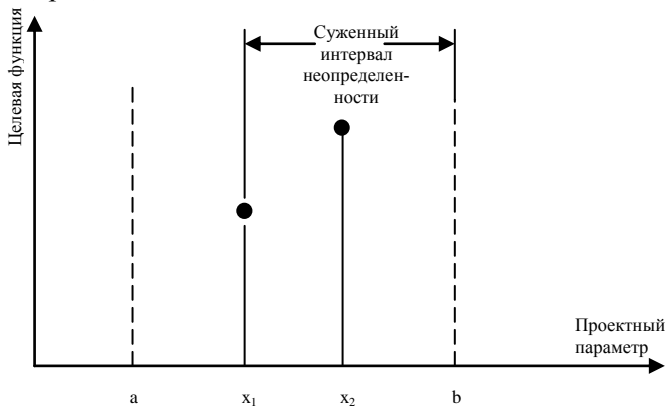


Рис 4.3.2. Сужение интервала неопределенности путем вычисления двух значений целевой функции

Задачу одномерной оптимизации можно поставить следующим образом. Значения проектного параметра x должны быть заключены в интервале $a \leq x \leq b$. Приступая к решению задачи, мы ничего не знаем о целевой функции, кроме того, что она унимодальная. Интервал значений x , в котором заключен оптимум, будем называть «интервалом неопределенности». В начале процесса оптимизации этот интервал имеет длину $b-a$ (рис. 4.3.1). Вычислив значения целевой функции M_1 и M_2 при значениях x_1 и x_2 в указанном интервале, сузим интервал неопределенности (рис. 4.3.2). Существует несколько способов систематического сужения интервала, которые излагаются ниже.

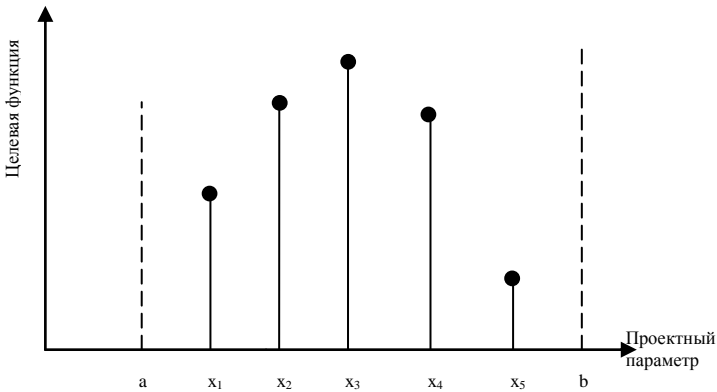
Общий поиск

Очевидно, наиболее естественным способом сужения интервала неопределенности для одномерной унимодальной функции является деление его на несколько равных частей с последующим вычис-

лением значений целевой функции в узлах полученной сетки (рис. 4.3.3). В результате интервал неопределенности сужается до двух шагов сетки. Обычно говорят о дроблении интервала неопределенности, которое характеризуется коэффициентом f . Разделив интервал неопределенности на N частей, получим $N+1$ узел, и тогда

$$f = \frac{2}{N+1}$$

Чтобы получить значение $f=0,01$, потребуется вычислить целевую функцию в 199 точках, а при $f=0,001$ $N=1999$. Ясно, что эффективность этого метода при уменьшении интервала неопределенности быстро падает. Сам собой напрашивается другой путь: чтобы получить $f=0,01$, вычислить



сначала функцию в 19 точках и получить $f=0,1$, а затем, вычислив еще 19 значений функции на сокращенном интервале неопределенности, получить $f=0,01$, сделав при этом всего 38, а не 199 вычислений. Таким образом, при некоторой изобретательности эффективность поиска можно резко повысить.

Деление интервала пополам

Пользуясь тем же приемом, но вычисляя значения функции в подинтервалах неодинаковое число раз, можно дополнительно повысить эффективность поиска. Вычисляя N значений функции на i последовательно сужаемых интервалах, получим для коэффициента дробления интервала неопределенности

$$f = \left(\frac{2}{N+1} \right)^i$$

При таком методе поиска целевую функцию приходится вычислять J раз, причем $J = N_i$. Весьма заманчиво найти оптимальное значение N , при котором J при заданном f минимально.

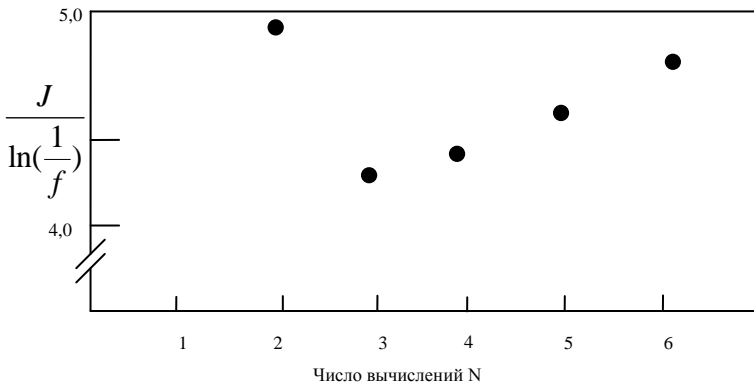


Рис. 4.3.4. Определение оптимального числа вычислений целевой функции при применении метода деления интервала пополам. С помощью соотношения $i=J/N$ и выражения для интервала неопределенности можно найти J :

$$J = \frac{N \ln(1/f)}{\ln\left(\frac{N+1}{2}\right)}$$

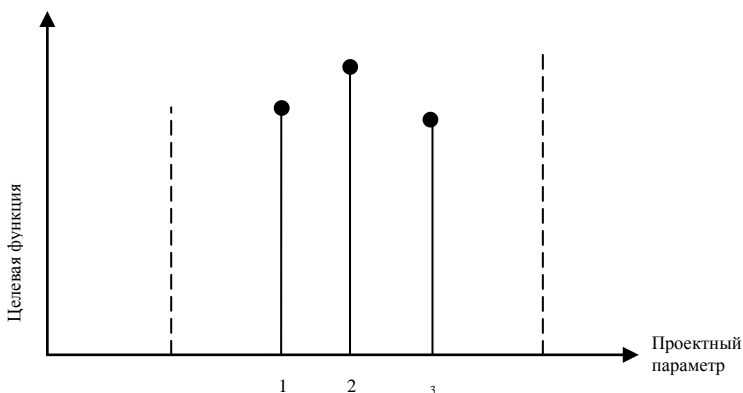


Рис. 6.9. Первый шаг при применении метода деления интервала пополам.

Если нанести значения $\frac{J}{\ln \frac{1}{f}}$ на график (рис. 4.3.4), оказывается,

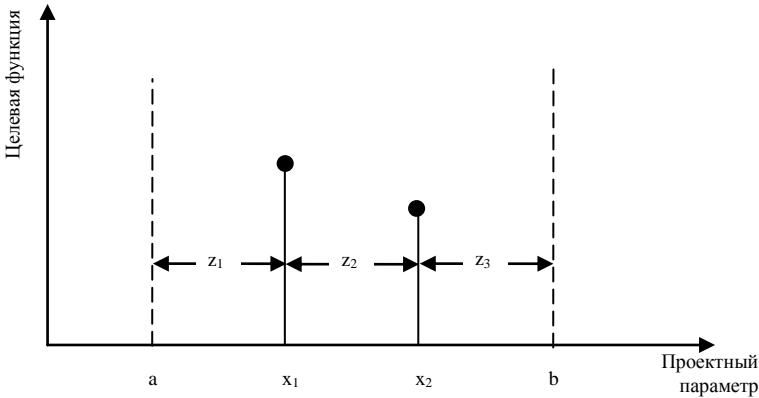
что минимум расположен где-то вблизи $N=3$. При этом на каждом интервале $f=0,5$. Поскольку интервал неопределенности делится каждый раз надвое, то метод и получил название метода деления интервала пополам. На рис. 4.3.6. показано, как три первых вычисленных значения функции позволяют сузить интервал неопределенности вдвое. Замечаем далее, что значение целевой функции в середине нового интервала уже известно. Поэтому для завершения поиска на следующем этапе потребуется вычислить только два (вместо трех) значения целевой функции. Это преимущество рассматриваемого метода сохраняется и в дальнейшем. В общем случае коэффициент дробления интервала неопределенности при $N \geq 3$ составляет

$$f = \frac{1}{2^{\frac{N-1}{2}}}$$

Метод дихотомии

Выше предполагалось, что значения целевой функции вычисляются при постоянном приращении проектного параметра. Ес-

ли снять это ограничение, то эффективность поиска можно повысить. Как уже отмечалось, вычисление целевой функции в двух точках



интервала неопределенности позволяет его сузить. Можно таким образом выбрать эти точки, что интервал неопределенности будет минимальным. На рис. 4.3.6. показаны обозначения, используемые в этой схеме. Если значение целевой функции при x_1 больше, чем при x_2 , то новый интервал неопределенности равен $Z_1 = z_1 + z_2$. В противном случае он определяется выражением $Z_2 = z_2 + z_3$. Задача состоит в том, чтобы одновременно минимизировать Z_1 и Z_2 , удовлетворив условиям

$$z_1 + z_2 + z_3 = Z,$$

$$z_1 > 0, z_2 > 0, z_3 > 0$$

Из равенства можно исключить z_2 . Тогда

$$Z - z_3 = \min, Z - z_1 = \min$$

Так как величина Z задана, то правые части этих уравнений будут тем меньше, чем больше z_1 и z_3 . Следовательно, оптимум соответствует условию

$$z_1 = z_3 = 0,5Z$$

Но тогда $z_2 = 0$, что противоречит условию $z_2 > 0$

Пусть z_2 имеет некоторое очень малое значение ε . Тогда из z_1 и z_3 вычтем по $\varepsilon/2$. В результате, после вычисления первой пары значений целевой функции при близких

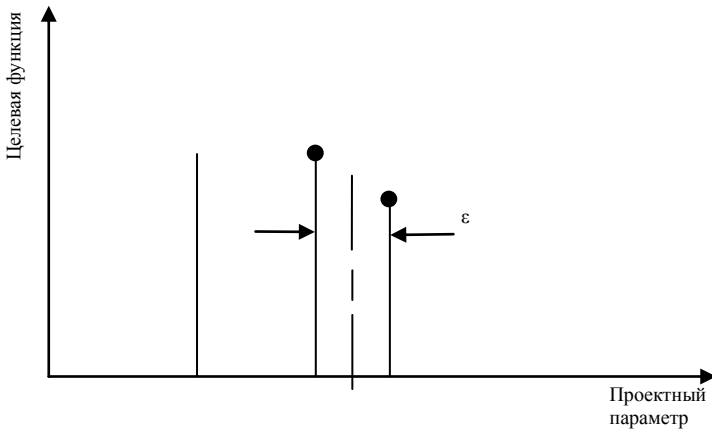


Рис. 4.3.7. Метод дихотомии.

значениях x интервал неопределенности сузится, как показано на рис.4.3.7, и коэффициент дробления будет равен

$$f = \frac{1}{2} + \frac{\varepsilon}{2}$$

В пределе, при $\varepsilon \rightarrow 0$, $f \rightarrow 1/2$. В дальнейшем при использовании метода дихотомии выполняются те же операции, что и при использовании метода деления интервала пополам. Отметим, однако, что для достижения одинаковых сужений интервала неопределенности метод дихотомии требует вычисления целевой функции в точках на одну меньше.

Метод золотого сечения

Из каждых трех значений целевой функции, вычисленных в интервале неопределенности, в дальнейшем используются только два, а третье не дает дополнительной информации и в дальнейшем не используется. В методе золотого сечения целевая функция вычисляется в точках интервала неопределенности, расположенных таким образом, чтобы каждое вычисленное значение целевой функции давало новую полезную информацию.

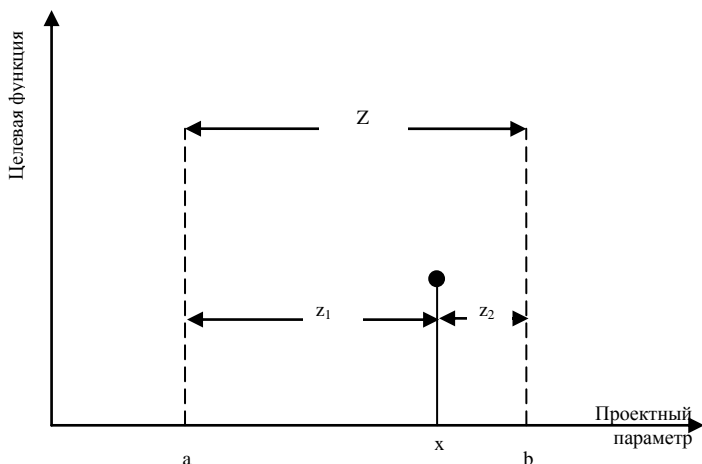


Рис. 4.3.8. Обозначения, используемые в методе золотого сечения

Сущность этого метода состоит в следующем. Интервал неопределенности делится на две неравные части так, что отношение длины большого отрезка к длине всего интервала равно отношению длины меньшего отрезка к длине большого отрезка. На рис.4.3.8 показан интервал неопределенности Z , состоящий из отрезков z_1 и z_2 , отношение длин которых определяется правилом золотого сечения

$$\frac{z_1}{Z} = \frac{z_2}{z_1}$$

Кроме того, $z_1 + z_2 = Z$. Из первого уравнения следует $z_1^2 = Zz_2$. Подставляя сюда значение Z из второго уравнения и деля обе части на z_1^2 , получаем

$$\left(\frac{z_2}{z_1} \right)^2 + \frac{z_2}{z_1} - 1 = 0$$

Решая это квадратное уравнение, находим для положительного корня значение

$$\frac{z_2}{z_1} = \frac{-1 + \sqrt{5}}{2} = 0,618$$

На рис. 4.3.9 показано деление интервала неопределенности в этом отношении и нанесены соответствующие значения целевой функции, которые позволяют уменьшить интервал неопределенности в $1/0,618$ раза. На этой стадии еще не видны преимущества метода золотого сечения по сравнению с методом дихотомии, однако они явно проявляются при дальнейшем делении интервала, так как оказывается, что одно из значений целевой функции, которое требуется вычислить на следующем шаге, уже известно.

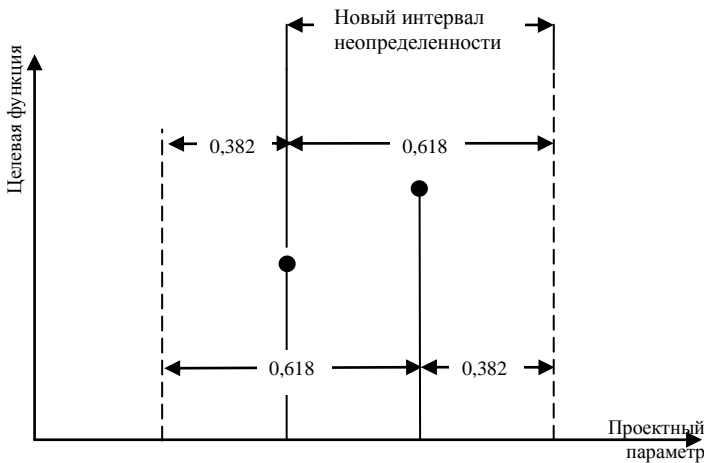


Рис. 4.3.9. Метод золотого сечения

Поэтому, чтобы уменьшить неопределенность еще в $1/0,618$ раза, потребуется дополнительно вычислить только одно значение целевой функции в точке, определяемой правилом золотого сечения. При $n > 2$ эффективность метода золотого сечения выше, чем у метода дихотомии, так как при каждом последующем вычислении целевой функции интервал неопределенности сокращается в $1/0,618$ раза. После вычисления N значений целевой функции коэффициент дробления интервала неопределенности составляет $f = 0,618^{N-1}$

Метод золотого сечения позволяет подметить интересную закономерность: наибольшее сокращение последующих интервалов неопределенности достигается при вычислении целевой функции в точках, равноудаленных от его центра. Если поступать таким образом и каждый раз, вычисляя целевую функцию, сокращать интервал неопределенности, то будут справедливы следующие соотношения:

$$Z_{J-2} = Z_{J-1} + Z_J, \quad 1 < J < N$$

где Z_J — длина интервала неопределенности после вычисления J -го значения целевой функции. Отметим, что золотого сечения существуют и другие методы поиска, основанные на вычислении целевой функции в точках, расположенных симметрично относительно центра интервала неопределенности. Для этих точек справедливы те же соотношения.

Метод Фибоначчи

Хотя метод золотого сечения и обладает высокой эффективностью, ясно, что он не является оптимальным при заданном числе вычислений целевой функции. Если конструктору заранее известно, что он сможет использовать лишь два значения целевой функции, то он, конечно, предпочтет метод дихотомии, который позволяет уменьшить интервал неопределенности сразу вдвое, а не в $1/0,618$ раза, как метод золотого сечения. Если есть возможность в процессе поиска оптимума изменять расположение точек, в которых вычисляются значения целевой функции, то можно соединить преимущества симметричного расположения точек, о которых говорилось выше, с преимуществами метода дихотомии и построить оптимальный алгоритм поиска. Пусть Z_n — длина интервала неопределенности после N -го шага. Условие симметрии имеет вид $Z_{J-2} = Z_{J-1} + Z_J, \quad 1 < J < N$

а условие вычисления последнего значения целевой функции на полученном методом дихотомии интервале δ записывается в виде $Z_{N-1} = 2Z_{N-\varepsilon}$

Из этих двух соотношений, возвращаясь назад, можно найти требуемую величину любого промежуточного интервала неопределенности и тем самым найти точки, в которых вычисляется целевая функция. Например,

$$Z_{N-3} = Z_{N-2} + Z_{N-1} = (Z_{N-1} + Z_N) + Z_{N-1} = 5Z_{N-2} - 2\varepsilon$$

и $Z_{N-4}=8Z_N-3\varepsilon$,

Общее выражение для произвольного интервала неопределенности имеет вид

$$Z_{N-K}=F_{K+1}Z_N-F_{K-1}\varepsilon$$

где коэффициенты F_j называются числами Фибоначчи и определяются следующим образом:

$$F_0=1, F_1=1, F_k=F_{k-1}+F_{k-2} \text{ при } K = 2,3,\dots$$

В табл. 6.1 приведены некоторые числа Фибоначчи. В общем случае величину последнего интервала неопределенности можно выразить в виде

$$Z_N = \frac{1}{F_N} + \frac{F_{N-2}}{F_N} \varepsilon$$

В пределе при $\varepsilon \rightarrow 0$ нижняя граница определяется величиной наименьшего интервала неопределенности, которую можно получить при заданном числе вычислений целевой функции.

Таблица 6.1

Числа Фибоначчи

k	F_k
0	1
1	1
2	2
3	3
4	5
5	8
6	13
7	21
8	34
9	55
10	89
11	144
12	233
13	377
14	610
15	987
16	1597

17	2584
18	4181
19	6765
20	10946

Применяя метод Фибоначчи, прежде всего решают, сколько значений целевой функции N может быть использовано. Затем, зная величину интервала неопределенности, выбирают распределение этих значений N в нем. Так как $Z_1 = Z_0 = 1$, то сначала вычисляют целевую функцию в точках, расположенных на расстояниях Z_2 от противоположных концов исходного интервала. В этом случае

$$Z_2 = \frac{F_{N-1}}{F_N} + \frac{(-1)^N}{F_N} \varepsilon$$

где ε — наименьшее приращение x , при котором значения целевой функции отличны друг от друга. На следующем шаге выбирают величину смещений Z_3 относительно Z_2 . В дальнейшем сохраняют подходящее значение интервала и повторяют процесс до тех пор, пока не будет вычислено N -е значение целевой функции.

Сравнение методов поиска по значениям коэффициента дробления интервала неопределенности f^1

Число вычислений целевой функции	Общий поиск	Деление интервала пополам	Метод дихотомии	Метод золотого сечения	Метод Фибоначчи
1	1,0	1,0	1,0	1,0	1,0
2	0,667	-	0,500	0,618	0,500
3	0,500	0,500	-	0,382	0,333
4	0,400	-	0,250	0,236	0,200
5	0,333	0,250	-	0,146	0,125
6	0,286	-	0,125	0,090	0,077
7	0,250	0,125	-	0,056	0,048
8	0,222	-	0,0625	0,0345	0,0294
9	0,200	0,0625	-	0,0213	0,0182
10	0,182	-	0,0312	0,0132	0,0112
11	0,167	0,0312	-	0,00813	0,00694
12	0,154	-	0,0156	0,00502	0,00429

13	0,143	0,0156	-	0,00311	0,00256
14	0,133	-	0,00781	0,00192	0,00164
15	0,125	0,00781	-	0,00119	0,00101
16	0,118	-	0,00391	0,00073 3	0,000626
17	0,111	0,00391	-	0,00045 3	0,000387
18	0,105	-	0,00195	0,00028 0	0,000239
19	0,100	0,00195	-	0,00017 3	0,000148
20	0,095	-	0,000976	0,00010 7	0,0000913

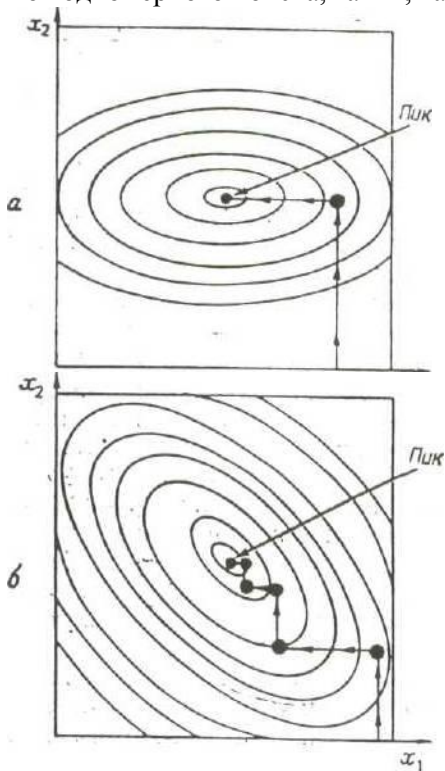
4.4 Методы многомерного поиска

По традиции методы оптимизации в многомерном пространстве можно разделить на регулярные и стохастические. В свою очередь регулярные методы подразделяются на две большие группы- прямые и косвенные. Прямые методы основаны на сравнении вычисляемых значений целевой функции в различных точках, а косвенные на использовании необходимых и достаточных условий математического определения максимума и минимума функций. Стратегия прямых методов-постепенное приближение к оптимуму; при использовании косвенных методов стремятся найти решение не исследуя неоптимальные точки.

Метод покоординатного подъема

Логическим развитием рассмотренной выше методики одномерного поиска было бы последовательное изменение каждого проектного параметра до тех пор, пока не будет достигнут максимум целевой функции. По завершении этой процедуры для всех переменных можно вернуться к первой и посмотреть, нельзя ли еще более усовершенствовать решение. Этот метод, называемый методом покоординатного подъема, не всегда позволяет найти оптимальное решение. На рис.4.4.1,а показана двумерная целевая функция, подходящая для решения задачи этим методом. Ее особенность состоит в том, что линии уровня близки по форме к окружностям или эллипсам, оси которых параллельны осям координат. Если же эти оси наклонены к осям координат (рис.4.4.1, б), то

эффективность алгоритма снижается, так как для нахождения оптимума приходится вычислять гораздо больше значений целевой функции. Метод покоординатного подъема совершенно неприменим, если линии уровня имеют точки излома (рис. 4.4.1.). Поскольку линии уровня такого типа весьма часто встречаются в инженерной практике, то прежде, чем воспользоваться указанным методом, следует убедиться, что решаемая задача не имеет подобного недостатка. Несмотря на это, метод покоординатного подъема часто используют на первой стадии решения задачи, применяя затем более сложные методы. К достоинствам метода покоординатного подъема следует отнести возможность использования простых алгоритмов одномерного поиска, таких, как метод золотого сечения.



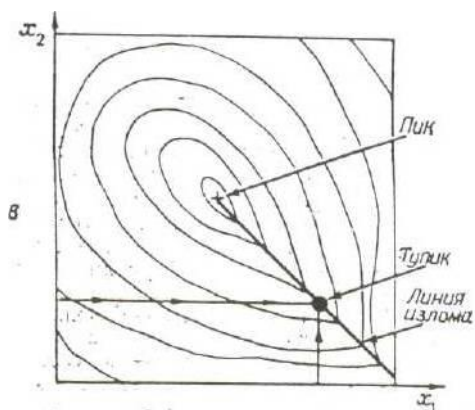


Рис. 4.4.1. метод покоординатного подъема

Метод исключения областей

Зная из главы 4.3, насколько эффективно методы одномерного поиска позволяют сокращать интервал неопределенности (одномерный или двумерный), можно попытаться применить ту же методику и к многомерному пространству. Один из наиболее очевидных методов исключения областей называется методом касательной к линии уровня, так как в нем используются касательные к линиям уровня целевой функции.

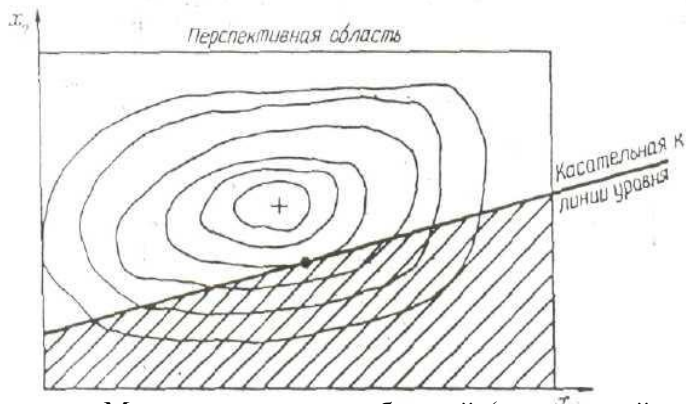


Рис. 4.4.2. Метод исключения областей (касательной к линии уровня) в случае выпуклых линий уровня.

Продemonстрируем этот метод на примере двумерной целевой функции, линии уровня которой показаны на рисх4.4.2. Пусть произвольно выбранная точка пространства проектирования лежит на линии уровня, проходящей несколько ниже пика, соответствующего оптимальному решению. Проведем через эту точку касательную к линии уровня. Сделать это нетрудно, так как касательная должна лежать в плоскости линии уровня и быть перпендикулярной локальному градиенту поверхности целевой функции. Если целевая функция достаточно гладкая и унимодальная, то касательная к линии уровня разделит пространство проектирования на две части, в одной из которых вероятность нахождения оптимума велика, а в другой мала. Пользуясь этим приемом в нескольких удачно выбранных точках, для которых известны значения целевой функции, можно существенно сузить область поиска. Однако осуществление этого алгоритма связано с некоторыми трудностями. Если линии уровня вогнутые, а не выпуклые, то может оказаться исключенной область, содержащая экстремум (рис. 4.4.3). Кроме того, оставшаяся после нескольких исключений область неопределенности может иметь конфигурацию, мало пригодную для применения других алгоритмов.

Одним из методов исключения является метод сеточного поиска, разработанный Мишке и дающий неплохие результаты. В этом случае суженная область неопределенности представляет собой гиперкуб - многомерный аналог квадрата или куба, - размеры которого можно определить заранее. Благодаря этому метод Мишке является одним из немногих методов многомерного поиска, эффективность которого поддается измерению. Чтобы лучше понять сущность этого метода, рассмотрим его

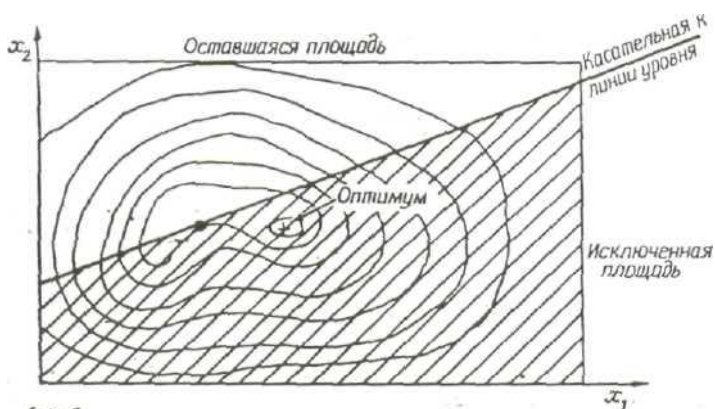


Рис.4.4.3 Метод исключения областей (касательной к линиям уровня) в случае вогнутых линий уровня.

для случая пространства проектирования, определяемого двумя переменными. Исходную область неопределенности в зависимости от размерности пространства отобразим на единичный квадрат, куб или гиперкуб. Это позволит вести поиск в нормированной области со стороны, равной единице. В гиперкубе построим сетку, образованную попарно симметричными взаимно ортогональными плоскостями, параллельными координатным направлениям, вдоль которых изменяются проектные параметры. Эти плоскости пересекаются по прямым, которые в свою очередь пересекаются в точках, называемых в дальнейшем узлами (рис. 4.4.4.). Вычислим значения целевой функции в узлах и в центре куба. В случае M проектных параметров получим 2^{M+1} значений целевой функции, из которых выберем наибольшее. Примем соответствующий узел за центр гиперкуба меньших размеров и продолжим исследование. Процесс продолжается до тех пор, пока не будет достигнута требуемая степень сужения интервала неопределенности. Если в области допустимых значений обозначить степень сужения вдоль какой-либо оси координат через r , то линейное сужение для b -мерного гиперкуба будет равно $f=r^b$, а число вычисленных значений целевой функции

$$N=b(2^M)+1$$

Мишке рекомендует выбирать r в интервале значений $2/3 < r < 1$. Он отмечает так же, что в

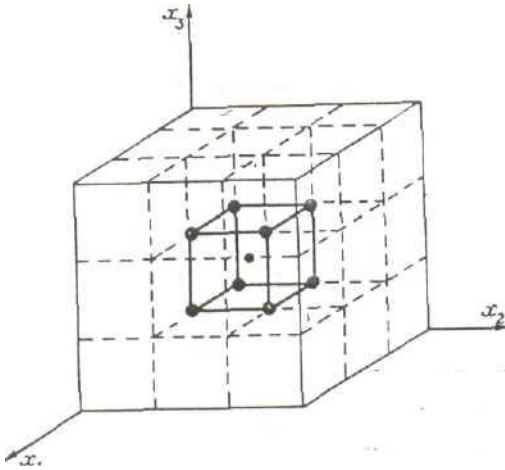


Рис.4.4.4 Сеточный метод поиска.

случае трех и более переменных большую эффективность обеспечивают не кубические, а звездообразные области.

Метод случайного поиска

Оригинальный подход, позволяющий обойти трудности применения детерминированных методов в случае многомерного пространства, предложен Бруксом и основан на случайном поиске. Пусть пространство проектирования представляет собой куб или гиперкуб со стороной, равной единице, и разделено на кубические ячейки путем деления на 10 равных частей каждой стороны куба, соответствующей одному из проектных параметров. При $N=2$ число ячеек равно 100, при $N=3$ оно равно 1000; в общем случае при N измерений число ячеек равно 10^N . Вероятность того, что выбранная наугад ячейка войдет в число 10% наиболее перспективных ячеек, равна 0,1, так как при $N=1$ нас будет интересовать одна ячейка из 10, при $N=2$ — одна из десяти лучших при общем количестве ячеек 100 и т. д. Вероятность того, что мы пропустим одну из 10% наиболее перспективных ячеек, составит 0,9. Если случайным образом выбрать две ячейки, то вероятность пропуска будет 0,92, т. е. 0,81. Вообще вероятность нахождения по крайней мере одной ячейки из наиболее перспективных, доля которых равна f , после N попыток составит

$$P=1-(1-f)^N$$

В табл. 4.4.1. указано, сколько ячеек надо выбрать случайным образом, чтобы обеспечить заданную вероятность при заданной доле наиболее перспективных ячеек. Из нее видно, что при случайной выборке 44 ячеек вероятность достижения $f=0,1$ составит 99%. Это очень неплохо, если вспомнить, что для 100%-ного обеспечения целевую функцию в случае пяти переменных пришлось бы вычислить 2 476 099 раз.

f	Вероятность			
	0,80	0,90	0,95	0,99
0,1	16	22	29	44
0,05	32	25	59	90
0,01	161	230	299	459
0,005	322	460	598	919

Метод случайного поиска имеет два преимущества. Во-первых, он пригоден для любой целевой функции независимо от того, является она унимодальной или нет. Во-вторых, вероятность успеха при попытках не зависит от размерности рассматриваемого пространства. Хотя этот метод не позволяет непосредственно найти оптимальное решение, он создает подходящие предпосылки для применения в дальнейшем других методов поиска. Поэтому его часто применяют в сочетании с одним или несколькими методами других типов.

Градиентные методы

Это наиболее часто применяемая группа методов косвенной многомерной оптимизации. Как известно, градиентом функции $F(\bar{x})$ векторного аргумента $\bar{x}(x_1, x_2, \dots, x_n)$ называется вектор, координаты которого служат частные производные по соответствующим переменным:

$$\nabla F = \left(\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \dots, \frac{\partial F}{\partial x_n} \right)$$

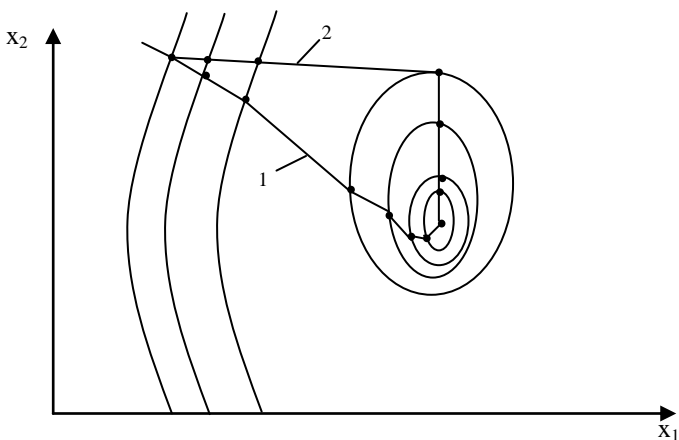
Известно, что вектор градиент совпадает с направлением наибыстрейшего возрастания целевой функции, то есть локально наилучшим является градиентное направление при максимизации или антиградиентное при минимизации. Методы поиска, в которых направления движения от итерации x^k и x^{k+1} определяется градиентом, вычисленным в точке x^k носят название градиентных методов. Градиентные методы отличаются друг от друга способом выбора шага вдоль вектора антиградиента. Простейшим способом движения из точки x^k по антиградиенту с постоянным шагом $h=\text{const}$ (ступенчатый наискорейший подъем) тоесть:

$$X^{k+1} = x^k - h \nabla F_k$$

При переходе через max начинается дробление шага. Иногда характер функции бывает достаточно сложным и частные производные вычисляются приближенно:

$$\frac{\partial F}{\partial x_i} = \frac{F(x_1, x_2, \dots, x_i + \Delta, \dots, x_N) - F(x_1, x_2, \dots, x_N)}{\Delta}$$

Основной недостаток метода - большой объем вычислений как значений самой функции, так и её производной на каждом шаге. В ряде случаев значительно более эффективным может оказаться так называемый метод "наискорейшего спуска". В этом методе производная вычисляется только на первом шаге. Дальнейшее движение осуществляется по этому направлению и величина шага рассчитывается с помощью методов одномерной оптимизации до определения оптимального значения функции на данном направлении. Затем вычисляется градиент в полученной точке и процедура повторяется. На рис 4.4.5. представлен ход процедуры поиска методом градиента (1) и методом "наискорейшего спуска" (2) и хотя количество шагов во втором случае может быть больше, за счет меньшего количества вычислений градиента, он может оказаться более эффективным.



Эффективность градиентных методов существенно снижается если гиперповерхность постоянного уровня вытянута в каком-либо направлении (овражная функция). При этом длина шага может оказаться сравнимой с погрешностями вычисления функций. Для повышения эффективности поиска наряду с градиентом используют так же матрицу вторых производных (гессиана), с помощью которой удастся выбирать направление поиска в большей мере соответствующего геометрии критерия оптимальности. Это значительно увеличивает объем вычислений в том числе и за счёт увеличения числа итераций.

СИМПЛЕКС-МЕТОД

Изложение этого метода начнем с пояснения того, что такое симплекс. Симплексом называется N -мерная замкнутая геометрическая фигура, ребра которой представляют собой прямые линии, пересекающиеся в $N+1$ вершине. В двумерном случае это треугольник, в трехмерном — тетраэдр. Схемы поиска с использованием симплексов основаны на слежении за изменением значений целевой функции в их вершинах. Главным в этих схемах является процесс отражения — нахождение вершины нового симплекса, расположенной симметрично относительно плоскости, проходящей через одну из сторон исходного симплекса. Выбор направления поиска вершины нового симплекса определяется положением той вершины исходного симплекса, в которой целевая функ-

ция имеет наихудшее значение (рис. 4.4.6.). Новая точка называется «дополнением» наихудшей точки. Если в только что полученной вершине нового симплекса значение целевой функции оказывается худшим, то алгоритм предусматривает возврат в исходную точку — вершину прежнего симплекса. Затем осуществляется переход к той вершине прежнего симплекса, в которой целевая функция имеет следующее по величине значение, и отыскивается точка, являющаяся её дополнением. Такой алгоритм обеспечивает систематическое смещение центра симплекса в направлении экстремума целевой функции.

Известен и более сложный метод -метод Нелдера-Мида, в котором помимо поиска вершин новых симплексов производится сжатие или растяжение их ребер. Этот алгоритм обладает большей точностью и обеспечивает локальное преобразование пространства проектирования, при котором достигается минимум унимодальной функции вида

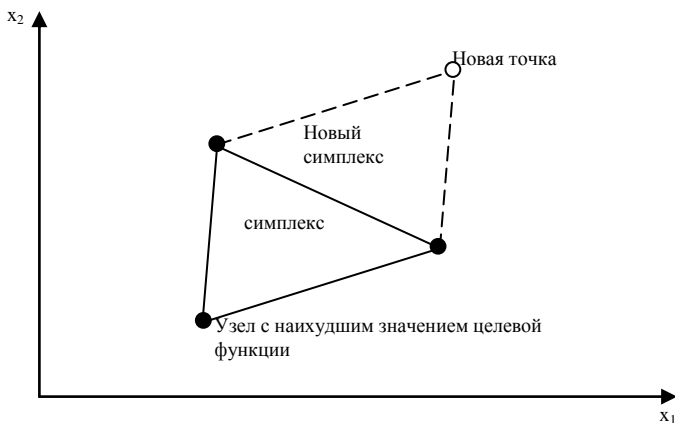
$$M = F(x_1, x_2, \dots, x_N)$$


Рис. 4.4.6. Симплекс-метод в двумерном пространстве