АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОРГАНИЗАЦИЯ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ИНСТИТУТ ЭКОНОМИКИ И УПРАВЛЕНИЯ»

Е.Н. Надеждин, Е.Е. Смирнова, В.С. Варзаков

МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ В ЭКОНОМИКЕ

УЧЕБНОЕ ПОСОБИЕ

Тула 2011

УДК 330.4(075.8)

ББК 65в631я73

Н 171

Надеждин Е.Н., Смирнова Е.Е., **Варзаков В.С.**
Математические методы и модели в экономике: учебное пособие для студентов экономических специальностей.- Тула: Автономная некоммерческая организация ВПО «Институт экономики и управления», 2011. - 249 с.

Учебное пособие посвящено изложению методов исследования операций и основ экономико-математических моделей применительно к задачам анализа экономических процессов и обоснования управленческих решений.

Учебное пособие разработано в соответствии требованиями ФГОУ по направлениям бакалавриата 080100 «Экономика», 080200 «Менеджмент» и 080500 «Бизнес-информатика» и предназначено для студентов, изучающих дисциплины «Методы моделирования и прогнозирования в экономике», «Экономико-математические методы и модели в логистике» и «Исследование операций». Пособие может быть использовано студентами при выполнении выпускных квалификационных работ, а также работниками аналитических служб предприятий в практической деятельности.

ISBN 978-5-9902636-4-2

Рецензенты: заместитель директора учреждения Российской академии образования «Институт информатизации образования», доктор педагогических наук, кандидат технических наук, профессор **О.А. Козлов;**
профессор кафедры автоматики и телемеханики Тульского государственного университета, доктор физико-математических наук, профессор **С.Д. Двоенко.**

ISBN 978-5-9902636-4-2

ОГЛАВЛЕНИЕ

Список основных сокращений

ЗЛП - задача линейного программирования;
ЗМП - задача математического программирования;
ЗНП - задача нелинейного программирования;
ЗПР - задача принятия решения;
ЗРВ - закон распределения вероятностей;
ИМ - имитационная модель;
ИО - исследование операций;
ЛПР - лицо, принимающее решение;
МДП - метод динамического программирования;
ММ - математическая модель;
МСИ - метод статистических испытаний;
ОС - обратная связь;
ОУ - объект управления;
ППП - пакет прикладных программ;
ПР - принятие решения;
ТМО - теория массового обслуживания;
СГ - сетевой график;
СМО - система массового обслуживания;
СПУ - сетевое планирование и управление;
СППР - система поддержки принятия решений;
ЦФ - целевая функция;
ЭС - экономическая система.

ВВЕДЕНИЕ

В настоящее время специалист (руководитель, менеджер, экономист, аналитик) в области экономики является одним из важных звеньев системы управления (предприятием, фирмой), от его профессиональной компетенции и инициативы во многом зависят стабильность экономического развития организации, перспективы бизнеса. Именно поэтому в условиях рыночных отношений важен высокий уровень фундаментальной подготовки выпускника экономического вуза. Освоение на хорошем прикладном уровне математического аппарата исследования операций, безусловно, будет способствовать быстрому профессиональному становлению в бизнесе выпускников АНО ВПО «Институт экономики и управления».

Актуальность издания учебного пособия определяется рядом положений, среди которых выделим следующие.

1. Возникновение и **быстрое развитие организаций** (предприятий, фирм) различных форм собственности, для успешного управления которыми в новых экономических условиях руководитель должен иметь хорошую фундаментальную подготовку в области математической статистики, прикладной математики и информационных технологий; владение аппаратом теории исследования операций позволяет руководителю (специалисту) корректно формулировать и оперативно решать прикладные задачи, связанные с оценкой эффективности системы управления, обоснованием и выбором оптимальных управленческих решений.

2. **Недостаток учебной литературы**, в которой были бы сбалансированы вопросы теории и практики исследования операций применительно к сфере экономики; перенос акцентов только в одну сторону в равной степени снижает потребительскую ценность представленного материала; подробное рассмотрение всех основных аспектов формализации экономических задач (от постановки до компьютерного решения) с анализом полученного решения, по нашему мнению, отвечает современным потребностям студентов.

3. **Появление** на рынке информационных технологий **нелицензионных программных продуктов**, отличающихся качеством формального представления экономических процессов и функциональными возможностями с точки зрения описания проблемных ситуаций в экономике; необоснованный выбор программных средств для решения прикладных задач может означать дополнительные материальные и временные затраты.

4. Определенная **сложность самостоятельного изучения** начинающими пользователями (студентами) унифицированных **пакетов прикладных программ** и программных средств; основные трудности здесь обусловлены недостатком опыта (умений и навыков) успешного применения программных продуктов и технологий при математической постановке задач исследования; переход от содержательной постановки задачи к её математическому представлению (модели) является нетривиальным и в настоящее время не может быть автоматизирован полностью.

5. **Наличие научно-методического задела** в виде общих методологических знаний и умений творческой деятельности, личного опыта в области экономики и прикладной информатики, позволяющих авторам в доступной для студентов форме изложить вопросы формализации прикладных задач в различных предметных областях: экономике, технике, кибернетике.

В качестве **объекта исследования** в пособии рассматриваются реальные экономические системы и процессы, характеризующиеся множеством параметров, факторов и условий. В задачах анализа, планирования, оптимизации исследуются реальные (или близкие к ним) конфликтные ситуации, используются конкретные статистические данные, которые заимствованы из официальных источников информации.

Предмет исследования определяют, с одной стороны, операционные модели, являющиеся формальным представлением экономических процессов, систем или их взаимодействия, и, с другой стороны, известные методы и алгоритмы исследования операций с использованием информационных технологий.

Основные задачи учебного пособия:

1. Изложение в доступной для понимания форме базовых понятий и принципов исследования операций, лежащих в основе построения и компьютерной реализации задач оценки, выбора и оптимизации параметров и характеристик существующих экономических процессов и систем;
2. Изложение теоретических основ методов исследования операций в объёме, необходимом и достаточном для самостоятельного изучения проблемных ситуаций, постановки и решения прикладных задач, обоснования и выбора оптимальных управленческих решений;
3. Показ на конкретных примерах методических особенностей компьютерного решения стандартных и нестандартных прикладных задач с применением аппарата теории исследования операций;
4. Ознакомление с методикой формализации прикладных задач на основе применения современных методических, программных и инструментальных средств и ППП.
5. Изложение методов исследования операций по следующей схеме «проблемная ситуация - теоретический анализ и обоснование – операционная модель задачи - метод решения – вычислительный алгоритм – прикладная задача - программная реализация»; реализация идея спиралевидной модели обучения студентов на примерах решения прикладных задач.

В основу материалов учебного пособия положены фундаментальные работы отечественных и зарубежных учёных: Н.Н. Моисеева, Е.С. Вентцель, В.Н. Волковой, Э. Мушика, Т. Саати, Х.А. Таха и др. - в областях системного анализа и исследования операций, а также методические материалы и многолетний опыт авторов, приобретенный при чтении лекций для студентов экономических и технических специальностей.

Структурно учебное пособие в к л ю ч а е т: введение, пять глав, заключение, список использованной литературы и два приложения.

В ***первой главе*** дана характеристика системного направления исследования сложных систем – теории исследования операций. Введены базовые понятия, раскрыты основные принципы и этапы операционного исследования сложных систем. Изучен предмет теории исследования операций. Приведена классификация задач исследования операций и даны примеры типовых проблемных ситуаций в экономике, предполагающих использование аппарата исследования операций.

Вторая глава посвящена изложению теории и методов решения задач математического программирования. Представлена формулировка обобщённой задачи математического программирования и проведена классификация задач математического программирования с учётом особенностей предметной области. Принимая во внимание широкое использование на практике линейных моделей экстремальных задач, в пособии подробно изложены такие методы линейного программирования как графический метод и симплексный метод. Приведены примеры решения этих задач в различных постановках.

В ***третьей главе*** представлены специальные экономико-математические модели. Изложена сущность моделей межотраслевого баланса и даны примеры их использования в экономическом анализе.

Четвертая глава посвящена стохастическому подходу к исследованию экономических процессов. Для построения стохастических моделей предлагаются марковские модели, аппарат теории массового обслуживания, метод статистических испытаний. Основное внимание уделено вопросам обоснования корректной математической формулировки (модели) задачи исследования. Представлены основные допущения, которые вводятся при формализации оптимизационных задач в рамках соответствующей математической схемы.

В **пятой главе** нашли отражение вопросы построения и анализа моделей задач принятия решений в условиях неопределённости. Здесь проведена классификация факторов, которые являются источниками неопределённостей при функционировании экономических систем. Рассмотрены задачи управления экономическими объектами в игровой постановке. Введены базовые понятия и дана классификация математических игр. Представлены основные критерии теории максимина: критерий Вальда, критерий Севиджа, критерий Гурвица и др. Показаны примеры использования критериев в задачах принятия решений.

В ***приложениях*** представлены глоссарий предметной области и задания для самостоятельной работы студентов.

В интересах наиболее глубокого усвоения материала и получения студентами навыков в применении методов моделирования и математического программирования к анализу экономических процессов внимание акцентируется на вопросах формализации и компьютерного решения прикладных задач с помощью офисных программ и унифицированных пакетов прикладных программ.

Учебное пособие подготовлено творческим коллективом авторов в составе: доктора технических наук, профессора **Е.Н. Надеждина** (введение, главы 1, 4 и 5, § 3.1), кандидата педагогических наук **Е.Е. Смирновой** (главы 2, 4 и 5, § 3.2 и приложения 1 и 2) и кандидата технических наук **В.С. Варзакова** (глава 3).

ГЛАВА 1. ПРЕДМЕТ И ЗАДАЧИ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ

1.1 Проблема принятия решения и её эволюция

Процесс принятия решения не нов - им интересовались еще жрецы и мудрецы глубокой древности, светила античных времен, его изучают и современные ученые. За всю историю своего существования люди пользовались различными способами принятия решения. Поэтому неудивительно, что с древних времен до настоящего времени проблемы решения входят в число актуальных проблем науки. При этом наибольшее внимание обращается на решение нестандартных задач, которые носят творческий характер.

Информационные источники, дошедшие до нас с древнейших времен, свидетельствуют, что уже тогда людей волновал вопрос, каким образом происходит принятие решения, особенно творческого, каков рациональный путь к выбору правильного поступка при новых, ранее не встречавшихся обстоятельствах, каким образом люди создают новое и полезное. Например, известно пособие по принятию решения, которое написал около четырех тысяч лет назад в Китае *И. Чинг*. Это пособие явилось оригинальным руководством в творческой работе. По мнению некоторых зарубежных специалистов (*Ф.Д. Баррет* и др.) известный трактат не потерял своего значения до наших дней и может быть рекомендовано современным хозяйственным руководителям и политическим деятелям (речь идет об управлении производством).

Несомненно, в те далекие времена научные данные о природе творческой деятельности, творческих решений были настолько скудны, что они не могли дать людям правильного объяснения этим процессам. Не находя правильного ответа на вопросы, люди объясняли их мистически. Так, древние греки и другие народы античного мира объясняли как силы природы, так и творческие способности их божественным происхождением. При таком объяснении основным орудием управления процессом принятия решения в неопределенных обстоятельствах или процессами творчества были молитвы музам - богиням творчества.

Однако и в те далекие времена люди искали более рациональные пути улучшения процесса принятия решения, особенно для таких случаев, когда было неизвестно, как изменятся обстоятельства. В ту пору предсказаниями будущего славился греческий город Дельфы со своим оракулом.

Вкратце суть этих предсказаний такова. Сначала с обстоятельствами дела досконально знакомились дельфийские жрецы. Затем они наблюдали за прорицательницей (*Пифией*), одурманенной выходящими из земли газами. Пифия выкрикивала отдельные слова и бессвязные фразы, вызывавшие у жрецов ассоциации, на основе которых они предсказывали будущее. Предсказание передавалось заинтересованному лицу или обнародовалось только после тщательного его обсуждения на совете.

Несмотря на многовековую историю изучения процесса решения и его принятия, термин «принятие решения» появился в научной литературе в 30-е годы XX-го столетия - в творческих работах по управлению частным производством для характеристики процессов децентрализации управления.

Современный этап развития науки управления характеризуется интенсивным поиском новых идей, подходов, методов и средств, способных повысить эффективность управления сложными социально-экономическими системами в условиях все возрастающего динамизма процессов, усложнения связей в системах и все более жесткого ограничения ресурсов. Одним из перспективных подходов является рассмотрение проблем управления с позиций принятия решений. Этот подход, связанный с использованием таких категорий, как «решение», «процесс принятия решений», «система принятия решений», в литературе называют концепцией принятия решений, теорией принятия решений, школой принятия решений. Актуальность и практическая ценность выделения и изучения проблем принятия решений в процессе управления определяются следующими причинами.

Во-первых, принятие решений занимает центральное место в процессе управления. Как известно, принятие решений наряду с прогнозированием, планированием, оценкой обстановки, исполнением решений, контролем и учетом является функцией управления. Центральная, важнейшая роль принятия решений определяется тем, что все другие функции управления направлены на формирование или реализацию решений. Кроме того, любую функцию управления технологически можно представить в виде последовательности решений. Например, при прогнозировании и планировании принимаются решения, связанные с выбором методов и средств, организацией работ, оценкой достоверности информации, выбором наиболее достоверного варианта прогноза и наилучшего варианта плана. Аналогичную цепочку решений можно построить и при рассмотрении других функций управления. Таким образом, функция принятия решений является с методологической и технологической точек зрения более общей, чем другие функции управления, поэтому в литературе иногда управление рассматривается как процесс принятия решений.

Во-вторых, принятие решений - это личная функция руководителя. Для руководителя любого ранга принятие решений является основной задачей, которую он обязан решать в процессе управления. Поэтому знание методов, технологии и средств решения этой задачи является необходимым элементом квалификации руководителя.

В-третьих, современной моделью функционирования организационных систем является система принятия решений. Эта модель согласует положительные стороны двух предшествующих теоретических моделей, имеющих структурно-описательный характер: механической модели организации как «полностью рациональной» системы; естественной модели организации как «живой» социальной системы.

Система принятия решений является третьей моделью организационных систем, в которой как первичный элемент рассматривается «решение». Основное развитие эта модель получила на рубеже 60-х годов прошедшего века. В данной модели рассматриваются рациональные принципы механической модели с учетом социальной и психологической специфики естественной модели. В решении объединяются объективные факторы информационного анализа проблем, проводимого на основе логического мышления, математических методов и ЭВМ, и субъективные психологические факторы лица, принимающего решение. Поэтому система принятия решений позволяет осуществить системный подход к исследованию сложных организационных систем, включающий социально-технологическую форму реализации процессов управления.

В-четвертых, подход, ориентированный на принятие решений, создает прочную базу для дальнейшего совершенствования автоматизированных систем информационного обеспечения и управления. Эти системы должны развиваться от автоматизации трудоемких рутинных учетно-расчетных задач до логико-аналитических задач формирования и обоснования вариантов решений. Рассмотрение организационной системы принятия решений, выделение в ней центров принятия решений, формирование информационных потоков, соответствующей структуры и их динамической перестройки в зависимости от решаемых проблем являются перспективным направлением совершенствования управления сложными социально-экономическими системами.

Изложенное показывает, что изучение проблем управления с единой методологической позиции принятия решений позволяет практически осуществить комплексный, системный подход к анализу функционирования организационных систем, технологии их управления, применению современных методов и технических средств.

В настоящее время существует достаточно большое, число современных научных дисциплин, посвященных проблеме принятия решений. К таким дисциплинам можно отнести и исследование операций.

Ситуацию, в которой происходит принятие решений, в общем случае характеризуют следующие основные черты.

1. *Наличие цели (целей).* Необходимость принятия решения диктуется наличием некоторой цели, которую нужно достичь, например: выполнить плановое задание, выбрать тип прибора, назначить план перевозок и т. д. Если же цель не поставлена, то не возникает и необходимость принимать какое-либо решение.

2. *Наличие альтернативных линий поведения.* Решения принимаются в условиях, когда существует более одного способа достижения цели, или иначе несколько альтернатив достижения цели. С различными альтернативами могут быть связаны различные затраты и разные вероятности достижения цели. Эти затраты и вероятности не всегда могут быть определены. Поэтому часто принятие решений сопряжено с неясностью и неопределенностью. Если же существует лишь одна линия поведения, то выбора нет и, следовательно, решение принимать не требуется: оно очевидно.

3. *Наличие ограничивающих факторов.* Решения обычно принимаются в условиях действия большого числа факторов, ограничивающих возможность выбора способов действий. Эти факторы называют *дисциплинирующими условиями*. Факторы, подлежащие учёту, можно условно разделить на три основные группы: экономические, технические и специальные. Под *экономическими* факторами понимают факторы, связанные с ресурсами: время, денежные средства, трудовые ресурсы, производственные возможности и т. п.

К *техническим* факторам обычно относят факторы, которые непосредственно связаны с инженерным анализом и выработкой требований к техническим характеристикам объектов: габаритам, массе, прочности, надежности и т. п. Наконец, *социальные* факторы, в том числе и чисто человеческие, выражают требования не только политической или социальной целесообразности осуществления той или иной альтернативы, но и этики.

Все перечисленные факторы накладывают ограничения на возможности достижения поставленной цели. Очевидно, что отсутствие ограничений существенно упрощает задачу принятия решения.

Таким образом, задача принятия решения (ЗПР) возникает в том и только в том случае, когда существует цель, которую нужно достичь, когда возможны различные способы ее достижения и существуют факторы, ограничивающие возможности достижения цели. Выяснение всех трех указанных элементов задачи принятия решений должно обязательно предшествовать её непосредственному решению. Во всех случаях задача принятия решений направлена на определение наилучшего (оптимального) или приемлемого способа действий для достижения одной или нескольких целей.

Под ***целью*** понимается в широком смысле идеальное представление желаемого состояния или результата деятельности. Если фактическое состояние не соответствует желаемому состоянию, то имеет место *проблемная ситуация*, или проблема, выработка плана устранения которой и составляет сущность задачи принятия решений.

Конечным результатом ЗПР является *решение*. Решение можно рассматривать как предписание к действию. С точки зрения содержания решением может быть стратегия управления экономической системой, способ действия, план работы, вариант инновационного проекта и т. д.

1.2 Основные понятия исследования операций

Одним из основных понятий теории принятия решений является *операция*. Под термином «***операция***» следует понимать *организованную деятельность в любой области жизни, объединенную единым замыслом, направленную на достижение определенной цели, имеющую характер повторяемости*, т.е. многократности. В дальнейшем ***операцией*** будем называть управляемое мероприятие, систему действий, объединённых единым замыслом и направленное на достижение какой-либо конкретной цели.

В данных формулировках подчеркиваются две особенности операции: её целевая направленность и повторяемость. Именно отсюда возникает возможность проводить исследования, касающиеся количественных сторон операции, общими научными путями с использованием методов теории вероятностей, статистики, и данных различных наук - физики, биологии, техники, экономического анализа и др.

Укажем примеры операций: а) производственная деятельность отрасли, выпускающей некоторую народнохозяйственную продукцию; б) формирование портфеля заказов фирмы; в) разработка плана транспортных перевозок материальных средств; г) совокупность мероприятий, направленных на реализацию бизнес-плана и т. д.

Изучение операций может проводиться как путем исследования оригинала самой операции, так и путем исследования *модели операции*. Основным методом исследования операций, особенно крупного масштаба, является метод моделирования. Получить аналитическое описание операции удаётся только в самых простых случаях. Постановка специальных экспериментов на реальных экономических системах с целью выбора оптимальных решений обычно сложна, связана с большими расходами и непредсказуемыми последствиями или просто нереальна. На практике значительно проще и рациональнее изучить закономерности и построить приближённую математическую модель экономического процесса и на основе изучения модели выбрать лучшее решение, отвечающее множеству нередко противоречивых требований и условий.

Таким образом, основной метод изучения операций крупного масштаба - исследование моделей операции, главным образом, моделей математических.

Второе важное понятие исследования операций – *оперирующая сторона*. Совокупность лиц, которые стремятся в данной операции к достижению некоторой цели, а также технических устройств, с помощью которых цель достигается, называется ***оперирующей стороной***.

В операции могут участвовать одна или несколько оперирующих сторон, преследующих различные, несовпадающие цели. Несовпадение целей оперирующих сторон создаёт конфликтную ситуацию. Подобные операции называются многосторонними или конфликтными. Так, например, в торговле исход операции зависит от деятельности двух сторон, преследующих противоположные цели: фирмы, стремящейся продать товар с прибылью, организации, приобретающей товар заданного качества по минимальной цене. Наряду с оперирующими сторонами в операции могут участвовать арбитры и природные силы, поведение которых в явном виде не подчинено стремлению к достижению цели операции.

Для достижения цели оперирующая сторона должна располагать некоторым запасом активных средств (ресурсов), используя или расходуя которые она может добиваться достижения цели. В качестве ресурсов в зависимости от сущности операции могут выступать: запасы сырья, рабочая сила, денежные средства, информационные и интеллектуальные ресурсы, торговые площади и т. п.

Операция является управляемым мероприятием. Оперирующая сторона управляет операцией, выбирая те или иные способы использования ресурсов – способ действий. В качестве синонимов термина «способ действия» часто используют следующие термины: альтернатива, стратегия, управление, решение. Возможности оперирующей стороны по управлению операцией всегда ограничены рядом естественных причин. Этот факт проявляется в наличии *ограничений* - дисциплинирующих условий – на выбор способов действий оперирующей стороны - стратегий. Стратегии, удовлетворяющие наложенным ограничениям, называются *всевозможными* или *до-*

пустимыми (в смысле заданных ограничений). Понятие «допустимые стратегии» является относительным: класс допустимых стратегий определяется наложенными ограничениями и изменяется, если изменяются ограничения.

Реализация той или иной допустимой стратегии оперирующей стороны обычно приводит к различным исходам операции. Чтобы сравнивать между собой качество различных стратегий, нужно иметь возможность оценивать соответствующие исходы операций. Исход операции оценивается с помощью некоторых критериев качества (критериев эффективности или критериев оптимальности). Критерий оптимальности является математическим выражением цели операции (математической моделью цели операции) позволяющим количественно определить (оценить) степень достижения этой цели. Стратегия, наилучшая в смысле выбранного критерия оптимальности, т. е. доставляющая ему требуемое экстремальное (максимальное или минимальное) значение, называется оптимальной стратегией. Синонимами этого термина являются термины «оптимальное решение», «оптимальное управление» и т. п.

Следует всегда иметь в виду, что понятие «оптимальная стратегия» является не абсолютным, а относительным, как и понятие «допустимая стратегия». Не существует оптимальной стратегии вообще, всякая оптимальная стратегия является наилучшей лишь в некотором узком, совершенно конкретном смысле, определенном критерием оптимальности. Одна и та же стратегия, оптимальная в смысле одного критерия, может оказаться далеко не оптимальной и даже очень плохой по другому критерию.

Поскольку значение критерия оптимальности в любой операции зависит от каких-либо величин, описывающих свойства операции, используемые ресурсы и т. д., то критерий оптимальности часто называют также *критериальной* или *целевой функцией* (функцией эффективности).

Следующее важное понятие исследования операций – *исследователь операции*. В составе оперирующей стороны специально выделяется и занимает особое место исследователь операции, или ***операционист***. Он принадлежит к оперирующей стороне и должен преследовать ту же цель, что и оперирующая сторона. Однако операционист не принимает окончательных решений по выбору способов действий, а лишь помогает в этом оперирующей стороне, предоставляя ей количественные основания для принятия решений. Иными словами, исследователь операции имеет право не решающего, а лишь совещательного голоса. Естественно, что поэтому он не должен нести ответственности за принятые решения и последствия от реализации предпринятых действий.

Суть работы исследователя операции состоит в детальном изучении сущности и специфики решаемой проблемы, определении всего набора допустимых стратегий, оценке их качества, сравнении их между собой и определении оптимальной стратегии. Исследование операции завершается рекомендациями по выбору оптимальной стратегии. Само же принятие решения, т. е. окончательный выбор стратегии и её реализация, выходят за рамки исследования и относится к компетенции ответственного лица - *руководителя операции* (РО).

Уточним содержание исследований, формирующих процесс принятия решений. Процессы принятия решений, реализуемые в самых различных сферах деятельности, имеют очень много общего, поэтому желательно создать некоторую универсальную, типовую схему, устанавливающую наиболее целесообразный набор и последовательность действий, проводимых при исследовании операций. В работах многих авторов по исследованию операций, системному анализу, управлению производством содержатся рекомендации по формированию состава и последовательности исследований в процессе принятия решений.

На основе анализа и обобщения этих рекомендаций можно предложить следующий *алгоритм типового процесса принятия решений*:

1) предварительное формулирование проблемы;

2) определение целей операции и выбор соответствующих критериев оптимальности;

3) выявление и формулирование дисциплинирующих условий;

4) составление возможно более полного списка альтернатив и предварительный их анализ с целью отбрасывания явно неэффективных;

5) сбор необходимой информации и прогнозирование изменений параметров операции в будущем;

6) точное формулирование постановки задачи;

7) разработку математической модели операции, позволяющей оценивать эффективность каждой альтернативы;

8) выбор метода решения задачи и разработка алгоритма решения;

9) оценку альтернатив и определение наиболее эффективных;

10) принятие решения ответственным руководителем;

11) выполнение решения и оценку результатов.

<u>Процесс принятия решений является сложной итеративной циклической процедурой</u>. Действительно, результат практически любого этапа исследований может повлиять на постановку задачи и привести к её изменению. В частности, даже практическое опробование принятого решения, если оно дает нежелательный результат, также является стимулом к пересмотру постановки задачи и поиску новых решений.

Структурная схема типового процесса исследования операций представлена на рис. 1.1. Указанные выше особенности процесса отображены с помощью обратных связей.

Отметим, что постановка задачи принятия решений индивидуальным лицом, принимающим решение (ЛПР), компактно может быть представлена в следующей символической форме:

$$\boxed{(S_0,\ T,\ R \parallel S,\ W, G,\ Y, f,\ P)} \tag{1.1}$$

т. е. в условиях исходной проблемной ситуации S_o, располагаемого времени T и ресурсов R необходимо доопределить ситуацию S_o множеством ситуаций S, сформулировать множество целей W, ограничений G, допустимых решений Y, произвести оценку предпочтения решений f и найти оптимальное или приемлемое решение Y из множества решений, руководствуясь сформулированным критерием выбора P.

Ранее было отмечено, что в процессе принятия решений осуществляется преобразование информации. На различных этапах процесса принятия решений происходит количественная и качественная переработка исходной информации с целью получения новой информации, концентрированное выражение которой находит свое отражение в решении. Переработка информации происходит в виде последовательных трёх фаз, для каждой из которых характерен свой уровень определенности решения задачи: *1) структуризация; 2) характеризация; 3) оптимизация.*

Структуризация заключается в выделении основных элементов решаемой задачи и установлении отношений между ними. В результате структуризации образуется логически упорядоченная система, позволяющая определить необходимую информацию и распределить её по элементам структуры. Результат структуризации отображается в виде схем, таблиц или формальной символической записи. Примерами структуризации являются построение дерева целей, формирование множества альтернативных решений и способов их реализации, установление взаимосвязанной системы гипотез о возможных событиях и т. п.

Характеризация заключается в определении системы характеристик, параметров и показателей, количественно описывающих решаемую задачу. Например, для дерева целей могут быть определены приоритеты, или относительные коэффициенты важности целей: для множества решений - предпочтения (полезности), для гипотез о возможных событиях - вероятности их свершения и т. д. Выполнение характеризации приводит к более полному и точному описанию решаемой задачи по сравнению с фазой структуризации и подготавливает исходные данные для выполнения оптимизации.

Оптимизация представляет собой поиск наилучшего решения задачи. Именно на этой фазе вся имеющаяся информация преобразуется в конечную форму, описывающую решение и показывающую, как зависят характеристики решения от исходных данных. Проведение оптимизации приводит к полной определенности решения задачи. Для решения задач в условиях неопределенности не всегда возможно проведение фазы оптимизации в строго формальном виде. Во многих случаях лицо, принимающее решение, (ЛПР) осуществляет оптимизацию в неявном виде, опираясь на некоторые общие принципы, профессиональный опыт и свои предположения.

Процесс принятия решения (ПР) требует полноценного информационного обеспечения, которое предполагает использование как имеющейся информации (априорной информации), так и получение, добывание с помощью экспериментов (опытов) новой послеопытной (апостериорной) информации.

Обобщенной характеристикой решения является *эффективность решения*. Эта характеристика включает эффект решения, определяющий степень достижения целей и стоимость решения - совокупность затрат ресурсов для принятия и реализации решения. Эффективность решения - это степень достижения целей, отнесенная к затратам на их достижение. Решение тем эффективнее, чем выше степень достижения целей и ниже стоимость затрат. ***Система принятия решений*** - организованная совокупность людей, методов, технических средств, информации и технологии ПР для достижения поставленных целей.

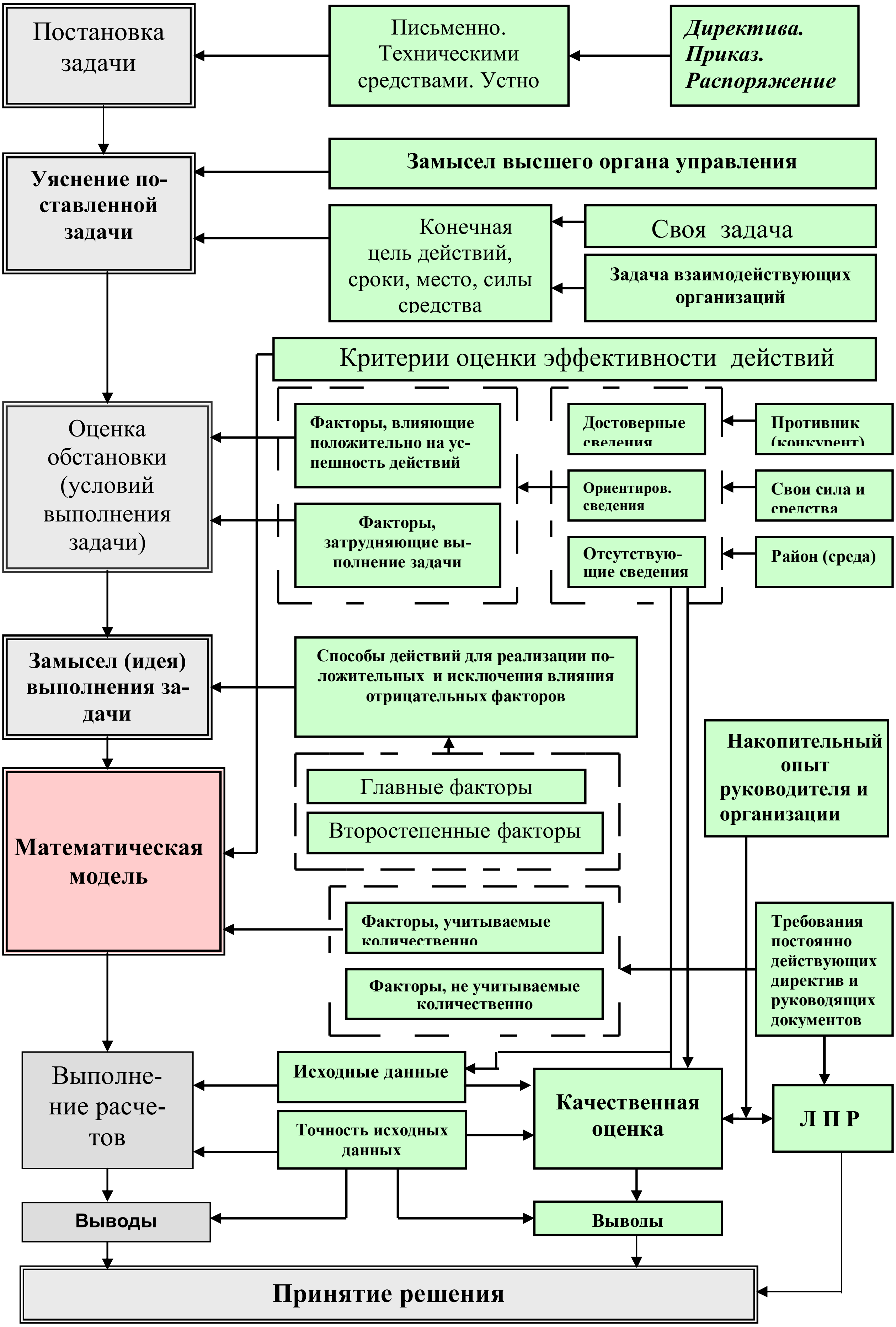


Рис. 1.1 - Структура типового процесса принятия решения

1.2.1 Классификация задач принятия решений

Классификацию ЗРП осуществим на основе опыта, накопленного в области технической и экономической кибернетики с учётом особенностей формальной модели (1.1). Для этого выделим следующие отдельные, важные *классификационные признаки* (рис. 1.2).

1. Число целей операции, преследуемых одной оперирующей стороной.

2. Наличие или отсутствие зависимости критерия оптимальности и дисциплинирующих условий от времени.

3. Наличие случайных и неопределенных факторов, влияющих на исход операции. Этот признак назван «определенность - риск - неопределенность».

По первому классификационному признаку ЗПР делятся на два больших класса: одноцелевые, или однокритериальные (скалярные), и многоцелевые, или многокритериальные, ЗПР.

Отметим, что многокритериальные задачи - задачи, которые состоят в поиске лучшего решения, удовлетворяющего нескольким несводимым друг с другом критериям. Для решения таких задач требуется их формализация, которая неизбежно связывается с экспертными оценками как самих критериев, так и взаимоотношений между ними. Термин «многокритериальные задачи» нельзя смешивать с термином «задачи векторной оптимизации»: в последнем случае речь идет о сопоставлении однородных критериев разных участников (такие задачи решаются с помощью критерия оптимальности, по Парето). Нельзя также оба эти термина смешивать с термином «многоэкстремальные задачи», для которых характерны не разные критерии, а наличие у целевой функции не только глобального, но и локальных экстремумов.

По второму классификационному признаку ЗПР делятся на два больших класса: статические и динамические.

В *статических* ЗПР критериальная функция и функция ограничений не зависят от времени. В статической экономико-математической модели значения всех переменных относятся к одному моменту времени. Статический подход к изучению экономики - это изучение состояния экономики на данный момент (статики) в отличие от динамического подхода, при котором экономика исследуется в развитии и движении. Динамические задачи сложнее статических. Модель является динамической, если как минимум одна её переменная относится к периоду времени, отличному от времени, к которому отнесены другие переменные.

Динамические задачи отличают две характерные особенности:

а) в качестве критерия оптимальности в динамических ЗПР выступает обычно не функция, как в статических ЗПР, а функционал, зависящий от функции времени, описывающий поведение некоторых динамических объектов, участвующих в операции;

б) в числе дисциплинирующих условий в динамических ЗПР обычно присутствуют так называемые дифференциальные связи. Они представляют собой дифференциальные уравнения, описывающие поведение динамических объектов, участвующих в операции.

Для математического описания динамических моделей используются системы линейных дифференциальных уравнений (в моделях с непрерывным временем), разностные уравнения (в моделях с дискретным временем), а также системы обыкновенных алгебраических уравнений.

В качестве примера динамической ЗПР можно привести задачу вывода летательного аппарата в заданную точку пространства с заданной точностью и за заданное время с минимальным расходом топлива. В настоящее время динамические ЗПР ещё не получили широкого применения в экономических исследованиях по причине сложности получения достоверных данных о бизнес-среде.

По третьему классификационному признаку - «определенность - риск - неопределенность» - ЗПР делятся на три больших класса.

1. Принятие решений *при определенности* (детерминированные ЗПР). Такие ЗПР характеризуются однозначной, детерминированной связью между принятым решением и его исходом. Это наиболее простой и наиболее изученный случай принятия решений, когда относительно каждой стратегии оперирующей стороны заранее, до проведения операции, известно, что она неизменно приводит к некоторому конкретному результату. В детерминированных ЗПР критерий оптимальности и дисциплинирующие условия зависят только от стратегий оперирующей стороны и фиксированных детерминированных факторов, т. е. от факторов, полностью известных оперирующей стороне.

2. Принятие решений в *условиях риска* (стохастические ЗПР). В этом случае каждая стратегия оперирующей стороны может привести к одному из множества возможных исходов, причём каждый исход имеет определённую вероятность появления. Предполагается, что ЛПР эти вероятности заранее, до определения операции, полностью известны. В некоторых случаях эти вероятности могут быть определены с любой требуемой для целей исследования степенью точности. В стохастических ЗПР критерий оптимальности зависит кроме стратегий оперирующей стороны и детерминированных факторов также от фиксированных стохастических факторов, т. е. от случайных факторов, законы распределения вероятностей которых априорно известны оперирующей стороне. Статистические характеристики (законы распределения, математические ожидания, дисперсии и т. п.) стохастических факторов, а также значения детерминированных факторов являются той исходной информацией, которая может быть использована исследователем операции при определении оптимальной стратегии.

Первое из приведенных названий рассматриваемого класса ЗПР - принятие решений при риске - связано со следующими обстоятельствами. Все случайные явления и процессы, сопровождающие операцию и влияющие на ее исход, могут быть хорошо изучены, и все их необходимые статистические характеристики полностью известны, однако исход каждой конкретной реализации операции заранее (до её проведения) неизвестен, случаен. В этом смысле оперирующая сторона всегда рискует (в большей или меньшей степени) получить не тот результат, на который она ориентируется, выбирая свою оптимальную стратегию в расчете на осредненные, статистические характеристики случайных факторов.

3. Принятие решений в условиях неопределенности. В данных ЗПР критерий оптимальности зависит кроме стратегий оперирующей стороны и фиксированных факторов также от неопределенных факторов, неподвластных оперирующей стороне и неизвестных ей в момент принятия решения (или известных с недостаточной для принятия решения точностью). В результате влияния неопределенных факторов каждая стратегия оперирующей стороны оказывается связанной с множеством возможных исходов, вероятности которых либо неизвестны оперирующей стороне (или известны с недостаточной для принятия решения точностью), либо вовсе не имеют смысла. Первое соответствует неопределенным факторам стохастической природы (т.е. недостаточно изученным стохастическим факторам, относительно которых отсутствует необходимая статистическая информация), второе - неопределенным факторам нестохастической природы.

Детерминированные ЗПР и ЗПР в условиях неопределенности можно считать предельными случаями ЗПР (т. е. полное знание и полное незнание). ЗПР, в которых имеется элемент риска, занимают некоторое промежуточное положение. Очевидно, что любой предельный случай всегда представляет собой большую или меньшую идеализацию реальной ситуации.

Классификацию ЗПР завершим указанием на математический аппарат, применяемый при решении ЗПР того или другого класса. Однокритериальные статические детерминированные ЗПР в своей общей постановке полностью совпадают с общей постановкой задачи математического программирования.

1.2.2 Исследование операций как научная дисциплина

Исследование операций (ИО) представляет собой научный метод выработки количественно обоснованных рекомендаций по принятию решений. Учитывая важность количественного фактора в исследовании операций и целенаправленность вырабатываемых рекомендаций исследование операций можно определить как теорию обоснования оптимальных решений.

Несмотря на различную природу операции могут быть описаны одними и теми же математическими моделями, более того, анализ этих моделей позволяет лучше понять суть того или иного явления и даже предсказать его дальнейшее развитие. Окружающий мир устроен (в информационном смысле) необычайно компактно, поскольку одна и та же информационная схема используется в самых различных физических (или других) проявлениях.

Благодаря наличию общих закономерностей в развитии самых различных систем их исследование возможно на основе математических методов. Исследование операций сегодня рассматривается как математический инструментарий, поддерживающий процесс принятия решений в разных областях человеческой деятельности, как совокупность методических средств, позволяющих обеспечить ЛПР необходимой количественной информацией, полученной научными методами. Исследование операций как научная дисциплина сформировалось на стыке математики и разнообразных социально-экономических дисциплин, и поэтому свой вклад в его становление внесли представители самых различных областей наук.

История возникновения исследования операций уходит корнями в далёкое прошлое. Так, ещё в 1885 году **Фредерик Тейлор** пришёл к выводу о возможности применения научного анализа в сфере производства. Проблема, рассмотренная им, на первый взгляд кажется тривиальной: «как оптимизировать работу землекопов?». Применение математического аппарата подтвердило несостоятельность принципа «Бери больше, кидай дальше и отдыхай, пока летит». Оказалось, что оптимальный вес грунта, позволяющий максимизировать количество перебрасываемого материала (при разумной экономии рабочей силы) в случае продолжительной работы значительно меньше того, что может поднять человек при максимальной нагрузке.

Пионером в области перевода сложных военно-стратегических задач на язык математики стал **Фредерик Ланчестер**. Одним из наиболее значительных результатов, полученных учёным, стало открытие в 1916 г. так называемого квадратичного закона, количественно связывающего достижение победы с двумя основными факторами: численным превосходством живой силы и эффективностью оружия. Было доказано, что при одновременном вступлении в бой численное превосходство в живой силе более важно, чем применение более совершенного вооружения, поскольку главную роль играет сосредоточение собственных войск и расчленение сил противника. Классическим примером использования квадратичного закона Ланчестера является победоносная тактика адмирала Нельсона в сражении при Трафальгаре.

В 1917 г. датский математик **А.К. Эрланг**, работавший в телефонной компании, сформулировал задачу минимизации потерь времени на установление телефонной связи. Полученные им результаты стали основополагающими принципами в теории телефонных сетей. Позднее формулы Эрланга (среднее время ожидания заявки вызова и др.) были приняты министерством связи Англии в качестве стандарта для расчёта эффективности телефонных соединений. Идеи Эрланга почти на полвека предвосхитили современные теории расчёта характеристик телефонных узлов.

В 1930 г. Г. Левинсон начал применять научный анализ к решению задач, возникающих в торговле. Методика исследования операций была использована для эффективности рекламы, размещения товаров, влияния конъюнктуры на номенклатуру и количества проданных товаров.

В годы второй мировой войны исследование операций широко применялось для планирования боевых действий. Так, специалисты по исследованию операций работали в штабе командования бомбардировочной операции США, дислоцированном в Англии. Ими исследовались многочисленные факторы, влияющие на эффективность бомбометания. Были выработаны рекомендации, приведшие к 4-х кратному повышению эффективности бомбардировок.

Вначале войны боевое патрулирование самолётов союзников для обнаружения кораблей и подводных лодок противника носило неорганизованный характер. Привлечение специалистов по исследованию операций позволило установить такие маршруты патрулирования и такое расписание полётов, при которых вероятность оставить объект незамеченным была сведена до минимума. Полученные рекомендации были применены для организации патрулирования над

Южной частью Атлантического океана с целью перехвата немецких кораблей с военными грузами. Из пяти вражеских кораблей, прорвавших блокаду, три были перехвачены на пути из Японии в Германию, один был обнаружен и уничтожен в Бискайском заливе и лишь одному кораблю удалось скрыться благодаря тщательной маскировке.

В годы войны все работы по использованию методов исследования операций были засекречены. По окончании второй мировой войны группы специалистов по исследованию операций продолжили свою работу в ВС США и Великобритании. Публикации ряда результатов вызвали всплеск общественного интереса к этому научному направлению. Возникла тенденция к применению методов ИО в коммерческой деятельности, в целях реорганизации производства, перевода промышленности на мирные рельсы. Сегодня на развитие математических методов ИО в экономике ассигнуются миллионы долларов.

Термин «***Исследование операций***» возник в результате буквального перевода с английского выражения *operations research*, введенного в конце 30-х годов XX века как условное наименование одного из подразделений британских ВВС, занимавшегося вопросами эффективного использования радиолокационных систем в общей системе противовоздушной обороны. Первоначально исследование операций было связано с решением задач военного содержания, но уже с конца 40-х годов оно используется для решения технико-экономических задач и задач управления на различных уровнях.

В 50-60-е годы на Западе создаются научные общества и центры исследования операций, выпускающие научные журналы, а ряд американских университетов включает эту дисциплину в свои учебные планы. В рамках ИО начинают формироваться отдельные самостоятельные направления – линейное программирование, выпуклое программирование, теория игр, теория массового обслуживания и др.

В настоящее время под ***исследованием операций*** понимают применение математических методов количественного обоснования решений в конкретной области целенаправленной человеческой деятельности (*Е.С. Вентцель*). Исследование операций как научное направление характерно для завершающих этапов жизненного цикла системы (эксплуатация, применение по назначению) и предполагает формализованное описание операции и количественный анализ факторов, определяющих достижение поставленных в операции задач.

Основной задачей ИО является <u>применение научных принципов и математических методов к исследованию функционирования систем с целью оценки характеристик и формирования рекомендаций по выбору оптимальных решений, обеспечивающих наиболее эффективное применение системы в конкретных условиях.</u>

Выделим **методические особенности исследования операций:**

1. Построение формализованной модели операции, предназначенной для получения количественных оценок альтернативных решений;
2. Необходимость охвата различных сторон деятельности;
3. Учёт прошлого опыта при исследовании аналогичных операций;
4. Использование результатов экспериментов, поставленных в различной форме (игры, испытания, учения и т. д.);

5. Направленность результата на количественное обоснование альтернатив и упрощение процесса принятия решения, а не на выдачу самого решения.

Объектом исследования в ИО является конкретная реально существующая (например, экономическая) система.

Предметом исследования в ИО являются характеристики или показатели, отражающие эффективность применения системы по целевому предназначению в конкретных условиях. Исследование операций позволяет количественно оценить эффективность работы системы в конкретных условиях и выработать рекомендации по улучшению её характеристик.

Содержательно всякая задача исследования операций является оптимизационной, т.е. состоит в выборе среди некоторого множества допустимых решений, которые можно в том или ином смысле квалифицировать как оптимальные. При этом допустимость каждого решения принимается в смысле его фактического осуществления, а оптимальность – в смысле его целесообразности.

Содержанием теоретического аспекта ИО является математический анализ оптимизационных задач и нахождение их оптимальных решений. Прикладной аспект ИО заключается в составлении (постановке) оптимизационных задач и в определении их оптимальных решений.

Постановка задачи ИО охватывает, прежде всего, формальное описание множества допустимых решений и критериев оптимального выбора. Оно должно соответствовать содержательным представлениям о возможном и целесообразном выборе в данных условиях. Проверка адекватности самих содержательных представлений объективной реальности и реализация решения уже выходят за пределы области интересов ИО.

Все решения (в том числе и оптимальные) принимаются всегда на основе информации, которой располагает принимающий решения субъект. Поэтому каждая задача ИО в своей постановке должна отражать структуру и динамику знаний ЛПР о множестве допустимых решений и о критерии оптимальности. Например, если принятие решения происходит в наперёд известном и не изменяющемся информационном состоянии, то задача называется статической. В таких условиях весь процесс принятия решения может быть сведён к единому мгновенному акту.

Выделим основные *этапы операционного исследования*:

1) наблюдение явления, сбор и систематизация исходных данных;

2) постановка задачи исследования;

3) разработка концептуальной (операционной) модели;

4) построение и преобразование математической модели к каноническому виду;

5) расчет характеристик или оптимизация параметров модели;

6) анализ выходных данных; если полученные результаты не удовлетворяют исследователя, то следует либо вернуться на этап 3 или 4, т.е. предложить для решения задачи другую математическую модель; либо вернуться на этап 2, т.е. собрать дополнительную информацию и корректно сформулировать задачу;

7) разработка рекомендаций по выбору рациональной структуры (алгоритма, параметров) системы или её компонентов.

Таким образом, операционное исследование, в общем случае, представляет собой итерационный процесс исследования, каждый следующий шаг которого приближает нас к решению конкретной проблемы. В центре операционного исследования находятся построение и расчет параметров математической модели (ММ) операции.

Математическая модель – это система математических соотношений, приближенно, в абстрактной форме описывающих изучаемый процесс или систему.

Экономико-математическая модель – это математическая модель, предназначенная для исследования экономической проблемы.

Проведение операционного исследования, построение и расчет математической модели позволяют проанализировать ситуацию и выбрать оптимальные решения по управлению ею или обосновать предложенные решения. Применение ММ необходимо в тех случаях, когда проблема сложна, зависит от большого числа факторов, по-разному влияющих на её решение. В этом случае непродуманное и научно не обоснованное решение может привести к серьезным последствиям. Примеров этому имеется немало, в частности в экономике. Использование математических методов и моделей позволяет осуществить предварительный выбор оптимальных или близких к ним вариантов решений по определенным критериям. Они научно обоснованы, и лицо, принимающее решение, может руководствоваться ими при выборе окончательного решения. Следует понимать, что не существует решений, оптимальных «вообще». Любое решение, полученное при расчете математической модели, оптимально по одному или нескольким критериям, предложенным поставщиком задачи и исследователем.

В настоящее время ММ применяются в задачах анализа, прогнозирования и выбора оптимальных решений в различных областях экономики. Это планирование и оперативное управление производством, управление персоналом, управление запасами, распределение материальных ресурсов, планировка и размещение активов, руководство инновационным проектом, формирование портфеля заказов и т. п.

Можно выделить следующие основные этапы построения ММ:

1. Определение цели, т.е. формулировка конечного результата, которого хотят добиться, решая поставленную задачу.

2. Определение параметров модели, т.е. заранее известных фиксированных факторов, на значения которых исследователь не влияет.

3. Формирование управляющих переменных, изменяя значение которых можно приближаться к поставленной цели. Значения управляющих переменных являются решениями задачи.

4. Определение области допустимых решений, т.е. тех ограничений, которым должны удовлетворять управляющие переменные.

5. Выявление неизвестных факторов, т.е. величин, которые могут изменяться случайным или неопределенным образом.

6. Выражение цели через управляющие переменные, параметры и неизвестные факторы, т.е. формирование целевой функции, называемой также критерием эффективности или критерием оптимальности задачи.

Введем следующие условные обозначения:

α - параметры модели;

x - управляющие переменные или решения;

X - область допустимых решений;

ξ - случайные или неопределенные факторы;

W - целевая функция или критерий эффективности (критерий оптимальности) $W = W(x, \alpha, \xi)$.

В соответствие с введенными терминами математическая модель задачи имеет следующий вид:

$$W = W(x, \alpha, \xi) \to \max(\min). \tag{1.2}$$

$$x \in X. \tag{1.3}$$

Решить задачу – это значит найти такое оптимальное решение $x^* \in X$*, чтобы при данных фиксированных параметрах* α *и с учётом неизвестных факторов* ξ *значение критерия эффективности* W *было по возможности максимальным (минимальным)*.

$$W^* = W(x^*, \alpha, \xi) \to \max_{x \in X}(\min) W(x, \alpha, \xi).$$

Таким образом, *оптимальное решение – это решение, наиболее предпочтительное перед другими решениями по определенному критерию эффективности (одному или нескольким).*

Перечислим основные принципы построения ММ.

1. Необходимо согласовать точность и подробность модели, во-первых, с точностью тех исходных данных, которыми располагает исследователь, и во-вторых, с теми результатами, которые требуется получить.

2. Математическая модель должна отражать существенные черты исследуемого явления и при этом не должна его сильно упрощать.

3. Математическая модель не может быть полностью адекватна реальному явлению, поэтому для его исследования лучше использовать несколько моделей, для построения которых применены разные математические методы. Если при этом получаются сходные результаты, то исследование заканчивается. Если результаты сильно различаются, то следует пересмотреть постановку задачи.

4. Любая сложная система всегда подвергается малым внешним и внутренним воздействиям, следовательно, математическая модель должна быть устойчивой, т.е. сохранять свои свойства и структуру при этих воздействиях.

На рис. 1.3 представлена классификация математических моделей.

По числу критериев эффективности математические модели делятся на однокритериальные и многокритериальные. Многокритериальные математические модели содержат два и более критерия.

По учёту неизвестных факторов математические модели делятся на детерминированные, стохастические и модели с элементами неопределенности.

В стохастических моделях *неизвестные факторы – это случайные величины, для которых известны функции распределения и различные статистические характеристики (математическое ожидание, дисперсия, среднеквадратическое отклонение* и т. п.). Среди стохастических можно выделить:

модели стохастического программирования, в которых либо в целевую функцию (1.2), либо в ограничения (1.3) входят случайные величины;

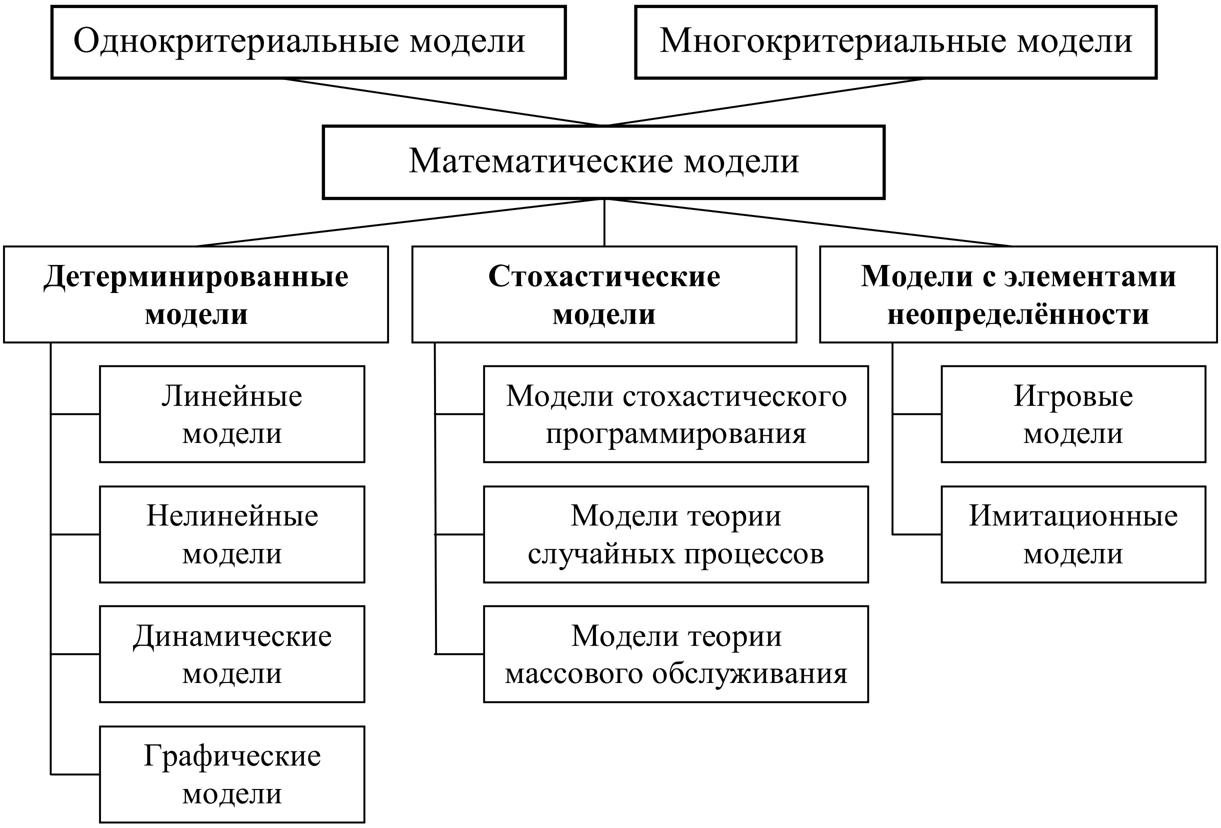


Рис. 1.3 – Классификационная схема математических моделей

модели теории случайных процессов, предназначенные для изучения процессов, состояние которых в каждый момент времени является случайной величиной;

модели теории массового обслуживания, в которой изучаются многоканальные системы, занятые обслуживанием требований; к стохастическим моделям можно отнести модели теории полезности, поиска и принятия решений.

Для моделирования ситуаций, зависящих от случайных факторов, для которых невозможно собрать статистические данные, используются модели с элементами неопределенности.

В моделях теории игр задача представляется в виде игры, в которой участвуют несколько игроков, преследующих разные цели, например, организация предприятия в условиях конкуренции.

В имитационных моделях реальный процесс разворачивается в машинном времени, и прослеживаются результаты случайных воздействий на него, например организация производственного процесса.

В детерминированных моделях неизвестные факторы не учитываются. По виду целевой функции и ограничений детерминированные модели делятся на группы: линейные, нелинейные, динамические и графические.

В линейных моделях целевая функция и ограничения линейны по управляющим переменным. Построение и расчёт линейных моделей являются наиболее развитым разделом математического моделирования, поэтому часто к ним стараются свести и другие задачи либо на этапе постановки, либо в процессе решения

Нелинейные модели – это модели, в которых либо целевая функция, либо какое-нибудь из ограничений (либо все ограничения) нелинейны по управляющим переменным.

Для нелинейных моделей нет единого метода расчета. В зависимости от вида нелинейности, свойств функции и ограничений можно предположить различные способы решения.

В динамических моделях в отличие от статистических линейных и нелинейных моделей учитывается фактор времени. Критерий оптимальности в динамических моделях может быть самого общего вида.

Графические модели используются тогда, когда задачу удобно представить в виде графической структуры.

1.2.3 Примеры задач исследования операций

Те параметры операционной модели, совокупность которых образует решение, называются *элементами решения*. В качестве элементов решения могут фигурировать различные числа, векторы, функции, физические признаки и т.д. Например, если составляется план перевозок однородных грузов из пунктов отправления $A_1, A_2, \ldots, A_m$ в пункты назначения $B_1, B_2, \ldots, B_n$, то элементами решения будут числа x_{ij}, показывающие, какое количество груза будет отправлено из i-го пункта отправления A_i в j-й пункт назначения B_j. Совокупность чисел $x_{11}, x_{12}, \ldots, x_{1n}, \ldots, x_{m1}, x_{m2}, \ldots, x_{mn}$ образует решение.

В простейших задачах исследования операций количество элементов решения может быть сравнительно невелико. Но в большинстве задач, имеющих практическое значение, число элементов решения достаточно велико.

Кроме элементов решения, которые в заданных пределах можно изменять в задаче исследования операций имеются априорно заданные «дисциплинирующие» условия, которые фиксированы с самого начала и нарушены быть не могут (например, грузоподъемность машины; размер планового задания; весовые характеристики оборудования и т. п.). В частности, к таким условиям относятся располагаемые ресурсы (материальные, технические, людские, информационные) и иные ограничения, налагаемые на решение. В своей совокупности они формируют так называемое «*множество возможных решений*». Обозначим это множество буквой X, а тот факт, что решение $\boldsymbol{x}$ принадлежит этому множеству, будем записывать в виде формулы: $x \in X$.

Речь идет о том, чтобы в множестве возможных решений X выделить те решения x, которые с той или иной точки зрения эффективнее (удачнее, предпочтительнее) других. Чтобы сравнивать между собой по эффективности разные решения, нужно иметь какой-то количественный критерий, так называемый *показатель эффективности*, его часто называют «***целевой функцией***». Этот показатель выбирается так, чтобы он отражал целевую направленность операции. «Лучшим» будет считаться то решение, которое в максимальной степени способствует достижению поставленной цели.

Чтобы выбрать показатель эффективности W, нужно, прежде всего, ответить на вопрос: *чего мы хотим, к чему стремимся*, предпринимая операцию? Выбирая решение, естественно, предпочитают такое, которое обращает показатель эффективности W в максимум (или же в минимум). Например, доход от операции хотелось бы обратить в максимум; если же показателем эффективности являются затраты, их желательно обратить в минимум. Если показатель эффективности желательно максимизировать, цель задачи укажут в виде $\boxed{W \to \max}$, а если минимизировать, то в виде: $\boxed{W \to \min}$.

Очень часто выполнение операции сопровождается действием случайных факторов (изменение курса валют, колебания спроса и предложения, действия конкурентов и т.д.). В таких случаях обычно в качестве показателя эффективности берется не сама переменная величина, которую хотелось бы максимизировать (минимизировать), а ее среднее значение (математическое ожидание).

В некоторых случаях операция, сопровождаемая случайными факторами, преследует какую-то вполне определенную цель A, которая может быть только достигнута полностью или совсем не достигнута (схема «да - нет»). Тогда в качестве показателя эффективности выбирается вероятность достижения этой цели $P(A)$.

Неправильный выбор показателя эффективности может привести к ошибочным решениям и к неоправданным затратам и потерям. Для иллюстрации принципов выбора показателя эффективности рассмотрим примеры.

Пример 1.1. **План снабжения предприятий.**

Имеется ряд предприятий, потребляющих известные виды сырья, и есть ряд сырьевых баз, которые могут поставлять это сырье предприятиям. Базы связаны с предприятиями какими-то путями сообщения (железнодорожными, водными, автомобильными, воздушными) со своими тарифами. Требуется разработать такой план снабжения предприятий сырьем (с какой базы, в каком количестве и какое сырье доставляется), чтобы потребности в сырье были обеспечены при минимальных расходах на перевозки.

Задача операции - обеспечить снабжение сырьем при минимальных расходах на перевозки. Показатель эффективности R - суммарные расходы на перевозки сырья за единицу времени, например, месяц. Цель операции - обеспечить $R \to \min$.

Пример 1.2. **Постройка участка магистрали.**

Сооружается участок железнодорожной магистрали. В распоряжении ЛПР - определенное количество ресурсов: людей, техники, стройматериалов и т.д. Требуется спланировать строительство (т.е. назначить очередность работ, распределить машины и людей по участкам пути, обеспечить ремонтные работы) так, чтобы оно было завершено в минимально возможный срок.

Естественным показателем эффективности является время завершения стройки, однако оно связано со случайными факторами (отказы техники, задержки в выполнении отдельных работ). Поэтому в качестве показателя эффективности целесообразно выбрать среднее ожидаемое время $\overline{T}$ окончания стройки. Цель операции – обеспечить $\overline{T} \to \min$.

Пример 1.3. **Продажа сезонных товаров.**

Для реализации определенной массы сезонных товаров создается сеть временных торговых точек. Требуется выбрать: число точек, их размещение, товарные запасы и количество персонала на каждой из них так, чтобы обеспечить максимальную экономическую эффективность распродажи.

В качестве показателя эффективности можно взять среднюю ожидаемую прибыль П от реализации товаров за сезон (П $\to$ max).

Пример 1.4. **Снегозащита дорог**

В условиях Восточной Сибири метели, заносящие снегом дороги, представляют серьезную помеху движению. Любой перерыв движения приводит к экономическим потерям. Существует ряд возможных способов снегозащиты (профиль дороги, защитные щиты и т. д.), каждый из которых требует известных затрат на сооружение и эксплуатацию. Известны господствующие направления ветров, есть данные о частоте и интенсивности снегопадов. Требуется разработать наиболее эффективные экономически средства снегозащиты с учетом потерь, связанных с заносами.

Речь идет о наиболее выгодном экономически плане снегозащиты, поэтому в качестве показателя эффективности можно выбрать средние за единицу времени (например, за год) расходы R. на содержание и эксплуатацию дорог, включая расходы, связанные как с сооружением защитных устройств, так и с расчисткой дорог и задержками транспорта. Цель операции: обеспечить $R \to \min$.

Пример 1.5. **Выборочный контроль продукции**

Завод выпускает определенного вида изделия. Для обеспечения их высокого качества организуется система выборочного контроля. Требуется разумно организовать контроль (т. е. выбрать размер контрольной партии, набор тестов, правила браковки и т. д.) так, чтобы обеспечить заданный уровень качества при минимальных расходах на контроль.

Естественный показатель эффективности, подсказанный формулировкой задачи, это средние ожидаемые расходы R на контроль за единицу времени, при условии, что система контроля обеспечивает заданный уровень качества, например, средний процент брака не выше заданного ($R \rightarrow \min$).

1.3 Сущность и принципы системного подхода

1.3.1 Сложные системы. Системный подход

Слово "система" образовано от греч. *systema*, что означает *целое, составленное из частей*. Для определённости дальнейшего рассмотрения примем в качестве основного следующее определение «*система*».

Система – упорядоченная совокупность (множество) взаимосвязанных объектов и (или) взаимодействующих элементов (объектов), предназначенная для достижения определённой цели.

Основными частями системы являются: вход, процесс (операция), выход, цель, обратная связь и ограничения (рис.1.4).

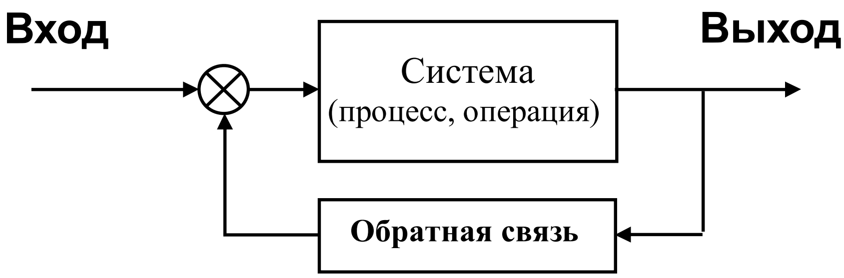


Рис.1.4 – Кибернетическая интерпретация системы

Понятия «вход», «выход» и «состояние» используются для описания взаимодействия системы с внешней средой. Воздействие внешней среды на систему характеризуется некоторыми показателями (или параметрами), которые называют входными параметрами (входами): $U=(u_1,u_2, \ldots, u_m)$, где m - размерность вектора входа. Входы преобразуются системой в параметры $Y=(y_1,y_2,\ldots, y_r)$, где n - размерность вектора выхода, которые характеризуют результаты процессов, протекающих в системе. Эти параметры называют выходными координатами (выходами) системы.

Состояние системы задаётся набором отдельных показателей – вектором состояния $Z=(z_1,z_2,\ldots,z_n)$, с достаточной полнотой характеризующего систему.

Системы, которые в течение определённого интервала времени переходят из одного состояния в другое, или функционируют, называют ***динамическими системами.***

Частями системы являются также процесс (операция) и обратная связь. Процесс переводит вход системы в выход, и в то же время процесс отражает внутреннюю структуру и свойства системы. Можно рассматривать процесс как совокупность операций, осуществляющих преобразование информации (и /или энергии, вещества), которые поступают в систему через входы.

Обратная связь представляет собой канал связи между выходом системы и её входом, который функционирует либо прямо, либо через другие элементы системы (например, через человека-оператора или орган управления).

Системы создаются, развиваются для достижения определённых целей. Под **целью** понимают конечное состояние, к которому система стремится в силу своего предназначения и структурной организации.

На рис.1.5 представлена укрупнённая функциональная схема динамической системы управления. Система управления независимо от назначения содержит: информационно-измерительную подсистему (ИИПС), управляющую под систему (устройство) (УПС), исполнительную подсистему (ИПС), обобщённый объект управления (ООУ), обратную связь (ОС) и дискриминатор (алгебраический сумматор).

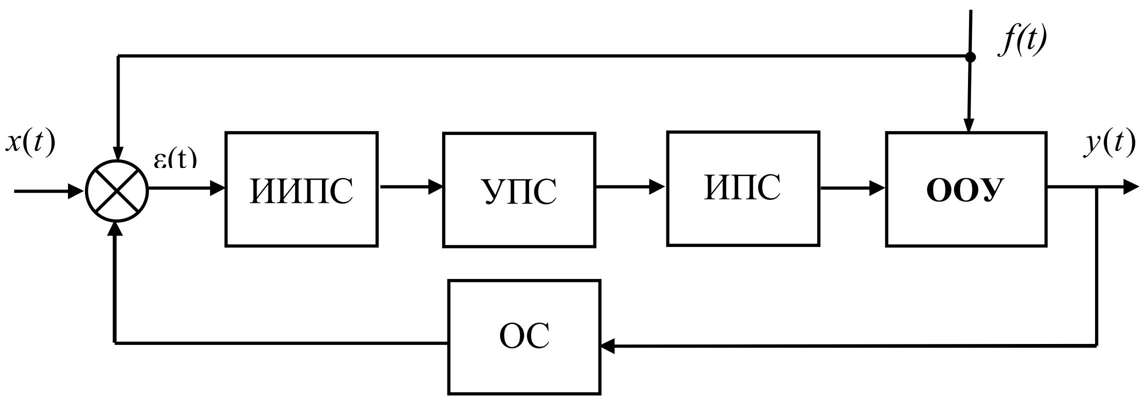


Рис. 1.5 - Функциональная схема динамической системы управления

На рис. 1.5 введены обозначения: $x(t)$ - входной сигнал; $f(t)$- возмущение (помеха), действующее на ИИПС и ООУ; $y(t)$ и $\hat{y}(t)$– выходной сигнал системы и его оценка - сигнал обратной связи, подаваемый на разностный вход сумматора; $u(t)$- управляющее воздействие (сигнал) $\varepsilon(t)$ - рассогласование (ошибка) системы. При этом текущая ошибка (рассогласование) $\varepsilon(t)$ системы определяется через отклонение оценки выходного сигнала от задающего воздействия и представляется в виде $\boxed{\varepsilon(t) = x(t) + f(t) - \hat{y}(t).}$

В данной схеме системы реализуется известный принцип кибернетики - *регулирование по отклонению*, согласно которому управляющее воздействие (сигнал) $u(t)$ формируется таким образом, чтобы мгновенное значение рассогласования (ошибки) $\varepsilon(t)$ в системе управления с течением времени стремилось к нулю.

Принципиальным для системы является то, что объекты (подсистемы, части) функционируют во времени "как единое целое" - каждый объект, подсистема, компонент работают ради единой цели, стоящей перед системой в целом.

Существенное место в понятии "система" занимает *принцип целостности*, согласно которому взаимосвязь и взаимодействие объектов порождает новые, системные свойства, не присущие его отдельным элементам.

Для системного подхода к исследованию, важное значение имеет внешняя среда. Под ***внешней средой*** понимают совокупность всех объектов, изменение свойств которых влияет на систему, а также тех объектов, чьи свойства меняются в результате изменения поведения системы. Внешняя среда может быть отделена от системы границей, фиксируемой в физическом или логическом смысле. В качестве внешней среды, по отношению к данной системе, могут выступать также другие взаимодействующие с ней системы, например, конкурирующие фирмы, системы управления высшего уровня иерархии.

Между системой и средой существует тесная взаимосвязь и взаимозависимость, заключающиеся в том, что система проявляет и формирует свои свойства в процессе взаимодействия со средой, являясь при этом его ведущим активным компонентом.

Неразрывно с понятием системы связано понятие "структура". В общем случае *структура* - это способ организации целого из составных частей. Под **структурой** системы будем понимать устойчивую упорядоченность в пространстве и во времени её элементов и связей. Структура является наиболее консервативной характеристикой системы: хотя состояние системы меняется, структура ее сохраняется неизменной длительное время.

В зависимости от характера организации в системе элементов и их связей можно выделить три основных типа структур: *сетевую, скелетную и централистскую*, отражающих последовательное повышение степени централизации системы.

В плане пространственной организации различают структуры: плоские и объемные; рассредоточенные, когда элементы равномерно распределены в пространстве; локально сосредоточенные при наличии сгущении элементов и сосредоточенные, когда имеется одно сгущение элементов.

По временному признаку выделяются экстенсивные структуры, в которых с течением времени происходит рост числа элементов, и интенсивные, в которых происходит рост числа связей и их мощности при неизменном составе элементов. Противоположные типы структур: редуцирующие и деградирующие. Еще один тип - стабильные структуры, в которых структура не меняется в течение всего периода "жизни" системы.

Обратная связь (ОС) - представляется собой канал связи между выходом и входом системы, функционирующей либо прямо, либо через другие элементы системы (например, через орган управления).

Системы создаются и развиваются для достижения определенной *цели*. Под ***целью*** понимают конечное состояние, к которому стремится система в силу своего назначения и структурной организации.

Таким образом, можно сформировать несколько признаков, позволяющих произвольный объект считать системой:

1) **Целостность и членимость** (система - целостная совокупность элементов). Это означает, что с одной стороны, система - целостное образование и, с другой - в ее составе отчетливо могут быть выделены целостные объекты (эле-

менты). Вне системы они могут обладать "системно значимыми" свойствами, а при вхождении в систему - системно определенными свойствами.

2) **Связи.** Наличие существенных устойчивых связей (отношений) между элементами и (их) свойствами, превосходящих по мощности (силе) связи со средой. К числу основных характеристик связи относятся: физическое наполнение, направленность, мощность и роль в системе.

По физическому наполнению связи можно подразделить на вещественные, энергетические, информационные, смешанные и ненаполненные (отношения). По направлению различают связи: прямые, обратные, контрсвязи и нейтральные. Важной характеристикой отношений и связей является их сила (или мощность).

Система существует как некоторое целостное образование тогда и только тогда, когда мощность (сила) существенных связей между элементами системы на интервале времени ($t > 0$) больше, чем мощность (сила) связей этих же элементов с окружающей средой.

Сравнительно просто оценивается мощность вещественных энергетических связей по интенсивности потоков вещества или энергии. Для информационных связей оценкой потенциальной мощности - действительная величина потока информации. Однако в общем случае при оценке мощности информационных связей необходимо учитывать качественные характеристики передаваемой информации. Однако в общем случае при оценке мощности информационных связей необходимо учитывать качественные характеристики передаваемой информации (ценность, полезность, верность и т.д.).

Роль связи в системе определяется характером ее влияния на ход процессов. В этом смысле различают связи: соединительные, ограничивающие, усиливающие (ослабляющие), запаздывающие (опережающие, мгновенные), селектирующие, преобразующие, положительные и отрицательные ОС, согласующие, координирующие и т. п.

3) **Организация.** Возникновение организации в системе - это актуализация (формирование) существенных связей элементов, упорядоченное распределение связей элементов во времени и пространстве.

При формировании складывается определенная структура системы, а свойства элементов трансформируются в функции (действия и поведение), связанные с 4-м свойством системы.

4) **Интегративные качества.** Наличие интегративных качеств показывает, что свойства системы хотя и зависят от свойств элементов, но не определяются ими полностью.

Таким образом, можно констатировать: <u>а) система не сводится к простой сумме совокупности элементов; б) расчленяя систему на отдельные части, изучая каждую из них только в отдельности, нельзя познать все свойства системы в целом.</u>

5) **Эффективность.** Под эффективностью системы понимают её способность выполнять заданные функции с заданным качеством.

Многообразие систем велико, и полной их классификации в настоящее время нет. Не выработаны также окончательно и принципы классификации.

В соответствии с признаком субстанциональности можно выделить три класса системы:

1. Е с т е с т в е н н ы е - системы, существующие в объективной действительности (неживой и живой природе, обществе). Это атом, молекула, живая клетка, организм, популяция, общество.

2. К о н ц е п т у а л ь н ы е - системы, отражающие реальную действительность, объективный мир (например, научные теории, гипотезы литературные произведения и т.п.).

3. И с к у с с т в е н н ы е с и с т е м ы - системы, созданные человеком.

Аналогичность процессов, протекающих в системах разного типа (технических, экономических, организационных социальных, биологических), и появившаяся на базе концепции "система" возможность переносить знания из одной области в другую, послужили толчком к созданию общей теории систем. **Общая теория систем** - научное направление, связанное с разработкой совокупности методологических, конкретно-научных и прикладных проблем исследования сложных систем произвольной природы. Общая Теория систем в основном связана с изучением систем, которые в современной литературе называют сложными (или большими) системами.

Сложная система - составной объект, части которого можно рассматривать как системы, связанные между собой заданными отношениями.

В общем случае сложная система имеет многоуровневую (иерархическую) структуру. Принято выделять три концепции иерархии: а) иерархия описания, или абстрагированная; б) иерархия сложности принимаемого решения, которая предусматривает использование термина "многослойная система принятия решения"; в) организационная иерархия, которая используется при исследовании и описании организационных систем.

Иерархия уровней описания обусловлена следующим положением: с одной стороны, необходимо иметь детальные описания системы, с другой стороны, это описание должно быть простым и обозримым. Выполнить эти требования одновременно можно лишь при использовании соответствующей иерархи описания системы. При этом система задается семейством моделей, каждая из которых описывает систему различных уровней абстрагирования. Уровни абстрагирования называют стратами.

Второе понятие иерархии связано с уровнями сложности принимаемых решений. В сложных ситуациях проблема принятия решения разбивается на множество последовательно решаемых более простых проблем так, что решение последних позволяет решить и исходную проблему. Такую иерархию называют иерархию слоев принимаемых решений, систему принятия решения - многослойной системой, а сам процесс расчленения исходной проблемы на более простые проблемы (задачи) - декомпозицией.

Организационные иерархии подразумевают, что система состоит из множества четко выделенных взаимодействующих подсистем. Подсистемы, находящиеся на верхних уровнях иерархии являются управляющими, а нижних уровнях - управляемыми.

Формирование структуры в общем случае также предполагает декомпозицию системы – расчленение ее на подсистемы. С другой стороны, если структурные отношения и состав подсистем (элементов) заданы, систему можно построить с помощью композиции (объединения) отдельных ее частей в единое целое.

Системный подход можно рассматривать как дальнейшее развитие на основе диалектического метода классических схем функционального и комплексного подходов исследования сложных систем.

В системном подходе можно выделить некоторые аспекты изучения систем, связанные в основном с их строением и функционированием: системно-элементный аспект предусматривает в качестве начального этапа исследования объекта (проблемы) как системы изучение его элементарного состава; системно-структурный аспект предусматривает изучение разнотипных связей, объединяющих элементы в систему; системно-функциональный аспект связан с изучением поведением частей систем и рассмотрением функционирования системы в целом; системно-исторический аспект предусматривает изучение системы с учетом ретроспективы ее развития.

Основные задачи системного подхода:

1. Разработка концептуальных (содержательных и формальных средств представления исследуемых объектов как систем;
2. Построение обобщённых моделей систем и моделей разных классов и свойств систем, включая модели динамики систем, их целенаправленного поведения, развития, иерархического построения процессов управления;
3. Исследование методологических оснований различных теорий систем.

Системный подход не имеет фиксированной предметной области. Он формирует характер, направление и стиль научного мышления при исследовании какого-либо процесса. При системном подходе исходят из того, что специфика сложного объекта (системы) не исчерпывается особенностями составляющих его элементов, а коренится, прежде всего, *в характере связей и отношений между ключевыми элементами*. Системный подход характеризуется рассмотрением взаимосвязи и взаимозависимости элементов внутренней и внешней среды разработки системы управления.

В практике управления сложными системами идея системного подхода используются в методологических средствах системного анализа.

Системный анализ появился как средство решения слабо структурированных проблем, т.е. задач, в которых преобладают качественные малоизвестные и неопределенные стороны.

Системный анализ - методология обоснования решения в условиях существенных неопределенностей, объединяющая общую схему системного подхода с аналитическим процессом принятия решения.

Объектом системного анализа являются системы определенного класса, включая процессы подготовки и принятия решений, рассматриваемые как системы правил, процедур и приемов.

Предмет системного анализа включает в себя общесистемные характеристики и взаимодействия исследуемой системы с окружающей средой. Привлечение методов системного анализа для решения задач проектирования сложных ИС необходимо прежде всего потому, что в процессе ПР приходится осуществлять выбор в условиях неопределенности, которая обусловлена наличием факторов, не поддающихся строгой количественной оценке.

Процедуры и методы системного анализа направлены, прежде всего, на выдвижение альтернативных вариантов решения проблемы, выявление масштабов неопределенности по каждому из вариантов и сопоставление вариантов по тем или иным критериям эффективности. Специалисты по системному анализу готовят и/или рекомендуют рациональные варианты решения, принятие же решения остается в компетенции соответствующего лица или органа.

Важнейшие **принципы системного анализа** сводятся к следующему: а) процесс ПР должен начинаться с выявления и четкого формулирования конечных целей; б) проблема рассматривается комплексно с выявлением всех последствий и взаимосвязи каждого частного решения; в) проводится анализ и выявление возможных альтернативных путей достижения цели с учетом всех существенных факторов; г) цели отдельных этапов и проектирования не должны противоречить целям всей программы исследования.

Анализ литературы в общей теории систем показывает, что можно указать принципиальную последовательность этапов проведения системного анализа: 1) уяснение задачи; 2) определение конечных целей; 3) разработка альтернатив, т.е. вариантов и средств достижения поставленных целей; 4) выявление потребных ресурсов и ограничений в них; 5) анализ взаимовлияния целей, альтернатив и ресурсов; 6) принятие решения; реализация решения.

Отметим, что <u>в системном анализе не только объект исследования, но и предмет исследования, рассматриваются как системы</u>.

На рис 1.6 представлена типовая схема этапов системного анализа с указанием связей между ними.

Основными теоретическими сферами системного анализа являются следующие направления: а) общие принципы проведения исследования сложных систем, и, в частности, интегрирование различных математических методов во взаимообусловленную совокупность методов системного анализа; б) проблемы сложности и неопределенности и методы их разрешения; в) проблема оценки предельных характеристик систем; г) принципы и способы (моделирования) сложных систем.

В настоящее время системный анализ использует широкий спектр разнообразных методов, которые можно объединить в следующие группы: эвристическое программирование, семиотический подход, методы аналогий, аналитические методы и имитационное моделирование.

Ведущей концепцией системного анализа служит **системный подход**. Важное значение имеет количественная оценка различных свойств, характеристик и факторов, поэтому основой всего системного анализа можно считать построение и исследование математических моделей. В системном анализе важную роль играет имитационное моделирование. Под ***имитационным модели-***

рованием будем понимать процесс формирования модели реальной системы и проведения на этой модели экспериментов в целях выявления существенных свойств системы и определения возможных путей ее создания, совершенствования и (или) эффективного использования.

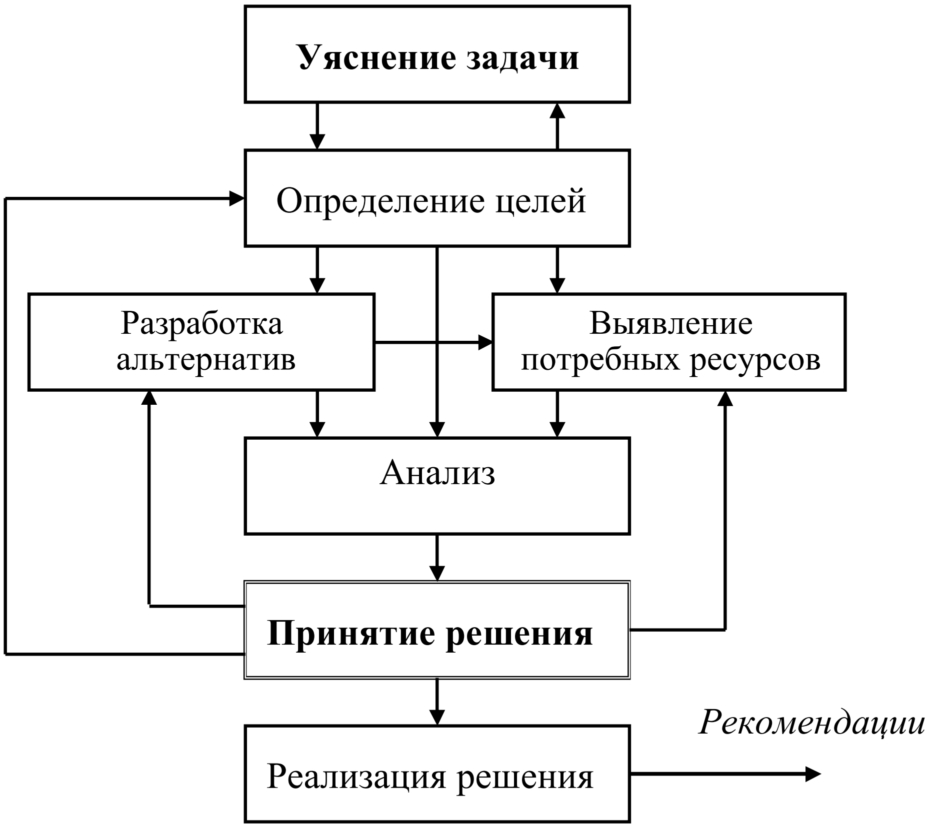


Рис.1.6 - Этапы системного анализа сложных систем

1.4 Прямые и обратные задачи исследования операций

Задачи исследования операций, встречающиеся в экономической практике, делятся на две категории: *а) прямые и б) обратные*.

Прямые задачи отвечают на вопрос: что будет, если в заданных условиях будет принято конкретное решение $x \in X$? В частности, чему будет равен, при данном решении x, выбранный показатель эффективности W (или же ряд таких показателей)?

Для решения такой задачи строится математическая модель, позволяющая выразить один или несколько показателей эффективности через заданные условия и элементы решения.

Обратные задачи отвечают на вопрос: как выбрать решение x для того, чтобы показатель эффективности W обратился в максимум?

Естественно, что модель прямой задачи ИО проще модели обратной задачи. Очевидно также, что для решения обратной задачи ИО, прежде всего, надо уметь решать прямую задачу. Для некоторых типов операций прямая задача решается настолько просто, что ею специально не занимаются. Для других типов операций построение ММ и вычисление показателя (показателей) эффективности само по себе далеко не тривиально (так, например, обстоит дело с прямыми задачами теории массового обслуживания).

Остановимся подробнее на обратных задачах. Если число возможных вариантов решения, образующих множество X, невелико, то можно попросту вычислить величину W для каждого из них, сравнить между собой полученные значения и непосредственно указать один или несколько оптимальных вариантов, для которых W достигает максимума. Такой способ нахождения оптимального решения называется «простым перебором». Однако, когда число возможных вариантов решения, образующих множество X, велико, поиск среди них оптимального «вслепую», простым перебором, затруднителен, а зачастую практически невозможен. В этих случаях применяются методы «направленного перебора», обладающие той общей особенностью, что оптимальное решение находится рядом последовательных «попыток» или «приближений», из которых каждое последующее приближает нас к искомому оптимальному.

Рассмотрим постановку задачи оптимизации решения (обратной задачи исследования операций) в общей форме.

Пусть имеется некоторая операция Q, на успех которой мы можем в какой-то мере влиять, выбирая тем или другим способом решение x (x - группа параметров). Пусть эффективность операции характеризуется одним показателем $W \to \max$.

Возьмем относительно простой, так называемый «детерминированный» случай, когда все условия операции полностью известны заранее, т. е. не содержат неопределенности. Тогда все факторы, от которых зависит успех операции, делятся на две группы:

1) заданные, заранее известные факторы (условия выполнения операции), которые мы для краткости обозначим через α;

2) зависящие от исследователя элементы решения, образующие в своей совокупности решение x.

Заметим, что первая группа факторов содержит, в частности, и ограничения, налагаемые на решение, т. е. определяет область возможных решений X.

Показатель эффективности W, который зависит от обеих групп факторов, запишем в виде формулы:

$$W = W(\alpha, x) \tag{1.4}$$

Будем считать, что вид зависимости (1.4) известен, т.е. прямая задача решена. Тогда обратная задача формулируется следующим образом.

При заданном комплексе условий α найти такое решение $x = x^*$, которое обращает показатель эффективности W в максимум.

Искомый максимум обозначим как:

$$W^* = \max_{x \in X}\{W(\alpha, x)\} \tag{1.5}$$

Задача (1.5) - типичная математическая задача нахождения максимума функции или функционала. Эта задача принадлежит к классу так называемых «вариационных задач», хорошо разработанных в математике. Самые простые из таких задач («задачи на максимум и минимум»). Чтобы найти максимум или минимум (короче, «экстремум») функции многих аргументов, надо продифференцировать её по всем аргументам (в данном случае - элементам решения), приравнять частные производные нулю, а затем решить полученную систему уравнений.

Известный классический метод в ИО имеет весьма ограниченное применение. Во-первых, когда аргументов много, задача решения системы уравнений зачастую оказывается не проще, а сложнее, чем непосредственный поиск экстремума. Во-вторых, когда на элементы решения наложены ограничения, экстремум часто достигается не в точке, где производные равны нулю (такой точки может вообще не быть), а где-то на границе области X. Возникают все специфические трудности так называемой «многомерной вариационной задачи при ограничениях». Кроме того, в некоторых задачах функция W вообще не имеет производных (например, задана только для целочисленных значений аргументов). Все это усложняет задачу поиска экстремума целевой функции.

Метод поиска экстремума и связанного с ним оптимального решения x^* должен всегда выбираться исходя из особенностей функции W и вида ограничений, накладываемых на решение.

1.5. Операционные модели

Напомним, что под операцией понимается совокупность взаимосвязанных действий, направленных на решение поставленной задачи в условиях сложившейся обстановки. Операция отражает взаимодействие всех средств, обеспечивающих решение поставленной задачи. Взаимодействие компонентов системы может быть представлено последовательными этапами – условно выделяемыми частями операции, направленными на решение промежуточной задачи операции.

В общем случае сложные системы могут быть представлены взаимосвязанными материальными компонентами, а цели – определёнными позициями, иерархически увязанными в «дерево» целей. Выделим две группы целей: абстрактные и материализованные. Особое значение для сложных экономических систем имеют материализованные цели, для которых желаемый результат функционирования системы может быть задан определённым уровнем обслуживания материального объекта.

Для сложных экономических систем, рассматриваемых в пособии, к основным компонентам операции можно отнести: исследуемую систему, материализованные цели, совокупность действий системы и целей, условия функционирования системы, реакции от других систем.

Для исследования операции устанавливается критерий эффективности, по которому определяется соответствие системы задачам операции.

Уровень поставленных задач отражает уровень рассматриваемой операции. К операциям высокого уровня можно отнести функционирование системы в составе группы сложных систем, операции в масштабе отрасли экономики и т. д. К операциям низкого уровня можно отнести функционирование подсистемы или её элементов на некотором этапе операции. В свою очередь уровень операции определяется условиями решаемых задач, составом задействованных средств, а также целью исследования. Правильное определение уровня операции является весьма важным для проведения качественного исследования. Для исследования эффективности моделируется операция этапа непосредственного применения системы, наиболее критичного к параметрам системы; этап эксплуатации может рассматриваться как частный случай.

Функционирование исследуемой системы в операции описывается абстрактной моделью. Выделяют вербальное (словесное) описание опера-ции, как схему операции, и формализованное (математическое) описание этой схемы, как модель операции. Описание операции является основой построения её математической модели.

В соответствии со схемой операции разрабатывается её ММ, т. е. математическое описание процессов, протекающих в операции с помощью основных компонентов: совокупности действий (*act*), отражающих функционирование каждой части ЭС; совокупности действий (*kon*), отражающих функционирование материализованной цели; условий операции *U*.

К общим принципам построения модели операции можно отнести следующие:

- в модели операции выделяются: система, цель, действия системы, действия цели и условия;
- цель должна рассматриваться как сложная система, активно реагирующая на каждое действие изучаемой системы;
- изучаемая система и цель должны быть представлены соответствующими компонентами иерархической структуры;
- уровень компонентов определяет характер модели операции: занижение уровня (компоненты низкого иерархического уровня) приводит к потере критичности и неоправданному усложнению модели; завышение уровня (компоненты высокого иерархического уровня) – к потере определяющих взаимосвязей;
- действия других систем и средств, не являющихся объектом непосредственного моделирования, в операции должны быть заменены соответствующими реакциями, определяемыми на основе системного подхода;
- в модели операции необходимо учитывать все компоненты, связанные с моделируемой системой функциональными связями;
- в модели операции должны отражаться ближайшие компоненты более высокого и более низкого иерархических уровней;
- схема действий, рассматриваемая в операции должна быть взаимозависимой; независимость действий должна специально оговариваться с обоснованием соответствующих положений.

При операционном моделировании выделяются действующее и обеспечивающее звенья системы.

Объектом операционного *моделирования* является схема операции и её основные компоненты: система, условия и действия, сопровождающие операцию.

Специфика ММ для исследования эффективности управления ЭС обусловлена: а) сжатыми сроками, отводимыми на исследование; б) возможностью детальной проработки вариантов проектируемого элемента; в) необходимостью обеспечения критичности к параметрам элемента; г) неформализуемостью ряда факторов; д) высокой степенью неопределенности внешней среды и т. д.

В связи с этим разработка и исследование математических моделей должны опираться на следующие положения:

а) блочный принцип формирования моделей на основе разработки типовых блоков с последующей их композицией;

б) метод статистических испытаний;

в) описание проектируемого элемента с помощью параметрических выражений и логических условий;

г) разработка типовых модулей целевой остановки и условий.

Разработка типовых блоков обеспечивает возможность быстрого построения и переналадки моделей для сокращения сроков моделирования и обеспечения адекватности модели исследуемому процессу. Оснащенность проектных организаций средствами автоматизированного проектирования обеспечивает возможность широкого использования метода статистического моделирования для исследования сложных объектов, изучение которых другими аналитическими методами затруднительно. Использование параметрических соотношений для увеличения числа исследуемых вариантов проектируемого элемента позволяет расширить область поиска рационального варианта, учесть различные ограничения, повысить качество исследования. Возможность параметрического анализа обусловлена разработкой специальных методов формирования множества вариантов и проектно-конструкторскими проработками. Разработка типовых модулей для исследования позволяет учесть ряд неопределенностей, упростить исследование и в то же время повысить его критичность.

В соответствии с типовой схемой операции модель проектируемого элемента в составе организационных структур управления ЭС представляется в виде блоков, описывающих параметры ДЗ и реакции ОЗ, а также блоков расчета параметров проектируемого элемента, представляемого в виде СД и СВ. Основой разработки моделей проектируемых элементов являются заданный облик ЭС и типовые структурные схемы, позволяющие сформировать множество вариантов в рамках фиксированной структуры. Формирование множества вариантов проводится на основе параметрического анализа, описываемого в модели типовыми блоками расчета проектных параметров элемента, получения и преобразования уравнений существования элемента, формирования альтернативных вариантов элемента для выбора наиболее рационального.

Модель операции разрабатывается на основе построенной схемы операции в виде ряда последовательных этапов и совокупности элементарных действий элементов. Каждое действие в схеме операции (независимо от того, какими средствами и на каком этапе оно выполняется) представляется пятью последовательными *типовыми фазами*: а) исходное состояние действую-щего элемента на заданный момент протекания операции; б) получение информации от средств разведки и целеуказания; в) распределение средств воздействия на выделенные объекты (цели); г) наведение; д) воздействие, связанное с непосредственным энергетическим контактом действующего элемента и объекта поражения. В соответствии с этой схемой разработка моделей эффективности базируется на формировании моделей типовых фаз. Модель фазы в каждом конкретном случае может быть построена на основе универсальной части, отражающей математическую сущность процесса, и конкретной части, учитывающей специфику исследуемого действия в конкретной задаче. *Универсальная часть* (типовой блок) представляет собой математически формализуемую часть модели фазы и может использоваться без изменения для диапазона систем и условий. *Конкретная часть* модели фазы разрабатывается в обобщенном виде и может подвергаться уточнениям для отражения специфики проектируемой системы. Методическое объединение отдельных моделей фаз составляет модель соответствующего этапа операции, а объединение этапов - модель всей операции. Основу такой композиции составляют объединяющие блоки (табл.1.1).

Таблица 1.1

Объединяющие блоки			
элементов		действий (фаз)	
Группирование	Взаимное влияние	Группирование	Взаимное влияние
Модель группы средств	Модель взаимного влияния	Модель этапа	Модель многократности

Типовые блоки отражают, как правило, элементарные независимые действия элементов в "дуэльной" ситуации. Поэтому при разработке модели операции в целом основу объединяющих блоков должно составлять отражение следующих факторов: особенности группового применения эле-ментов на основе задания модели группы средств (ГС) и модели учета взаимного влияния элементов; объединение разнообразных действий на основе разработки моделей этапов операции и моделей учета многократности и взаимного влияния действий. В состав модели операции включаются параметры условий применения. Условия первой группы объединяются по модульному принципу.

Целевая обстановка, определенная на уровне анализа систем, корректируется для задач проектной эффективности, а модели противодействия строятся на основе существующих более емких моделей с учетом использования методов проектной эффективности. При построении моделей противодействия, которые должны в общем случае характеризовать возможный ответ на каждое действие проектируемой системы, основное внимание должно быть сосредоточено на неавтономных ее действиях, которым в принципе может быть противопоставлено противодействие.

Остановимся на общих принципах разработки этой части модели. Параметры объектов внешней среды, представляемые в модели, определяются с одной стороны параметрами СВ (например, для СВ с избирательным воздействием объект должен представляться как сложная функционирующая система) и с другой стороны - задачами, поставленными перед проектируемой системой.

Определение средств противодействия (состава и характеристик) при задании условий первой группы проводится на основе соответствующих моделей противодействия с использованием данных о состоянии и перспективах развития средств, а также методов анализа противодействия. Определение полного перечня СП возможно на основе всех пяти фаз каждого действия. В моделях противодействия необходимо учитывать временные характеристики функционирования СП; привязку СП к объектам целевой обстановки; возможность перераспределения СП между объектами в интересах ЦО в целом; динамику возможного развития СП. С учетом указанных факторов определяются необходимые характеристики СП (качественные и количественные), зависящие от параметров СВ.

Таким образом, схема функционирования и характеристики ЦО, состав и характеристики СП, взаимная их увязка составляют основу модели условий первой группы, которая должна содержать следующие данные: а) по объектам ЦО - тип, расположение, основные характеристики; б) по СП - тип, расположение, количество, основные характеристики, дополнительные сведения (например, данные по обеспечивающим средствам). При разработке этой модели необходимо учитывать, что ЦО формируется, как правило, на уровне анализа систем и корректируется с учетом задач проектной эффективности; объективно существует неоднозначность противодействия, связанная как с неопределенностью самих СП, так и с возможной реакцией; данная ЦО является исходной для дальнейшего построения на ее основе модулей.

Условия второй группы (пассивные), как правило, задаются в виде соответствующих распределений параметров (дискретных или непрерывных), характеризующих учитываемые в модели условия функционирования проектируемой системы. При упрощенном задании условий этой группы можно выделить два частных случая: определение средних значений параметров, описывающих условия (при этом понятие "средний" устанавливается для каждого конкретного параметра в конкретном исследовании); определение граничных (предельных) значений параметров, характе-ризующих диапазон условий, установленный заказывающей организацией. При задании уровня граничных условий необходимо иметь в виду, что некоторое его снижение может повысить эффективность системы за счет более рационального распределения ресурсов. Такое снижение требований по условиям применения системы должно быть согласовано с потребителем на основе соответствующего обоснования.

На основе модели операции с помощью средств решения задачи по исходным данным определяются массивы числовой информации, содержащие оценку показателей эффективности. Эти массивы информации и составляют результаты исследования эффективности, анализ которых позволяет сформулировать рекомендации по выбору рационального варианта проектируемого элемента системы.

На основе ММ устанавливается зависимость показателя эффективности W от параметров системы, в состав которой входит тот или иной вариант проектируемого элемента из множества $\{\alpha\}$, и параметров условий применения $\{\beta\}$ и $\{U\}$: $W = F[\{\alpha\}, \{\beta\}, \{U\}]$.

Если условия $\{\beta\}$, $\{U\}$ являются заданными детерминированными, то задача выбора рационального варианта проектируемого элемента сводится к *типичной вариационной задаче*. Для задач проектной эффективности характерно наличие неопределенных условий применения. В исследовании операций разработаны подходы к выбору решения в условиях неопределенности. Возможны три случая задания неопределенных факторов:

1) заданы законом распределения вероятностей (ЗРВ);

2) неизвестны, заданы диапазоном и имеют пассивный характер (условия второй группы);

3) неизвестны, заданы диапазоном и имеют активный характер, т.е. могут изменяться в зависимости от выбора $\{\alpha\}$ (условия первой группы).

Если неопределенные факторы заданы ЗРВ, то их учет может осуществляться с помощью следующих двух приемов.

1. Замена случайных параметров их математическими ожиданиями (сведение к детерминированной схеме). Здесь предусматривается определение математического ожидания (МОЖ) неопределенных параметров M[U], определение и оптимизация по $\{\alpha_\mu\}$, $\mu = \overline{1, r}$ зависимости $W = F[\{\alpha\}, \{\beta\}, \{U\}]$, *в которой параметры* $\{U\}$ заменяются их МОЖ. Далее осуществляется оптимизация этой зависимости по вектору параметров $\{\alpha_\mu\}$: $W = F[\{\alpha_\mu\}, \{\beta\}, M\{U\}] \to \max$.

2. "Взвешивание" результата расчета W по вероятности (оптимизация "в среднем"). Этим приемом эффективность W определяется в соответствии с зависимостями:

$$W = \sum_{\nu=1}^{N} P(U_\nu) \cdot W(U_\nu) \quad \text{- для дискретных параметров } U;$$

$$W = \int_{U_н}^{U_в} W(U) \cdot f(U) \cdot dU \quad \text{- для непрерывных параметров } U.$$

Здесь $P(U_v)$ - вероятность того, что случайный параметр примет значение U_v; $W(U_v)$ - эффективность системы в условиях $U \equiv U_v$.

Выделенный прием может использоваться: 1) в приближённых расчетах; 2) в случаях, если диапазон изменения параметра U невелик; 3) в случаях, если зависимость W(U) линейная или близка к ней. "Взвешивание" по вероятности позволяет получить более точные оценки.

Для задач проектной эффективности более характерны неопределенные факторы, распределение которых неизвестно. В этом случае выбор рационального варианта может осуществляться на основе анализа игровой матрицы, в которой стратегиями являются варианты проектируемого элемента и условий, а мерой платежа - показатель эффективности.

Выбор рационального варианта в этом случае усложняется отсутствием уверенности в полноте распределения условий и невозможностью выбора решения в смешанных стратегиях, так как создается только один вариант проектируемого элемента. Для случая, когда условия носят неопределённый характер, но являются пассивными, в методах исследования операций для выбора рационального варианта используют критерии максимина.

Разработка операционной модели, предназначенной для расчёта (или оптимизации) характеристик, исследования эффективности системотехнического комплекса (СТК), осуществляется в следующей последовательности.

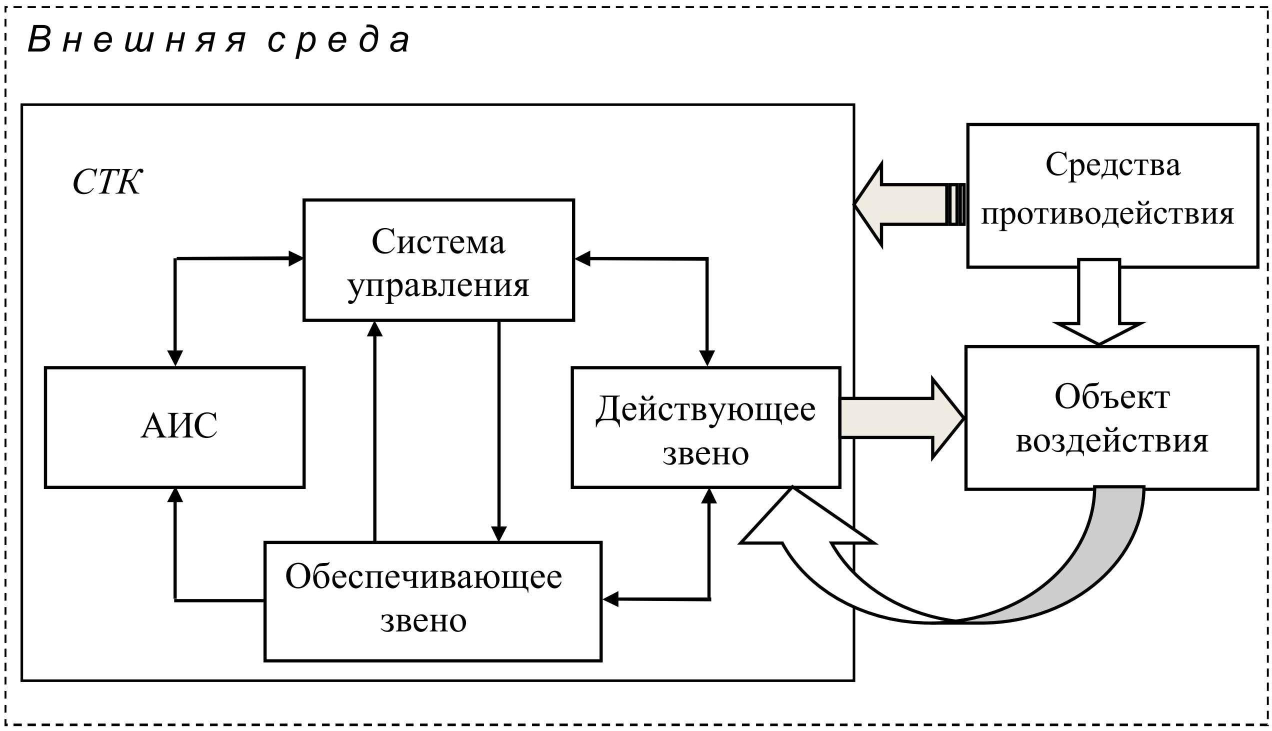


Рис. 1.7 - Блок-схема операционной модели

ЭТАП I. ПОСТАНОВКА ЗАДАЧИ ИССЛЕДОВАНИЯ

1.1. Формулировка целей и назначения модели.

1.2. Определение объектов, участвующих в операции.

1.3. Разработка принципиальной схемы обстановки.

1.4. Формализация процесса функционирования элементов СТК.

1.5. Формулировка оперативно-тактических ограничений модели.

1.6. Разработка укрупненной блок-схемы операционной модели.

ЭТАП II. ОБОСНОВАНИЕ МАТЕМАТИЧЕСКОЙ СХЕМЫ МОДЕЛИ

2.1. Систематизированное представление постановки задачи моделирования с учетом априорной информации и ожидаемых результатов.

2.2. Формулировка показателя эффективности (или критерия оптимизации) СТК и его элементов.

2.3. Выбор математической схемы с учетом заданных требований и ограничений; уточнение входных, выходных и промежуточных переменных; введение необходимых допущений и аппроксимаций.

2.4. Разработка логической блок-схемы операционной модели и её сетевого графа.

2.5. Определение дополнительных зависимостей и условий для контроля адекватности операционной модели.

ЭТАП III. РАЗРАБОТКА АЛГОРИТМА ОПЕРАЦИОННОЙ МОДЕЛИ

3.1. Обоснование математических зависимостей каждого блока.

3.2. Разработка частных алгоритмов блоков.

3.3. Обоснование алгоритмов и схем сопряжения блоков.

3.4. Разработка общего алгоритма решения задачи.

3.5. Определение состава и порядка использования унифицированных блоков и программных модулей.

3.6. Разработка алгоритмов и схем адаптации (настройки) алгоритмов отдельных блоков на конкретную предметную область.

ЭТАП IV. РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

4.1. Обоснованный выбор языка программирования, общего и специального математического обеспечения.

4.2. Запись на языке программирования алгоритмов каждого блока и их отладка.

4.3. Комплексирование рабочих программ операционной модели в пакет прикладных программ.

4.4. Введение в состав ППП вспомогательных (сервисных) программных модулей (в частности, модулей графического сопровождения решения, статистической обработки результатов и т. д.).

4.5. Автоматизация решения многовариантных задач исследования.

4.6. Экспериментальная проверка (тестирование) работы алгоритма и ППП операционной модели путем решения контрольных задач.

4.7. Корректировка алгоритмов или процедур в соответствии с результатами экспериментальной проверки.

4.8. Разработка инструкций пользователю ППП по решению задач исследования.

1.6 Контрольные вопросы

1. Укажите особенности экономической системы как объекта системного исследования. Укажите факторы, которые затрудняют строгое математическое описание экономических процессов.

2. На примере конкретной экономической системы укажите план содержательного описания – обобщённой модели объекта исследования.

3. Дайте характеристику объекта и предмета исследования операций.

4. В чём проявляется отличие исследование операций от теории моделирования сложных систем?

5. Поясните сущность системного анализа – как методологии исследования сложных систем.

6. Приведите примеры прикладных задач исследования операций.

7. Какие математические методы используются в задачах исследования операций?

8. Каковы функции операциониста в задачах исследования операций?

9. Сформулируйте и приведите пример прямой задачи исследования операций.

10. Сформулируйте и приведите пример обратной задачи исследования операций.

11. Перечислите исходные данные, необходимые для построения операционной схемы?

12. Каким требованиям должна отвечать операционная модель?

13. Укажите факторы, которые на практике ведут к возникновению неопределённости в задачах ИО.

14. Перечислите этапы разработки операционной модели экономической системы.

15. Каким образом учитывается время в задачах исследования операций?

16. Назовите показатели экономической системы (процесса), которые могут быть определены по результатам операционного моделирования.

ГЛАВА 2. ЗАДАЧИ МАТЕМАТИЧЕСКОГО ПРОГРАММИРОВАНИЯ

2.1 Общая характеристика задач математического программирования

Общая задача математического программирования (ЗМП) может быть сформулирована следующим образом. *Требуется найти значения n переменных* $x_1, x_2, ..., x_n$, которые удовлетворяют системе из m соотношений:

$$\boxed{g(x_1, x_2, ..., x_n)\ \{\le, =, \ge\}\ b_i; \quad i \in \overline{1, m}}, \tag{2.1}$$

где b_i - заданные константы; $g_i(\cdot)$ - заданные функции, и максимизируют заданную функцию

$$\boxed{F = F(x_1, x_2, ..., x_n) \to \max}. \tag{2.2}$$

Соотношения (2.1) в ЗМП называются *ограничениями*. Совокупность этих ограничений образует область D_x допустимых значений переменных $x_1, x_2, ..., x_n$. Очевидно, что эта область не должна быть пуста. В противном случае исчезла бы свобода (хотя и ограниченная) в выборе значений переменных, которая позволяет сравнивать значения функции в различных точках и определять искомый максимум. Специфика системы ограничений составляет принципиальную особенность ЗМП.

Для краткости изложения в дальнейшем будем пользоваться следующей условной формой записи ЗМП в постановке (2.1), (2.2):

$$\begin{cases} F(x) \to \max; \\ g(x)\ \{\le, =, \ge\}\ b_i; \quad i \in \overline{1, m}, \end{cases} \tag{2.3}$$

где $x = (x_1, x_2, ..., x_n)$ - n- мерный вектор искомых (управляемых, оптимизируемых) переменных (параметров).

Постановку ЗМП, представленную соотношениями (2.1) и (2.2), дополним следующими замечаниями.

1. В каждом отдельном ограничении из системы (2.1) сохраняется только один из знаков $\le, =$ *или* $\ge$, причём разные ограничения могут иметь и разные знаки.

2. Величины n *и* m (n - число переменных в задаче, m - число ограничений) в общей ЗМП между собой не связаны и могут принимать любые значения при условии $n \ge 1, m \ge 0$, причём допускается, что $m \le \{<, =, >\} n$. Исключение составляет случай, когда все m ограничений (2.1) являются строгими равенствами; в этом случае величины должны быть связаны соотношением $0 \le m \le n$, т.е. число уравнений связи должно быть не больше числа переменных. В частном случае m может быть и нулём, так что общая постановка (2.1), (2.2) включает и случай, когда все ограничения (2.1) отсутствуют.

3. В ЗМП различают и по-разному оформляют два вида ограничений: областные и функциональные. Каждое областное ограничение относится только к одной переменной, ограничивая область её изменения. *Областное ограничение* может иметь, например, следующий вид:

$$x_j \ge b_j. \tag{2.4}$$

Областное ограничение представляет собой частный случай ограничения (2.4), когда функция $g_i(x)$ имеет простейший вид $g_i(x) = x_j$.

В ЗМП встречаются и двухсторонние областные ограничения:

$$b_{\min j} \le x_j \ge b_{\max j}. \tag{2.5}$$

Одно двухстороннее ограничение равносильно двум односторонним ограничениям, имеющим вид, подобный выражению (2.4): $x_j \ge b_{\min j}$ *и* $x_j \le b_{\max j}$.

4. Частным случаем областных ограничений являются так называемые ограничения на знак переменных. Эти ограничения выражают требования, состоящие в том, что некоторые или все переменные задачи должны удовлетворять условию неотрицательности. Условия неотрицательности переменных имеют вид:

$$x_j \ge 0;\ j = \overline{1,p};\quad p \le n. \tag{2.6}$$

Требование (2.6) возникло в сфере приложений ЗМП к решению практических задач принятия решений и отражает тот факт, что в большинстве задач проектирования переменные не могут принимать отрицательных значений вследствие их физической сути.

5. В ряде ЗМП в качестве ограничений дополнительно к требованию (2.6) может вводиться условие, по которому некоторые или все переменные могут принимать лишь некоторые дискретные значения, например, только целочисленные. Это условие в частном случае может иметь вид

$$x_j = 0,\ \Delta_j,\ 2\cdot\Delta_j,\ 3\cdot\Delta_j, \ldots,\quad j \in \overline{1,k};\quad k \le n, \tag{2.7}$$

где Δ_j - некоторая константа.

Условия, подобные (2.7), можно назвать требованиями дискретности или ограничениями на непрерывность переменных. Если в частном случае в условии (2.7) параметр $\Delta_j = 1$, то условие превращается в требование целочисленности переменных:

$$x_j = 0,\ 1,\ 2,\ 3, \ldots,\quad j \in \overline{1,k};\quad k \le n. \tag{2.8}$$

В дальнейшем условимся обозначение вида (2.1) оставлять лишь для собственно функциональных ограничений. Наличие же областных ограничений, требований неотрицательности и дискретности переменных будем отмечать особо с помощью обозначений вида (2.4)-(2.8).

В настоящее время не существует единого метода, одинаково пригодного для решения всех видов ЗМП. В зависимости от вида целевых функций $F(x)$ и ограничений $g_i(x)$ среди ЗМП выделяют частные типы задач, для решения которых разработаны специальные методы.

Рассмотрим классификацию ЗМП и дадим краткую характеристику методов их решения. В основу традиционной классификации задач математического программирования положены различия в характере критериальной функции $F(x)$ и функции ограничения $g_i(x)$.

Задачи математического программирования подразделяют на два больших класса: *классические* и *неклассические*. Основным признаком такого деления выступает дифференцируемость критериальной функции и функциональных ограничений.

1. Классические ЗМП (классические задачи оптимизации). К ним относятся задачи, которые удовлетворяют совокупности следующих признаков: а) непрерывности критериальной функции $F(x)$ и функциональных ограничений $g_i(x)$ и наличия у них непрерывных частных производных по крайней мере до второго порядка; б) отсутствия среди функциональных ограничений неравенств, что влечёт за собой требование $m \le n$; в) отсутствия областных ограничений и требований неотрицательности; г) отсутствия требований дискретности переменных. Если хотя бы одно из этих требований не удовлетворяется, задача относится к неклассическим задачам.

Следует отметить, что из перечисленных требований принципиальным являются лишь первое и последнее. Если же не выполняются другие требования, то за счёт введения вспомогательных переменных и, следовательно, увеличения размерности задачи эти требования всегда могут быть удовлетворены. Под *размерностью задачи* здесь понимается число n компонентов в векторе $\{x_j\}$ оптимизируемых параметров (число переменных). Увеличение размерности задачи обычно усложняет процесс её решения.

Классические ЗМП, в свою очередь, подразделяются на два подкласса по признаку отсутствия или наличия ограничений: *задачи поиска безусловного экстремума и задачи поиска условного экстремума с ограничениями – равенствами.*

2. Неклассические ЗМП (неклассические задачи оптимизации). В таких задачах на значения вектора параметров $x = \{x_j\}$, кроме функциональных ограничений вида $g_i(x)\ \{\le, =, \ge\} b_i;\ \ i \in \overline{1,m}$, где $g_i(x)$ - заданные функции, обычно накладываются ограничения на знак, выражающие требования неотрицательности всех или некоторой части компонент вектора $x = \{x_j\}$: $x_j \ge 0,\ i \in \overline{1,p},\ \ p \le n$.

В дальнейшем для простоты записи будем считать, что требование (2.6) предъявляется ко всем компонентам вектора $x = \{x_j\}$, т.е. $p = n$. Если в частном случае это не так, то всегда можно за счёт введения дополнительных переменных перейти к указанному случаю $p = n$.

Общая постановка неклассической непрерывной (т.е. не имеющей требований дискретности) ЗМП обычно записывается в виде

$$\boxed{F(x) \to \max(\min);\quad g_i(x) \ge b_i,\quad i \in \overline{1,m};\quad x_j \ge 0,\quad j \in \overline{1,n}.} \tag{2.9}$$

Неклассические ЗМП обычно подразделяют на два класса: *специальные и неспециальные.* К специальным ЗМП относят такие, для которых вследствие каких-то специфических особенностей их критериальных (целевых) функций и функциональных ограничений разработаны специальные методы решения, например: задача линейного программирования. Специфическая особенность этой задачи – линейность критериальной функции и функциональных ограничений.

Совокупность специальных методов в работах по математическому программированию получила и другое название – непрямые методы поиска экстремума. В этом названии нашёл отражение тот факт, что траектория поиска экстремума этими методами проходит не непосредственно в направлении экстремума, а “окольными путями” с учётом специфических особенностей критериальной функции и функциональных ограничений. Например, при решении задачи линейного программирования (ЗЛП) симплекс-методом траектория поиска начинается в некоторой начальной опорной точке на границе области допустимых значений вектора параметров и затем переходит из одной опорной точки в другую в направлении возрастания критериальной функции.

Неспециальные ЗМП решают с использованием неспециальных методов поиска экстремума. Неспециальные методы имеют и другое название – прямые методы поиска экстремума. Эти методы являются универсальными в том смысле, что при их разработке учитывались не какие-либо специфические особенности конкретной задачи оптимизации, а лишь самые общие соображения, связанные с задачей отыскания экстремума. Прямые методы применимы для решения как специальных, так и неспециальных задач математического программирования.

Разделение ЗМП на специальные и неспециальные является условным, отражающим современное состояние теории и практики математического программирования. С ростом достижений в области практики класс специальных ЗМП будет непрерывно расширяться. Выделим основные типы специальных неклассических задач математического программирования.

1 Задача линейного программирования

ЗЛП характеризуется тем, что функции $F(x)$ и $g_i(x)$ являются линейными относительно векторного аргумента $\{x\}$. Общая постановка ЗЛП может быть представлена в виде:

$$\left.\begin{aligned} &F(x)=\sum_{j=1}^{n} c_j \cdot x_j \to \max; \\ &g_i(x)=\sum_{i=1}^{n} a_{ij} \cdot x_j = b_i; \quad i \in \overline{1,m}; \\ &x_j \ge 0; \quad j \in \overline{1,n}, \end{aligned}\right\} \tag{2.10}$$

где b_j, c_j, a_{ij} - заданные постоянные величины.

Все остальные задачи математического программирования, которые нельзя свести к постановке (2.10), являются нелинейными ЗМП.

ЗЛП представляют собой хорошо разработанный класс задач. В практике исследования экономических систем и процессов линейные детерминированные модели, как правило, дают, довольно грубое приближение к реальной задаче. Более детальный анализ предметной области позволяет обнаружить скрытые нелинейные и стохастические явления. Напомним, что *нелинейным* явлением называют такое явление, в котором отсутствует прямая пропорциональность между причиной и результатом.

2 Задачи квадратичного программирования

Эти задачи характеризуются квадратичной зависимостью критериальной функции $F(x)$ и линейной зависимостью функциональных ограничений $g(x)$ от вектора параметров $x=[x_1,x_2,...,x_n]$. Общая постановка задачи квадратичного программирования имеет вид:

$$F(x)=\sum_{j=1}^{n} c_j \cdot x_j + \sum_{j=1}^{n}\sum_{k=1}^{n} d_{jk} \cdot x_j \cdot x_k \to \max;$$
$$g_i(x)=\sum_{j=1}^{n} a_{ij} \cdot x_j \ge b_i;\ i \in \overline{1,m};\quad x_j \ge 0;\ j \in \overline{1,n}, \qquad (2.11)$$

где b_i, c_j, a_{ij}, d_{jk} - заданные постоянные величины.

Для решения задач этого типа разработаны специальные методы, использующие ***теорию Куна-Таккера.***

3 Задачи выпуклого программирования

В этих задачах критериальная функция $F(x)$ и функции ограничений $g_i(x)$ относятся к классу выпуклых функций. Общая постановка задачи выпуклого программирования по форме совпадает с общей постановкой (2.11) неклассической ЗМП. Для решения задач выпуклого программирования успешно применяются методы возможных направлений. Отметим, что рассмотренные задачи линейного и квадратичного программирования (п.1 и п.2) представляют собой частные случаи задачи выпуклого программирования, поскольку линейные и квадратичные функции принадлежат к классу выпуклых функций. Следовательно, методы, разработанные применительно к задачам выпуклого программирования, применимы также для решения задач линейного и квадратичного программирования, но не наоборот.

4 Задачи с сепарабельными критериальными функциями и линейными ограничениями

Сепарабельной критериальной функцией называется такая функция n переменных, которая может быть представлена в виде суммы n функций одной переменной:

$$F(x)=F(x_1,x_2,...,x_n)=\sum_{j=1}^{n} f_j(x_j) = f_1(x_1)+f_2(x_2)+...+f_n(x_n) \qquad (2.12)$$

или произведения n функций одной переменной

$$F(x)=F(x_1,x_2,...,x_n)=\prod_{j=1}^{n} f_j(x_j) = f_1(x_1)\cdot f_2(x_2)\cdot ...\cdot f_n(x_n). \qquad (2.13)$$

В случае (2.12) целевая функция называется *аддитивной*, а в случае (2.13) – *мультипликативной сепарабельной* функцией. В формуле (2.13) Символ $\prod_{j=1}^{n}$ означает произведение n целевых функций.

<u>Постановка задачи оптимизации</u> имеет вид:

$$F(x)=F(x_1,x_2,...,x_n)=\left\{\sum_{j=1}^{n} f_j(x_j);\ \prod_{j=1}^{n} f_j(x_j)\right\}\to \max;$$
$$g_i(x_i)=\sum_{j=1}^{n} a_{ij}\cdot x_j \ge b_i;\ i\in\overline{1,m};\quad x_j \ge 0,\ j\in\overline{1,n}. \qquad (2.14)$$

Для решения подобных задач может быть использован *метод динамического программирования Беллмана*.

5 Задачи дискретного программирования

К таким задачам может быть отнесена любая ЗМП, в которой имеются дополнительные ограничения на вектор управляемых параметров $x = [x_1, ..., x_n]$, состоящие в требовании, чтобы все (*полностью дискретные задачи*) или некоторые (*частично дискретные задачи*) компоненты вектора x принимали только дискретные не отрицательные значения:

$$\left.\begin{array}{l} x_k = 0, \Delta_k, 2\cdot\Delta_k, 3\cdot\Delta_k, ..., r\cdot\Delta_k; \\ k \in \overline{1, p}; \quad p \le n; \quad x_j \ge 0; \quad j \in \overline{(p+1), n}, \end{array}\right\} \qquad (2.15)$$

где n - размерность вектора x; Δ_k - некоторый дискрет (положительная величина).

Частным случаем задач дискретного программирования, как отмечалось ранее, являются задачи целочисленного программирования. В них принимается условие $\Delta_k = 1$.

2.2 Классические задачи линейного программирования

Большое число экономических задач сводится к линейным математическим моделям. Традиционно оптимизационные линейные ММ называются моделями линейного программирования.

В общем виде задача линейного программирования формулируется следующим образом.

Максимизировать (минимизировать) функцию $f = \sum_{j=1}^{n} c_j \cdot x_j$

при ограничениях

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j \le b_i \\ \sum_{j=1}^{n} a_{ij} x_j \ge b_i \\ \sum_{j=1}^{n} a_{ij} x_j = b_i \end{cases} \qquad (2.16)$$

где x_j - управляющие переменные; a_{ij}, b_i - параметры; f - целевая функция или критерий эффективности задачи. Функция f и ограничения (2.16) – линейны.

Решить задачу линейного программирования – это значит найти значения управляющих переменных x_j, удовлетворяющих ограничениям (2.16), при которых целевая функция f принимает минимальное или максимальное значение.

В зависимости от вида целевой функции f и ограничений (2.16) можно выделить *несколько типов* задач линейного программирования или линейных моделей: общая линейная задача, транспортная задача, задача о назначениях.

Пример 2.1

Предприятие производит изделия трех видов, поставляет их заказчикам и реализует на рынке. Заказчикам требуется 1000 изделий первого вида, 2000 изделий второго вида и 2500 изделий третьего вида.

Условия спроса на рынке ограничивают число изделий первого вида 2000 единицами, второго – 3000 и третьего – 5000 единицами. Для изготовления изделий используется 4 типа ресурсов. Количество ресурсов, потребляемых для производства одного изделия, общее количество ресурсов и прибыль от реализации одного изделия каждого вида заданы в табл. 2.1.

Таблица 2.1

Тип ресурсов	Вид изделий			Всего ресурсов
	1	2	3	
1	500	300	1000	25000000
2	1000	200	100	30000000
3	150	300	200	20000000
4	200	200	400	40000000
Прибыль	20	40	50	

Как организовать производство так, чтобы:

1) обеспечить заказчиков; 2) не допустить затоваривания; 3) получить максимальную прибыль.

Решение:

Цель: получение максимальной прибыли;

Параметры задачи: все числовые данные; Управляющие переменные: x_1 - число изделий первого вида; x_2 - число изделий второго вида; x_3 - число изделий третьего вида.

Ограничения: обеспечить заказчиков, не превысить запас ресурсов, не допустить затоваривания рынка.

В соответствии с этими ограничениями выпишем область допустимых решений задачи:

$$\begin{cases} x_1 \ge 1000; \\ x_2 \ge 2000; \\ x_3 \ge 2500; \\ x_1 \le 2000; \\ x_2 \le 3000; \\ x_3 \le 5000; \\ 500x_1 + 300x_2 + 1000x_3 \le 25000000; \\ 100x_1 + 200x_2 + 100x_3 \le 30000000; \\ 150x_1 + 300x_2 + 200x_3 \le 20000000; \\ 100x_1 + 200x_2 + 400x_3 \le 40000000. \end{cases}$$

Первые три неравенства в системе соответствуют спросу заказчиков. 4-6 неравенства формализуют спрос на рынке. Последние четыре неравенства соответствуют ограничениям по ресурсам.

Целевая функция имеет: $\boxed{F = 20x_1 + 40x_2 + 50x_3 \to \max}$.

Методы решения задачи линейного программирования

1) Графический метод решения задачи линейного программирования.

Если число переменных в задаче линейного программирования равно двум, а ограничениями является система неравенств, то задачу можно решать графическим методом.

Пример 2.2

При продаже двух видов товара используется четыре типа ресурсов. Норма затрат ресурсов на реализацию единицы товара, общий объем каждого ресурса заданы в табл. 2.2.

Таблица 2.2

Ресурсы	Норма затрат ресурсов на товары		Общее кол-во ресурсов
	1-го вида	2-го вида	
1	2	2	12
2	1	2	8
3	4	0	16
4	0	4	12

Прибыль от реализации одной единицы товара первого вида составляет 2 усл. ед., второго вида – 3 усл. ед. Требуется найти оптимальный план реализации товаров, обеспечивающий торговому предприятию максимальную прибыль.

Решение. Математическая модель задачи имеет следующий вид.

Цель: получение максимальной прибыли;

Параметры задачи: все числовые данные;

Управляемые переменные: x_1 - количество реализуемых изделий первого вида; x_2 - количество реализуемых изделий второго вида.

Ограничения: обеспечить заказчиков, не превысить запас ресурсов, не допустить затоваривания рынка.

В соответствии с этими ограничениями выпишем область допустимых решений задачи:

$$\begin{cases} 2x_1 + 2x_2 \le 12; \\ x_1 + 2x_2 \le 8; \\ 4x_1 \le 16; \\ 4x_2 \le 12; \\ x_1 \ge 0; \\ x_2 \ge 0. \end{cases}$$

Целевая функция имеет: $\boxed{F = 2x_1 + 3x_2 \to \max}$.

Построим в плоскости X_1OX_2 (рис. 2.1) область допустимых значений параметров, соответствующую системе ограничений:

$$\begin{cases} x_1 + x_2 \le 6; \ \ x_1 + 2 \cdot x_2 \le 8; \\ x_1 \le 4; \ \ x_2 \le 3; \\ x_1 \ge 0; \ \ x_2 \ge 0. \end{cases}$$

Построим вектор $\vec{\nabla}(2;3)$ и прямую L, перпендикулярную этому вектору. Передвинем прямую L в направлении вектора $\vec{\nabla}(2;3)$ пока не найдем точку, проходя через которую прямая L не будет пересекать область допустимых значений $x_1 = 4$; $x_2 = 2$. $F = 4 \cdot 2 + 3 \cdot 2 = 14$.

Задача решена.

Изложенный выше графический метод применим для решения задач линейного программирования следующего вида: $\boxed{f = c_1 \cdot x_1 + c_2 \cdot x_2 \to \max(\min)}$

$$\begin{cases} a_{i1} \cdot x_1 + a_{i2} \cdot x_2 \le b_i, \quad i = \overline{1, m_1} \\ a_{i1} \cdot x_1 + a_{i2} \cdot x_2 \ge b_i, \quad i = \overline{m_1 + 1, m} \end{cases}$$

Укажем **алгоритм решения** ЗЛП графическим методом:

Шаг 1. Записывают уравнения прямых, соответствующих ограничениям и строят их на плоскости X_1OX_2.

Шаг 2. Определяют области, в которых выполняются ограничения задачи. Для этого выбирают произвольную точку на плоскости X_1OX_2 и подставляют её координаты в первую часть одного из неравенств. Если неравенство верно, то искомая полуплоскость находится с той же стороны от прямой, что и точка; в противном случае искомая полуплоскость лежит с противоположной стороны от прямой. Эти действия последовательно выполняются для всех неравенств.

Шаг 3. Определяют область допустимых решений задачи как область пересечения m полуплоскостей, соответствующих m ограничениям задачи.

Шаг 4. Определяют направление возрастания (убывания) целевой функции f. Строят вектор-нормаль $\vec{n}(c_1, c_2)$, его направление показывает направление возрастания функции f, в противоположном направлении функция убывает.

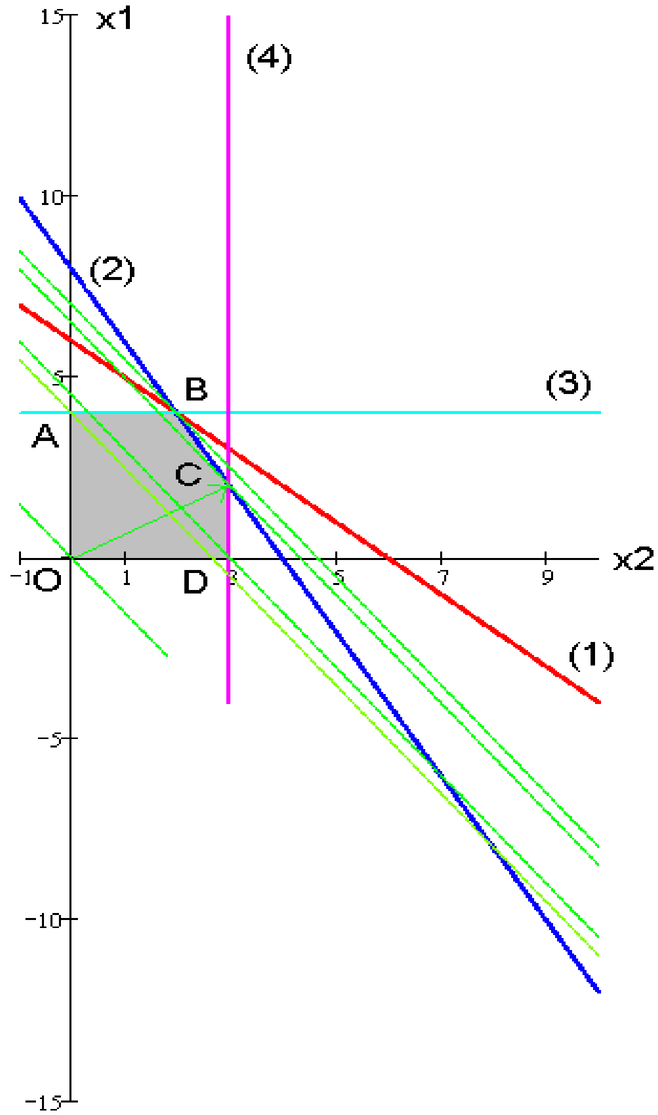


Рис. 2.1 - Область допустимых значений параметров

Шаг 5. Определяют граничную точку или точки области допустимых решений, в которых целевая функция принимает максимальное или минимальное значение.

Шаг 6. Вычисляют значения найденной точки, решая совместно уравнения, задающие прямые, на пересечении которых находится эта точка, или определяя уравнение граничной прямой области допустимых решений, с которой совпадает линия уровня целевой функции.

Пример 2.3

Решите задачу линейного программирования графическим методом.

$F = x_1 + x_2 \to \max$;

$$\begin{cases} x_1 + 2x_2 \le 14; \\ -5x_1 + 3x_2 \le 15; \\ 4x_1 + 6x_2 \ge 24; \\ x_1, x_2 \ge 0. \end{cases}$$

Решение. С использованием системы неравенств определим область допустимых значений управляемых параметров (рис.2.2). Построим вектор $\vec{\nabla}(1;1)$ и прямую L, перпендикулярную этому вектору. Передвинем прямую L в направлении вектора $\vec{\nabla}(1;1)$ пока не найдем точку, проходя через которую прямая L не будет пересекать область допустимых значений.

Задача решена.

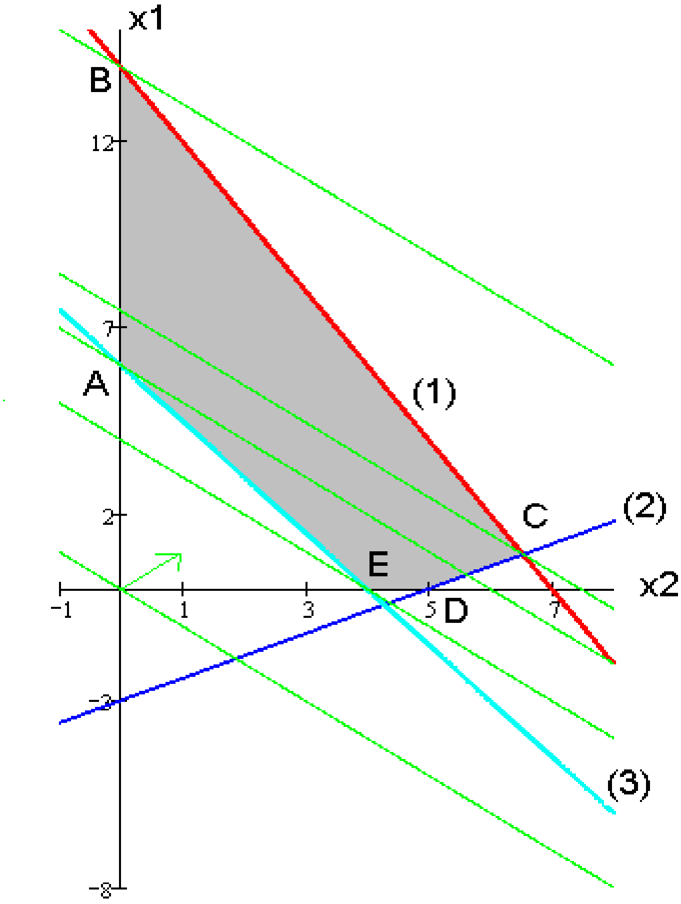


Рис. 2.2 - Область допустимых значений параметров

Пример 2.4

Решите задачу линейного программирования графическим методом

$$F = -2x_1 + x_2 \to \min;$$

$$\begin{cases} 3x_1 - 2x_2 \le 12; \\ -x_1 + 2x_2 \le 8; \\ 2x_1 + 3x_2 \ge 6; \\ x_1, x_2 \ge 0. \end{cases}$$

Решение. Построим вектор $\vec{\nabla}(1;-2)$ и прямую L, перпендикулярную этому вектору (рис. 2.3). Передвинем прямую L в противоположном направлении вектора $\vec{\nabla}(1;-2)$ пока не найдем точку, проходя через которую прямая L не будет пересекать область допустимых значений.

Задача решена.

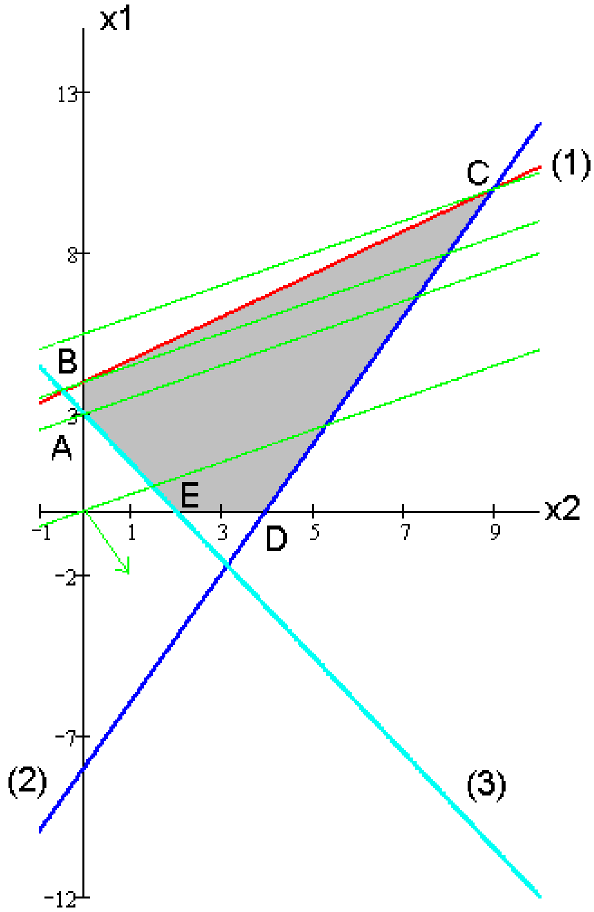


Рис. 2.3 - Область допустимых значений параметров

1. Симплекс-метод

В случае двух переменных область допустимых решений, как правило, представляет собой замкнутый многоугольник. Для n переменных, областью допустимых решений является многомерный многогранник. Оптимальное решение, как правило, это вершина (граничная точка) такого многогранника. Симплекс-метод заключается в последовательном целенаправленном обходе вершин симплекса. В каждой следующей граничной точке симплекса значение целевой функции, в общем случае, улучшается.

Для применения симплекс-метода задачу следует записать в канонической форме:

$$f = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max$$

$$\begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + \ldots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n = b_m \\ x_j \geq 0. \end{cases}$$

В канонической форме записи все переменные неотрицательны, ограничениями являются уравнения, и требуется найти такие значения x_j, при которых целевая функция имеет максимум. Переход к канонической форме записи производится с помощью действий:

1) если требуется найти минимум функции f, то заменяя f на $-f$, переходят к задаче максимизации, так как $\min f = -\max(-f)$;

2) если ограничение содержит неравенство со знаком $\leq$, то от него переходят к уравнению, добавляя в левую часть ограничения дополнительную неотрицательную переменную;

3) если ограничение содержит неравенство со знаком $\geq$, то от него переходят к уравнению, вычитая из левой части дополнительную неотрицательную переменную;

4) если в задаче какая-либо из переменных произвольна, то от неё избавляются, заменяя её разностью двух неотрицательных переменных.

Пример 2.5

Представить в канонической форме задачу: $f_1 = 5x_1 + 2x_2 - 3x_3 \to \min$

$$\begin{cases} 2x_1 - 3x_2 + x_3 \geq 10 \\ x_1 - 8x_2 - 2x_3 \leq 7 \\ 5x_1 + 2x_2 + 7x_3 = 20 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

Решение. Запишем целевую функцию в виде: $f = -5x_1 - 2x_2 + 3x_3 \to \max$.

Вычтем неотрицательную переменную x_4 из левой части первого неравенства и добавим дополнительную неотрицательную переменную x_5 к левой части второго неравенства. Произвольную переменную x_3 заменим разностью двух неотрицательных переменных $x_3 = x_6 - x_7$.

В результате получим каноническую форму записи:

$$f_1 = 5x_1 + 2x_2 - 3x_3 \to \min$$

$$\begin{cases} 2x_1 - 3x_2 + x_6 - x_7 - x_4 = 10 \\ x_1 - 8x_2 - 2x_6 + 2x_7 + x_5 = 7 \\ 5x_1 + 2x_2 + 7x_6 - 7x_7 = 20 \\ x_1, x_2, x_4, x_5, x_6, x_7 \geq 0 \end{cases}$$

Задача решена.

Симплекс-метод является методом направленного перебора решений системы. Каждое следующее решение улучшает значение целевой функции. Симплекс-метод включает два этапа:

1) определение начального решения, удовлетворяющего ограничениям;

2) последовательное улучшение начального решения и получение оптимального решения задачи.

Укажем **алгоритм симплекс-метода:**

Шаг 1. Получение начального решения. Выбирают m переменных, называемых базисными и обладающих следующим свойством: они входят с коэффициентом 1 только в одно уравнение и с коэффициентом 0 в остальные уравнения системы. Остальные $n-m$ переменных называют свободными.

Шаг 2. Выражение функции f только через свободные переменные.

Шаг 3. Проверка решения на оптимальность.

Шаг 4. Получение нового решения.

4.1 Выбор переменной, вводимой в список базисных переменных.

4.2 Выбор переменной, выводимой из списка базисных переменных.

4.3 Выполнение симплекс преобразования.

Переход на шаг 3.

Пример 2.6

Решить задачу линейного программирования: $f = 5x_1 - 2x_2 + 3x_3 \to \max$

$$\begin{cases} 3x_1 + 3x_2 - x_3 \le 15 \\ x_1 + 3x_3 \le 7 \\ -2x_1 + 8x_2 \le 20 \end{cases}$$

Решение. Запишем задачу в каноническом виде, вводя дополнительные переменные x_4, x_5, x_6.

$$\begin{cases} 3x_1 + 3x_2 - x_3 + x_4 = 15 \\ x_1 + 3x_3 + x_5 = 7 \\ -2x_1 + 8x_2 + x_6 = 20 \\ x_j \ge 0 \end{cases}$$

Начальное решение: $X_0\{x_1 = 0;\, x_2 = 0;\, x_3 = 0;\, x_4 = 15;\, x_5 = 7;\, x_6 = 20\}$.

Шаг 1. Свободные переменные: x_1, x_2, x_3. Базисные переменные: x_4, x_5, x_6.

Функция $f = 5x_1 - 2x_2 + 3x_3$ уже выражена через свободные переменные.

Введение переменной в список базисных переменных означает, что ей приписывается отличное от нуля положительное значение, т.е. её значение увеличивается. Из формулы для целевой функции $f = 5x_1 - 2x_2 + 3x_3$ видно, что увеличение значения x_2 приводит только к уменьшению f, т.е. переменную x_2 бессмысленно вводить в список базисных переменных. Увеличение переменных x_1 и x_3 приводит к увеличению значения f, при этом на большую величину значение изменяется с увеличением x_1, следовательно, переменная x_1 должна быть базисной переменной.

Для определения переменной, выводимой из списка базисных переменных, надо в соответствии с алгоритмом симплекс-метода найти отношения элементов столбца свободных членов к элементам разрешающего столбца и среди них выбрать минимальное:

$$\min\left\{\frac{15}{3}; \frac{7}{1}; \frac{20}{-2}\right\} = \{5; 7; +\infty\} = 5\ ,$$

следовательно, из списка базисных переменных надо вывести x_4.

Аналогичным образом проводятся шаги 2 и 3.

Шаг 2. Свободные переменные: x_4, x_2, x_3. Базисные переменные: x_1, x_5, x_6.

Выразим базисные переменные через свободные переменные.

$$x_1 = \frac{15 - 3x_2 + x_3 - x_4}{3};$$

$$x_5 = 7 - x_1 - 3x_3 = 7 - \frac{15 - 3x_2 + x_3 - x_4}{3} - 3x_3 = \frac{-8 + 3x_2 - 10x_3 + x_4}{3};$$

$$x_6 = 20 + 2x_1 - 8x_2 = 20 + \frac{30 - 6x_2 + 2x_3 - 2x_4}{3} - 8x_2 = \frac{90 - 30x_2 - 2x_3 - 2x_4}{3}.$$

Выразим целевую функцию через базисные переменные:

$$f = \frac{75 - 15x_2 + 5x_3 - 5x_4}{3} - 2x_2 + 3x_3 = \frac{75 - 30x_2 - 5x_4 + 14x_3}{3}.$$

Переменная x_3 вводится в базисные переменные.

Переменная x_5 выводится из базисных, т.к. $\min\left\{\frac{15}{-1};\frac{7}{3};\frac{20}{0}\right\} = \frac{7}{3}$.

Шаг 3. Свободные переменные: x_4, x_2, x_5. Базисные переменные: x_1, x_3, x_6.

Выразим базисные переменные через свободные.

$$x_1 = \frac{7 - 9x_2 - 3x_4 - x_5 + 45}{10} = \frac{52 - 9x_2 - 3x_4 - x_5}{10};$$

$$x_3 = \frac{3x_2 + x_4 - 3x_5 + 6}{10};$$

$$x_6 = 20 + 2x_1 - 8x_2 = \frac{152 - 57x_2 - 3x_4 - x_5}{5}.$$

Выразим целевую функцию через базисные переменные:

$$f = \frac{52 - 9x_2 - 3x_3 - x_4}{3} - 2x_2 + \frac{9x_2 + 3x_4 - 9x_5 + 18}{10} = \frac{278 - 66x_2 - 12x_4 - 9x_5}{10}.$$

Так как все коэффициенты в целевой функции при свободных коэффициентах отрицательны, то найдено оптимальное решение $f = 27{,}8$ при $X_3\{x_1 = 5{,}2;\ x_2 = 0;\ x_3 = 0{,}6;\ x_4 = 0;\ x_5 = 0;\ x_6 = 30{,}4\}$.

Задача решена.

Пример 2.7

Решить задачу линейного программирования:

$$f = y_1 + 4y_2 + y_3 \to \max$$

$$\begin{cases} y_1 + y_2 + y_3 \le 3 \\ 4y_1 + 2y_2 \le 2 \end{cases}$$

Решение. Запишем задачу в каноническом виде

$$\begin{cases} y_1 + y_2 + y_3 + y_4 = 3 \\ 4y_1 + 2y_2 + y_5 = 2 \end{cases}$$

Начальное решение имеет вид: $Y_0\{y_1 = 0;\ y_2 = 0;\ y_3 = 0;\ y_4 = 3;\ y_5 = 2\}$.

Шаг 1. *Свободные переменные*: y_1, y_2, y_3.

Базисные переменные: y_4, y_5.

Функция $f = y_1 + 4y_2 + y_3$ уже выражена через свободные переменные.

Выразим базисные переменные через свободные переменные

$$y_4 = 3 - y_1 - y_2 - y_3; \quad y_5 = 2 - 2y_2 - 4y_1.$$

Переменная y_2 вводится в базисные переменные.

Переменная y_5 выводится из базисных, так как $\min\left\{\frac{3}{1}; \frac{2}{2}\right\} = 1$.

Шаг 2. Свободные переменные: y_4, y_2.

Базисные переменные: y_1, y_3, y_5.

Выразим базисные переменные через свободные переменные

$$y_2 = \frac{2 - 4x_1 - y_5}{2};$$

$$y_4 = 3 - y_1 - y_3 - y_2 = 3 - y_1 - y_3 - \frac{2 - 4y_1 - y_5}{2} = \frac{4 + 2y_1 - 2y_3 + y_4}{2}.$$

Недопустимое базисное решение: $Y_2\{y_1 = 0;\ y_2 = 1;\ y_3 = 0;\ y_4 = 2;\ y_5 = 0\}$.

Выразим целевую функцию через базисные переменные:

$$f = y_1 + 2(2 - 4y_1 - y_5) + y_3 = 4 - 7y_1 - 2y_5 + y_3.$$

Переменная y_3 вводится в базисные переменные.

Переменная y_4 выводится из базисных, т.к. $\min\left\{\frac{3}{1}; \frac{2}{0}\right\} = 3$.

Шаг 3. Свободные переменные: y_3, y_2.

Базисные переменные: y_1, y_4, y_5.

Выразим базисные переменные через свободные переменные

$$y_2 = \frac{2 - 4x_1 - y_5}{2}; \quad y_3 = 3 - y_1 - y_4 - \frac{2 - y_5 - 4y_1}{2} = \frac{4 + 2y_1 - 2y_4 + y_5}{2}.$$

Выразим целевую функцию через базисные переменные:

$$f = y_1 + 2(2 - 4y_1 - y_5) + 2 + y_1 - y_4 + \frac{y_5}{2} = 6 - 6y_1 - y_4 - \frac{3}{2}y_5.$$

Так как все коэффициенты в целевой функции при свободных коэффициентах отрицательны, то нами получено оптимальное решение $f = 6$ при $Y_3\{y_1 = 0;\ y_2 = 1;\ y_3 = 3;\ y_4 = 0;\ y_5 = 0\}$.

Задача решена.

Пример 2.8

Пусть требуется составить математическую модель указанной ниже задачи, а затем решить её графическим методом, симплекс-методом, методом симплекс-таблиц. Предприятие изготавливает и продает краску двух видов: для внутренних и внешних работ. Для производства краски используется два исходных продукта А и В. Расходы продуктов А и В на 1 т соответствующих красок и запасы этих продуктов на складе приведены в табл. 2.3.

Цена за 1 т краски для внутренних работ составляет 2000 руб., краска для наружных работ продается по 1000 руб. за 1 т. *Требуется определить*: какое количество краски каждого вида следует производить предприятию, чтобы получить максимальный доход?

Таблица 2.3

Исходный продукт	Расход продуктов (в тоннах на 1 т краски)		Запас продукта на складе (т)
	Краска для внутренних работ	Краска для внешних работ	
A	1	2	3
B	3	1	3

Решение. Составим математическую модель задачи.

<u>Переменные задачи</u>. Обозначим: x_1 - количество производимой краски для внутренних работ; x_2 - соответствующее количество краски для наружных работ.

<u>Ограничения</u>, которым должны удовлетворять переменные задачи: $x_1, x_2 \geq 0$; по расходу продукта А: $x_1 + 2x_2 \leq 3$; по расходу продукта В: $3x_1 + x_2 \leq 3$.

В левых частях последних двух неравенств определены расходы продуктов А и В, а в правых частях неравенств записаны запасы этих продуктов.

Целевая функция задачи. Обозначим через Z - доход от продажи краски (в тыс. рублей), тогда целевая функция задачи записывается так: $Z = 2x_1 + x_2$.

В рассматриваемом примере содержатся только две переменные x_1 и x_2, поэтому задачу можно решить графически.

В системе координат X_1OX_2 (рис. 2.4) построим область допустимых значений управляемых переменных, отвечающую системе неравенств

$$\begin{cases} x_1 \leq 3 - 2 \cdot x_2; \quad x_1 \leq \dfrac{3 - x_2}{3}; \\ x_1, x_2 \geq 0. \end{cases}$$

Построим вектор $\vec{\nabla}(1;2)$ и прямую L, перпендикулярную этому вектору. Передвинем прямую L в направлении вектора $\vec{\nabla}(1;2)$ пока не найдем точку, проходя через которую прямая L не будет пересекать область допустимых значений

$$x_1 = \frac{3}{5}; \; x_2 = \frac{6}{5}. \quad Z = \frac{3}{5} \cdot 2 + \frac{6}{5} = \frac{12}{5}.$$

Решим данную задачу **симплекс-методом.** Запишем задачу в каноническом виде, вводя дополнительные переменные x_3, x_4.

$$\begin{cases} x_1 + 2x_2 + x_3 = 3 \\ 3x_1 + x_2 + x_4 = 3 \\ x_j \geq 0 \end{cases}$$

Начальное решение: $X_0\{x_1 = 0;\, x_2 = 0;\, x_3 = 3;\, x_4 = 3\}$.

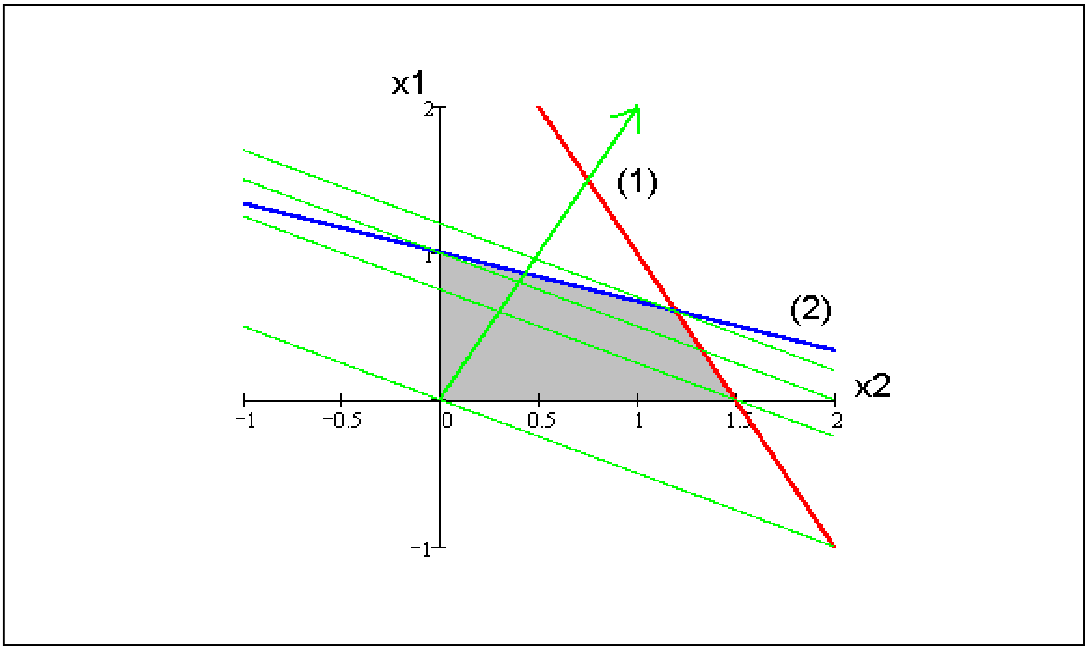


Рис. 2.4 - Область допустимых значений параметров

Шаг 1. Свободные переменные: x_1, x_2. Базисные переменные: x_3, x_4.

Функция $Z = 2x_1 + x_2$ уже выражена через свободные переменные.

Увеличение переменных x_1 и x_2 приводит к увеличению значения f, при этом на большую величину значение изменяется с увеличением x_1, следовательно переменная x_1 должна быть базисной переменной.

Для определения переменной, выводимой из списка базисных переменных, надо в соответствии с алгоритмом симплекс-метода найти отношения элементов столбца свободных членов к элементам разрешающего столбца и среди них выбрать минимальное:

$$\min\left\{\frac{3}{1}; \frac{3}{3}\right\} = 1,$$

следовательно, из списка базисных переменных надо вывести x_3.

Шаг 2. Свободные переменные: x_4, x_2. Базисные переменные: x_3, x_1.

Выразим базисные переменные через свободные переменные

$$x_1 = \frac{3 - x_2 - x_4}{3}; \quad x_3 = \frac{6 - 5x_2 + x_4}{3}.$$

Выразим целевую функцию через базисные переменные:

$$Z = \frac{6 - 3x_2 - x_4}{3} + x_2 = \frac{6 + x_2 - x_4}{3}.$$

Переменная x_2 вводится в базисные переменные. Переменная x_3 выводится из базисных, так как $\min\left\{\frac{3}{2}; \frac{3}{1}\right\} = \frac{3}{2}$.

Шаг 3. Свободные переменные: x_4, x_3. Базисные переменные: x_2, x_1.

Выразим базисные переменные через свободные переменные

$$x_1 = \frac{3 + x_3 - 2x_4}{5}; \quad x_2 = \frac{-3x_3 + x_4 + 6}{5}.$$

Выразим ЦФ через базисные переменные: $Z = \frac{12 - x_3 - 3x_4}{5}$. Так как все коэффициенты в ЦФ при свободных коэффициентах отрицательны, то найдено оптимальное решение $Z = \frac{12}{5}$ при условии $X_3\left\{x_1 = \frac{3}{5};\, x_2 = \frac{6}{5};\, x_3 = 0;\, x_4 = 0\right\}$.

Решим задачу с использованием симплекс-таблиц.

Симплекс-таблица составляется из коэффициентов при переменных x_1, x_2, x_3, x_4 и чисел, стоящих в правых частях уравнений – ограничений задачи; в первой строке записываются элементы первого уравнения, во второй – второго. В последней строке симплекс-таблицы записываются коэффициенты и правая часть целевой функции. Таким образом, симплекс-таблица содержит две строки коэффициентов (по числу ограничений задачи) и строку коэффициентов целевой функции. Число столбцов в симплекс-таблице равно числу переменных задачи плюс один столбец правых частей равенств:

	x_1	x_2	x_3	x_4	b
(1)	1	2	1	0	3
(2)	3	1	0	1	3
(3)	-2	-1	0	0	0

Симплекс-таблица определяет частное решение системы уравнений-ограничений:

$$\begin{cases} x_1 + 2x_2 + x_3 = 3 \\ 3x_1 + x_2 + x_4 = 3, \end{cases}$$

при котором свободные переменные равны нулю ($x_1 = 0, x_2 = 0$), а базисные переменные равны правым частям соответствующих строк ($x_3 = 3, x_4 = 3$).

Значение ЦФ Z всегда равно числу, стоящему в правом нижнем углу таблицы ($Z = 2 \cdot 0 + 1 \cdot 0 = 0$). Первая симплекс-таблица соответствует начальному решению ЗЛП ($x_1 = 0, x_2 = 0, x_3 = 3, x_4 = 3, Z = 0$).

Симплекс-метод заключается в последовательном перемещении по вершинам многоугольника допустимых решений. Каждой вершине соответствует своя симплекс-таблица, которая получается из предыдущей при помощи симплекс-преобразования. Симплекс-преобразованию предшествует выбор разрешающей строки и разрешающего столбца.

В качестве разрешающего столбца берут столбец, у которого коэффициент в строке ЦФ является отрицательным и наибольшим по модулю. Если в данной симплекс-таблице строка ЦФ не содержат отрицательных коэффициентов, то решение ЗЛП получено, и симплекс-таблица определяет решение задачи, при котором целевая функция Z принимает максимальное значение.

Разрешающая строка определяется по отношениям коэффициентов столбца b к соответствующим коэффициентам разрешающего столбца. Разрешающей будет строка, для которой это отношение минимально. При этом для нулевых и отрицательных коэффициентов разрешающего столбца отношения не вычисляются (и соответствующие строки не могут быть разрешающими).

Для первой симплекс-таблицы разрешающим столбцом является первый столбец (свободная переменная x_1 будет преобразована в базисную). Среди отношений коэффициентов столбца b к коэффициентам разрешающего столбца: $\frac{3}{1}$ и $\frac{3}{3}$ минимальным будет отношение $\frac{3}{3}$: разрешающей строкой будет вторая строка (базисная переменная x_4 будет преобразована в свободную).

x_1	x_2	x_3	x_4	b
1	2	1	0	3
$\longrightarrow \boxed{3}$	1	0	1	3
-2	-1	0	0	0

На пересечении разрешающего столбца и разрешающей строки находится разрешающий элемент. Процедура симплекс-преобразования состоит в том, чтобы на месте разрешающего элемента получить единицу, а все остальные элементы разрешающего столбца сделать нулевыми. При этом допускается выполнение только двух операций со строками симплекс-таблицы:

а) разрешающую строку можно делить (умножать) на любое число;

б) из любой строки можно вычитать элементы разрешающей строки или к любой строке можно прибавлять элементы разрешающей строки.

Шаг 1. Выполним преобразование первой симплекс-таблицы.

1) Делим элементы разрешающей строки на 3

x_1	x_2	x_3	x_4	b
1	2	1	0	3
1	1/3	0	1/3	1
-2	-1	0	0	0

2) Из элементов первой строки вычитаем элементы второй (разрешающей) строки (при этом первый элемент разрешающего столбца равен нулю)

x_1	x_2	x_3	x_4	b
0	5/3	1	-1/3	2
1	1/3	0	1/3	1
-2	-1	0	0	0

3) К элементам третьей строки прибавляем элементы второй (разрешающей) строки, предварительно умножив их на два (при этом третий элемент разрешающего столбца будет равен нулю):

x_1	x_2	x_3	x_4	b
0	5/3	1	-1/3	2
1	1/3	0	1/3	1
0	-1/3	0	2/3	2

Преобразование закончено. Полученной симплекс-таблице соответствует следующее решение: базисные переменные $x_1 = 1, x_3 = 2$ (их значения определяются с помощью стрелок, указанных в симплекс-таблицах); свободные переменные: $x_2 = 0, x_4 = 0$. Значение целевой функции $Z = 2$ (указано в правом нижнем углу таблицы).

Шаг 2. Так как в строке коэффициентов ЦФ есть отрицательный коэффициент (-1/3 во втором столбце), то преобразование продолжается. Второй столбец является разрешающим (свободная переменная x_2 переводится в базисную), минимальным среди отношений: $\frac{2}{5/3} = \frac{6}{5}$ и $\frac{1}{1/3} = 3$ является первое число, следовательно, разрешающей строкой является первая строка (базисная переменная x_3 переводится в свободную).

x_1	x_2	x_3	x_4	b
0	$\boxed{5/3}$	1	-1/3	2
1	1/3	0	1/3	1
0	-1/3	0	2/3	2

Выполнив симплекс-преобразование, получим:

x_1	x_2	x_3	x_4	b
0	1	3/5	-1/5	6/5
1	0	0	1/3	3/5
0	0	1/5	3/5	12/5

Шаг 3. Так как в строке коэффициентов целевой функции нет отрицательных, решение задачи закончено.

Оптимальное решение таково: базисные переменные: $x_1^* = \frac{3}{5}$; $x_2^* = \frac{6}{5}$;

свободные переменные: $x_3^* = 0$; $x_4^* = 0$. Максимальное значение дохода (целевой функции): $Z^* = \frac{12}{5}$.

Задача решена.

Двойственная задача линейного программирования

Рассмотрим задачу линейного программирования следующего вида:

$$f = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max$$

$$\begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n \le b_1 \\ a_{21}x_1 + \ldots + a_{2n}x_n \le b_2 \\ \vdots \\ a_{n1}x_1 + \ldots + a_{nn}x_n \le b_n \\ x_j \ge 0. \end{cases}$$

В задаче *требуется* максимизировать целевую функцию.

Двойственная задача линейного программирования имеет вид:

$$g = b_1y_1 + b_2y_2 + \ldots + b_my_m \to \min$$

$$\begin{cases} a_{11}y_1 + \ldots + a_{m1}y_m \ge c_1 \\ a_{12}y_1 + \ldots + a_{m2}y_m \ge c_2 \\ \vdots \\ a_{1n}y_1 + \ldots + a_{mn}y_m \ge c_n \\ y_j \ge 0. \end{cases}$$

В двойственной задаче требуется найти минимум целевой функции, ограничения – неравенства со знаком $\ge$.

Теорема двойственности: *если одна из взаимно двойственных задач имеет оптимальное решение x^*, то другая также имеет оптимальное решение y^*. При этом соответствующие им оптимальные значения целевых функций $f^* = f(x^*)$ и $g^* = g(y^*)$ равны.*

Пример 2.9

Сформулируйте двойственную задачу по отношению к ЗЛП:

$$F = 3x_1 + 3x_2 - 4x_3 \to \max$$

$$\begin{cases} 2x_1 + x_2 - 3x_3 \ge 18 \\ 4x_1 - 5x_2 \le 12 \\ 3x_1 - 2x_2 + x_3 \ge 14 \\ x_1, x_2, x_3 \ge 0. \end{cases}$$

Решение. Двойственная задача имеет вид: $\boxed{G = 18y_1 + 12y_2 + 14y_3 \to \min}$

$$\begin{cases} 2y_1 + 4y_2 + 3y_3 \le 3 \\ y_1 - 5y_2 - 2y_3 \ge 3 \\ 3y_1 + y_3 \le -4 \\ y_1, y_2, y_3 \ge 0. \end{cases}$$

Задача решена.

Реализация решения ЗМП с помощью ППП Excel

В **Excel** имеется процедура «Поиск решения» (Solver), реализующая стандартные численные методы решения задач нелинейного программирования: метод Ньютона и метод сопряженных градиентов, а также симплекс-метод для решения задач линейного программирования. Использование указанной процедуры для пользователя, который знаком с теоретическими основами численных методов оптимизации и обладает опытом корректной формулировки ЗМП, не представляет особого труда.

Необходимо последовательно выполнить следующие действия.

1. Ввести данные и базовые формулы в рабочий лист Excel.

2. Вызвать процедуру «*Поиск решения*» из меню *Сервис*.

3. Указать адрес ячейки, содержащей формулу для вычисления целевой функции и задать вид оптимизации (максимум, минимум, заданное значение).

4. Указать адреса рабочих (изменяемых) ячеек, которые будут содержать значения переменных задачи.

5. Задать ограничения, которым должны удовлетворять переменные.

6. Используя опцию *Параметры,* задать численный метод для решения задачи и указать параметры этой процедуры.

7. Нажать клавишу «*Выполнить»*.

Замечание. Поиск решения – это надстройка Excel, и она должна быть предварительно загружена. Чтобы загрузить эту процедуру нужно выполнить команду *Сервис* $\to$ *Надстройки*, установить флажок у надстройки *Поиск решения* и нажать кнопку *ОК*.

Решение примера 2.8 на основе электронных таблиц

1. Введение данных примера в таблицу **Excel** (рис. 2.5).

	A	B	C	D	E	F
1		Переменные				
2	Имя	Краска 1	Краска 2	Доход		
3						
4		2	1			
5		Ограничения				
6	Ресурс			Расход	Запасы	
7	A	1	2		3	
8	B	3	1		3	
9						
10						

Рис. 2.5

На рис. 2.5 «Краска 1» обозначает краску для внутренних работ, «Краска 2» - краску для наружных работ.

Для переменных задачи x_1 и x_2 отведены ячейки B3 и C3, которые называются рабочими, или изменяемыми ячейками. В изменяемые ячейки будут записаны оптимальные значения переменных.

В ячейку D4 вводится формула для вычисления целевой функции задачи (дохода) $Z = 2x_1 + x_2$. Для этого надо выполнить следующие действия:

1) перевести курсор в D4;

2) курсор на кнопку f_x (мастер функций);

3) в появившемся окне выбрать в левом столбце – «*Математические*», а в правом столбце – «СУММПРОИЗВ» (рис. 2.6);

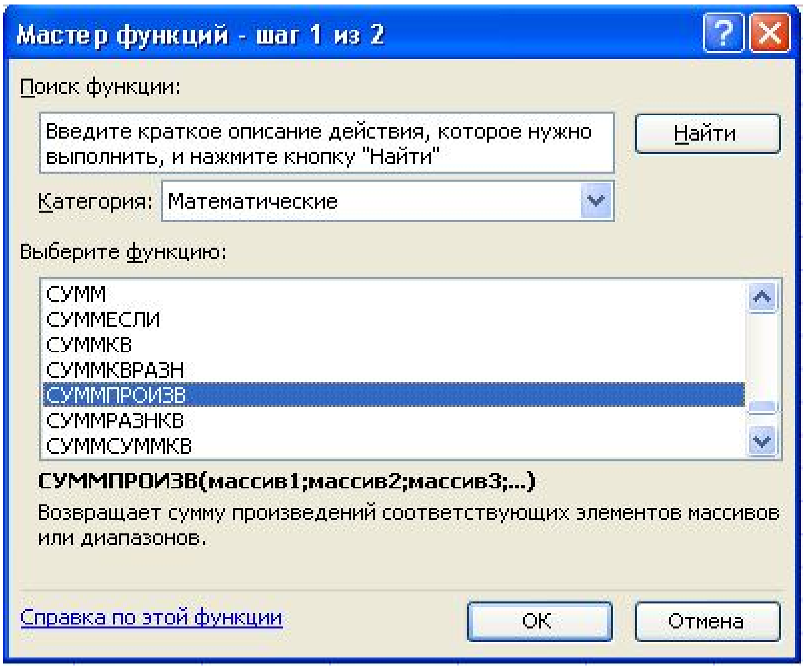


Рис. 2.6

4) в окне мастера функции нажать *ОК*, в появившемся окне (рис. 2.7) в поле «массив 1» ввести (протаскивая курсор мыши по ячейкам) адреса изменяемых ячеек B3:C3. В поле «массив 2» вводятся адреса ячеек, содержащих цены на краски B4:C4, после нажать *Готово*.

В ячейку D7 вводится формула для вычисления израсходованного количества продукта А: $x_1 + 2x_2$, а в ячейку D8 вводится формула для израсходованного количества продукта В: $3x_1 + x_2$. Обе формулы вводятся аналогично целевой функции (рис. 2.7 и 2.8).

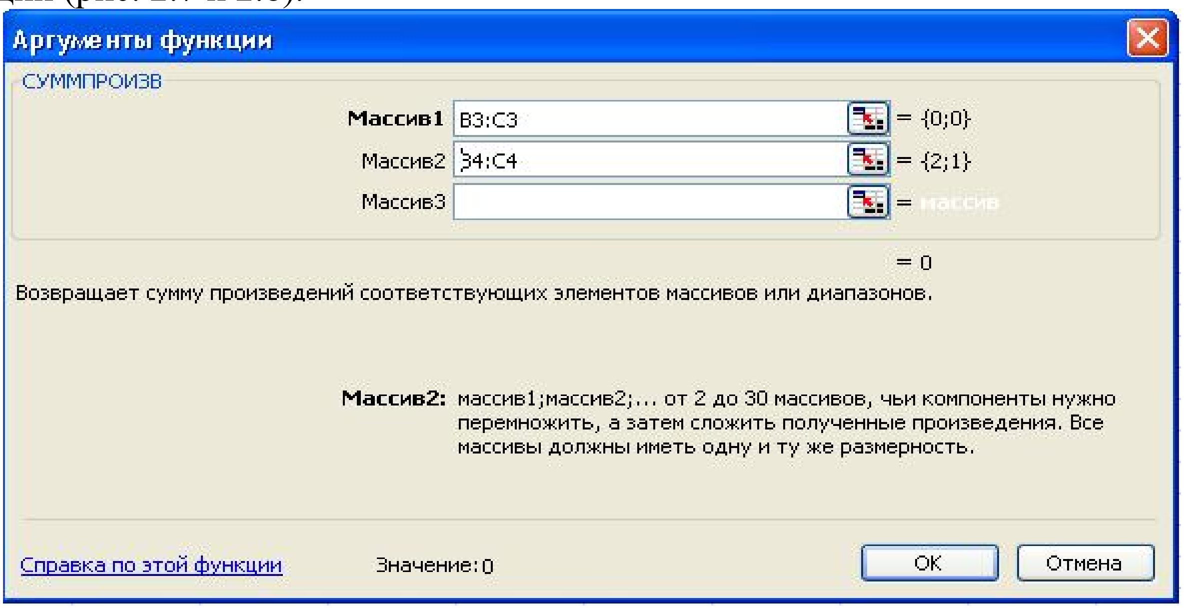


Рис. 2.7

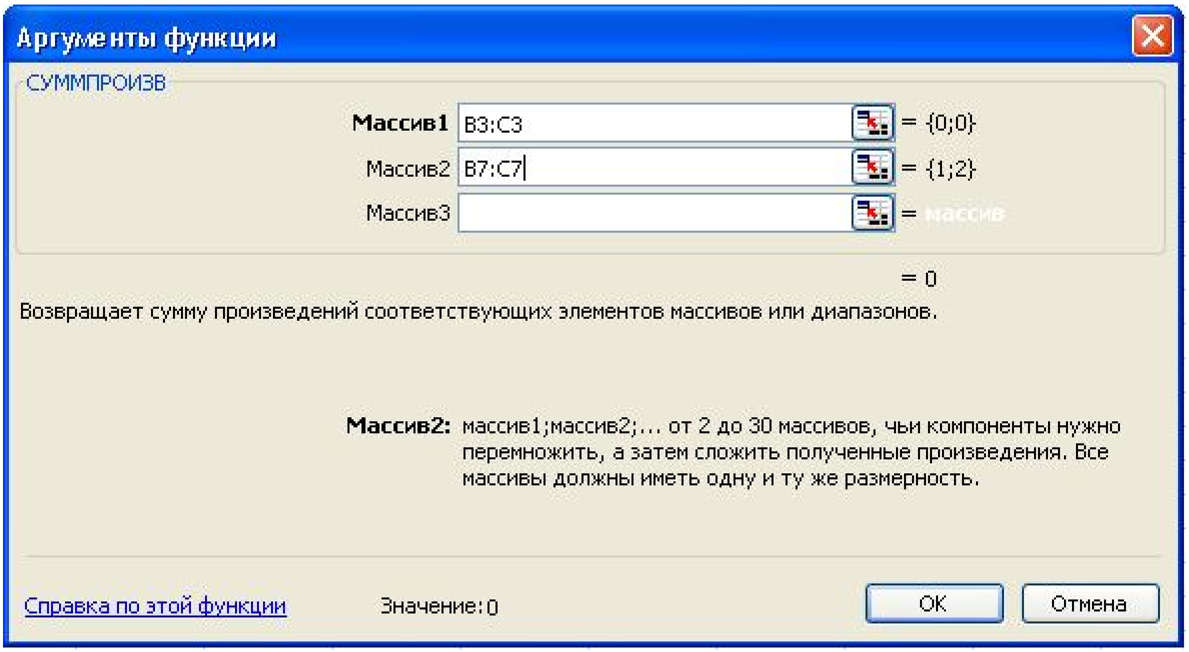


Рис. 2.8

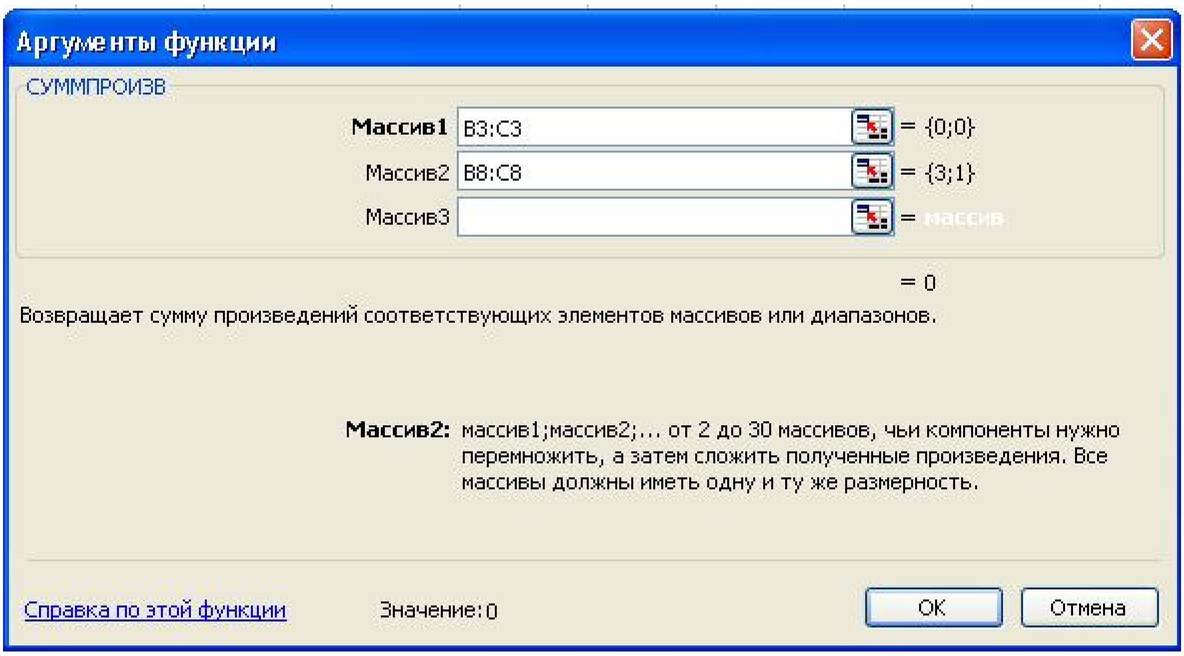


Рис. 2.9

Проверить результаты ввода можно следующим образом: при установке курсора в ячейку D4 в строке ввода должно появиться:

«СУММПРОИЗВ(B3:C3;B4:C4)».

В ячейку D7: «СУММПРОИЗВ(B3:C3;B7:C7)».

В ячейку D8: «СУММПРОИЗВ(B3:C3;B8:C8)».

Окончательно после ввода формул и данных экран имеет вид (рис. 2.10).

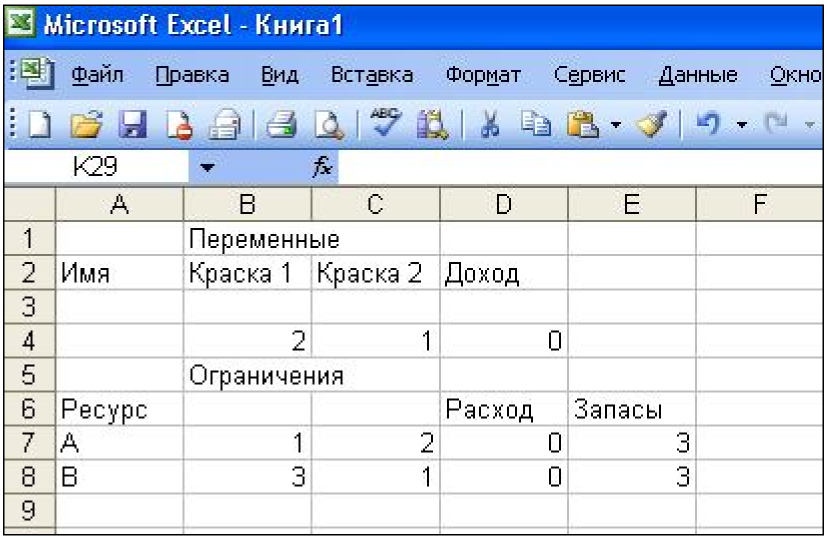


Рис. 2.10

2. Работа в окне «Поиск решения»

В меню «Сервис» выбираем процедуру «Поиск решения».

В появившемся окне (рис. 2.11) нужно установить адрес ячейки D4, содержащей формулу для вычисления целевой функции, значение целевой функции – максимальное, адреса изменяемых ячеек: B3:C3.

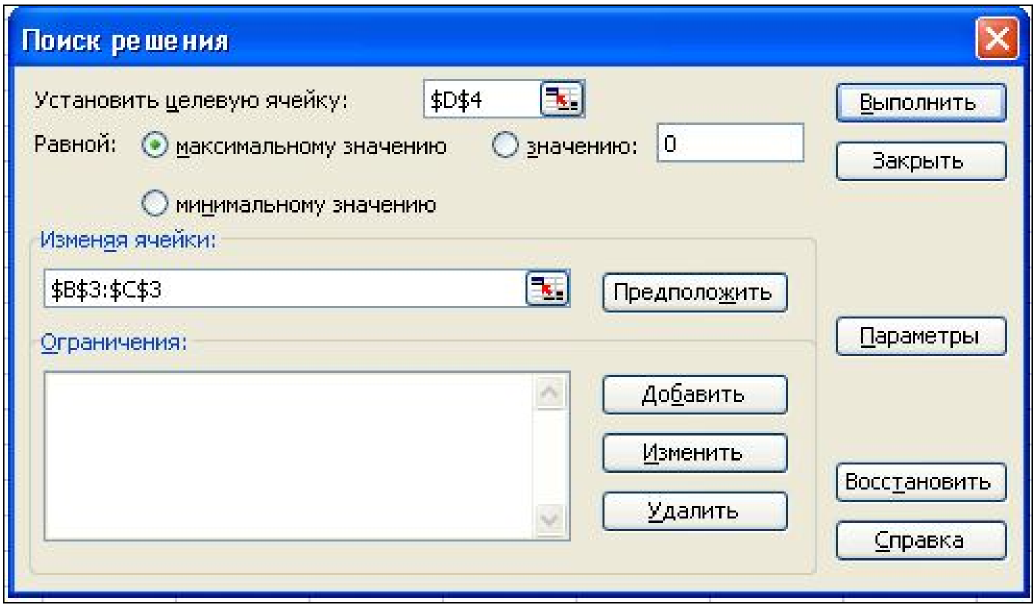


Рис. 2.11

Чтобы ввести ограничения задачи, нужно нажать кнопку «Добавить». В появившемся диалоговом окне (рис. 2.12) слева ввести адрес D7 (израсходованное количество продукта А), затем выбрать знак <= и в правой части ввести количество продукта А на складе, равное 3 (или адрес ячейки E7).

После ввода нажать кнопку «Добавить» и аналогично ввести второе ограничение: D8<=3. Снова нажать кнопку «Добавить» и ввести ограничение: B3:C3>=0 (соответствующее ограничению $x_1, x_2 >= 0$). После ввода полученного ограничения нажать ОК. После ввода всех ограничений окно «Поиска решений» будет иметь следующий вид (рис. 2.13).

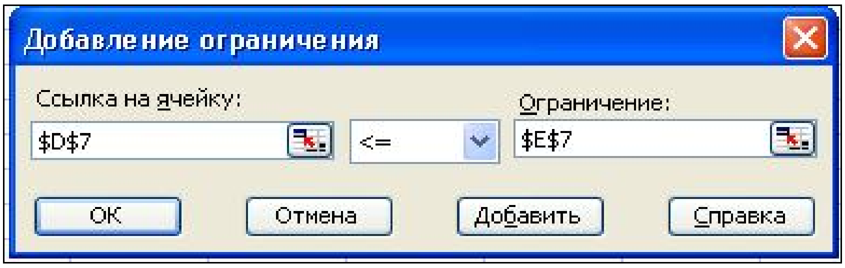


Рис. 2.12

3. Настройка параметров решения задачи

В окне «Поиск решения» нажать кнопку «*Параметры*», в появившемся окне (рис. 2.14) установить флажок в пункте «Линейная модель». В этой случае при решении задачи будет использоваться симплекс-метод. Остальные значения можно оставить без изменения. После нажать кнопку ОК.

Для решения задачи в окне «Параметры поиска решения» нажать кнопку «*Выполнить*». Если решение найдено, то появляется диалоговое окно для выбора вида отчета (рис. 2.15).

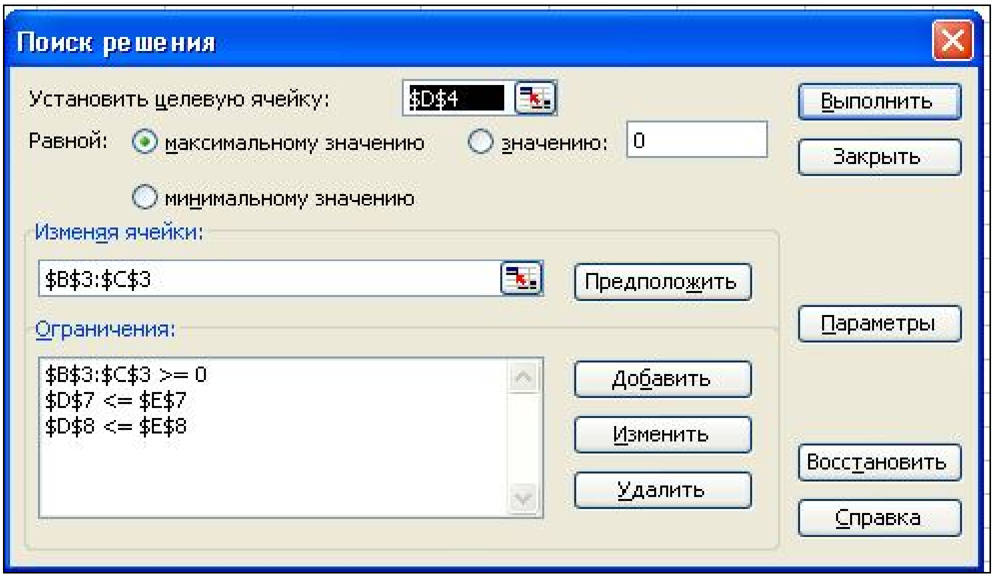


Рис. 2.13

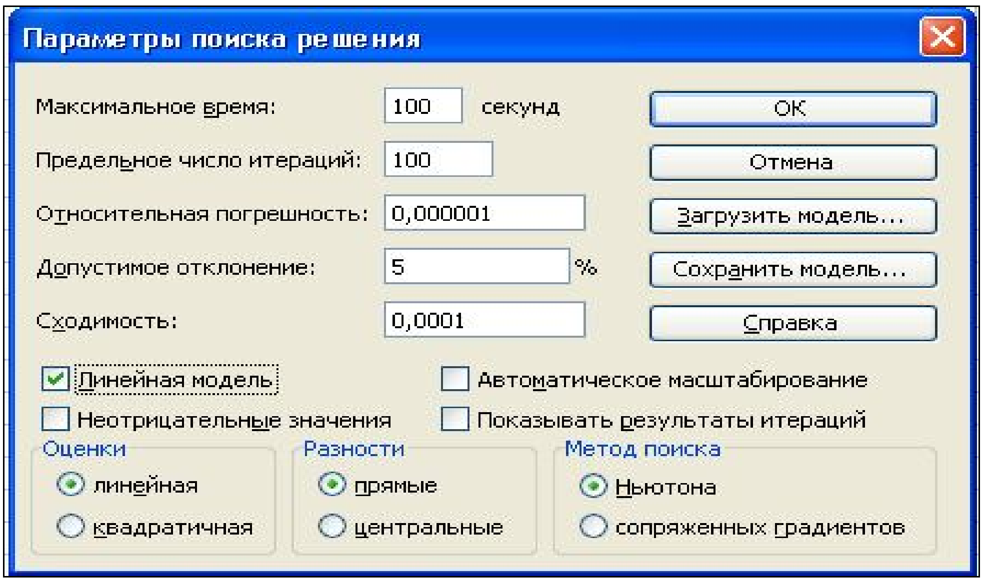


Рис. 2.14

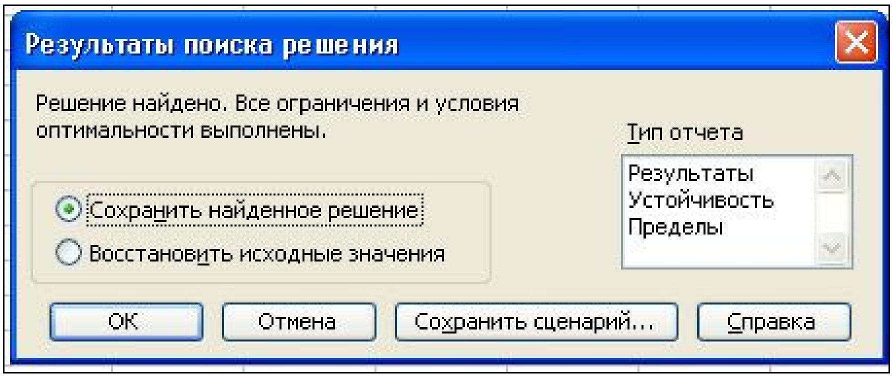


Рис. 2.15

Для просмотра результатов выбираем тип отчета: «Результаты» и нажимаем кнопку ОК. В появившихся трех таблицах (рис. 2.2.16) приводятся результаты поиска. Из таблиц видно, что в оптимальном решении:

производство краски 1=B3=0,6;

производство краски 2=C3=1,2;

при этом доход = D4=2,4;

расход ресурса А=D7=3;

расход ресурса В=D8=3.

Целевая ячейка (Максимум)

	Ячейка	Имя	Исходное значение	Результат	
	$D$4	Доход	0	2,4	
Изменяемые ячейки					
	Ячейка	**Имя**	**Исходное значение**	**Результат**	
	$B$3	Краска 1	0	0,6	
	$C$3	Краска 2	0	1,2	

Ограничения

	Ячейка	Имя	Значение	Формула	Статус	Разница
	$D$7	А Расход	3	$D$7<=$E$7	связанное	0
	$D$8	В Расход	3	$D$8<=$E$8	связанное	0
	$B$3	Краска 1	0,6	$B$3>=0	не связан.	0,6
	$C$3	Краска 2	1,2	$C$3>=0	не связан.	1,2

Рис. 2.16

Таким образом, оба ресурса являются дефицитными (соответствующие этим ресурсам ограничения называются связанными).

«Отчет по результатам» состоит из трех таблиц (рис. 2.16):

- в таблице 1 приводятся сведения о целевой функции;
- в таблице 2 приводятся значения переменных задачи;

- в таблице 3 показаны результаты поиска для ограничений задачи.

Первоначальная таблица EXCEL заполняется результатами, полученными при решении (рис. 2.17).

	A	B	C	D	E	F
1		Переменные				
2	Имя	Краска 1	Краска 2	Доход		
3		0,6	1,2			
4		2	1	2,4		
5		Ограничения				
6	Ресурс			Расход	Запасы	
7	A	1	2	3	3	
8	B	3	1	3	3	
9						

Рис. 2.17

Реализация задачи в вычислительной среде Mathcad (v.13).

Решение задачи в Mathcad 13 показано на рис. 2.18.

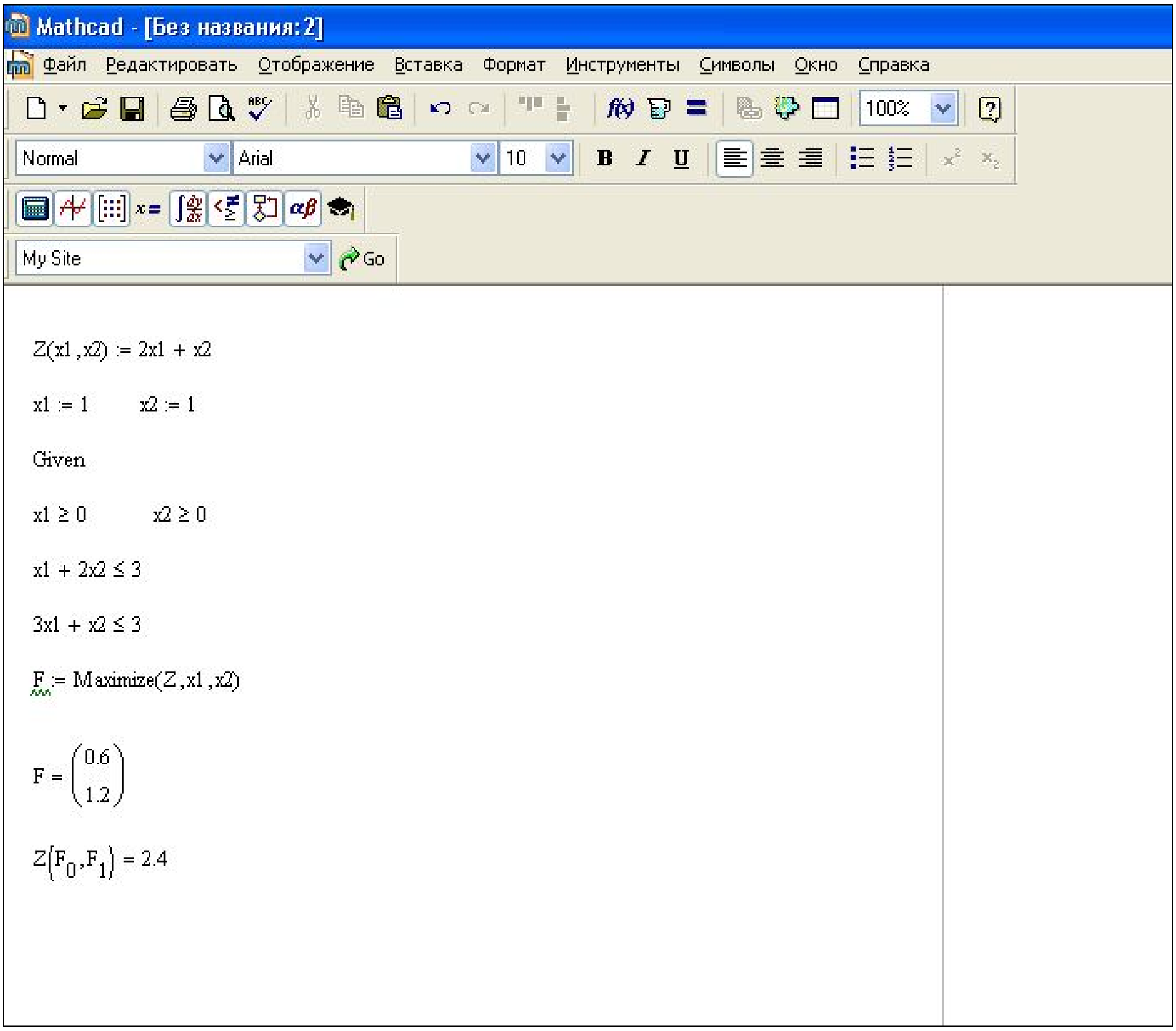


Рис. 2.18

Целочисленное программирование. Метод Гомори

Если управляемы переменные в ЗЛП определяют количество единиц неделимой продукции, то оптимальное решение должно быть получено в целых числах. К задачам такого типа относится большое число экономических задач, например, распределение производственных заказов между предприятиями, оптимальный раскрой материалов, определение загрузки оборудования, распределение транспортных средств по рейсам, задачи производства и реализации неделимой продукции. Если единица составляет малую часть от общего количества, например при планировании массового и крупносерийного производства, то для нахождения оптимального решения применяют обычный симплекс-метод и округляют полученное решение до целого. В противном случае, например, при планировании производства или реализации автомобилей, округление может привести к решению, отличному от оптимального.

Линейные задачи, решение которых должно быть получено в целых числах, называют задачами целочисленного программирования.

Математическая модель задачи целочисленного программирования имеет следующий вид:

$$f = \sum_{j=1}^{n} c_j x_j \to \max(\min).$$

$$\begin{cases} \sum\limits_{j=1}^{n} a_{ij} x_j \le b_i, \quad i = \overline{1, m_1}; \\ \sum\limits_{j=1}^{n} a_{ij} x_j \ge b_i, \quad i = \overline{m_1 + 1, m_2}; \\ \sum\limits_{j=1}^{n} a_{ij} x_j = b_i, \quad i = \overline{m_2 + 1, m}; \\ x_j \in Z, \quad j = \overline{1, n}, \end{cases}$$

где Z - множество целых чисел.

Для решения задачи целочисленного программирования может быть применен *метод Гомори*. Решение задач по методу Гомори включает 2 этапа.

Этап 1. Решение исходной задачи обычным симплекс-методом и проверка решения на целочисленность. Если решение содержит хотя бы одно дробное значение, то переходят к этапу 2, в противном случае расчет заканчивается.

Этап 2. Составление дополнительного ограничения (сечения) и решение расширенной задачи обычным симплекс-методом. Дополнительное ограничение (сечение) отсекает нецелочисленные решения.

Сечение обладает следующими двумя свойствами:

1) любое целочисленное решение ему удовлетворяет;

2) любое нецелочисленное решение задачи ему не удовлетворяет.

Сечение составляется следующим образом.

Пусть выполнен этап 1; получено

$X = \{x_1 = b_1, x_2 = b_2, \ldots, x_i = b_i, \ldots, x_m = b_m, x_{m+1} = 0, \ldots, x_n = 0\}$; b_i - дробное число.

Рассмотрим i-е ограничение:

$$b_i = x_i + a_{im+1}x_{m+1} + a_{im+2}x_{m+2} + \ldots + a_{in}x_n .$$

Так как b_i - дробное, а в правой части все переменные целые, то хотя бы одно значение a_{ij}, $j = \overline{m+1, n}$ должно быть дробным.

Возьмем дробную часть от левой и правой частей ограничения.
Обозначим через $\{r\}$ дробную часть числа r.

Дробная часть суммы не превосходит суммы дробных частей слагаемых, поэтому

$$\{x_i + a_{im+1}x_{m+1} + a_{im+2}x_{m+2} + \ldots + a_{in}x_n\} \le \{x_i\} + \{a_{im+1}x_{m+1}\} + \{a_{im+2}x_{m+2}\} + \ldots + \{a_{in}x_n\}.$$

Дробная часть произведения не превосходит произведения целого на дробную часть, следовательно:

$$\{x_i\} + \{a_{im+1}x_{m+1}\} + \{a_{im+2}x_{m+2}\} + \ldots + \{a_{in}x_n\} \le x_{m+1}\{a_{im+1}\} + x_{m+2}\{a_{im+2}\} + \ldots + x_n\{a_{in}\}.$$

В результате имеем

$$\{u_i\} \le x_{m+1}\{a_{im+1}\} + x_{m+2}\{a_{im+2}\} + \ldots + x_n\{a_{in}\}.$$

Обозначим

$$\{a_{ij}\} = q_{ij}, \quad \{b_j\} = q_i .$$

Тогда из последнего неравенства получаем

$$q_{im+1}x_{m+1} + q_{im+2}x_{m+2} + \ldots + q_{in}x_n \ge q_i .$$

Вычитая из левой части неравенства дополнительную неотрицательную переменную, переходим к уравнению

$$q_{im+1}x_{m+1} + q_{im+2}x_{m+2} + \ldots + q_{in}x_n - x_{n+1} = q_i; \quad x_{n+1} \ge 0 .$$

Путём добавления полученного ограничения к исходной задаче получаем задачу большей размерности. Эту задачу решают обычным симплекс-методом, т.е. переходят к этапу 1.

Если при решении задачи симплекс-методом имеется несколько дробных решений, то дополнительное ограничение следует составлять для значения, имеющего максимальную дробную часть.

Пример 2.10

Необходимо определить количество часов, отводимых на различные виды занятий по учебной дисциплине с учётом ограничений.

Пусть рассматривается экспериментальный курс «Методы математического моделирования» с заданным объёмом V=150 академических часов. Выделим шесть видов учебных занятий (N=6): 1. Лекции; 2. Практические занятия (*по методам решения типовых задач*); 3. Лабораторные работы (*практикум по компьютерному решению задач*); 4.Самостоятельная работа; 5. Консультации; 6. Творческие задания (проекты). Предположим, что на самостоятельную работу и консультации (согласно опыта) отводится примерно 30% всего времени обучения.

Требуется определить такой *план* распределения учебного времени x_j*, j=1,…,6, отводимых на каждый вид занятий в учебно-тематическом плане (УТП), при котором достигается наибольший эффект обучения.

Решение

Эффективность обучения в общем случае характеризуется векторным показателем и зависит от множества факторов. Пусть показатель эффективности обучения задан с помощью линейной формы

$$f(x_1,\ldots,x_N) = \beta_1 x_1 + \ldots + \beta_N x_N\,, \qquad (2.20)$$

где $\beta_j \quad (j = 1,\ldots,N)$ - априорно известные весовые коэффициенты вида занятий; $\sum_{j=1}^{N} \beta_j = 1$ - нормировочное условие;

$x_j \quad (j = 1,\ldots,N)$ - часы, отводимые на j-й вид занятий в УТП.

На значения $x_j \quad (j = 1,\ldots,N)$ помимо известного требования $\sum_j x_j = V$ дополнительно накладываются ограничения вида

$$g_i(x_1, x_2, \ldots x_N) \le c_j\,, \quad j = 1,\ldots,m\,, \qquad (2.21)$$

отражающие допустимые суммарные затраты, например, финансовые, на проведение полного курса обучения.

В указанной интерпретации исходная задача оптимизации УТП сводится к известной в практике математического программирования «задаче о рюкзаке». Её решение может быть получено на основе применения традиционных методов линейного программирования.

Для обоснованного выбора весовых коэффициентов $\beta_j \quad (j = 1,\ldots,N)$ воспользуемся методом экспертных оценок с представлением результатов в матричной форме записи.

Эксперты оценивают (сравнивают) все формы занятий по множеству частных показателей. Ограничимся рассмотрением шести показателей:

W_1 - трудоёмкость методической подготовки преподавателя;

W_2 - коэффициент корреляции, отражающий степень информационного взаимодействия в системе «преподаватель-студент»;

W_3 - коэффициент познавательной активности обучаемого;

W_4 - коэффициент теоретической значимости занятия;

W_5 - степень развития практических навыков по предмету;

W_6 - коэффициент творческой активности студентов;

Каждый эксперт попарно оценивает все варианты занятий между собой по показателям W_j (j=1…N). Сравнивая каждый раз лишь два варианта, эксперт по одному из показателей определяет отношение между ними как «больше» (>), «меньше» (<) или «равно» (=). Для удобства количественного анализа результатов указанные экспертные оценки будем отражать соответственно числами 1,5; 0,5 и 1,0.

Методику оптимизации структуры УТП представим в виде последовательности действий.

1. Отбор группы экспертов.
2. Группа экспертов (N>3) путем попарного сравнения показателей между собой выявляют степень значимости (относительный вес) β_j каждого показателя W_j. При этом заполняется табл. 2.4.

3. На основании принятой системы сравнения составляется квадратная матрица $A = \left\| a_{ij} \right\|, \quad a_{ij} = \begin{cases} 1{,}5 & \text{при} \quad X_i > X_j; \\ 1{,}0 & \text{при} \quad X_i = X_j; \\ 0{,}5 & \text{при} \quad X_i < X_j. \end{cases}$

где a_{ij} - числовая мера, определяющая степень превосходства объекта i над объектом j при сравнении по одному из показателей.

4. На основе усреднения результатов экспертной оценки вычисляются относительные веса показателей W_j, которые отражаются в табл. 2.5.
5. Группа экспертов сравнивает все предложенные формы обучения по каждому показателю W_j, j=1,…,N, раздельно. Операции производятся по той же методике, что и нахождение относительных весов показателей. Заполняется табл. 2.5.
6. Итерационным методом находятся относительные веса q_i для каждого показателя; в частности, на первом шаге итерации относительные веса показателей q_i определяются по формуле:

$$q_i = \frac{\sum_{j=1}^{N} a_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}}, \quad i = 1, \ldots, N;$$

здесь a_{ij} – среднее значение оценки предпочтения объекта i над объектом j, полученное на текущем шаге приближения.

Таблица 2.4

Сравниваемые показатели	Эксперты				
	I	II	III	IV	Система сравнения
W_1 и W_2	1,5	1,0	1,0	1,5	1,5
W_1 и W_3	0,5	1,5	0,5	1,0	1,0
W_1 и W_4	0,5	0,5	0,5	0,5	0,5
W_1 и W_5	0,5	0,5	0,5	0,5	0,5
W_1 и W_6	0,5	1,0	0,5	1,0	0,5
W_2 и W_3	1,0	1,5	1,0	0,5	1,0
W_2 и W_4	0,5	0,5	1,0	1,0	0,5
W_2 и W_5	0,5	1,0	0,5	0,5	0,5
W_2 и W_6	0,5	0,5	0,5	0,5	0,5
W_3 и W_4	1,0	0,5	1,0	1,0	1,0
W_3 и W_5	1,0	0,5	1,0	1,0	1,0
W_3 и W_6	1,0	1,5	0,5	1,0	1,0
W_4 и W_5	1,0	1,5	1,0	1,0	1,0
W_4 и W_6	1,5	1,5	1,0	1,5	1,5
W_5 и W_6	1,5	1,0	1,0	1,0	1,5

Таблица 2.5

Сравниваемые формы обучения	Показатели					
	W_1	W_2	W_3	W_4	W_5	W_6
1 и 2	1,5	0,5	0,5	1,5	0,5	0,5
1 и 3	1,5	0,5	0,5	1,5	0,5	0,5
1 и 4	1,5	1,5	1,0	1,5	0,5	0,5
1 и 5	1,5	0,5	0,5	1,5	0,5	0,5
1 и 6	0,5	0,5	0,5	1,0	0,5	0,5
2 и 3	1,0	1,0	1,0	1,0	1,5	1,0
2 и 4	1,5	1,5	1,0	1,0	1,5	0,5
2 и 5	1,5	0,5	1,0	1,0	1,5	1,5
2 и 6	0,5	0,5	0,5	1,0	1,0	0,5
3 и 4	1,5	1,5	1,0	1,0	1,5	1,0
3 и 5	1,5	1,0	1,0	1,0	1,5	1,5
3 и 6	0,5	0,5	0,5	1,0	0,5	0,5
4 и 5	0,5	0,5	1,0	1,0	1,0	1,5
4 и 6	0,5	0,5	0,5	1,0	0,5	0,5
5 и 6	0,5	1,0	0,5	0,5	0,5	0,5

7. Составляется итоговая таблица (табл. 2.6), отражающая результаты оценки коэффициентов важности форм занятий.

Таблица 2.6

Виды занятий	Показатели						Суммарный приоритет (коэффициенты β_j)
	W_1	W_2	W_3	W_4	W_5	W_6	
	Степень значимости						
	0,1332	0,1114	0,1743	0,2107	0,2107	0,1598	
1	0,2087	0,.1275	0,1111	0,2226	0,0989	0,0986	0,1449
2	0,1603	0,1719	0,1641	0,1524	0,2306	0,1613	0,1756
3	0,1603	0,1927	0,1641	0,1524	0,1742	0,1788	0,1688
4	0,1002	0,1065	0,1544	0,1524	0,1328	0,1938	0,1432
5	0,1201	0,1958	0,1641	0,1369	0,1328	0,1188	0,1422
6	0,2506	0,2055	0,2422	0,1845	0,2306	0,2486	0,2257

После вычисления значений коэффициентов β_j и конкретизации целевой функции (2.4) исходную задачу оптимизации представим в виде:

Определить:

$$\boxed{x^* = \arg\max_{x \in D_x} \sum_{j=1}^{N=6} \beta_j \cdot x_j \,,} \tag{2.22}$$

если область D_x варьирования параметров $x_1,\ldots,x_6$ задаётся системой ограничений:

$$\begin{cases} x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = V; \\ x_1 + x_2 + x_3 + x_6 = \dfrac{2 \cdot V}{3}; \\ A_1 x_1 + A_2 x_2 + A_3 x_3 + A_6 x_6 \le 120 \cdot V; \\ x_2 + x_3 \le 3x_1; \quad x_i \ge 1, \quad i = 1,\ldots,6. \end{cases} \tag{2.23}$$

Здесь $A_1 = 120;\ A_2 = A_3 = 110;\ A_6 = 300$ - коэффициенты затрат (в рублях).

Рассмотренная задача линейного программирования решена в вычислительной среде Mathcad. Экстремальное значение целевой функции и координаты экстремума определяют оптимальное решение исходной задачи:

f*(x*)=25,997; x*=(16; 43; 5; 45; 5; 36).

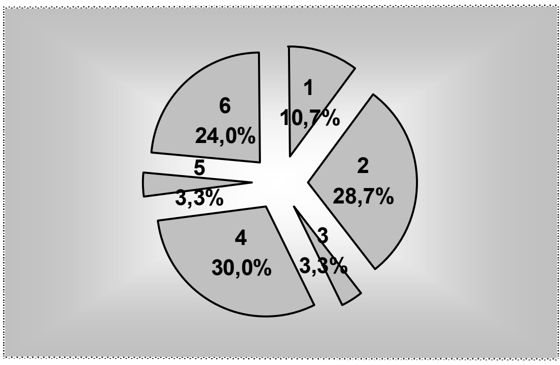
В результате вычислительного эксперимента определена оптимальная в смысле принятой схемы формализации структура УТП (табл. 2.7 и рис. 2.19).

В основе предложенного методического подхода лежит применение метода экспертных оценок, что позволяет ранжировать показатели, количественно оценить относительный вес (важность) выделенных видов занятий и интерпретировать цель оптимизации УТП курса с помощью аддитивной целевой функции вида (2.22). В результате преобразований исходную задачу оптимизации учебно-тематического плана удалось свести к канонической формулировке линейной задачи дискретного программирования (2.22) при наличии ограничений (2.23) с последующим её решением известным численным методом.

Таблица 2.7. Результаты оптимизации УТП.

Формы занятий	Кол-во часов
1. Лекции	16
2. Практические занятия	43
3. Лабораторные работы	5
4. Самостоятельная работа	45
5. Консультации	5
6. Творческие задания	36

Рис. 2.19 - Оптимальная структура УТП.



<u>Задача решена</u>.

2.3 Специальные задачи линейного программирования

Среди задач линейной оптимизации могут быть выделены два класса задач со специальной структурой: *транспортная задача и задача о назначениях*. Эти задачи используются для моделирования и оптимизации экономических проблем, связанных с формированием оптимального плана перевозок, оптимального распределения индивидуальных контрактов на транспортировки, составления оптимального штатного расписания, определения оптимальной специализации предприятий, рабочих участков и станков, оптимального назначения кандидатов на работы, оптимального использования торговых агентов. Критериальной функцией в данных задачах является линейная функция, ограничения также линейны, поэтому для решения могут применяться методы линейной оптимизации, например симплекс-метод. Однако специальная структура таких задач позволяет предложить более удобные методы их решения.

2.3.1 Транспортная задача

Построим транспортную модель на примере конкретной задачи.

Пример 2.10

Четыре предприятия экономического района для производства продукции используют некоторое сырье. Спрос на сырье каждого из предприятий соответственно составляет: 120, 50, 190 и 110 усл. ед. Сырье сосредоточено в трех местах. Предположения поставщиков сырья соответственно равны: 160, 140 и 170 усл. ед. На каждое предприятие сырье может завозиться от любого поставщика. Тарифы перевозок заранее известны и задаются матрицей

$$C = \begin{pmatrix} 7 & 8 & 1 & 2 \\ 4 & 5 & 9 & 8 \\ 9 & 2 & 3 & 6 \end{pmatrix}.$$

В i-й строке j-го столбца матрицы C указан тариф на перевозку сырья от i-го поставщика j-му потребителю, $i = 1,2,3;\ j = 1,2,3,4$. Под тарифом здесь понимается стоимость перевозки единицы сырья. ***Требуется*** составить *план перевозок*, при котором общая стоимость перевозок минимальна.

<u>Построение математической модели</u>.

Цель задачи состоит в минимизации суммарной стоимости на перевозки. Эта цель может быть достигнута с помощью оптимальной организации перевозок сырья. Следовательно, за неизвестные параметры можно принять количество сырья, перевозимого от каждого поставщика каждому потребителю.

Пусть x_{ij} - количество сырья, перевозимого от i-го поставщика j-му потребителю. Параметры задачи – число поставщиков и потребителей, предложение и спрос сырья в каждом пункте, тарифы на перевозки.

Ограничения задачи – это ограничения на предложение и спрос сырья. Предложения сырья всех поставщиков не должны быть меньше суммарного спроса на него во всех пунктах потребления. В данной задаче имеет место точное равенство между предложением и спросом

$$120+50+190+110=160+140+170=470.$$

Количество сырья, вывозимого от каждого поставщика, должно быть равно наличному количеству сырья. Количество сырья, доставленное каждому потребителю, должно равняться его спросу. Последнее ограничение - условие неотрицательности x_{ij}.

Целевая функция отражает суммарные затраты S на перевозку, равные сумме произведений тарифов на перевозку на количество перевозимого сырья от каждого поставщика каждому потребителю.

Окончательно математическая модель задачи примет вид

$$S=7x_{11}+8x_{12}+x_{13}+2x_{14}+4x_{21}+5x_{22}+9x_{23}+8x_{24}+9x_{31}+2x_{32}+3x_{33}+6x_{34}\to\min.$$

$$\begin{cases} x_{11}+x_{12}+x_{13}+x_{14}=160; \\ x_{21}+x_{22}+x_{23}+x_{24}=140; \\ x_{31}+x_{32}+x_{33}+x_{34}=170; \\ x_{11}+x_{21}+x_{31}=120; \\ x_{12}+x_{22}+x_{32}=50; \\ x_{13}+x_{23}+x_{33}=190; \\ x_{14}+x_{24}+x_{34}=110; \\ 160+140+170=120+50+190+110; \\ x_{ij}\ge 0, \quad i=1,2,3; \quad j=1,2,3,4. \end{cases}$$

Целевая функция и ограничения линейны, т.е. данная задача относится к ЗЛП, однако, благодаря особой структуре, эта задача получила специальное название: *транспортная задача* или *транспортная модель*.

В общем случае формулировка транспортной задачи имеет *следующий вид*: дано m поставщиков продукции одного вида и n потребителей; предложение каждого i-го поставщика составляет a_i единиц, $i=\overline{1,m}$; спрос каждого j-го потребителя - b_j единиц, $j=\overline{1,n}$; тарифы перевозок равны c_{ij}, $i=\overline{1,m}$; $j=\overline{1,n}$. *Требуется определить оптимальный план перевозок* продукции (т. е. количество продукции, перевозимой от каждого поставщика каждому потребителю), при котором суммарная стоимость перевозок минимальна. Заметим, что транспортная модель строится при условии линейной зависимости стоимости перевозок от количества перевозимой продукции.

Пусть x_{ij} - количество продукции, перевозимой от i-го поставщика j-му потребителю. Тогда формально транспортная задача записывается следующим образом:

$$\boxed{S=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}\to\min} \tag{2.30}$$

$$\begin{cases} \sum\limits_{j=1}^{n} x_{ij} \le a_i, \quad i = \overline{1,m}; \\ \sum\limits_{i=1}^{m} x_{ij} \ge b_j, \quad j = \overline{1,n}; \\ \sum\limits_{i=1}^{m} a_i \ge \sum\limits_{j=1}^{n} b_j; \\ x_{ij} \ge 0. \end{cases} \tag{2.31}$$

Совокупность чисел (x_{ij}), удовлетворяющая ограничениям (2.31), называется *планом перевозок* или планом транспортной задачи.

Решить транспортную задачу – это значит найти такие значения x_{ij}, которые удовлетворяют ограничениям (2.31) и доставляют минимум целевой функции (2.30). Целевая функция (2.30) определяет суммарную стоимость перевозок. Первое ограничение в (2.31) соответствуют тому, что количество продукции, вывозимой от i-го поставщика, не должно превосходить предложения i-го поставщика (для всех поставщиков). Второе ограничение в (2.31) соответствует тому, что количество продукции, ввозимой j-му потребителю, должно полностью удовлетворять спрос j-го потребителя (для всех потребителей). Третье ограничение в (2.31) соответствует тому, что суммарное предложение не должно быть меньше суммарного спроса.

Задача (2.30) – (2.31) называется ***несбалансированной транспортной моделью (задачей).***

Задача (2.30) – (2.31), в которой первые три ограничения в (2.31) имеют вид равенств, называется ***сбалансированной транспортной моделью (задачей)***.

Покажем, что любую несбалансированную транспортную модель можно свести к сбалансированной. Пусть суммарное предложение больше суммарного спроса, т.е.

$$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j . \tag{2.32}$$

Введем фиктивного $(n+1)$-го потребителя, спрос которого

$$b_{n+1} = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j ,$$

а тариф на перевозку этому потребителю от всех поставщиков равен 0.

$$c_{in+1} = 0, \quad i = \overline{1,m} .$$

Очевидно, при этом первые два неравенства в (2.31) перейдут в равенства, и к ним добавится ограничение (равенство) для $(n+1)$-го пункта потребления.

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n+1} b_j .$$

Естественно, что в реальных задачах суммарное предложение может быть меньше суммарного спроса, т.е.

$$\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j . \qquad (2.33)$$

Транспортные задачи, содержащие ограничение (2.33), также являются несбалансированными и могут быть сведены к сбалансированным с помощью ввода фиктивного $(m+1)$-го поставщика, предложение которого

$$a_{m+1} = \sum_{j=1}^{n} b_j - \sum_{i=1}^{m} a_i ;$$

стоимость перевозки от $(m+1)$-го поставщика нулевая,

$$c_{m+1\,j} = 0, \quad j = \overline{1,n} .$$

Неравенство (2.33) перейдет в равенство

$$\sum_{i=1}^{m+1} a_i = \sum_{j=1}^{n} b_j .$$

Рассмотрим сбалансированную транспортную задачу

$$S = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \to \min \qquad (2.34)$$

$$\begin{cases} \sum_{j=1}^{n} x_{ij} = a_i , \quad i = \overline{1,m}; \\ \sum_{i=1}^{m} x_{ij} = b_j , \quad j = \overline{1,n}; \\ \sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j ; \\ x_{ij} \ge 0. \end{cases} \qquad (2.35)$$

Как отмечалось выше, для решения задачи может быть применен симплекс-метод, но её особая структура (все ограничения имеют вид равенств, в которые неизвестные входят с коэффициентами, равными 1), позволяет решать ее более простыми методами. Для решения транспортной задачи составляют вспомогательную таблицу 2.8.

Таблица 2.8

Номер поставщика	Номер потребителя						Предложение
	1	2	…	j	…	n	
1	c_{11}	c_{12}	…	c_{1j}	…	c_{1n}	a_1
2	c_{21}	c_{22}	…	c_{2j}	…	c_{2n}	a_2
…	…	…	…	…	…	…	…
i	c_{i1}	c_{i2}	…	c_{ij}	…	c_{in}	a_j
…	…	…	…	…	…	…	…
m	c_{m1}	c_{m2}	…	c_{mj}	…	c_{mn}	a_m
Спрос	b_1	b_2	…	b_j	…	b_n	

В левой колонке и верхней строке таблицы записаны соответственно номера поставщиков и потребителей. В правой колонке и нижней строке записаны, соответственно, предложения каждого поставщика и спрос каждого потребителя. В правом верхнем углу клетки, стоящей на пересечении i-й строки и j-го столбца стоит тариф c_{ij} на перевозку от i-го поставщика к j-му потребителю.

Решение транспортной задачи записывают в клетки транспортной таблицы: на пересечении i-й строки и j-го столбца записывается значение x_{ij}.

Решение транспортной задачи состоит из двух этапов:

Этап 1. Нахождение начального плана перевозок (x_{ij}), удовлетворяющего ограничениям (2.35);

Этап 2. Улучшение начального плана перевозок и получение оптимального плана перевозок (x_{ij}), доставляющего минимум функции (2.34).

Заметим, что общее число неизвестных в транспортной задаче равно $m \times n$. Первые два уравнения в (2.35) не являются линейно-независимыми, так как их правые части связаны условием (третье уравнение в (2.35). Число линейно-независимых уравнений в ограничениях транспортной задачи равно, следовательно, не $m+n$, а $m+n-1$. Таким образом, *число неизвестных больше числа связывающих их уравнений так же, как и в основной задаче линейного программирования.*

Первые два уравнения в системе (2.35) можно разрешить относительно $m+n-1$ базисных переменных. Остальные $mn-(m+n-1)$ переменных являются свободными. Каждое решение транспортной задачи находят следующим образом: свободные $mn-(m+n-1)$ переменные полагаются равными нулю, а базисные $m+n-1$ переменные находят из системы ограничений (2.35). Полученное решение проверяют на оптимальность. Если решение неоптимально, то осуществляют переход к новому решению путем изменения списка базисных переменных. Эти действия повторяют до тех пор, пока не будет получено оптимальное решение, доставляющее минимум целевой функции (2.34).

Рассмотрим три метода нахождения начального решения транспортной задачи: метод «северо-западного» угла, метод минимального элемента и метод Фогеля.

Метод «северо-западного» угла

Шаг 1. Составляют транспортную таблицу.

Шаг 2. Транспортную таблицу начинают заполнять с левого верхнего (северо-западного) угла. При заполнении двигаются по строке вправо и по столбцу вниз. В клетку, находящуюся на пересечении первой строки и первого столбца, помещается максимально возможное число единиц продукции, разрешенное ограничениями на предложение и спрос: $x_{11} = \min(a_1, b_1)$. Пусть, например, $a_1 < b_1$, тогда $x_{11} = a_1$ и предложение первого поставщика полностью исчерпано. Первая строка вычеркивается, и двигаются по столбцу вниз. В клетку, находящуюся на пересечении первого столбца и второй строки, помещается максимально возможное число единиц продукции, разрешенное ограничениями на предложение и спрос: $x_{21} = \min(a_2, b_1 - a_1)$. Если например, $b_1 - a_1 \le a_2$, то $x_{21} = b_1 - a_1$. Спрос первого потребителя удовлетворен. Первый столбец вычеркивают и двигаются по второй строке вправо. Заполнив клетку, стоящую на пересечении второй строки и второго столбца, переходят к заполнению следующей третьей клетки второй строки, либо второго столбца. Процесс продолжают до тех пор, пока не исчерпается предложение и не удовлетворится спрос. Последняя заполненная клетка находится в последнем n-м столбце и последней m-й строке.

Пример 2.11

Определить начальное решение по методу «северо-западного» угла для транспортной задачи из примера 2.2.1.

Решение. Транспортная таблица имеет следующий вид (табл. 2.9):

Таблица 2.9

№ п/п	1	2	3	4	Предложение
1	120 [7]	40 [8]	[1]	[2]	160
2	[4]	10 [5]	130 [9]	[8]	140
3	[9]	[2]	60 [3]	110 [6]	170
Спрос	120	50	190	110	

В первую клетку помещают: $x_{11} = \min(160{,}120) = 120$. Спрос первого потребителя полностью удовлетворен, первый столбец вычеркивают. Остаток сырья в первом пункте составляет: 160-120=40 усл.ед. Двигаемся по первой строке вправо $x_{21} = \min(160-120{,}50) = 40$. Предложение поставщика исчерпано, первая строка вычеркивается. Второму потребителю не хватает 50-40=10 усл. ед. Двигаемся по второму столбцу вниз $x_{22} = \min(140{,}50-40) = 10$. Второй столбец вычеркивается. Двигаемся по второй строке вправо $x_{23} = \min(140-10{,}90) = 130$. Вторая строка вычеркивается. Двигаемся по третьему столбцу вниз $x_{33} = \min(170{,}190-130) = 60$. Спрос третьего потребителя удовлетворен. Двигаемся по третьей строке вправо $x_{34} = \min(170-160{,}110) = 110$. Таблица заполнена. Число ненулевых значений x_{ij} равно 6. Число базисных переменных задачи 3+4-1=6. Остальные 3*4-6=6 переменных являются свободными, их значения равны нулю.

Начальный план перевозок имеет вид:

$$X_1 = \begin{pmatrix} x_{11} = 120 & x_{12} = 40 & x_{13} = 0 & x_{14} = 0 \\ x_{21} = 0 & x_{22} = 10 & x_{23} = 130 & x_{24} = 0 \\ x_{31} = 0 & x_{32} = 0 & x_{33} = 60 & x_{34} = 110 \end{pmatrix}$$

Стоимость перевозок по этому плану составляет

$$S_1 = 120 \cdot 7 + 40 \cdot 8 + 10 \cdot 5 + 130 \cdot 9 + 60 \cdot 3 + 110 \cdot 6 = 3220.$$

<u>*Задача решена.*</u>

Метод «северо-западного» угла – наиболее простой метод нахождения начального решения. План перевозок, полученный по этому методу, обычно бывает достаточно близок оптимальному.

Метод минимального элемента

Шаг 1. Составляют транспортную таблицу.

Шаг 2. Выбирают клетку таблицы, которой соответствует минимальное значение тарифа, и переходят на шаг 3.

Шаг 3. В выбранную клетку аналогично методу «северо-западного» угла помещают максимально возможное число единиц продукции, разрешенное ограничениями на предложение и спрос. После этого, если предложение производителя исчерпано, вычеркивают соответствующую строку; если спрос удовлетворен, вычеркивают соответствующий столбец.

Если все клетки заполнены и вычеркнуты, то план перевозок построен. В противном случае переходят к шагу 2 без учета заполненных и вычеркнутых клеток.

Пример 2.12

Определить начальное решение по методу минимального элемента для транспортной задачи из примера 2.11.

Решение задачи

Решение записано в табл. 2.10.

Таблица 2.10

№ п/п	1	2	3	4	Предложение
1	7	8	160 1	2	160
2	120 4	5	9	20 8	140
3	9	50 2	30 3	90 6	170
Спрос	120	50	190	110	

Минимальный тариф $c_{13} = 1$, $x_{13} = \min(160{,}190) = 160$. Первую строку вычеркивают. Минимальный тариф для оставшихся клеток $c_{32} = 2$, $x_{32} = \min(170{,}50) = 50$. Второй столбец вычеркивают.

Для оставшихся клеток минимальный тариф:

$c_{33} = 3$, $x_{33} = \min(170 - 50{,}190 - 160) = 30$. Третий столбец вычеркивают.

Для оставшихся клеток минимальный тариф:

$c_{21} = 4$, $x_{21} = \min(140{,}120) = 120$. Первый столбец вычеркивают.

Для оставшихся клеток минимальный тариф:

$c_{34} = 6$, $x_{34} = \min(170 - 50 - 30{,}110) = 90$. Для одной оставшейся клетки $x_{24} = \min(140 - 120{,}110 - 90) = 20$.

План перевозок, полученный по методу минимального элемента имеет вид

$$X_1 = \begin{pmatrix} x_{11} = 0 & x_{12} = 0 & x_{13} = 160 & x_{14} = 0 \\ x_{21} = 120 & x_{22} = 0 & x_{23} = 0 & x_{24} = 20 \\ x_{31} = 0 & x_{32} = 50 & x_{33} = 30 & x_{34} = 90 \end{pmatrix}$$

Стоимость перевозок по этому плану составляет

$$S_1 = 160 \cdot 1 + 120 \cdot 4 + 20 \cdot 8 + 50 \cdot 2 + 30 \cdot 3 + 90 \cdot 6 = 1530\,.$$

Задача решена.

Стоимость перевозок, полученных по методу минимального элемента, обычно бывает меньше стоимости перевозок, полученных по методу «северо-западного» угла.

Метод Фогеля

Шаг 1. Составляют транспортную таблицу.

Шаг 2. Для каждой строки и каждого столбца транспортной таблицы определяют разность между наименьшим тарифом и ближайшим к нему значением. Переход к шагу 3.

Шаг 3. В строке или столбце, которым соответствует наибольшая разность, выбирают клетку с наименьшим тарифом. Переход к шагу 4.

Шаг 4. В выбранную клетку, аналогично предыдущим методам, записывают максимально возможное число единиц продукции, которое разрешается ограничениями на предложение и спрос. После этого вычеркивают либо строку, если предложение поставщика исчерпано, либо столбец, если спрос потребителя удовлетворен.

Если все клетки таблицы заполнены или вычеркнуты, то план перевозок построен. В противном случае переходят к шагу 2 без учета вычеркнутых и заполненных клеток.

В методе Фогеля используются штрафы, взимаемые за неудачный выбор маршрута. Рассчитанные на шаге 2 разности между двумя уровнями затрат на перевозку являются штрафами за неверно выбранный маршрут перевозки.

Пример 2.13

Определим начальное решение по методу Фогеля для транспортной задачи из примера 2.11 (табл. 2.2.4).

Решение

Таблица 2.11

№	1	2	3	4	Предложение	Разность по строкам			
1	^{7}	^{8}	50 ^{1}	110 ^{2}	160	1	$\boxed{6}$	-	-
2	120 ^{4}	20 ^{5}	^{9}	^{8}	140	1	1	1	1
3	^{9}	30 ^{2}	140 ^{3}	^{6}	170	1	1	1	$\boxed{7}$
Спрос	120	50	190	110					
Разность по столбцам	3	3	2	$\boxed{4}$					
	3	3	2	-					
	5	3	$\boxed{6}$	-					
	5	3	-	-					

Разности по строкам будем записывать в правой части табл. 2.11, разности по столбцам – внизу табл. 2.11. Максимальную разность будем отмечать квадратиком. Наименьший тариф в первой строке равен 1. Ближайший к нему равен 2. Разность равна 1. Наименьший тариф во второй строке 4. Ближайшее к нему значение 5. В третьей строке 2 и 3, соответственно. Разности по всем строкам равны 1.

В первом столбце наименьший тариф $c_{21}=4$. Ближайшее значение $c_{11}=7$, $c_{11}-c_{21}=7-4=3$. Во втором столбце наименьшее значение $c_{32}=2$. Ближайшее значение $c_{22}=5$, $c_{22}-c_{32}=5-2=3$.

Третий столбец: $c_{13}=1$, $c_{33}=3$, $c_{33}-c_{13}=3-1=2$.

Четвертый столбец: $c_{14}=2$, $c_{34}=6$, $c_{34}-c_{14}=6-2=4$.

Максимальная из всех разностей 4 находится в четвертом столбце. В этом столбце клетка с наименьшим тарифом $c_{14}=2$ находится в первой строке. В эту клетку помещаем максимально возможное значение: $x_{14}=\min(110{,}160)=110$. Четвертый потребитель полностью удовлетворил свой спрос, и четвертый столбец вычеркиваем.

Повторяем предыдущие действия без учета вычеркнутых и заполненных клеток.

Первая строка: минимальный тариф $c_{13}=1$. Ближайшее значение $c_{11}=7$, $c_{11}-c_{13}=7-1=6$.

Вторая строка: минимальный тариф $c_{21}=4$. Ближайшее значение $c_{22}=5$, $c_{22}-c_{21}=5-4=1$.

Третья строка: минимальный тариф $c_{32}=2$. Ближайшее значение $c_{33}=3$, $c_{33}-c_{32}=3-2=1$.

Первый столбец: минимальный тариф $c_{21}=4$. Ближайшее значение $c_{11}=7$, $c_{11}-c_{21}=7-4=3$.

Второй столбец: минимальный тариф $c_{32}=2$. Ближайшее значение $c_{22}=5$, $c_{22}-c_{32}=5-2=3$.

Третий столбец: минимальный тариф $c_{13}=1$. Ближайшее значение $c_{33}=3$, $c_{33}-c_{13}=3-1=2$.

Максимальная разность равна 6 и стоит в первой клетке. Минимальный тариф в первой строке $c_{13}=1$. В эту клетку помещаем $x_{13}=\min(160-110{,}190)=50$. Вычеркиваем первую строку. Повторяем все действия без учета первой строки и четвертого столбца.

Вторая строка: минимальный тариф $c_{21}=4$. Ближайшее значение $c_{22}=5$, $c_{22}-c_{21}=5-4=1$.

Третья строка: минимальный тариф $c_{32}=2$. Ближайшее значение $c_{33}=3$, $c_{33}-c_{32}=3-2=1$.

Первый столбец: минимальный тариф $c_{21}=4$. Ближайшее значение $c_{31}=9$, $c_{31}-c_{21}=9-4=5$.

Второй столбец: минимальный тариф $c_{32}=2$. Ближайшее значение $c_{22}=5$, $c_{22}-c_{32}=5-2=3$.

Третий столбец: минимальный тариф $c_{33}=3$. Ближайшее значение $c_{23}=9$, $c_{23}-c_{33}=9-3=6$.

Максимальная разность равна 6 и стоит в третьем столбце. Минимальный из оставшихся тарифов в этом столбце $c_{33} = 3$, $x_{33} = \min(170,190-50) = 140$. Спрос третьего потребителя удовлетворен, третий столбец вычеркиваем.

Вновь составляем разности для невычеркнутых строк и столбцов.

Вторая строка: минимальный тариф $c_{21} = 4$. Ближайшее значение $c_{22} = 5$, $c_{22} - c_{21} = 5 - 4 = 1$.

Третья строка: минимальный тариф $c_{32} = 2$. Ближайшее значение $c_{31} = 9$, $c_{31} - c_{32} = 9 - 2 = 7$.

Первый столбец: минимальный тариф $c_{21} = 4$. Ближайшее значение $c_{31} = 9$, $c_{31} - c_{21} = 9 - 4 = 5$.

Второй столбец: минимальный тариф $c_{32} = 2$. Ближайшее значение $c_{22} = 5$, $c_{22} - c_{32} = 5 - 2 = 3$.

Максимальная разность стоит в третьей строке. Минимальный из оставшихся тарифов в этой строке $c_{32} = 2$, $x_{32} = \min(170-140,50) = 30$.

Предложение поставщика исчерпано, и третью строку вычеркиваем. Осталась одна строка транспортной таблицы. Это вторая строка. В этой строке сначала заполняем клетку с наименьшим тарифом $c_{21} = 4$, $x_{21} = \min(140,120) = 120$. Оставшееся предложение второго поставщика записываем в единственную свободную клетку $x_{22} = \min(140-120,50-30) = 20$.

Полученный по методу Фогеля план перевозок имеет вид

$$X_1 = \begin{pmatrix} x_{11} = 0 & x_{12} = 0 & x_{13} = 50 & x_{14} = 110 \\ x_{21} = 120 & x_{22} = 20 & x_{23} = 0 & x_{24} = 0 \\ x_{31} = 0 & x_{32} = 30 & x_{33} = 140 & x_{34} = 0 \end{pmatrix}.$$

Затраты на перевозку по этому плану составляют

$$S_3 = 50 \cdot 1 + 110 \cdot 2 + 120 \cdot 4 + 20 \cdot 5 + 30 \cdot 2 + 140 \cdot 3 = 1330.$$

Задача решена.

Таким образом, для одной и той же транспортной задачи получены различные начальные планы перевозок, построенные с использованием разных методов. При этом затраты на перевозки составляют соответственно: $S_1 = 3220$, $S_2 = 1530$, $S_3 = 1330$.

Метод Фогеля наиболее трудоемкий, однако начальный план перевозок, построенный с его использованием, обычно бывает близок к оптимальному плану, а в некоторых случаях является оптимальным планом.

Оптимальный план транспортной задачи. Метод потенциалов.

Заметим, что для всех полученных решений число заполненных (отличных от нуля) клеток транспортной таблицы в точности равно числу базисных переменных задачи, т.е. равно 6.

Если при решении транспортной задачи число заполненных клеток транспортной таблицы равно $m+n-1$, где m - число производителей, n - число потребителей, то план перевозок невырожденный.

Если число заполненных клеток транспортной таблицы меньше $m+n-1$, то план перевозок вырожденный. Вырожденный план перевозок получится, если на каком-то шаге одновременно удовлетворяется спрос потребителя и исчерпывается предложение соответствующего поставщика, т.е. одновременно вычеркивается строка и столбец.

Для нахождения оптимального плана перевозок необходимо уметь оценивать полученный план на оптимальность. Для оценки плана на оптимальность вводится понятие косвенных затрат. Косвенные затраты – это затраты, получаемые для маршрутов, по которым не осуществляются перевозки при данном плане. Рассчитанные косвенные затраты сравниваются с реальными затратами, которые имели бы место, если бы перевозки по данным маршрутам осуществлялись. Если для всех невыбранных маршрутов косвенные затраты не больше реальных, то данный план перевозок является оптимальным. Если хотя бы для одного маршрута косвенные затраты больше реальных, то план перевозок может быть улучшен путем введения в него данного маршрута. Ввод нового маршрута в план перевозок соответствует вводу в список базисных переменных переменной транспортной задачи, соответствующей этому маршруту. Эти рассуждения лежат в основе ряда методов, применяемых для нахождения оптимального плана на перевозок. Рассмотрим один из них – метод потенциалов.

Получение оптимального плана транспортной задачи с использованием метода потенциалов

Шаг 1. Получение начального плана перевозок по методу «северо-западного» угла, минимального элемента, Фогеля или любым другим методом.

Шаг 2. Проверка плана на невырожденность. Если полученный план вырожденный, формально заполняют нулями некоторые из свободных клеток так, чтобы общее число занятых клеток было равно $m+n-1$. Нули надо расставлять так, чтобы не образовался замкнутый цикл из занятых клеток.

Шаг 3. Проверка плана на оптимальность.

Шаг 3.1. Определение потенциалов производителей и потребителей. Составляют систему уравнений для заполненных клеток транспортной таблицы: $U_i + V_j = C_{ij}$, где i, j - номера строк и столбцов на пересечении которых стоят заполненные клетки, U_i - потенциал i-го поставщика, V_j - потенциал j-го потребителя, C_{ij} - тариф на перевозку из пункта i в пункт j. Число уравнений в системе равно $m+n-1$, а число неизвестных U_i и V_j равно $m+n$. Для решения данной системы одно из неизвестных выбирают произвольно. Обычно полагают $U_i = 0$. Решая систему уравнений, находят значения потенциалов U_i и V_j.

Шаг 3.2. Определение суммы потенциалов (косвенных тарифов) для свободных клеток: $C1_{qp} = U_q + V_p$, где q и p - номера строк и столбцов, на пересечении которых стоит свободная клетка, U_q - потенциал q-го поставщика, V_p - потенциал p - го потребителя, $C1_{qp}$ - косвенные тарифы.

Шаг 3.3. Проверка на оптимальность.

Для каждой свободной клетки транспортной таблицы составляется разность между $C1_{qp}$ и C_{qp} (косвенным и реальным тарифами) $\Delta_{qp} = C1_{qp} - C_{qp}$. Если все $\Delta_{qp} \leq 0$, то полученный план оптимален. Если хотя бы для одной свободной клетки $\Delta_{qp} > 0$, то план может быть улучшен. Переход к шагу 4.

Шаг 4. Улучшение плана.

Шаг 4.1. Выбор переменной, вводимой в список базисных переменных.

Выбирают клетку, которой соответствует максимальное положительное значение разности, полученной на шаге 3.3. Если имеется несколько одинаковых значений, то из них выбирают любое.

Переменная транспортной задачи, соответствующая этой клетке, вводится в список базисных переменных, т.е. данная клетка транспортной таблицы заполняется.

Шаг 4.2. Выбор переменной, выводимой из списка базисных переменных.

Заполнение клетки, выбранной на шаге 4.1, происходит следующим образом. Строят цикл, начинающийся и заканчивающийся в выбранной свободной клетке, содержащий в качестве вершин заполненные клетки таблицы и состоящий из горизонтальных и вертикальных отрезков. При этом в каждой клетке таблицы, являющейся вершиной цикла, соединяют обязательно горизонтальный и вертикальный отрезки. В свободной клетке условно ставят знак «+», а в остальных вершинах цикла, чередуясь, ставят «-» и «+». Затем происходит перераспределение продукции по циклу. Для этого выбирают клетку со знаком «-», которой соответствует наименьшее число единиц продукции. Это значение прибавляют к значениям, стоящим в клетках со знаком «+», и отнимают от значений, стоящих в клетках со знаком «-». При таком перераспределении общий баланс не изменяется. Свободная клетка заполняется. А клетка со знаком «-», которой соответствует наименьшее количество продукции, становится свободной; соответствующую ей переменную исключают из списка базисных.

Для нового плана повторяют все действия, т.е. переходят к шагу 2.

Пример 2.14: Найти оптимальный план перевозок для транспортной задачи из примера 2.11.

Решение

В качестве начального плана выберем план, найденный по методу минимального элемента (табл. 2.12).

Таблица 2.12

№ п/п	1	2	3	4	Предложение
1	7	8	160 [1] −	+ [2]	160
2	120 [4]	5	9	20 [8]	140
3	9	50 [2]	30 [3] +	90 − [6]	170
Спрос	120	50	190	110	

Число заполненных клеток равно 4+3-1=6, т.е. данный план невырожденный. Определим потенциалы производителей и потребителей, составив уравнения $U_i + V_j = C_{ij}$ для заполненных клеток.

$U_1 + V_3 = 1$; $U_2 + V_1 = 4$; $U_2 + V_4 = 8$; $U_3 + V_2 = 2$; $U_3 + V_3 = 3$; $U_3 + V_4 = 6$.

$U_1 = 0$; $U_4 = 4$; $U_3 = 2$;

$V_1 = 0$; $V_2 = 0$; $V_3 = 1$; $V_4 = 4$.

Составим разности для свободных клеток

$$\Delta_{11} = (U_1 + V_1) - C_{11} = (0 + 0) - 7 = -7;$$
$$\Delta_{12} = (U_1 + V_2) - C_{12} = (0 + 0) - 8 = -8;$$
$$\Delta_{14} = (U_1 + V_4) - C_{14} = (0 + 4) - 2 = 2;$$
$$\Delta_{22} = (U_2 + V_2) - C_{22} = (4 + 0) - 5 = -1;$$
$$\Delta_{23} = (U_2 + V_3) - C_{23} = (4 + 1) - 9 = -4;$$
$$\Delta_{31} = (U_3 + V_1) - C_{31} = (2 + 0) - 9 = -7.$$

Получена положительная разность $\Delta_{14} = 2$. Запомним клетку первой строки и четвертого столбца. Строим цикл, начинающийся и заканчивающийся в этой клетке. Вершинами цикла являются клетки: (3;4), (3;3), (1;3). В клетке (1;4) ставим «+», в клетке (3;4) «-», в клетке (3;3) «+», в клетке (1;3) «-». Перераспределяем продукцию по циклу. Минимальное значение для клеток со знаком «-» находится в клетке (3;4) $x_{34} = 90$. Отнимем 90 от значений, стоящих в клетках со знаком «-», и прибавляем к значениям, стоящим в клетках со знаком «+». Получаем новый план перевозок, представленный в следующей транспортной таблице (табл. 2.13):

Таблица 2.13

№ п.п	1	2	3	4	Предложение
1	^{7}	^{8}	70 ^{1} -	+ 90 ^{2}	160
2	120 ^{4}	^{5}	+ ^{9}	- 20 ^{8}	140
3	^{9}	50 ^{2}	120 ^{3} +	^{6}	170
Спрос	120	50	190	110	

Новый план невырожденный. Проверим его на оптимальность.

$U_1 + V_3 = 1$; $U_1 + V_4 = 2$; $U_2 + V_1 = 4$; $U_2 + V_4 = 8$; $U_3 + V_2 = 2$; $U_3 + V_3 = 3$.

$U_1 = 0$; $U_2 = 6$; $U_3 = 2$;

$V_1 = -2$; $V_2 = 0$; $V_3 = 1$; $V_4 = 2$.

Составим разности для свободных клеток

$$\Delta_{11} = (U_1 + V_1) - C_{11} = -2 - 7 = -9;$$
$$\Delta_{12} = (U_1 + V_2) - C_{12} = 0 - 8 = -8;$$
$$\Delta_{22} = (U_2 + V_2) - C_{22} = 6 - 5 = 1;$$
$$\Delta_{23} = (U_2 + V_3) - C_{23} = 7 - 9 = -2;$$
$$\Delta_{31} = (U_3 + V_1) - C_{31} = 0 - 9 = -9;$$
$$\Delta_{34} = (U_3 + V_4) - C_{34} = 4 - 6 = -2.$$

Положительная разность $\Delta_{22}=1$. Заполним клетку (2;2). Цикл будет содержать клетка: (2;2), (3;2), (3;3), (1;3), (1;4), (2;4), (2;2).

Минимальное значение $x_{24}=20$ для клеток со знаком «-». Перераспределив продукцию по циклу, получим новый план перевозок, представленный в табл. 2.14.

Таблица 2.14

№ п/п	1	2	3	4	Предложение
1	7	8	50 1	110 2	160
2	120 4	20 5	9	8	140
3	9	30 2	140 3	90 6	170
Спрос	120	50	190	110	

Проверим полученный невырожденный план на оптимальность.

$U_1+V_3=1$; $U_1+V_4=2$; $U_2+V_1=4$; $U_2+V_2=2$; $U_3+V_2=2$; $U_3+V_3=3$.

$U_1=0$; $U_2=5$; $U_3=2$;

$V_1=-1$; $V_2=0$; $V_3=1$; $V_4=2$.

Составим разности для свободных клеток

$$\Delta_{11}=(U_1+V_1)-C_{11}=-1-7=-8;$$
$$\Delta_{12}=(U_1+V_2)-C_{12}=0-8=-8;$$
$$\Delta_{22}=(U_2+V_2)-C_{22}=6-9=-3;$$
$$\Delta_{23}=(U_2+V_3)-C_{23}=7-8=-1;$$
$$\Delta_{31}=(U_3+V_1)-C_{31}=1-9=-8;$$
$$\Delta_{34}=(U_3+V_4)-C_{34}=4-6=-2.$$

Все разности отрицательные, следовательно, получен оптимальный план.

$$X^*=\begin{pmatrix} x_{11}=0 & x_{12}=0 & x_{13}=50 & x_{14}=110 \\ x_{21}=120 & x_{22}=20 & x_{23}=0 & x_{24}=0 \\ x_{31}=0 & x_{32}=30 & x_{33}=140 & x_{34}=0 \end{pmatrix}.$$

Затраты на перевозку по этому плану составляют

$$S_3=50\cdot1+110\cdot2+120\cdot4+20\cdot5+30\cdot2+140\cdot3=1330.$$

Заметим, что этот план совпадает с начальным планом, найденный по методу Фогеля.

Задача решена.

Пример 2.15

Фирма обслуживающая туристов прибывающих на отдых, должна разместить их в 4 отелях: «Морской», «Солнечный», «Слава» и «Уютный», в которых забронировано соответственно 5, 15, 15 и 10 мест. Пятнадцать туристов прибывают по железной дороге, двадцать пять прилетают очередным рейсом в аэропорт, а пять человек прибудут на теплоходе на морской вокзал. Транспортные расходы при перевозке из пунктов прибытия в отели приведены в табл. 2.15.

Таблица 2.15

Исходный пункт, i		Пункт назначения (отели), j			
		Морской	Солнечный	Слава	Уютный
		1	2	3	4
Железнодорожный вокзал	1	10	0	20	11
Аэропорт	2	12	7	9	20
Морской вокзал	3	0	14	16	18

В условиях жесткой конкуренции фирма должна минимизировать свои расходы, значительную часть которых составляют именно транспортные расходы. Требуется определить такой план перевозки туристов из пунктов прибытия в отели, при котором суммарные транспортные расходы будут минимальны и все туристы будут размещены в отелях.

Решение. Рассмотрим математическую модель задачи.

1. *Переменные задачи*. Обозначим количество туристов, которые будут перевозиться из пункта i в отель j как x_{ij} ($i = 1,2,3;\ j = 1,2,3,4$). Это переменные задачи, значения которых должны быть определены в процессе решения. Например, x_{23} - это число туристов, которое должно быть перевезено из аэропорта (пункт 2) в отель «Слава» (пункт 3). В задаче содержится 12 переменных.

2. *Ограничения на переменные задачи*. Очевидно, что все переменные задачи неотрицательные и целые числа, т. е.

$$x_{ij} \geq 0, \quad x_{ij} \text{ - целые числа, где } i = 1,2,3;\ j = 1,2,3,4.$$

Кроме этого, должны удовлетворяться следующие условия. Число туристов, вывозимых с железнодорожного вокзала (пункт 1) равно 15, поэтому

$$x_{11} + x_{12} + x_{13} + x_{14} = \sum_{j=1}^{4} x_{1j} = 15.$$

Аналогично, для аэропорта (пункт 2):

$$x_{21} + x_{22} + x_{23} + x_{24} = \sum_{j=1}^{4} x_{2j} = 25,$$

и для морского вокзала (пункт 3):

$$x_{31} + x_{32} + x_{33} + x_{34} = \sum_{j=1}^{4} x_{3j} = 5.$$

По условию задачи в отеле «Морской» (пункт 1) забронировано 5 мест, поэтому

$$x_{11} + x_{21} + x_{31} = \sum_{i=1}^{3} x_{i1} = 5.$$

Аналогично, для отеля «Солнечный» (пункт 2):

$$x_{12} + x_{22} + x_{32} = \sum_{i=1}^{3} x_{i2} = 15.$$

Для отеля «Слава» (пункт 3):

$$x_{13} + x_{23} + x_{33} = \sum_{i=1}^{3} x_{i3} = 15.$$

Для отеля «Уютный» (пункт 4):

$$x_{14} + x_{24} + x_{34} = \sum_{i=1}^{3} x_{i4} = 10.$$

Обычно транспортная задача записывается в виде таблицы, где в ячейках помещаются переменные задачи x_{ij}, а в правом верхнем углу ячейки стоят стоимости перевозки из пункта i в пункт j. В крайнем правом столбце и нижней строке таблицы записываются числа определяющие ограничения задачи.

3. *Целевая функция*. транспортные расходы на перевозку туристов в отели вычисляются по формуле

$$Z = \sum_{i=1}^{3}\sum_{j=1}^{4} c_{ij}x_{ij} = 10 \cdot x_{11} + 0 \cdot x_{12} + \ldots + 18 \cdot x_{34}.$$

Окончательно транспортная задача имеет вид: нужно найти такие целые значения переменных x_{ij}, при которых целевая функция будет иметь минимальное значение и будут выполнены ограничения:

$$\begin{cases} \sum_{j=1}^{4} x_{1j} = 15; \quad \sum_{j=1}^{4} x_{2j} = 25; \quad \sum_{j=1}^{4} x_{3j} = 5; \\ \sum_{i=1}^{3} x_{i1} = 5; \quad \sum_{i=1}^{3} x_{i2} = 15; \quad \sum_{i=1}^{3} x_{i3} = 15; \\ \sum_{i=1}^{3} x_{i4} = 10; \quad x_{ij} \ge 0. \end{cases} \tag{2.36}$$

Реализация транспортной задачи с помощью ППП Excel.

При решении транспортной задачи в Excel задача должна быть предварительно сбалансирована.

1. Ввод данных. Вводим исходные данные в ячейки Excel (рис. 2.20).

В ячейках B3:E5 введены стоимости перевозок. В ячейках F3:F5 находится число прибывающих туристов, а в ячейках B6:E6 находится число мест в отелях. Ячейки B8:E10 – рабочие (изменяемые) ячейки, в которых будут вычисляться значения переменных задачи x_{ij}.

В ячейках F8:F10 нужно записать формулы для вычисления левых частей первых трех ограничений (2.36):

в F8 должна быть сумма ячеек B8:E8;

в F9 должна быть сумма ячеек B9:E9;

в F10 должна быть сумма ячеек B10:E10.

Формулы для вычисления последних четырех левых частей ограничений (2.36) введем в ячейки B11:E11:

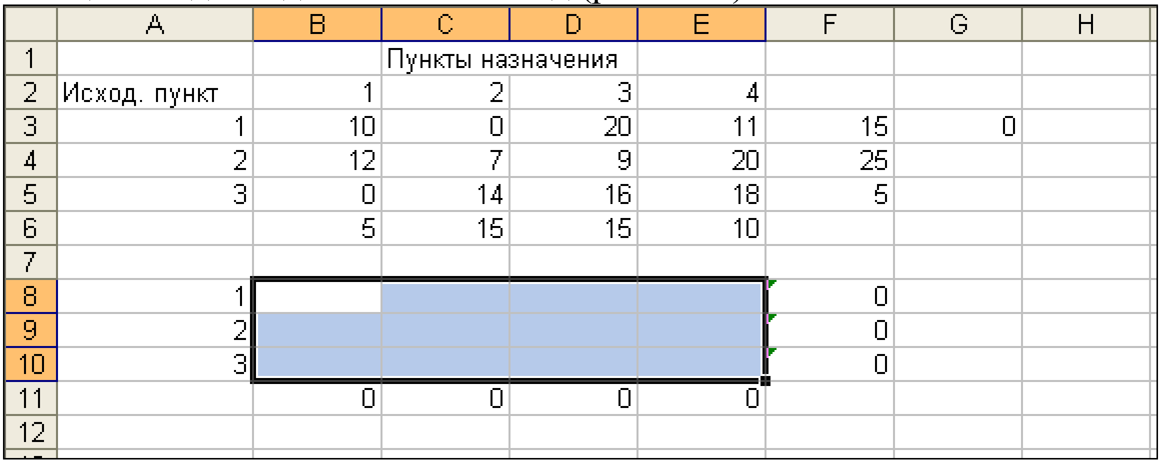
в B11 должна быть сумма ячеек B8:B10;

в C11 должна быть сумма ячеек C8:C10;

в D11 должна быть сумма ячеек D8:D10;

в E11 должна быть сумма ячеек E8:E10;
Целевую функцию поместим в ячейку G3:
G3:СУММПРОИЗВ(B3:E5;B8:E10).

Таблица исходных данных имеет вид (рис. 2.20).



	A	B	C	D	E	F	G	H
1			Пункты назначения					
2	Исход. пункт	1	2	3	4			
3	1	10	0	20	11	15	0	
4	2	12	7	9	20	25		
5	3	0	14	16	18	5		
6		5	15	15	10			
7								
8	1					0		
9	2					0		
10	3					0		
11		0	0	0	0			
12								

Рис. 2.20

2. Заполнение окна процедуры «Поиск решения».
Целевая функция: G3.
Значение целевой функции: min.
Изменяемые ячейки: B8:E10.
Ограничения задачи:
F8:F10=F3:F5;
B11:E11=B6:E6;
B8:E10≥ 0 и B8:E10 – целые числа.

В окне «Параметры» установить «Линейная модель», что соответствует решению задачи симплекс-методом. Результаты заполнения окна показаны на рис. 2.21.

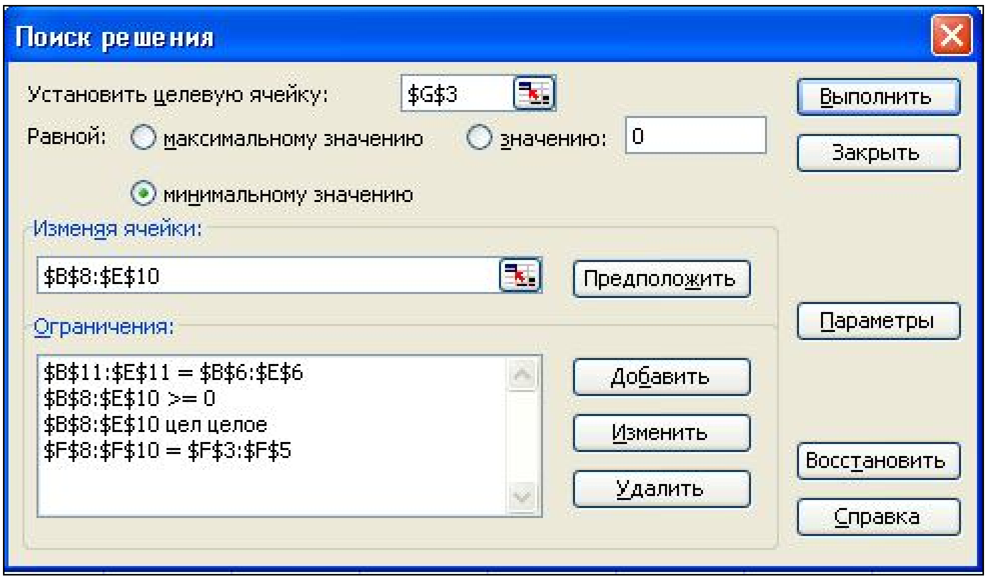


Рис. 2.21

3. Выполнив процедуру «Поиск решения», получим следующие результаты (рис. 2.22):

	A	B	C	D	E	F	G	H
1			Пункты назначения					
2	Исход. пункт	1	2	3	4			
3	1	10	0	20	11	15	315	
4	2	12	7	9	20	25		
5	3	0	14	16	18	5		
6		5	15	15	10			
7								
8	1	0	5	0	10	15		
9	2	0	10	15	0	25		
10	3	5	0	0	0	5		
11		5	15	15	10			
12								

Рис. 2.22

Таким образом, с железнодорожного вокзала (исходный пункт 1) следует 10 туристов отвезти в отель «Уютный» (пункт 4) и 5 туристов в отель «Солнечный» (пункт назначения 2) и 15 туристов в отель «Слава» (пункт назначения 3); туристов прибывающих на морской вокзал (исходный пункт 3) нужно отправить в отель «Морской» (пункт назначения 1). Все эти результаты показаны в конечной таблице (рис. 2.22). При этом суммарная стоимость транспортных расходов составит **315** руб.

Реализация задачи с помощью Mathcad 13.

Решение задачи в Mathcad 13 показано на рис. 2.23.

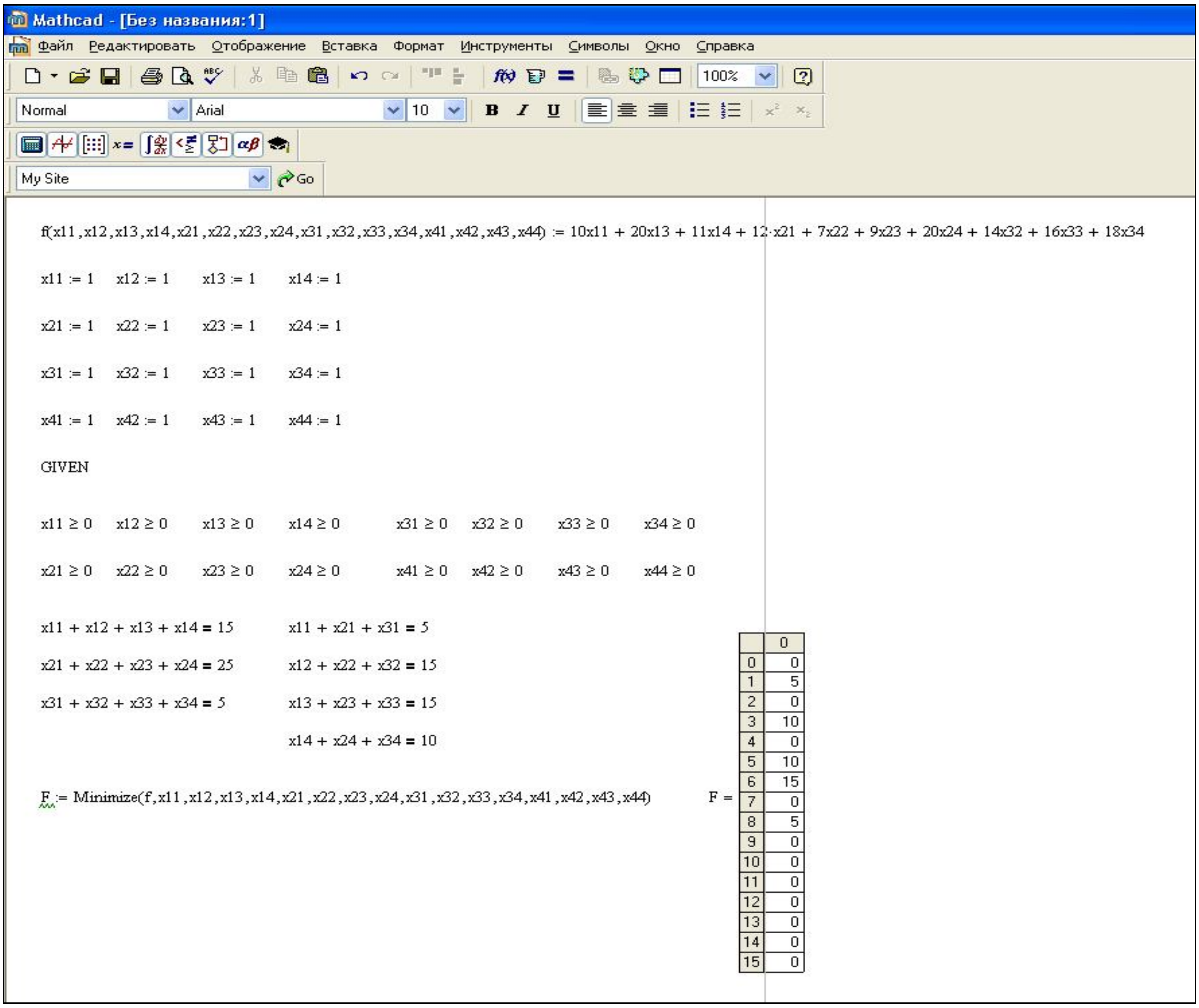


Рис. 2.23

2.3.2 Задача о назначениях

В общем виде задача о назначениях формулируется следующим образом.

Имеется n работ и n кандидатов для их выполнения. Затраты i-го кандидата на выполнение j-й работы равны c_{ij}. Каждый кандидат может быть назначен только на одну работу, и каждая работа может быть выполнена только одним кандидатом. Требуется найти назначение кандидатов на работы, при котором суммарные затраты на выполнение работ минимальны.

<u>Определим формальную модель данной задачи</u>. Пусть x_{ij} - переменная, значение которой равно 1, если i-й кандидат выполняет j-ю работу, и 0 – в противном случае. Тогда условие о том, что каждый кандидат выполняет только одну работу, запишется в виде

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i = \overline{1,n}.$$

Условие о том, что каждая работа может выполняться одним кандидатом, запишется в виде

$$\sum_{i=1}^{n} x_{ij} = 1, \quad j = \overline{1,n}.$$

Целевая функция задачи имеет вид

$$C = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} x_{ij}.$$

В функцию входят только те значения c_{ij}, для которых x_{ij} отличны от 0, т.е. входят затраты, соответствующие назначенным работам.

Математическая модель выглядит следующим образом:

$$\boxed{C = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} x_{ij} \to \min} \tag{2.37}$$

$$\begin{cases} \sum_{j=1}^{n} x_{ij} = 1, \quad i = \overline{1,n}; \\ \sum_{i=1}^{m} x_{ij} = 1, \quad j = \overline{1,n}; \\ \sum_{i=1}^{m} a_i \geq \sum_{j=1}^{n} b_j; \\ x_{ij} \in \{0;1\}. \end{cases} \tag{2.38}$$

Решить задачу о назначениях – значит найти x_{ij}, удовлетворяющие (2.2.8) и доставляющие минимум функции (2.37). Задача (2.37) – (2.38) является задачей линейного программирования (целевая функция линейна, ограничения линейны) и может быть решена симплекс-методом. Также задача (2.37) – (2.38) – это транспортная задача, в которой правые части ограничений равны 1, а переменные могут принимать только два значения. Однако относительная простая форма задачи позволила разработать для её решения достаточно простые методы, один из которых – венгерский.

Венгерский метод решения задачи о назначениях

Для решения задачи о назначениях составляют таблицу 2.16:

Таблица 2.16

№	1	2	…	j	…	n
1	c_{11}	c_{12}	…	c_{1j}	…	c_{1n}
2	c_{21}	c_{22}	…	c_{2j}	…	c_{2n}
…	…	…	…	…	…	…
i	c_{i1}	c_{i2}	…	c_{ij}	…	c_{in}
…	…	…	…	…	…	…
n	c_{m1}	c_{m2}	…	c_{mj}	…	c_{nn}

В левой колонке записаны номера кандидатов, в верхней строке – номера работ. В i-й строке j-м столбце стоят затраты на выполнение i-м кандидатом j-й работы.

В венгерском методе используется *следующий принцип*: оптимальность решения задачи о назначениях не нарушается при уменьшении (увеличении) элементов строки (столбца) на одну и ту же величину. Решение считается оптимальным, если все измененные таким образом затраты $c'_{ij} \geq 0$ и можно отыскать такой набор x_{ij}, что

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c'_{ij} x_{ij} = 0 .$$

Алгоритм метода содержит следующие шаги.

Шаг 1. Получение нулей в каждой строке. Для этого в каждой строке определяют наименьший элемент, и его значение отнимают от всех элементов этой строки. Переход к шагу 2.

Шаг 2. Получение нулей в каждом столбце. В преобразованной таблице в каждом столбце определяют минимальный элемент, и его значение вычитают из всех элементов этого столбца. Переход к шагу 3.

Шаг 3. Поиск оптимального решения. Просматривают строку, содержащую наименьшее число нулей. Отмечают один из нулей этой строки и зачёркивают все остальные нули этой строки и того столбца, в котором находится отмеченный нуль. Аналогичные операции последовательно проводят для всех строк. Если назначение, которое получено при всех отмеченных нулях, является полным (т.е. число отмеченных нулей равно n), то решение является оптимальным, в противном случае следует переходить к шагу 4.

Шаг 4. Поиск минимального набора строк и столбцов, содержащих все нули. Для этого необходимо отметить:

1) все строки, в которых не имеется ни одного отмеченного нуля;
2) все столбцы, содержащие перечёркнутый нуль хотя бы в одной из отмеченных строк;
3) все строки, содержащие отмеченные нули хотя бы в одном из отмеченных столбцов.

Действия п.п. 2 и 3 повторяются поочередно до тех пор, пока есть что отмечать. После этого необходимо зачеркнуть каждую непомеченную строку и каждый помеченный столбец.

Цель этого шага – провести минимальное число горизонтальных и вертикальных прямых, пересекающих по крайней мере один раз все нули.

Шаг 5. Перестановка некоторых нулей.

Взять наименьшее число из тех клеток, через которые проведены прямые. Вычесть его из каждого числа, стоящего в вычеркнутых строках. Эта операция не изменяет оптимального решения, после чего весь цикл расчета повторить, начиная с шага 3.

Пример 2.16

Институт получил гранты на выполнение четырех исследовательских проектов. Выходные результаты первого проекта являются входными данными для второго проекта, выходные результаты второго проекта – это входные данные для третьего проекта, результаты третьего проекта используются для работы над четвертым проектом. В качестве научных руководителей проектов рассматриваются кандидатуры четырех учёных, обладающих различным опытом и способностями. Каждый учёный оценил время, необходимое ему для реализации проекта.

Матрица затрат времени приведена ниже

$$T = \begin{pmatrix} 3 & 7 & 5 & 8 \\ 2 & 4 & 4 & 5 \\ 4 & 7 & 2 & 8 \\ 9 & 7 & 3 & 8 \end{pmatrix}.$$

В i-й строке j-м столбце матрицы T стоит время на выполнение i-м учёным j-го проекта.

Продолжительность времени задана в месяцах. *Требуется выбрать* научного руководителя для выполнения каждого проекта так, чтобы суммарное время выполнения всех проектов (при заданном качестве) было минимальным.

Решение

Данная задача, очевидно, является задачей о назначениях. В качестве работ рассматриваются исследовательские проекты, в качестве кандидатов – учёные, претендующие на роль научных руководителей.

Введем переменные x_{ij}.

$$x_{ij} = \begin{cases} 1, & \text{если } i-\text{й ученый} - \text{научный руководитель } j-\text{проекта}, \\ 0, & \text{в противном случае.} \end{cases}$$

Целевая функция задачи имеет вид

$$C = 3x_{11} + 7x_{12} + 5x_{13} + 8x_{14} + 2x_{21} + 4x_{22} + 4x_{23} + 5x_{24} + 4x_{31} + 7x_{32} + 2x_{33} + 8x_{34} +$$
$$+ 9x_{41} + 7x_{42} + 3x_{43} + 8x_{44} \to \min.$$

Решим задачу венгерским методом, используя приведенную ниже таблицу. В i-й строке j-м столбце этой таблицы стоит время t_{ij} на выполнение j-го проекта i-м учёным. Выберем в каждой строке минимальный элемент и запишем его в правом столбце таблице 2.17.

Таблица 2.17

№	1	2	3	4	
1	3	7	5	8	3
2	2	4	4	5	2
3	4	7	2	8	2
4	9	7	3	8	3

Вычтем минимальные элементы из соответствующих строк, перейдем к новой таблице, в которой найдем минимальные значения в каждом столбце и запишем их в нижней строке таблицы 2.18.

Таблица 2.18

№	1	2	3	4
1	0	4	2	5
2	0	2	2	3

3	2	5	0	6
4	6	4	0	5
	0	2	0	3

Отнимем минимальные элементы из соответствующих столбцов. Перейдем к табл. 2.19.

Таблица 2.19

№	1	2	3	4
1	0 ●	2	2	2
2	~~0~~	0 ●	2	~~0~~
3	2	3	0 ●	3
4	6	2	~~0~~	2

В таблице 2.19 сделаем назначения. Строками, содержащими наименьшее число нулей (один нуль), являются первая, третья и четвертая строки. Отметим кружочком нуль первой строки. Вычеркнем нуль из первого столбца. Это вычеркивание означает, что, так как первый учёный назначен научным руководителем первого проекта, второй учёный уже не может быть назначен. Отмечаем нуль в третьей строке и вычеркиваем нуль, стоящий в четвертой строке в третьем столбце, что соответствует тому, что четвертый учёный уже не может быть назначен научным руководителем третьего проекта.

Отмечаем любой из нулей второй строки (действуя по порядку, отмечаем нуль, стоящий во втором столбце) и вычеркиваем нуль, стоящий в четвертом столбце. Это вычёркивание означает, что, так как второй учёный назначен научным руководителем второго проекта, то он не может быть выбран для выполнения четвертого проекта.

Число отмеченных нулей равно 3, т.е. назначение не является полным. Перейдем к шагу 4 алгоритма.

Найдем минимальный набор строк и столбцов, содержащий все нули (табл. 2.20).

Таблица 2.20

№	1	2	3	4	
1	0 ●	2	2	2	
2	~~0~~	0 ●	2	~~0~~	
3	2	3	0 ●	3	●
4	6	2	~~0~~	2	●
			●		

Отметим кружочком четвертую строку, не содержащую ни одного отмеченного нуля. Отметим третий столбец, содержащий отмеченный нуль в третьем столбце. Кроме третьего столбца, больше нет столбцов, содержащих перечёркнутые нули в отмеченных строках. Вычеркнем отмеченный столбец и неотмеченные строки. В оставшихся клетках минимальный элемент равен 2. Вычтем его из каждого числа невычеркнутых (1,2,4) столбцов. Получим табл. 2.21.

Таблица 2.21

№	1	2	3	4
1	-2	0	2	0
2	-2	-2	2	-2
3	0	1	0	1
4	4	0	0	0

Теперь прибавим 2 к каждому числу вычеркнутых строк в преобразованной таблице. Получим табл. 2.22.

Таблица 2.22

№	1	2	3	4
1	0 ●	2	4	2
2	$\cancel{0}$	0 ●	4	$\cancel{0}$
3	$\cancel{0}$	1	0 ●	1
4	4	$\cancel{0}$	$\cancel{0}$	0 ●

Вновь сделаем назначение, отметив по порядку нули в табл. 2.22. Это назначение является полным, так как число отмеченных нулей равно 4. Получено следующее назначение:

первый учёный назначается научным руководителем первого проекта: $x_{11} = 1$;

второй учёный – научный руководитель второго проекта: $x_{22} = 1$;

третий учёный – научный руководитель третьего проекта: $x_{33} = 1$;

четвертый учёный – научный руководитель четвертого проекта: $x_{44} = 1$.

Время выполнения четырех проектов: $C = 3 + 4 + 2 + 8 = 17$.

Данное назначение не единственное. Если во второй строке сначала отметить не второй, а четвёртый нуль, получим следующее значение (табл. 2.23).

Таблица 2.23

№	1	2	3	4
1	0 ●	2	4	2
2	$\cancel{0}$	$\cancel{0}$	4	0 ●
3	$\cancel{0}$	1	0 ●	1
4	4	0 ●	$\cancel{0}$	$\cancel{0}$

первый учёный назначается научным руководителем первого проекта: $x_{11} = 1$;

второй учёный – научный руководитель четвертого проекта: $x_{24} = 1$;

третий учёный – научный руководитель третьего проекта: $x_{33} = 1$;

четвертый учёный – научный руководитель второго проекта: $x_{42} = 1$.

Время выполнения четырех проектов: $C = 3 + 4 + 2 + 8 = 17$.

Таким образом, получены два оптимальных варианта назначения, которым соответствуют минимальные времена выполнения проектов.

Задача решена.

Заметим, что результат, полученный по венгерскому методу, останется прежним, если в алгоритме заменить строки на столбцы, и наоборот.

Пример 2.17

В конкурсе на занятие пяти вакансий $(V_1, V_2, V_3, V_4, V_5)$ участвуют семь претендентов ($P_1, P_2, P_3, P_4, P_5, P_6, P_7$). Результаты тестирования каждого претендента, на соответствующие вакансии, даны в виде матрицы – С (табл. 2.24). Принимаем условие, что тестирование производилось по десятибалльной шкале.

Определить, какого претендента и на какую вакансию следует принять, причем так, чтобы сумма баллов отобранных претендентов оказалось максимальной.

Таблица 2.24

	V_1	V_2	V_3	V_4	V_5
P_1	7	5	7	6	7
P_2	6	4	8	4	9
P_3	8	6	4	3	8
P_4	7	7	8	5	7
P_5	5	9	7	9	5
P_6	6	8	6	4	7
P_7	7	7	8	6	4

Решение задачи

Построим математическую модель задачи

1. *Переменные задачи*. Введем переменные x_{ij}, принимающие 2 значения:

$x_{ij} = 0$, если i-й претендент (P_i) не принимается на j-ю вакансию (V_j).

$x_{ij} = 1$, если i-й претендент (P_i) принимается на j-ю вакансию (V_j).

2. *Ограничения на переменные задачи*. Очевидно, что все переменные задачи неотрицательные и целые числа: $x_{ij} \geq 0$ и x_{ij} - целые. Кроме того, так как каждый претендент может занять только одну вакансию и все вакансии должны быть заняты, должны удовлетворяться следующие ограничения:

$$\sum_{i=1}^{7} x_{ij} = 1, \quad j = 1,2,\ldots,7\,, \quad \sum_{j=1}^{5} x_{ij} = 1, \quad i = 1,2,\ldots,5\,.$$

Другими словами, в матрице (x_{ij}) суммы элементов по каждой строке и суммы элементов по каждому столбцу должны быть равны единицам. Это условие означает, что выбор претендентов должен быть таким, чтобы в матрице (x_{ij}), представляющей решение задачи, было бы по одной единице в каждой строке и по одной единице в каждом столбце, остальные элементы матрицы должны равняться нулю.

3. Целевая функция в задаче о назначениях.

Необходимо выбрать претендентов так, чтобы суммарное число очков, набранное ими было бы максимальным. Суммарное число набранных очков вычисляется по формуле:

$$Z = \sum_{i=1}^{7}\sum_{j=1}^{5} c_{ij}x_{ij} = 7 \cdot x_{11} + 5 \cdot x_{12} + \ldots + 4 \cdot x_{75}\,.$$

Математическая модель задачи записывается так:

$$\text{н а й т и} \quad \boxed{Z = \sum_{i=1}^{7}\sum_{j=1}^{5} c_{ij}x_{ij} \to \max}$$

при ограничениях:

$$\sum_{i=1}^{7} x_{ij} = 1, \quad j = 1,2,\ldots,7\,,$$

$$\sum_{j=1}^{5} x_{ij} = 1, \quad i = 1,2,\ldots,5\,.$$

Таким образом, задача о назначениях есть частный случай транспортной задачи.

Реализация задачи о назначениях с помощью ППП Excel.

1. Ввод данных. Вводим данные задачи в Excel, при этом нужно ввести два столбца (6 и 7) с нулевыми значениями для сбалансирования задачи. Результаты заполнения таблицы представлены на рис. 2.24.

	A	B	C	D	E	F	G	H	I	J	K
1				Баллы							
2	Претенденты			Вакансии							
3		1	2	3	4	5	6	7		Сумма баллов	
4	1	7	5	7	6	7	0	0		0	
5	2	6	4	8	4	9	0	0			
6	3	8	6	4	3	8	0	0			
7	4	7	7	8	5	7	0	0			
8	5	5	9	7	9	5	0	0			
9	6	6	8	6	4	7	0	0			
10	7	7	7	8	6	4	0	0			
11											
12											
13		1	2	3	4	5	6	7			
14	1								0		
15	2								0		
16	3								0		
17	4								0		
18	5								0		
19	6								0		
20	7								0		
21		0	0	0	0	0	0	0			
22											

Рис. 2.24

В ячейках B4:F10 введены результаты тестирования претендентов, а в ячейках G4:H10 введены нули, что соответствует фиктивным вакансиям.

Ячейки B14:H20 являются изменяемыми ячейками для нашей процедуры.

В ячейках B21:H21 находятся суммы значений соответствующих столбцов изменяемых ячеек. Так, в ячейке B21 находится сумма ячеек B14:B20.

Аналогично в ячейках:

В C21 находится сумма ячеек находится сумма ячеек C14:C20;

В D21 находится сумма ячеек находится сумма ячеек D14:D20;

В E21 находится сумма ячеек находится сумма ячеек E14:E20;

В F21 находится сумма ячеек находится сумма ячеек F14:F20;

В G21 находится сумма ячеек находится сумма ячеек G14:G20;

В H21 находится сумма ячеек находится сумма ячеек H14:H20.

В ячейках I14:I20 находятся суммы значений соответствующих строк изменяемых ячеек. Так, в ячейке I14 находится сумма ячеек B14:H14. Аналогично в ячейках:

в I15 находится сумма ячеек B15:H15;

в I16 находится сумма ячеек B16:H16;

в I17 находится сумма ячеек B17:H17;

в I18 находится сумма ячеек B18:H18;

в I19 находится сумма ячеек B19:H19;

в I20 находится сумма ячеек B20:H20.

Целевая функция заносится в ячейку J3 и вычисляется по формуле «СУММПРОИЗВ(B4:P10;B14:H20)».

2. Заполнение окна процедуры «Поиск решения».

Целевая функция: J3.

Значение целевой функции: max.

Изменяемые ячейки: B14:H20.

Ограничения задачи:

B21:H21=1 и I14:I20=1 (все свободные рабочие места должны быть заняты и все претенденты размещены по вакансиям);

B14:H20 ≥ 0 (изменяемые ячейки должны иметь положительные значения);

B14:H20 – двоичные числа.

В окне «Параметры» установить «Линейная модель», что соответствует решению задачи симплекс-методом. Результаты заполнения окна показаны на рис. 2.25.

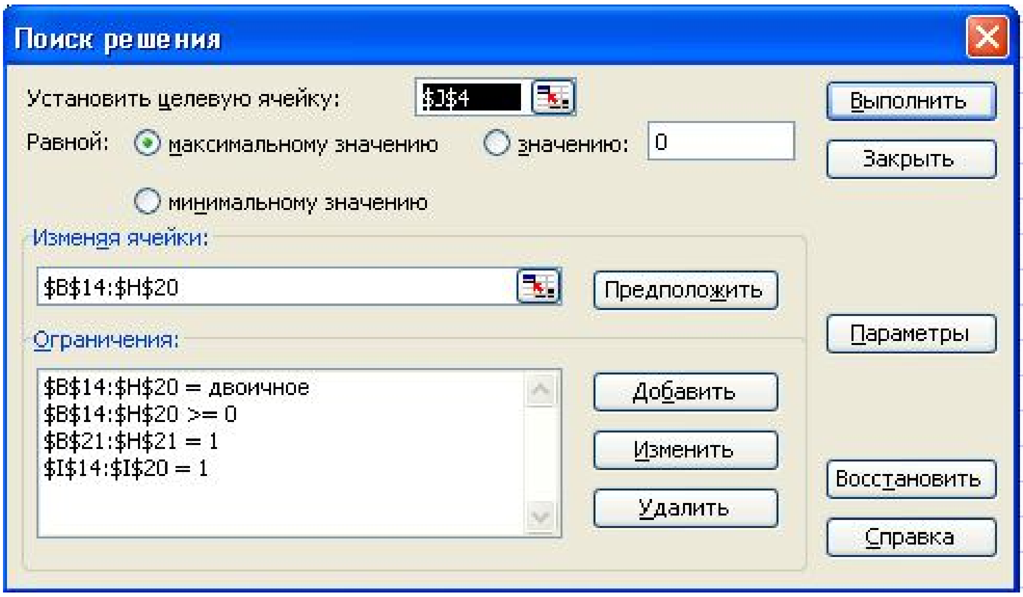


Рис. 2.25

3. Выполнив процедуру «Поиск решения», получим в первоначальной таблице следующие результаты (рис. 2.26).

	A	B	C	D	E	F	G	H	I	J	K
1				Баллы							
2	Претенденты			Вакансии							
3		1	2	3	4	5	6	7		Сумма баллов	
4	1	7	5	7	6	7	0	0		42	
5	2	6	4	8	4	9	0	0			
6	3	8	6	4	3	8	0	0			
7	4	7	7	8	5	7	0	0			
8	5	5	9	7	9	5	0	0			
9	6	6	8	6	4	7	0	0			
10	7	7	7	8	6	4	0	0			
11											
12											
13		1	2	3	4	5	6	7			
14	1	0	0	0	0	0	0	1	1		
15	2	0	0	0	0	1	0	0	1		
16	3	1	0	0	0	0	0	0	1		
17	4	0	0	1	0	0	0	0	1		
18	5	0	0	0	1	0	0	0	1		
19	6	0	1	0	0	0	0	0	1		
20	7	0	0	0	0	0	1	2,22E-16	1		
21		1	1	1	1	1	1	1			
22											

Рис. 2.26

Задачи по теме «Специальные задачи линейного программирования»

Пример 2.18: Формирование оптимального штата фирмы.

Фирма набирает штат сотрудников. Она располагает n группами различных должностей по b_j вакантных единиц в каждой группе, $j = \overline{1, n}$. Кандидаты для занятия должностей проходят тестирование, по результатам которого их разделяют на m групп по a_i кандидатов в каждой группе, $i = \overline{1, m}$. Для каждого кандидата из i-й группы требуются определенные затраты c_{ij} на обучение для занятия j-й должности. В частности, некоторые $c_{ij} = 0$, т.е. кандидат полностью соответствует должности, или $c_{ij} = \infty$, т.е. кандидат вообще не может занять данную должность. *Требуется* распределить кандидатов на должности, затратив минимальные средства на их обучение.

Решение

Предположим, что общее число кандидатов соответствует числу вакантных должностей. Тогда данная задача соответствует транспортной модели. В роли поставщиков выступают группы кандидатов, а в роли потребителей – группы должностей. Предложением является число кандидатов в каждой группе, спросом – число вакансий в каждой группе должностей. В качестве тарифов на перевозки рассматриваются затраты на переобучение. Управляющие переменные x_{ij} - это количество кандидатов из i-й группы, назначенных на j-ю должность.

Математическая модель записывается в виде

$$\boxed{C = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \to \min}$$

$$\begin{cases} \sum_{j=1}^{n} x_{ij} = a_i, \quad i = \overline{1,m}; \\ \sum_{i=1}^{m} x_{ij} = b_j, \quad j = \overline{1,n}; \\ \sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j; \\ x_{ij} \geq 0. \end{cases}$$

Методы решения этой задачи такие же, как и транспортной задачи.

Типовые задачи

Задача 2.1. Продукция определенного типа производится в городах A_1, A_2, A_3 и потребляется в городах B_1, B_2, B_3 и B_4.

В табл. 2.25 указаны: объем производства, спрос, стоимость перевозки единицы продукции.

Требуется составить оптимальный план перевозки продукции, при котором стоимость всех перевозок будет минимальна.

Предварительно следует проверить, сбалансирована ли данная транспортная задача. Если задача не сбалансирована, то нужно ввести фиктивных потребителей или производителей, добавляя к исходной таблице столбцы или строки.

Таблица 2.25

Производители	Потребители				Объем производства
	B_1	B_2	B_3	B_4	
A_1	20	47	31	13	49
A_2	3	38	44	10	18
A_3	11	32	46	17	68
Спрос	45	30	10	45	

Таблица 2.26

Производители	Потребители				Объем производства
	B_1	B_2	B_3	B_4	
A_1	31	13	45	35	48
A_2	38	44	10	33	48
A_3	20	47	31	13	44
Спрос	40	41	45	44	

Таблица 2.27

Производители	Потребители				Объем производства
	B_1	B_2	B_3	B_4	
A_1	45	35	7	43	49
A_2	44	10	33	46	41
A_3	42	41	2	38	68
Спрос	44	12	88	44	

Задача 2. В конкурсе на занятие пяти вакансий (V_1, V_2, V_3, V_4, V_5) участвуют семь претендентов ($P_1, P_2, P_3, P_4, P_5, P_6, P_7$). Результаты тестирования каждого претендента, в случае занятия им одной вакансии, даны в виде матрицы (тестирование производилось по десятибалльной системе).

Определить, какого претендента и на какую вакансию следует принять, причем так, чтобы сумма баллов оказалась максимальной.

Таблица 2.28

	V_1	V_2	V_3	V_4	V_5
P_1	6	4	6	3	6
P_2	4	5	3	6	7
P_3	8	7	8	5	8
P_4	5	5	6	8	4
P_5	8	7	5	4	9
P_6	9	5	6	6	4
P_7	5	8	8	6	5

Таблица 2.29

	V_1	V_2	V_3	V_4	V_5
P_1	4	7	5	7	7
P_2	6	7	3	6	8
P_3	3	4	5	8	7
P_4	7	9	8	5	6
P_5	4	6	4	8	4
P_6	5	4	9	6	5
P_7	7	5	4	7	4

Таблица 2.30

	V_1	V_2	V_3	V_4	V_5
P_1	6	5	6	4	6
P_2	4	7	4	7	6
P_3	8	4	7	3	7
P_4	6	6	7	7	4
P_5	5	9	4	6	4
P_6	6	4	9	4	5
P_7	4	6	5	4	9

2.3.3 Задача оптимального использования ресурсов

В распоряжении предприятия имеется определённое количество ресурсов: рабочая сила, финансы, сырьё, оборудование, производственные площади и т.п. Например, это ресурсы трёх видов: рабочая сила (b_1 чел./дней), сырьё (b_2 кг) и оборудование (b_3 станко/час). Предприятие может выпускать однотипные изделия четырёх видов. Информация о количестве единиц каждого ресурса, необходимых для производства одного изделия каждого вида, и доходах, получаемых предприятием от единицы каждого вида товаров, приведена в табл. 2.31.

Таблица 2.31

Ресурсы	Нормы расхода ресурсов на единицу изделия				Наличие ресурсов
	Изделие 1	Изделие 2	Изделие 3	Изделие 4	
Труд	a_{11}	a_{12}	a_{13}	a_{14}	b_1
Сырьё	a_{21}	a_{22}	a_{23}	a_{24}	b_2
Оборудование	a_{31}	a_{32}	a_{33}	a_{34}	b_3
Цена 1 ед. изделия (тыс. руб.)	c_1	c_2	c_3	c_4	

Требуется найти такой план выпуска продукции, при котором общая стоимость продукции будет минимальной.

Экономико-математическая модель задачи

Обозначим через переменные x_1, x_2, x_3 *и* x_4 число изделий каждого вида. В соответствии с условием задачи целевая функция должна отражать общую стоимость выпускаемой продукции, которая должна быть максимальной:

$$F(x_1, x_2, x_3, x_4) = c_1 \cdot x_1 + c_2 \cdot x_2 + c_3 \cdot x_3 + c_4 \cdot x_4 \rightarrow \max .$$

Ограничения по ресурсам представим системой неравенств:

$$\begin{cases} a_{11} \cdot x_1 + a_{12} \cdot x_2 + a_{13} \cdot x_3 + a_{14} \cdot x_4 \le b_1; \\ a_{21} \cdot x_1 + a_{22} \cdot x_2 + a_{23} \cdot x_3 + a_{24} \cdot x_4 \le b_2; \\ a_{31} \cdot x_1 + a_{32} \cdot x_2 + a_{33} \cdot x_3 + a_{34} \cdot x_4 \le b_1; \\ x_1 \ge 0; \quad x_2 \ge 0; \quad x_3 \ge 0; \quad x_4 \ge 0. \end{cases}$$

Задача о размещении производственных заказов

В планируемом периоде предприятию необходимо обеспечить производство М тыс. однородных новых изделий, которые могут выпускать четыре филиала. Для освоения этого нового вида изделий выделены капитальные вложения в размере С млн. рублей. Разработанные для каждого филиала предприятия проекты освоения нового вида изделия характеризуются величинами удельных капитальных вложений и себестоимостью единицы продукции в соответствии с табл. 2.32.

Таблица 2.32

Показатели	Филиалы предприятия			
	1	2	3	4
Себестоимость производства изделия, руб.	d_1	d_2	d_3	d_4
Удельные капиталовложения, руб.	c_1	c_2	c_3	c_4

2.4 Задачи нелинейного программирования

2.4.1 Классификация и особенности методов нелинейного программирования

В общем виде ***задача нелинейного программирования*** (ЗНП) формулируется следующим образом:

$$\boxed{f(x_1, x_2, \ldots, x_n) \to \max(\min)}. \tag{2.39}$$

$$\begin{cases} g(x_1, x_2, \ldots, x_n) \le b_i, & i = \overline{1, m_1}; \\ g(x_1, x_2, \ldots, x_n) \ge b_i, & i = \overline{m_1 + 1, m_2}; \\ g(x_1, x_2, \ldots, x_n) = b_i, & i = \overline{m_2 + 1, m}, \end{cases} \tag{2.40}$$

где x_j - управляющие переменные или решения задачи нелинейного программирования, $j = \overline{1, n}$;

b_i - фиксированные параметры, $i = \overline{1, m}$;

$f, g_i, i = \overline{1, n}$ - заданные функции от n переменных.

Если f и g_i линейны, то соотношения (2.39) и (2.40) определяют модель задачи линейного программирования.

Решить задачу нелинейного программирования – это значит найти такие значения управляющих переменных x_j, $j=\overline{1,n}$, которые удовлетворяют системе ограничений (2.40) и доставляют максимум (или минимум) функции f.

Для задачи нелинейного программирования, в отличие от линейных задач, нет единого метода решения. В зависимости от вида целевой функции (2.39) и ограничений (2.40) разработано несколько специальных методов решения, к которым относятся: методы множителей Лагранжа, квадратичное и выпуклое программирование, градиентные методы, ряд приближенных методов решения, графический метод. Заметим, что на практике представление экономических задач в виде моделей нелинейного программирования часто бывает довольно искусственным.

Нелинейное программирование (НП) представляет собой раздел в теории математического программирования, предметом которого является изучение экстремальных задач с нелинейными целевыми функциями и ограничениями.

В инженерной практике под нелинейным программированием обычно понимают методы формализации и решения задач параметрической оптимизации нелинейных целевых (критериальных) функций в условиях нелинейных функциональных ограничений. В общем случае при решении прикладных задач нелинейного программирования производится поиск условного экстремума нелинейной целевой функции в области заданных ограничений.

В теории нелинейного программирования сложилось несколько подходов к классификации методов оптимизации.

На рис. 2.27 и 2.28 представлены схемы, поясняющие общую классификацию методов нелинейного программирования. Среди методов НП большое практическое значение имеют *методы поисковой оптимизации*, в которых процедура поиска оптимального решения x^* связана с проведением испытаний в точках x^ν; при этом оптимальное значение вектора параметров x^* определяется с помощью рекуррентных соотношений вида

$$x^* = P[x^0, F(x^0), \varphi(x^{k-1}, F(x^{k-1}), \varphi(x^{k-2}, F(x^{k-2})], \quad k = 1,2,3,...$$

где $x^0 = (x_1^0, x_2^0, ..., x_n^0)^T$ - начальная (опорная) точка.

Если траектория поиска однозначно определяется выбором начальной точки и алгоритмом решения, то такой метод поиска называется **детерминированным**. Если траектория поиска экстремума зависит также и от реализаций некоторого случайного (псевдослучайного) процесса, то такой метод называется **стохастическим**. Методы поиска экстремального значения ЦФ можно разделить на две группы:

1) методы, основанные на использовании первых или вторых частных производных ЦФ и функций ограничений задачи; методы этой группы получили название **градиентных методов поиска**; к ним относятся: методы градиента и наискорейшего спуска, метод сопряженных градиентов, метод проектирования градиента и другие.

2) методы, не использующие информацию о производных ЦФ и функций

ограничений; эти методы получили название **методов прямого поиска**; к ним относятся: метод покоординатного спуска (метод Гаусса-Зайделя), метод вращения координат, метод конфигураций, методы случайного поиска и некоторые другие.

В настоящее время наибольшее практическое применение и теоретическое обоснование получили методы поиска, относящиеся к первой группе. Первые и вторые частные производные от функций определяют градиент и кривизну исследуемой функции. Напомним, что *градиент некоторой непрерывной функции* нескольких переменных представляет собой вектор, составляющими которого являются частные производные этой функции по отдельным переменным; градиент по определению указывает направление, в котором функция возрастает с наибольшей скоростью. Свойство уменьшения величины градиента по мере приближения к экстремуму используют, например, при оптимизации по методу градиента. Это легко проследить по рис 2.29, наблюдая за изменением наклона касательных к кривой $Q(\alpha)$. Для указанного способа характерно плавное движение по направлению к экстремуму $Q^* = Q(\alpha^*)$.

В **методе наискорейшего спуска** движение происходит по направлению вектора градиента до тех пор, пока производная ЦФ по этому направлению не обращается в ноль. Затем вновь определяется направление градиента и движение осуществляется вдоль этого вектора, пока не обратится в ноль производная от ЦФ по данному направлению, и так далее до тех пор, пока не будет достигнут экстремум. Для рассматриваемого метода характерен быстрый выход в окрестность экстремума. Одним из эффективных современных методов безусловной минимизации функций является **метод сопряженных градиентов** (МСГ). Его можно рассматривать как оптимальную реализацию градиентного метода применительно к квадратичной функции. При минимизации произвольной квадратичной функции МСГ позволяет получить точное решение за число шагов, не превышающее число оптимизируемых параметров.

Методы проектирования градиента обобщает градиентный метод на случай минимизации функции при наличии ограничений. Каждый шаг алгоритма поиска состоит при этом из двух этапов – градиентной фазы и фазы восстановления. В градиентной фазе предполагается, что ограничения задачи удовлетворяются, а через опорную точку в пространстве переменных проводятся гиперплоскости, касательные к поверхностям, образованным функциями ограничений, и градиент ЦФ проектируется на пересечение этих гиперплоскостей. Вдоль полученной проекции осуществляется поиск точки с наименьшим значением ЦФ. Полученная точка может не удовлетворять ограничениям задачи. Поэтому в фазе восстановления осуществляется корректировка приращения ЦФ таким образом, чтобы с требуемой точностью удовлетворять ограничениям.

На практике применение градиентных методов ограничивается необходимостью явного определения частных производных ЦФ. Оценка производных с помощью конечных разностей может нарушить сходимость метода к оптимальному решению. Достоинства прямых методов оптимизации состоят в следующем: а) на ЦФ задачи не накладывается никаких ограничений, кроме свойства её непрерывности; б) не требуется непрерывность функций ограничений; в) надежность и простота вычислительной схемы; г) малый объем подготовительной работы при формализации задач НП.

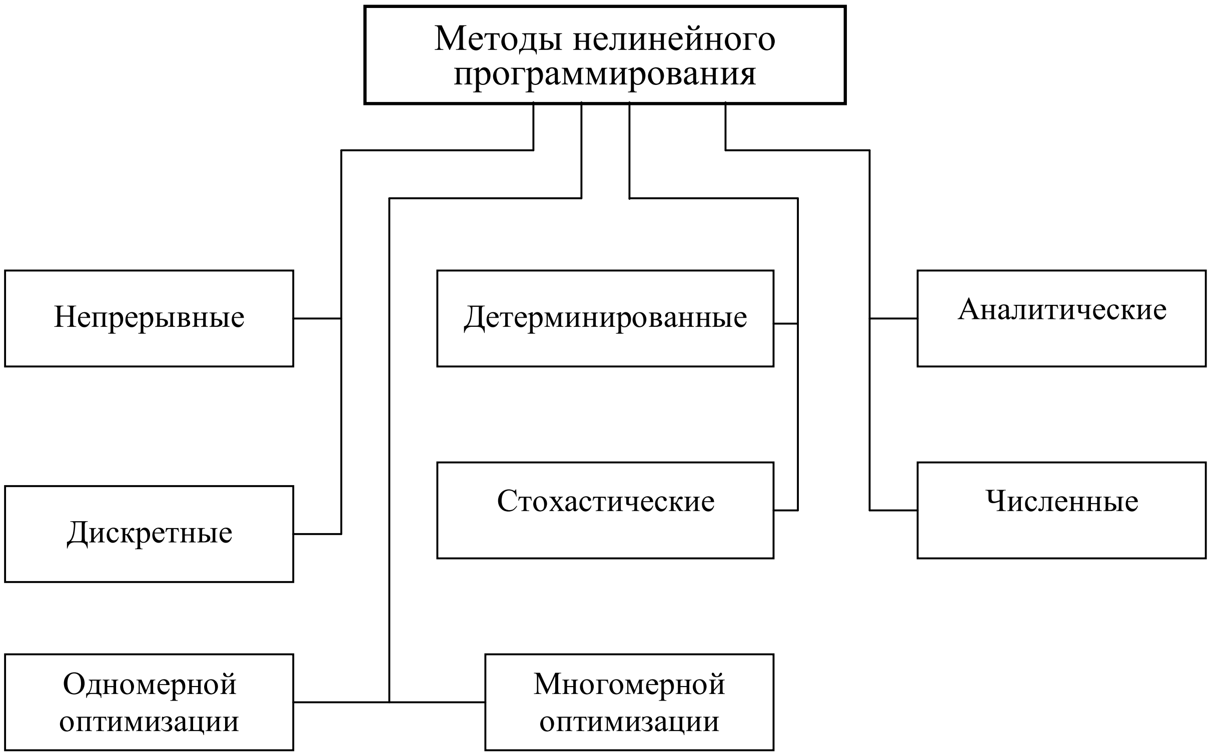


Рис. 2.27 – Классификация методов нелинейного программирования

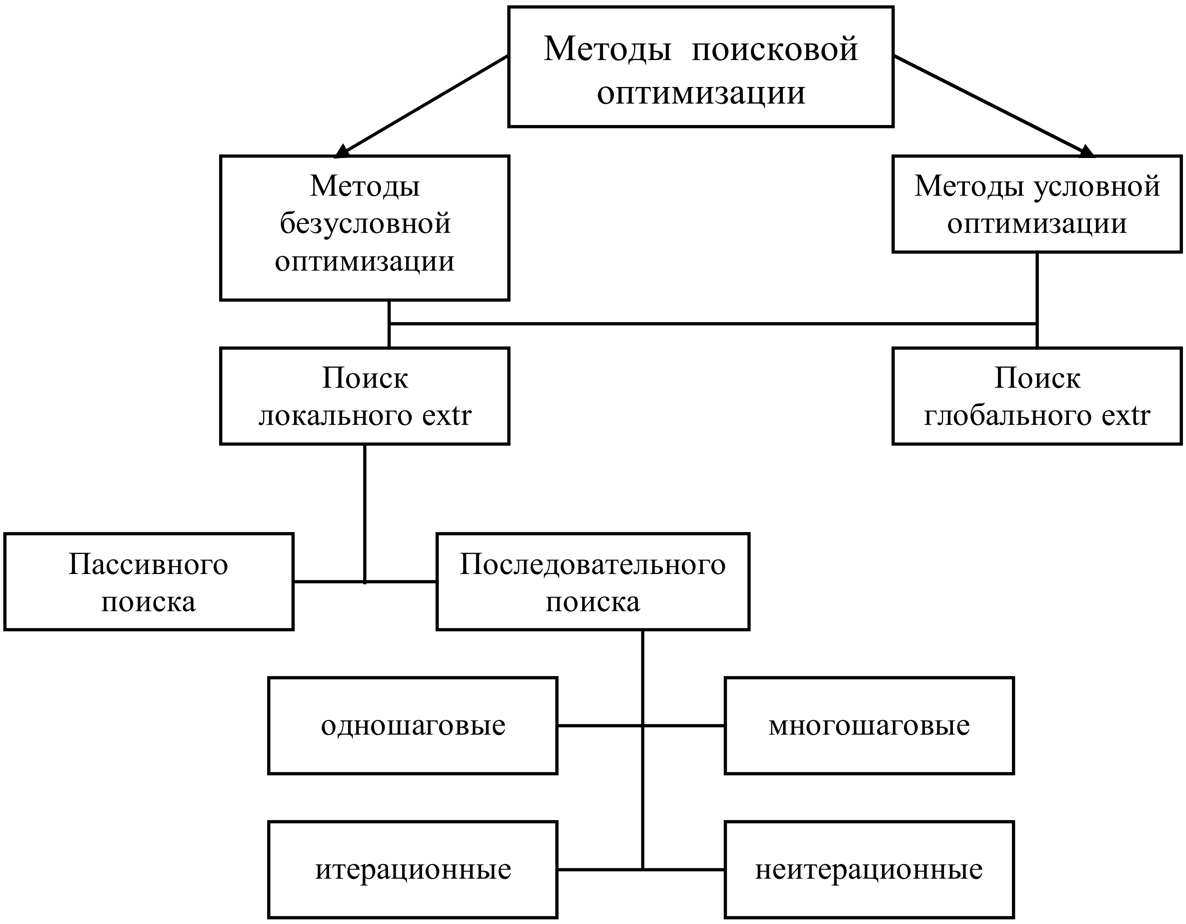


Рис. 2.28 – Классификация поисковых методов оптимизации

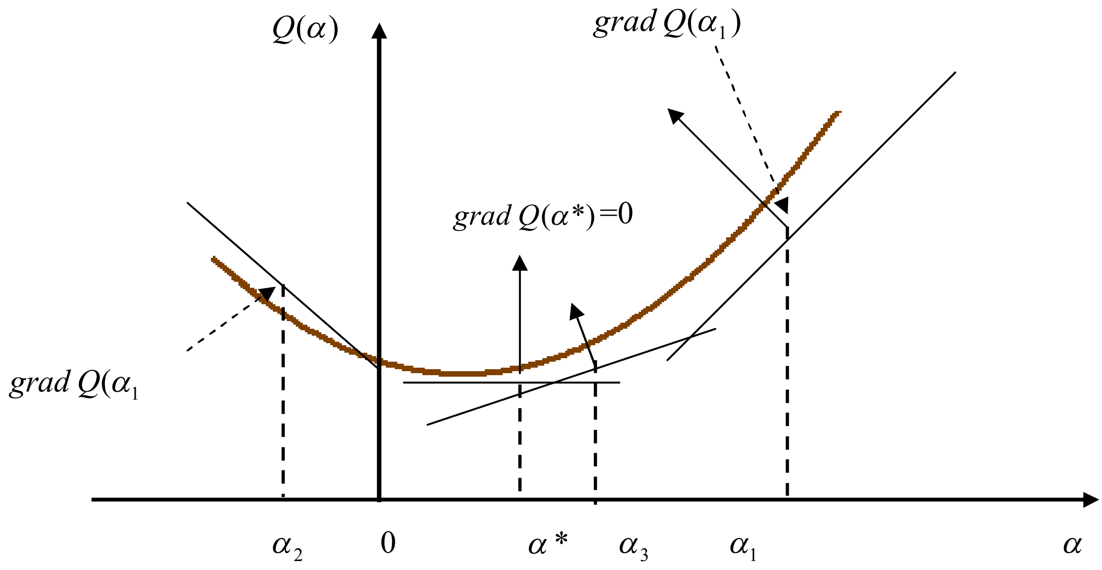


Рис. 2.29 - Траектория поиска экстремума функции $Q(\alpha)$

В качестве общего недостатка методов прямого поиска укажем проблему остановки процесса вычислений. Эта трудность, как правило, преодолевается с применением интерактивной технологии решения задач оптимизации.

2.4.2 Методы регулярного поиска безусловного экстремума

Метод Гаусса-Зайделя. Сущность этого метода заключается в поочередном нахождении частных экстремумов. Сначала изменяется, например, параметр α_1 (остальные сохраняются постоянными) и отыскивается частный экстремум по этому параметру $\left(\frac{\partial Q}{\partial \alpha_1}=0\right)$. После этого он фиксируется, начинается изменение параметра α_2 и отыскивается второй частный экстремум $\left(\frac{\partial Q}{\partial \alpha_2}=0\right)$ и т.д. После того как будут найдены все частные экстремумы, вновь изменяется параметр α_1 и отыскивается экстремум $\frac{\partial Q}{\partial \alpha_1}=0$. Затем определяются и другие частные экстремумы $\left(\frac{\partial Q}{\partial \alpha_2},\ldots,\frac{\partial Q}{\partial \alpha_m}\right)$.

Процесс поиска носит циклический итерационный характер и заканчивается, когда все частные производные целевой функции $\frac{\partial Q}{\partial \alpha_i}$, $i=1,2,\ldots,m$ становятся меньше некоторого порогового значения, определяемого чувствительностью измерительных устройств.

Процесс движения системы к экстремуму для простого случая двух варьируемых параметров иллюстрируется рис. 2.30. На этом рисунке функция

$Q(\alpha_1, \alpha_2)$ изображена линиями равных значений Q_i $(Q_1 < Q_2 < Q_3 \ldots)$. Предположим, что в некоторый момент времени система находилась в точке A и параметр α_1 увеличивается $(\alpha_2 = const)$. Увеличение α_1 будет происходить до тех пор, пока производная $\dfrac{\partial Q}{\partial \alpha_1}$ не обратится в нуль. Это происходит в точке B, где прямая $\alpha_2 = const$ становится касательной к одной из линий равных значений Q. В точке B происходит переключение системы на изменение второго параметра α_2, а первый параметр сохраняется при этом неизменным $(\alpha_1 = \alpha_{1B} = const)$. В точке C при $\alpha_2 = \alpha_{2C}$ достигается второй частный экстремум $\left(\dfrac{\partial Q}{\partial \alpha_2} = 0\right)$. После этого опять начинает изменяться параметр α_1 и т.д.

Процесс поиска заканчивается лишь тогда, когда параметры α_1, α_2 принимают значения, лежащие в заштрихованной области экстремальных значений параметров. Размеры этой области определяются чувствительностью измерительных средств. Как видно из рис. 2.30 (линия $ABCD\ldots$), движение системы в область экстремума происходит не по кратчайшему пути, что увеличивает время поиска, однако метод Гаусса-Зайделя может быть легко реализован.

Метод градиента. При использовании этого метода в каждый данный момент времени измеряются все составляющие градиента функции $Q(\alpha_1, \ldots, \alpha_m)$, т.е. определяется направление быстрейшего возрастания этой функции и обеспечивается движение точки $(\alpha_1, \ldots, \alpha_m)$ в этом направлении. Для этого все варьируемые параметры системы $\alpha_1, \ldots, \alpha_m$ изменяются со скоростями, пропорциональными составляющим $gradQ$:

$$\frac{\partial \alpha_1}{\partial t} = K_1 \frac{\partial Q}{\partial \alpha_1}; \quad \frac{\partial \alpha_2}{\partial t} = K_2 \frac{\partial Q}{\partial \alpha_2}, \quad , \frac{\partial \alpha_m}{\partial t} = K_m \frac{\partial Q}{\partial \alpha_m}.$$

Движение точки (α_1, α_2) к области экстремальных значений при двух варьируемых параметрах иллюстрируется рис. 2.30. Траектория этой точки (кривая $A'B'C'$) нормальна к кривым равных значений функции $Q(\alpha)$.

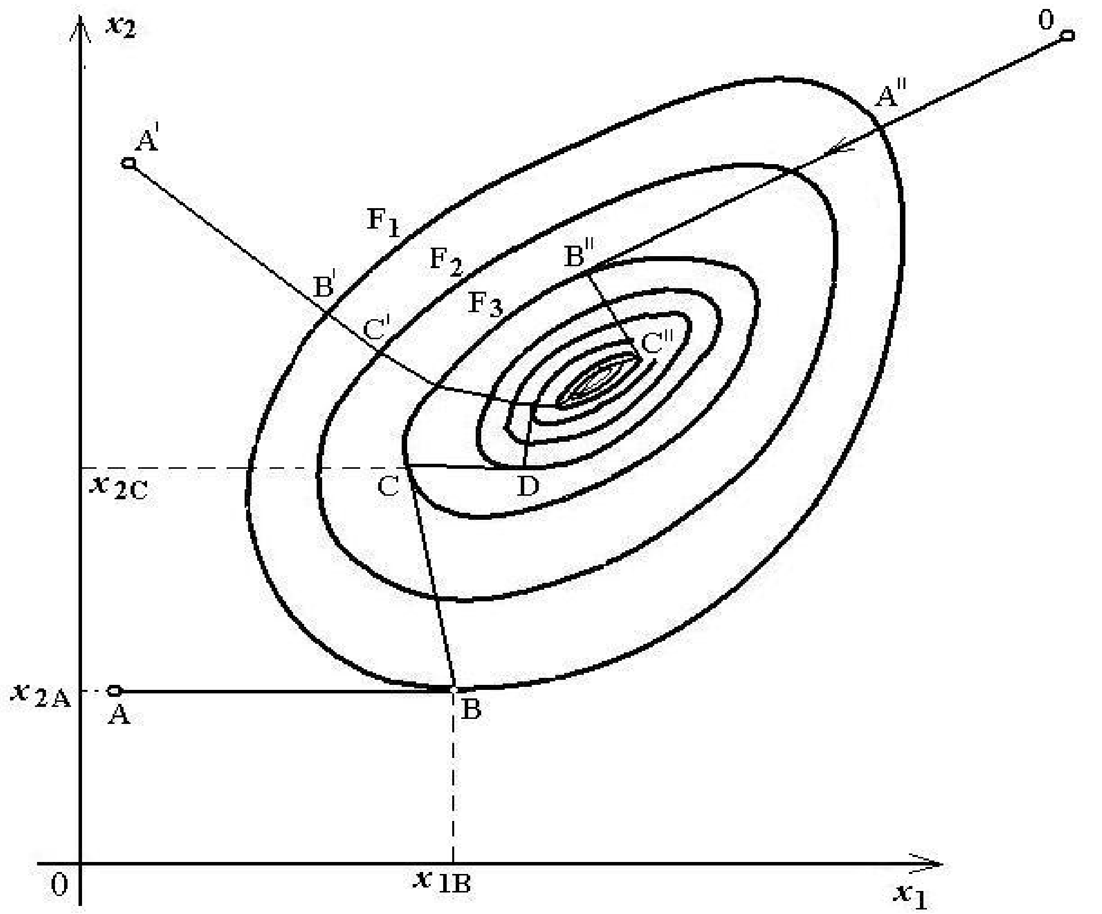


Рис. 2.30 - Схема движения к экстремуму для случая двух параметров

Метод наискорейшего спуска. При использовании метода наискорейшего спуска поиск экстремума производится следующим образом. В начальной точке A'' на рис. 2.30 определяется направление вектора градиента Q. Это направление (отрезок l) нормально к кривой постоянных значений Q_1. После этого осуществляется движение системы в найденном направлении до тех пор, пока производная функции Q_1, взятая по этому направлению, $\frac{\partial Q_1}{\partial l}$ не обратиться в нуль, т.е. пока не наступит частный экстремум (точка B'' на рис. 2.30, где прямая движения является касательной к кривой Q_3). В точке B'' вновь определяется направление вектора градиента (перпендикулярно направлению l) и осуществляется движение системы в этом новом направлении до наступления другого частного экстремума (точка C'') и т. д. до выхода системы в область экстремума.

Достоинством метода наискорейшего спуска является относительно малое число вычислительных операций, требуемых для выхода в область экстремума.

Метод сопряжённых градиентов

Метод сопряжённых градиентов (МСГ) относится к группе методов регулярного поиска и использует квадратичное приближение целевой функции. Этот метод является многошаговым, достаточно простым для реализации на ЭВМ, и в то же время он обладает достаточно высокой скоростью сходимости:

число итераций в данном случае равно числу неизвестных параметров. Особенностью метода является то, что он пригоден только для оптимизации детерминированных гладких и выпуклых функций.

Поиск минимума функции осуществляется с помощью следующей итерационной процедуры:

$$x[k] = x[k] - \gamma[k] \cdot g[k], \tag{2.41}$$

где $g[k]$ - направление поиска, задаваемое методом сопряжённых градиентов в виде

$$g[k] = grad\ F(\ x[k] + \alpha[k] \cdot g[k-1]). \tag{2.42}$$

Коэффициент $\gamma[k]$, определяющий величину шага, находится из условия минимума функции $F(x)$ в данном направлении $g[k]$, т.е. из условия

$$F(\ x[k] - \gamma[k] \cdot g[k]) \to \underset{\gamma[k]}{min}. \tag{2.43}$$

Отсюда следует, что

$$\left\langle\ g[k], \nabla F(x[k+1]\ \right\rangle = 0. \tag{2.44}$$

Здесь и далее знак $\langle \cdot,\cdot \rangle$ означает скалярное произведение.

Из (2.42) видно, что направление поиска $g[k]$ на k-й итерации зависит от направления поиска на $(k-1)$-й итерации и градиента целевой функции $F(x)$ в точке $x[k]$. При вычислении $g[k]$ по формуле (2.42) основные трудности заключаются в определении коэффициента $\alpha[k]$. При выводе формул для вычисления $\alpha[k]$ необходимо получение максимальной величины разности $F(x[k]) - F(x[k+1])$ по данным, полученным на $(k+1)$- шаге поиска. При этом используется квадратичное приближение минимизируемого функционала

$$F(x[k+1]) = F\ (x[k] - \gamma[k] \cdot g[k]) = F(x[k]) - \langle \gamma[k] \cdot g[k] \rangle,$$

$$\nabla F(x[k]) + \frac{1}{2} \cdot \gamma^2[k] \cdot \langle H[k] \cdot g[k],\ g[k] \rangle\ ,$$

где $H[k]$ - матрица вторых производных функции $F(x)$ в точке $x[k]$.

Проводя минимизацию выражения (2.44) по параметру $\gamma[k]$, а затем минимизацию $\langle H[k] \cdot g[k],\ g[k] \rangle$ по $\alpha[k]$, получим следующее выражение для $\alpha[k]$:

$$\alpha[k] = -\frac{\langle \nabla F(x[k],\ H[k] \cdot g[k-1] \rangle}{\langle g[k],\ H[k] \cdot g[k-1] \rangle}. \tag{2.45}$$

Необходимость вычисления матрицы вторых производных ограничивает использование формулы (2.45). Если на интервале $(x[k], x[k-1])$ матрица $H[k]$ изменяется незначительно, то для приближённого вычисления параметра $\alpha[k]$ можно воспользоваться формулой

$$\alpha[k] \approx \frac{\langle\ \nabla F(x[k],\ \nabla F(x[k]) - \nabla F(x[k-1])\ \rangle}{[\nabla F(x[k-1])]^2}. \tag{2.46}$$

Известно, что при минимизации квадратичных функционалов с помощью МСГ градиенты в точках $x[k]$ и $x[k-1]$ ортогональны, т.е.
$\langle\ \nabla F(x[k],\ \nabla F(x[k-1])\ \rangle = 0$. Поэтому формулу (2.46) можно записать так:

$$\alpha[k] \approx \frac{[\nabla F(x[k]]^2}{[\nabla F(x[k-1])]^2}. \qquad (2.47)$$

Кроме того, при минимизации квадратичных функционалов направления поиска удовлетворяют соотношению

$$\langle g[i], Bg[j] \rangle = 0, \quad i, j = 1,2,3,...,n-1; \quad i \neq j, \qquad (2.48)$$

где B - симметрическая матрица квадратичного функционала. Векторы $g[k]$ *и* $g[k-1]$, удовлетворяющие условию (2.48), называются *сопряжёнными*.

Рассмотренный метод регулярного поиска можно рекомендовать для решения задач многопараметрической оптимизации квадратичных функционалов, отражающих качество автоматического управления объектом. При наличии случайных погрешностей при расчёте значений функционала сходимость рассмотренного метода снижается.

2.4.3 Методы случайного поиска

Пусть рассматривается задача максимизации целевой функции F(x) на множестве параметров $x \in D_x$, $x=(x_1,x_2,...,x_n)^T$.

Методы случайного поиска используют в следующих случаях:

- при большой размерности вектора параметров ($n \geq 3$);
- если значения ЦФ вычисляются неточно, и направление градиента определить невозможно;
- на управляемые параметры накладываются областные двухсторонние ограничения:

$$x_{i\min} \leq x_i \leq x_{i\max}, \; i = \overline{1,n}.$$

Методы случайного поиска (МСП) основаны на использовании датчика случайных чисел для формирования составляющих вектора параметров x^k на очередном k-м шаге поиска. Выделяют две группы МСП: методы слепого поиска и методы направленного поиска.

Метод «слепого» случайного поиска

Метод связан с проведением серии N опытов (проб, шагов, итераций) по определению максимального значения ЦФ для различных вариантов вектора параметров, который формируется с помощью датчика случайных чисел с учетом областных ограничений. Опыт считается удачным, если выполняются функциональные ограничения, заданные системой неравенств.

После выполнения N удачных опытов выбирается лучший результат $J^* = F(x^*)$ в соответствии с целью оптимизации, который и принимается за экстремум ЦФ, одновременно определяются и его координаты $x^* = (x_1^*,...,x_n^*)^T$.

Рассмотрим вычислительную схему метода.

ШАГ 1. $k = 1$; $y^{[0]} = 0$; задание для каждого параметра x_i, $i = 1,...,n$ *диапазона варьирования* $[x_{min}; x_{max}]$.

ШАГ 2. Генерация вектора случайных чисел с прямоугольным законом распределения $\{\beta_i\} \in [0;1]$, $i = \overline{1,n}$.

ШАГ 3. Задание вектора управляемых параметров на k–м шаге поиска по формуле:

$$x_i^{[k]} = x_{i\min} + (x_{i\max} - x_{i\min}) \cdot \beta_i^{[k]}, \qquad i = \overline{1,n}.$$

ШАГ 4. Вычисление значения ЦФ

$$y^{[k]} = F(x^{[k]}).$$

ШАГ 5. Сравнение значений

$$y^{[k]} > y^{[k-1]}.$$

Если условие выполняется, то переход на шаг 7.

Если условие не выполняется – переход на шаг 6.

ШАГ 6. Идентификация

$$y^{[k]} = y^{[k-1]}; \;\; x^{[k]} = x^{[k-1]}.$$

ШАГ 7. Анализ условия $k \geq N$.

Если условие выполняется – выход из цикла проведения опытов и переход на шаг 8. Иначе – $k = k+1$ и возврат на шаг 2.

ШАГ 8. Фиксация результатов, полученных на последнем шаге поиска (в k-м опыте)

$$F^* = y^{[k]}(x^*); \quad x^* = x^{[k]}.$$

Метод направленного случайного поиска (базовый вариант)

Пусть $x^{[0]}$ – произвольная точка в области допустимых значений D_x. Из точки $x[0] = (x_1^{[0]},...,x_n^{[0]})^T$ совершается движение с шагом h в случайном направлении с равномерным распределением. Движение представляющей точки описывается в виде

$$x^{[k+1]} = \begin{cases} x^{[k]} + h\varepsilon_k, & \textit{если} \quad F(x^{k+1}) > F(x^k); \\ x^{[k]} - h\varepsilon_k, & \textit{если} \quad F(x^{k+1}) \leq F(x^k). \end{cases}$$

Здесь $h=(h_1,...,h_n)^T$ – шаг приращения параметров, обычно принимается $h_i=const$.

$\varepsilon_k = (\varepsilon_1^k,...,\varepsilon_n^k)^T$, $\varepsilon_i^k = \Delta x_i^k = x_i^{k+1} - x_i^k$ - приращение i-го параметра на $(k+1)$-м шаге поиска.

Выход из цикла поиска экстремума ЦФ осуществляется в случае, если после проведения нескольких шагов абсолютное приращение ЦФ станет меньше априорно заданной ошибки $\Delta F(x) \leq \delta F$.

Если ЦФ $F(x)$ имеет больше одного экстремума и нужно найти наименьший из них, то ни один из алгоритмов регулярного поиска не решит поставленную задачу. Для этого используют алгоритмы глобального случайного поиска.

Метод глобального случайного поиска при правильной организации является достаточно эффективным средством отыскания глобального экстремума. Состоит он в следующем. В области поиска F_2 в соответствии с заданной, например, нормальной, плотностью распределения $P(x|m,\sigma^2)$ генерируются случайные точки $x_1, x_2, ...$. Здесь m - математическое ожидание этого распределения, а σ^2 - максимальная дисперсия его компонент. Смысл алгоритма состоит в том, что m располагается в точке с минимальным, найденным ранее, значением минимизируемой функции, $m_k = x_k*,\quad где,\quad J(x_i*) = \min_x J(x)$.

Алгоритм поиска экстремума записывается в виде рекуррентной формулы для $x*_k \quad и \quad J*_k$:

$$x*_k = \begin{cases} x*_{k-1} \quad и \quad J*_k = J*_{k-1}, & если\, J(x_k) \ge J*_{k-1}; \\ x_k \quad и \quad J*_k = J(x_k), & если\, J(x_k) < J*_{k-1}. \end{cases} \tag{2.49}$$

Отметим, что этот алгоритм гарантирует отыскание глобального экстремума, если плотность распределения случайных точек не равна нулю ни в одной точке области поиска D, т. е.

$$P\{x|m,\sigma^2\} \neq 0 \quad для\; x \in D.$$

Действительно, так как вероятность случайного попадания в окрестность глобального экстремума в этом случае конечна, то это событие рано или поздно реализуется. Так как сходимость алгоритма поиска относительно медленная, поэтому при реализации этого алгоритма следует принимать меры для увеличения скорости его сходимости.

Случайные пробы x определяются m-мерным нормальным законом распределения вероятностей $P(x*_k, \sigma^2{}_k)$, где $x*_k$ - математическое ожидание этого распределения, соответствующее наилучшей оценке глобального экстремума на k`-м шаге; $\sigma^2{}_k$ – дисперсия. Таким образом, в памяти алгоритма на каждом k-м шаге хранятся параметры $\{x_k^*, \sigma^2{}_k, J_k^*\}$, которые изменяются следующим образом.

Рекуррентные формулы для x_k^* и J_k^* сохраняют вид согласно (2.49), а СКО σ_k изменяется в соответствии с правилом

$$\sigma_k = \begin{cases} \sigma_0, & если\;\; J(x_k) < J*_{k-1}, \\ \sigma_{k-1} - f(\sigma_{k-1}), & если\;\; J(x_k) \ge J*_{k-1} \end{cases} \tag{2.50}$$

где σ_0^2 - исходная (заданная априори) дисперсия, $f(\sigma_k)$ - функция, определяющая процедуру уменьшения СКО σ_k (в частности, $f(\sigma_k) = q \cdot \sigma_k$, $(1 \le q \le 1)$.

Согласно предложенной схеме алгоритм работает следующим образом. Случайные пробы производятся в соответствии с алгоритмом формирования случайных чисел с нормальным ЗРВ; математическое ожидание вектора параметров $\{x\}$ соответствует наилучшей предыдущей пробе, а дисперсия σ_x^2 уменьшается при неудачных пробах и увеличивается до априорно заданной величины σ_0 при удачных пробах. Уменьшение дисперсии в данном случае связано с необходимостью уточнения найденного локального экстремума, а увеличение – с необходимостью изучения новой ситуации для отыскания наилучшей точки в расширенной области.

Изложенный метод отражает «осторожный подход» к задаче поиска глобального экстремума ЦФ и обеспечивает его гарантированное определение при большом числе проб ($N \to \infty$) и соответствующем выборе функции $f(\sigma)$. При $f(\sigma)=0$ рассмотренный алгоритм поиска экстремума вырождается в случайный перебор (слепой поиск) с нормальным распределением проб.

На рис. 2.31 представлена схема алгоритма определения минимума аналитически заданной целевой функции $J(x)$ и её координат x^* на основе метода глобального случайного поиска.

Геометрическую интерпретацию задачи нелинейного программирования дадим с помощью графического метода её решения.

Рассмотрим ЗНП, содержащую две переменные

$$f(x_1, x_2) \to \max(\min). \tag{2.51}$$

$$\begin{cases} g(x_1, x_2) \le b_i, & i = \overline{1, m_1}; \\ g(x_1, x_2) \ge b_i, & i = \overline{m_1 + 1, m_2}; \\ g(x_1, x_2) = b_i, & i = \overline{m_2 + 1, m}. \end{cases} \tag{2.52}$$

Система ограничений (2.52) определяет в n-мерном пространстве некоторую область, которая является областью допустимых решений задачи.

Решить задачу нелинейного программирования – это значит найти точку области допустимых решений (2.52), через которую проходит линия $f(x_1, x_2) = C$ наивысшего (низшего) уровня.

Указанная точка может находиться как на границе, так и внутри области допустимых решений (2.52), в отличие от задач линейного программирования.

Аналогично линейным задачам задачи нелинейного программирования, удобно решать графически, когда целевая функция и ограничения содержат две переменные.

Алгоритм решения задач нелинейного программирования графическим методом:

Шаг 1. На плоскости x_1Ox_2 строят область допустимых решений, определенную ограничениями (2.52). Если область пуста, т. е. ограничения несовместны, то задача (2.51) - (2.52) не имеет решения. В противном случае переходят к шагу 2.

Шаг 2. Строят линии уровня функции $f(x_1, x_2) = C$, где C - некоторая константа. Переход к шагу 3.

Шаг 3. Определяют направление возрастания (при максимизации), убывания (при минимизации) функции f.

Шаг 4. Находят точку области допустимых решений, через которую проходит линии уровня $f(x_1, x_2) = C$ с наибольшим (при максимизации), наименьшим (при минимизации) значением C или устанавливают неограниченность функции на области допустимых решений.

Шаг 5. Определяют значения x_1, x_2 для точки, найденной на шаге 4, и величину функции f в этой точке.

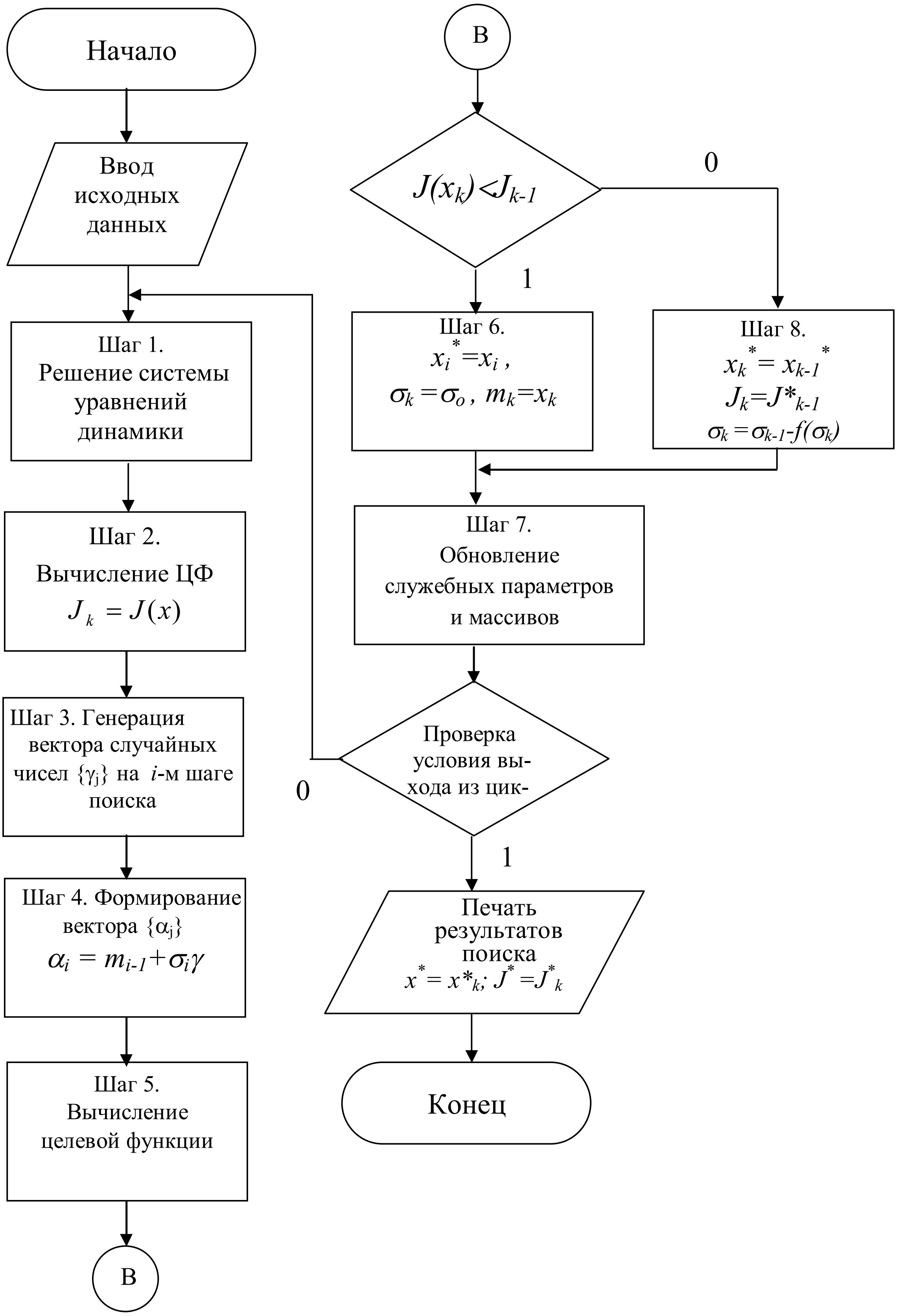


Рис. 2.31 - Схема алгоритма случайного глобального поиска

Пример 2.19

Решить задачу нелинейного программирования $f = 2x_1 + 3x_2 \to \max(\min)$

$$\begin{cases} x_1 x_2 \geq 3; \\ x_1^2 + x_2^2 \leq 16; \\ x_1 \geq 0; \\ x_2 \geq 0. \end{cases}$$

Решение

В соответствии с алгоритмом построим на плоскости x_1Ox_2 область допустимых решений (рис. 2.32).

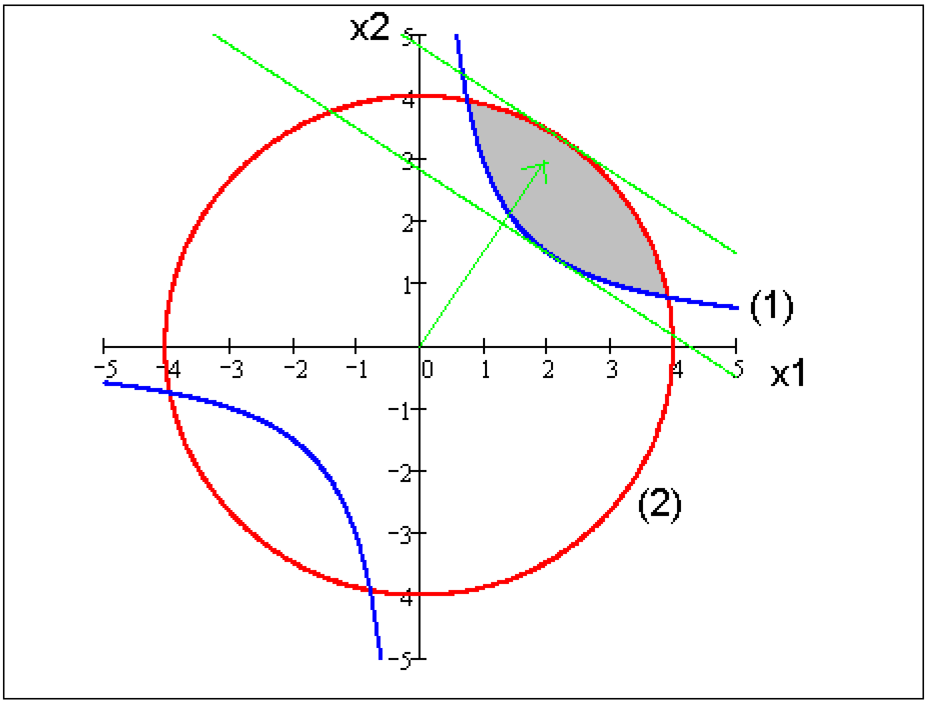


Рис. 2.32

Ограничения $x_1 \geq 0$, $x_2 \geq 0$ выделяют на плоскости x_1Ox_2 первую четверть. Границей полуплоскости, соответствующей первому ограничению является гипербола $x_2 = 3/x_1$. Неравенство выполняется для точек, лежащих выше гиперболы. Границей полуплоскости, определяемой вторым ограничением, является окружность с центром в точке (0; 0) и радиусом, равным 4.

Функция возрастает в направлении вектора-нормали $\vec{n}$ с координатами $(2;3)$, и её линии уровня расположены перпендикулярно вектору-нормали $\vec{n}$. Таким образом, максимум достигается в точке A, а минимум – в точке B.

Заметим, что в точке A совпадают тангенсы углов наклона касательной к окружности $x_1^2 + x_2^2 = 16$ и прямой $2x_1 + 3x_2 = C_1$ к оси Ox_1. Тангенсы углов наклона касательной и прямой к оси Ox_1 определяются значениями производных по x_1 соответствующих функций.

Для прямой $x_2 = -\frac{2}{3}x_1 + \frac{C_1}{3}$ тангенс угла наклона равен $(-\frac{2}{3})$. Продифференцируем выражение $x_1^2 + x_2^2 = 16$ как неявную функцию от x_1. В итоге вычислим

$$2x_1 + 2x_2 x_2' = 0, \quad x_2' = -\frac{x_1}{x_2}.$$

Приравниваем значения тангенсов, откуда получаем

$$-\frac{x_1}{x_2} = -\frac{2}{3}, \quad 3x_1 - 2x_2 = 0.$$

К этому уравнению добавим уравнение окружности, которой принадлежит точка A. Получаем систему уравнений

$$\begin{cases} 3x_1 - 2x_2 = 0; \\ x_1^2 + x_2^2 = 16. \end{cases}$$

Решив систему уравнений, найдем оптимальное решение

$$x_1 = \frac{12}{\sqrt{3}}; \quad x_2 = \frac{12}{\sqrt{13}}; \quad f_{\max} = \frac{52}{\sqrt{3}}.$$

Аналогично определим координату точки B, в которой тангенс угла наклона к оси Ox_1 прямой $2x_1 + 3x_2 = C_2$ совпадает с тангенсом угла наклона касательной к функции $x_1 x_2 = 3$.

$$x_2' = -\frac{3}{x_1^2}.$$

Получаем уравнение

$$-\frac{3}{x_1^2} = -\frac{2}{3}.$$

Вторым соотношением для нахождения координат точки является уравнение гиперболы, которой принадлежит точка B:

$$\begin{cases} -\dfrac{3}{x_1^2} = -\dfrac{2}{3}; \\ x_1 x_2 = 3. \end{cases}$$

Из последней системы уравнений найдем оптимальное решение, соответствующее минимальному значению f,

$$x_1 = \frac{3}{\sqrt{2}}; \quad x_2 = \sqrt{2}; \quad f_{\min} = \frac{12}{\sqrt{2}}.$$

Задача решена.

Метод множителей Лагранжа

Пусть требуется решить задачу нелинейного программирования следующего вида:

$$f(x_1, x_2, \ldots, x_n) \to \max(\min). \tag{2.53}$$

$$\begin{cases} g_1(x_1, x_2, \ldots, x_n) = b_1; \\ g_2(x_1, x_2, \ldots, x_n) = b_2; \\ \ldots \\ g_m(x_1, x_2, \ldots, x_n) = b_m, \end{cases} \tag{2.54}$$

где f и g_i, $i = \overline{1, n}$, - функции параметров $x_1, \ldots, x_n$.

Для решения поставленной задачи может быть применен *метод множителей Лагранжа*. Рассмотрим особенности известного метода на примере задачи максимизации функции двух переменных

$$f(x_1,x_2)\to\max,\quad g(x_1,x_2)=b.$$

На плоскости x_1Ox_2 уравнение $g(x_1,x_2)=b$ определяет график некоторой функции. На нём указывают несколько линий уровня некоторой функции $f(x_1,x_2)$ и выбранное в качестве примера направление её возрастания.

В некоторой точке A, в которой функция f достигает максимального значения, совпадают касательные линии к графикам функции

$$f(x_1,x_2)=C \text{ и } g(x_1,x_2)=b.$$

Следовательно, в точке A векторы-нормали к функциям $f(x_1,x_2)=C$ и $g(x_1,x_2)=b$ пропорциональны. Обозначим эти векторы соответственно через k и l. Получаем $\vec{l}=\lambda\vec{k}$, где λ - некоторый коэффициент пропорциональности. Координатами векторов $\vec{k}$ и $\vec{l}$ являются значения частных производных функций f и g соответственно в точке A.

$$\vec{l}\left(\frac{\partial f}{\partial x_1};\frac{\partial f}{\partial x_2}\right);\quad \vec{k}\left(\frac{\partial g}{\partial x_1};\frac{\partial g}{\partial x_2}\right).$$

Из условия пропорциональности в точке A имеем

$$\frac{\partial f}{\partial x_1}=\lambda\frac{\partial g}{\partial x_1};\quad \frac{\partial f}{\partial x_2}=\lambda\frac{\partial g}{\partial x_2}.$$

Для определения значений x_1,x_2, при которых функция f достигает максимума к этим уравнениям надо добавить условие принадлежности точки A графику функции $g(x_1,x_2)=b$.

Окончательно получаем систему уравнений, определяющую оптимальное решение поставленной задачи

$$\begin{cases}\dfrac{\partial f}{\partial x_1}=\lambda\dfrac{\partial g}{\partial x_1};\\ \dfrac{\partial f}{\partial x_2}=\lambda\dfrac{\partial g}{\partial x_2};\\ g(x_1,x_2)=b.\end{cases}$$

Введем вспомогательную функцию – функцию Лагранжа

$$F(x_1,x_2,\lambda)=f(x_1,x_2)+\lambda(b-g(x_1,x_2)).$$

Тогда последнюю систему уравнений можно записать в виде

$$\begin{cases}\dfrac{\partial F(x_1,x_2,\lambda)}{\partial x_1}=\dfrac{\partial f(x_1,x_2)}{\partial x_1}-\lambda\dfrac{\partial g(x_1,x_2)}{\partial x_1}=0;\\ \dfrac{\partial F(x_1,x_2,\lambda)}{\partial x_2}=\dfrac{\partial f(x_1,x_2)}{\partial x_2}-\lambda\dfrac{\partial g(x_1,x_2)}{\partial x_2}=0;\\ \dfrac{\partial F}{\partial\lambda}=b-g(x_1,x_2)=0.\end{cases}$$

Алгоритм метода множителей Лагранжа:

Шаг 1. Составляют функцию Лагранжа

$$F(x_1,\ldots,x_n,\lambda_1,\ldots,\lambda_m)=f(x_1,\ldots,x_n)+\sum_{i=1}^{m}\lambda_i(b_i-g_i(x_1,\ldots,x_n)).$$

Шаг 2. Находят частные производные функции Лагранжа по x_j и λ_j, $j=\overline{1,n}$; $i=\overline{1,m}$ и приравнивают их к нулю

$$\begin{cases} \dfrac{\partial F}{\partial x_j} = \dfrac{\partial f}{\partial x_j} - \sum\limits_{i=1}^{m} \lambda_i \dfrac{\partial g}{\partial x_j} = 0; \quad j=\overline{1,n} \\ g_i(x_1,\ldots,x_n) = b_i; \quad i=\overline{1,m}. \end{cases} \qquad (2.55)$$

Шаг 3. Решают систему (2.55) и определяют точки, в которых функция $f(x_1,\ldots,x_n)$ может иметь экстремум.

Шаг 4. Проверяют полученные на шаге 3 точки на экстремум и определяют экстремальное значении функции f в найденной точке.

Пример 2.20

Фирма реализует автомобили двумя способами: через магазин и через торговых агентов. При реализации x_1 автомобилей через магазин расходы на реализацию составляют $4x_1 + x_1^2$ у.е., а при продаже x_2 автомобилей через торговых агентов расходы составляют x_2^2 у.е. Найти оптимальный способ реализации автомобилей, минимизирующий суммарные расходы, если общее число предназначенных для продажи автомобилей составляет 200 штук.

Решение задачи

Составим математическую модель задачи.

Целью является минимизация суммарных расходов $R = 4x_1 + x_1^2 + x_2^2$.

Управляющие переменные – это число автомобилей, реализуемых первым и вторым способом: x_1 и x_2, соответственно (всего - 200 штук).

Окончательно математическая модель имеет следующий вид:

$$R = 4x_1 + x_1^2 + x_2^2 \to \min .$$

$$\begin{cases} x_1 + x_2 = 200, \\ x_1 \ge 0, \\ x_2 \ge 0. \end{cases}$$

Для решения задачи воспользуемся методом множителей Лагранжа.

Функция Лагранжа имеет вид:

$$F(x_1, x_2, \lambda) = 4x_1 + x_1^2 + x_2^2 + \lambda(200 - x_1 - x_2).$$

Найдем частные производные функции F по переменным x_1, x_2 и λ и приравниваем их к нулю. Получим следующую систему уравнений:

$$\begin{cases} \dfrac{\partial F}{\partial x_1} = 2x_1 + 4 - \lambda = 0; \\ \dfrac{\partial F}{\partial x_2} = 2x_2 - \lambda = 0; \\ \dfrac{\partial F}{\partial \lambda} = 200 - x_1 - x_2 = 0. \end{cases}$$

Решая систему уравнений, найдем

$$x_1 = 99,\ x_2 = 101,\ \lambda = 202,\ f(x_1, x_2) = 20398.$$

Определитель, составленный из вторых частных производных функций f по x_1, x_2, имеет вид:

$$\begin{vmatrix} \dfrac{\partial^2 f}{\partial x_1 \partial x_1} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2 \partial x_2} \end{vmatrix} = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 > 0 .$$

Следовательно, по теореме о достаточном условии существования условного экстремума функция f в точке с координатами $x_1 = 99$, $x_2 = 101$ действительно имеет экстремум

$$\frac{\partial^2 f}{\partial x_1^2} = 2 > 0 ,$$

следовательно в этой точке функция f имеет условный экстремум.

Таким образом, для получения минимальных расходов, нужно реализовать 99 автомобилей через магазин и 101 автомобиль через торговых агентов. При этом расходы на реализацию составят **20398** у.е.

Данную задачу можно было решить и графическим методом (рис. 2.33).

Областью допустимых решений задачи является отрезок AB, линиями уровня функции $f = x_1^2 + 4x_1 + x_2^2 = (x_1 + 2)^2 + x_2^2 - 4$ являются концентрические окружности $(x_1 + 2)^2 + x_2^2 = C$ с центром в точке $x_1 = -2$, $x_2 = 0$ и радиусом $\sqrt{C}$.

Из рис. 2.33 видно, что минимальное значение функции, принадлежащее области допустимых решений, достигается в точке E, в которой совпадают угловой коэффициент прямой $x_2 = 200 - x_1$ и касательной к окружности $(x_1 + 2)^2 + x_2^2 = C$ к оси Ox_1.

Продифференцировав последнее уравнение по x_1, получим

$$2(x_1 + 2) + 2x_2 x_2' = 0 , \quad x_2' = -\frac{x_1 + 2}{x_2} .$$

Приравняем последнее выражение значению углового коэффициента прямой и добавим к этому уравнению уравнение прямой, которой принадлежит точка E.

$$\begin{cases} -\dfrac{x_1 + 2}{x_2} = -1, \\ x_1 + x_2 = 200 . \end{cases}$$

Решив последнюю систему уравнений, найдем экстремальное значение целевой функции и его координаты: $\boxed{f(x_1, x_2) = 20398}$, $x_1 = 99$, $x_2 = 101$, $\lambda = 202$.

Задача решена.

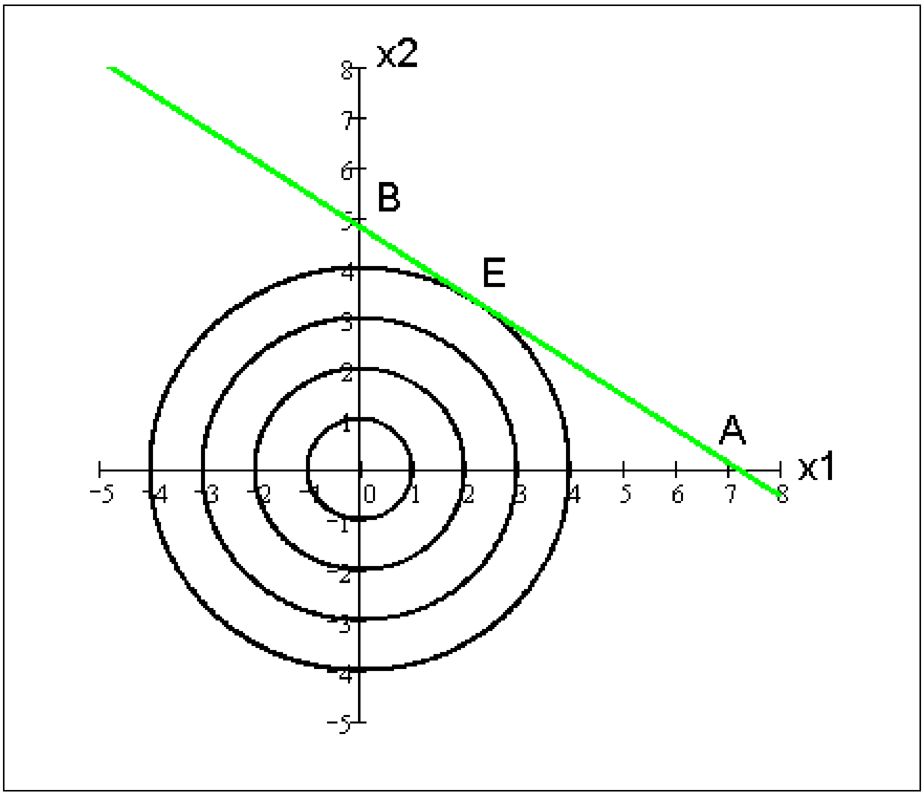


Рис. 2.33

2.5 Задачи динамического программирования

Динамическое программирование можно рассматривать как специальный вычислительный метод решения задач оптимизации управления динамическими системами, позволяющий представить процесс оптимизации в виде последовательности отдельных этапов (шагов).

Динамическое программирование – это вычислительный метод для решения задач определенной структуры. Динамическое программирование возникло и сформировалось в 1950-1953 гг. благодаря работам американского учёного ***Р. Беллмана*** и его сотрудников. Первые задачи, которые привели к появлению вычислительного метода динамического программирования (МДП), являлись динамическими задачами управления запасами.

Основу метода динамического программирования составляет **принцип оптимальности**, утверждающий, что *каков бы ни был путь достижения некоторого состояния системы, последующие решения должны принадлежать оптимальной траектории для оставшейся части пути, начинающейся с этого состояния*. Применение этого принципа на практике позволяет получить все используемые в динамическом программировании функциональные рекуррентные соотношения.

Сущность вычислительного метода рассмотрим на следующем примере. Пусть требуется найти

$$\boxed{\max_{x_1,\ldots,x_n} z = \sum_{i=1}^{n} f_i(x_i)} \qquad (2.56)$$

при условиях

$$\sum_{j=1}^{n} a_j x_j \le b; \quad a_j > 0, \quad x_j \ge 0\,. \tag{2.57}$$

Целевая функция задачи является суммой функций от одной переменной. Такая функция называется *аддитивной*. Если все $f_i(x_i)$ - выпуклые (вогнутые), то для решения может быть применен *метод множителей Лагранжа*. Однако, если функция является многоэкстремальной (содержит несколько локальных максимумов), то известный метод дает лишь одно из таких решений. Для нахождения глобального максимума классический метод множителей Лагранжа неприменим.

Рассмотрим метод, обеспечивающий решение исходной задачи. Считаем все $\{a_j\}$, $j = 1,\dots,n$ и b целыми числами. Предположим также, что в задаче все переменные $\{x_i\}\,(i = n)$ могут принимать только целочисленные значения.

Введем следующие значения. Через z^* обозначим абсолютный максимум z при условии $\sum_{j=1}^{n} a_j x_j \le b$. Выбираем значения x_n и, зафиксировав его, максимизируем z по всем остальным переменным $x_1,\dots,x_{n-1}$. Предположим, что такая максимизация проведена для всех возможных значений x_n. Тогда z^* будет наибольшим из всех возможных значений z. Формально этот процесс записывается так:

$$\max_{x_1,\dots,x_{n-1}}\left(\sum_{j=1}^{n} f_j(x_j)\right) = f_n(x_n) + \max_{x_1,\dots,x_{n-1}}\left(\sum_{j=1}^{n-1} f_j(x_j)\right), \tag{2.58}$$

причем

$$\sum_{j=1}^{n-1} a_j x_j \le b - a_n x_n\,.$$

Учитывая, что $\sum_{j=1}^{n-1} f_j(x_j)$ для неотрицательных целых чисел, удовлетворяющих условию $\sum_{j=1}^{n-1} a_j x_j \le b - a_n x_n$, зависит от ($b - a_n x_n$), введём обозначение

$$\max_{x_1,\dots,x_{n-1}}\left(\sum_{j=1}^{n-1} f_j(x_j)\right) = \Lambda_{n-1}(b - a_n x_n). \tag{2.59}$$

Допустим, что мы вычислили $\Lambda_{n-1}(b - a_n x_n)$ для всех допустимых значений $x_n = \left\{0,1,\dots,\left[\dfrac{b}{a_n}\right]\right\}$, где $\left[\dfrac{b}{a_n}\right]$ обозначает целую часть $\dfrac{b}{a_n}$. Очевидно, что

$$z^* = \max_{x_n \ge 0}\left[f_n(x_n) + \Lambda_{n-1}(b - a_n x_n)\right]. \tag{2.60}$$

Для вычисления (2.60) определяем значения $f_n(x_n)\Lambda_{n-1}(b - a_n x_n)$ для всех допустимых значений x_n и выбираем максимальное. Одновременно находим и x_n^*. Таким образом, если бы была известна функция $\Lambda_{n-1}(b - a_n x_n)$, то вся задача свелась бы к задаче с одной переменной.

Покажем, как можно вычислять $\Lambda_{n-1}(b - a_n x_n)$. Очевидно, что

$$\Lambda_{n-1}(\xi)=\max_{x_1,\ldots,x_{n-1}}\left(\sum_{j=1}^{n-1} f_j(x_j); \sum_{j=1}^{n-1} a_j x_j\right)\le \xi\ .$$

Рассуждая аналогично, получаем

$$\Lambda_{n-1}(\xi)=\max_{x_{n-1}}\left(f_{n-1}(x_{n-1})+\Lambda_{n-2}(\xi-a_{n-1}x_{n-1})\right), \qquad (2.61)$$

где

$$\Lambda_{n-2}(\xi-a_{n-1}x_{n-1})=\max_{x_1,\ldots,x_{n-1}}\left(\sum_{j=1}^{n-2} f_j(x_j)\right), \qquad (2.62)$$

причем максимум вычисляем по всем неотрицательным целым $x_1,\ldots,x_{n-2}$, удовлетворяющим условию $\sum_{j=1}^{n-2} a_j x_j \le \xi - a_{n-1}x_{n-1}$. Далее вычисляем $\Lambda_{n-2}(\xi), \Lambda_{n-3}(\xi)$ и т. д., пока на последнем шаге не придём к результату

$$\Lambda_1(\xi)=\max_{0\le x_1\le\left[\frac{\xi}{a_1}\right]} f_1(x_1). \qquad (2.63)$$

Чтобы решить задачу, процесс вычислений необходимо вести *в обратном порядке*, начиная с $\Lambda_1(\xi)$. Зафиксировав начало интервала и изменив верхний его конец ξ, вычисляем

$$\Lambda_1(\xi)=\max f_1(x_1), \quad 0\le x_1 \le \left[\frac{\xi}{a_1}\right]$$

для всех значений $\xi = 0, 1, ..., b$.

Оптимальное решение на первом шаге обозначим через $\hat{x}_1(\xi)$. Строим таблицу динамического программирования следующего вида (табл. 5.1). Вычислив $\Lambda_1(\xi)$, найдем $\Lambda_2(\xi)$, используя соотношение

$$\Lambda_2(\xi)=\max_{0\le x_2\le\left[\frac{\xi}{a_2}\right]}\left(f_2(x_2)+\Lambda_1(\xi-a_2x_2)\right).$$

Далее вычислим последовательно $\Lambda_2(\xi)$ для всех $\xi = 0, 1,..., b$, используя при этом результаты табл. 2.33.

Таблица 2.33

ξ	$\Lambda_1(\xi)$	$x_1(\xi)$
0		
1		
…	…	…
b		

Обозначим $\varphi_2(0;\xi)=f_2(0)+\Lambda_1(\xi)$. Тогда получим

$$\varphi_2(1;\xi)=f_2(1)+\Lambda_1(\xi-a_2),\ldots,\quad \varphi_2\left(\frac{\xi}{a_2};\xi\right)=f_2\left(\frac{\xi}{a_2}\right)+\Lambda_1\left(\xi-a_2\left[\frac{\xi}{a_2}\right]\right).$$

Наибольшее из этих чисел и есть $\Lambda_2(\xi)$. Одновременно находим $x_2(\xi)$. Затем строим таблицу, аналогичную табл. 2.33, для $\Lambda_2(\xi)$ и $x_2(\xi)$. Так продолжается до вычисления $\Lambda_{n-1}(\xi)$ для $\xi = 0, 1, \ldots, b$. Функцию $\Lambda_n(\xi)$ табулировать не нужно, так как достаточно определить лишь $\Lambda_n(b) = z^*$. Одновременно находим и оптимальное значение для переменной $x_n(b)$.

Для определения значений $x_{n-1}^*, \dots, x_1^*$ следует использовать уже вычисленные таблицы $(n-1)$-го; $(n-2)$-го и т. д. шагов.

Из предыдущей таблицы $(n-1)$-го шага находим $x_{n-1}^{\bullet} = \hat{x}_{n-1}\left(b - a_n x_n^*\right)$.

Для этого берем $\xi = b - a_n x_n^*$. Аналогично определяем

$$x_{n-2}^{\bullet} = \hat{x}_{n-2}\left(b - a_n x_n^* - a_{n-1} x_{n-1}^*\right).$$

Как видим, динамическое программирование представляет собой направленный последовательный перебор вариантов, который приводит к нахождению глобального максимума. Для применения МДП необходимо табулировать функции $\Lambda_1(\xi), \dots, \Lambda_{n-1}(\xi)$ для всех допустимых значений ξ.

Сравним по числу необходимых операций динамического программирования с простым перебором. Для упрощения расчета примем, что все переменные равны $a_1 = a_2 = \ldots = a_n = 1$.

При простом переборе число возможных вариантов решений (при условии целочисленности всех переменных) равно числу способов, которыми можно закрепить b число одинаковых целей за n числом средств поражения.

Это число составляет

$$C_{n+b-1}^{b} = \frac{(n+b-1)!}{b!\,(n-1)!}.$$

Оценим число операций, требуемых для решения этой задачи методом динамического программирования.

Для вычисления $\Lambda_k(\xi)$ при фиксированном ξ необходимо провести $(\xi+1)$ вычислений. Поэтому для заполнения одной таблицы $\Lambda_k(\xi)$ необходимо проделать

$$\sum_{\xi=0}^{b}(\xi+1) = \frac{(b+1)(b+2)}{2!} \text{ операций.}$$

Следовательно, для вычисления всех функций $\Lambda_1(\xi), \dots, \Lambda_{n-1}(\xi)$ необходимо $(n-1)\dfrac{(b+1)\cdot(b+2)}{2!}$ операций.

С учетом вычислений функции $\Lambda_n(b)$ общее число операций

$$N_{\Sigma} = (n-1)\frac{(b+1)\cdot(b+2)}{2!} + (b+1) = \frac{(b+1)\cdot[(n-1)\cdot(b+2)+2]}{2!}.$$

Подведем итоги. Рассмотренную задачу можно трактовать как задачу распределения с одним ограниченным источником сырья

$$\sum_{j=1}^{n} a_j x_j \le b,$$

где x_j - количество сырья, используемое в j-м способе производства. Тогда $f_j(x_j)$ - доход от переработки j-м способом x_j единиц сырья. Поэтому $\Lambda_k(\xi)$ можно рассматривать как максимальный доход от первых k способов производства, когда общее количество сырья равно ξ единиц. Поэтому данная задача представляет собой n-шаговый процесс принятия решений, где на j-м шаге принимается решение, какое количество сырья из общего его объема следует направить на переработку по j-му способу. Как видим, структура задачи не изменяется от числа шагов, т.е. задача инвариантна относительно n. Решение для k-шаговой задачи получается из решения для $(k-1)$-шаговой задачи путем добавления k-го шага и использования результата, полученного для всех предыдущих шагов.

Следовательно, сущность динамического подхода состоит в замене решения данной n-шаговой задачи последовательностью задач: одношаговой, двухшаговой и т. д.

Выделим основные свойства задач, к которым возможно применить принцип динамического программирования:

1. Задача должна допускать интерпретацию как n-шаговый процесс принятия решений.
2. Задача должна быть определена для любого числа шагов и иметь структуру, не зависящую от их числа.
3. При рассмотрении k-шаговой задачи должно быть задано некоторое множество параметров, описывающих состояние системы, от которых зависят оптимальные значения переменных. Причем это множество не должно изменяться при увеличении числа шагов.
4. Выбор решения (управления) на k-м шаге не должен оказывать влияния на предыдущие решения, кроме необходимого пересчета переменных.

Пусть $\overline{\xi}$ - вектор параметров, описывающих состояние процесса (вектор состояния). Тогда $\Lambda_k(\overline{\xi})$ - оптимальное значение целевой функции для k-шагового процесса при условии $\xi \cdot \Lambda_k(\overline{\xi})$ будем называть функцией состояния.

Пусть $\overline{X}$ - вектор переменных, подлежащих выбору на k-м шаге. Тогда для задач, к которым можно применить метод динамического программирования, должно выполняться следующее рекуррентное соотношение:

$$\Lambda_k(\overline{\xi}) = \max_{\overline{X}_k}\left\{f(\overline{X}_k;\xi) + \Lambda_{k-1}\left[T(\overline{\xi};\overline{X}_k)\right]\right\}, \qquad (2.64)$$

где $T(\overline{\xi};\overline{X}_k)$ - вектор состояния предыдущего $(k-1)$-го шага при условиях $\overline{\xi}$ и $\overline{X}_k$.

Сформулируем теперь принцип оптимальности Р.Беллмана. Оптимальная стратегия обладает *следующим свойством*: каковы бы ни были начальное состояние ξ и начальная стратегия $\overline{X}_1$, последующие стратегии (решения) должны быть оптимальны по отношению к состоянию $T(\overline{\xi};\overline{X}_1)$, получающемуся в результате предыдущего решения.

Динамическое программирование (иначе «*динамическое планирование*») есть особый метод оптимизации решений, специально приспособленный к так называемым «многошаговым» (или «многоэтапным») операциям.

Представим себе некоторую операцию Q, распадающуюся на ряд последовательных «шагов» или «этапов», - например, деятельность предприятия в течение ряда лет; разработка бизнес плана фирмы; последовательность тестов, применяемых при подготовке студентов к Интернет-тестированию. Некоторые операции (подобно вышеприведенным) расчленяются на шаги естественно; в некоторых разделение (декомпозицию) приходится вводить искусственно - скажем, процесс инвестиционного анализа инновационного проекта можно условно разбить на этапы, каждый из которых занимает какое-то время Δt.

Итак, рассмотрим операцию Q, состоящую из m шагов (этапов). Пусть эффективность операции характеризуется каким-то показателем W («выигрыш»). Предположим, что выигрыш W за всю операцию складывается из выигрышей на отдельных шагах:

$$W = \sum_{i=1}^{m} w_i , \tag{2.65}$$

где w - выигрыш на i-м шаге.

Если показатель W обладает таким свойством, то его называют сепарабельной критериальной функцией аддитивного типа.

Операция Q, о которой идет речь, представляет собой управляемый процесс, т. е. мы можем выбирать какие-то параметры, влияющие на его ход и исход, причем на каждом шаге выбирается какое-то решение, от которого зависит выигрыш на данном шаге и выигрыш за операцию в целом. Будем называть это решение «шаговым управлением». Совокупность всех шаговых управлений представляет собой управление операцией в целом. Обозначим его буквой x, а шаговые управления - буквами $x_1, x_2, \ldots, x_m$:

$$x = (x_1, x_2, \ldots, x_m) \tag{2.66}$$

Требуется найти такое управление x, при котором выигрыш W обращается в максимум:

$$W = \sum_{i=1}^{m} w_i \to \max \tag{2.67}$$

То управление x^*, при котором этот максимум достигается, будем называть *оптимальным управлением*. Такое управление состоит из совокупности оптимальных шаговых управлений:

$$x^* = (x_1^*, x_2^*, \ldots, x_m^*). \tag{2.68}$$

Тот максимальный выигрыш, который достигается при этом управлении обозначается W^*:

$$W^* = \max_x \{W(x)\}. \tag{2.69}$$

Рассмотрим несколько примеров многошаговых операций и для каждого из них поясним, что понимается под «управлением» и каков «выигрыш» (показатель эффективности) W.

Пример 2.21

Планируется деятельность группы промышленных предприятий $П_1, П_2, \ldots П_к$ на период m хозяйственных лет (m-летку). В начале периода на развитие группы выделены какие-то средства M, которые должны быть как-то распределены между предприятиями. В процессе работы предприятия вложенные в него средства частично расходуются (амортизируются), а частично сохраняются и снова могут быть перераспределены. Каждое предприятие за год приносит доход, зависящий от того, сколько средств в него вложено. В начале каждого хозяйственного года имеющиеся в наличии средства перераспределяются между предприятиями. Ставится вопрос: какое количество средств в начале каждого года нужно выделять каждому предприятию, чтобы суммарный доход за *m* лет был максимальным?

Выигрыш W (суммарный доход) представляет собой сумму доходов на отдельных шагах (годах):

$$W = \sum_{i=1}^{m} w_i \tag{2.70}$$

и, следовательно, обладает свойством аддитивности.

Управление x_i на i-м шаге состоит в том, что в начале i-го года предприятиям выделяются какие-то средства $x_{i1}, x_{i2}, \ldots, x_{ik}$ (первый индекс - номер шага, второй - номер предприятия). Таким образом, шаговое управление есть вектор с k составляющими. Величины w_i в формуле (2.70) зависят от количества вложенных в предприятия средств.

Управление x всей операцией состоит из совокупности всех шаговых управлений: $x = (x_1, x_2, \ldots, x_m)$.

Требуется найти такое распределение средств по предприятиям и по годам (оптимальное управление x^*), при котором величина W обращается в максимум.

В этом примере шаговые управления были векторами; в последующих примерах они будут проще и выражаться просто числами.

Пример 2.22

Владелец автомашины эксплуатирует ее в течение m лет. В начале каждого года он может принять одно из трех решений:

1) продать машину и заменить ее новой;
2) ремонтировать ее и продолжать эксплуатацию;
3) продолжать эксплуатацию без ремонта.

Шаговое управление - выбор одного из этих трех решений. Непосредственно числами они не выражаются, но можно приписать первому численное значение 1, второму 2, третьему 3. Какие нужно принять решения по годам (т. е. как чередовать управления 1,2,3), чтобы суммарные расходы на эксплуатацию, ремонт и приобретение новых машин были минимальны?

Показатель эффективности (в данном случае это не «выигрыш», а «проигрыш») равен

$$W = \sum_{i=1}^{m} w_i\,, \tag{2.71}$$

где w_i - расходы в i-м году. Величину W требуется обратить в минимум.

Управление операцией в целом представляет собой какую-то комбинацию чисел 1, 2, 3, например:

$$x = (3,3,2,2,2,1,3,\ldots),$$

что означает: первые два года эксплуатировать машину без ремонта, последующие три года ее ремонтировать, в начале шестого года продать, купить новую, затем снова эксплуатировать без ремонта и т. д. Любое управление представляет собой вектор (совокупность чисел):

$$x = (j_1, j_2, \dots, j_m), \tag{2.72}$$

где каждое из чисел $j_1, j_2, \dots, j_m$ имеет одно из трех значений: 1, 2 или 3. Нужно выбрать совокупность чисел (2.72), при которой величина (2.71) минимальна.

Пример 2.23

Прокладывается участок железнодорожного пути между пунктами А и B (рис. 2.33). Местность пересеченная, включает лесистые зоны, холмы, болота, реку, через которую надо строить мост. Требуется так провести дорогу из A в B, чтобы суммарные затраты на сооружение участка были минимальны.

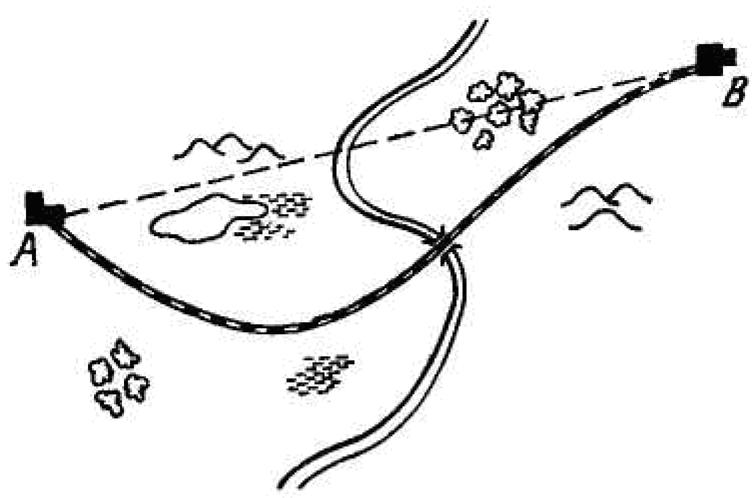


Рис. 2.33

В этой задаче нет естественного разделения процесса на шаги: его приходится вводить искусственно, для чего, например, можно отрезок AB разделить на m частей, провести через точки деления прямые, перпендикулярные AB, и считать за «шаг» переход с одной такой прямой на другую. Если провести их достаточно близко друг от друга, то можно считать на каждом шаге участок пути прямолинейным. Шаговое управление на i-м шаге представляет собой угол φ_i, который составляет участок пути с прямой AB. Управление всей операцией состоит из совокупности шаговых управлений:

$$x = (\varphi_1, \varphi_2, \dots, \varphi_m).$$

Требуется выбрать такое (оптимальное) управление x^*, при котором суммарные затраты на сооружение всех участков минимальны:

$$W = \sum_{i=1}^{m} w_i \to \min .$$

Любую многошаговую задачу можно решать по-разному: либо искать сразу все элементы решения на всех m шагах, либо же строить оптимальное управление шаг за шагом, па каждом этапе расчета оптимизируя только один шаг. Обычно второй способ оптимизации оказывается проще, чем первый, особенно при большом числе шагов.

Такая идея постепенной, пошаговой оптимизации и лежит в основе метода динамического программирования. Оптимизация одного шага, как правило, проще оптимизации всего процесса: на практике лучше много раз решить сравнительно простую задачу, чем один раз - сложную.

Однако, планируя многошаговую операцию, надо выбирать управление на каждом шаге с учетом всех его будущих последствий на еще предстоящих шагах. Управление на i-м шаге выбирается не так, чтобы выигрыш именно на данном шаге был максимален, а так, чтобы была максимальна сумма выигрышей на всех оставшихся до конца шагах плюс данный.

Процесс динамического программирования обычно разворачивается от конца к началу: прежде всего планируется последний, m-й шаг. А как его спланировать, если неизвестно, чем кончился предпоследний? Другими словами, нет сведений об условиях, в которых приступают к последнему шагу.

Планируя последний шаг, нужно сделать разные предположения о том, чем кончился предпоследний, $(m-1)$-й шаг, и для каждого из этих предположений найти условное оптимальное управление на m-м шаге («условное» потому, что оно выбирается исходя из условия, что предпоследний шаг кончился так-то и так-то).

Предположим, что мы это сделали, и для каждого из возможных исходов предпоследнего шага знаем условное оптимальное управление и соответствующий ему условный оптимальный выигрыш на m-м шаге. Теперь мы можем оптимизировать управление на предпоследнем, $(m-1)$-м шаге. Снова сделаем все возможные предположения о том, чем кончился предыдущий, $(m-2)$-й шаг, и для каждого из этих предположений найдем такое управление на $(m-1)$-м шаге, при котором выигрыш за последние два шага (из которых m-й уже оптимизирован!) максимален. Так мы находится для каждого исхода $(m-2)$-го шага условное оптимальное управление на $(m-1)$-м шаге и условный оптимальный выигрыш на двух последних шагах. Далее, «пятясь назад», оптимизируем управление на $(m-2)$-м шаге и т. д., пока не дойдем до первого.

Предположим, что все условные оптимальные управления и условные оптимальные выигрыши за весь «хвост» процесса (на всех шагах, начиная от данного и до конца) известны. Это значит: мы знаем, что надо делать, как управлять на данном шаге и что мы за это получим на «хвосте», в каком бы состоянии пи был процесс к началу шага. Теперь можем построить уже не условно оптимальное, а просто оптимальное управление x^* и найти не условно оптимальный, а просто оптимальный выигрыш W^*.

Пусть известно, в каком состоянии S_0 была управляемая система (объект управления S) в начале первого шага. Тогда можно выбрать оптимальное управление x_1^* на первом шаге. Применив его, изменим состояние системы на некоторое новое S_1^*, в этом состоянии мы подошли ко второму шагу. Тогда нам тоже известно условное оптимальное управление x_2^*, которое к концу второго шага переводит систему в состояние S_2^*,- и т. д. Что касается оптимального выигрыша W^* за всю операцию, то он нам уже известен: ведь именно на основе его максимальности мы выбирали управление на первом шаге.

Таким образом, в процессе оптимизации управления многошаговый процесс «проходится» дважды: первый раз - от конца к началу, в результате чего находятся условные оптимальные управления и условные оптимальные выигрыши за оставшийся «хвост» процесса; второй раз - от начала к концу; при этом остается только «прочитать» уже готовые рекомендации и найти безусловное оптимальное управление x^*, состоящее из оптимальных шаговых управлении $x_1^*, x_2^*, \ldots, x_m^*$.

Первый этап - условной оптимизации - несравненно сложнее и длительнее второго. Второй этап почти не требует дополнительных вычислений.

Рассмотрим примеры решения типовых задач динамического программирования.

Пример 2.24

Прокладка наивыгоднейшего пути между двумя пунктами. Рассмотрим пример 2.23 в упрощенных условиях. Нужно соорудить путь, соединяющий два пункта A и B, из которых второй лежит к северо-востоку от первого.

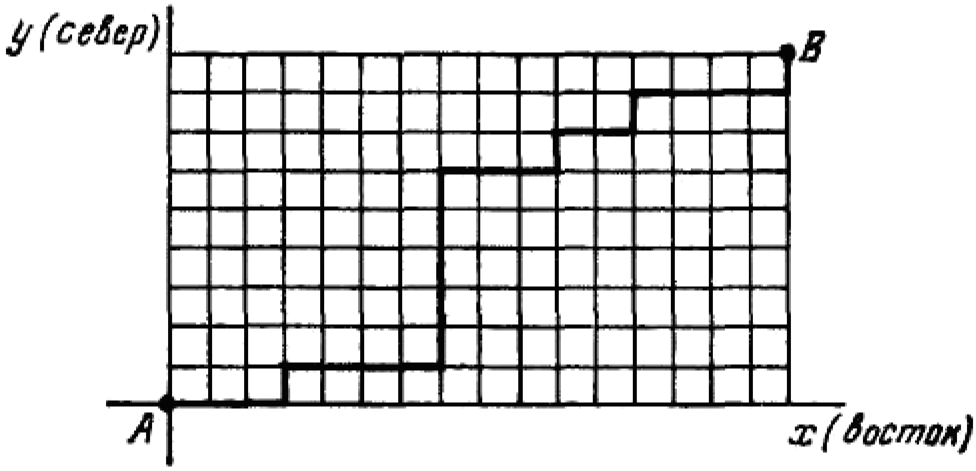


Рис. 2.34

Для простоты допустим, что прокладка пути состоит из ряда шагов, и на каждом шаге мы можем двигаться либо строго на восток, либо строго на север; любой путь из A в B представляет собой ступенчатую ломаную линию, отрезки которой параллельны одной из координатных осей (рис. 2.34). Затраты на сооружение каждого из таких отрезков известны. Требуется проложить такой путь из A в B, при котором суммарные затраты минимальны.

Решение

Решение можно производить одним из двух способов: либо перебрать все возможные варианты пути, и выбрать тот, на котором затраты минимальны (а при большом числе отрезков это выполнить трудно); либо разделить процесс перехода из A в B на отдельные шаги (один шаг - один отрезок) и оптимизировать управление по шагам. Второй способ, безусловно, более удобен. Тут сказываются преимущества целенаправленного, организованного поиска решения перед «слепым» перебором.

Продемонстрируем, как это делается, на конкретном примере. Разделим расстояние от A до B в восточном направлении, скажем, на 7 частей, а в северном - на 5 частей (в принципе дробление может быть сколь угодно мелким). Тогда любой путь из A в B состоит из $m = 7 + 5 = 12$ отрезков, направленных на восток или на север (рис. 2.35). Проставим на каждом из отрезков число, выражающее (в каких-то условных единицах) стоимость прокладки пути по этому отрезку. Требуется выбрать такой путь из A в B, для которого сумма чисел, стоящих на отрезках, минимальна.

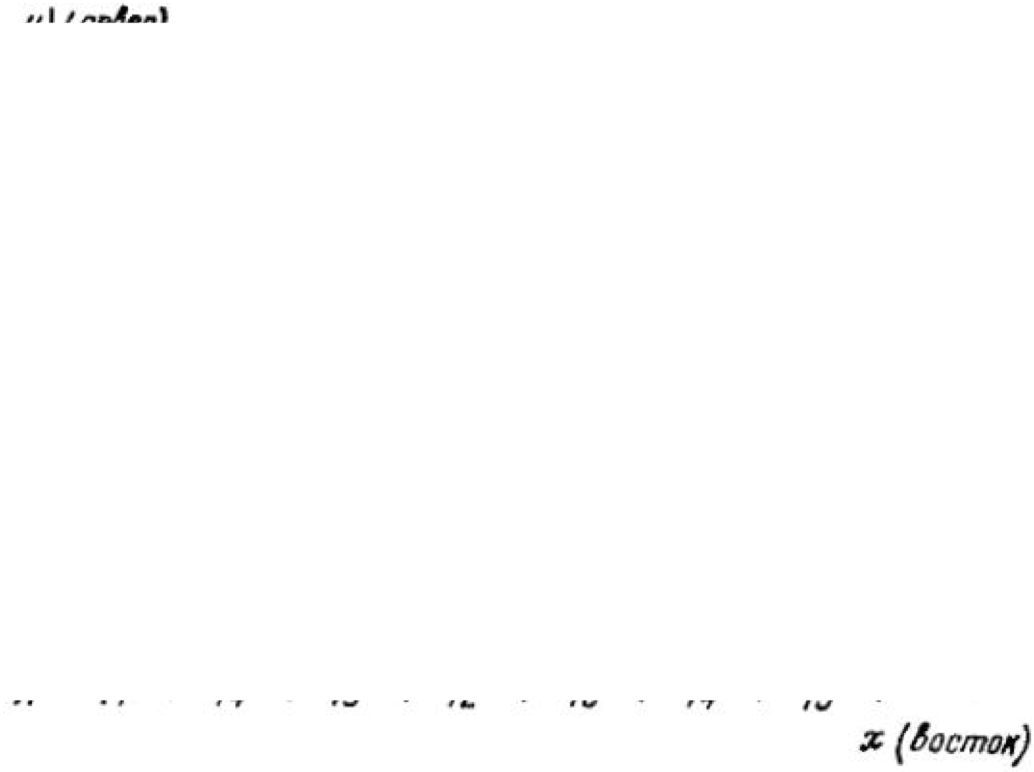


Рис. 2.35

Будем рассматривать сооружаемый путь как управляемую систему S, перемещающуюся под влиянием управления из начального состояния A в конечное B. Состояние этой системы перед началом каждого шага будет характеризоваться двумя координатами: восточной (x) и северной (y), обе - целочисленные ($0 \le x \le 7, 0 \le y \le 5$). Для каждого из состояний системы (узловой точки прямоугольной сетки на рис. 2.35) необходимо найти условное оптимальное управление: идти из этой точки на север (управление «c») или на восток (управление «в»). Выбирается это управление так, чтобы стоимость всех оставшихся до конца шагов (включая данный) была минимальна. Эту стоимость мы по-прежнему будем называть «условным оптимальным выигрышем» (хотя в данном случае это не «выигрыш», а «проигрыш») для данного состояния системы S перед началом очередного шага.

Процедуру условной оптимизации будем разворачивать в обратном направлении - от конца к началу. Прежде всего, произведем условную оптимизацию последнего, 12-го шага. Рассмотрим отдельно правый верхний угол прямоугольной сетки (рис. 2.36).

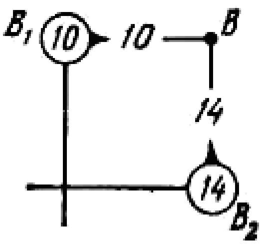


Рис. 2.36

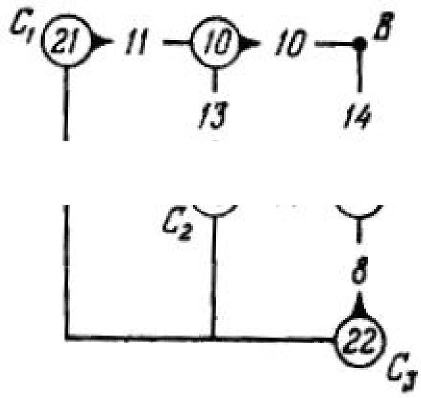


Рис. 2.37

Где мы можем находиться после 11-го шага? Только там, откуда за один (последний) шаг можно попасть в B, т. е. в одной из точек B_1 или B_2. Если мы находимся в точке B_1, у нас нет выбора (управление вынужденное): надо идти на восток, и это обойдется нам в 10 единиц. Запишем это число 10 в кружке у точки B_1, а оптимальное управление покажем короткой стрелкой, исходящей из B_1 и направленной на восток. Для точки B_2 управление тоже вынужденное (север), расход до конца равен 14, мы его запишем в кружке у точки B_2. Таким образом, условная оптимизация последнего шага сделана, и условный оптимальный выигрыш для каждой из точек B_1, B_2 найден и записан в соответствующем кружке.

Теперь оптимизируем предпоследний (11-й) шаг. После 10-го шага мы могли оказаться в одной из точек C_1, C_2, C_3 (рис. 2.37). Найдем для каждой из них условное оптимальное управление и условный оптимальный выигрыш. Для точки C_1 управление вынужденное: идти на восток; обойдется это до конца в 21 единицу (11 на данном шаге, плюс 10, записанных в кружке при B_1). Число 21 записываем в кружке при точке C_1. Для точки C_2 управление уже не вынужденное: мы можем идти как на восток, так и на север. В первом случае мы затратим на данном шаге 14 единиц и от B_2 до конца - еще 14, всего 28 единиц. Если пойдем на север, затратим 13 + 10, всего 23 единицы. Значит, условное оптимальное управление в точке C_2 - идти на север (отмечаем это стрелкой, а число 23 записываем в кружке у C_2). Для точки C_3 управление снова вынужденное («с»), обойдется это до конца в 22 единицы (ставим стрелку на север, число 22 записываем в кружке при C_3).

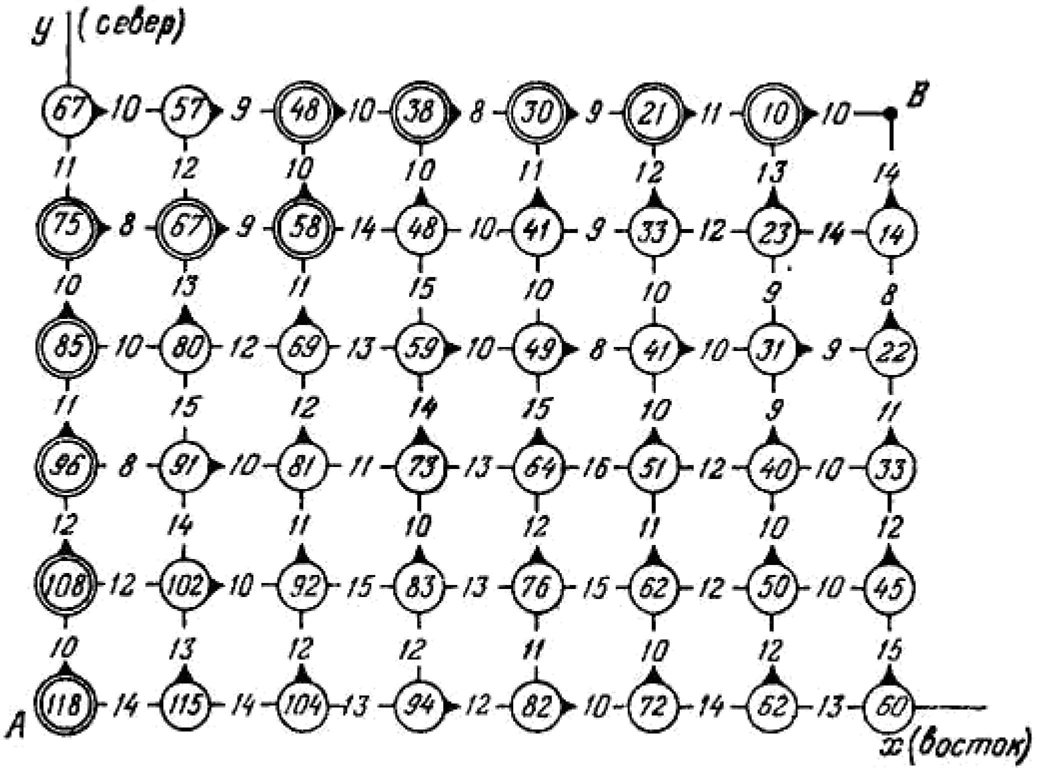


Рис. 2.38

Аналогично, «пятясь» от предпоследнего шага назад, найдем для каждой точки с целочисленными координатами условное оптимальное управление («с» или «в»), которое обозначим стрелкой, и условный оптимальный выигрыш (расход до конца пути), который запишем в кружке. Вычисляется он так: расход на данном шаге складывается с уже оптимизированным расходом, записанным в кружке, куда ведет стрелка. Таким образом, на каждом шаге мы оптимизируем только этот шаг, а следующие за ним - уже оптимизированы. Конечный результат процедуры оптимизации показан на рис. 2.38.

Таким образом, условная оптимизация уже выполнена: в какой бы из узловых точек мы ни находились, мы уже знаем, куда идти (стрелка) и во что нам обойдется путь до конца (число в кружке). В кружке при точке A записан оптимальный выигрыш на все сооружение пути из A в B: $W^* = 118$.

Теперь остается построить безусловное оптимальное управление — траекторию, ведущую из A и B самым дешевым способом. Для этого нужно только «слушаться стрелок», т. е. прочитать, что они предписывают делать на каждом шаге. Такая оптимальная траектория отмечена на рис. 2.38 дважды обведенными кружками. Соответствующее безусловное оптимальное управление будет определяться моделью $x^* = (c, c, c, c, в, в, c, в, в, в, в)$, т. е. первые четыре шага мы должны делать на север, следующие два - на восток, затем опять один на север и остальные пять - на восток.

Задача решена.

Заметим, что в ходе условной оптимизации мы можем столкнуться со случаем, когда оба управления для какой-то точки на плоскости являются оптимальными, т. е. приводят к одинаковому расходу средств от этой точки до конца, Например, в точке с координатами (5; 1) оба управления «с» и «в» являются оптимальными и дают расход до конца равным 62. Из них мы произвольно выбираем любое (в нашем случае мы выбрали «с»; с тем же успехом мы могли бы выбрать «в»). Такие случаи неоднозначного выбора оптимального управления постоянно встречаются в динамическом программировании; в дальнейшем мы специально отмечать их не будем, а попросту выберем произвольно любой из равноценных вариантов. От этого произвола, разумеется, может зависеть оптимальное управление всем процессом, но не оптимальный выигрыш. Вообще, в задачах динамического программирования (как и в задачах линейного) решение далеко не всегда единственное.

Пример 2.25

Задача о распределении ресурсов. В нашем распоряжении имеется какой-то запас средств (ресурсов) K, который должен быть распределен между m предприятиями $\Pi_1, \Pi_2, \ldots, \Pi_m$. Каждое из предприятий Π_i при вложении в него каких-то средств x приносит доход, зависящий от x, т. е. представляющий собой какую-то функцию $\varphi_i(x)$. Все функции $\varphi_i(x)$ ($i = 1,2,\ldots,m$) заданы (разумеется, эти функции - не убывающие). Нужно распределить средства K между предприятиями, чтобы в сумме они дали максимальный доход.

Решение задачи

Данная задача решается методом динамического программирования. Хотя в своей постановке она не содержит упоминания о времени, можно все же операцию распределения средств мысленно развернуть в какой-то последовательности, считая за первый шаг вложение средств в предприятие Π_1, за второй - в Π_2 и т. д. Управляемая система S в данном случае - средства или ресурсы, которые распределяются. Состояние системы S перед каждым шагом характеризуется одним числом S - наличным запасом еще не вложенных средств. В этой задаче «шаговыми управлениями» являются средства $x_1, x_2, \ldots, x_m$, выделяемые предприятиям. Требуется найти оптимальное управление, т. е. такую совокупность чисел $x_1, x_2, \ldots, x_m$, при которой суммарный доход максимален:

$$W = \sum_{i=1}^{m} \varphi_i(x_i) \to \max .$$

Решим эту задачу сначала в общем, формульном виде, а потом - для конкретных числовых данных. Найдем для каждого i-го шага условный оптимальный выигрыш (от этого шага и до конца), если мы подошли к данному шагу с запасом средств S. Обозначим условный оптимальный выигрыш $W_i(S)$, а соответствующее ему условное оптимальное управление — средства, вкладываемые в i-е предприятие, - $x_i(S)$.

Начнем оптимизацию с последнего, m-го шага. Пусть мы подошли к этому шагу с остатком средств S. Очевидно, вложить всю сумму S целиком в предприятие Π_m. Поэтому условное оптимальное управление на m-м шаге: отдать последнему предприятию все имеющиеся средства S, т. е. $x_m(S)=S$, а условный оптимальный выигрыш $W_m(S)=\varphi_m(S)$.

Задаваясь набором значений S (располагая их достаточно тесно), мы для каждого значения S будем знать $x_m(S)$ и $W_m(S)$. Последний шаг оптимизирован.

Перейдем к предпоследнему, $(m-1)$-му шагу. Пусть мы подошли к нему с запасом средств S. Обозначим $W_{m-1}(S)$ условный оптимальный выигрыш на двух последних шагах: $(m-1)$-м и m-м (который уже оптимизирован). Если мы выделим на $(m-1)$-м шаге $(m-1)$-му предприятию средства x, то на последний шаг останется $S-x$. Наш выигрыш на двух последних шагах будет равен

$$\varphi_{m-1}(x)+W_m(S-x),$$

и нужно найти такое x, при котором этот выигрыш максимален:

$$W_{m-1}(S)=\max_{x\le S}\{\varphi_{m-1}(x)+W_m(S-x)\}.$$

Этот максимум и есть условный оптимальный выигрыш за два последних шага, а то значение x, при котором этот максимум достигается, - условное оптимальное управление на $(m-1)$-м шаге.

Далее оптимизируем $(m-2)$-й, $(m-3)$-й и т. д. шаги. Вообще, для любого i-го шага будем находить условный оптимальный выигрыш за все шаги с этого и до конца по формуле

$$W_i(S)=\max_{x\le S}\{\varphi_i(x)+W_{i+1}(S-x)\}$$

и соответствующее ему условное оптимальное управление $x_i(S)$ - то значение x, при котором этот максимум достигается.

Продолжая таким образом, дойдем, наконец, до 1-го предприятия $П_1$. Здесь нам не нужно будет варьировать значения S; мы точно знаем, что запас средств перед первым шагом равен K:

$$W^* = W_1(K)=\max_{x\le K}\{\varphi_1(x)+W_2(K-x)\} \tag{2.73}$$

Итак, максимальный выигрыш (доход) от всех предприятий найден. Значение x, при котором достигается максимум (2.73), и есть оптимальное управление x_1^* на 1-м шаге. После того как мы вложим эти средства в 1-е предприятие, у нас, их останется $K-x_1^*$. Выделяем второму предприятию оптимальное количество средств: $x_2^*=x_2(K-x_1^*)$, и т. д. до конца.

Рассмотрим численный пример. Исходный запас средств $K=10$ (условных единиц), и требуется его оптимальным образом распределить между пятью предприятиями ($m=5$). Для простоты предположим, что вкладываются только целые количества средств. Функции дохода $\varphi_i(x)$ заданы в таблице 2.34.

Таблица 2.34

x	$\varphi_1(x)$	$\varphi_2(x)$	$\varphi_3(x)$	$\varphi_4(x)$	$\varphi_5(x)$
1	0.5	0.1	0.6	0.3	1.0
2	1.0	0.5	1.1	0.6	1.2
3	1.4	1.2	1.2	1.3	1.3

4	2.0	1.8	1.4	1.4	1.3
5	2.5	2.5	1.6	1.5	1.3
6	2.8	2.9	1.7	1.5	1.3
7	3.0	3.5	1.8	1.5	1.3
8	3.0	3.5	1.8	1.5	1.3

В каждом столбце, начиная с какой-то суммы вложений, доходы перестают возрастать (реально это соответствует тому, что каждое предприятие способно «освоить» лишь ограниченное количество средств).

Произведем условную оптимизацию так, как это было описано выше, начиная с последнего, 5-го шага. Каждый раз, когда мы подходим к очередному шагу, имея запас средств S, мы пробуем выделить на этот шаг то или другое количество средств, берем выигрыш на данном шаге по таблице 2.34, складываем с уже оптимизированным выигрышем на всех последующих шагах до конца (учитывая, что средств у нас осталось уже меньше, как раз на такое количество средств, которое мы выделили) и находим то вложение, на котором эта сумма достигает максимума. Это вложение и есть условное оптимальное управление на данном шаге, а сам максимум - условный оптимальный выигрыш.

В таблице 2.35 даны результаты условной оптимизации по всем шагам. Таблица построена так: в первом столбце даются значения запаса средств S, с которым мы подходим к данному шагу. Далее таблица разделена на пять пар столбцов, соответственно номеру шага. В первом столбце каждой пары приводится значение условного оптимального управления, во втором - условного оптимального выигрыша. Таблица заполняется слева направо, сверху вниз. Решение на пятом - последнем - шаге вынужденное: выделяются все средства; на всех остальных шагах решение приходится оптимизировать. В результате последовательной оптимизации 5-го, 4-го, 3-го, 2-го и 1-го шагов мы получим полный список всех рекомендаций по оптимальному управлению и безусловный оптимальный выигрыш W^* за всю операцию - в данном случае он равен 5,6. В последних двух столбцах таблицы 2.34 заполнена только одна строка, так как состояние системы перед началом первого шага нам в точности известно: $S_0 = K = 10$. Оптимальные управления на всех шагах выделены рамкой. Таким образом, мы получили окончательный вывод: надо выделить первому предприятию две единицы из десяти, второму - пять единиц, третьему - две, четвертому — ни одной, пятому - одну единицу. При этом распределении доход будет максимален и равен 5,6.

Рассмотрим как заполняется таблица 2.35 на примере $x_3(7)$. Предположим, что все шаги после третьего (4-й и 5-й) уже оптимизированы, т. е. Заполнены две первые пары столбцов таблицы 2.5.2. Найдем $x_3(7)$ и $W_3(7)$. Для этого составим вспомогательную таблицу 2.35. В первом ее столбце перечислены все возможные вложения x на третьем шаге, не превосходящие $S = 7$. Во втором столбце - то, что останется после такого вложения от запаса средств $S = 7$. В третьем столбце - выигрыш на третьем шаге от вложения средств x в третье предприятие (заполняется по столбцу $\varphi_3(x)$ таблицы 2.34).

Таблица 2.35

S	$i=5$		$i=4$		$i=3$		$i=2$		$i=1$	
	$x_5(S)$	$W_5(S)$	$x_4(S)$	$W_4(S)$	$x_3(S)$	$W_3(S)$	$x_2(S)$	$W_2(S)$	$x_1(S)$	$W_1(S)$
1	$\boxed{1}$	1.0	$\boxed{0}$	1.0	0	1.0	0	1.0		
2	2	1.2	1	1.3	1	1.6	0	1.6		
3	3	1.3	2	1.6	$\boxed{2}$	2.1	0	2.1		
4	4	1.3	3	2.3	2	2.4	0	2.4		
5	5	1.3	3	2.5	1	2.9	0	2.9		
6	6	1.3	4	2.6	2	3.4	5	3.5		
7	7	1.3	5	2.7	2	3.6	5	4.1		
8	8	1.3	5	2.8	4	3.7	$\boxed{5}$	4.6		
9	9	1.3	6	2.8	5	3.9	7	5.1		
10	10	1.3	7	2.8	5	4.1	7	5.6	$\boxed{2}$	$\boxed{5.6}$

В четвертом столбце - оптимальный выигрыш на всех оставшихся шагах (четвертом и пятом) при условии, что мы подошли к четвертому шагу с оставшимися средствами (заполняется по столбцу $i=4$ таблицы 2.35). В пятом столбце - сумма двух выигрышей: шагового и оптимизированного дальнейшего при данном вложении x в третий шаг.

Таблица 2.36

x	$7-x$	$\varphi_3(x)$	$W_4(7-x)$	$\varphi_3(x)+W_4(7-x)$
7	0	1.8	0	1.8
6	1	1.7	1.0	2.7
5	2	1.6	1.3	2.9
4	3	1.4	1.6	3.0
3	4	1.2	2.3	3.5
$\boxed{2}$	5	1.1	2.5	$\boxed{3.6}$
1	6	0.6	2.6	3.2
0	7	0	2.7	2.7

Из всех выигрышей последнего столбца выбирается максимальный (табл. 2.36) $W_3(7)=3{,}6$, а соответствующее ему управление - $x_3(7)=2$.

Задача решена.

Пример 2.26

Задача о загрузке машины. Опираясь на МДП, можно с успехом решать ряд задач оптимизации, в том числе: некоторые задачи целочисленного программирования. Заметим, что целочисленность решений, затрудняющая решение классических ЗЛП, в данном случае несколько упрощает процедуру.

В качестве примера рассмотрим задачу о загрузке машины: *имеется определенный набор предметов $\Pi_1,\ldots,\Pi_n$ (каждый в единственном экземпляре); известны их веса $q_1,\ldots,q_n$ и стоимости $c_1,\ldots,c_n$.* Грузоподъемность машины равна

Q. *Требуется определить*, какие из предметов нужно взять в машину, чтобы их суммарная стоимость (при суммарном весе $\leq Q$) была максимальна?

Решение

Нетрудно заметить, что эта задача, в сущности, аналогична предыдущей (распределение ресурсов между n предприятиями). В самом деле, процесс загрузки машины представим состоящим из n шагов; на каждом шаге принимают решение: брать данный предмет в машину или нет. Управление на i-м шаге равно единице, если данный i-й предмет берут, и нулю - если не берут.

Чтобы характеризовать состояние системы перед очередным шагом, можно использовать вес S, который еще остался в нашем распоряжении до конца (полной загрузки машины) после того, как предыдущие шаги выполнены (какие-то предметы погружены в машину). Для каждого из значений S *надо* найти $W_i(S)$ - суммарную максимальную стоимость предметов, которыми можно «догрузить» машину при данном значении S, и положить $x_i(S)=1$, если данный i-й предмет берут в машину, и $x_i(S)=0$, если не берут.

Рассмотрим конкретный числовой пример: имеется шесть предметов, веса и стоимости которых указаны в табл. 2.37.

Таблица 2.37

Предмет $П_i$	$П_1$	$П_2$	$П_3$	$П_4$	$П_5$	$П_6$
Вес q_i	4	7	11	12	16	20
Стоимость c_i	7	10	15	20	27	34

Суммарная грузоподъемность машины $Q=35$ единиц веса. Требуется указать номера предметов, которые нужно включить в груз, чтобы их суммарная стоимость была максимальна.

Как и ранее, будем придавать S только целые значения. Условная оптимизация решения показана в таблице 2.38, где в каждой строке для соответствующего номера шага (номера предмета) приведены: условное оптимальное управление x_i (0 или 1) и условный оптимальный выигрыш W_i (стоимость всех оставшихся до конца предметов при оптимальном управлении на всех шагах).

В таблице 2.38 выделены: оптимальный выигрыш $W^*=57$ и оптимальные шаговые управления, при которых этот выигрыш достигается: $x_1^*=0$, $x_2^*=1$, $x_3^*=0$, $x_4^*=1$, $x_5^*=1$, $x_6^*=0$, т. е. загрузить машину надо предметами 2, 4 и 5, суммарный вес которых равен в точности 35.

В рассмотренном примере решение можно найти «простым перебором», пробуя все возможные комбинации предметов, проверяя выполнимость ограничений и выбирая ту, для которой стоимость максимальна. Однако при большом числе предметов такой подход является нерациональным: число комбинаций неумеренно растет при увеличении числа предметов.

Задача решена.

Таблица 2.38

S	$i=6$		$i=5$		$i=4$		$i=3$		$i=2$		$i=1$	
	x_i	W_i	x_i	W_i	x_i	W_i	x_i	W_i	x_i	W_i	x_i	W_i
0	0	0	0	0	0	0	0	0	0	0		
1	0	0	0	0	0	0	0	0	0	0		
2	0	0	0	0	0	0	0	0	0	0		
3	0	0	0	0	0	0	0	0	0	0		
4	0	0	0	0	0	0	0	0	0	0		
5	0	0	0	0	0	0	0	0	0	0		
6	0	0	0	0	0	0	0	0	0	0		
7	0	0	0	0	0	0	0	0	1	10		
8	0	0	0	0	0	0	0	0	1	10		
9	0	0	0	0	0	0	0	0	1	10		
10	0	0	0	0	0	0	0	0	1	10		
11	0	0	0	0	0	0	1	15	0	15		
12	0	0	0	0	1	20	0	20	0	20		
13	0	0	0	0	1	20	0	20	0	20		
14	0	0	0	0	1	20	0	20	0	20		
15	0	0	0	0	1	20	0	20	0	20		
16	0	0	1	27	0	27	0	27	0	27		
17	0	0	1	27	0	27	0	27	0	27		
18	0	0	1	27	0	27	0	27	0	27		
19	0	0	1	27	0	27	0	27	1	30		
20	1	34	0	34	0	34	0	34	0	34		
21	1	34	0	34	0	34	0	34	0	34		
22	1	34	0	34	0	34	0	34	0	34		
23	1	34	0	34	0	34	1	35	1	37		
24	1	34	0	34	0	34	1	35	1	37		
25	1	34	0	34	0	34	1	35	1	37		
26	1	34	0	34	0	34	1	35	1	37		
27	1	34	0	34	0	34	1	42	1	44		
28	1	34	0	34	1	47	0	47	0	47		
29	1	34	0	34	1	47	0	47	0	47		
30	1	34	0	34	1	47	0	47	0	47		
31	1	34	0	34	1	47	1	49	0	49		
32	1	34	0	34	1	54	0	54	0	54		
33	1	34	0	34	1	54	0	54	0	54		
34	1	34	0	34	1	54	0	54	0	54		
35	1	34	0	34	1	54	0	54	**1**	57	**0**	**57**

Общая модель задачи динамического программирования

Рассмотренные выше простейшие задачи динамического программирования дают понятие об общей идее метода: пошаговая оптимизация, проходимая в одном направлении «условно», в другом - «безусловно». Метод динамического программирования в настоящее время является мощным и плодотворным методом оптимизации управления, который в известной степени инвариантен требованию целочисленности решения, нелинейность целевой функции. Но в отличие от линейного программирования динамическое программирование не сводится к какой-либо стандартной вычислительной процедуре; оно может быть передано на машину только после того, как записаны соответствующие формулы, а это часто бывает не так-то легко.

В основе МДП лежит принцип оптимальности Беллмана, формулирующийся следующим образом: *управление на каждом шаге надо выбирать так, чтобы оптимальной была сумма выигрышей на всех оставшихся до конца процесса шагах, включая выигрыш на данном шаге*.

Поясним это правило. При решении задачи динамического программирования на каждом шаге выбирается управление, которое должно привести к оптимальному выигрышу. Если считать все шаги независимыми друг от друга, то оптимальным шаговым управлением будет то управление, которое приносит максимальный выигрыш именно на данном шаге. Но, например, при покупке новой техники взамен устаревшей на её приобретение затрачиваются определенные средства. Поэтому прибыль от её эксплуатации вначале может быть небольшой. Однако в следующие годы новая техника будет приносить большую прибыль. И наоборот, если руководитель примет решение оставить старую технику для получения прибыли в текущем году, то в дальнейшем это приведет к значительным убыткам. Данный пример демонстрирует следующий факт: в многошаговых процессах все шаги зависят друг от друга, и, следовательно, управление на каждом конкретном шаге надо выбирать с учётом его будущих воздействий на весь процесс.

Другой момент, который следует учитывать при выборе управления на данном шаге, - это возможные варианты окончания предыдущего шага. Эти варианты определяют состояние процесса. Например, при определении количества средств, вкладываемых в предприятие в i-м году, необходимо знать, сколько средств осталось в наличии к этому году и какая прибыль получена в предыдущем $(i-1)$-м году. Таким образом, при выборе шагового управления необходимо учитывать: 1) возможные исходы предыдущего шага и 2) влияние управления на все оставшиеся до конца процесса шаги.

Укажем особенности составления ММ динамического программирования. Дополнительно введем следующие условные обозначения:

s - состояние процесса;

S_i - множество возможных состояний процесса перед i-м шагом;

W_i - выигрыш с i-го шага до конца процесса, $i=\overline{1,m}$.

Можно определить следующие основные этапы составления математической модели задачи динамического программирования.

1. Разбиение задачи на шаги (этапы). Шаг не должен быть слишком мелким, чтобы не проводить лишних расчётов и не должен быть слишком большим, усложняющим процесс шаговой оптимизации.

2. Выбор переменных, характеризующих состояние s моделируемого процесса перед каждым шагом, и выявление налагаемых на них ограничений. В качестве таких переменных следует брать факторы, представляющие интерес для исследователя, например, годовую прибыль при планировании деятельности предприятия.

3. Определение множества шаговых управлений x_i, $i=\overline{1,m}$ и налагаемых на них ограничений, т.е. области допустимых управлений X.

4. Определение выигрыша

$$\varphi_i(s,x_i),$$

который принесет на i-м шаге управление x_i, если система перед этим находилась в состоянии s.

5. Определение состояния s', в которое переходит система из состояния s под влиянием управления x_i,

$$s'=f_i(s,x_i), \tag{2.74}$$

где f_i - функция перехода на i-м шаге из состояния s в состояние s'.

6. Составление уравнения, определяющего условный оптимальный выигрыш на последнем шаге, для состояния s моделируемого процесса

$$W_m(s)=\max_{x_m\in X}\{\varphi_m(s,x_m)\}. \tag{2.75}$$

7. Составление основного функционального уравнения динамического программирования, определяющего условный оптимальный выигрыш для данного состояния s с i-го шага и до конца процесса через уже известный условный оптимальный выигрыш с $(i+1)$-го шага и до конца:

$$W_i(s)=\max_{x_m\in X}\{\varphi_i(s,x_i)+W_{i+1}(f_i(s,x_i))\}. \tag{2.76}$$

В уравнении (2.76) в уже известную функцию $W_{i+1}(s)$, характеризующую условный оптимальный выигрыш с $(i+1)$-го шага до конца процесса, вместо состояния s подставлено новое состояние $s'=f_i(s,x_i)$, в которое система переходит на i-м шаге под влиянием управления x_i.

Этапы решения задачи динамического программирования

После того как выполнены пункты 1-7 и математическая модель составлена, приступают к её расчёту. Укажем основные этапы решения задачи динамического программирования.

1. Определение множества возможных состояний S_m для последнего шага.

2. Проведение условной оптимизации для каждого состояния $s\in S_m$ на последнем m-м шаге по формуле (2.75) и определение условного оптимального управления $x_m(s)$, $s\in S_m$.

3. Определение множества возможных состояний S_i для i-го шага, $i=2,3,\dots,m-1$.

4. Проведение условной оптимизации i-го шага, $i = 2,3,\ldots,m-1$ для каждого состояния $s \in S_i$ по формуле (2.76) и определение условного оптимального управления $x_i(s)$, $s \in S_i$, $i = 2,3,\ldots,m-1$.

5. Определение начального состояния системы s_1, оптимального выигрыша $W_1(s_1)$ и оптимального управления $x_1(s_1)$ по формуле (2.76) при $i = 1$. Это есть оптимальный выигрыш для всей задачи $W^* = W_1(x_1^*)$.

6. Проведение безусловной оптимизации управления. Для проведения безусловной оптимизации необходимо найденное на первом шаге оптимальное управление $x_1^* = x_1(s_1)$ подставить в формулу (2.74) и определить следующее состояние системы $s_2 = f_2(s_1, x_1^*)$. Для измененного состояния найти оптимальное управление $x_2^* = x_2(s_2)$, подставить в формулу (2.74) и т. д. Для i-го состояния s_i найти $s_{i+1} = f_{i+1}(s_i, x_i^*)$ и $x_{i+1}^*(s_{i+1})$ и т. д.

Выбор оптимальной стратегии замены оборудования как задача динамического программирования

В общем виде проблема ставится следующим образом: *определить* оптимальную стратегию использования оборудования в период времени длительностью m лет, причём прибыль за каждые i лет, $i = \overline{1,m}$ от использования оборудования возраста t лет должна быть максимальной.

Известны: $r(t)$ - выручка от реализации продукции, произведенной за год на оборудовании возраста t лет, $l(t)$ - годовые затраты, зависящие от возраста оборудования t, $c(t)$ - остаточная стоимость оборудования возраста t лет, P - стоимость нового оборудования. Под возрастом оборудования понимается период эксплуатации оборудования после последней замены, выраженный в годах.

Для построения математической модели последовательно выполняются этапы, сформулированные ниже.

1. Определение числа шагов. Число шагов равно числу лет, в течение которых эксплуатируется оборудование.

2. Определение состояний системы. Состояние системы характеризуется возрастом оборудования t; $t = \overline{0,m}$.

3. Определение управлений. В начале i-го шага, $i = \overline{1,m}$ может быть выбрано одно из двух управлений: заменять или не заменять оборудование. Каждому варианту управления приписывается число

$$x_i = \begin{cases} 0, & \textit{если оборудование не заменяется;} \\ 1, & \textit{если оборудование заменяется} \end{cases}$$

4. Определение функции выигрыша на i-м шаге. Функция выигрыша на i-м шаге – это прибыль от использования оборудования к концу i-го года эксплуатации, $t = \overline{0,m}$, $i = \overline{1,m}$.

$$\varphi_i(t) = \begin{cases} r(t) - l(t), & \textit{если оборудование в начале i – го года не заменяется;} \\ c(t) - p + r(0) - l(0), & \textit{если оборудование заменяется.} \end{cases}$$

Таким образом, если оборудование не продаётся, то прибыль от его использования – это разность между стоимостью произведенной продукции и эксплуатационными издержками. При замене оборудования прибыль составляет разность между остаточной стоимостью оборудования и стоимостью нового оборудования, к которой прибавляется разность между стоимостью продукции и эксплуатационными издержками для нового оборудования, возраст которой в начале i-го шага составляет 0 лет.

5. Определение функции изменения состояния.

$$f_i(t)=\begin{cases} t+1, & если\ x_i=0; \\ 1, & если\ x_i=1. \end{cases}$$

6. Составление функционального уравнения для $i=m$.

$$W_m(t)=\max_{x_m\in\{0,1\}}\begin{cases} r(t)-l(t); \\ c(t)-p+r(0)-l(0), \end{cases}$$

где $W_i(t)$ - прибыль от использования оборудования возраста t лет с i-го шага (с конца i-го года) до конца периода эксплуатации.

$W_{i+1}(t+1)$ - прибыль от использования оборудования возраста $t+1$ год с $(i+1)$-го шага до конца периода эксплуатации.

Таким образом, математическая модель задачи построена. Расчёт модели проведем для конкретного примера.

Пример 2.27

$m=12$, $p=10$, $c(t)=0$, $r(t)-l(t)=\varphi(t)$.

Значения $\varphi(t)$ заданы в табл. 2.39.

Таблица 2.39

t	0	1	2	3	4	5	6	7	8	9	10	11	12
$\varphi(t)$	10	9	8	7	6	5	4	3	2	1	0	0	0

Для данного примера функциональные уравнения будут иметь вид

$$W_m(t)=\max_{x_m\in\{0,1\}}\begin{cases} \varphi(t); \\ -p+\varphi(0), \end{cases} \qquad W_i(t)=\max_{x_i\in\{0,1\}}\begin{cases} \varphi(t)-W_{i+1}(t+1); \\ -p+\varphi(0)-W_{i+1}(1). \end{cases}$$

Для решения данной задачи заполняется табл. 2.40.

В левой колонке таблицы записываются возможные состояния системы $t=\overline{0,12}$, в верхней строке – номера шагов $i=\overline{0,12}$. Для каждого шага определяют условные оптимальные управления $x_i(t)$ и условный оптимальный выигрыш $W_i(t)$ с i-го шага и до конца для оборудования возраста t лет.

Поясним, как заполняется таблица для нескольких шагов.

1. Условная оптимизация начинается с последнего 12-го шага. Для $i=12$ рассматриваются возможные состояния системы $t=0,1,\ldots,12$. Функциональное уравнение на 12-м шаге имеет вид

$$W_{12}(t)=\max_{x_m\in\{0,1\}}\begin{cases} \varphi(t); \\ -p+\varphi(0), \end{cases}$$

1) $t=0$

$$W_{12}(0)=\max_{x_m\in\{0,1\}}\begin{cases}10;\\-10+10\end{cases}=10;\quad x_{12}(0)=0.$$

2) $t=1$

$$W_{12}(1)=\max_{x_m\in\{0,1\}}\begin{cases}9;\\-10+10\end{cases}=9;\quad x_{12}(1)=0.$$

…

10) $t=9$

$$W_{12}(9)=\max_{x_m\in\{0,1\}}\begin{cases}1;\\-10+10\end{cases}=1;\quad x_{12}(9)=0.$$

11) $t=10$

$$W_{12}(10)=\max_{x_m\in\{0,1\}}\begin{cases}0;\\-10+10\end{cases}=0;\quad x_{12}(10)=0,\ x_{12}(10)=1.$$

…

13) $t=12$

$$W_{12}(12)=\max_{x_m\in\{0,1\}}\begin{cases}0;\\-10+10\end{cases}=0;\quad x_{12}(12)=0,\ x_{12}(12)=1.$$

Таким образом, на 12-м шаге оборудование возраста 0-9 лет заменять не надо. Оборудование возраста 10-12 лет можно заменить или продолжить его эксплуатацию, так как для $t=10,11,12$ имеется два условных оптимальных управления 1 и 0.

По результатам расчётов заполняются два столбца таблицы, соответствующие $i=12$.

2. Условная оптимизация 11-го шага.

Для $i=11$ рассматриваются все возможные состояния системы $t=0,1,\ldots,12$. Функциональное уравнение на 11-м шаге имеет вид

$$W_{11}(t)=\max_{x_m\in\{0,1\}}\begin{cases}\varphi(t)+W_{12}(t+1);\\-p+\varphi(0)+W_{12}(1).\end{cases}$$

1) $t=0$

$$W_{11}(0)=\max_{x_m\in\{0,1\}}\begin{cases}\varphi(0)+W_{12}(1);\\-p+\varphi(0)+W_{12}(1)\end{cases}=\max_{x_m\in\{0,1\}}\begin{cases}10+9;\\-10+10+9\end{cases}=19;\quad x_{11}(0)=0.$$

2) $t=1$

$$W_{11}(1)=\max_{x_m\in\{0,1\}}\begin{cases}\varphi(1)+W_{12}(2);\\-p+\varphi(0)+W_{12}(1)\end{cases}=\max_{x_m\in\{0,1\}}\begin{cases}9+8;\\-10+10+9\end{cases}=17;\quad x_{11}(1)=0.$$

…

6) $t=5$

$$W_{11}(5)=\max_{x_m\in\{0,1\}}\begin{cases}\varphi(5)+W_{12}(6);\\-p+\varphi(0)+W_{12}(1)\end{cases}=\max_{x_m\in\{0,1\}}\begin{cases}5+4;\\-10+10+9\end{cases}=9;\qquad x_{11}(5)=0;\qquad x_{11}(5)=1.$$

Таблица 2.40

t	$i=12$		$i=11$		$i=10$		$i=9$		$i=8$		$i=7$		$i=6$		$i=5$		$i=4$		$i=3$		$i=2$		$i=2$	
	x_{12}	W_{12}	x_{11}	W_{11}	x_{10}	W_{10}	x_9	W_9	x_8	W_8	x_7	W_7	x_6	W_6	x_5	W_5	x_4	W_4	x_3	W_3	x_2	W_2	x_1	W_1
0	0	10	0	19	0	27	0	34	0	40	0	45	0	51	0	58	0	64	0	70	0	75	**0**	**82**
1	0	9	0	17	**0**	**24**	0	30	0	35	0	41	**0**	**48**	0	54	0	60	0	65	**0**	**72**	0	78
2	0	8	**0**	**15**	0	21	0	26	0	32	**0**	**39**	0	45	0	51	0	56	**0**	**63**	0	69	0	75
3	**0**	**7**	0	13	0	18	0/1	24	**0**	**31**	0	37	0	43	0/1	48	**0**	**55**	0	61	0	67	0	73
4	0	6	0	11	1	17	**1**	**24**	0/1	30	0	35	0/1	41	**1**	**48**	0/1	54	0/1	60	0	66	1	72
5	0	5	0/1	9	1	17	1	24	1	30	0/1	35	1	41	1	48	1	54	1	60	0/1	65	1	72
6	0	4	1	9	1	17	1	24	1	30	1	35	1	41	1	48	1	54	1	60	1	65	1	72
7	0	3	1	9	1	17	1	24	1	30	1	35	1	41	1	48	1	54	1	60	1	65	1	72
8	0	2	1	9	1	17	1	24	1	30	1	35	1	41	1	48	1	54	1	60	1	65	1	72
9	0	1	1	9	1	17	1	24	1	30	1	35	1	41	1	48	1	54	1	60	1	65	1	72
10	0/1	0	1	9	1	17	1	24	1	30	1	35	1	41	1	48	1	54	1	60	1	65	1	72
11	0/1	0	1	9	1	17	1	24	1	30	1	35	1	41	1	48	1	54	1	60	1	65	1	72
12	0/1	0	1	9	1	17	1	24	1	30	1	35	1	41	1	48	1	54	1	60	1	65	1	72

7) $t=6$

$$W_{11}(6)=\max_{x_m\in\{0,1\}}\begin{cases}\varphi(6)+W_{12}(7);\\ -p+\varphi(0)+W_{12}(1)\end{cases}=\max_{x_m\in\{0,1\}}\begin{cases}4+3;\\ -10+10+9\end{cases}=9;\quad x_{11}(6)=1.$$

… … …

13) $t=12$

$$W_{11}(12)=\max_{x_m\in\{0,1\}}\begin{cases}\varphi(12);\\ -p+\varphi(0)+W_{12}(1)\end{cases}=\max_{x_m\in\{0,1\}}\begin{cases}0;\\ -10+10+9\end{cases}=9;\quad x_{11}(12)=1.$$

Таким образом, на 11-м шаге не следует заменять оборудование возраста 0-4 года. Для оборудования возраста 5 лет возможны две стратегии использования: заменить или продолжать эксплуатировать.

Начиная с 6-го года оборудование следует заменить. По результатам расчётов заполняются два столбца таблицы, соответствующие $i=11$.

3. $i=10$

1) $t=0$

$$W_{10}(0)=\max_{x_m\in\{0,1\}}\begin{cases}\varphi(0)+W_{11}(1);\\ -p+\varphi(0)+W_{11}(1)\end{cases}=\max_{x_m\in\{0,1\}}\begin{cases}10+17;\\ -10+10+17\end{cases}=27;\quad x_{10}(0)=0.$$

2) $t=1$

$$W_{10}(1)=\max_{x_m\in\{0,1\}}\begin{cases}\varphi(1)+W_{11}(2);\\ -p+\varphi(0)+W_{11}(1)\end{cases}=\max_{x_m\in\{0,1\}}\begin{cases}9+15;\\ -10+10+17\end{cases}=24;\quad x_{10}(1)=0.$$

3) $t=2$

$$W_{10}(2)=\max_{x_m\in\{0,1\}}\begin{cases}\varphi(2)+W_{11}(3);\\ -p+\varphi(0)+W_{11}(1)\end{cases}=\max_{x_m\in\{0,1\}}\begin{cases}8+13;\\ -10+10+17\end{cases}=21;\quad x_{10}(2)=0.$$

4) $t=3$

$$W_{10}(3)=\max_{x_m\in\{0,1\}}\begin{cases}\varphi(3)+W_{11}(4);\\ -p+\varphi(0)+W_{11}(1)\end{cases}=\max_{x_m\in\{0,1\}}\begin{cases}7+11;\\ -10+10+17\end{cases}=18;\quad x_{10}(3)=0.$$

5) $t=4$

$$W_{10}(4)=\max_{x_m\in\{0,1\}}\begin{cases}\varphi(4)+W_{11}(5);\\ -p+\varphi(0)+W_{11}(1)\end{cases}=\max_{x_m\in\{0,1\}}\begin{cases}6+9;\\ -10+10+17\end{cases}=17;\quad x_{10}(4)=1.$$

… … …

13) $t=12$

$$W_{10}(0)=\max_{x_m\in\{0,1\}}\begin{cases}\varphi(12)\\ -p+\varphi(0)+W_{12}(1)\end{cases}=\max_{x_m\in\{0,1\}}\begin{cases}0\\ -10+10+17\end{cases}=17;\quad x_{10}(12)=1.$$

На 10-м шаге не следует заменять оборудование возраста 0-3 года. Начиная с 4-го года, оборудование следует заменять, так как новое оборудование приносит большую прибыль.

По результатам расчётов заполняются два столбца, соответствующие $i=10$.

Аналогичным образом заполняются остальные девять столбцов таблицы. При расчётах $W_{i+1}(t)$ на каждом шаге значения $\varphi(t)$ для каждого $t=\overline{0,12}$ берутся из таблицы исходных данных, а значения $W_i(t)$ - из последнего, заполненного на предыдущем шаге столбца.

Этап условной оптимизации заканчивается после заполнения табл. 2.40

Безусловная оптимизация начинается с первого шага.

Предположим, что на первом шаге $i=1$ имеется новое оборудование, возраст которого 0 лет.

Для $t=t_1=0$ оптимальный выигрыш составляет $W_1(0)=82$. Это значение соответствует максимальной прибыли от использования нового оборудования в течение 12 лет.

$W^*=W_1(0)=82$.

Выигрышу $W_1(0)=82$ соответствует безусловное оптимальное управление $x_1(0)=0$.

Для $i=2$ $t_2=t_1+1=1$.

Безусловное оптимальное управление $x_2(1)=0$.

Для $i=3$ $t_3=t_2+1=2$.

Безусловное оптимальное управление $x_3(2)=0$.

И, далее, соответственно

$i=4$; $t_4=t_3+1=3$; $x_4(3)=0$

$i=5$; $t_5=t_4+1=4$; $x_5(4)=1$

$i=6$; $t_6=1$; $x_6(1)=0$

$i=7$; $t_7=t_6+1=2$; $x_7(2)=0$

$i=8$; $t_8=t_7+1=3$; $x_8(3)=0$

$i=9$; $t_9=t_8+1=4$; $x_9(4)=1$

$i=10$; $t_{10}=1$; $x_{10}(1)=0$

$i=11$; $t_{11}=t_{10}+1=2$; $x_{11}(2)=0$

$i=12$; $t_{12}=t_{11}+1=3$; $x_{12}(3)=0$.

Управление, составляющие оптимальную стратегию использования оборудования, выделены в табл. 2.40 жирным шрифтом.

В рамках данной задачи оптимальная стратегия заключается в замене оборудования при достижении им возраста 4-х лет. Аналогичным образом можно определить оптимальную стратегию использования оборудования любого возраста.

Задача решена.

Пример 2.28. **Самостоятельная работа на тему:** «Оптимальное распределение инвестиций как задача динамического программирования».

Задание. Инвестор выделяет средства в размере D условных единиц, которые должны быть распределены между m-предприятиями. Каждое i-е предприятие при инвестировании в него средств x приносит прибыль $\varphi_i(x)$ у.е., $i=\overline{1,m}$. Необходимо выбрать оптимальное распределение инвестиций между предприятиями, обеспечивающее максимальную прибыль.

В качестве выигрыша W в задаче рассматривается прибыль, приносимая m предприятиями.

Построение математической модели.

1. Определение числа шагов. Число шагов m равно числу предприятий, в которые осуществляется инвестирование.

2. Определение состояний системы. Состояние системы на каждом шаге характеризуется количеством средств s, имеющихся в наличии перед данным шагом, $s \le D$.

3. Выбор шаговых управлений. Управлением на i-м шаге x_i, $i = \overline{1,m}$ является количество средств, инвестируемых в i-е предприятие.

4. Функция выигрыша на i-м шаге $\varphi_i(x_i)$ - это прибыль, которую приносит i-е предприятие при инвестировании в него средств x_i.

$$W = \sum_{i=1}^{m} \varphi_i(x_i),$$

следовательно, данная задача может быть решена методом динамического программирования.

5. Определение функции перехода в новое состояние.

$$f_i(s,x) = s - x .$$

Таким образом, если на i-м шаге система находилась в состоянии s, а выбрано управление x, то на $i+1$ шаге система будет находиться в состоянии $s-x$. Другими словами, если в наличии имеются средства в размере s усл. ед., и в i-е предприятие инвестируется x у.е., то для дальнейшего инвестирования остаётся $s-x$ у.е.

6. Составление функционального уравнения для $i = m$.

$$W_m(s) = \varphi_m(s),\ x_m(s) = s .$$

На последнем шаге, т.е. перед инвестированием средств в последнее предприятие, условное оптимальное управление соответствует количеству средств, имеющихся в наличии, т.е. определяет, сколько средств осталось, сколько надо вложить в последнее предприятие. Условный оптимальный выигрыш равен доходу, приносимому последним предприятием.

7. Составление функционального уравнения

$$W_i(s) = \max_{x \le s}\{\varphi_i(x) + W_{i+1}(s-x)\}.$$

Поясним данное уравнение. Пусть перед i-м шагом у инвестора остались средства в размере s у. е. Тогда x у. е. он может вложить в i-е предприятие, при этом оно принесет доход $\varphi_i(x)$, а оставшиеся $s-x$ у. е. – в остальные предприятия с $i+1$-го до m-го. Условный оптимальный выигрыш от такого вложения $W_{i+1}(s-x)$. Оптимальным будет то условное управление x, при котором сумма $\varphi_i(x)$ и $W_{i+1}(s-x)$ максимальна.

Пример расчёт.

Пусть заданы $D = 5000$, $m = 3$. Значение $\varphi_i(x)$, $i = \overline{1,3}$ заданы в таблице 2.41.

Таблица 2.41

x, тыс. у. е.	$\varphi_1(x)$, тыс. у. е.	$\varphi_2(x)$, тыс. у. е.	$\varphi_3(x)$, тыс. у. е.
1	1.5	2	1.7
2	2	2.1	2.4
3	2.5	2.3	2.7
4	3	3.5	3.2
5	3.6	4	3.5

Для $x_1 > x_2$ $\varphi_i(x_1) \geq \varphi_i(x_2)$, $i = \overline{1,5}$.

В задаче сделано предположение, что вкладываются средств объёмом в тысячу условных единиц.

Решение

Проведем условную оптимизацию, результаты которой отразим в табл. 2.42.

Таблица 2.42

S	$i=3$		$i=2$		$i=1$	
	$x_3(s)$	$W_3(s)$	$W_2(s)$	$W_2(s)$	$W_1(s)$	$W_1(s)$
1	1	1,7	0	2,0		
2	**2**	2,4	1	3,7		
3	3	2,7	**1**	4,4		
4	4	3,2	1	4,7		
5	5	3,5	1/4	5,2	**2**	**6,4**

В первой колонке таблицы записываются возможные состояния системы $s = \overline{1,5}$, в верхней строке – номера шагов $i = \overline{1,3}$. На каждом шаге определяются условные оптимальные управления $x_i(s)$ и условные оптимальные выигрыши $W_i(s)$.

1. Проведение условной оптимизации для последнего шага $i = 3$. Функциональное уравнение на последнем шаге имеет вид $W_3(s) = \varphi_3(s)$, $x_3(s) = s$,

Поэтому два столбца табл. 2.42, соответствующие $i = 3$, заполняются автоматически по таблице исходных данных.

2. Условная оптимизация для $i = 2$.

Функциональное уравнение имеет вид $W_2(s) = \max\limits_{x \leq s}\{\varphi_2(x) + W_3(s - x)\}$.

Для проведения условной оптимизации заполним ряд вспомогательных таблиц (табл. 2.43 – 2.48), соответствующих различным значениям s, т.е. различным исходам окончания предыдущего шага.

1) $s = 1$

Таблица 2.43

x	$1-x$	$\varphi_2(x)$	$W_3(1-x)$	$\varphi_2(x)+W_3(1-x)$
0	1	0	1.7	1.7
1	0	2	0	2

$\max\limits_{x\le 1}\{1.7;2\}=2$, следовательно

$W_2(1)=2$;

$x_2(1)=1$.

2) $s=2$

Таблица 2.44

x	$2-x$	$\varphi_2(x)$	$W_3(2-x)$	$\varphi_2(x)+W_3(2-x)$
0	2	0	2.4	2.4
1	1	2	1.7	3.7
2	0	2.1	0	2.1

$\max\limits_{x\le 2}\{2.4;3.7;2.1\}=3.7$, следовательно

$W_2(2)=3.7$;

$x_2(2)=1$.

3) $s=3$

Таблица 2.45

x	$3-x$	$\varphi_2(x)$	$W_3(3-x)$	$\varphi_2(x)+W_3(3-x)$
0	3	0	2.7	2.7
1	2	2	2.4	4.4
2	1	2.1	1.7	3.8
3	0	2.3	0	2.3

$\max\limits_{x\le 3}\{2.7;4.4;3.8;2.3\}=4.4$, следовательно

$W_2(3)=4.4$;

$x_2(3)=1$.

4) $s=4$

Таблица 2.46

x	$4-x$	$\varphi_2(x)$	$W_3(4-x)$	$\varphi_2(x)+W_3(4-x)$
0	4	0	3.2	3.2
1	3	2	2.7	4.7
2	2	2.1	2.4	4.5
3	1	2.3	1.7	4
4	0	3.5	0	3.5

$\max\limits_{x\le 4}\{3.2;4.7;4.5;4;3.5\}=4.7$, следовательно

$W_2(4)=4.7$;

$x_2(4)=1$.

5) $s=5$

Таблица 2.47

x	$5-x$	$\varphi_2(x)$	$W_3(5-x)$	$\varphi_2(x)+W_3(5-x)$
0	5	0	3.5	3.5
1	4	2	3.2	5.2
2	3	2.1	2.7	4.8
3	2	2.3	1.4	4.7
4	1	3.5	1.7	5.2
5	0	4	0	4

$\max\limits_{x\le 5}\{3.5;5.2;4.8;4.7;5.2;4\}=5.2$, следовательно

Для $s=5$ $W_2(5)=5.2$ возможны два условных оптимальных управления $x_2(5)=1$ и $x_2(5)=4$.

3. Условная оптимизация для $i=1$.

Перед первым шагом состояние системы известно.

$s=D$ тыс. у. е., и условную оптимизацию следует проводить только для этого значения $s=5$

Таблица 2.48

x	$5-x$	$\varphi_2(x)$	$W_2(5-x)$	$\varphi_2(x)+W_2(5-x)$
0	5	0	5.2	5.2
1	4	1.5	4.7	6.2
2	3	2	4.4	6.4
3	2	2.5	3.7	6.2
4	1	3	2	5
5	0	3.6	0	3.6

$\max\limits_{x\le 5}\{5.2;6.2;6.4;6.2;5;3.6\}=6.4$, следовательно, $W_1(5)=6.4$; $x_1(5)=2$.

Оптимальная прибыль, приносимая тремя предприятиями при инвестировании в них 5 тыс. у. е., равна 6,4 тыс. у. е. $W^*=W_1(5)=6.4$.

Проведем безусловную оптимизацию и представим результаты в таблице.

Пусть $i=1$; $s_1=5$; $W_1(5)=6.4$; $x_1^*=x_1(5)=2$.

Для $i=2$ имеем $s_2=s_1-x_1=5-2=3$. $W_2(3)=4.4$; $x_2^*=x_2(3)=1$.

Для $i=3$ имеем $s_3=s_2-x_2=3-1=2$.

$W_3(2)=2.4$; $x_3^*=x_3(2)=2$. $x^*=(2;1;2)$.

Таким образом, для получения максимальной прибыли в размере 6400 у. е. следует по 2000 у. е. вложить в первое и третье предприятия и 1000 у. е. – во второе предприятие.

Задача решена.

2.6 Контрольные вопросы и задачи

1. Поясните основные отличия классических моделей задач линейного и нелинейного программирования.

2. Укажите стандартную форму записи для ЗЛП: $F = -2x_1 + x_2 + 5x_3 \to \max$

$$\begin{cases} 4x_1 + 2x_2 + 5x_3 \le 12 \\ 6x_1 - 3x_2 + 4x_3 = 18 \\ 3x_1 + 3x_2 - 2x_3 \ge 16 \\ x_1, x_2, x_3 \ge 0. \end{cases}$$

3. Сформулируйте двойственную задачу по отношению к ЗЛП:

$$F = x_1 - 2x_2 + 5x_3 \to \max$$

$$\begin{cases} 2x_1 + 2x_2 + 4x_3 \le 18 \\ 2x_1 + x_2 - 3x_3 \le 20 \\ 5x_1 - 3x_2 + 6x_3 \ge 19 \\ x_1, x_2, x_3 \ge 0. \end{cases}$$

4. Исходная задача линейного программирования имеет оптимальный план со значением целевой функции $F_{\max} = 10$. Какое из чисел является значением целевой функции $F^*_{\min}$ двойственной задачи?

5. Фирма производит два вида продуктов K_1 и K_2. Для изготовления продуктов применяются машины A, B, C, D. Время необходимое для изготовления продуктов K_1 и K_2 на равных машинах, допустимое время использования машин, а также прибыль от продажи продуктов приведены в таблице:

Машины	Допустимое время (в часах)	Необходимое время (в часах)	
		K_1	K_2
A	4	0	1
B	7	4	1
C	5	2	1
D	10	6	1
Прибыль от продажи продуктов, тыс. руб.		10	4

Необходимо выполнить следующее.

1) Определить графическим методом, какое количество каждого продукта необходимо произвести, чтобы прибыль была максимальной?

2) Решить задачу симплекс-методом.

3) Для данной задачи записать двойственную ЗЛП.

6. Предприятие, располагающее ресурсами сырья четырёх видов A, B, C, D, может производить продукцию двух видов P_1, P_2. В таблице указаны затраты

ресурсов на изготовление 1 т продукции, объём ресурсов и прибыль, получаемая от продажи 1 т соответствующей продукции.

Вид сырья	Вид продукции		Объём ресурсов, т.
	P_1	P_2	
A	0	1	5
B	1	0	4
C	2	1	9
D	2	1	6
Прибыль руб.	2	5	

а) Определить ассортимент выпускаемой продукции, при котором полученная прибыль будет максимальной? Решить задачу геометрическим способом.

б) Решить задачу симплекс методом.

в) Для задачи написать двойственную ЗЛП.

6. Для изготовления двух видов изделий A и B завод использует в качестве сырья алюминий и медь. На изготовление изделий заняты токарные и фрезерные станки. Исходные данные задачи приведены в таблице:

Вид ресурсов	Объём ресурсов	Нормы расхода на 1 изделие	
		A	B
Алюминий, кг.	12	3	2
Медь, кг.	20	1	4
Токарные станки, станко-час.	7	2	1
Фрезерные станки, станко-час.	3	1	0
Прибыль на 1 изделие, тыс. руб.		4	1

а) Определить ассортимент выпускаемой продукции, при котором прибыль будет максимальной? Решить задачу геометрическим способом.

б) Решить задачу симплекс методом.

в) Для задачи написать двойственную ЗЛП.

7. Фирма производит два вида продуктов K_1 и K_2. Для изготовления продуктов применяются машины A, B, C, D. Время необходимое для изготовления продуктов K_1 и K_2 на равных машинах, допустимое время использования машин, а также прибыль от продажи продуктов приведены в таблице:

Машины	Допустимое время (в часах)	Необходимое время (в часах)	
		K_1	K_2
A	4	0	1
B	5	2	1
C	7	4	1

D	3	2	0
Прибыль на 1 изделие, тыс. руб.		2	3

а) какое количество каждого продукта необходимо произвести, чтобы прибыль была максимальной? б) Решить задачу графическим методом и симплекс-методом; в) для данной задачи сформулировать двойственную ЗЛП.

8. Для изготовления двух видов продукции P_1, P_2 используются четыре вида сырья: A, B, C, D. Запасы сырья ограничены: B – 6 т., C – 8 т., D – 5 т., A – 12 т. Нормы расхода сырья на изготовление 1 т. Продукции приведены в таблице. Составить такой план выпуска продукции, чтобы при её реализации получить максимальную прибыль.

Вид сырья	Количество сырья на 1т продукции	
	P_1	P_2
A	2	3
B	1	1
C	2	1
D	0	1
Прибыль руб.	2	5

Требуется:

а) решить задачу геометрическим способом; б) решить задачу симплекс методом; в) сформулировать двойственную ЗЛП.

9. Фирма выпускает два вида продукции. В процессе производства используются три технологические операции. При изготовления второго изделия технологическая операция №2 не выполняется. Время выполнения операции (в часах) приводится в таблице.

Изделие	Технологические операции		
	1	2	3
1	1	3	1
2	2	-	4

Фонд рабочего времени ограничен:

- для первой операции – 12 часов;
- для второй операции – 9 часов;
- для третьей операции – 6 часов;

Изучение рынка показало, что ожидаемая прибыль от продажи одного изделия видов 1 и 2 соответственно равна 4 и 7 руб.

а) Каков наиболее выгодный суточный объём производства каждого вида продукции? Решить задачу геометрическим способом.

б) Решить задачу симплекс методом.

в) Для задачи написать двойственную ЗЛП.

10. Фирма производит два продукта A и B, рынок сбыта которых не ограничен. Каждый продукт должен быть обработан машинами 1, 2 и 3. Время обработки для каждого из изделий A и B приведено в таблице:

Продукт	Машина		
	1	2	3
A	5	4	2
B	6	3	4

Время работы машин 1, 2, 3 соответственно 42, 38 и 38 час. в неделю. Прибыль от изделий *A* и *B* составляет соответственно 8 и 6 тыс. рублей.

Требуется:

а) определить недельные нормы выпуска изделий *A* и *B* и рассчитать максимальную прибыль; решить задачу геометрическим способом;

б) решить задачу симплекс методом;

в) для исходной задачи сформулировать двойственную ЗЛП.

11. Решить задачу планирования объёмов выпуска трёх видов товаров x_1, x_2 и x_3, при которых обеспечивается максимальный объём продаж $F(x_1, x_2, x_3)$ при наличии ограничений на расход сырья и трудозатраты (приложение 2, таблица, вариант 24). Задача решена в вычислительной среде Mathcad (v.14) с использованием стандартной процедуры **Maximize(F,x1,x2,x3).**

Рис. 2.38 – Фрагмент программы решения ЗЛП в среде Mathcad

ГЛАВА 3. СПЕЦИАЛЬНЫЕ ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МОДЕЛИ

3.1 Экономико-математическая модель межотраслевого баланса (модель Леонтьева)

Рассмотрим наиболее простой вариант межотраслевого баланса, которую называют моделью «*затраты-выпуск*». Анализ модели «затраты-выпуск» сводится к составлению и решению системы линейных алгебраических уравнений, в которых параметрами являются коэффициенты затрат на производство продукции.

Пусть весь производственный сектор нарогного хозяйства подразделяется на «чистых отраслей». «*Чистая отрасль*» - это условное понятие – некоторая часть хозяйства, более или менее цельная, например, энергетика, машиностроение, сельское хозяйство.

Пусть $x_{i,j}$ - объём продукции отрасли i, расходуемый в отрасли j; X_i - объём производства отрасли i за данный промежуток времени (так называемый *валовый выпуск* продукции i); Y_i - объём потребления продукции отрасли в непроизводственной сфере (объём конечного потребления); Z_j - условно чистая продукция, которая включает оплату труда, чистый доход и амортизацию.

Единицы измерения всех указанных величин могут быть или натуральными (кубометры, килограммы, штуки и т. п.) или стоимостными. В зависимости от этого различают *натуральный* и *стоимостной* межотраслевые балансы. Ниже будем рассматривать стоимостной баланс.

В табл. 3.1 представлена принципиальная схема межотраслевого баланса в стоимостном выражении. Рассматривая схему баланса по столбцам, можно заметить, что итог материальных затрат любой потребляющей отрасли и её условно чистой продукции равен валовой продукции этой отрасли. Данный вывод можно записать в виде соотношения:

$$X_j = \sum_{i=1}^{n} x_{ij} + Z_i, \quad j = 1,...,n. \qquad (3.1)$$

Напомним, что величина условно чистой продукции Z_j равна сумме амортизации, оплаты труда и чистого дохода отрасли j. Соотношение (3.1) охватывает систему из n уравнений, отражающих стоимостной состав продукции всех отраслей материальной сферы.

Рассматривая схему МОБ по строкам для каждой производящей отрасли, отметим, что валовая продукция той или иной отрасли равна сумме материальных затрат потребляющих её продукцию отраслей и конечной продукции данной отрасли:

$$X_j = \sum_{j=1}^{n} x_{ij} + Y_i, \quad i = 1,...,n. \qquad (3.2)$$

Таблица 3.1

Производящие отрасли	Потребляющие отрасли				Конечный продукт	Валовый продукт
	1	2	…	n		
1	X_{11}	X_{12}	…	X_{1n}	Y_1	X_1
2	X_{21}	X_{22}	…	X_{2n}	Y_2	X_2
…	…	…	…	…	…	…
n	X_{n1}	X_{n2}	…	X_{nn}	Y_n	X_n
Условно чистая продукция	Z_1	Z_2	…	Z_n	$\sum_{i=1}^{n} X_i = \sum_{j=1}^{n} X_j$	
Валовой продукт	X_1	X_2	…	X_n		$\sum_{i=1}^{n} X_i = \sum_{j=1}^{n} X_j$

Формула (3.2) описывает систему из n уравнений, которые называются уравнениями распределения продукции отраслей материального производства по направлениям использования.

Балансовый характер таблицы выражается в том, что

$$\sum_{i=1}^{n} X_i = \sum_{j=1}^{n} X_j , \quad \sum_{i=1}^{n} Y_i = \sum_{j=1}^{n} Z_j .$$

Коэффициенты прямых материальных затрат. Основу экономико-математической модели МОБ составляет технологическая матрица прямых затрат А (a_{ij}).

Коэффициент прямых материальных затрат a_{ij} показывает, сколько необходимо единиц продукции отрасли i для производства единицы продукции отрасли j, если учитывать только прямые затраты:

$$a_{ij} = x_{ij} / X_j , \; i, j = 1,..., n . \tag{3.3}$$

Введём два важных предположения, необходимых для рассмотрения модели Леонтьева: 1) сложившуюся технологию производства считаем неизменной, тогда матрица $A = (a_{ij})$ постоянна; 2) постулируем свойство линейности существующих технологий – для выпуска отраслью j любого объёма продукции X_j необходимо затратить продукцию отрасли i в количестве $a_{ij} \cdot X_j$, т.е. материальные издержки пропорциональны объёму производимой продукции:

$$x_{ij} = a_{ij} \cdot X_j. \tag{3.4}$$

Подставляя выражение (3.4) в балансовое соотношение (3.2), получаем

$$X_i = \sum_{i=1}^{n} a_{ij} \cdot X_i + Y_i , \tag{3.5}$$

или в матричной форме

$$X=AX+Y. \tag{3.6}$$

С помощью этой модели можно выполнять три вида плановых расчётов:

а) задавая для каждой отрасли величины валовой продукции (X_j), можно определить объёмы конечной продукции каждой отрасли (Y_i):

$$Y=(E-A) \cdot X; \tag{3.7}$$

б) задавая величины конечной продукции всех отраслей (Y_i), можно определить величины валовой продукции каждой отрасли (X_i):

$$X=(E-A)^{-1} \cdot Y; \tag{3.8}$$

в) задавая для ряда отраслей величины валовой продукции, а для всех остальных отраслей – объёмы конечной продукции, можно найти величины конечной продукции первых отраслей и объёмы валовой продукции вторых.

В формулах (3.7) и (3.8) символ E обозначает единичную матрицу размера n , а выражение $(E-A)^{-1}$ - матрицу, обратную матрице $(E-A)$. Если определитель матрицы $(E-A)$ не равен нулю, т.е. эта матрица невырожденная, то существует обратная к ней матрица. Обозначим обратную матрицу через $B=(E-A)^{-1}$, тогда систему уравнений в матричной форме (3.8) можно записать в виде $X = B \cdot Y$.

Элементы матрицы B называются коэффициентами полных материальных затрат. Они показывают, сколько всего нужно произвести продукции отрасли *i* для выпуска в сферу конечного использования продукции отрасли *j*. Плановые расчёты по модели Леонтьева можно выполнять, если соблюдается условие продуктивности.

Неотрицательную матрицу A будем называть продуктивной, если существует такой неотрицательный вектор $X \geq 0$, что $X > A \cdot X$. Данное условие означает, что существование положительного вектора конечной продукции $Y > 0$ для модели межотраслевого баланса (3.6).

Для того чтобы матрица коэффициентов прямых материальных затрат A была продуктивной, **необходимо и достаточно**, чтобы выполнялось одно из перечисленных ниже условий:

1) матрица $(E-A)$ неотрицательно обратима, т.е. существует обратная матрица $(E-A)^{-1} \geq 0$;

2) матричный ряд $E + A + A^2 + A^3 + ... = \sum_{k=1}^{\infty} A^k$ сходится, причём его сумма равна обратной матрице $(E-A)^{-1}$:

$$B=(E-A)^{-1} = E + A + A^2 + A^3 + ...$$

3) наибольшее по модулю собственное значение λ матрицы A, т.е. решение характеристического уравнения $|\lambda \cdot E - A| = 0$, строго меньше единицы;

4) все главные миноры матрицы $(E - A)$, т.е. определители матриц, образованные элементами первых строк и первых столбцов этой матрицы, порядка от 1 до n, положительны.

Более простым способом проверки продуктивности матрицы A является ограничение на величину её нормы, в данном случае на величину наибольшей из сумм элементов матрицы A в каждом столбце. Если норма матрицы A строго меньше 1, то эта матрица продуктивна. Данное условие является достаточным, но не необходимым условием продуктивности, поэтому матрица A может оказаться продуктивной и в случае, когда её норма больше единицы.

Пример 3.1. Даны коэффициенты прямых затрат a_{ij} и конечный продукт Y_i для трёхотраслевой экономической системы:

$$A = \begin{bmatrix} 0{,}3 & 0{,}1 & 0{,}4 \\ 0{,}2 & 0{,}5 & 0{,}0 \\ 0{,}3 & 0{,}1 & 0{,}2 \end{bmatrix}, \qquad Y = \begin{bmatrix} 200 \\ 100 \\ 300 \end{bmatrix}.$$

Требуется определить:

1) коэффициенты полных затрат; вектор валового выпуска;
2) межотраслевые поставки продукции;
3) проверить продуктивность матрицы A;
4) заполнить схему межотраслевого баланса.

Для решения задачи воспользуемся функциями электронных таблиц Excel. В табл. 3.2 приведены результаты расчётов по первым трём пунктам задачи.

Для вычисления обратной матрицы необходимо:

- выделить диапазон ячеек для размещения обратной матрицы;
- выбрать функцию МОБР в категории Математические;
- ввести диапазон ячеек, в которых содержится матрица $(E - A)$;
- нажать клавиши CTRL+SHIFT+ENTER.

В ячейки **B6:D8** запишем элементы матрицы $(E - A)$. Массив $(E - A)$ задан как диапазон ячеек. Выделим диапазон **B10:D12** для размещения обратной матрицы $B = (E - A)^{-1}$ и введём формулу для вычислений МОБР (**B6:D8**). Затем следует нажать клавиши CTRL+SHIFT+ENTER.

Все элементы матрицы коэффициентов полных затрат **B** неотрицательны, следовательно, матрица **A** продуктивна.

2. Вычислим вектор валового выпуска **X** по формуле $X = B \cdot Y$.

Для умножения матриц необходимо:

- выделить диапазон ячеек для размещения результата умножения матриц;
- выбрать функцию МУМНОЖ в категории Математические;
- ввести диапазоны ячеек, где содержатся матрицы **B** и **Y**;

- нажать клавиши CTRL+SHIFT+ENTER.

В ячейки G10:G12 запишем элементы вектора конечного продукта **Y**. Выделим диапазон B15:B17 для размещения вектора валового выпуска **X**, вычисляемого по формуле $X = B \cdot Y$. Затем вводим формулу для вычислений МУМНОЖ (B10:D12, G10:G12). Далее нажимают клавиши CTRL+SHIFT+ENTER.

3. Межотраслевые поставки x_{ij} вычисляют по формуле $x_{ij} = a_{ij} \cdot X_j$.

4. Заполняют схему межотраслевого баланса (табл. 3.3).

Таблица 3.2

	A	B	C	D	E	F	G
1							
2		0,3	0,1	0,4			
3	A	0,2	0,5	0			
4		0,3	0,1	0,2			
5							
6		0,7	-0,1	-0,4			
7	E-A	-0,2	0,5	0			
8		-0,3	-0,1	0,8			
9	1)						
10		2,040816	0,612245	1,020408163			200
11	B	0,816327	2,244898	0,408163265	Y		100
12		0,867347	0,510204	1,683673469			300
13							
14	2)						
15		775,5102					
16	X	510,2041					
17		729,5918					
18							
19	3)						
20		232,6531	51,02041	291,8367347			
21	X	155,102	255,102	0			
22		232,6531	51,02041	145,9183673			

Таблица 3.3

Производящие отрасли	Потребляющие отрасли			Конечный продукт	Валовый продукт
	1	2	3		
1	232,6	51,0	291,8	200	775,3
2	155,1	255,0	0,0	100	510,1
3	232,6	51,1	145,9	300	729,6
Условно чистая продукция	155,0	153,1	291,9	600	
Валовый продукт	775,3	510,1	729,6		2015

3. 2 Анализ экономических показателей на основе межотраслевых балансовых моделей

К числу важнейших аналитических возможностей балансового метода относится определение прямых и полных затрат труда на единицу продукции и разработка на этой основе балансовых продуктово-трудовых моделей. При этом исходной моделью служит отчётный межпродуктовый баланс в натуральном выражении. В отдельной строке баланса даётся распределение затрат живого труда в производстве всех видов продукции. Предполагается, что трудовые затраты выражены в единицах труда одинаковой степени сложности.

Пусть L_j - затраты живого труда в производстве продукта j, X_j - объём производства этого продукта (валовой выпуск). Тогда прямые затраты труда на единицу продукции вида j (коэффициент прямой трудоёмкости) можно задать следующей формулой:

$$t_j = \frac{L_j}{X_j}; \quad j = 1,..., n. \tag{3.9}$$

Введём понятие *полных затрат труда* как суммы прямых затрат живого труда и затрат овеществлённого труда, перенесённых на продукт через израсходованные средства производства. Обозначим величину полных затрат труда на единицу продукции вида j через T_j, тогда произведения вида $a_{ij} \cdot T_i$ будут отражать затраты овеществлённого труда, перенесённого на единицу продукта j через средство производства i. Будем предполагать, что коэффициенты прямых материальных затрат a_{ij} выражены в натуральных единицах. Тогда полные трудовые затраты на единицу продукции вида j (*коэффициент полной трудоёмкости*) будут равны

$$T_j = \sum_{i=1}^{n} a_{ij} \cdot T_i + t_j; \quad j = 1,..., n. \tag{3.10}$$

Введём вектор-строку коэффициентов прямой трудоёмкости $\bar{t} = (t_1, t_2,..., t_n)$ и вектор-строку коэффициентов полной трудоёмкости $\overline{T} = (T_1, T_2,..., T_n)$. Тогда с помощью матрицы коэффициентов прямых материальных затрат A (в натуральном выражении) систему уравнений (3.10) можно переписать в матричном виде

$$\overline{T} = \overline{T} \cdot A + \bar{t}. \tag{3.11}$$

Выполнив очевидные матричные преобразования с использованием единичной матрицы E: $\overline{T} - \overline{T} \cdot A = \overline{T} \cdot E - \overline{T} \cdot A = \overline{T}(E - A) = \bar{t}$, получим следующее соотношение для вектора коэффициентов полной трудоёмкости:

$$\overline{T} = \bar{t} \cdot (E - A)^{-1}. \tag{3.12}$$

Матрица $(E - A)^{-1}$ представляет собой матрицу коэффициентов полных материальных затрат - B, поэтому равенство (3.12) можно переписать в виде

$$\overline{T} = \bar{t} \cdot B. \tag{3.13}$$

Обозначим через L величину совокупных затрат живого труда по всем видам продукции, которая с учётом формулы (3.9) будет равна

$$L = \sum_{j=1}^{n} L_j = \sum_{j=1}^{n} t_j X_j = \bar{t} \cdot \overline{X}. \qquad (3.14)$$

Используя соотношения (3.9),(3.13) и (3.14), приходим к следующему равенству:

$$\bar{t} \cdot \overline{X} = \overline{T} \cdot \overline{Y}. \qquad (3.15)$$

Здесь $\bar{t}$ *и* $\overline{T}$ - векторы-строки коэффициентов прямой и полной трудоёмкости, а $\overline{X}$ *и* $\overline{Y}$ - векторы столбцы валовой и конечной продукции соответственно.

Соотношение (3.15) представляет собой основное балансовое равенство в теории межотраслевого баланса труда. В данном случае его конкретное экономическое содержание заключается в том, что стоимость конечной продукции, оцененной по полным затратам труда, равна совокупным затратам живого труда. Сопоставляя потребительский эффект различных взаимозаменяемых продуктов с полными трудовыми затратами на их выпуск, можно судить о сравнительной эффективности их производства. Показатели полной трудоёмкости выявляют структуру затрат на выпуск различных видов продукции, и, прежде всего, соотношение между затратами живого и овеществлённого труда.

На основе коэффициентов прямой и полной трудоёмкости могут быть разработаны межотраслевые и межпродуктовые балансы затрат труда и использования трудовых ресурсов. Схематично эти балансы строятся по общему типу матричных моделей, однако все показатели в них (межотраслевые связи, конечный продукт, условно чистая продукция и др.) выражены в трудовых измерителях.

Пример 3.2

Пусть в дополнение к исходным данным примера 3.1 заданы в некоторых единицах измерения затраты живого труда (трудовые ресурсы) в трёх отраслях: $L_1 = 1160$, $L_2 = 460$, $L_3 = 875$. Требуется определить коэффициенты прямой и полной трудоёмкости и составить межотраслевой баланс затрат труда.

Решение.

1. Опираясь на формулу (3.9) и результаты примера 3.1, определяем коэффициенты прямой трудоёмкости:

$$t_1 = 1160/775{,}3 = 1{,}5; \quad t_2 = 460/510{,}1 = 0{,}9; \quad t_3 = 875/729{,}6 = 1{,}2.$$

2. По формуле (3.13), где B - матрица коэффициента полных материальных затрат, вычисленная в примере 3.1, находим коэффициенты полной трудоёмкости:

$$\overline{T} = (1{,}5;\ 0{,}9;\ 1{,}2) \times \begin{bmatrix} 2{,}041 & 0{,}612 & 1{,}020 \\ 0{,}816 & 2{,}245 & 0{,}408 \\ 0{,}867 & 0{,}510 & 1{,}684 \end{bmatrix} = (4{,}84;\ 3{,}55;\ 3{,}92).$$

3. Умножая строки 1-3 первого и второго квадрантов межотраслевого материального баланса, построенного на примере 3.1, на соответствующие коэф-

фициенты прямой трудоёмкости, **получим искомую схему межотраслевого баланса труда** (в трудовых измерителях). Результаты расчётов представлены в табл. 3.4.

Таблица 3.4

Производящие отрасли	Потребляющие отрасли				
	Межотраслевые затраты овеществлённого труда			Затраты труда на конечную продукцию	Затраты труда в отраслях (трудовые ресурсы)
	1	2	3		
1	348,9	76,5	437,7	300,0	1163,0
2	139,6	229,5	0,0	90,0	459,1
3	279,1	61,2	175,1	360,0	875,5

3.3 Модели управления запасами

Современные (логистические) технологии в области управления запасами, применяемые производителями направлены в основном на минимизацию материальных запасов. Примерами таких систем являются следующие методы:

- МРП (Materials Requirements Planing) – планирование потребности в материалах – система планирования производственных ресурсов.
- «Канбан» – метод, обеспечивающий оперативное регулирование количества произведенной продукции на каждой стадии поточного производства.
- «Джаст ин тайм» (Just-in-time) – «точно вовремя» – общий организационный подход, с помощью которого, в результате учитывающего детали спроса, точного управления, значительно сокращаются запасы и тем самым длительность производственного цикла.
- ОПТ – (Optimized Production Technologies) – оптимизированные производственные технологии.
- ДРП (Distribution Requirements Planing) – система управления и планирования распределения продукции.

Последние новшества в сфере производства таковы: дифференциация продукции на возможно более поздней стадии производства на базе использования максимально однотипных комплектующих; использование выгод массового производства не на стадии сборки, на стадии изготовления комплектующих изделий; стремление к максимальному удовлетворению потребностей клиента на этапе выбора товара для производства. Все это требует гибкости производства на цеховом уровне, достигаемой как за счет расширения возможностей по переналадке оборудования, так и благодаря применению новых методов управления запасами - «Канбан» и «точно в срок».

Суть системы «Канбан» состоит в том, чтобы наличные запасы по своему количеству соответствовали потребностям начальной стадии производственного процесса, а не накапливались, как прежде.

Одним из методов сокращения запасов, повышения гибкости производства и возможности противостояния возрастающей конкуренции стал метод «точно в срок», получивший наибольшее распространение в США и странах Западной Европы. Сущность метода базируется на трех предпосылках.

Во-первых, предполагается, что заявкам потребителей готовой продукции должны соответствовать не её предварительно накопленные запасы, а производственные мощности, готовые перерабатывать сырье и материалы, поступающие почти «с колес». Вследствие этого объем производственных запасов, квалифицируемый как замороженные мощности, минимизируется. *Во-вторых*, в условиях минимальных запасов необходима непрерывная рационализация организации и управления производством, ибо высокий объем запасов смягчает последствия от упущений в области менеджмента, нивелирует неиспользуемые производственные мощности, ненадежную работу поставщиков и посредников. *В-третьих*, для оценки эффективности производственного процесса, помимо уровня затрат и производительности фондов, следует учитывать срок реализации заявки, так называемую длительность полного производственного цикла. Короткие сроки реализации заявок облегчают управление предприятием и способствуют росту его конкурентоспособности благодаря возможности оперативного и гибкого реагирования на изменения внешних условий.

В противоположность традиционным методам управления, при методе «точно в срок» централизованное планирование касается только последнего звена логистической цепи, т. е. склада готовой продукции. Все другие производственные и снабженческие единицы получают распоряжения непосредственно от звена, находящегося ближе к концу логистической цепи. К примеру, склад готовых изделий дал заявку (что равнозначно выдаче производственного задания) на определенное число изделий в монтажный цех, монтажный цех отдает распоряжение об изготовлении подузлов цехам обработки.

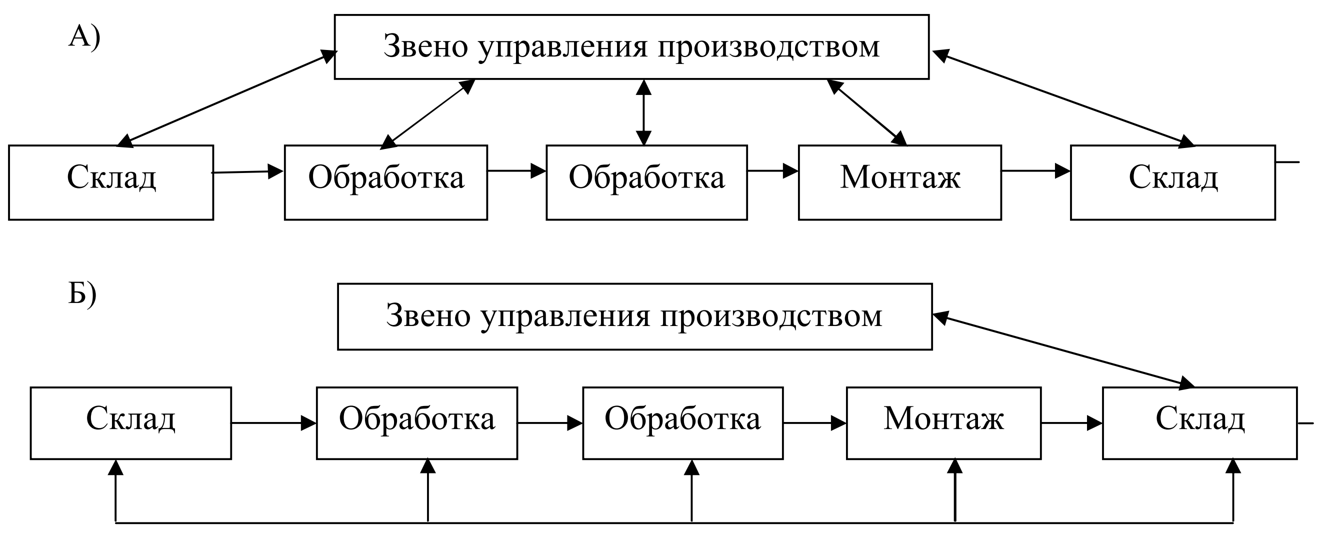


Рис. 3.1 - Схема управления производством: А - в традиционной системе; Б - в системе «точно в срок»

Это означает, что производственное задание всегда выдается потребителю, использующему (или обрабатывающему) данную деталь. Тем самым материалопоток от «источника» к «потребителю» предваряется потоком информации в обратном направлении, т. е. производству «точно в срок» предшествует информация «точно в срок».

Практика показывает, что для эффективного внедрения стратегии «точно в срок» необходимо изменение способа мышления целого коллектива, занимающегося вопросами производства и сбыта. Традиционный стереотип мышления типа «чем больше, тем лучше» должен быть заменён схемой «чем меньше, тем лучше», если речь идет об уровне запасов, использовании производственных мощностей, продолжительности производственного цикла или о величине партии продукции.

Результаты анализа, проведенного по внедрению концепции «точно в срок» на западноевропейских предприятиях, являются многообещающими. Усредненные данные, полученные более чем на 100 обследованных объектах (отдельные проекты функционируют на фирмах непрерывно от 2 до 5 лет), таковы:

• запасы незавершенного производства сократились более чем на 80%;

• запасы готовой продукции снизились примерно на 33%;

• объем непроизводственных запасов (материалов и полученных по кооперации деталей) колебался от 4 часов до 2 дней по сравнению с 5-15 днями до внедрения метода «точно в срок»;

• продолжительность производственного цикла (срок реализации заданий всей логистической цепи) сократилась примерно на 40%;

• производственные издержки снизились на 10-20%;

• значительно повысилась гибкость производства.

Затраты, связанные с подготовкой и внедрением стратегии «точно в срок», относительно невелики и окупались, как правило, уже через несколько месяцев функционирования этой системы. Использование стратегии «точно в срок» дает и другие выгоды, в том числе неэкономического характера. Использование принципов системы «точно в срок» оказывает также положительное влияние на долгосрочную инвестиционную политику предприятия, которая в данном случае отдает предпочтение машинам и оборудованию, связанным с гибкой автоматизацией производственных, транспортных и контрольных процессов.

В течение последних 15 лет в промышленно развитых странах было разработано множество моделей, имеющих отношение к различным вопросам управления запасами. При помощи моделирования доказывается эффективность применяемых мер в процессе производства или выполнения производственной программы, поскольку могут быть измерены периоды прохождения продукта через всю технологическую линию. При помощи моделирования можно также проверить проекты гибких производственных участков, обслуживаемых автоматическими транспортными средствами, оценить затраты на материально-техническое снабжение производства. Проектирование складов с применением компьютера дает возможность получить информацию об их оптимальной системе, величине необходимых Капиталовложений и затратах на эксплуатацию складов.

Фирмы часто используют математические модели для выбора уровня запасов путем балансирования затрат на подготовительные операции или расходов на выполнение заказа и сопоставления затрат при дефиците запасов с затратами на их хранение. Затраты на хранение запасов включают в себя не только затраты на содержание пасов на складе, издержки вследствие порчи продукции и морального износа, но и упущенную выгоду, т. е. норму прибыли, которую можно было бы получить, используя другие возможности инвестирования при эквивалентном риске.

Один из вариантов снижения риска при хранении запасов - использование технологий, основанных на внедрении систем гибкого производства, его роботизации. В данном случае преимуществом касается сокращение времени и затрат на подготовительные операций. Это делает экономически выгодным изготовление изделий небольшими партиями, что особенно важно в условиях жесткой конвенции и постоянных изменений требований рынка. При этом временно существенно снижается и риск морального устаревания запасов.

Таблица 3.5. Сравнение современных систем управления запасами

Название системы	Основные преимущества	Недостатки
MRP - Manufacturing Recourse Planning, США	Снижение издержек производства за счет уменьшения складских заделов, сокращения сроков изготовления продукции и их соблюдение. Размеры снижения запасов на складах составляют в среднем 20% и более.	Непременное условие - точность исходных данных, что требует реорганизации информационных систем. Система не обеспечивает координированного учета множества требований, поступающих извне, а поэтому не позволяет действительно объективно определить оптимальные размеры партий готовой продукции. В результате пропадает возможность достоверно оценивать степень эффективности принимаемых решений и реальные размеры экономии от принимаемых вариантов поставок и использования материальных ресурсов.
MAP-Material Availability Planning, США	Система реального обеспечения материальными ресурсами, т.к. материальное планирование осуществляется на базе дискретного потока данных относительно фактически поступающих заказов на поставку	Сильная зависимость от факторов внешнего окружения в части поставки материальных ресурсов - их стоимости, сроков поставки, а также «перекрестного» воздействия множества факторов и требований, которые в системе MAP считаются неоп-

	продукции. Процесс определения размеров партий и продуктовой структуры выпуска продукции приобретает динамичный характер и протекает под воздействием оценки главного фактора: величины затрат на материальные ресурсы, поступающих из внешних источников.	ределенными.
Kanban - Канбан, Япония	- Канбан, Япония Система оперативного планирования производственных запасов и материальных потоков между отдельными производственными операциями. Главное правило - межоперационная поставка исключительно доброкачественных бездефектных деталей и полуфабрикатов. Цель внедрения - исключить запасы и незавершенное производство, обеспечить большую гибкость производства, возможность приспособления к изменяющимся требованиям рынка.	План производства определенного количества деталей и полуфабрикатов на каждой предшествующей технологической стадии определяется заданием производственного участка, выполняющим последующую стадию при данной производственной программе предприятия. Канбан можно определить как вытягивающую (в отличие от MRP и MAP, которые являются подталкивающими) систему планирования, информация в которой идет от конечной точки непосредственного производства к предыдущему участку работы.
Jast-in-time - «точно в срок», Япония	Сырье, полуфабрикаты, комплектующие подаются небольшими партиями непосредственно в нужные точки производственного процесса, минуя складские помещения; готовая продукция также отгружается потребителям непосредственно по мере завершения производства. Основной принцип данной системы - выработка и поставка продукции точно в заданный срок и не ранее. В результате -сокращение производственных запасов и расстояний транспортировки.	Необходимое условие - выравнивание производства, т.е. если для какого-либо процесса производства детали будут поступать в разные промежутки времени или неравными по количеству партиями, то на предшествующем этапе производства должно быть задействовано столько оборудования и рабочей силы, чтобы можно было удовлетворить максимальный спрос. Реализация системы невозможна при централизованном планировании, необходимое условие - внедрение системы канбан.

Далее рассмотрим систему управления запасами - планирование материальных потребностей (MRP), которая характеризуется зависимым спросом и используется для прогнозирования зависимого спроса. Изделие имеет зависимый спрос, если его использование прямо связано с планом производства др. изделий. Например, спрос на автомобили зимой зависит от выпуска шипованных шин, или спрос на лекарства зависит от эпидемии, спрос на хирургические инструменты зависит от частоты выполнения операций.

Цель планирования материальных потребностей *заключается* в том, чтобы иметь в запасах только то, что непосредственно требуется для выполнения планов текущего производства.

Система планирования материальных потребностей нуждается в информации трех видов. Для примера предположим, что производитель детских трехколесных велосипедов пользуется системой планирования материальных потребностей для управления запасами колес определенного размера. Предположим также, что анализ потребности в таких колесах проводится в конце февраля. При этом потребуются три вида исходных данных:

1. План производств велосипедов с колесами данного типа. Допустим, предприятие намерено произвести 1000 велосипедов в третью неделю апреля.
2. Спецификация материалов для велосипедов с указанием деталей и их количества, требующегося для сборки одного велосипеда. В нашем случае на каждый велосипед требуется три колеса.
3. Инвентаризационные данные по данной позиции с зависимым спросом. В частности, необходимо знать:

- количество, имеющееся в запасах на данный момент. Например, имеется 150 колес;
- заказанное количество и ожидаемый срок получения заказа. Допустим, у поставщика заказано 1800 колес, ожидаемый срок прибытия заказа – вторая неделя марта;
- время реализации заказа. Колеса обычно поступают через две недели после размещения заказа на поставку.

Анализ при планировании потребности проходит в три этапа:

1. Суммарная потребность (или позиция) рассчитывается на основе плана производства и спецификации материалов. В нашем случае в третьей неделе апреля потребуется 3000 колес (3 колеса на один велосипед * 1000 велосипедов).

2. Чистая потребность рассчитывается путем вычитания из показателя суммарной потребности количества, имеющегося в наличии, и заказанного количества со сроком поставки, отвечающим плану производства. Поскольку 150 колес имеется на складе и 1800 заказанных колес прибудут в марте, чистая потребность на третью неделю апреля составит 1050 колес (3000 1950).

3. С учетом сроков реализации заказов планируется время размещения заказа таким образом, чтобы удовлетворить чистую потребность к планируемой дате начала производства. Поскольку срок выполнения заказа на колеса составляет две недели, то заказ на 1050 колес должен быть размещен в первую неделю апреля.

В примере со сборкой велосипеда система планирования материальных потребностей достаточно проста.

Практическое использование и реализация системы планирования материальных потребностей в условиях производства, когда требуются сотни и даже тысячи различных наименований, представляет определенные трудности.

При планировании материальных потребностей используются три вида исходных данных: спецификация материалов (комплектующих), требующихся для изготовления продукции; инвентаризационные данные по этому виду материалов, включающие количество, имеющееся на данный момент, заказанное количество, ожидаемый срок получения и время реализации заказа.

Из вышесказанного следует, что управление запасами с зависимым спросом значительно проще и подчинено планам производства. Сложность может быть вызвана только широкой номенклатурой и ассортиментом запасов, которая решается при наличии информационной системы на предприятии.

Достоинством MRP является способность своевременно и точно осуществлять репланирование (во многом благодаря тому, что система компьютеризирована). Способность такой системы учитывать происходящие изменения известна под названием восстанавливающее MRP . Восстанавливающее MRP использует целую программу с представлением всех вычислений, позволяющих получить новый план чистых потребностей. Следует отметить, что внесение всех изменений не всегда целесообразно, так как частое внесение изменений приводит к нервозности в работе системы. Производственные менеджеры должны оценивать значимость и последствия изменения прежде чем вносить его в MRP

Основной целью работы MRP является производство такого количества, которое необходимо без хранения на складе и ожидания дальнейших заказов. Имеется множество путей определения размера партий изделий, деталей и узлов в MRP. Наиболее часто используются следующие методы размерно-объемных расчетов.

• партия за партией – производится точно столько, сколько требуется, при этом заказы и хранения запасов равны, а затраты переналадок зависят от количества переналадок, что отражается в плане чистой потребности в материалах;

• размер экономического заказа – EOQ более предпочтительно использовать там, где существует относительно постоянный независимый спрос, не требующий изучения, такой подход усредняет спрос в пределах рассматриваемого временного горизонта;

• последовательное балансирование по отдельным периодам (PPB) – более динамичный подход к выравниванию затрат переналадки и хранения PPB использует дополнительную информацию с учетом представлений о величине запасов в будущем. PPB пытается сбалансировать затраты переналадки на основе данных о спросе. Ключевым здесь является понятие об отдельном экономичном периоде (EPP), который измеряется отношением затрат на переналадку к затратам на хранение. PPB будет стремиться к некоторому увеличению потребности так, чтобы число отдельных периодов аппроксимировало EPP.

• алгоритм Вагнера – Витина является моделью динамического программирования, добавляющей некоторую сложность в расчет размера партии. Эта модель предполагает наличие временного горизонта, за которым отсутствует дополнительная чистая потребность. Такая техника часто используется на практике, но она связана с большими интеллектуальными затратами и требует глубоких знаний в области программирования.

Следует отметить, что в реальности все размеры партий, рассчитанные на основе приведенных методов не точны, так как производственная система не в состоянии быстро реагировать на частые изменения. На практике наиболее эффективен метод партия за партией, так как размер партии может быть изменен с учетом множества различных ограничений. Этот метод обеспечивает наиболее экономичные результаты. Однако там, где затраты переналадки значительны и спрос более или менее постоянен, удовлетворительные результаты обеспечивают метод PPB, алгоритм Вагнера - Витина, а также EOQ техника.

Безусловно, система планирования запасов MRP , учитывающая зависимый спрос, дает множество преимуществ. К ним относятся:

- возможность поддерживать низкий уровень материальных запасов производства;
- возможность отслеживать материально-производственные потребности;
- возможность оценивать данные по материальным потребностям производства, полученные из конкретного контрольного графика производственного процесса;
- возможность распределения времени и сроков производства.

Более совершенной и развитой является система планирования потребности материалов с обратной связью, которая обеспечивает обратной связью производственное планирование и систему управления запасами. Система объединяет обратную связь план по мощности, производственный график и достаточно удаленное во времени планирование производства.

Важную роль в планировании производства с обратной связью играют загрузочные рапорты, которые показывают потребности ресурсов для всех текущих назначений в соответствии с планом и ожидаемыми распоряжениями в каждом рабочем центре.

Система MRP с обратной связью позволяет планировщику перераспределить работу между временными периодами, чтобы сгладить загрузку или, по крайней мере, разбросать ее в пределах мощности. Затем можно перерасписать обработку всех элементов плана, определяющего чистую потребность. Существуют следующие тактики сглаживания загрузки и минимизации изменений времен длительности обработки:

1. Запараллеливание – перекрытие исполнения операций с различной полнотой перекрытия, которое понижает время обработки и основано на передаче отдельных единиц на следующую операцию, не ожидая окончания обработки всей партии на предыдущей операции.

2. Операционное расщепление предусматривает размещение партии на обработку на двух различных станках, выполняющих одну и ту же операцию. Это увеличивает суммарное время переналадки, но в результате израсходованное время уменьшается, поскольку обработку на каждом станке проходит только часть первоначальной партии.

3. Расщепление партии приводит к нарушению установленного порядка движения партии в соответствии с расписанием обработки по ходу технологического процесса.

Прогнозировать независимый спрос сложнее, чем зависимый. Для управления запасами с независимым спросом применяются две системы управления:

1. Система с фиксированным количеством заказа;
2. Система с фиксированным интервалом времени.

В системах с *фиксированным количеством заказа* постоянно контролируется уровень запасов. Когда количество падает ниже установленного уровня, выдается заказ на пополнение запасов. Заказывается всегда одно и то же количество. Таким образом, фиксированными величинами в этой системе является уровень, при котором повторяется заказ, и заказываемое количество. Системы с фиксированными количествами заказа являются наиболее приемлемыми для запасов со следующими характеристиками:

- высокая стоимость предметов снабжения;
- высокие издержки хранения материально-технических запасов;
- высокий уровень ущерба, возникающего в случае отсутствия запасов;
- скидка с цены в зависимости от заказываемого количества;
- относительно непредсказуемый или случайный характер спроса.

Привлекательность такой системы заказа заключается в простоте механизма ее действия. Главный недостаток применения системы состоит в том, что заказ производится без изучения ожидаемой потребности. Может сложиться такая ситуация, что еще долго после того, как сделан заказ, потребность в нем не возникнет, и в результате запас не потребляется. Или наоборот: спрос все возрастает и не может быть удовлетворен имеющимся в наличии запасом. Заказ с твердо установленным количеством заказанного применяется только в тех случаях, когда суммы затрат на запас плюс затраты на заказ должны быть минимальными.

Система с фиксированным количеством заказа требует соблюдения следующих правил контроля:

• заказывать следующую партию можно в том случае, когда количество наличного запаса достигнет нижней точки заказа;

• необходимо заказывать оптимальный объем партии заказа;

• критерием оптимизации становится минимум совокупных затрат на хранение запасов и повторение заказа. Данный критерий учитывает три фактора, влияющих не величину совокупных затрат:

а) используемая площадь складских помещений;

б) издержки на хранение запасов;

в) стоимость оформления заказа.

В современной практике управления запасами на предприятиях активно используются различные модели с независимым спросом:

- модель экономического (по количеству) заказа (EOQ);
- модель производственного (по количеству) заказа;
- модель заказа с резервным запасом;

- модель с дисконтируемым количеством;
- системы с фиксированным периодом (BQ системы).

Цель большинства моделей управления запасами – это минимизировать суммарные затраты и свести к минимуму отрицательные последствия при накоплении и дефиците.

К главным затратам на управление относятся затраты на размещение, пополнение и хранение, остальные затраты носят постоянный характер.

Основная модель управления заказами (EOQ) предусматривает ряд допущений, например,

- спрос известен и постоянен;
- время между размещением заказа и его исполнением известно и постоянно;
- заказ поступает полностью, т.е. одной партией и в одно время;
- понижение (дисконт) количества невозможно;
- изменяются только затраты на перезаказ или размещение заказа;
- дефицит запасов исключен, если заказ размещен вовремя.

Суть модели экономического заказа (EOQ) заключается в одноразовом пополнении запаса и нулевом времени исполнения заказа. Заказ удовлетворяется в тот момент, когда на него поступила заявка и прибывает одновременно полностью, т.е. уровень запаса совершает прыжок от 0 до Q.

Пример. Предприятие при нулевом запасе на складе сделало заказ на 200 единиц комплектующих. Все комплектующие поступили одновременно и уровень запаса повысился от О до 200 ед. Так как спрос постоянен во времени, запас со склада убывает с постоянной скоростью, когда он снизится до О, выдается новый заказ. Такой процесс повторяется во времени постоянно. Схематически, (см. рис. 3.2)

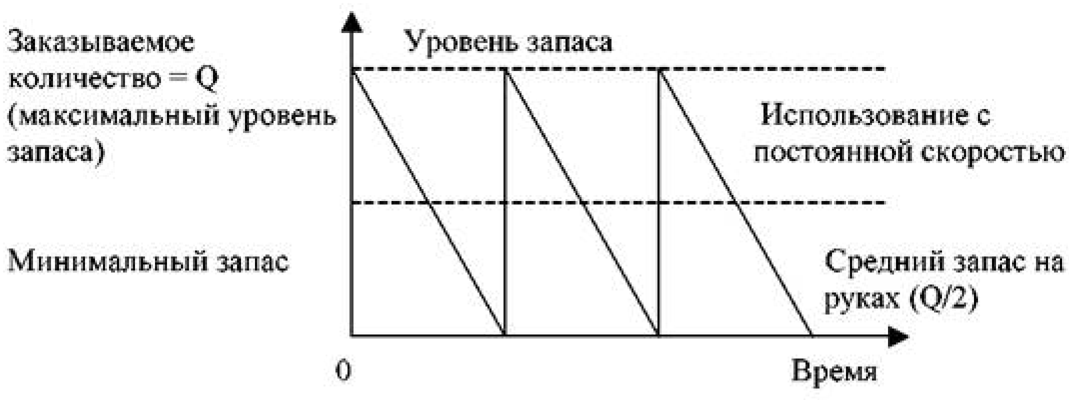


Рис. 3.2 - Модель экономического заказа

К достоинствам EOQ модели следует отнести ее надежность, так как она дает положительный результат даже при значительном изменении параметров. При этом следует отметить, что установление точной цены заказа и затрат хранения запасов часто затруднительно. Удобство этой модели также в том, что общие затраты изменяются незначительно.

После определения оптимальной величины заказа, необходимо определить, когда заказывать. Простые модели управления запасами исходят из того, что получение заказа должно быть немедленным, то есть заказывать нужно в тот момент, когда уровень запаса достигнет 0. Однако время между размещением и получением заказа, называемое временем выполнения заказа или временем доставки, может составлять как несколько часов, так и несколько месяцев. Таким образом, решение о том, когда заказывать, выражаемое термином точка перезаказа, определяется уровнем запаса, по достижении которого должен быть размещен заказ.

Точка перезаказа (ROP) можно представить следующим равенством.

ROP = (Дневная потребность) * (d - время выполнения нового заказа в днях).

Рис. 3.3 - Точка перезаказа

Уравнение для ROP означает, что спрос, однороден и постоянен.

Ежедневный спрос d определяется отношением годового спроса (Д) деленного на число рабочих дней в году

$$d = \frac{Д}{T_{раб.дн.}}$$

Модель производственного (по количеству) заказа подходит для использования в производственном процессе, когда запасы наращиваются постепенно и показатель экономического уровня заказа уже предположительно установлен. Эта модель подтверждает, что оптимальная величина заказа Q обеспечена равенством затрат на заказ и хранение.

Модель заказа с резервным запасом применяется, когда на предприятии возрастает спрос на материалы и удается избежать их дефицита, используя страховой запас. Модели, отражающие такое состояние производства, называются моделями заказа с резервным запасом или моделями, планирующими нехватку запаса. На рис. 3.4 показан уровень запаса как функция времени.

Чтобы увеличить объемы продаж, многие поставщики предлагают своим партнерам дисконтирование по количеству. Количественный дисконт (скидка)

– это снижение цены единицы P , когда товар покупается в больших количествах. Типичное расписание количественного дисконта представлено в табл.3.6

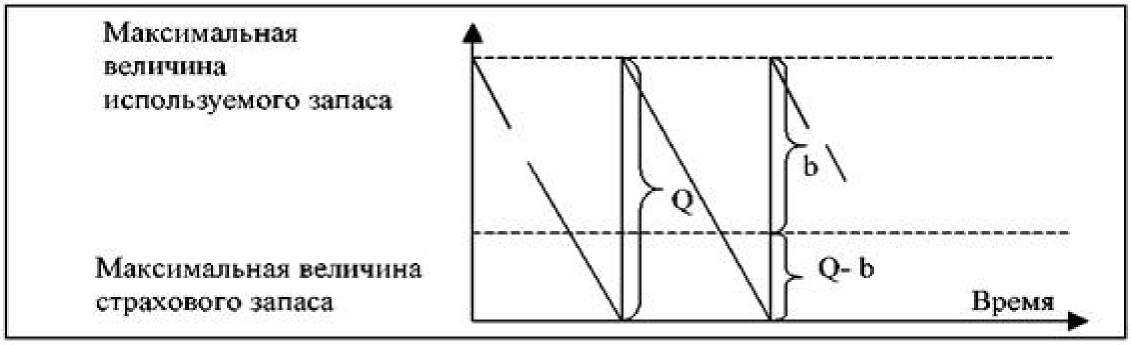


Рис. 3.4 - Модель заказа с резервным запасом

Таблица 3.6

Номер дисконта	Дисконтируемое количество	Дисконт, %	Дисконтная цена, P, $
1	От 0 до 999	0	5,00
2	1000- 1999	4	4,80
3	2000 и выше	5	4,75

В рассмотренных выше моделях запасов основная цель заключалась в минимизации общих затрат. Поскольку стоимость единицы для третьего дисконта из таблицы является наименьшей, может появиться искушение сделать заказ в 2000 ед. или больше, чтобы сэкономить на понижении цены изделия. Однако при этом можно не достичь минимизации общих затрат на запасы, так как при увеличении количества заказа растут затраты на хранение. Наибольший эффект достигается, когда значение количественного дисконта рассматривается между понижающейся стоимостью продукта и увеличивающимися затратами на хранение. С включением затрат на приобретение продукта в расчет уравнение, определяющее общие годовые затраты примет вид:

$$TC = DS/Q + QH/2 + PD.$$

Здесь: D - годовой спрос в единицах; S - затраты заказа или переналадки;
P - цена единицы изделия; H - затраты хранения единицы за год.

Затем определяется количество, которое будет соответствовать минимальным общим годовым затратам. Рассмотрим следующий пример.

Предприятие пользуется дисконтными скидками для оптовых покупателей. Дисконтное расписание представлено в табл. 3.6. Затраты заказа составляют $ 49 на заказ, годовой спрос равен 5000 ед. товара, и текущие затраты запаса изменяются в проценте от стоимости I, который равен 20 %. Определим, какое заказываемое количество минимизирует общие затраты запаса.

1. Для каждого значения дисконта рассчитываем величину Q *, используя следующее уравнение: $Q^* = sgr(2 \cdot D \cdot S / I \cdot P)$.

Здесь затраты хранения (H = IP) выражены в виде процента I от цены единицы продукта P вместо того, чтобы рассматривать их как постоянную величину, приходящуюся на единицу продукта в год H.

Q_1* = sqr (2(5000)(49)/(0.2)/(5.00)) = 700 ед.

Q_2* = sqr (2(5000)(49)/(0.2)/(4.80)) = 714 ед.

Q_3* = sqr (2(5000)(49)/(0.2)/(4.75)) = 718 ед.

2. Для любого дисконта, если заказываемое количество слишком мало, чтобы быть дисконтированным, изменим заказываемое количество в сторону его увеличения до ближайшей минимальной величины, которую затем можно будет продисконтировать, Q_1* находится между 0 и 999, поэтому оно не должно быть увеличено. Q_2* находится ниже значений заказов, входящих в диапазон от 1000 до 1999, поэтому оно должно быть увеличено до 1000. То же можно сказать и о Q_3*, оно должно быть увеличено до 2000.

Если заказываемое количество меньше ранжируемого количества, соответствующего дисконтированию, то необходимо иметь ввиду, что ранжируемое количество при соответствующем ему дисконте обеспечивает и более низкие общие затраты.

3. Произведем расчет для всех заказываемых количеств, используя уравнения общих затрат. Результат представлен в табл. 3.7 .

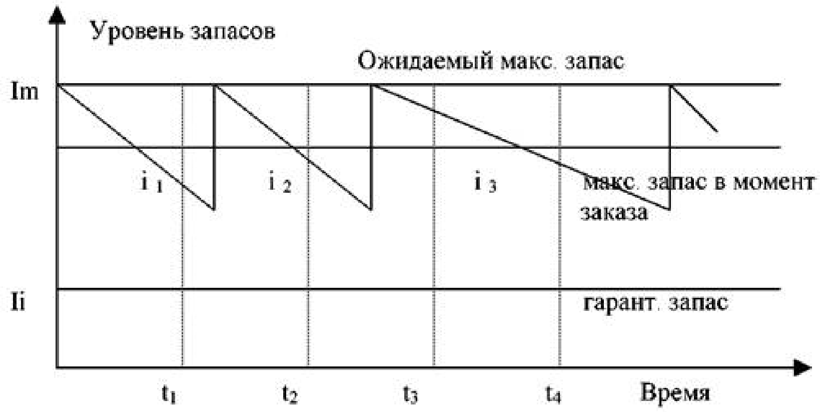
Таблица 3.7

Номер дисконта	Цена единицы, $	Заказываемое количество	Годовые затраты на товар, $	Годовые затраты на заказ, $	Годовые затраты хранения $	Общие затраты, $
1	5.00	700	25000	350	350	25700
2	4.80	1000	24000	245	480	24725
3	4.75	2000	23750	122.5	950	24822.5

4. Отберем то Q *, которое соответствует самым низким общим затратам. Согласно данным табл. 3.7, это заказ, равный 1000 ед.

Рассмотрим систему управления запасами с независимым спросом **с фиксированным интервалом времени**, характеризующуюся тремя параметрами: Im максимальный ожидаемый запас; Ii максимальный запас в момент заказа на пополнение запаса; t период контроля запасов.

Суть системы управления запасами с фиксированным интервалом времени заключается в том, что устанавливается некоторая промежуточная величина запаса Ii таким образом, если в контрольной точке (t_1 , t_2 , t_3 , t_4 и т. д.) имеющийся запас заключен между Ii и максимально ожидаемым запасом Im, то заказ не производится. Заказы производятся только тогда, когда запас равен или меньше Ii. (см. рис. 3.5).



i_1, i_2, i_3 - точка заказа

t_1, t_2, t_3, t_4 – моменты времени в которые производится контроль запасов.

Рис. 3.5 - Управление запасами с фиксированным интервалом времени

В результате использования такой системы можно ожидать уменьшения количества заказов очень малых размеров при одновременном появлении нескольких довольно больших заказов. Сумма дополнительных затрат (по сравнению с партией оптимального размера) в случае заказов большими партиями не столь велика, как в случае заказов малыми партиями.

Вторая система (с фиксированным интервалом времени между заказами) предусматривает следующую последовательность операций подготовки заказа:

• устанавливается периодичность контроля запасов, ориентированная на график поставок поставщика;

• рассчитывается величина требуемых запасов как сумма количеств, продаваемых за период контроля запасов и за время ожидания поставки и количе ств в страховом запасе;

• составляется и выполняется график проведения контроля уровня запасов;

• принимается решение о размере заказа – указывается количество деталей;

• заказ высылается в установленный графиком день.

В системе с фиксированным интервалом времени заказы на восполнение размещаются с заданной периодичностью, например, один раз в две недели. Заказываемое количество непостоянно и зависит от имеющегося остатка. Эта система более подходит для предметов материально-технического снабжения со следующими характеристиками:

• малоценные предметы;

• низкие затраты на хранение материально-технических запасов;

• незначительные издержки, если даже запасы закончились;

• один из многих предметов, закупаемых у одного и того же поставщика;

• скидка с цены зависит от стоимости заказов на несколько предметов.

Системы с фиксированным интервалом времени применяются, например, при управлении запасами канцелярских товаров или бакалейных продуктов в магазине.

При системе с фиксированным интервалом времени между заказами устанавливаются строго определенные сроки представления заказа, тем самым решается вопрос когда? Сначала нужно ответить на вопрос сколько? Для этого устанавливается и фиксируется в карточках учета или памяти компьютера величина требуемого запаса для каждой детали. Наличный запас плюс ожидаемое пополнение по предыдущему заказу становятся достаточными для удовлетворения спроса до следующего пополнения запаса:

Требуемый запас = Количество, расходуемое за период контроля + Количество, расходуемое за период поставки + Страховой запас.

Из этого следует формула: **РЗ=ТЗ - (НЗ+ОП)**, где РЗ - размер заказа; ТЗ - требуемый запас; НЗ - наличный запас; ОП - ожидаемое пополнение.

При работе по системе пополнения запасов через фиксированные интервалы времени ничто не препятствует использованию оптимальных заказов, особенно для деталей, имеющих стабильный спрос в период между поставками. Оптовики стараются точнее рассчитывать и своевременно корректировать величины требуемых запасов деталей – это способствует точности заказов и минимизации расходов по закупкам.

3.4 Контрольные вопросы и задачи

1. Раскройте основную идею балансовой модели «затраты – выпуск».
2. В чём заключаются допущения, вводимые при построении МОБ?
3. Дайте характеристику основных понятий МОБ.
4. Поясните принципиальную схему межотраслевого баланса.
5. Раскройте экономическое содержание показателей прямой и полной трудоёмкости продукции.
6. Какие существуют проблемы управления запасами?
7. Расскажите о системах управления запасами
8. Как рассчитать оптимальный размер запасов?
9. Как рассчитать страховой запас?
10. Назовите основные параметры различных систем управления запасами.

11. **Задание 1**. Известно, что затраты на выполнение заказа (поставку единицы продукции) составляют 15 ден. ед.; годовое потребление 12000 ед.; годовые затраты на хранение продукции 0,1 ден. ед.; размер партии поставки: 100, 200, 400, 600, 800, 1000 ед.; годовое производство 15000 ед.; издержки, обусловленные дефицитом составляют 0,4 ден. ед. ***Требуется:***

а) вычислить оптимальный размер закупаемой партии и построить график;

б) определить оптимальный размер заказываемой партии при пополнении заказа на конечный интервал.

в) рассчитать оптимальный размер партии в условиях дефицита.

12. **Задание 2.** Известно, что годовой спрос составляет 10000 ед., затраты, связанные с доставкой продукции, равны 20,0 дол/ед; цена единицы продукции составляет 1,4 дол/ед.; затраты на содержание запасов равны 40% от цены ед. продукции.

Определить:

1. Оптимальный размер партии поставки.
2. Цену, которую должен установить поставщик при поставке продукции партиями по 450 ед.
3. Оптимальный размер производимой партии на предприятии при производстве 150000 ед. в год.

ГЛАВА 4. СТОХАСТИЧЕСКИЕ МОДЕЛИ

4.1 Случайные процессы. Простейший поток событий

Пусть имеется некоторая физическая система *S,* которая с течением времени меняет свое состояние (переходит из одного состояния в другое), причем заранее неизвестным, случайным образом. Тогда мы будем говорить, что в системе S протекает *случайный процесс*.

Случайным процессом $X(t)$ называется процесс, значение которого при любом фиксированном $t = t_0$ является случайной величиной $X(t_0)$. Случайная величина $X(t_0)$, в которую обращается случайный процесс при $t = t_0$, называется *сечением случайного процесса*, соответствующим данному значению аргумента t.

Под «физической системой» можно понимать что угодно: офис банка, конвейер, информационно-справочную систему, торговую точку, страховую компанию, предприятие, отрасль промышленности, выставочный зал и т. д.

Большинству процессов, протекающих в реальных экономических системах, свойственны, в той или иной мере, черты случайности, неопределенности.

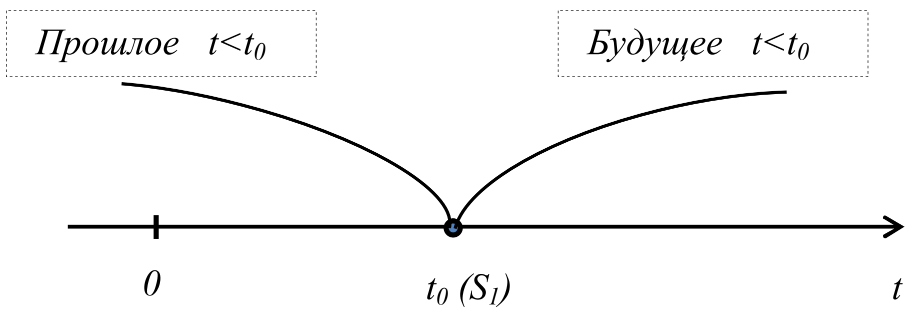


Рис. 4.1

Случайный процесс, протекающий в системе, называется ***марковским***, если *для любого момента времени t_0 вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент t_0 и не зависят от того, когда и как система пришла в это состояние.*

Пусть в настоящий момент t_0 (см. рис. 4.1) система находится в определенном состоянии S_o. Мы наблюдаем процесс со стороны и в момент t_0 знаем состояние системы S_o и всю предысторию процесса, все, что было при $t < t_0$. Нас, естественно, интересует будущее ($t > t_0$).

Для марковского случайного процесса «вероятностное предсказание» оказывается гораздо проще, чем для немарковского. Если процесс - марковский, то предсказывать можно, учитывая только настоящее состояние системы S_0 и забыв о его «предыстории» (поведении системы при $t < t_0$). Само состояние S_o, разумеется, зависит от прошлого, но как только оно достигнуто, характеристики системы в прошлом можно не учитывать. Другими словами, в марковском процессе «будущее зависит от прошлого только через настоящее».

На практике марковские процессы в чистом виде обычно не встречаются, они являются идеализацией реальных процессов. Однако, на практике нередко приходится иметь дело с процессами, для которых влиянием «предыстории» можно пренебречь. При количественном анализе таких процессов полезными являются *марковские модели*.

В исследовании операций большое значение имеют так называемые марковские случайные процессы с дискретными состояниями и непрерывным временем. Процесс называется *процессом с дискретными состояниями*, если его возможные состояния $S_1, S_2, S_3, \ldots$ можно заранее перечислить (перенумеровать), и переход системы из состояния в состояние происходит «скачком», практически мгновенно. Процесс называется *процессом с непрерывным временем*, если моменты возможных переходов из состояния в состояние не фиксированы заранее, а неопределенны, случайны, если переход может осуществиться, в принципе, в любой момент. Будем рассматривать только процессы с дискретными состояниями и непрерывным временем.

Пример 4.1. Автоматическое устройство сортировки грузов S на складе состоит из двух узлов, каждый из которых в случайный момент времени может выйти из строя (отказать), после чего мгновенно начинается ремонт узла, тоже продолжающийся заранее неизвестное, случайное время. Возможные состояния системы можно перечислить: S_1 - оба узла исправны, S_2 - первый узел ремонтируется, второй исправен, S_3 - второй узел ремонтируется, первый исправен, S_4 - оба узла ремонтируются.

Переходы системы S из состояния в состояние происходят практически мгновенно, в случайные моменты выхода из строя того или другого узла или окончания ремонта.

При анализе случайных процессов с дискретными состояниями удобно пользоваться геометрической схемой - *графом состояний*. Состояния системы изображаются прямоугольниками (или кругами), а возможные переходы из состояния в состояние - стрелками, соединяющими состояния. Далее будем изображать состояния системы прямоугольниками, в которых записаны обозначения состояний: $S_1, S_2, S_3, \ldots, S_n$.

Построим граф состояний для рассмотренного выше примера (см. рис. 4.2). Стрелка, направленная из S_1 в S_2, означает переход в момент отказа первого узла; стрелка, направленная обратно, из S_2 в S_1,- переход в момент окончания ремонта этого узла. Остальные стрелки объясняются аналогично.

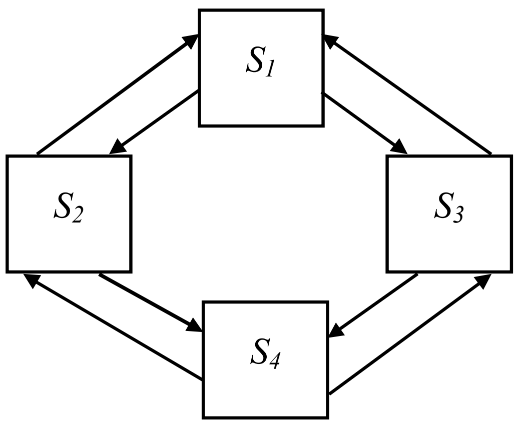


Рис. 4.2

Если процесс, протекающий в системе с дискретными состояниями и непрерывным временем, является марковским, то для его описания можно построить довольно простую математическую модель.

Потоком событий называется последовательность однородных событий, следующих одно за другим в какие-то случайные моменты времени. Например: поток заказов на доставку товаров; поток железнодорожных составов, поступающих на сортировочную станцию; поток квартир, выставляемых на продажу и т.д. Поток событий можно наглядно изобразить рядом точек на оси времени O_t (рис. 4.3); при этом положение каждой из них случайно, и на рисунке изображена только какая-то одна реализация потока.

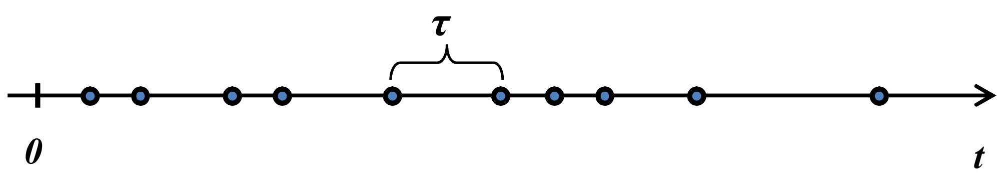


Рис. 4.3

В теории вероятностей «событием» (или «случайным событием») называется какой-то исход опыта, обладающий той или другой вероятностью. События, образующие поток, сами по себе вероятностями не обладают; вероятностями обладают другие, производные от них события, например: «на участок времени τ (рис. 4.3) попадет ровно два события», или «на участок времени Δt попадет хотя бы одно событие», или «промежуток времени между двумя соседними событиями будет не меньше *t»*.

Важной характеристикой потока событий является его *интенсивность* λ - среднее число событий, приходящееся на единицу времени. Интенсивность потока может быть как постоянной (λ = const), так и переменной, зависящей от времени *t.* Например, поток посетителей супермаркета в предвыходные дни интенсивнее, чем в будние дни.

Поток событий называется *регулярным*, если события следуют одно за другим через определенные, равные промежутки времени. На практике чаще встречаются потоки не регулярные, со случайными интервалами.

Поток событий называется *стационарным*, если его вероятностные характеристики не зависят от времени. В частности, интенсивность λ стационарного потока должна быть постоянной. Это отнюдь не значит, что фактическое число событий, появляющееся в единицу времени, постоянно, - нет, поток неизбежно (если только он не регулярный) имеет какие-то случайные сгущения и разрежения, как, например, показано на рис. 4.3. Важно, что для стационарного потока эти сгущения и разрежения не носят закономерного характера: на один участок единичной длины может попасть больше, а на другой - меньше событий, но среднее число событий, приходящееся на единицу времени, постоянно и от времени не зависит.

Как правило, отклонения от стационарности могут быть объяснены какими-то физическими причинами. Например, естественно, интенсивность потока клиентов салона мод, утром меньше, чем днем и вечером. Увеличение интенсивности потока покупателей, приходящих в магазин, в часы после окончания рабочего дня тоже имеет физическое объяснение. Если поток событий имеет тенденцию к явно выраженным сгущениям и разрежениям (особенно периодическим), можно сформулировать гипотезу о закономерности и постараться найти физическую причину, объясняющую данную закономерность.

На практике часто встречаются потоки событий, которые (по крайней мере на ограниченном участке времени) могут считаться стационарными. Например, поток вызовов, поступающих к диспетчеру такси между 10 и 12 часами, практически стационарен; тот же поток в течение суток уже не стационарен.

Поток событий называется *потоком без последействия*, если для любых двух непересекающихся участков времени τ_1 и τ_2 (см. рис. 4.4) число событий, попадающих на один из них, не зависит от того, сколько событий попало на другой. По сути это означает, что события, образующие поток, появляются в те или другие моменты времени независимо друг от друга, вызванные каждое своими собственными причинами. Например, поток пассажиров, входящих в метро, практически не имеет последействия. А вот поток покупателей, отходящих от прилавка с купленными товарами, уже имеет последействие (хотя бы потому, что интервал по времени между отдельными покупателями не может быть меньше, чем минимальное время t_0 обслуживания каждого из них).

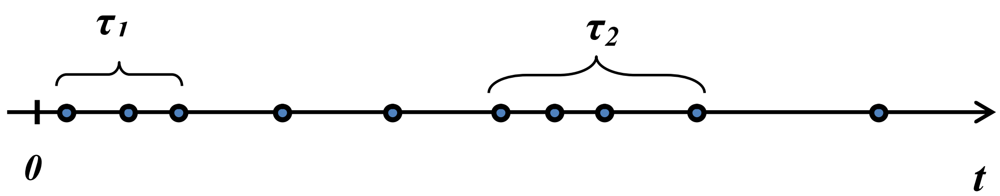


Рис. 4.4

Аналогично обстоит дело и с потоком поездов, подходящих к станции (между ними всегда существует какой-то минимальный интервал t_0, выбираемый из соображений безопасности). Впрочем, если минимальный интервал между событиями много меньше среднего интервала между ними $\bar{t} = 1/\lambda$,, иногда наличием последействия можно пренебречь.

Поток событий называется *ординарным*, если события в нем появляются поодиночке, а не группами по нескольку сразу. Например, поток юридических лиц, обращающихся за услугами в консалтинговую фирму, обычно ординарен, чего нельзя сказать о потоке клиентов, направляющихся в банк в первой декаде месяца для получения пенсии. Поток поездов, подходящих к станции, ординарен, а поток вагонов - неординарен. Если поток событий ординарен, то вероятностью попадания на малый участок времени Δt двух или более событий можно пренебречь.

Поток событий называется ***простейшим*** (или *стационарным пуассоновским*), если он обладает сразу *тремя свойствами: стационарен, ординарен и не имеет последействия.* Название «простейший» связано с тем, что процессы, связанные с простейшими потоками, имеют наиболее простое математическое описание. Отметим, самый простой, на первый взгляд, регулярный поток не является «простейшим», так как обладает последействием: моменты появления событий в таком потоке связаны жесткой, функциональной зависимостью. Без специальных усилий по поддержанию его регулярности такой поток обычно не создается.

Простейший поток играет среди других потоков особую роль, в чем-то подобную роли нормального закона среди других законов распределения. А именно, при наложении (суперпозиции) достаточно большого числа независимых, стационарных и ординарных потоков (сравнимых между собой по интенсивности) получается поток, близкий к простейшему.

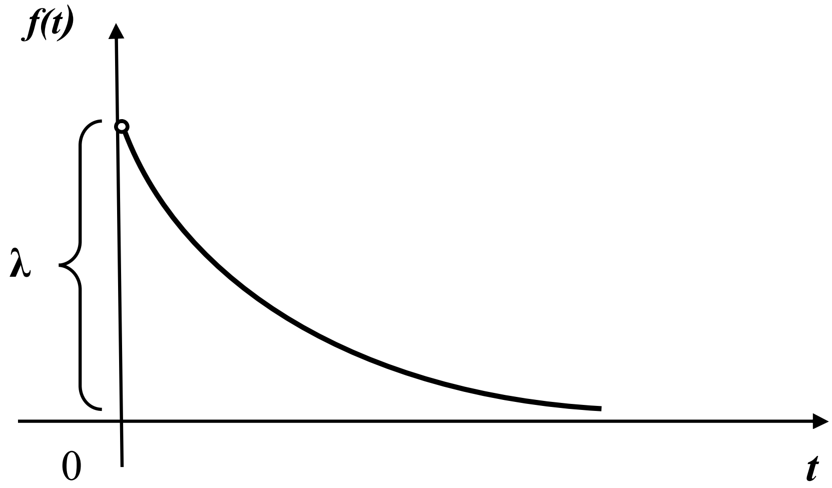


Рис. 4.5

Известно, что для простейшего потока с интенсивностью λ интервал T между соседними событиями имеет так называемое показательное распределение с плотностью (см. рис. 4.5)

$$f(t)=\lambda\cdot e^{-\lambda t} \quad (t>0) \tag{4.1}$$

Величина λ в формуле (4.1) называется параметром показательного закона. Для случайной величины T, имеющей показательное распределение, математическое ожидание $m=m_T$ есть величина, обратная параметру λ, а среднее квадратическое отклонение $\sigma_т$ равно математическому ожиданию (МОЖ):

$$m_T=\sigma_T=1/\lambda\ . \tag{4.2}$$

В теории вероятностей в качестве «меры случайности» неотрицательной случайной величины нередко рассматривают так называемый коэффициент вариации:

$$v_T=\sigma_T/m_T\ . \tag{4.3}$$

Из формул (4.2), (4.3) следует, что для показательного распределения v_T= 1, т. е. для простейшего потока событий коэффициент вариации интервалов между событиями равен единице.

Очевидно, что для регулярного потока событий, у которого интервал между событиями не случаен ($\sigma_т$ = 0), коэффициент вариации равен нулю. Для большинства потоков событий, встречающихся на практике, коэффициент вариации интервалов между событиями заключен между нулем и единицей и может служить некоторой мерой «степени регулярности» потока: чем v_T ближе к нулю, тем «регулярнее» поток. Простейший поток - это «наименее регулярный» из встречающихся на практике потоков.

В расчетах, связанных с потоками событий, удобно пользоваться понятием «элемента вероятности». Рассмотрим на оси Ot простейший поток с интенсивностью λ и произвольно расположенный участок времени Δt. *Элементом вероятности* называется вероятность попадания на этот участок хотя бы одного события потока. Легко доказать, что элемент вероятности (с точностью до малых величин более высокого порядка по сравнению с Δt) равен

$$p_{\Delta t} = \lambda \cdot \Delta t, \tag{4.4}$$

т.е. для простейшего потока элемент вероятности равен интенсивности потока, умноженной на длину элементарного участка. Заметим, что элемент вероятности, в силу отсутствия последействия, совершенно не зависит от того, сколько событий и когда появлялись ранее.

Отметим, что с точностью до величин высшего порядка малости вероятность появления хотя бы одного события на элементарном участке Δt равна вероятности появления ровно одного события на этом участке. Это вытекает из ординарности простейшего потока.

Поток событий называется *рекуррентным* (иначе - «*потоком Пальма*»), если он стационарен, ординарен, а интервалы времени между событиями $T_1, T_2, T_3, \ldots$ (рис. 4.6) представляют собой независимые случайные величины с одинаковым произвольным распределением (например, с плотностью, показанной на рис.4.7). Приведем пример рекуррентного потока. Транспортный робот на конвейере работает непрерывно до своего отказа (выхода из строя); отказавший элемент оперативно заменяется новым. Если отдельные экземпляры элемента выходят из строя независимо друг от друга, то поток отказов (он же поток «замен» или «восстановлений») будет рекуррентным.

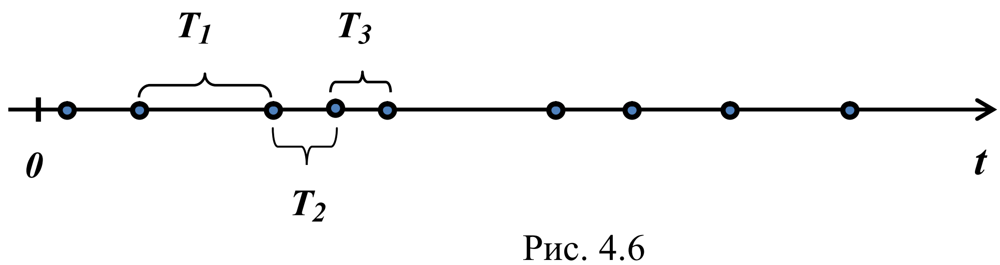


Рис. 4.6

Другой пример. Менеджер торгового зала непрерывно занят обслуживанием покупателей (как это бывает в часы максимальной нагрузки). Обслуживание покупателя продолжается случайное время T. Тогда поток обслуженных покупателей будет рекуррентным (если считать, что времена обслуживания отдельных покупателей независимы и, например, настойчивость одного из клиентов не скажется на времени обслуживания других).

Очевидно, простейший поток представляет собой частный случай рекуррентного потока, когда интервалы между событиями имеют показательное распределение (4.1). Другим частным (вырожденным) случаем рекуррентного потока является регулярный поток событий, где интервалы вообще не случайны, постоянны.

Множество рекуррентных потоков событий, обладающих разной степенью упорядоченности, можно получить «просеиванием» простейшего потока. Пусть, например, в банк поступает простейший поток посетителей, а у входа стоит менеджер, направляющий первого посетителя - к первому оператору, второго - ко второму оператору и т. д. Если число операторов равно n, то к каждому из них поступает так называемый «поток Эрланга n-го порядка» клиентов. Такой поток получается из простейшего, если сохранять в потоке каждое n-е событие, а промежуточные - отбрасывать. Простейший поток есть не что иное, как поток Эрланга первого порядка. Можно показать, что при таком просеивании простейшего потока коэффициент вариации интервалов уменьшается, и при увеличении порядка n поток Эрланга приближается к регулярному. Коэффициент вариации интервалов между событиями потока Эрланга n-го порядка равен $v_T^{(n)} = 1/\sqrt{n}$. Потоки Эрланга образуют целую гамму потоков с различной степенью упорядоченности - от «полного беспорядка» (простейший поток) до полной упорядоченности (регулярный поток).

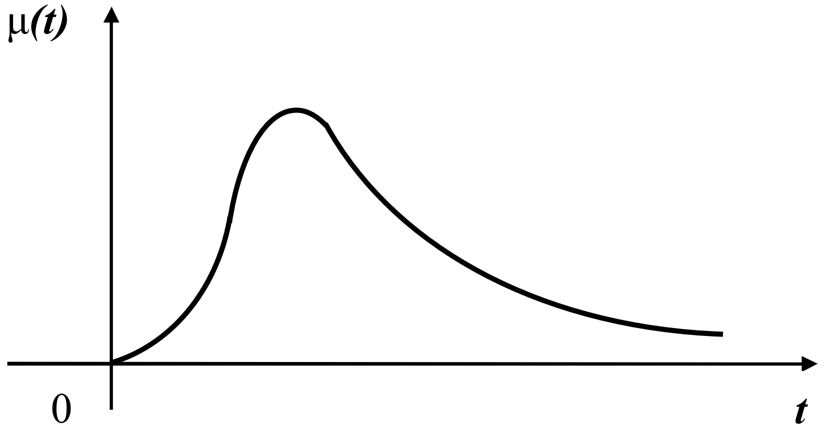


Рис. 4.7

4.2 Вероятностные модели динамических систем

При рассмотрении марковских процессов с дискретными состояниями и непрерывным временем удобно представить себе, что все переходы системы S из состояния в состояние происходят под действием каких-то потоков событий (поток вызовов, поток отказов, поток восстановлений и т. д.). Если все потоки событий, переводящие систему S из состояния в состояние,- простейшие, то процесс, протекающий в системе, будет марковским. Это обусловлено тем, что простейший поток не обладает последействием: в нем «будущее» не зависит от «прошлого».

Если система S находится в каком-то состоянии S_i, из которого есть непосредственный переход в другое состояние S_j (стрелка, ведущая из S_i в S_j на графе состояний), то будем представлять так, как будто на систему, пока она находится в состоянии S_i, действует простейший поток событий, переводящий её по стрелке $S_i \rightarrow S_j$. Как только появится первое событие этого потока, происходит «перескок» системы из состояния S_i в состояние S_j.

Для наглядности удобно на графе состояний у каждой стрелки указывать интенсивность того потока событий, который переводит систему по данной стрелке. Обозначим через $\lambda_{i\,j}$ интенсивность потока событий, переводящего систему из состояния S_i в S_j. На рис. 4.8 показан граф состояний с проставленными у стрелок интенсивностями; такой *граф* называют *размеченным*.

Построим размеченный граф состояний для промышленного робота, состоящего из двух узлов. Пусть выделены четыре состояния системы: S_0 - оба узла исправны, S_1 - первый узел ремонтируется, второй исправен, S_2 - второй узел ремонтируется, первый исправен, S_3 - оба узла ремонтируются.

Интенсивности потоков событий, переводящих систему из состояния в состояние, будем вычислять, предполагая, что среднее время ремонта узла не зависит от того, ремонтируется ли один узел или оба сразу.

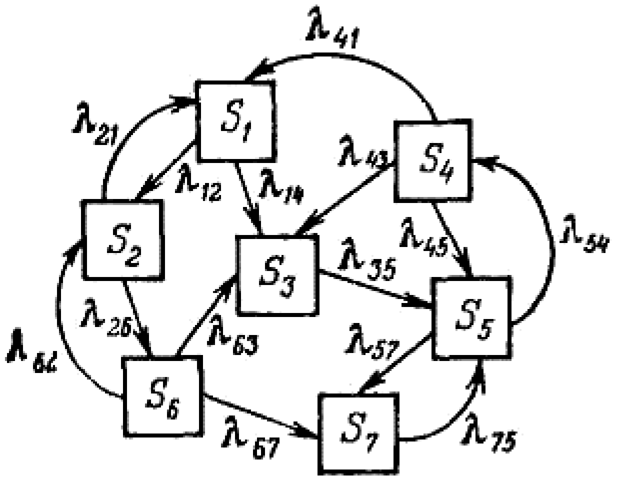


Рис. 4.8

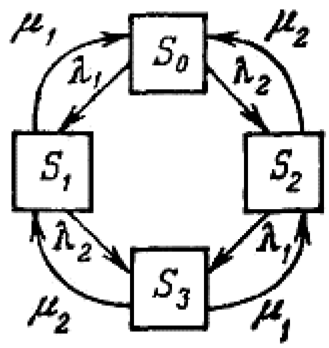


Рис. 4.9

Это будет так, если ремонтом каждого узла занят отдельный специалист. Найдем все интенсивности потоков событий, переводящих систему из состояния в состояние. Пусть система находится в состоянии S_1. Какой поток событий переводит ее в состояние S_2? Очевидно, поток отказов первого узла. Его интенсивность λ_1 равна единице, деленной на среднее время безотказной работы первого узла. Очевидно, что поток «окончаний ремонтов» первого узла переводит систему обратно из S_2 в S_1. Его интенсивность μ_1 равна единице, деленной на среднее время ремонта первого узла. Аналогично вычисляются интенсивности потоков событий, переводящих систему по всем стрелкам графа рис. 4.9.

Имея в своем распоряжении размеченный граф состояний системы, можно построить математическую модель данного процесса.

Пусть рассматривается система S, имеющая n возможных состояний S_1, S_2, ..., S_n. *Указанные состояния образуют полную группу событий.* Назовем вероятностью i-го состояния вероятность $p_i\ (t)$ того, что в момент t система будет находиться в состоянии S_i. Очевидно, что для любого момента времени t сумма всех вероятностей состояний равна единице:

$$\sum_{i=1}^{n} p_i(t) = 1 \qquad (4.5)$$

С использованием размеченного графа состояний можно найти все вероятности состояний $p_i\ (t)$ как функции времени. Для этого составляются и решаются так называемые *уравнения Колмогорова* - особого вида дифференциальные уравнения, в которых неизвестными функциями являются вероятности состояний.

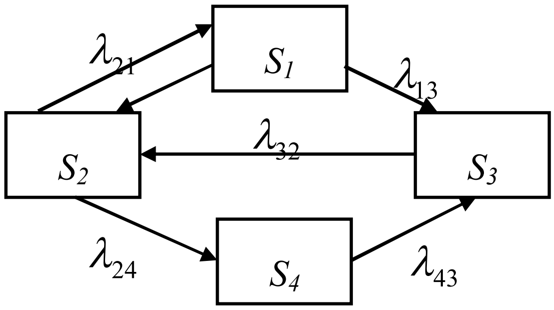


Рис. 4.10

Покажем на конкретном примере, как эти уравнения составляются. Пусть система S имеет четыре состояния: $S_1, S_2, S_3, S_4,$ размеченный граф которых показан на рис. 4.10. Рассмотрим одну из вероятностей состояний, например $p_1(t)$. Это - вероятность того, что в момент t система будет в состоянии S_1. Придадим t малое приращение Δt и найдем $p_1(t + \Delta t)$ - вероятность того, что в момент $(t + \Delta t)$ система будет в состоянии S_1. Это может произойти двумя способами:

1) в момент t система уже была в состоянии S_1, а за время Δt не вышла из него;
2) в момент t система была в состоянии S_2, а за время Δt перешла из него в S_1.

Найдем вероятность первого варианта. Вероятность того, что в момент t система была в состоянии S_1, равна $p_1(t)$. Эту вероятность нужно умножить на вероятность того, что, находившись в момент t в состоянии S_1, система за время Δt не перейдет из него ни в S_2, ни в S_3. Суммарный поток событий, выводящий систему из состояния S_1, тоже будет простейшим, с интенсивностью $\lambda_{12} + \lambda_{13}$ (при наложении - суперпозиции - двух простейших потоков получается опять простейший поток, так как свойства стационарности, ординарности и отсутствия последействия сохраняются). Значит, вероятность того, что за время Δt система выйдет из состояния S_{1t} равна $(\lambda_{12} + \lambda_{13})\Delta t$; вероятность того, что не выйдет: $1 - (\lambda_{12} + \lambda_{13})\Delta t$. Отсюда вероятность первого варианта равна $p_1(t) \cdot [1 - (\lambda_{12} + \lambda_{13})\ \Delta t]$.

Найдем вероятность второго варианта. Она равна вероятности того, что в момент t система будет в состоянии S_2, а за время Δt перейдет из него в состояние S_1, т. е. она равна $p_2(t)\lambda_{21}\Delta t$.

Складывая вероятности обоих вариантов (по правилу сложения вероятностей), получим:

$$p_1(t + \Delta t) = p_1(t) \cdot [1 - (\lambda_{12} + \lambda_{13}) \cdot \Delta t] + p_2(t) \cdot \lambda_{21} \cdot \Delta t \ .$$

Раскроем квадратные скобки, перенесем $p_1(t)$ в левую часть и разделим обе части на Δt:

$$\frac{p_1(t+\Delta t)-p_1(t)}{\Delta t}=\lambda_{21}p_2(t)-(\lambda_{12}+\lambda_{13})\,p_1(t).$$

Устремим Δt к нулю и слева получим в пределе производную функции $p_1(t)$. Таким образом, запишем дифференциальное уравнение для $p_1(t)$:

$$\frac{dp_1(t)}{dt}=\lambda_{21}\cdot p_2(t)-(\lambda_{12}+\lambda_{13})\cdot p_1(t)\,. \qquad (4.6)$$

Рассуждая аналогично для всех остальных состояний, напишем еще три дифференциальных уравнения. Присоединяя к ним уравнение (4.6), получим систему дифференциальных уравнений для вероятностей состояний:

$$\left.\begin{aligned}
\frac{dp_1}{dt}&=\lambda_{21}\cdot p_1-(\lambda_{12}+\lambda_{13})p_1\,;\\
\frac{dp_2}{dt}&=\lambda_{12}\cdot p_1+\lambda_{32}p_3-(\lambda_{24}+\lambda_{21})p_2\,;\\
\frac{dp_3}{dt}&=\lambda_{31}\cdot p_1+\lambda_{43}\cdot p_4-\lambda_{32}\cdot p_3\,;\\
\frac{dp_4}{dt}&=\lambda_{24}p_2-\lambda_{43}\cdot p_4\,.
\end{aligned}\right\} \qquad (4.7)$$

Это система четырех линейных дифференциальных уравнений с четырьмя неизвестными функциями p_1, p_2, p_3, p_4. Заметим, что одно из них можно отбросить, пользуясь тем, что $p_1+p_2+p_3+p_4=1$: выразить любую из вероятностей p_i через другие, это выражение подставить в (4.7), а соответствующее уравнение с производной $\frac{dp_i}{dt}$ отбросить.

Сформулируем общее правило составления уравнений Колмогорова. В левой части каждого из них стоит производная вероятности какого-то (i-го) состояния. В правой части - сумма произведений вероятностей всех состояний, *из которых идут стрелки в данное состояние,* на интенсивности соответствующих потоков событий, минус суммарная интенсивность всех потоков, *выводящих систему из данного состояния,* умноженная на вероятность данного (i-го) состояния.

Пользуясь этим правилом, запишем уравнения Колмогорова для системы S, размеченный граф состояний которой дан на рис. 4.9:

$$\left.\begin{aligned}
\frac{dp_1}{dt}&=\mu_1p_2+\mu_2p_3-(\lambda_1+\lambda_2)p_1;\\
\frac{dp_2}{dt}&=\lambda_1p_1+\mu_2p_4-(\lambda_2+\mu_1)p_2;\\
\frac{dp_3}{dt}&=\lambda_2p_1+\mu_1p_4-(\lambda_1+\mu_2)p_3;\\
\frac{dp_4}{dt}&=\lambda_2p_2+\lambda_1p_3-(\mu_1+\mu_2)p_4.
\end{aligned}\right\} \qquad (4.8)$$

Чтобы решить уравнения Колмогорова и найти вероятности состояний, прежде всего, надо задать начальные условия. Если точно известно начальное состояние системы S_i, то в начальный момент (при $t = 0$) $p_i(0) = 1$, а все остальные начальные вероятности равны нулю. Так, например, уравнения (4.8) естественно решать при начальных условиях $p_0(0) = 1$, $p_1(0) = p_2(0) = p_3(0) = 0$.

Линейные дифференциальные уравнения с постоянными коэффициентами можно решать аналитически, это удобно в случаях, когда число состояний (и, соответственно, уравнений) не превосходит двух. Если количество уравнений в модели больше трёх, для их решения применяют унифицированные компьютерные программы, реализующие численные методы интегрирования. Таким образом, уравнения Колмогорова дают возможность найти вероятности выделенных состояний как функций времени.

Уточним, что будет происходить с вероятностями состояний при $t \to \infty$? Будут ли $p_1(t)$, $p_2(t)$,... стремиться к каким-то пределам? Если эти пределы существуют и не зависят от начального состояния системы, то они называются *финальными вероятностями состояний*. В теории случайных процессов доказывается, что *если число n состояний системы конечно, и из каждого из них можно (за конечное число шагов) перейти в любое другое, то финальные вероятности существуют*.

Предположим, что это условие выполнено и финальные вероятности существуют:

$$\lim_{t\to\infty} p_i(t) = p_i, \quad i = 1,2,\ldots,n.. \tag{4.9}$$

Финальные вероятности обозначают через p_1, p_2, …., как и сами вероятности состояний, но понимая под ними уже не переменные величины (функции времени), а постоянные числа. Очевидно, что эти вероятности образуют в сумме единицу:

$$\sum_{i=1}^{n} p_i = 1. \tag{4.10}$$

При $t \to \infty$ в системе S устанавливается предельный стационарный режим, в ходе которого система случайным образом меняет свои состояния, но их вероятности уже не зависят от времени. Финальную вероятность состояния S_i можно истолковать как среднее относительное время пребывания системы в этом состоянии. Например, если система S имеет три состояния S_1 S_2, S_3 и их финальные вероятности равны 0,2, 0,3 и 0,5, это значит, что в предельном, стационарном режиме система в среднем две десятых времени проводит в состоянии S_1, три десятых - в состоянии S_2 и половину времени - в состоянии S_3.

Если вероятности p_1, p_2,... постоянны, то их производные равны нулю. Значит, *чтобы найти финальные вероятности, нужно все левые части в уравнениях Колмогорова положить равными нулю и решить полученную систему* уже не дифференциальных, а *линейных алгебраических уравнений*. Можно и не писать уравнений Колмогорова, а прямо по графу состояний написать систему линейных алгебраических уравнений. Если перенести отрицательный член каждого уравнения из правой части в левую, то получим сразу систему уравнений, где слева стоит финальная вероятность данного состояния p_i, умноженная на суммарную интенсивность всех потоков, ведущих из данного состояния, а справа - сумма произведений интенсивностей всех потоков, входящих в i-е состояние, на вероятности тех состояний, из которых эти потоки исходят.

Пользуясь этим правилом, напишем линейные алгебраические уравнения для финальных вероятностей состояний системы, граф состояний которой дан на рис. 4.9:

$$\left.\begin{aligned} (\lambda_1 + \lambda_2)\, p_0 &= \mu_1 p_1 + \mu_2 p_2, \\ (\lambda_2 + \mu_1)\, p_1 &= \lambda_1 p_0 + \mu_2 p_3, \\ (\lambda_1 + \mu_2)\, p_2 &= \lambda_2 p_0 + \mu_1 p_3, \\ (\mu_1 + \mu_2)\, p_3 &= \lambda_2 p_1 + \lambda_1 p_2. \end{aligned}\right\} \qquad (4.11)$$

Эту систему четырех уравнений с четырьмя неизвестными p_0, p_1, p_3, p_4, можно решить аналитически. Однако, уравнения (4.11) однородны (не имеют свободного члена) и, значит, определяют неизвестные только с точностью до произвольного множителя. Воспользуемся так называемым *нормировочным условием*:

$$p_1 + p_2 + p_3 + p_4 = 1 \qquad (4.12)$$

и с его помощью решим систему уравнений. При этом одно (любое) из уравнений можно отбросить (оно вытекает как следствие из остальных).

Зададимся численными значениями интенсивностей $\lambda_1 = 1$, $\lambda_2 = 2$, $\mu_1 = 2$, $\mu_2 = 3$ и решим систему (4.11). Заменим четвертое уравнение нормировочным условием (4.12). Уравнения примут вид:

$$\begin{cases} 3 \cdot p_1 = 2 \cdot p_2 + 3 \cdot p_3; \quad 4 \cdot p_2 = p_1 + 3 \cdot p_4; \\ 4 \cdot p_3 = 2 \cdot p_1 + 2 \cdot p_4; \quad p_2 + p_2 + p_3 + p_4 = 1. \end{cases} \qquad (4.13)$$

Решая уравнения совместно, получим:

$$p_1 = 0{,}40; \quad p_2 = 0{,}20; \quad p_3 = 0{,}27; \quad p_4 = 0{,}13.$$

Отсюда следует, что в предельном, стационарном режиме система S в среднем 40% времени будет проводить в состоянии S_1 (оба узла исправны), 20% - в состоянии S_2 (первый узел ремонтируется, второй работает), 27% - в состоянии S_3 (второй узел ремонтируется, первый работает) и 13% - в состоянии S_4 . Знание финальных вероятностей может помочь оценить среднюю эффективность работы системы и загрузку ремонтных органов.

Предположим, что система S в состоянии S_1 (полностью исправная) приносит в единицу времени доход 8 (условных единиц), в состоянии S_2 - доход 3, в состоянии S_3 - доход 5, в состоянии S_4 - вообще не приносит дохода. Тогда в предельном, стационарном режиме средний доход в единицу времени будет

$$D = 0{,}40 \cdot 8 + 0{,}2 \cdot 3 + 0{,}27 \cdot 5 = 5{,}15\,.$$

Теперь оценим загрузку ремонтных органов, занятых ремонтом узлов 1 и 2. Узел 1 ремонтируется долю времени, равную $p_2 + p_4 = 0{,}20 + 0{,}13 = 0{,}33$. Узел 2 ремонтируется долю времени $p_3 + p_4 = 0{,}40$.

4.3 Системы массового обслуживания

4.3.1 Сущность и классификация систем массового обслуживания

На практике часто приходится сталкиваться с работой систем организационного управления, в которых имеют место повторяющиеся процессы обслуживания однородных объектов. Такие системы в исследовании операций называют **системами массового обслуживания** (СМО). В качестве примеров СМО можно указать: офис банка, в котором обслуживаются клиенты, справочные бюро, супермаркеты и т.п.

Каждая СМО состоит из какого-то числа обслуживающих единиц (или «приборов»), которые называют **каналами обслуживания**. Каналами могут быть: линии связи, рабочие точки, кассиры, продавцы, лифты, автомашины и др. В зависимости от числа каналов обслуживания различают одноканальные и многоканальные СМО. Всякая СМО предназначена для обслуживания потока заявок (или «требований»), поступающих в случайные моменты времени. Обслуживание заявки продолжается, в общем случае, случайное время $T_{об}$, после чего канал освобождается и готов к приему следующей заявки. Процесс работы СМО представляет собой случайный процесс с дискретными состояниями и непрерывным временем; состояние СМО меняется скачком в моменты появления каких-то событий (или прихода новой заявки, или окончания обслуживания, или момента, когда заявка, которой надоело ждать, покидает очередь).

СМО можно рассматривать как аналитическую модель реально существующей системы в терминах теории массового обслуживания. Исследования, проводимые на такой модели, имеют целью изучение процесса функционирования реальной системы, определение характеристик системы, которые обеспечивают заданное качество её функционирования. СМО включает в себя (рис.4.11) входящий поток требований (заявок, вызовов, клиентов), нуждающихся в «обслуживании», и механизм (алгоритм), осуществляющий это «обслуживание».

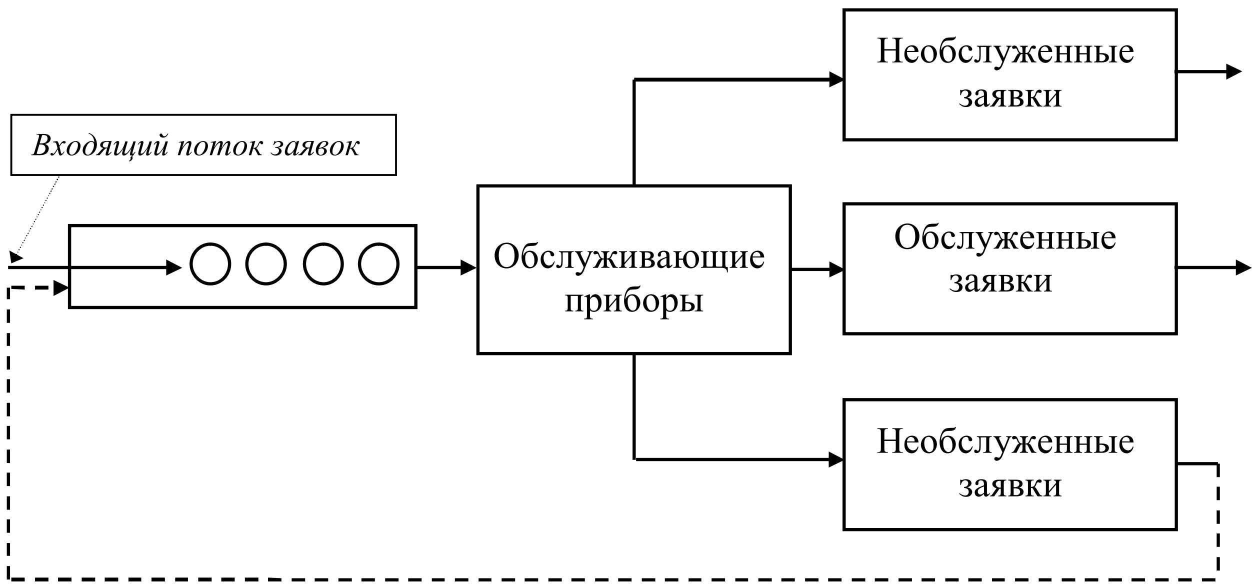


Рис. 4.11 – Укрупненная структура СМО

Теория массового обслуживания (ТМО) представляет собой раздел теории вероятностей, изучающий потоки требований на обслуживание, поступающие в системы обслуживания и выходящие из них, а также характеристики обслуживания и их зависимость от правил (дисциплины) обслуживания. Целью исследования являются рациональный выбор структуры системы обслуживания и характеристик процесса обслуживания.

Предмет **теории массового обслуживания** - построение математических моделей, связывающих заданные условия работы СМО (число каналов, их производительность, правила работы, характер потока заявок) с интересующими нас характеристиками - показателями эффективности СМО, описывающими, с той или другой точки зрения, ее способность справляться с потоком заявок. В качестве таких показателей (в зависимости от обстановки и целей исследования) могут применяться разные величины, например: среднее число заявок, обслуживаемых СМО в единицу времени; среднее число занятых каналов; среднее число заявок в очереди и среднее время ожидания обслуживания; вероятность того, что число заявок в очереди превысит каков-то значение, и т. д. Среди заданных условий работы СМО специально не выделяют элементов решения: ими могут быть, например, число каналов, их производительность, режим работы СМО и т.д. Важно уметь решать прямые задачи ИО, а обратные - ставятся и решаются в зависимости от того, какие именно параметры нужно выбирать или изменять. Общей особенностью всех задач, рассматриваемых в ТМО, является случайный характер исследуемых явлений и наличие повторяющихся алгоритмов обработки заявок.

Математический анализ работы СМО существенно облегчается, если процесс работы представляет собой марковский процесс. Для этого достаточно, чтобы все потоки событий, переводящие систему из состояния в состояние (потоки заявок, потоки «обслуживании»), были *простейшими*. Если это свойство нарушается, то математическое описание процесса становится гораздо сложнее и довести его до явных, аналитических формул удается лишь в редких случаях. Аппарат марковской теории массового обслуживания полезен для приближенного описания работы СМО даже в тех случаях, когда потоки событий - не простейшие. Во многих случаях для принятия разумного решения по организации работы СМО вовсе и не требуется точного знания всех ее характеристик - зачастую достаточно и приближенного, ориентировочного.

Системы массового обслуживания делятся на типы (или классы) по ряду признаков. Первое деление: СМО с отказами и СМО с очередью. В СМО с отказами заявка, поступившая в момент, когда все каналы заняты, получает отказ, покидает СМО и в дальнейшем процессе обслуживания не участвует. Примеры СМО с отказами встречаются в телефонии: заявка на разговор, пришедшая в момент, когда все каналы связи заняты, получает отказ и покидает СМО необслуженной. В СМО с очередью заявка, пришедшая в момент, когда все каналы заняты, не уходят, а становится в очередь и ожидает возможности быть обслуженной. На практике чаще встречаются (и имеют большее значение) СМО с очередью; недаром теория массового обслуживания имеет второе название: «теория очередей».

СМО с очередью подразделяются на разные виды, в зависимости от того, как организована очередь - ограничена она или не ограничена. Ограничения могут касаться как длины очереди, так и времени ожидания (так называемые «СМО с нетерпеливыми заявками»). При анализе СМО должна учитываться также и «дисциплина обслуживания» - заявки могут обслуживаться либо в порядке поступления (раньше пришла, раньше обслуживается), либо в случайном порядке. Нередко встречается так называемое обслуживание с приоритетом — некоторые заявки обслуживаются вне очереди. Приоритет может быть как абсолютным - когда заявка с более высоким приоритетом «вытесняет» из-под обслуживания заявку с низшим (например, пришедший в парикмахерскую клиент высокого ранга прогоняет с кресла обыкновенного клиента), так и относительным - когда начатое обслуживание доводится до конца, а заявка с более высоким приоритетом имеет лишь право на лучшее место в очереди.

Существуют СМО с так называемым многофазовым обслуживанием, состоящим из нескольких последовательных этапов или «фаз» (например, покупатель, пришедший в магазин, должен сначала выбрать товар, затем оплатить его в кассе, затем получить на контроле).

Кроме этих признаков, СМО делятся на два класса: «открытые» и «замкнутые». В открытой СМО характеристики потока заявок не зависят от того, в каком состоянии сама СМО (сколько каналов занято). В замкнутой СМО - зависят. Например, если один рабочий обслуживает группу станков, время от времени требующих наладки, то интенсивность потока «требований» со стороны станков зависит от того, сколько их уже неисправно и ждет наладки. Это - пример замкнутой СМО. Классификация СМО далеко не ограничивается приведенными их разновидностями, но мы ограничимся ими.

Оптимизация работы СМО может производиться с разных позиций: с точки зрения организаторов (или владельцев) СМО или с точки зрения обслуживаемых клиентов. С первой точки зрения желательно «выжать все, что возможно» из СМО и добиться того, чтобы ее каналы были предельно загружены. С точки зрения клиентов желательно всемерное уменьшение очередей, которые зачастую приводят к бессмысленной трате сил и времени и, в конечном итоге, к понижению производительности труда.

При решении задач оптимизации в теории массового обслуживания необходим системный подход, полное и комплексное рассмотрение всех последствий каждого решения. Например, с точки зрения клиентов СМО желательно увеличение числа каналов обслуживания; но ведь работу каждого канала надо оплачивать, что удорожает обслуживание. Построение математической модели позволяет решить оптимизационную задачу о выборе рационального числа каналов с учетом влияния всех существенных факторов. Поэтому в задачах массового обслуживания обычно не выделяют какого-либо одного показателя эффективности, а сразу формулируют эти задачи как многокритериальные.

4.3.2 Схема гибели и размножения. Формула Литтла

1. Схема гибели в размножения. Имея в распоряжении размеченный граф состояний, можно составить уравнения Колмогорова для вероятностей состояний, а затем записать и решить алгебраические уравнения для финальных вероятностей. Для некоторых частных случаев удается последние уравнения решить заранее, в буквенном виде. В частности, это удается сделать, если граф состояний системы представляет собой так называемую «*схему гибели и размножения*».

Граф состояний для схемы гибели и размножения имеет вид, показанный на рис. 4.12. Особенность этого графа в том, что все состояния системы можно вытянуть в одну цепочку, в которой каждое из средних состояний *(S_1, S_2, ..., S_{n-1})* связано прямой и обратной стрелкой с каждым из соседних состояний - правым и левым, а крайние состояния *(S_0, S_n)* - только с одним соседним состоянием. Термин «схема гибели и размножения» ведет начало от биологических задач, где подобной схемой описывается изменение численности популяции.

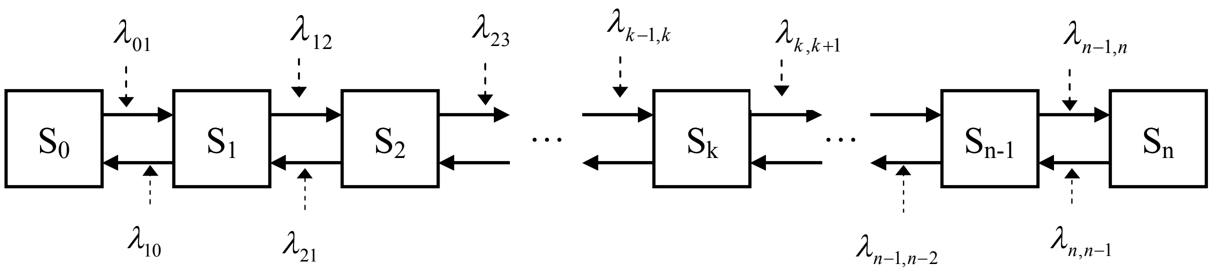


Рис. 4.12

Схема гибели и размножения достаточно часто встречается в разных задачах экономической практики. Опираясь на рекомендации теории массового обслуживания, полезно получить формулы для расчёта финальных вероятностей состояний.

Предположим, что все потоки событий, переводящие систему по стрелкам графа - простейшие. Для краткости будем называть и систему *S* и протекающий в ней процесс - простейшими.

Пользуясь графом рис. 4.12, составим и решим алгебраические уравнения для финальных вероятностей состояний. Их существование вытекает из того, что из каждого состояния можно перейти в каждое другое, и число состояний конечно.

Для первого состояния S_0 имеем:

$$\lambda_{01}\, p_0 = \lambda_{10}\, p_1 . \tag{4.14}$$

Для второго состояния *S_1*: $(\lambda_{12} + \lambda_{10})p_1 = \lambda_{01}p_0 + \lambda_{21}p_2$.

В силу (4.14) последнее соотношение приводится к виду $\lambda_{12}\, p_1 = \lambda_{21}\, p_2$,

далее, аналогично $\lambda_{23}\, p_2 = \lambda_{32}\, p_3$ и вообще $\lambda_{k-1,k} = \lambda_{k,k-1} p_k$, где *k* принимает все значения от 0 до *n.* Итак, финальные вероятности $p_0, p_1, ..., p_n$ удовлетворяют уравнениям

$$\left.\begin{array}{c}\lambda_{01}\,p_0 \;=\; \lambda_{10}\,p_1\\ \lambda_{12}\,p_1 = \lambda_{21}\,p_2\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots.\\ \lambda_{k-1,k}\,p_{k-1} = \lambda_{k,k-1}\,p_k\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ \lambda_{n-1,n}\,p_{n-1} = \lambda_{n,n-1}\,p_n\end{array}\right\} \qquad (4.15)$$

кроме того, надо учесть нормировочное условие $p_0+p_1+p_2+...+p_n=1$.

Решим эту систему уравнений. Из первого уравнения (4.15) выразим p_1 через p_0:

$$p_1=\frac{\lambda_{01}}{\lambda_{10}}\,p_0\,. \qquad (4.16)$$

Из второго, с учетом (4.16), получим:

$$p_2=\frac{\lambda_{12}}{\lambda_{21}}\,p_1=\frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}}\,p_0\,, \qquad (4.17)$$

из третьего, с учетом (4.17),

$$p_3=\frac{\lambda_{23}\lambda_{12}\lambda_{01}}{\lambda_{32}\lambda_{21}\lambda_{10}}\,p_0 \qquad (4.18)$$

и вообще, для любого k:

$$p_k=\frac{\lambda_{k-1,k}\ldots\lambda_{12}\lambda_{01}}{\lambda_{k,k-1}\ldots\lambda_{21}\lambda_{10}}\,p_0\,. \qquad (4.19)$$

Обратим внимание на формулу (4.19). В числителе стоит произведение всех интенсивностей, стоящих у стрелок, ведущих слева направо (с начала и до данного состояния S_k), а в знаменателе - произведение всех интенсивностей, стоящих у стрелок, ведущих справа налево (с начала и до S_0).

Таким образом, все вероятности состояний p_0, p_1, ..., p_n могут быть выражены через одну из них (p_0). Подставим эти выражения в нормировочное условие. Получим, вынося за скобку p_0:

$$p_0\left(1+\frac{\lambda_{01}}{\lambda_{10}}+\frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}}+\ldots+\frac{\lambda_{n-1,n}\ldots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\ldots\lambda_{21}\lambda_{10}}\right)=1\,.$$

Отсюда следует выражение для вероятности p_0:

$$p_0=\left(1+\frac{\lambda_{01}}{\lambda_{10}}+\frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}}+\ldots+\frac{\lambda_{n-1,n}\ldots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\ldots\lambda_{21}\lambda_{10}}\right)^{-1}. \qquad (4.20)$$

Все остальные вероятности выражены через p_0 /см. формулы (4.16)-(4.19)/. Заметим, что коэффициенты при p_0 в каждой из них представляют собой не что иное, как последовательные члены ряда, стоящего после единицы в формуле (4.20). Значит, при вычислении p_0 все эти коэффициенты будут определены.

Полученные выше формулы полезны при решении простейших задач теории массового обслуживания.

2. Формула Литтла. Теперь выведем формулу, связывающую (для предельного, стационарного режима) среднее число заявок $L_{сист}$, находящихся в системе массового обслуживания (т. е. обслуживаемых или стоящих в очереди), и среднее время пребывания заявки в системе $W_{сист}$.

Рассмотрим СМО (одноканальную, многоканальную, марковскую, немарковскую, с неограниченной или с ограниченной очередью) и связанные с нею два потока событий: поток заявок, прибывающих в СМО, и поток заявок, покидающих СМО. Если в системе установился предельный, стационарный режим, то среднее число заявок, прибывающих в СМО за единицу времени, равно среднему числу заявок, покидающих ее: оба потока имеют одну и ту же интенсивность λ.

Обозначим: $X(t)$ - число заявок, прибывших в СМО до момента t, $Y(t)$ - число заявок, покинувших СМО до момента t.

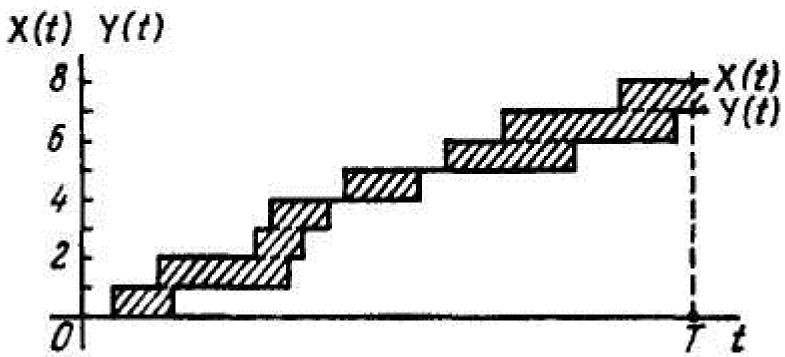


Рис. 4.13

Указанные функции являются случайными и меняются скачком (увеличиваются на единицу) в моменты приходов заявок $X(t)$ и уходов заявок $Y(t)$. Вид функций $X(t)$ и $Y(t)$ показан на рис. 4.13; обе линии - ступенчатые, верхняя - $X(t)$, нижняя - $Y(t)$. Очевидно, что для любого момента t их разность $Z(t)= X(t)-Y(t)$ - есть число заявок, находящихся в СМО. Когда линии $X(t)$ и $Y(t)$ сливаются, в системе нет заявок.

Рассмотрим большой промежуток времени T (мысленно продолжив график далеко за пределы чертежа) и вычислим для него среднее число заявок, находящихся в СМО. Оно будет равно интегралу от функции $Z(t)$ на этом промежутке, деленному на длину интервала T:

$$L_{сист} = \frac{1}{T}\int_0^T Z(t)dt \qquad (4.21)$$

Но этот интеграл представляет собой не что иное, как площадь фигуры, заштрихованной на рис. 4.13. Фигура состоит из прямоугольников, каждый из которых имеет высоту, равную единице, и основание, равное времени пребывания в системе соответствующей заявки (первой, второй и т. д.). Обозначим эти времена как $t_1, t_2, \ldots$ Обратим внимание, что под конец промежутка T некоторые прямоугольники войдут в заштрихованную фигуру не полностью, а частично, но при достаточно большом T это не меняет ситуацию. Таким образом, можно считать, что

$$\int_0^T Z(t)\cdot dt = \sum_i t_i \qquad (4.22)$$

где сумма распространяется на все заявки, пришедшие за время *T*.

Разделим правую и левую часть (4.22) на длину интервала *T.* Получим, с учетом (4.21),

$$L_{сист} = \frac{1}{T}\sum_i t_i \tag{4.23}$$

Разделим и умножим правую часть (4.23) на интенсивность λ:

$$L_{сист} = \frac{1}{T\lambda}\sum_i t_i \cdot \lambda \ .$$

Произведение $T \cdot \lambda$ есть не что иное, как среднее число заявок, пришедших за время *T.* Разделим сумму всех времен t_i на среднее число заявок и получим среднее время пребывания заявки в системе $W_{сист}$. Итак,

$$L_{сист} = \lambda W_{сист} \ ,$$

откуда

$$\boxed{W_{сист} = \frac{1}{\lambda} L_{сист}}. \tag{4.24}$$

Выражение (4.24) есть известная **формула Литтла**: для любой СМО, при любом характере потока заявок, при любом распределении времени обслуживания, при любой дисциплине обслуживания *среднее время пребывания заявки в системе равно среднему числу* заявок в *системе, деленному на интенсивность потока заявок.*

Таким же образом выводится вторая формула Литтла, связывающая среднее время пребывания заявки в очереди $W_{оч}$ и среднее число заявок в очереди $L_{оч}$:

$$W_{оч} = \frac{1}{\lambda} L_{оч} \ . \tag{4.25}$$

Для этого достаточно вместо нижней линии на рис. 4.13 взять функцию *U(t)* - количество заявок, ушедших до момента *t* не из системы, а из очереди (если заявка, пришедшая в систему, не становится в очередь, а сразу идет под обслуживание, можно все же считать, что она становится в очередь, но находится в ней нулевое время). Формулы Литтла (4.24) и (4.25) играют большую роль в теории массового обслуживания.

Рассмотрим некоторые простейшие СМО и получим выражения для их характеристик (показателей эффективности). Все потоки событий, переводящие СМО из состояния в состояние, в данном параграфе будем считать простейшими. В их числе будет и так называемый «поток обслуживании». Под ним понимается поток заявок, обслуживаемых одним непрерывно занятым каналом. В этом потоке интервал между событиями, как и всегда в простейшем потоке, имеет показательное распределение. В данном параграфе показательное распределение времени обслуживания будет рассматриваться применительно к простейшей СМО.

1. Многоканальная СМО с отказами (*задача Эрланга*). Рассмотрим одну из первых по времени, «классических» задач теории массового обслуживания, которая возникла из практических нужд телефонии и была решена в начале XX века датским инженером ***А.К.Эрлангом***. Задача ставится следующим образом.

Имеется n каналов (линий связи), на которые поступает поток заявок с интенсивностью К. Поток обслуживании имеет интенсивность λ (величина, обратная среднему времени обслуживания $t_{об}$*). Найти финальные вероятности состояний СМО, а также следующие характеристики её эффективности*:

A - абсолютную пропускную способность, т. е. среднее число заявок, обслуживаемых в единицу времени;

Q - относительную пропускную способность, т. е. среднюю долю пришедших заявок, обслуживаемых системой;

$P_{отк}$ - вероятность отказа, т.е. вероятность того, что заявка покинет СМО необслуженной;

k - среднее число занятых каналов.

Решение

Состояния СМО будем нумеровать по числу заявок, находящихся в системе (в данном случае оно совпадает с числом занятых каналов): S_0 - в СМО нет ни одной заявки, S_1 - в СМО находится одна заявка (один канал занят, остальные свободны), …, S_k - в СМО находится *k* заявок (*k* каналов заняты, остальные свободны), …, S_n - в СМО находится *n* заявок (все *n* каналов заняты). Граф состояний СМО соответствует схеме гибели и размножения (рис. 4.12). Разметим этот граф - проставим у стрелок интенсивности потоков событий. Из S_o в S_1 систему переводит поток заявок с интенсивностью λ, (как только приходит заявка, система перескакивает из S_0 в S_1). Тот же поток заявок переводит систему из любого левого состояния в соседнее правое (см. верхние стрелки на рис. 4.14).

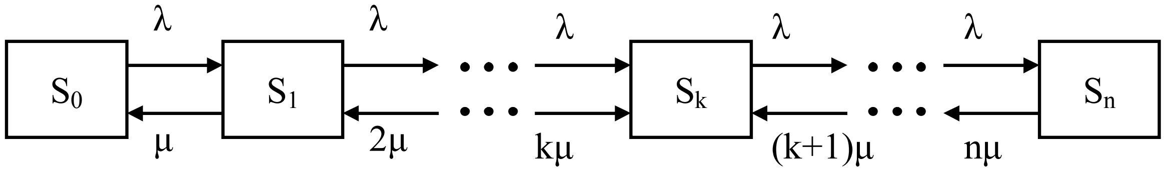


Рис. 4.14

Проставим интенсивности у нижних стрелок. Пусть система находится в состоянии S_1 (работает один канал). Он производит μ обслуживаний в единицу времени. Проставляем у стрелки $S_1 \rightarrow S_0$ интенсивность μ. Теперь представим себе, что система находится в состоянии S_2 (работают два канала). Чтобы ей перейти в S_1, нужно, чтобы либо закончил обслуживание первый канал, либо второй; суммарная интенсивность их потоков обслуживании равна *2*μ*;* проставляем её у соответствующей стрелки. Суммарный поток обслуживаний, создаваемый тремя каналами, имеет интенсивность 3μ, а *k* каналами - *k*μ. Проставляем эти интенсивности у нижних стрелок на рис. 4.14.

Зная все интенсивности, воспользуемся уже готовыми формулами (4.19), (4.20) для финальных вероятностей в схеме гибели и размножения. По формуле (4.20) получим:

$$p_0 = \left(\begin{array}{l} 1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2\mu^2} + \frac{\lambda^3}{2 \cdot 3\mu^3} + \ldots \\ \qquad\qquad \ldots + \frac{\lambda^k}{k!\mu^k} + \ldots + \frac{\lambda^n}{n!\mu^n} \end{array} \right)^{-1} \qquad (4.25)$$

Члены разложения $\frac{\lambda}{\mu}, \frac{\lambda^2}{2\mu^2}, \ldots, \frac{\lambda^n}{n!\mu^n}$ - будут представлять собой коэффициенты при p_0 в выражениях для $p_1, p_2, \ldots, p_n$:

$$p_1 = \frac{\lambda}{\mu} p_0, \; p_2 = \frac{\lambda^2}{2\mu^2} p_0, \ldots, \; p_k = \frac{\lambda^k}{k!\mu^k} p_0, \ldots, p_n = \frac{\lambda^n}{n!\mu^n} p_n. \qquad (4.26)$$

Заметим, что в формулы (4.25), (4.26) интенсивности λ и μ входят не по отдельности, а только в виде отношения $\frac{\lambda}{\mu}$. Введём обозначение

$$\frac{\lambda}{\mu} = \rho \qquad (4.27)$$

и будем называть величину ρ «*приведенной интенсивностью потока заявок*». Смысл этой переменной - среднее число заявок, приходящее за среднее время обслуживания одной заявки. Пользуясь этим обозначением, перепишем формулы (4.25), (4.26) в виде:

$$p_0 = \left(1 + \rho + \frac{\rho^2}{2!} + \ldots + \frac{\rho^k}{k!} + \ldots + \frac{\rho^n}{n!} \right)^{-1}; \qquad (4.28)$$

$$p_1 = \rho p_0, p_2 = \frac{\rho^2}{2!} p_0, \ldots, p_k = \frac{\rho^k}{k!} p_0, \ldots, p_n = \frac{\rho^n}{n!} p_0. \qquad (4.29)$$

Формулы (4.28), (4.29) для финальных вероятностей состояний называются *формулами Эрланга*.

Таким образом, формулы для финальных вероятностей найдены. По ним можно вычислить характеристики эффективности СМО. Сначала найдем $P_{отк}$ - вероятность того, что пришедшая заявка получит отказ (не будет обслужена). Для этого нужно, чтобы все *n* каналов были заняты, значит,

$$P_{отк} = p_n = \frac{\rho^n}{n!} p_0. \qquad (4.30)$$

Отсюда находим относительную пропускную способность - вероятность того, что заявка будет обслужена:

$$Q = 1 - P_{отк} = 1 - \frac{\rho^n}{n!} p_0. \qquad (4.31)$$

Абсолютную пропускную способность получим, умножая интенсивность потока заявок λ на Q:

$$A = \lambda Q = \lambda \left(1 - \frac{\rho^n}{n!} p_0 \right). \qquad (4.31)$$

Осталось найти среднее число занятых каналов $\bar{k}$. Эту величину можно найти как математическое ожидание дискретной случайной величины с возможными значениями 0, 1, ..., *n* и вероятностями этих значений $p_0, p_1, \ldots, p_n$:

$$\bar{k} = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + \ldots + n \cdot p_n.$$

Подставляя сюда выражения (4.29) для p_k (k = 0, 1, ..., n) и выполняя соответствующие преобразования, можно получить формулу для расчёта $\bar{k}$. Воспользуемся другим приёмом. Нам известна абсолютная пропускная способность *А*. Это - интенсивность потока обслуженных системой заявок. Каждый занятый канал в единицу времени обслуживает в среднем μ, заявок. Значит, среднее число занятых каналов равно

$$\bar{k} = A / \mu \tag{4.32}$$

или, учитывая (4.31), получим

$$\bar{k} = \rho\left(1 - \frac{\rho^n}{n!} p_0\right). \tag{4.33}$$

2. Одноканальная СМО с неограниченной очередью. На практике довольно часто встречаются одноканальные СМО с очередью. В ТМО одноканальные СМО с очередью занимают особое место. Применительно к таким СМО получено большинство аналитических формул для немарковских систем.

Пусть имеется одноканальная СМО с очередью, на которую не наложено никаких ограничений (ни по длине очереди, ни по времени ожидания). На эту СМО поступает поток заявок с интенсивностью λ; поток обслуживании имеет интенсивность μ, обратную среднему времени обслуживания заявки $\bar{t}_{об}$. Требуется найти финальные вероятности состояний СМО, а также характеристики ее эффективности:

$L_{сист}$ - среднее число заявок в системе;

$W_{сист}$ - среднее время пребывания заявки в системе;

$L_{оч}$ - среднее число заявок в очереди;

$W_{оч}$ - среднее время пребывания заявки в очереди;

$P_{зан}$ - вероятность того, что канал занят (степень загрузки канала).

Что касается абсолютной *А* и относительной *Q* пропускной способности , то вычислять их нет надобности, поскольку очередь заявок неограниченна, и каждая заявка рано или поздно будет обслужена, поэтому *А* = λ, по той же причине $Q = 1$.

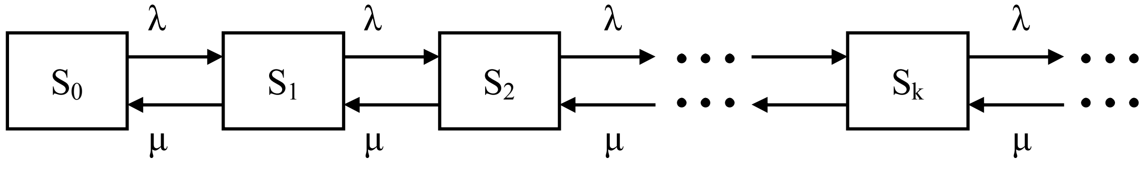


Рис. 4.15

<u>Решение.</u> Состояния системы будем нумеровать по числу заявок, находящихся в СМО: S_0 - канал свободен, S_1 - канал занят (обслуживает заявку), очереди нет, S_2 - канал занят, одна заявка стоит в очереди, …., S_k - канал занят, *k*-1 заявок стоят в очереди, … .

Теоретически число состояний ничем не ограничено (бесконечно). Граф состояний имеет вид, показанный на рис. 4.15. Это - схема гибели и размножения, но с бесконечным числом состояний. По всем стрелкам поток заявок с интенсивностью λ переводит систему слева направо, а справа налево - поток обслуживаний с интенсивностью μ.

Отметим, что финальные вероятности для рассматриваемой СМО существуют не всегда, а только когда система не перегружена. Можно доказать, что если ρ строго меньше единицы (ρ<1), то финальные вероятности существуют, а при ρ ≥ 1 очередь при $t \to \infty$ растет неограниченно. Особенно «непонятным» кажется этот факт при ρ=1. При ρ = 1 СМО справляется с потоком заявок, только если поток этот - регулярен, и время обслуживания - тоже не случайное, равное интервалу между заявками. В этом «идеальном» случае очереди в СМО вообще не будет, канал будет непрерывно занят и будет регулярно выпускать обслуженные заявки. Но стоит только потоку заявок или потоку обслуживании стать случайными - и очередь уже будет расти до бесконечности. На практике этого не происходит только потому, что «бесконечное число заявок в очереди» - абстракция.

Формулы для финальных вероятностей в схеме гибели и размножения выводились для случая конечного числа состояний, однако их можно обобщить и для бесконечного числа состояний. Подсчитаем финальные вероятности состояний по формулам (4.19), (4.20). В нашем случае число слагаемых в формуле (4.20) будет бесконечным. Получим выражение для p_0:

$$p_0 = \left[1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2 + \ldots + \left(\frac{\lambda}{\mu}\right)^k + \ldots\right]^{-1} = \left(1 + \rho + \rho^2 + \ldots + \rho^k + \ldots\right)^{-1}. \qquad (4.34)$$

Ряд в формуле (4.34) представляет собой геометрическую прогрессию. При ρ<1 ряд сходится - это бесконечно убывающая геометрическая прогрессия со знаменателем ρ. При ρ ≥ 1 ряд расходится (что является косвенным, хотя и не строгим доказательством того, что финальные вероятности состояний p_0, $p_1, \ldots, p_k, \ldots$ существуют только при ρ < 1). Теперь предположим, что это условие выполнено, и ρ < 1. Суммируя прогрессию в (4.34), имеем

$$1 + \rho + \rho^2 + \ldots + \rho^k + \ldots = \frac{1}{1-\rho},$$

откуда,

$$p_0 = 1 - \rho. \qquad (4.35)$$

Вероятности $p_1, p_2, \ldots, p_k, \ldots$ вычислим по формулам:

$$p_1 = \rho \cdot p_0,\ p_2 = \rho^2 \cdot p_0, \ldots,\ p_k = \rho^k \cdot p_0, \ldots,$$

откуда с учетом (4.35), найдем окончательно:

$$p_1 = \rho \cdot (1-\rho),\ p_2 = \rho^2 \cdot (1-\rho), \ldots,\ p_k = \rho^k \cdot (1-\rho), \ldots, \qquad (4.36)$$

Как видно, вероятности p_0, $p_1 \ldots$, p_k, ... образуют геометрическую прогрессию со знаменателем ρ. Максимальной является p_0 - вероятность того, что канал будет вообще свободен. Как бы ни была нагружена система с очередью, если она справляется с потоком заявок (ρ <1), то самое вероятное число заявок в системе будет равно 0.

Найдем среднее число заявок в СМО $L_{сист}$. Случайная величина Z - число заявок в системе - имеет возможные значения 0, 1, 2, ..., k, ... с вероятностями $p_0, p_1, p_2, ..., p_k$, ... Её математическое ожидание равно

$$L_c = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + ... + k \cdot p_k + ... = \sum_{k=1}^{\infty} k \cdot p_k \quad . \tag{4.37}$$

Подчеркнем, что сумма берется не от 0 до ∞, а от 1 до ∞, так как нулевой член равен нулю. Подставим в формулу (4.37) выражение для p_k (4.36):

$$L_c = \sum_{k=1}^{\infty} k \cdot \rho^k \cdot (1-\rho).$$

Теперь вынесем за знак суммы общий множитель $\rho(1-\rho)$:

$$L_c = \rho \cdot (1-\rho) \cdot \sum_{k=1}^{\infty} k \cdot \rho^{k-1}$$

Отметим, что $k\rho^{k-1}$ есть производная от выражения ρ^k; следовательно

$$L_c = \rho \cdot (1-\rho) \cdot \sum_{k=1}^{\infty} \frac{d}{d\rho} k \cdot \rho^k .$$

Меняя местами операции дифференцирования и суммирования, получим:

$$L_c = \rho \cdot (1-\rho) \cdot \frac{d}{d\rho} \sum_{k=1}^{\infty} \rho^k . \tag{4.38}$$

Но сумма в формуле (4.38) есть сумма бесконечно убывающей геометрической прогрессии с первым членом ρ и знаменателем ρ; эта сумма равна $\frac{\rho}{1-\rho}$, а ее производная $\frac{1}{(1-\rho)^2}$. Подставляя это выражение в (4.38), получим:

$$L_c = \frac{\rho}{1-\rho}. \tag{4.39}$$

Применим формулу Литтла (4.24), найдем среднее время пребывания заявки в системе:

$$L_c = \frac{\rho}{\lambda \cdot (1-\rho)}. \tag{4.40}$$

Определим среднее число заявок в очереди $L_{оч}$. Число заявок в очереди равно числу заявок в системе минус число заявок, находящихся под обслуживанием. Значит (по правилу сложения математических ожиданий), среднее число заявок в очереди $L_{оч}$ равно среднему числу заявок в системе $L_{сист}$ минус среднее число заявок под обслуживанием. Число заявок под обслуживанием может быть либо нулем (если канал свободен), либо единицей (если он занят). Математическое ожидание такой случайной величины равно вероятности $P_{зан}$ того, что канал занят. Очевидно, $P_{зан}$ равно единице минус вероятность p_0 того, что канал свободен: $P_{зан} = 1 - p_0 = \rho$. Следовательно, среднее число заявок под обслуживанием равно $L_{об} = \rho$, отсюда

$$L_{оч} = L_c - \rho = \frac{\rho}{1-\rho} - \rho$$

и окончательно

$$L_{оч} = \frac{\rho^2}{1-\rho}. \qquad (4.41)$$

По формуле Литтла (4.25) найдем среднее время пребывания заявки в очереди:

$$W_{оч} = \frac{\rho^2}{\lambda \cdot (1-\rho)}. \qquad (4.42)$$

Таким образом, все характеристики эффективности СМО найдены.

3. Многоканальная СМО с неограниченной очередью.

Аналогично задаче 2 решается задача об n-канальной СМО с неограниченной очередью. Нумерация состояний осуществляется по числу заявок, находящихся в системе: S_0 - в СМО заявок нет (все каналы свободны), S_1 - занят один канал, остальные свободны, S_2 - занято два канала, остальные свободны, … , S_k - занято $к$ каналов, остальные свободны, …, S_n - заняты все n каналов (очереди нет), S_{n+1} - заняты все n каналов, одна заявка стоит в очереди, …, S_{n+r} - заняты все n каналов, r заявок стоит в очереди.

Граф состояний показан на рис. 4.16. Предлагаем читателю самому обдумать и обосновать значения интенсивностей, проставленных у стрелок. Граф рис. 4.16 есть схема гибели и размножения, но с бесконечным числом состояний.

Сообщим без доказательства естественное условие существования финальных вероятностей: $\rho/n<1$. Если $\rho/n>1$, очередь растет до бесконечности.

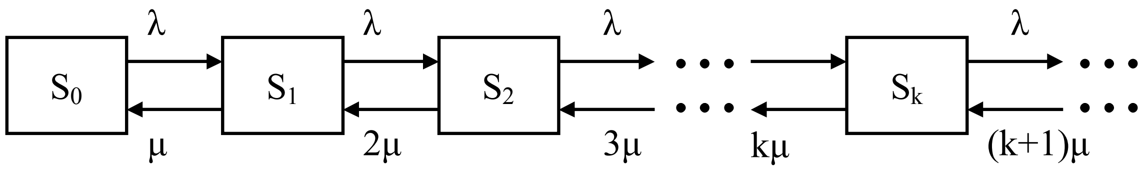


Рис. 4.16

Предположим, что условие $\rho/n<1$ выполнено, и финальные вероятности существуют. Применяя все те же формулы (4.20), (4.19) для схемы гибели и размножения, найдем эти финальные вероятности. В выражении для p_0 будет стоять ряд членов, содержащих факториалы, плюс сумма бесконечно убывающей геометрической прогрессии со знаменателем ρ/n. Суммируя ее, найдем

$$\left.\begin{aligned}
&p_0=(1+\frac{\rho}{1!}+\frac{\rho^3}{2!}+...+\frac{\rho^n}{n!}+\frac{\rho^{n+1}}{n!\cdot(n-\rho)})^{-1},\\
&p_1=\frac{\rho}{1!}\cdot p_0,...,\ p_k=\frac{\rho^k}{k!}\cdot p_0,...,\ p_n=\frac{\rho^n}{n!}\cdot p_0,\\
&p_{n+1}=\frac{\rho^{n+1}}{n\cdot n!}\cdot p_0,...,\ p_{n+r}=\frac{\rho^{n+r}}{n^r\cdot n!}\cdot p_0,....
\end{aligned}\right\} \tag{4.43}$$

Теперь найдем характеристики эффективности СМО. Из них находится среднее число занятых каналов $\bar{k}=\lambda/\mu=\rho$ (это вообще справедливо для любой СМО с неограниченной очередью). Найдем среднее число заявок в системе L_c и среднее число заявок в очереди $L_{оч}$. Из них можно вычислить второе, по формуле $L_{оч}=\sum_{r=1}^{\infty} rp_{n+r}$, выполняя соответствующие преобразования по образцу задачи 2 (с дифференцированием ряда), получим:

$$L_{оч}=\frac{\rho^{n+1}\cdot p_0}{n\cdot n!\cdot(1-\rho/n)^2}. \tag{4.44}$$

Прибавляя к нему среднее число заявок под обслуживанием (или среднее число занятых каналов) $\bar{k}=\rho$, получим:

$$L_c=L_{оч}+\rho. \tag{4.45}$$

Разделив выражения для $L_{оч}$ и $L_с$ на λ, по формуле Литтла получим средние времена пребывания заявки в очереди и в системе:

$$W_{оч}=\frac{L_{оч}}{\lambda},\quad W_c=\frac{L_c}{\lambda}. \tag{4.46}$$

4. Одноканальная СМО с ограниченной очередью.

Задача отличается от задачи 2 только тем, что число заявок в очереди ограничено (не может превосходить некоторого заданного *m).* Если новая заявка приходит в момент, когда все места в очереди заняты, она покидает СМО необслуженной (получает отказ). Надо найти финальные вероятности состояний (кстати, они в этой задаче существуют при любом ρ - ведь число состояний конечно), вероятность отказа $P_{отк}$, абсолютную пропускную способность *A,* вероятность того, что канал занят $P_{зан}$, среднюю длину очереди $L_{оч}$ среднее число заявок в СМО L_c, среднее время ожидания в очереди $W_{оч}$, среднее время пребывания заявки в СМО W_c. При вычислении характеристик очереди можно пользоваться тем же приемом, какой мы применяли в задаче 2, с той разницей, что суммировать надо не бесконечную прогрессию, а конечную.

5. **Замкнутая СМО с одним каналом и *m* источниками заявок.**

Для конкретности поставим задачу в следующей форме: один рабочий обслуживает *m* станков, каждый из которых время от времени требует наладки. Интенсивность потока требований каждого работающего станка равна λ. Если станок вышел из строя в момент, когда рабочий свободен, он сразу же поступает на обслуживание. Если он вышел из строя в момент, когда рабочий занят, он становится в очередь и ждет, пока рабочий освободится. Среднее время наладки станка $\overline{t_{об}} = 1/\mu$. Интенсивность потока заявок, поступающих к рабочему, зависит от того, сколько станков работает. Если работает *k* станков, она равна *kλ*. Найти финальные вероятности состояний, среднее число работающих станков и вероятность того, что рабочий будет занят.

Заметим, что и в этой СМО финальные вероятности будут существовать при любых значениях λ, и $\mu = 1/\overline{t_{об}}$, так как число состояний системы конечно.

4.4 Метод статистических испытаний

4.4.1 Теоретические основания метода статистических испытаний

Наибольшее распространение при исследовании сложных динамических систем получил *метод статистических испытаний* (МСИ). Его называют также *методом Монте-Карло*.

Метод статистических испытаний *заключается* в воспроизведении исследуемого физического процесса при помощи вероятностной ММ и в вычислении характеристик этого процесса. МСИ *основан* на многократном воспроизведении исследуемого процесса (проведении испытаний построенной математической модели) с последующей статистической обработкой полученных данных для получения характеристик рассматриваемого процесса в виде статистических оценок его параметров. Величина рассеивания этих параметров определяет степень приближения решения задачи в вероятностном смысле.

Теоретическую основу МСИ составляют базовые положения теории вероятностей и математической статистики.

1. Предельная теорема теории вероятностей. Когда случайная величина (результаты опыта) определяются суммой большого числа случайных величин (или факторов), то она подчиняется *нормальному закону распределения* (НЗР) вероятностей независимо от закона распределения слагаемых, если последние независимы.

2. Теорема Чебышева устанавливает связь между средним арифметическим конечного числа значений случайной величины и её математическим ожиданием.

При достаточно большом числе независимых опытов среднее арифметическое наблюдаемых значений случайной величины сходится по вероятности к её математическому ожиданию.

3. Теорема Бернулли устанавливает связь между частотой появления события и его вероятностью.

При неограниченном увеличении числа опытов N частота появления события A сходится по вероятности к значению вероятности p

$$\mathrm{P}\left\{\left|p^* - p\right| < \xi\right\} > 1 - \delta\ \},$$

где $p^* \approx k / N$ p^* - частота появления события А в N опытах;

k - число появлений события А в N опытах;

ξ и δ - малые положительные числа.

Метод статистических испытаний обладает следующими **достоинствами**:

- универсальность применения к любым динамическим системам (моделям), в том числе и стохастическим системам;
- относительно простая и гибкая вычислительная схема, удобная для алгоритмизации и программирования;
- устойчивость результата по отношению к возможным ошибкам при реализации отдельных опытов;
- наличие апробированной методики (алгоритмов) оценки точности получаемых результатов;
- универсальность метода, которая заключается в возможности использования для исследования нелинейных систем и процессов, для которых отсутствуют корректные аналитические описания;
- возможность сопряжения МСИ с другими методами вычислительной математики.

Сущность МСИ сводится к введению случайных реализаций входных случайных функций $x_j(t)$ и/или $f_j(t)$ на соответствующие входы исследуемой системы (или ее модели). На каждый из входов системы при одном испытании должна быть подана одна реализация входного возмущения, при этом будет получена реализация каждой из выходных координат. Многократно повторяя подобные испытания, получают для каждой из выходных координат совокупность реализаций. Затем путем статистической обработки полученных результатов определяют законы распределения выходных координат или отдельные характеристики этих законов.

Для воспроизведения и ввода входных возмущений наряду с использованием реальных записей их реализаций применяется физическое или математическое моделирование случайных функций и параметров. С этой целью создано большое число разнообразных датчиков случайных функций и случайных величин, а также программ (датчиков) для получения на ЭВМ последовательности псевдослучайных чисел, на основе которых синтезируются реализации случайных функций. Очевидными достоинствами МСИ являются универсальность и простота. Метод может быть использован применительно к любым нелинейным системам, причем сложность вычислительной схемы метода не зависит от сложности объекта исследования. Метод допускает использование не только математических моделей систем, но также и полунатурных моделей, содержащих отдельные блоки реальной системы. В принципе метод может быть реализован непосредственно на самой системе, если только технически возможны ввод случайных возмущений в систему и измерение её параметров.

Для априорно известной (или ранее разработанной) ММ исследование системы на основе МСИ строится по типовой вычислительной схеме. Упрощённый вариант такой схемы показан на рис. 4.17.

Количество $N_{оп}$ опытов (реализаций) с ММ зависит от заданных требований к точности вычисления оценок характеристик и определяется по правилам математической статистики.

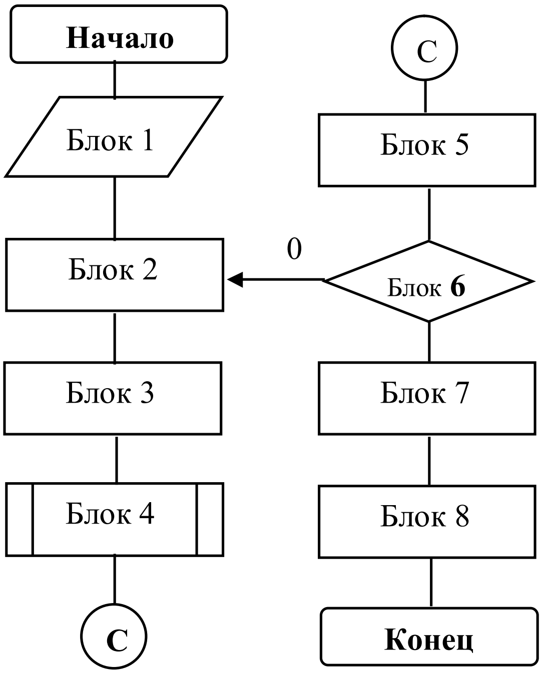


Рис. 4.17 - Вычислительная схема МСИ

Блок 1 – Ввод исходных данных.
Блок 2 – Организация цикла перебора опытов (реализаций) k=1,2,…,N;
Блок 3 – Генерация входных воздействий с помощью ДСЧ и формирующих фильтров;
Блок 4 – Имитация процесса путём решения функциональных уравнений ДС;
Блок 5 – Накопление (фиксация) результатов в каждом опыте;
Блок 6 – Проверка условий окончания цикла перебора опытов: $\mathbf{k \geq N}$ **?**
Блок 7 – Статистическая обработка результатов вычислительного эксперимента;
Блок 8 – Вывод результатов в заданном формате. Построение и аппроксимация гистограмм распределения характеристик системы стандартными ЗРВ.

Цикл (переход «6-2») обеспечивает последовательное проведение статистических опытов с ММ с учетом случайной составляющей заданного комплекса факторов, т.е. воздействий определенного класса или различных условий функционирования ДС. При этом в блоке 1 автоматически (или с помощью исследователя) формируется план моделирования, в котором конкретизируются условия проведения каждого эксперимента.

С окончанием перебора всего множества реализаций осуществляется статистическая обработка результатов статистического моделирования. Производится расчет характеристик и показателей ЭС, которые затем представляются в унифицированной форме (таблицы, графики, гистограммы и т.п.).

Основным недостатком МСИ является необходимость накопления больших массивов информации о выходных координатах системы, что связано с выполнением значительного объема вычислений для получения приемлемой ошибки моделирования. Для того чтобы получить законы распределения выходных координат системы или хотя бы их отдельные характеристики с достаточной для практики точностью, требуется вычислить сотни значений этих координат.

Следует подчеркнуть, что при росте объема статистической выборки наряду с ростом степени уверенности в правильности определения результата всегда будет оставаться и определенная степень риска получения ошибочных данных. Необходимые оценки могут быть выполнены по известным статистическим критериям.

Предположим, что на вход системы в каждом опыте подается реализация $x_i(t)$ случайной величины $X(t)$ и в результате с выхода снимается соответствующая реализация $y_i(t)$ случайной функции $Y(t)$. Выполнив N опытов, можно получить оценки $\hat{m}_y(t)$ и $\hat{D}_y(t)$ математического ожидания $m_y(t)$ и дисперсии $D_y(t)$ выходной переменной системы $Y(t)$, пользуясь известными формулами математической статистики:

$$\hat{m}_y(t) = \frac{1}{N}\sum_{i=1}^{N} y_i(t), \quad D_y(t) = \frac{1}{N-1}\sum_{i=1}^{N}\left[y_i(t) - m_y(t)\right]^2 \tag{4.47}$$

Точность получаемых результатов можно оценивать средними квадратическими отклонениями оценок $\hat{m}_y(t)$ и $\hat{D}_y(t)$. Из теории вероятности известно, что дисперсии оценок определяются формулами

$$D\left[\hat{m}_y(t)\right] = \frac{D_y(t)}{N}, \quad D\left[\hat{D}_y(t)\right] = \frac{2D_y^{\,2}(t)}{N-1}.$$

Соответственно СКО оценок определяются формулами

$$\sigma_{m_y} = \sqrt{\frac{D_y(t)}{N}} = \frac{\sigma_y(t)}{\sqrt{N}}, \quad \sigma_{D_y} = \sqrt{\frac{2}{N-1}} D_y(t). \tag{4.48}$$

В качестве характеристик точности оценок (4.47) можно применять их относительные средние квадратичные отклонения:

$$\frac{\sigma_{m_y}}{\sigma_y(t)} = \frac{1}{\sqrt{N}}, \quad \frac{\sigma_{D_y}}{D_y(t)} = \sqrt{\frac{2}{N-1}}$$

При необходимости оценить абсолютные СКО оценок $\hat{m}_y(t)$ и $\hat{D}_y(t)$ следует заменить в формулах (4.48) неизвестную величину $D_y(t)$ её оценкой $\hat{D}_y(t)$.

Для более полной оценки точности полученных результатов можно вычислить доверительные вероятности для заданных границ оценок $\hat{m}_y(t)$ и $\hat{D}_y(t)$, т.е. вероятности различных отклонений оценок от соответствующих вероятностных характеристик. Имея в виду, что при достаточно большом числе опытов $n \gg 1$ законы распределения оценок $m_y(t)$ и $D_y(t)$ близки к нормальному, на основании известных формул теории вероятностей получим

$$\begin{aligned} \alpha_1 &= P\left[\left|\hat{m}_y(t) - m_y(t)\right| < \varepsilon_1\right] = 2\Phi\left(\frac{\varepsilon_1}{\sigma_{m_y}}\right), \\ \alpha_2 &= P\left[\left|D_y^{*}(t) - D_y(t)\right| < \varepsilon_2\right] = 2\Phi\left(\frac{\varepsilon_2}{\sigma_{D_y}}\right) \end{aligned} \tag{4.49}$$

Здесь ε_1 и ε_2 - отклонения оценок $\hat{m}_y(t)$ и $\hat{D}_y(t)$ от истинных значений;

α_1 и α_2 - соответствующие доверительные вероятности.

Подставляя в (4.49) выражения (4.48) средних квадратических отклонений оценок, получим

$$\alpha_1 = 2\Phi\left(\frac{\varepsilon_1}{\sigma_y(t)}\sqrt{N}\right), \quad \alpha_2 = 2\Phi\left(\frac{\varepsilon_2}{D_y(t)}\sqrt{\frac{N-1}{2}}\right)$$

Вводя относительные отклонения оценок

$$\nu_1 = \frac{\varepsilon_1}{\sigma_y(t)}, \quad \nu_2 = \frac{\varepsilon_2}{D_y(t)},$$

получим окончательно $\alpha_1 = 2\Phi\left(\nu_1\sqrt{N}\right)$, $\alpha_2 = 2\Phi\left(\nu_2\sqrt{\frac{N-1}{2}}\right)$.

В табл. 4.1 и 4.2 приведены зависимости требуемого числа испытаний N от заданных доверительных вероятностей α_1 и α_2 и относительных отклонений ν_1 и ν_2. Относительные отклонения ν_1 и ν_2 определяют точность полученных результатов, а доверительные вероятности α_1 и α_2 - их надежность.

Таблица 4.1. Требуемое число испытаний для определения МОЖ

α_1 \ ν_1	**0,2**	**0,15**	**0,10**	**0,05**	**0,01**
0,5	18	31	70	231	7000
0,7	27	47	108	431	10800
0,8	41	73	164	651	16400
0,9	68	121	272	1090	27200

Таблица 4.2. Требуемое число испытаний для определения дисперсии

α_2 \ ν_2	**0,2**	**0,15**	**0,10**	**0,05**	**0,01**
0,5	37	63	141	563	14000
0,7	55	95	217	863	21600
0,8	83	147	239	1300	32800
0,9	137	243	545	2180	54400

Из таблиц видно, что с ростом требований точности и надежности результатов необходимое число испытаний резко возрастает. При этом повышение требований надежности ведет к менее быстрому росту необходимого числа испытаний, чем повышение требований к точности. Поэтому МСИ может быть практически использован только в тех случаях, когда требуется получить точность порядка (15…20)%. И лишь в единичных случаях, когда требуется получить высокую точность, можно применить этот метод при очень большом числе испытаний. Кроме того, следует учитывать, что с увеличением числа испытаний и в целом времени вычислений растут инструментальные ошибки, которые в некоторых случаях могут привести к нецелесообразности МСИ.

Применение МСИ в задачах исследования операций основано на формировании случайных чисел и их последовательностей в следующих вероятностных схемах:

- моделирование независимых и зависимых испытаний в схеме случайных событий;
- получение случайных чисел с заданным законом распределения.

Получение случайных чисел в любой из вероятностных схем базируется на построении (генерации) в ЭВМ последовательности равномерно распределенных чисел.

При моделировании случайных величин с заданным законом распределения вероятностей используется в основном метод Неймана-Голенко, а также метод замены переменных (метод обратных функций). Однако, указанные методы не позволяют реализовать многие законы распределения за сравнительно небольшое количество машинных операций. Поэтому применяются приближенные способы преобразования случайных чисел, равномерно распределенных в интервале $(0,1)$. Такие преобразования выполняют:

- с помощью линейно-кусочной аппроксимации законов распределения;
- с помощью отбора случайных чисел δ_{ki} из множества $\{\delta_k\}$ таких, которые удовлетворяли бы заданному распределению;
- по методу суперпозиции;
- на основании предельных теорем теории вероятностей.

Результаты наблюдений случайных величин каждого из параметров моделируемого процесса $x_1,\ldots,x_i$ после проведения n реализаций модели образуют эмпирическое распределение. Обработка на ЭВМ этих результатов позволяет получить в процессе моделирования численные характеристики эмпирического распределения (статистические оценки искомых параметров):

- вероятность появления события (непрерывного или дискретного) - p;
- математическое ожидание случайной величины - m_x;
- дисперсию или среднее квадратичное отклонение - σ_x;
- показатели асимметрии S_x и эксцесса E_x распределения;
- корреляционный момент двух случайных величин x и y - k_{xy}.

Для построения гистограммы эмпирического распределения в ЭВМ рассчитываются относительная частота (частость) случайной величины по формуле

$$\overline{p}_i = n_i / N, \quad i = \overline{1,m},$$

где m - выбранное число интервалов (разрядов) диапазона изменения случайной величины x;

n_i - число наблюдений случайной величины x, попавших в i-й разряд.

Указанная формула применима как для непрерывных, так и для дискретных случайных величин. При равенстве всех разрядов гистограмма строится непосредственно по данным об относительной частоте. В противном случае она строится по относительной плотности p_i':

$$p_i' = \frac{n_i}{nh_i},$$

где h_i - величина i-го разряда.

Формулы для оценок основных параметров эмпирического распределения случайной величины x в МСИ на ЭВМ приведены в табл. 4.4.

Оценки $m_x = \overline{x}, \sigma^2, \sigma_1^{\ 2}, k_{xy}, k^1{}_{xy}$ и S_2 являются несмещенными состоятельными оценками теоретического распределения, остальные оценки состоятельные, но несмещенные.

Определение точности статистических оценок искомых параметров, зависящих от числа характеристик моделируемого случайного процесса и от числа реализаций модели на ЭВМ, производится с помощью известных методов. Основными из них являются:

- определение среднего квадратичного отклонения σ_x;
- рассчитанного параметра $\overline{x}$ от теоретически точного x;
- определение доверительного интервала 2ε, полученных оценок $\overline{x}$ на основании заданных доверительных вероятностей α.

Оценка точности частости события, как статистической вероятности его появления, производится, исходя из требуемой надежности оценки α и произведенного числа реализаций N по формуле

$$\xi_p = t_\alpha \frac{S}{\sqrt{N}} = t_\alpha \sqrt{\frac{\overline{p}\left(1-\overline{p}\right)}{N}}, \tag{4.50}$$

где t_α - параметр распределения Стьюдента, значения которого табулированы и зависят от значений α и n (для $\alpha = 0{,}95$ и $n > 40$ $t\alpha = 1{,}96$); α - заданная доверительная вероятность.

Необходимое число реализаций N, обеспечивающее заданную точность и надежность результата, определяется обращенными формулами, которые устанавливают зависимость между числом испытаний n, доверительной вероятностью α и доверительным интервалом 2ε искомого параметра.

Для вычисления числа N на практике применяют следующие формулы:

- при определении математического ожидания:

$$N = t_{\ \alpha} \cdot \frac{\sigma_x}{\xi_{\overline{x}^2}}; \tag{4.51}$$

- при определении вероятности события:

$$N = t_\alpha^{\ 2} \Big/ \xi_{\overline{p}^2}\, \overline{p}\left(1-\overline{p}\right); \tag{4.52}$$

- при распределении дисперсии:

а) в общем случае

$$N = \frac{t_\alpha\left(m_4 - \sigma_x^{\ 4}\right)}{\xi_\sigma} \tag{4.53}$$

б) в случае

$$N = t_\alpha \frac{2\sigma_x^{\ 4}}{\varepsilon_\sigma}\ . \tag{4.54}$$

Поскольку до начала испытаний σ_x и p неизвестны, то сначала целесообразно на ЭВМ осуществить небольшое число реализаций модели $n_0 = 20 - 30$ и получить приближенные значения S^* и p^*. Затем по n_0 и заданному α и t_α. После чего по формулам (4.50) и (4.51) вычислить необходимое значение n.

В табл. 2.1. для доверительной вероятности $\alpha = 0{,}95$ сведены значения необходимого количества реализаций n для различных p и ζ_p (ζ_p - относительная погрешность), полученные при использовании формулы (4.51).

Таблица 4.3

Вероятность		Заданная погрешность		
p	P	$\zeta_p = 0{,}05$	$\zeta_p = 0{,}02$	$\zeta_p = 0{,}01$
0,1	0,9	140	900	3600
0,2	0,8	250	1500	6200
0,3	0,7	330	2100	8400
0,4	0,6	380	2300	9400
0,5		390	2400	9800

Таким образом, основным недостатком МСИ является необходимость проведения большого количества испытаний для обеспечения заданной точности оценок параметров.

Ошибка МСИ при заданной доверительной вероятности уменьшается обратно пропорционально величине $\sqrt{N}$ (N - количество испытаний). С увеличением доверительной вероятности необходимое число опытов также возрастает.

Таблица 4.4 Формулы для расчета статистических характеристик ($n \equiv N$)

№ п/п	Параметр	Аналитическое выражение
1	Математическое ожидание	$m_x = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
2	Дисперсия	$\begin{cases} \sigma_x^2 = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left[x_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right]^2, \text{ при } n > 40, \\ \sigma_x^2 = \frac{n}{n-1}\sigma^2 \quad \text{при } n \le 40 \end{cases}$
3	Среднее квадратическое отклонение (для нормального распределения)	$\sigma = \pm\sqrt{\sigma^2}$ при $n > 40$ $\sigma_1 = \sqrt{\frac{n}{n-1}}\sigma$ при $n \le 40$ $\sigma_2 = \left(1 + \frac{1}{4n} + \frac{9}{32 \cdot n^2}\right)\sigma_1$ при $n \ge 10$

4	Корреляционная функция случайных величин X и Y	$k_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \frac{1}{n^2}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i$ при $n > 40$. $k'_{xy} = \frac{n}{n-1} k_{xy}$ при $n \le 40$
5	Показатель асимметрии	$s_k = \frac{m_3}{\sigma^3} = \frac{1}{n\sigma^3}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^3$
6	Показатель эксцесса	$E_x = \frac{m_4}{\sigma^4} - 3 = \frac{1}{n\sigma^4}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^4 - 3$

Алгоритмы формирования случайных воздействий

На входе и выходе функциональных элементов ЭС наряду с детерминированными действуют и случайные возмущения, которые представляются для имитации последовательностью случайных чисел. Такие выборки случайных чисел, имеющие определенные статистические характеристики, вырабатываются специальной программой, которая носит название датчика псевдослучайных чисел (ДСЧ). Как правило, ДСЧ входит в математическое обеспечение ЭВМ. С помощью рекуррентных вычислительных схем реализовано достаточное количество алгоритмов генерирования псевдослучайных чисел. Псевдослучайными эти числа называются потому, что фактически они, даже пройдя все статистические тесты на случайность и заданный закон распределения, остаются полностью детерминированными. Это значит, что при одних и тех же исходных данных ДСЧ генерирует одинаковые последовательности чисел. Последнее дает возможность проводить испытания различных моделей ЭС при наличии на входах одних и тех же случайных сигналов.

Каждый ДСЧ характеризуется законом распределения генерируемых чисел, а также своей повторяемостью. Последнее объясняется тем, что существующие алгоритмы получения случайных чисел в силу конечности разрядной сетки ЭВМ дают конечную последовательность случайных чисел, период которой определяет число возможных циклов обращения к ДСЧ.

Основой для построения ДСЧ с различными законами распределения вероятностей является датчик с равномерным распределением чисел в диапазоне [0...1]. Напомним, что случайная величина μ называется равномерно распределенной на отрезке [a, b], если ее плотность вероятности постоянна на этом отрезке и равна нулю вне его, т.е.

$$\varphi(\mu) = 1/(b-a) \quad \text{при} \quad a=<\mu<=b \quad \text{и} \quad \varphi(\mu) = 0 \quad \text{при} \quad \mu<a \quad \mu>b.$$

Так как график этой функции изображается в виде прямоугольника, то такое распределение также называют прямоугольным. Числовые характеристики равномерного закона – МОЖ и дисперсия соответственно равны - $(a+b)/2$ и $(b-a)^2/12$.

Отметим важное свойство суммы N независимых равномерно распределенных случайных величин: распределение этой суммы очень быстро (по мере роста числа слагаемых) приближается к нормальному закону. Указанное свойство используется, в частности, при статистическом моделировании нормально распределенных наблюдений.

Механизм формирования нормально распределенных случайных величин заключается в следующем. Значения исследуемой случайной величины формируются под воздействием очень большого числа независимых случайных факторов, причем сила воздействия каждого отдельного фактора мала и не может превалировать среди остальных, а характер воздействия - аддитивный. Функция плотности случайных величин этого соответствует нормальному ЗРВ. ДСЧ с равномерным и нормальным распределением являются базовыми в задачах статистического анализа.

Для моделирования случайных величин **z** с нормальным ЗРВ используется следующее выражение (при M>8):

$$z = \frac{1}{M}\sum_{i=1}^{M}\left[x_i - \frac{1}{2}\right] \quad .$$

К использованию на практике может быть рекомендована другая схема генерации ортонормированных случайных чисел z (для которых $m_z=0$ и $D_z=1$):

$\mathbf{z= 2,506628\ x + 2,624935\ x^3 + 5,772536\ x^5 / (1,0\text{-}3,1\ x^2)}$

x - псевдослучайное число с прямоугольным ЗРВ на отрезке (0,1);

z - псевдослучайное число с нормальным ЗРВ.

Шумовой сигнал, сформированный на основе ДСЧ, генерирующего последовательность чисел с центрированным нормальным распределением ($m_z=0$ и D_z=const), называется "белым" шумом. Длительность элементарных импульсов такого сигнала мала, а амплитуда - распределена по нормальному закону. Спектр "белого" шума можно считать равномерным в широком диапазоне частот. Используя генератор "белого" шума, можно сформировать выборку нормально распределенных чисел с заданными корреляционными свойствами. Например, если пропустить "белый" шум через низкочастотный фильтр (Ф), имеющий передаточную функцию вида

$$W_ф(S)=K_ф/(T_фS+1),$$

то получим выборку чисел с КФ экспоненциального вида. Соответствующая схема моделирования показана на рис.4.18 (ФФ - формирующий фильтр).

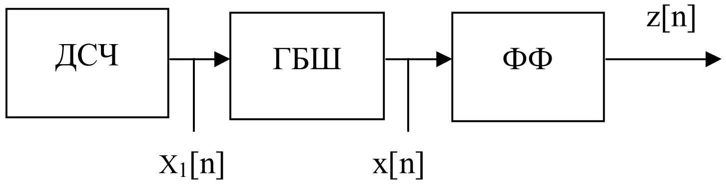


Рис. 4.18

Для моделирования случайных последовательностей, обладающих теми или иными корреляционными свойствами, что характерно для имитации сигналов на входах узлов систем управления, используют эффект прохождения случайных сигналов через линейное звено с заданной импульсной характеристикой. Для образования сигнала в этом случае приходится производить операцию свертки входного сигнала с импульсной характеристикой, что вызывает большие затраты времени. На практике имитацию случайных сигналов осуществляют с помощью рекуррентных (марковских) алгоритмов, согласно которым очередное значение случайной последовательности можно найти, зная несколько его предыдущих значений.

Пусть, например, требуется получить последовательность нормальных случайных чисел с корреляционной функцией (КФ) вида:

$$K(\tau) = \sigma^2 \cdot \exp(-|\tau| \cdot \omega^*) = \sigma^2 \cdot \rho$$

где ω^* - эффективная ширина спектра процесса.

Моделирующий алгоритм описывается формулой $z[n] = a_0 x_1[n] + b_1 z[n]$.
Здесь $a_0 = \sigma \cdot \sqrt{1-\rho^2}$ - множитель входного процесса; b1= ρ - коэффициент обратной связи; $\rho = \exp(-\gamma^*)$ - значение случайной величины, полученное в предыдущем цикле вычислений при этом $\gamma^* = \omega^* \Delta t$; Δt - шаг дискретизации процесса.

Для реализации сложных КФ используется более полный рекуррентный алгоритм, использующий результаты вычислений предыдущих циклов.

При исследовании функционирования ЭС схема применения МСИ состоит в следующем.

1. Определяются исходные данные (детерминированные и случайные), влияющие на функционировании исследуемой системы, а также совокупность параметров, определяющих выходные величины, характеризующие эффективность данной системы.

2. Исследуемая ЭС расчленяется на более простые, самостоятельно функционирующие подсистемы, каждая их которых описывается своими случайными характеристиками и своей группой входных и выходных величин; такая подсистема в общей системе представляет отдельный блок.

3. Разрабатываются ММ каждого блока, а затем всего исследуемого процесса в целом; на основании опыта или имеющейся информации выбираются ЗРВ случайных величин, определяющих простые случайные явления в каждой подсистеме.

4. Разрабатываются алгоритмы получения случайных величин, подчиняющихся ЗРВ, определенным в п.3, по известным формулам.

5. Разрабатываются алгоритмы определения и анализа статистических оценок искомых характеристик исследуемой ЭС в зависимости от цели, стоящей перед системой; при наличии в составе программного обеспечения ЭВМ стандартных программ математической статистики указанные выше действия

(п. 5) сводятся к выбору соответствующих процедур, необходимых для решения данной задачи.

6. Осуществляется композиция частных блоков алгоритма и его формализация на алгоритмическом языке.

7. Обосновывается необходимое число реализаций N при проведении статистических испытаний или экспериментов.

8. Разрабатывается план исследований данного процесса с помощью МСИ на ЭВМ и организуется его осуществление.

9. Осуществляется статистическая обработка и аппроксимация результатов вычислительного эксперимента.

4.5 Контрольные вопросы и задачи

1. В чём заключается свойство потока событий «отсутствие последействия»?
2. Какими свойствами обладает простейший поток событий?
3. Поясните сущность понятия «система массового обслуживания». Приведите примеры СМО в различных сферах экономики.
4. В чём заключается метод статистических испытаний?
5. Поясните механизм записи уравнений Колмогорова для размеченного графа состояний системы.
6. Какое свойство лежит в основе нормировочного условия для вероятностей состояний системы?
7. В чём проявляется универсальность метода статистических испытаний?
8. Каковы недостатки МСИ?
9. Поясните механизм формирования последовательности случайных чисел с заданными корреляционными свойствами.

10. *Задача 4.1***.** Имеется станция связи с тремя каналами (n = 3), интенсивность потока заявок $\lambda = 1{,}5$ (заявки в минуту); среднее время обслуживания одной заявки $\bar{t}_{об} = 2$ (мин.), все потоки событий - простейшие. Найти финальные вероятности состояний и характеристики эффективности СМО: A, Q, $P_{отк}$, $\bar{k}$.

Ответы для контроля решения: $p_0 = 1/13$, $p_1 = 3/13$, $p_2 = 9/26$, $p_3 = 9/26 \approx 0{,}346$, $A \approx 0{,}981$, $Q \approx 0{,}654$, $P_{отк} \approx 0{,}346$, $\bar{k} \approx 1{,}96$.

Из ответов видно, что СМО в значительной мере перегружена: из трех каналов занято в среднем около двух, а из поступающих заявок около 35% остаются необслуженными. Предлагается выяснить: сколько потребуется каналов для того, чтобы удовлетворить не менее 80% поступающих заявок? И какая доля каналов при этом будет простаивать?

ГЛАВА 5. МОДЕЛИ С ЭЛЕМЕНТАМИ НЕОПРЕДЕЛЁННОСТИ

5.1 Задачи принятия решений в условиях неопределённости

В главе 1 рассмотрена обратная задача ИО в детерминированном случае, когда показатель эффективности W зависит только от двух групп факторов: заданных, заранее известных α и элементов решения x. Реальные задачи исследования операций содержат помимо этих двух групп еще одну - неизвестные факторы, которые в совокупности обозначим переменой ξ. Итак, показатель эффективности W зависит от всех трех групп факторов:

$$W = W(\alpha, x, \xi). \tag{5.1}$$

Так как величина W зависит от неизвестных факторов ξ, то даже при заданных α и x она уже не может быть вычислена и остаётся неопределенной. Задача поиска оптимального решения тоже теряет определенность. В явном виде максимизировать неизвестную величину W невозможно.

Изменим *формулировку задачи.* При заданных условиях α с учетом неизвестных факторов ξ требуется найти такое решение $x \in X$, которое, по возможности, обеспечивает максимальное значение показателя эффективности W.

Наличие неопределенных факторов ξ переводит задачу в новое качество: она превращается в *задачу о выборе решения в условиях неопределенности*.

Если условия операции неизвестны, то невозможно оптимизировать решение, чего удаётся добиться при наличии необходимой информации. Отметим, что любое решение, принятое в условиях неопределенности, безусловно, хуже решения, принятого в заранее известных условиях.

Задача принятия решения в условиях неопределенности на каждом шагу встречается нам в жизни. Характерным примером является приятие финансовых решений на фондовой бирже и на конкурсе инновационных проектов.

Для того чтобы управленческие решения в условиях неопределённости принимать обоснованно, современная наука располагает рядом приемов. Каким из них воспользоваться - зависит от того, какова природа неизвестных факторов ξ, откуда они возникают и кем контролируются.

Прежде всего, рассмотрим наиболее благоприятный для исследования вид частичной неопределенности. Это случай, когда неизвестные факторы ξ представляют собой обычные объекты изучения теории вероятностей - случайные величины (или случайные функции), статистические характеристики которых известны или в принципе могут быть получены одним из известных способов. Такие задачи исследования операций будем называть ***стохастическими задачами***, а присущую им неопределенность - стохастической неопределенностью.

Приведем пример стохастической задачи исследования операций. Организуется система технического обслуживания автотранспортных средств личного пользования с целью снижения аварийных отказов и уменьшения простоев техники за счет неисправностей и ремонтов. Отказы подсистем автомобилей отечественного производства, длительности ремонтов и профилактик носят случайный характер. Характеристики всех случайных факторов, входящих в задачу, могут быть получены, если собрать соответствующую статистику.

Рассмотрим более подробно вид стохастической неопределенности. Пусть неизвестные факторы ξ представляют собой случайные величины с какими-то, в принципе известными, вероятностными характеристиками - законами распределения вероятностей, математическими ожиданиями, дисперсиями и т. п. Тогда показатель эффективности W, зависящий от этих факторов, тоже будет величиной случайной. Максимизировать случайную величину невозможно: при любом решении x она остается случайной, неконтролируемой. Поэтому один из приёмов состоит в том, что случайные факторы ξ заменяют средними значениями (математическими ожиданиями). Тогда задача становится детерминированной и может быть решена обычными методами.

В тех случаях, когда влияние случайности на интересующий исход операции является существенным, указанный переход к детерминированной задаче неправомерен.

Рассмотрим такую операцию Q, где факторы ξ существенно случайны и заметно влияют на показатель эффективности W, который представляет собой функционал.

Возникает мысль: надо взять в качестве показателя эффективности среднее значение (математическое ожидание) этой случайной величины $\overline{W} = M[W]$ и выбрать такое решение x, при котором этот усредненный по условиям показатель обращается в максимум:

$$\overline{W} = M[W(\alpha, x, \xi)] \to \max . \tag{5.2}$$

В большинстве случаев такой подход, который назовем «оптимизацией в среднем», вполне оправдан. А как же с элементом неопределенности? Конечно, в какой-то мере он сохраняется. Эффективность каждой отдельной операции, проводимой при конкретных значениях случайных факторов ξ, может сильно отличаться от ожидаемой как в большую, так и в меньшую сторону. Такая «оптимизация в среднем» часто применяется на практике в стохастических задачах исследования операций. Чтобы этот прием был законным, нужно, чтобы операция обладала свойством повторяемости, и недостаток информации о значениях и динамике показателя эффективности можно было бы компенсировать за счёт статистических данных. Например, если мы предпринимаем длинный ряд однородных операций, с целью получить максимальный доход, то доходы от отдельных операций суммируются, «минус» в одном случае покрывается «плюсом» в другом.

Особенно осторожным надо быть с «оптимизацией в среднем», когда речь идет не о повторяемой, массовой операции, а о единичной, «уникальной». Все зависит от того, к каким последствиям может привести неудача данной операции, т.е. слишком малое значение показателя эффективности W; иногда оно может означать попросту катастрофу.

Стохастическая неопределенность - это почти определенность, если только известны вероятностные характеристики входящих в задачу случайных факторов. Гораздо хуже обстоит дело, когда неизвестные факторы ξ не могут быть изучены и описаны статистическими методами. Это бывает в двух случаях: а) распределение вероятностей для параметров ξ в принципе существует, но к моменту принятия решения не может быть получено; б) распределение вероятностей для параметров ξ вообще не существует.

Пример ситуации типа а): проектируется информационно-вычислительная система, предназначенная для обслуживания каких-то случайных потоков требований (запросов). Вероятностные характеристики этих потоков требований в принципе могли бы быть получены из статистики, если бы данная ИВС (или аналогичная ей) уже существовала и функционировала достаточно долгое время. Рекомендуется применить следующий прием: оставить некоторые элементы решения x свободными, изменяемыми. Затем выбрать для начала какой-то вариант решения, зная заведомо, что он не самый лучший, и пустить систему в ход, а потом, по мере накопления опыта, целенаправленно изменять свободные параметры решения, добиваясь того, чтобы эффективность, не уменьшалась, а увеличивалась. Такие совершенствующиеся в процессе применения алгоритмы управления называются адаптивными. Преимущество адаптивных алгоритмов в том, что они не только избавляют от предварительного сбора статистики, но и перестраиваются в ответ на изменение обстановки. По мере накопления опыта такой алгоритм постепенно улучшается.

Обратимся к случаю, когда у неопределенных факторов ξ вообще не существует вероятностных характеристик; другими словами, когда их нельзя считать «случайными» в обычном смысле слова.

Поясним, что под термином «случайное явление» в теории вероятностей принято понимать явление, относящееся к классу повторяемых и, главное, обладающее свойством статистической устойчивости. При повторении однородных опытов, исход которых случаен, их средние характеристики проявляют тенденцию к устойчивости, стабилизируются. Частоты событий приближаются к их вероятностям, средние арифметические - к математическим ожиданиям.

Выделим случай полной неопределенности, когда факторы ξ заранее неизвестны, не имеет смысла говорить об их «законах распределения» или других вероятностных характеристиках.

Пусть выбирается решение x в некоторой операции Q, об условиях которой полностью отсутствуют сведения, можно сделать лишь предположения. Зададим какие-то, более или менее правдоподобные, значения параметров ξ. Тогда задача перейдет в категорию детерминированных и может быть решена обычными методами.

Будет ли найденное решение корректным для других условий ξ? Как правило, нет. Поэтому ценность его - сугубо ограниченная. В данном случае разумно будет выбрать не решение x, оптимальное для каких-то условий ξ, а некое компромиссное решение, которое, не будучи оптимальным, может быть, ни для каких условий, будет все же приемлемым в целом их диапазоне.

В настоящее время полноценной научной теории компромисса не существует, хотя некоторые попытки в этом направлении в теории игр и статистических решений делаются. Обычно окончательный выбор компромиссного решения осуществляется человеком. Опираясь на предварительные расчеты, в ходе которых решается большое число прямых задач исследования операций для разных условий ξ и разных вариантов решения x, он может оценить сильные и слабые стороны каждого варианта и на этой основе сделать выбор. Для этого необязательно знать точный «условный» оптимум для каждой совокупности условий ξ. Математические вариационные методы в данном случае отступают на задний план.

Подчеркнем еще одну полезную функцию предварительных математических расчетов в задачах с полной неопределенностью»: они помогают заранее отбросить те решения $x \in X$, которые при любых условиях ξ уступают другим, т. е. оказываются неконкурентоспособными. В ряде случаев это помогает существенно сузить множество X, иногда - свести его к небольшому числу вариантов, которые легко могут быть просмотрены и оценены человеком в поисках удачного компромисса.

При рассмотрении задач исследования операций с полной неопределенностью полезно сравнивать разные подходы, разные точки зрения. Выделим так называемую позицию крайнего пессимизма. Она сводится к тому, что, принимая решение в условиях полной неопределенности, надо всегда рассчитывать на худшее и принимать то решение, которое дает максимальный эффект в наихудших условиях. Если в этих условиях мы получаем выигрыш $W = \overline{W}$, то можно гарантировать, что в любых других выигрыш будет не меньше. Принцип гарантированного результата привлекателен тем, что дает четкую постановку задачи оптимизации и возможность ее решения корректными математическими методами. Область его применения - по преимуществу так называемые «конфликтные ситуации», в которых условия зависят от сознательно действующего лица («разумного противника»), отвечающего на любое решение наихудшим для нас образом.

В нейтральных ситуациях принцип «гарантированного выигрыша» не является единственно возможным, но может быть рассмотрен наряду с другими. Пользуясь им, нельзя забывать, что эта точка зрения - крайняя, что на ее основе можно выбрать осторожное, «перестраховочное» решение, которое не всегда будет разумным. Можно представить себе, например, руководителя фирмы, который всякое свое решение будет принимать исходя из гипотезы, что его конкурент необычайно умен, хитер и изворотлив и на каждое его действие немедленно ответит наихудшим для него образом. Вряд ли такому руководителю будет сопутствовать удача. Напротив, в любой конкретной ситуации нужно стараться угадать, в чем слаб и некомпетентен конкурент, и стараться «обвести его вокруг пальца». Тем менее уместен крайне пессимистический подход в ситуациях, где стороне, принимающей решение, не противостоят никакие враждебные силы. Расчеты, основанные на точке зрения «крайнего пессимизма», всегда должны корректироваться разумной долей оптимизма.

Известен в теории принятия решений *метод экспертных оценок*. Он часто применяется в задачах, связанных с прогнозированием в условиях полной неопределённости, например при оценке инновационных проектов по нескольким показателям. Идея метода сводится к следующему: собирается коллектив сведущих, компетентных в данной области людей, и каждому из них предлагается ответить на группу вопросов. Затем полученные ответы обрабатываются наподобие статистического материала. Результаты обработки, разумеется, сохраняют субъективный характер, но в гораздо меньшей степени, чем, если бы мнение высказывал один эксперт. Подобного рода экспертные оценки для неизвестных условий могут быть применены при решении задач исследования операций при полной неопределённости. Каждый из экспертов на глаз оценивает степень правдоподобия различных вариантов условий ξ, приписывая им какие-то субъективные вероятности. Несмотря на субъективный характер оценок каждого эксперта, усредняя оценки целого коллектива, можно получить нечто, болев объективное и полезное (кстати, оценки разных экспертов расходятся не так сильно, как можно было бы ожидать). Таким образом, задача сводится к классической стохастической задаче.

Сделаем одно общее замечание. При обосновании решения в условиях неопределенности, что элемент неопределенности сохраняется. Поэтому нельзя предъявлять к точности решений слишком высокие требования. Вместо того чтобы указать одно-единственное, в точности «оптимальное» (с какой-то точки зрения) решение, лучше выделить целую область «приемлемых» решений, которые оказываются несущественно хуже других, какой бы точкой зрения мы ни пользовались. В пределах этой области и должны производить свой окончательный выбор ответственные за это люди. Исследователь, предлагая им рекомендации по выбору решения, всегда должен одновременно указывать точки зрения, из которых вытекают те или другие рекомендации.

4.2 Основные понятия теории игр

Ситуация называется *конфликтной*, если в ней сталкиваются интересы нескольких (обычно двух) лиц, преследующих противоположные цели. Каждая из сторон может проводить ряд мероприятий для достижения своих целей, причем успех одной стороны, как правило, означает неудачу (проигрыш) другой.

Авторами первого фундаментального работы по теории игр *Дж. фон Нейман и О. Моргенштерн* /1944 г./ («Теория игр и экономическое поведение» - М. Наука, 1970). Теория разрабатывалась применительно к анализу конфликтных ситуаций в вопросах экономики, когда при наличии свободной конкуренции в роли борющихся сторон выступают торговые фирмы, промышленные предприятия и т. п.

Укажем примеры конфликтных ситуаций, связанных с принятием решения в условиях неопределенности:

1. Выбор рационального варианта обслуживания клиентов банка;
2. Выбор метода решения задачи анализа (синтеза) в условиях неполноты исходных данных или некорректной постановки задачи;
3. Обоснование рациональной структуры портфеля заказов;
4. Задача выбора оптимального способа презентации;

5. Многоальтернативные выборы в парламент;
6. Арбитражные споры;
7. Спортивные состязания.

Упрощенная формализованная модель конфликтной ситуации называется **игрой**, а конфликтующие стороны - **игроками**. В дальнейшем ограничимся рассмотрением парных игр, в которых участвуют только две конфликтующие стороны.

Следует различать понятия игры и понятие индивидуальной партии игры. Игра представляет собой совокупность правил, описывающих поведение игроков. Каждый случай разыгрывания игры некоторым конкретным образом от начала и до конца представляет собой партию игры. Элементами игры являются ходы. Правила игры предусматривают, какова должна быть последовательность ходов, и указывают характер каждого хода. Ходы бывают личные и случайные. Личный ход представляет собой выбор игроком одного из заданного множества вариантов. Решение, принятое игроком при личном ходе, называется *выбором.*

Случайный ход – это выбор одного из множества вариантов с помощью специального механизма случайного выбора (пример случайных ходов – бросание монеты, сдача карт, бросание кубика с метками). Выбор, осуществляемый при случайном ходе, называется исходом этого хода.

Структура правил игры:

1. Для первого хода указывается: будет это личный или случайный ход.

а) для личного хода: указываются имеющиеся варианты и номер игрока, который делает выбор;

б) для случайного хода: перечисляются имеющиеся варианты и вероятности их выбора.

2. Для каждого следующего хода правила определяют в зависимости от выборов и исходов предыдущих ходов:

а) будет этот ход личным или случайным;

б) для личного хода: выделяется игрок, который делает выбор; возможные варианты, из которых делается выбор; информация относительно выборов и исходов предыдущих ходов.

3. Определяются в зависимости от выборов и исходов, следующих друг за другом ходов (в зависимости от хода игры):

а) условия окончания игры;

б) методику подсчета выигрыша (проигрыша) каждого из игроков.

Стратегия игрока представляет собой однозначное описание его выбора в каждой возможной ситуации, при которой он должен сделать личный ход.

Если игра состоит только из личных ходов, то исход игры будет определен, если каждый из игроков выбрал свою стратегию. Однако если в игре имеются случайные ходы, то игра будет носить вероятностный характер и выбор стратегий игроков еще не определит окончательный исход игры.

Формальное описание игры двух лиц

Пусть рассматривается парная игра. Обозначим через A и B множества или пространства возможных стратегий, которыми могут пользоваться первый и второй игроки.

Величины $a \in A$ и $b \in B$ будут означать конкретные стратегии первого и второго игроков. Для удобства введения в рассмотрение случайных ходов будем считать, что в игре принимает участие третий игрок, который и делает случайные ходы, пользуясь для этого соответствующим механизмом случайного выбора. Обозначим через H пространство стратегий этого игрока.

Любая стратегия $h \in H$ третьего игрока, представляющая собой конкретную последовательность всех случайных ходов в партии, будет происходить с некоторой вероятностью $P(h)$, которую легко подсчитать, зная вероятности каждого случайного хода в этой последовательности.

$P(h)$ представляет собой распределение вероятностей на пространстве H, т.е. удовлетворяет условиям

$$\mathrm{P(h)} \geq 0; \qquad \sum_{h \in H} p(h) = 1.$$

Обозначим через q некоторый вариант игры (одну возможную партию). Этот вариант будет определен, если выбраны стратегии игроков a и b и стратегия случайных ходов h. Следовательно, конкретная партия q представляет собой тройку величин: $q(a, b, h)$.

Результатом партии является выигрыш или проигрыш каждого из игроков. Для удобства выигрыши и проигрыши будем оценивать каким-либо числом - υ.

Рассмотрим одну из конкретных партий $q(a, b, h)$ и обозначим через $L_a(a, b, h)$ и $L_b(a, b, h)$ проигрыши (потери) первого и второго игроков соответственно. При этом выигрыши рассматриваем как отрицательные проигрыши. Общая сумма проигрышей обоих игроков равна:

$$\sum L = L_a(a,b,h) + L_b(a,b,h).$$

В дальнейшем ограничимся рассмотрением игр с нулевой суммой, у которых $\sum L = 0$. В таких играх проигрыш одного игрока равен выигрышу другого игрока.

В играх *с нулевой суммой* можно ограничиться рассмотрением выигрыша первого игрока (проигрыша второго игрока):

$$L_a(a,b,h) = -L_b(a,b,h) = L(a,b,h).$$

Поскольку стратегия h является случайной, то при выбранных стратегиях a и b потери $L(a,b,h)$ будут случайной величиной с распределением вероятности $P(h)$ на пространстве H. Поэтому оценить выбранные стратегии a и b можно лишь путем усреднения потерь $L(a,b,h)$ по всему пространству H, т. е. введя средних потерь $L(a,b)$:

$$L(a,b) = \sum_{h \in H} L(a,b,h) \cdot p(h).$$

Игра будет определена, если перечислены все возможные стратегии игроков, т.е. заданы пространства A и B, и $\forall \ a \in A$ и $b \in B$ определены потери $L(a,b)$.

Таким образом, приходим к следующему формальному определению игры.

Игра G определяется тройкой $G = (A, B, L)$, где A и B – некоторые пространства; L – ограниченная числовая функция, определенная на прямом произведении $A \times B$. Точки $a \in A$ и $b \in B$ – называются стратегиями первого и второго игроков, а функция L называется функцией потерь.

Игры, в которых каждый игрок имеет конечное число стратегий (конечные игры), удобно задавать в виде матрицы потерь (матриц игры, платежной матрицы).

Пусть $G = (A, B, L)$ – некоторая игра, в которой $A = \{a_1, ..., a_m\}$ и $B = \{b_1, ..., b_n\}$. Тогда матрица порядка $m \times n$

$$C = || C_{ij} || = \begin{Vmatrix} c_{11} & c_{12} & ... & c_{1n} \\ ... & ... & ... & ... \\ c_{m1} & c_{m2} & ... & c_{mn} \end{Vmatrix},$$

в которой $C_{ij} = L(a_i, b_j)$ называется *матрицей игры С.*

Для того чтобы описание игры было законченным, необходимо указать цели, которыми руководствуются игроки при выборе стратегии.

Первый игрок стремиться обеспечить себе наибольший выигрыш, т.е. максимизировать функцию $L(a,b)$.

Таким образом, цели игроков оказываются прямо противоположными. Основной трудностью является то, что ни один из игроков не контролирует полностью значение $L(a,b)$, т.к. первый игрок распоряжается только значением a, а второй игрок только значением b.

Преодоление указанной трудности, т.е. определение наиболее рационального способа ведения игры каждым игроком и составляет сущность теории игр.

Рассмотрим пример игры $(m \times n)$, $m = n = 4$.

Таблица 5.1

	b_1	b_2	b_3	b_4	$X(a)$
a_1	$\underline{7}$	2	$\underline{5}$	1	1
a_2	2	2	3	4	2
a_3	5	3	4	4	3^*
a_4	3	2	1	$\underline{6}$	4
$Y(b)$	7	3^*	5	6	

Если первый игрок применяет стратегию a_k, то он обеспечивает для себя гарантированный выигрыш $X(a_k)$, равный наименьшему элементу множества $L(a_к,b)$, $X(a_k) = \min L(a_к,b)$.

В теории игр предполагается, что игроки X и Y действуют достаточно осторожно, избегая необоснованности риска. В этом случае первый игрок X выбрав такую стратегию $a \in A$, которая соответствует максимальному из чисел $X(a)$. Обозначая гарантированный выигрыш первого игрока через α, назовем его нижней ценой игры

$$\alpha = \max_i \min_j L(a,b) = \max_i X(a).$$

Аналогично введем определение верхней чистой цены игры

$$\beta = \min_b Y(b) = \max_i \min_j L(a,b),$$

α - это гарантированный выигрыш первого игрока,

β - это гарантированный проигрыш второго игрока.

Отметим, что всегда выполняется условие $\alpha \le \beta$ для $\forall\ a \in A$ и $b \in B$ в игре $G = (A, B, L)$.

Рассмотрим решение матричной игры в чистых стратегиях. Если каждый из игроков выбирает однозначно с вероятностью 1 некоторую стратегию, то говорят, что он пользуется чистой стратегией. В таком случае решение игры будет в чистых стратегиях. В частности, в чистых стратегиях имеет решение игра с «седловой» точкой.

Рассмотрим парную игру с платежной матрицей $\{\mathbf{C_{ij}}\}$ размером $(\mathbf{m}\times\mathbf{n})$. Поскольку интересы игроков противоположны, то примем: игрок I стремится максимизировать свой выигрыш, а игрок II – наоборот, минимизировать проигрыш. *Требуется найти решение игры*, которое состоит в определении наилучшей стратегии каждым игроком.

Найдем в каждой строке элемент $\boldsymbol{\alpha_i} = \min_j \boldsymbol{C_{ij}}$, $\boldsymbol{i} = \overline{1,\boldsymbol{m}}$. Вычислим $\max_i \boldsymbol{\alpha_i} = \max_i \min_j \boldsymbol{C_{ij}} = \alpha$ называется нижней ценой игры. Нижняя цена игры является гарантированным выигрышем игрока I при любой стратегии игрока II.

Аналогично определяется по каждому столбцу элемент $\beta_j = \max_i \boldsymbol{C_{ij}}$, $\boldsymbol{j} = \overline{1,\boldsymbol{n}}$. Величина $\min_j \beta_j = \min_j \max_i C_{ij} = \beta$ называется верхней ценой игры. Верхняя цена игры является гарантированным проигрышем игрока II при любой стратегии игрока I. Можно показать, что для любой матрицы игры $\|C_{ij}\|$ выполняется равенство $\beta \ge \alpha$.

Если $\alpha = \beta$, то соответствующий элемент матрицы $\boldsymbol{C} = \|\boldsymbol{C_{ij}}\|$, определяемый как $\boldsymbol{a^*_{ij}} = \max_j \min_i \boldsymbol{C_{ij}} = \alpha = \beta$, называется «*седловой*» точкой матрицы $\boldsymbol{C}$.

Седловая точка является минимальным элементом соответствующей строки максимальным элементом соответствующего столбца.

Действительно, точка $\boldsymbol{a^*_{ij}}$ является точкой равновесия, определяя однозначно оптимальные стратегии A_{iopt} и B_{jopt} игроков I и II. При отклонении стратегии игрока I от A_{iopt} его выигрыш будет уменьшаться, и, соответственно, при отклонении стратегии игрока II от B_{jopt} его проигрыш увеличивается. Она определяет средний выигрыш игрока I и средний проигрыш игрока II при использовании ими оптимальных стратегий.

Пример 5.1. Задана матрица игры вида

A_i \ B_j	B_1	B_2	B_3	α_i
A_1	2	1	2	1
A_2	5	2	3	2
A_3	4	0	4	0
β_j	5	2	4	$\max \alpha_i = 2$ / $\min \beta_i = 2$

В данном случае $\max\limits_i \alpha_i = \min\limits_j \beta_j = 2$. Элемент C_{22} – является седловой точкой, а стратегии A_2 - B_2 – оптимальные стратегии игроков A и B. Цена игры равна $\nu = \alpha = \beta = 2$.

Решение игры в смешанных стратегиях

Если игрок выбирает одну из стратегий $\{A_i\}$, $i = \overline{1,m}$ в соответствии с распределением вероятностей $\overline{P} = \{p_i\}$, где p_i - вероятность выбора стратегии A_i, то игрок применяет смешанную стратегию. В этом случае вектор вероятностей $\overline{P} = (p_1,..., p_m)$ определяет смешанную стратегию игрока I.

Аналогично $\overline{Q} = \{q_j\}$, $j = \overline{1,n}$ (q_j - вероятность выбора стратегии B_j игроком II) называется смешанной стратегией игрока II. $\overline{Q} = (q_1,..., q_n)$.

Если $\overline{P}$ и $\overline{Q}$ - соответственно смешанные стратегии игроков I и II, то ожидаемый выигрыш игрока I определяется соотношением:

$$M(\overline{P},\overline{Q}) = P^T \cdot C \cdot Q = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} p_i q_j .$$

$M(\overline{P},\overline{Q})$ - платёжная функция.

Определим условия, при которых игра имеет решение в смешанных стратегиях.

Основная теорема теории матричных игр

Каждая матричная игра с нулевой суммой имеет решение в чистых или в смешанных стратегиях, т. е. существуют такие $\overline{P}_0$ и $\overline{Q}_0$, при которых справедливо неравенство:

$$M(\overline{P},\overline{Q}_0) \le M(\overline{P}_0,\overline{Q}_0) \le M(\overline{P}_0,\overline{Q}).$$

Здесь $(\overline{\boldsymbol{P}}_0,\overline{\boldsymbol{Q}}_0)$ - седловая точка платежной функции $\boldsymbol{M}(\overline{\boldsymbol{P}},\overline{\boldsymbol{Q}})$, а величина $\boldsymbol{M}(\overline{\boldsymbol{P}}_0,\overline{\boldsymbol{Q}}_0) = \nu$ - цена игры.

Седловая точка $(\overline{\boldsymbol{P}}_0,\overline{\boldsymbol{Q}}_0)$ обладает следующим свойством:

$$\boldsymbol{M}(\overline{\boldsymbol{P}}_0,\overline{\boldsymbol{Q}}_0) = \max_{P}\min_{Q}\boldsymbol{M}(\boldsymbol{P},\boldsymbol{Q}) = \min_{Q}\max_{P}\boldsymbol{M}(\boldsymbol{P},\boldsymbol{Q}).$$

Следовательно, она является равновесной точкой игры и определяет оптимальные смешанные стратегии игроков $\overline{\boldsymbol{P}}_0 = \overline{\boldsymbol{P}}_{opt}$, $\overline{\boldsymbol{Q}}_0 = \overline{\boldsymbol{Q}}_{opt}$.

Утверждение, что платежная функция $\boldsymbol{M}(\boldsymbol{P},\boldsymbol{Q})$ действительно имеет седловую точку и поэтому всегда существует решение игры, известно как основная теорема матричных игр. Теорема была сформулирована и доказана ***Дж. фон Нейманом и О. Моргенштерном*** (1944 г.).

Методы теории игр применимы в задачах анализа эффективности элементов комплексов РАВ на этапе проектирования, когда рассматривается парная игра: проектируемая система – активные условия.

Основные особенности применения методов теории игр в задачах проектной эффективности заключается в задании стратегии сторон, в выборе решений (для случая решения игры в смешанных стратегиях), в анализе матрицы игры.

В качестве стратегий со стороны проектируемой системы рассматриваются варианты проектируемого элемента, из которых необходимо выбрать наиболее рациональный.

Пусть A_i, $i=\overline{1,m}$ - варианты проектируемого элемента. В качестве стратегий второй стороны рассматриваются варианты условий A_j, $j=\overline{1,n}$, характеризующиеся параметрами средств противодействия, стратегиями их применения и вариантами реакций на применяемые системой решения. Предположим, что на основе математической модели операции получена матрица игры (платёжная матрица) W(m×n), элементы которой a_{ij} характеризуют эффективность системы при выборе i - го варианта проектируемого элемента и при j- -м варианте условий.

Решение может быть в чистых стратегиях, когда есть седловая точка, условие наличия которой имеет вид $\alpha = \beta$, где $\alpha = \max_i \min_j a_{ij}$ - нижняя цена игры, а $\beta = \min_j \max_i a_{ij}$- верхняя цена игры. В этом случае рациональным будет единственный вариант проектируемого элемента, который может быть принят к конструкторской реализации.

Решение в смешанных стратегиях состоит в реализации чистых стратегий распределениями: для проектируемой системы в виде вектора – столбца $SA = \{p_i\}$, $\sum_{i=1}^{m} p_i = 1$, а для условий – в виде вектора - строки $SA = \{q_j\}$, $\sum_{j=1}^{m} q_j = 1$.

Здесь p_i –вероятность (частота) выбора стратегии A_i; q_i –вероятность (частота) выбора стратегии B_j.

Эффективность системы при использовании смешанных стратегий (платёжная функция) примет вид:

$W(SA,SB) = (SA)^T \cdot W \cdot (SB)^T = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} p_i q_j$ где $(SA)^T$ и $(SB)^T$- транспонированные матрицы SA и SB.

Платёжная функция W(SA,SB) для матричной игры с нулевой суммой, когда выигрыш одной стороны равен проигрышу другой, всегда имеет седловую точку. Это означает, что всегда существует решение в чистых или смешанных стратегиях. Если SA_0 и SB_0 – оптимальные смешанные стратегии, то выполняется соотношение

$$W(SA, SB_0) \le W(SA_0, SB_0) \le W(SA_0, SB).$$

Здесь $W(SA_0,SB_0)$ – цена игры, т.е. отклонение от смешанных стратегий SA_0 и SB_0 для любой из сторон не выгодно.

Если получено решение в смешанных стратегиях, то *возможны следующие случаи* принятия окончательного решения:

1) для дальнейшего проектирования выбирается тот вариант, который гарантирует максимальную эффективность (выбор по максиминной стратегии);

2) выбирается вариант; используемый в смешанной стратегии с максимальной частотой;

3) реализуется несколько вариантов комплектования элемента с частотами, соответствующими смешанной стратегии;

4) реализуются компенсационные возможности элемента (за счёт перераспределения ресурсов), обеспечивающие рациональные параметры элемента в зависимости от конкретного варианта условий.

Для численного решения матричных игр могут быть использованы следующие методы: а) графо-аналитический; б) метод статистических испытаний; в) методы линейного программирования.

5.3 Классические критерии теории максимина

Рассмотрим группу классических критериев (теории максимина) при-нятия решений в условиях неопределенности (без проведения экспериментов).

1. Максиминный **критерий Вальда (крайнего пессимизма)**

Согласно данного критерия выбирается стратегия, для которой *min* выигрыш максимален

$$\boxed{W = \max_i \min_j C_{ij}}.$$

2. Минимаксный **критерий Севиджа**

$$\boxed{S = \min_i \max_j r_{ij}}.$$

Выбирается та стратегия a_i, при которой *min* является величиной риска r_{ij} в самой неблагоприятной ситуации игры с природой.

3. **Критерий** *пессимизма-оптимизма* **Гурвица**. Этот критерий при выборе решения рекомендует руководствоваться некоторым средним результатом, характеризующим состояние между крайним пессимизмом и крайним оптимизмом. Согласно этому критерию стратегия в матрице C выбирается в соответствии со значением

$$H_C = \max_{1\le i\le m}\left\{x\cdot\min_{1\le j\le n} c_{i,j} + (1-x)\cdot\max_{1\le j\le n} c_{i,j}\right\}$$

При $x=0$ критерий Гурвица совпадает с максимаксным критерием, а при $x=1$ - с критерием Вальда. Покажем процедуру применения данного критерия для матрицы C (5.1) при $x=0{,}5$:

- для первой стратегии $(i=1)$ $H_1 = 0{,}5\cdot(\min\limits_{1\le j\le n} c_{i,j} + \max\limits_{1\le j\le n} c_{i,j}) = 5$;
- для второй стратегии $(i=2)$ $H_2 = 0{,}5\cdot(\min\limits_{1\le j\le n} c_{i,j} + \max\limits_{1\le j\le n} c_{i,j}) = 5{,}5$;
- для третьей стратегии $(i=3)$ $H_3 = 0{,}5\cdot(\min\limits_{1\le j\le n} c_{i,j} + \max\limits_{1\le j\le n} c_{i,j}) = 4$;

Тогда $H_C = \max\limits_{1\le i\le m}\{0{,}5\cdot(\min\limits_{1\le j\le n} c_{i,j} + \max\limits_{1\le j\le n} c_{i,j})\} = 5{,}5$; , т. е. оптимальной является вторая стратегия A_2.

Применительно к матрице рисков R критерий пессимизма-оптимизма Гурвица имеет вид: $H_R = \max\limits_{1\le i\le m}\left\{x\cdot\min\limits_{1\le j\le n} r_{i,j} + (1-x)\cdot\max\limits_{1\le j\le n} r_{i,j}\right\}$. При $x=0$ выбор стратегии игрока 1 осуществляется по условию наименьшего из всех возможных рисков ($\min\limits_{i,j} r_{i,j}$); при $x=1$- по критерию минимаксного риска Севиджа.

В случае, когда по принятому критерию рекомендуется к использованию несколько стратегий, выбор между ними может делаться по дополнительному критерию, например в расчет могут приниматься средние квадратичные отклонения от средних выигрышей при каждой стратегии. На практике выбор существенно зависит от склонности ЛПР к риску.

В заключение приведем результаты применения рассмотренных выше критериев на примере следующей матрицы выигрышей:

C=

	B_1	B_2	B_3	B_4
A_1	20	30	15	15
A_2	75	20	35	20
A_3	25	80	25	25
A_4	85	5	45	5

Для **игрока 1** лучшими являются стратегии:

- по критерию Вальда – A_3;
- по критерию Сэвиджа – A_2 и A_3;
- по критерию Гурвица (при $x=0{,}6$) – A_3;

- по критерию максимакса – A_4 .

Поскольку стратегия A_3 , фигурирует в качестве оптимальной по трем критериям выбора из четырех испытанных, степень её надежности можно признать достаточно высокой для того, чтобы рекомендовать эту стратегию к практическому применению.

Таким образом, в случае отсутствия информации о вероятностях состоянии среды теория игр не дает однозначных и математически строгих рекомендаций по выбору критериев принятия решений. Это объясняется в большей мере не слабостью теории, а неопределенностью самой ситуации. Единственный разумный выход в подобных случаях - попытаться получить дополнительную информацию, например, путем проведения исследований или экспериментов. В отсутствие дополнительной информации принимаемые решения теоретически недостаточно обоснованы и в значительной мере субъективны.

Хотя применение математических методов в играх с природой не дает абсолютно достоверного результата и последний в определенной степени является субъективным (вследствие произвольности выбора критерия принятия решения), оно тем не менее создает некоторое упорядочение имеющихся в распоряжении ЛПР данных: задаются множество состояний природы, альтернативные решения, выигрыши и потери при различных сочетаниях состояния «среда - решение». Такое упорядочение представлений о проблеме само по себе способствует повышению качества анализа альтернатив и обоснования принимаемых решений.

5.4 Контрольные вопросы и задачи

1. Раскройте сущность понятий «конфликтная ситуация» и «игра».
2. Приведите примеры конфликтных ситуаций в области малого и среднего бизнеса. Укажите примеры конфликтных ситуаций, в которых принимают участие три стороны.
3. Поясните особенности модели конфликтной ситуации в виде конечной матричной игры?
4. В чём заключается основная теорема матричных игр?
5. Что означает обнаружение «седловой» точки в матричной игре?
6. Какие рекомендации для финансовых агентов на фондовой бирже можно сформулировать с использованием методов теории игр.
7. В чём проявляется гибкость управленческого решения, полученного с помощью критерия Лапласа?
8. Перечислите способы определения значений элементов матрицы эффективности - платёжной матрицы.
9. Для заданной матрицы эффективности игры С (4х4) (см. табл.):
   а) определить нижнюю и верхнюю цену игры;
   б) найти решение игры по критерию Севиджа

С =

	B_1	B_2	B_3	B_4
A_1	20	30	15	15
A_2	75	20	35	20
A_3	25	80	25	25
A_4	85	5	45	5

10. Для заданной матрицы эффективности игры С (4х4) (см. табл.)
    а) найти решение игры по критерию Вальда;
    б) найти решение игры по критерию Гурвица, если коэффициент х={0,25; 0,50; 0,75};
    д) сравнить результаты оценки решений по различным критериям.

	B_1	B_2	B_3	B_4
A_1	10	30	75	55
A_2	20	65	15	45
A_3	15	40	25	60
A_4	90	5	50	35

ЗАКЛЮЧЕНИЕ

В области исследования экономических систем можно выделить ряд проблемных вопросов, начиная от самых общих концептуальных и кончая специальными, типичными для конкретной группы систем и объектов. Современное состояние исследования операций как научной дисциплины, методы которой имеют глубокое теоретическое обоснование, позволяет с общих методологических позиций найти наилучшие решения многих прикладных задач в области количественного анализа экономических систем и процессов.

В настоящее время наибольший научный интерес вызывают вопросы подготовки и принятия компромиссных управленческих решений, представления и решения логистических задач, разработки и исследования операционных моделей конфликтных ситуаций, задачи анализа рисков инновационных проектов, оптимизации портфеля заказов и др.

Исследование операций наряду с другими научными дисциплинами (эконометрика, инвестиционный анализ, моделирование систем, экономический анализ) даёт в руки экономиста (специалиста аналитической службы) мощный инструментарий, который способен облегчить декомпозицию сложных экономических проблем и найти продуктивное решение возникающих исследовательских и практико-ориентированных задач.

Систематическое изучение теоретических основ и приложений методов исследования операций позволит будущим специалистам экономических специальностей взвешенно и профессионально подходить к анализу и решению нестандартных многоплановых задач при развитии инновационной экономики России.

СПИСОК ЛИТЕРАТУРЫ

Основная литература:

1. Афанасьев М.Ю., Суворов Б.П. Исследование операций в экономике. Модели, задачи, решения. - М.: Инфра-М, 2003.- 444 с.

2. Вентцель Е.С. Исследование операций. Задачи, принципы, методология: учебное пособие; 2- изд-е.- М.: Высшая школа, 2001.- 208 с.

3. Замков О.О., Черемных Ю.А., Толстопятенко А.В. Математические методы в экономике: учебник. 2-е изд. Под общ. ред. А.В. Сидоровича.- М.: Изд-во «Дело и Сервис», 1999.- 368 с.

4. Исследование операций в экономике: учебное пособие для вузов; Под ред. Н.Ш. Кремера / Н.Ш. Кремер и др.- М.: ЮНИТИ, 2004.- 407 с.

5. Конюховский П.В. Математические методы исследования операций в экономике.- Санкт-Петербург: Питер, 2002.- 208 с.

6. Мухин В.И. Основы теории управления: учебник для вузов.- М.: Экзамен, 2003.-256 с.

7. Надеждин Е.Н., Бушуев В.Д. Методы моделирования в задачах исследования систем организационного управления: монография. Под ред. Е.Н. Надеждина.- Тула: АНО ВПО «Институт экономики и управления», 2011.- 280 с.

8. Неруш Ю.М. Логистика: учебник. – М.: ТК Велби, Проспект, 2008. - 520 с.

9. Экономико-математические методы и прикладные модели: учебное пособие для вузов / Под ред В.В.Федосеева.- М.: ЮНИТИ, 2000.- 391 с.

10. Экономический анализ: ситуации, тесты, примеры, задачи, выбор оптимальных решений, финансовое прогнозирование: учебное пособие / Под ред. М.И. Баканова и А.Д. Шеремета.- М.: Финансы и статистика, 2001.- 656 с.

Дополнительная литература:

1. Бусленко Н.П. Моделирование сложных систем.- М.: Наука. Гл. ред. физ.-мат. лит., 1978.-400 с.

2. Варфоломеев В.И. Алгоритмическое моделирование элементов экономических систем. Практикум. – М.: Финансы и статистика, 2000.

3. Вентцель Е. С., Овчаров Л.А. Теория случайных процессов и её инженерные приложения. Учебное пособие. 3-е изд., перераб. и доп.- М.: Издательский центр «Академия», 2003, 432 с.

4. Гермейер Ю.Б. Введение в исследование операций.- М.: Наука, 1971.- 384 с.

5. Гмурман В. Б. Теория вероятностей и математическая статистика.— М.: Высшая школа, 2000.

6. Гнеденко Б. В., Коваленко П. Н. Введение в теорию массового обслуживания.- М.: Наука, 1966.

7. Дьяконов В.П. Mathcad 2000: Учебный курс.– СПб.: Питер, 2001.–

592 с.

8. Козловский В.А., Маркина Т.В., Марков В.М. Производственный и оперативный менеджмент. - М, СПб.- 1998.

9. Косоруков О.А., Мищенко А.В. Исследование операций. Учебник для вузов.- М.: Экзамен, 2003.- 448 с.

10. Кузин Л.Т. Основы кибернетики. Т. 2.: Основы кибернетических моделей.- М.: Энергия, 1979.

11. Мину М. Математическое программирование: Теория и алгоритмы.- М.: Наука, 1990.- 488 с.

12. Надеждин Е.Н. Моделирование систем управления ракетно-артиллерийского вооружения: учебное пособие.- Тула: Изд-во Тульского ВАИУ, 1995. - 167 с.

13. Надеждин Е.Н. Математические основы проектирования роботизированных образцов ракетно-артиллерийского вооружения. Ч. I: Методы декомпозиции и принятия решений: учебное пособие.- Тула: Изд-во Тульского ВАИУ, 1997. - 161 с.

14. Овчаров Л. А. Прикладные задачи теории массового обслуживания.- М.: Машиностроение, 1969.

15. Орлова И.В. Экономико-математическое моделирование.: практическое пособие по решению задач.- М.: вузовский учебник, 2004.- 144 с.

16. Саати Т. Л. Элементы теории массового обслуживания и ее приложения.- М.: Советское радио, 1971.

17. Советов Б.Я., Яковлев С.А. Моделирование систем: учебник для вузов. - М.: Высшая школа, 1985.- 271 с.

18. Таха Х. Введение в исследование операций: в 2-х книгах. Кн. 2. Пер. с англ. – М.: Мир, 1985.- 496 с.

19. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере. / Под ред. Ф.Э. Фигурнова.- М.: ИНФАМ , 1998.- 528 с.

20. Хазанова Л.Э. Математические методы в экономике: учебное пособие.- 2-е изд., испр. и перераб..- М.: Изд-во БЕК, 2002.- 144 с.

ПРИЛОЖЕНИЕ 1: ГЛОССАРИЙ

№ п/п	Новые понятия	Содержание
1	**Выборка**	- некоторое количество наблюдений, отобранных из генеральной совокупности.
2	**Задача**	- *в теории решений* – логическое высказывание вида «Дано В, требуется С», где В – исходное состояние объекта управления, С – целевое состояние; - *в хозяйственно-управленческой практике* – задание, поручение, даваемое руководителем работнику аппарата управления или вышестоящим органом управления подчинённой организации или предприятию для исполнения.
3	**Значение оценки**	- число, полученное в результате применения оценки к конкретной выборке.
4	**Исследование операций**	- применение математических, количественных методов для обоснования решений во всех областях целенаправленной человеческой деятельности (*Е.С. Вентцель*); - применение математических методов для моделирования систем и анализа их характеристик (*Таха Х.*)
5	**Математическая модель**	- количественное отображение на принятом формальном языке в рамках заданных допущений наиболее существенных свойств и отношений экономической системы; - совокупность переменных и связей между ними в форме уравнений, описывающих зависимость между наблюдаемыми переменными.
6	**Метод**	- (от *греч. путь*) - систематизированная совокупность шагов, действий, которые необходимо предпринять, чтобы решить определённую задачу или достичь определённой цели.
7	**Метод наименьших квадратов (МНК)**	- метод нахождения оценок параметров регрессии, основанный на минимизации суммы квадратов остатков всех наблюдений $\sum q_i^2$.
8	**Метод Монте-Карло**	- численный метод, применяемый для моделирования случайных величин и функций, вероятностные характеристики которых совпадают с решениями аналитических задач.
9	**Методика**	- совокупность методов обучения чему-нибудь, практическое выполнение чего-нибудь; последовательность действий, операций или алгоритмов, отражающих специфику предметной области и направленных на практическое (эффективное)

		применение одного или нескольких методов к решению определённого класса задач.
10	**Моделирование**	- процесс исследования реальной системы, включающий построение модели, целенаправленное изучение её свойств и перенос полученных знаний на исследуемую систему.
11	**Наблюдение**	- наблюдаемое значение случайной величины или набора случайных величин.
12	**Несмещённая оценка**	- оценка, имеющая нулевое смещение.
13	**Операция**	- последовательность действий, объединённых общим замыслом и направленных на достижение конкретного результата.
14	**Проблема**	- несоответствие между существующим и требуемым (целевым) состоянием системы при данном состоянии среды в рассматриваемый момент времени.
15	**Решение**	- определённый выбор зависящих от организатора параметров.
16	**Система**	- целостная совокупность взаимосвязанных частей или подсистем, взаимодействующая с внешней средой и, как правило, предназначенная для достижения определённой цели (целей).
17	**Системный анализ**	- методология количественного обоснования решений, основанная на структуризации систем и количественном сравнении альтернатив с использованием принципов системного подхода и методов теории принятия решений.
18	**Ситуация**	- в теории решений – логическое высказывание вида «дано В», где В - исходные условия, а цель явно не определена.
19	**Смещение оценки**	- разность между МОЖ оценки и истинным значением оцениваемого параметра.
20	**Состоятельная оценка**	- оценка, у которой смещение и дисперсия стремятся к нулю при увеличении объёма выборки.
21	**Управление**	- процесс формирования целенаправленного поведения системы посредством информационных воздействий, вырабатываемых человеком (группой людей) или устройством.
22	**Цель**	- ситуация или область ситуаций, которая должна быть достигнута при функционировании системы за определенный промежуток времени.
23	**Экономико-математическая модель**	- достаточно точное описание исследуемого экономического процесса или объекта с помощью математического аппарата.
24	**Эффективная оценка**	- несмещённая оценка, имеющая наименьшую дисперсию среди всех несмещённых оценок.

ПРИЛОЖЕНИЕ 2: ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

ЗАДАНИЕ 1. Найти решение ЗЛП **(варианты 1-22)** графическим методом с определением максимального и минимального значения целевой функции $F(x_1, x_2)$, где x_1 и x_2 – оптимизируемые параметры, с учётом системы ограничений на переменные x_1 и x_2. Построить и проанализировать графическое изображение области допустимых значений (ОДЗ) ЦФ в системе координат $\mathbf{x_2 0 x_1}$. Указать условия, при которых на практике может быть использован графический метод решения ЗЛП. Результаты вычислений проверить в вычислительной среде Mathcad с использованием стандартных процедур оптимизации **Maximize(F,x1,x2)** и **Minimize(F,x1,x2)** (см. задание 11, с.165, рис. 2.38) или в офисной программе EXCEL.

В отчёте о самостоятельной работе ***отразить*** а) формулировку ЗЛП; б) график ОДЗ изменяемых параметров; в) результаты решения ЗЛП и комментарии к ним; г) фрагмент компьютерной программы; д) выводы.

Таблица

Номер варианта	ЦФ $F(x_1, x_2)$	Ограничения
1	$4x_1 + 3x_2$	$4x_1+2x_2 \le 15$; $2{,}5x_1+3x_2 \ge 10$; $x_1+x_2 \le 10$; $x_1 \ge 2$; $x_2 \ge 1$.
2	$1 + 2x_1 + 3x_2$	$4x_1+2x_2 \le 15$; $2{,}5x_1+3x_2 \ge 10$; $x_1+x_2 \le 8$; $x_1 \ge 2$; $x_2 \ge 2$.
3	$1 + 3x_1 + 2x_2$	$3x_1+2x_2 \le 21$; $1{,}5x_1+3x_2 \ge 10$; $x_1+x_2 \le 9$; $x_1 \ge 1$; $x_2 \ge 1$.
4	$3x_1 + 5x_2$	$x_1+1{,}5x_2 \ge 5$; $3x_1+x_2 \le 12$; $2x_1+3x_2 \le 12$; $x_1, x_2 \ge 0$; $x_2 \le 5$.
5	$2 + 4x_1 -3x_2$	$x_1+2x_2 \ge 2$; $2x_1+x_2 \le 10$; $2x_1-x_2 \le 1$; $x_1, x_2 \ge 0$.
6	$2x_1 + x_2$	$x_1+x_2 \ge 3$; $2x_1+3x_2 \le 15$; $3x_1+2x_2 \le 10$; $x_1 \ge 1$; $0 \le x_2 \le 4$
7	$5 - 2x_1 + 4x_2$	$x_1+2x_2 \le 12$; $-3x_1+2x_2 \le 8$; $x_1+3x_2 \ge 6$; $x_1, x_2 \ge 1$.
8	$5x_1 + 10x_2$	$2x_1+x_2 \le 10$; $x_1+2x_2 \ge 6$; $x_1 \ge 1$; $2x_2 \ge 3$.
9	$5 + 3x_1 - 2x_2$	$-x_1+2x_2 \le 5$; $7x_1+5x_2 \le 35$; $x_1+x_2 \ge 3$; $x_1 \ge 0$; $x_2 \ge 0$.
10	$2x_1 - x_2 + 1$	$x_1+x_2 \ge 6$; $3x_1+2x_2 \le 20$; $x_1 \ge 0$; $x_2 \ge 1$.
11	$x_1 + 3x_2$;	$3x_1+5x_2 \le 20$; $2x_1+x_2 \le 12$; $x_1- x_2 \ge 1$; $x_1 \ge 0$; $x_2 \ge 0$.
12	$5x_1 + 5x_2$	$x_1+2x_2 \le 10$; $2x_1+2x_2 \ge 5$; $x_1 \ge 1$; $x_2 \ge 1$.
13	$3 + x_1 +10x_2$	$2x_1+5x_2 \le 20$; $x_1+2x_2 \ge 5$; $x_1 \ge 1$; $x_2 \ge 1$.
14	$2 - 2x_1 + 5x_2$	$7x_1+2x_2 \ge 12$; $x_1+5x_2 \le 30$; $x_1+x_2 \le 8$; $x_1 \ge 1$; $x_2 \ge 1$.
15	$2x_1 + 2x_2$	$2x_1+5x_2 \le 10$; $3x_1+x_2 \ge 5$; $x_1+ x_2 \le 6$; $x_1 \ge 0$; $x_2 \ge 1$.
16	$5x_1 + 3x_2$	$3x_1+5x_2 \le 15$; $5x_1+2x_2 \le 10$; $x_1+x_2 \ge 3$; $x_1 \ge 0$; $x_2 \ge 0$.
17	$1-2x_1 + 5x_2$	$3x_1+2x_2 \ge 12$; $x_1+3x_2 \le 30$; $x_1+x_2 \le 8$; $x_1 \ge 1$; $x_2 \ge 0$.
18	$2x_1 + 3x_2$	$3x_1+3x_2 \le 21$; $x_1+x_2 \ge 1$; $x_1+2x_2 \le 8$; $x_1 \ge 0$; $x_2 \ge 0$.
19	$2 + 2x_1 + 2x_2$	$3x_1+3x_2 \le 15$; $x_1+4x_2 \ge 4$; $x_1 \ge 0$; $x_2 \ge 0$.
20	$3x_1 + 5x_2$	$x_1+3x_2 \le 12$; $2x_1+5x_2 \ge 15$; $2x_1 \ge 1$; $x_2 \ge 1$.
21	$10x_1 + 5x_2$	$2x_1+x_2 \le 10$; $x_1+x_2 \ge 2$; $x_1+2x_2 \le 10$; $x_1 \ge 1$; $x_2 \ge 1$.
22	$3+x_1 + x_2$	$x_1+2x_2 \le 10$; $x_1+x_2 \ge 2$; $2x_1+x_2 \le 12$; $x_1 \ge 0$; $x_2 \ge 0$.
23	$10x_1 + 5x_2$	$2x_1+x_2 \le 10$; $x_1+2x_2 \ge 3$; $2x_1+x_2 \ge 3$; $x_1 \ge 1$; $x_2 \ge 1$.
24	$4x_1 + 3x_2 + 2x_3$	$4x_1+3{,}4x_2+2x_3 \le 390$; $2x_1+x_2+3x_3 \le 100$; $x_1 \ge 10$; $x_2 \ge 10$; $x_3 \ge 10$
25	$2x_1 + 4x_2 + 6x_3$	$3x_1+3x_2 \le 25$; $x_1+2x_2 \ge 5$; $2x_1+x_3 \le 12$; $x_1 \ge 2$; $x_2 \ge 2$; $x_3 \ge 2$.

ЗАДАНИЕ 2. Необходимо с использованием рекомендаций теории марковских цепей с непрерывным временем построить и проанализировать вероятностную модель экономической системы с конечным числом состояний и непрерывным временем перехода. Размеченный граф переходов представлен на **рис.1 и рис.2**.

Деятельность ЭС рассматривается на интервале времени $t \in [0;\ 90]$ суток. Интенсивности переходов $\lambda_{i,j}$ из состояния S_i в состояние S_j указаны **в табл. 1.** (**варианты 1…20**) и **в табл. 2.** (**варианты 21…40**).

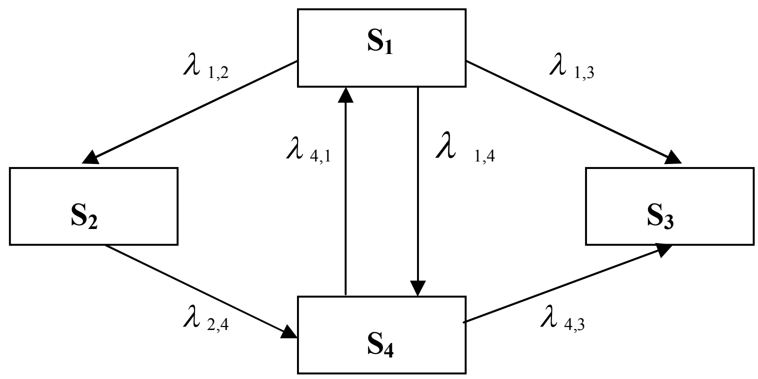


Рис.1 - Размеченный граф переходов ЭС

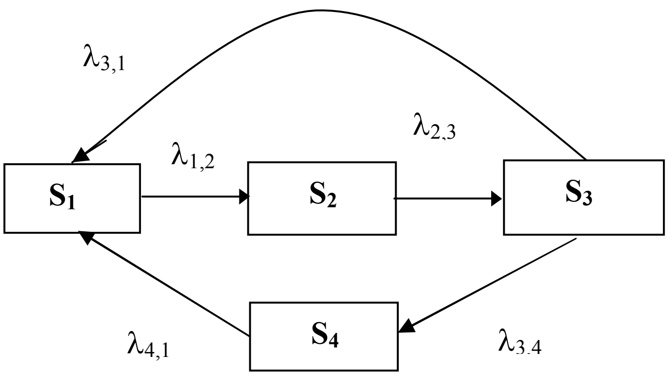


Рис.2 - Размеченный граф переходов ЭС

Начальные условия: $P_1(t_0)=1$; $P_i(t_0)=0$, $i = 2,\ldots,4$.

Порядок выполнения работы:

1. Получить в аналитическом виде систему линейных дифференциальных уравнений Колмогорова, соответствующих размеченному графу переходов.
2. Решить аналитически систему уравнений и определить предельные значения вероятностей состояний ЭС для установившегося режима.
3. Составить программу (в среде Mathcad) численного решения системы уравнений Колмогорова методом Рунге-Кутта с фиксированным шагом (шаг интегрирования – (0,02…0,20) суток – выбрать самостоятельно) на заданном интервале времени ΔT . Дополнительные данные для моделирования получить у преподавателя.
4. Решить (по вариантам) систему дифференциальных уравнений относительно вероятностей состояний $P_i(t)$ с учётом заданных значений интенсивностей перехода при известных начальных условиях (НУ) $P_i(t_0)$, $i = 1,\ldots,4$. Для решения использовать стандартный метод Рунге-Кутта с фиксированным шагом.
5. Построить графики изменения вероятностей $P_i(t)$, $i=1,\ldots,4$ на заданном временном интервале. Определить наиболее вероятное состояние ЭС в заданный момент времени на интервале ΔT =90 суток.
6. Провести анализ результатов решения и сделать выводы.
7. Привести пример ЭС, для которой можно выделить 4 функциональных состояния (раскрыть их физический смысл), дать характеристику деятельности ЭС.

Таблица 1: Интенсивности переходов ($1/\,сут$)

Варианты	$\lambda_{1,2}$	$\lambda_{2,4}$	$\lambda_{1,4}$	$\lambda_{4,1}$	$\lambda_{4,3}$	$\lambda_{1,3}$
1	2,0	0,5	3,0	0,25	1,0	2,0
2	3,0	0,25	3,0	0,25	1,1	0,3
3	4,0	0,35	3,0	0,25	0,3	0,4
4	2,5	0,45	3,0	0,25	0,4	0,35
5	1,5	0,55	3,0	0,25	0,5	0,5
6	2,0	0,65	3,0	0,25	0,6	0,7
7	2,0	0,75	3,0	0,25	0,7	0,6
8	2,5	0,85	3,0	0,25	0,8	0,8
9	3,0	0,95	3,0	0,25	0,9	0,9
10	1,5	0,5	3,0	0,25	1,0	1,0
11	2,0	0,45	2,5	0,25	1,0	1,1
12	2,2	0,45	2,5	0,28	1,2	0,5
13	1,5	0,45	2,5	0,28	1,0	1,3
14	1,5	0,45	2,0	0,25	1,0	1,5
15	2,5	0,4	2,5	0,28	1,5	1,0
16	2,2	0,65	2,5	0,35	1,2	0,5
17	1,8	0,75	3,5	0,25	1,5	1,0
18	1,5	0,5	3,0	0,05	1,0	1,5
19	1,8	0,5	2,5	0,25	1,4	1,1
20	1,5	0,65	3,2	0,15	1,2	1,2

Таблица 2: Интенсивности переходов ($1/\,сут$)

Варианты	$\lambda_{1,2}$	$\lambda_{2,3}$	$\lambda_{3,1}$	$\lambda_{4,1}$	$\lambda_{3,4}$
19	2,0	0,5	3,0	0,25	1,0
20	3,0	0,25	3,0	0,25	1,1
21	4,0	0,35	3,0	0,25	0,3
22	2,5	0,45	3,0	0,25	0,4
23	1,5	0,55	3,0	0,25	0,5
24	2,0	0,65	3,0	0,25	0,6
25	2,0	0,75	3,0	0,25	0,7
26	2,5	0,85	3,0	0,25	0,8
27	3,0	0,95	3,0	0,25	0,9
28	1,5	0,5	3,4	0,25	1,0
30	2,0	0,45	2,5	0,25	1,0
31	2,2	0,45	2,5	0,28	1,2
32	1,5	0,45	2,5	0,28	1,5
33	1,5	0,45	2,0	0,25	1,2
34	2,5	0,4	2,5	0,28	1,5
35	1,5	0,5	3,0	0,25	1,0
36	2,1	0,55	2,1	0,45	0,9
37	2,3	0,25	2,5	0,65	1,6
38	2,0	0,45	2,0	0,25	1,2
39	2,5	0,25	2,5	0,65	1,8
40	2,0	0,45	2,0	0,25	1,2